



**Federal Aviation
Administration**

DOT/FAA/AM-06/14
Office of Aerospace Medicine
Washington, DC 20591

A Rapid and Inexpensive PCR-Based STR Genotyping Method for Identifying Forensic Specimens

Doris M. Kupfer¹
Mark Huggins²
Brandt Cassidy³
Nicole Vu³
Dennis Burian¹
Dennis V. Canfield¹

¹Civil Aerospace Medical Institute
P.O. Box 25082
Oklahoma City, OK 73125

²Advancia
655 Research Parkway
Oklahoma City, OK 73104

³DNA Solutions
840 Research Parkway
Oklahoma City, OK 73104

June 2006

Final Report

NOTICE

This document is disseminated under the sponsorship of the U.S. Department of Transportation in the interest of information exchange. The United States Government assumes no liability for the contents thereof.

This publication and all Office of Aerospace Medicine technical reports are available in full-text from the Civil Aerospace Medical Institute's publications

Web site:

www.faa.gov/library/reports/medical/oamtechreports/index.cfm

Technical Report Documentation Page

1. Report No. DOT/FAA/AM-06/14		2. Government Accession No.		3. Recipient's Catalog No.	
4. Title and Subtitle A Rapid and Inexpensive PCR-Based STR Genotyping Method for Identifying Forensic Specimens				5. Report Date June 2006	
				6. Performing Organization Code	
7. Author(s) Doris M. Kupfer, ¹ Mark Huggins, ² Brandt Cassidy, ³ Nicole Vu, ³ Dennis Burian, ¹ and Dennis V. Canfield ¹				8. Performing Organization Report No.	
9. Performing Organization Name and Address ¹ FAA Civil Aerospace Medical Institute ² Advancia P.O. Box 25082 655 Research Parkway Oklahoma City, OK 73125 Oklahoma City, OK 73104				10. Work Unit No. (TRAIS)	
				11. Contract or Grant No.	
12. Sponsoring Agency name and Address Office of Aerospace Medicine Federal Aviation Administration 800 Independence Ave., S.W. Washington, DC 20591				13. Type of Report and Period Covered	
				14. Sponsoring Agency Code	
15. Supplemental Notes					
16. Abstract Situations arise where forensic samples can be inadvertently misidentified during field collection. Samples identified in the field as being from the same person that return conflicting toxicological results may suggest such misidentification. Polymerase chain reaction (PCR)-based human identity testing provides a reliable and independent method to confirm sample identification. In the study presented here, the Federal Bureau of Investigation's human Combined DNA Identity System (CODIS) loci were used on a rapid, inexpensive microfluidics chip electrophoresis platform to confirm the identity of forensic samples from an aircraft accident site. Absolute allele identification was not achieved with this method, but it was found to be suitable for comparative analysis, as demonstrated by validation of the results and conclusions from capillary electrophoresis. The nine CODIS Short Tandem Repeat loci and a gender discrimination locus used in this study have a greater than 5E+07 matching probability suitable for small sample sizes.					
17. Key Words Forensic Science, Aerospace Medicine, DNA Typing, Short Tandem Repeats, Microfluidics Chip				18. Distribution Statement Document is available to the public through the Defense Technical Information Center, Ft. Belvoir, VA 22060; and the National Technical Information Service, Springfield, VA 22161	
19. Security Classif. (of this report) Unclassified		20. Security Classif. (of this page) Unclassified		21. No. of Pages 18	22. Price

ACKNOWLEDGMENTS

We thank Dr. Marco Scarpetta, Orchid Cellmark, for his comments and help with power of discrimination calculations.

A RAPID AND INEXPENSIVE PCR-BASED STR GENOTYPING METHOD FOR IDENTIFYING FORENSIC SPECIMENS

INTRODUCTION

The possibility exists that samples recovered from a fatal accident site can be misidentified. Results from routine toxicological or other biochemical testing of samples that conflict with the preliminary identification suggest this possibility but are not definitive. Genotyping, a DNA-based analysis, which is not a standard part of sample evaluation, can serve as an additional and independent test for sample identification. Capillary electrophoresis is the method often used for analysis of the polymerase chain reaction (PCR) products generated in genotyping tests (1). This method requires expensive, dedicated instrumentation, software for analysis, and trained personnel potentially available only in specialized facilities. There is a need for a more accessible and inexpensive human identity testing method in instances where genotyping is necessary only as a confirmatory test and there are a minimal number of samples.

Short tandem repeat (STR) loci are polymorphic regions found in the genome that are used as genetic markers for human identity testing. Typing of STR loci by polymerase chain reaction (PCR) is becoming a standard for nuclear DNA genotyping analysis. The FBI's Combined DNA Index System (CODIS) has defined a core set of STR markers for human identity testing (2) which are highly discriminatory over a wide range of ethnicities. Robust primer sequences for these loci have been made publicly available (3) and are available online through the National Institute of Standards and Testing (NIST) Website, STRbase (www.cstl.nist.gov/div831/strbase/) (4). The probability of samples from different subjects having identical alleles for all of the CODIS loci is very small, leading to the use of these loci in forensic evaluations (1, 5).

Results will be presented here describing a qualitative method of genotyping using the Agilent 2100 Bioanalyzer as the microfluidics platform for separation of PCR products amplified from STR loci selected from tetrameric CODIS STR loci. The study demonstrates the novel use of this instrument for rapid, inexpensive electrophoresis and analysis of STR PCR products to allow discrimination among samples from aviation accidents. Accuracy and reproducibility were demonstrated with control samples from a variety of tissues and blood.

The method was applied to an aviation case study where initial sample identification did not correlate with subsequent biochemical and toxicological evaluation. A paternity group sample set previously used for capillary electrophoresis (CE) system validation also was tested. Finally, the findings were validated on a CE system with agreement in all cases.

MATERIALS AND METHODS

Genomic DNA extraction. Tissue or blood was stored frozen at -20°C . Blood was thawed and a 200 μl aliquot used for genomic DNA isolation with the QIAamp DNA mini kit, #51304 (Qiagen; Valencia, CA) using the protocol provided by the manufacturer. Tissue samples included kidney, lung, liver, muscle, and skin. A scalpel was used to cut and mince approximately 25 mg from the frozen tissue. The minced tissue was processed as directed by the manufacturer using the ChargeSwitch gDNA Mini Tissue kit, #CS11204 (Invitrogen; Carlsbad, CA). Yield was determined by A260 on a NanoDrop Spectrophotometer (Model ND-1000, NanoDrop Technologies; Wilmington, DE). Quality was examined on a Bioanalyzer DNA7500 chip (Agilent Technologies; Palo Alto, CA). Samples showing presence of material equal to or greater than 10Kbp were used for PCR.

Sample sets. The 15 control samples were from three individuals. Samples C1-C5 were from Subject 1, Samples C6-C10 from Subject 2, and Samples C11-C15 were from Subject 3. Each control sample set contained a sample from blood, liver, lung, kidney and muscle. The aviation case study contained five samples, labeled T1-T5. They were from kidney, skin, kidney, muscle, and lung, respectively. The paternity set was made up of four blood samples, 1304-1307 from the College of American Pathologists, AABB/CAP PARF Survey B 2005.

PCR Primers and Amplification. Sequences for the ten primer pairs used in the study were as published (3) with modifications as shown at STRbase, www.cstl.nist.gov/biotech/strbase/promega_primers.htm (4). Nine of these are from the 13 CODIS core STR loci (6) and one, amelogenin, is a sex-determination locus (5). Table 1 lists the STR loci and primer sequences used for amplification and the expected product sizes. Primers were synthesized (Integrated DNA Technologies; Coralville, IA) and diluted as needed.

Table 1. Genetic loci and associated primer pair sequences used for PCR in the study (4). See Materials & Methods for PCR conditions.

Locus*	Primer A	Primer B	Product size range (bp)	Repeat Structure
Amel	ACC TCA TCC TGG GCA CCC TGG TT	AGG CTT GAG GCC AAC CAT CAG	212, 218	NA
CSF1PO	AAC CTG AGT CTG CCA AGG ACT AGC	TTC CAC ACA CCA CTG GCC ATC TTC	291-327	[AGAT]n
D3S1358	ACT GCA GTC CAA TCT GGG T	ATG AAA TCA ACA GAG GCT TGC	111-143	TCTA[TCTG]n[TCTA]n
D5S818	GGT GAT TTT CCT CCT TGG TAT CC	AGC CAC AGT TTA CAA CAT TTG TAT CT	123-155	[AGAT]n
D7S820	ATG TTG GTC AGG CTG ACT ATG	GAT TCC ACA TTT ATC CTC ATT GAC	211-251	[GATA]n
D13S317	ACA GAA GTC TGG GAT GTG GAG GA	GGC AGC CCA AAA AGA CAG A	157-201	[TATC]n
D16S539	GGG GGT CTA AGA GCT TGT AAA AAG	GTT TGT GTG TGC ATC TGT AAG CAT GTA TC	264-304	[GATA]n
TH01	ATT CAA AGG GTA TCT GGG CTC TGG	GTG GGC TGA AAA GCT CCC GAT TAT	171-215	[AATG]n
TPOX	ACT GGC ACA GAA CAG GCA CTT AGG	GGA GGA ACT GGG AAC CAC ACA GGT TA	220-256	[AATG]n
vWA [†]	GCC CTA GTG GAT GAT AAG AAT AAT CAG TAT GTG	GGA CAG ATG ATA AAT ACA TAG GAT GGA TGG	123-181	[TCTA]n[TCTG]n[TCTA]n (TCCA[TCTA]n)

*Amel (Amelogenin) is a sex-determination locus, all others listed are tetrameric STR loci

[†]vWA repeat sequence in () is not always present

Amplification was performed using reaction conditions recommended in the Promega PowerPlex 2.1 system (7) except that the buffer was 1x Amplitaq gold reaction buffer, 0.1% TritonX-100, and 0.2mM each dNTP in a 25µl final volume. Cycling was performed on a GeneAmp 9600 thermal cycler (PerkinElmer; Wellesley, MA) using the recommended cycling profile for 40 cycles:

An initial 95° C 11 min, 96° C 1 min; then 94° C 30sec, ramp 68 sec to 60° C, hold for 30 sec, ramp 50 sec to 70° C, hold for 45 sec, for 10 cycles. Then 90° C 30 sec, ramp 60 sec to 60° C, hold for 30 sec, ramp 50 sec to 70° C, hold for 45 sec, for 30 cycles. Then 60° C for 30 min, 4° C soak.

Preliminary testing was performed to optimize the reactions for primer and template concentration. The conditions selected for the study were 1-5ng template DNA and 20 µmol each primer for a 25µl reaction. For multiplexed reactions, conditions were unchanged except that the template concentrations were increased two-fold.

STR Electrophoresis and Analysis. No cleanup of the amplification products was performed. Detection of a 1µl aliquot from a 1:4 dilution in distilled water of each amplification product was performed on a 2100 Bioanalyzer using Expert version B.01.02.S1136 software and DNA 500 Labchip kits (Agilent Technologies; Palo Alto, CA). Expert software was used to overlay the electropherograms and assign PCR product length. Product lengths were calculated by the software based on product migration times compared to spiked-in size standards. Capillary electrophoresis was performed on a 310 Genetic Analyzer (Applied Biosystems; Foster City, CA). Amplification and labeling of PCR products for capillary electrophoresis were done using the PowerPlex 16 System (Promega Corp.; Madison, WI).

RESULTS AND DISCUSSION

The goal of this study was to develop a simple DNA-based protocol that would allow an independent evaluation of accident site samples in cases where toxicological testing results did not agree with the assigned sample identity. PCR is the recognized method of choice because it requires small amounts of starting DNA template, is highly reproducible, and is relatively insensitive to degradation for the amplicon sizes used in this study. An extensive human genotyping system has been established using PCR of well-characterized, short repetitive genomic regions with standardized primers. The PCR primers were selected from the well-characterized FBI human CODIS STR (short tandem repeats) loci. STRbase was used as the source of the CODIS primer sequences (www.cstl.nist.gov/div831/strbase/) (4).

An Agilent 2100 Bioanalyzer was used for the electrophoresis and analysis of PCR samples as a less expensive, convenient alternative to capillary electrophoresis. The Bioanalyzer uses microfluidics chips to perform electrophoresis in the presence of a fluorescent, intercalating dye. In the study presented here, this instrument was used for the separation and analysis of unpurified, unlabelled tetrameric STR PCR products from three sets of samples; first, a control panel of tissue and blood samples from three individuals; second, a case study where toxicological testing results conflicted with sample identification made at the accident site; and finally, a paternity sample set that had previously been examined by capillary electrophoresis (CE).

Nine primer sets from the CODIS STR loci set (6) and a gender identification locus, amelogenin (5) were selected for PCR to have an acceptable power of discrimination. Method development was performed using the control panel. The control panel consisted of liver, lung, kidney, muscle, and blood from each of three subjects. DNA was extracted from all 15 samples and the optimum primer and template concentrations were determined with the standard thermocycling protocol for PCR given in the Materials and Methods section.

There were two primary concerns with the development of the assay: reproducibility across tissue types at each locus, and assay sensitivity. Efficacy of the Bioanalyzer for fragment sizing has been demonstrated (8) with DNA7500 chips, which have a 100-7500 basepair (bp) optimum range. This study showed reproducible sizing results for both well-to-well and day-to-day replicates using a single 380 bp PCR product and a commercial plasmid digest. The DNA 500 chips used in this study have a specified size range of 25-500 bp and at least a five basepair resolution. The CODIS loci used in the study are tetrameric. This defined the resolution necessary for the STR assay to be four basepairs. Table 1 lists the loci and primers selected for the study and the expected product size ranges. In the preliminary tests, a commercial 25 bp ladder containing 11 species was used, as well as control sample PCR products, to more carefully examine the reproducibility and resolution of the DNA 500 chip. The electropherograms from the test showed that there was a high level of intra-chip and inter-chip reproducibility. Table 2A shows that the CV (coefficient of variation) of the well-to-well basepair values for the 25 bp ladder were low, ranging from 0-0.46% over 6 replicates of 11 species/well. The size of the 50 bp species was called the least accurately, showing a 10% difference from the expected size (45 bp instead of 50 bp). However, species in the size range of the STR loci used in this study (approximately 100-300bp) averaged approximately 2% larger than the

Table 2. Well-to-well and inter chip comparison of basepair sizes from ladder standards or PCR products run on 2100 Bioanalyzer DNA 500 chips.

A. Comparison among multiple runs with the Promega 25 bp ladder, all sizes are in basepairs (bp) showing standard deviation between wells and t-Test results between chips.

	Chip 1			Chip 2			Chip 1vs2			% Difference from Expected		
	Well		%CV*	Well		%CV	t-Statistic	Mean	Expected	%		
10	11	12		10	11	12						
45	45	45	0.00	45	45	45	0.00	45	50	-10		
74	74	74	0.00	75	75	75	0.00	74.5	75	-0.7		
101	101	101	0.00	103	102	102	0.46	101.5	100	1.5		
128	127	127	0.37	129	129	128	0.37	128	125	2.4		
153	153	153	0.00	155	155	155	0.00	154	150	2.7		
178	178	178	0.00	180	180	180	0.00	179	175	2.3		
203	202	202	0.23	205	205	204	0.23	203.5	200	1.8		
229	228	228	0.21	231	231	230	0.20	229.5	225	2		
254	253	253	0.19	257	257	256	0.18	255	250	2		
281	280	279	0.29	284	283	283	0.17	281.7	275	2.4		
307	306	305	0.27	311	311	310	0.15	308.3	300	2.8		

* %CV coefficient of variation

Table 2B. Comparison of the two amelogenin PCR products with sizes in basepairs. %CV is used to compare intrachip results, t-test results compare peaks 1 and 2 between chips.

Well	Chip 1		Chip 2	
	Peak1	Peak2	Peak1	Peak2
2	205	209	206	210
3	205	209	206	210
4	204	209	206	210
5	205	209	206	210
6	205	209	207	210
7	204	208	207	210
8	204	208	206	210
9	203	208	206	210
%CV	0.24	0.22	0.22	0.00

t-Test (between chips)	
Peak 1	Peak 2
0.000383	0.000136

expected sizes. The amelogenin PCR products from the control panel also showed a low CV of 0-0.24% for peaks 1 and 2 (Table 2B). This indicated a good well-to-well correlation. Paired t-Tests were performed to compare data between chips and showed statistically significant differences for both the 25 bp ladder and the amelogenin PCR products ($p < 0.05$, Tables 2A and B).

Accuracy of the Bioanalyzer depended on the size of the molecule but was well within specifications for the chip. The 25bp ladder species were called consistently within 2% of the expected size but were generally called larger than the expected size (Table 2A), except for the 50bp species, which was consistently called smaller. Table 2B shows that the amelogenin PCR products were consistently called within 4% of the expected size by the Expert software. These have been published as being 212 and 218 bp, respectively (5) but migrate consistently as seven to nine basepairs smaller in our hands on the Bioanalyzer platform. The difference in migration may be because the PCR reaction was not purified or denatured before electrophoresis or to the presence of “ragged” single stranded ends on the PCR products, compared to the 25bp ladder species in the commercial product. This suggested that the Bioanalyzer was limited in the detection of absolute sizes of the genotyping PCR products. However, the high level of reproducibility of the intra-chip size calls also suggested that reliable comparative analysis would be possible and would allow an accurate determination of forensic sample identification.

Reproducibility between different samples from the same subject was evaluated. DNA isolated from control

samples C3 and C5 (isolated from lung and muscle) was subjected to PCR, and the products were analyzed on the Bioanalyzer in replicate wells. No significant difference in fragment lengths called by the Expert software was seen when the replicates were compared across tissue types (data not shown). This suggested that the isolation protocol yielded consistent template material regardless of starting tissue.

Based on our sample identification need, the design of the assay was for rapid evaluation and relative comparison of PCR products, rather than absolute identification of the specific alleles displayed by the samples. Due to statistical significance of inter-chip sensitivity, all loci comparisons were done between samples electrophoresed on the same chip (Table 2).

All 15 of the control samples were used for PCR with the ten primer pairs using the conditions listed in Materials and Methods and subjected to electrophoresis on the Bioanalyzer. The results were evaluated using the Expert software. The electrophoresis results, both aligned sizes (in basepairs) and electropherograms, were examined for patterns to discriminate between samples. It was noted that the occurrence of singlets (homozygous alleles) versus doublets (heterozygous alleles) was sufficient to discriminate between some samples such that the three control subjects (samples C1-C15) could be identified based on the pattern of homozygous alleles alone. This is a useful starting point for data analysis; however, homozygosity was not necessary to determine sample identities since differences at biallelic loci were observed.

The aviation case study contained five samples, T1-T5. PCR was performed using the established conditions with each of the ten primer sets, followed by electrophoresis on the Bioanalyzer. The results showed that T5 had homozygous alleles for the CSF1PO, D5S818, and THO1 loci. This was sufficient to identify Sample T5 as being unique from T1-T4 (Table 3A). The basepair sizes assigned to the PCR products suggested that the remaining four samples could be divided into two groups. Samples T1 and T3 fell into one group and Samples T2 and T4 fell into a second. Samples in each group potentially belong to the same individual based on the assigned basepair sizes for the PCR products (Table 3A). Examining overlays of the electropherograms confirmed that Samples T1 and T3 displayed the same peaks for all loci tested and differed from T2 and T4. Samples T2 and T4 also displayed the same peaks for all loci tested. This confirmed the assignment to three groups. Figure 1 is an electropherogram overlay for the vWA locus PCR products from samples T1-T5, showing an example of the relative peak migration for the aviation case study samples. Figure 1A clearly shows the difference in migration pattern between peaks from samples T1, T2, and T5. Figure 1B shows electrophoretic migration patterns clearly demonstrating that samples T1 and T3 contained the same vWA alleles that differed from samples T2 and T4, which in turn, display the same vWA alleles. See Table 3A for the corresponding assigned basepair measurements and assessment. Combining the human identity testing information with the toxicological data suggested that the original assignment of samples T2, T4, and T5 to the same individual was an error. Sample T5 was clearly from a different individual. Samples T1 and T3 were blind controls from the same individual included to confirm the efficacy of the assay for quality assurance.

To further validate the assay, the aviation case study samples were examined by CE (Materials and Methods) to confirm the Bioanalyzer conclusions, and a set of four samples making up a paternity group (samples 1304-1307), which had previously been examined by CE, was tested using the Bioanalyzer. The CE results are summarized in Table 3, along with the Bioanalyzer basepair assignments and analysis summary. Listed in Table 3A are the results of the aviation case study by sample with both the base pair size assigned by the Expert software ("bp" column) and the allele assignment made by the Genotyper software following capillary electrophoresis on an ABI 310 ("al" column). Results were summarized using only the microfluidics chip electrophoresis data and agreed with the CE results at all loci except CSF1PO and Tho1. Alleles 11 and 12 for locus CSF1PO were not resolved sufficiently to be called with the Bioanalyzer software. Manual examination of the electropherograms showed

a partially resolved doublet. In the studies reported here, this did not change the final conclusion for the samples. The poor resolution of these alleles is thought to be due to the larger sizes of the CSF1PO PCR products. Alleles 11 and 12 are reported as 311 and 315 bp respectively. Heterozygous samples from all other loci used in the study with alleles differing by four basepairs were easily resolved. For Bioanalyzer genotyping, it is expected that choosing loci with PCR products of less than 300bp will have better sensitivity and accuracy. Alleles 9 and 9.3 for the Tho1 locus could not be resolved. Even with manual observation, a doublet call could not be made. There is only a three basepair size difference between these, 195 and 198 bp, respectively. However, as shown in this study, it is unlikely that an inability to resolve these two alleles will affect the overall accuracy of the assay across a wide range of loci.

The paternity group samples were known to be a mother (1304), daughter (1305), and potential fathers (1306 and 1307). The microfluidics chip results (Table 3B) clearly identified four individuals, and did not support 1307 as a potential father but did support paternity for 1306 for all loci tested. These findings agreed with the CE results.

The ability of the Bioanalyzer to resolve products from multiplexed PCR reactions was tested. The limited resolution of the Bioanalyzer chip and unlabeled PCR products prevented the high degree of multiplexing possible with the CE systems. However, a series of five duplex reactions were successfully resolved. This reduced reagent and microchip usage by half. CSF1PO and Tho1 reactions could be cycled and electrophoresed together, as could D13S17 and TPOX, D3S1358 and amelogenin, D16S539 and D5S818, and finally, vWA and D7S820 (see Material and Methods for conditions) with no loss of resolution and no overlap in the electrophoresed PCR products. Figure 2 shows an example of this duplex reaction set. Secondary products can be seen in the D5S818 and D16S539 panel. However, the height and migration time clearly show these, as well as the shoulder seen in the TPOX peak, as artifacts. The duplexed loci were combined for optimum separation but are not the only multiplex reactions. A number of additional combinations were tested (data not shown), and there appeared to be no excessive secondary products or inhibition of amplification due to the presence of multiple sets of primers at the concentrations reported here. Indeed, a triplex of D5S818, CSF1PO, and THO1 amplifies and separates well.

To address the sensitivity of the assay to contaminated template samples that could result in a miscalled genotype, PCR products from a mixture of two templates were examined. Samples 1304 and 1306 were mixed

Table 3. Bioanalyzer (bp) and capillary electrophoresis (al) results for samples in the Case Study and Paternity Group. The Bioanalyzer summary was supported by the CE results in all cases, see text.

		A. The aviation case study results with summary determined from the Bioanalyzer results.												B. The paternity group results with summary and paternity determination from the Bioanalyzer results.															
		Case Study												Paternity Group															
Locus	T1	T2			T3			T4			T5			Summary	1304	1305		1306		1307		Summary	Paternity						
		bp*	al	bp	al	bp	al	bp	al	bp	al	bp	al			bp	al	bp	al	bp	al								
Amel	204 209	X Y	204 209	X Y	204 208	X Y	204 208	X Y	204 208	X Y	204 208	X Y	204 208	X Y	All male	206	X	205	X	205	X	205	X	205	X	205	X	04,05 female 06,07 male	NA [†]
CSF1PO															1=3, 2=4 5 unique														
	313 322	11 13	311 13	11 13	312 11	11 13	313 11	11 13	310 10	10	312 11	11	316 11,12	11,12		318	11,12	318	12								05, 06 unique	not 07	
D3S1358																													
	140 143	16 17	131 135	14 15	132 136	14 15	131 131	14 15	132 14	131 131	14 15	131 14	14	1=3, 2=4 5 unique	135	15	134	15	130	14	138	16	138	16	138	16	all unique	not 07	
D5S818																													
	143 148	11 12	143 147	11 12	143 146	11 12	143 146	11 12	143 146	11 12	143 146	11 12	142 146	11 12	1=3, 2=4 5 unique	139	10	142	11	141	11	141	11	141	11	141	11	04, 06 unique	ND [‡]
D7S820																													
	229 237	9 11	233 10	10	229 9	9	228 9	9	228 9	9	228 9	9	228 9	9	1=3=5, 2=4	235	10	234	10	238	11	238	11	238	11	238	11	all unique	ND
D13S317																													
															1=3, 2=4	255	new										04, 07 unique	not 07	

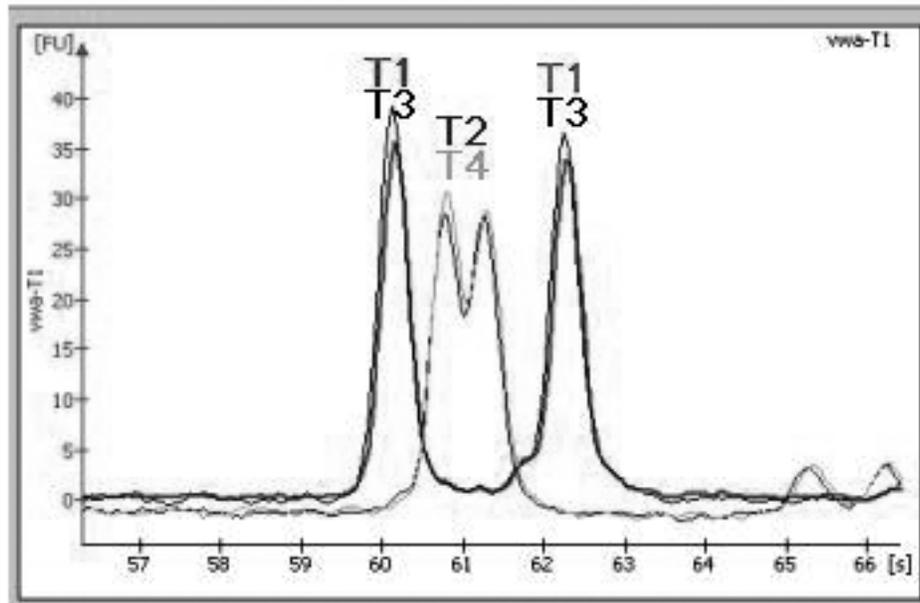
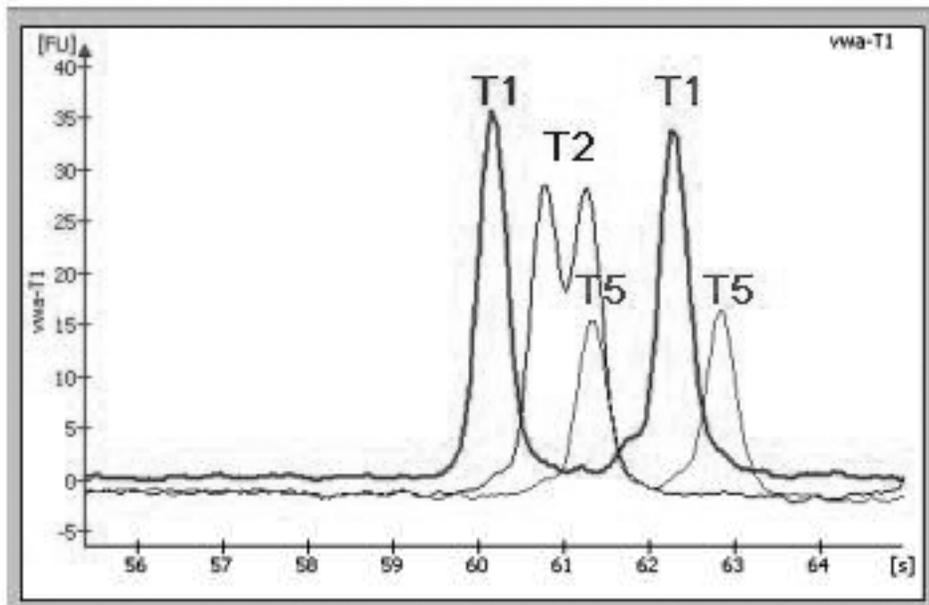


Figure 1. Electropherogram overlays of locus vWA PCR products. The X-axis is migration time, the Y-axis is fluorescence units. Overlay of samples T1, T2, T5 show their unique migration patterns. Overlay of samples T1, T3 and T2, T4 respectively, show matching migration patterns, suggesting T1=T3 and T2=T4.

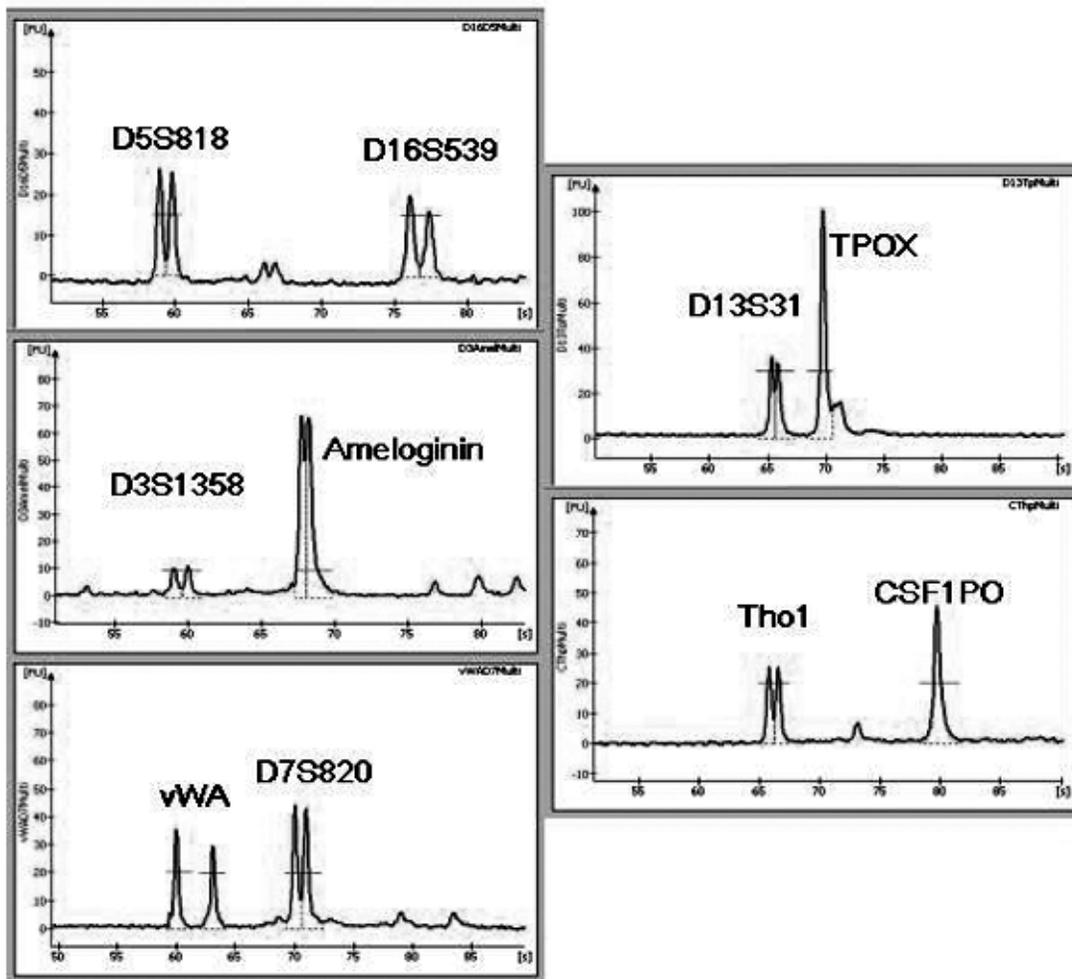


Figure 2. Duplex PCR electropherograms for the ten STR loci used in the study for Control sample 2. PCR performed using conditions described in Materials and Methods. Peaks with intersecting horizontal lines were considered valid products for analysis by the Bioanalyzer software. The X-axis is migration time, the Y-axis is fluorescence units.

in three proportions (Fig. 3A). Both the mixed and un-mixed samples were subjected to PCR with D3S1358 primers and the products subjected to electrophoresis on the microfluidics chip, as above. Control samples C1 and C14 were similarly mixed and PCR was performed with D7S820 primers. In Figure 3 the electrophoresis profiles from both of these series are shown. Figure 3A shows the presence of additional peaks or shoulders in the 1304/1306 mixtures, clear evidence of more than one template source. However, the mixture of samples C1 and C14 shown in Figure 3B has a confusing profile that could result in a false genotype assignment in both the 75:25 and 50:50 C14:C1 mixtures due to the more intense homozygous C1 peak. The significantly lower left-most peak contributed by C14 and the shoulder on the C1 peak (also contributed by C14) would suggest the possibility of contaminated template. Therefore, care should be taken to look for greatly differing peak heights as indicators of sample contamination. These results suggest that only heavily contaminated samples will result in confusing results since contaminating peaks could not be seen in 10:90 and 90:10 mixtures of these samples (data not shown). Taken together, Figures 3A and B suggest that in the event of contaminated samples an ambiguous reading would likely be clearly diagnosed by results at other loci.

The statistical power of the assay was estimated using data from the Einum and Scarpetta study for three North American ethnic groups (9). The nine STR loci

were determined to provide a combined matching probability of approximately $9.5E+07$ to $3E+08$ if all tetrameric alleles could be detected with the Bioanalyzer protocol. Removing the CSF1PO locus from the matching probability calculation naturally reduces the matching probability (Table 4) but the assay retains an acceptable level of power for small, defined populations. The assay was not able to discriminate between the fairly common 9 and 9.3 ThoI alleles (see Table 3). Binning of these two alleles, using the formula for power of discrimination (10), provided a correction for the ThoI locus (M. Scarpetta, personal communication) giving a corrected matching probability of approximately $5E+07$ to $1.9E+08$ (Table 4). It has been observed that the ThoI 9.3 and 10 alleles that differ by one basepair are difficult to resolve on some platforms (11, 12). Allele 10 is relatively rare and has not been detected to date in our samples sets, but a binning of alleles 9.3 and 10 (performed as for the 9 and 9.3 allele binning) showed a small change in allelic frequencies from 0.8372 with no binning to 0.8355 for the Black population, 0.8873 to 0.8824 for the Caucasian population, and 0.8295 to 0.8272 for the Hispanic population. Binning these alleles would have only a minimal effect on the overall matching probability of this assay. The amelogenin gender discrimination locus also was included in the study and does provide additional power, but because it is not an STR locus, it was not included in the Einum and Scarpetta data.

Table 4. Estimation of Matching Probability for the nine STR loci used in the study based on PD values from Einum and Scarpetta (9).

STR locus	Matching Probability (1/1-PD) [†]		
	N. American Blacks N=8000	N. American Caucasians N=8000	N. American Hispanic N=1000
CSF1PO*	12.66	8.26	8.06
D5S818	9.71	7.14	8.26
D7S820	12.05	16.13	13.51
D13S317	7.35	13.33	17.86
D16S539	14.49	11.36	12.82
THOI [‡]	9.71	11.90	8.62
TPOX	11.63	5.32	6.76
vWA	17.24	10.31	12.99
Total MP	3.07E+08	9.42E+07	1.56E+08
w/o CSF1PO	2.43E+07	1.14E+07	1.93E+07
w/o THOI & CSF1PO	2.50E+06	9.57E+05	2.24E+06

*Cannot discriminate all tetrameric alleles for this locus

[†]PD Power of Discrimination, see text

[‡]Includes correction for inability to discriminate alleles 9 and 9.3, see text.

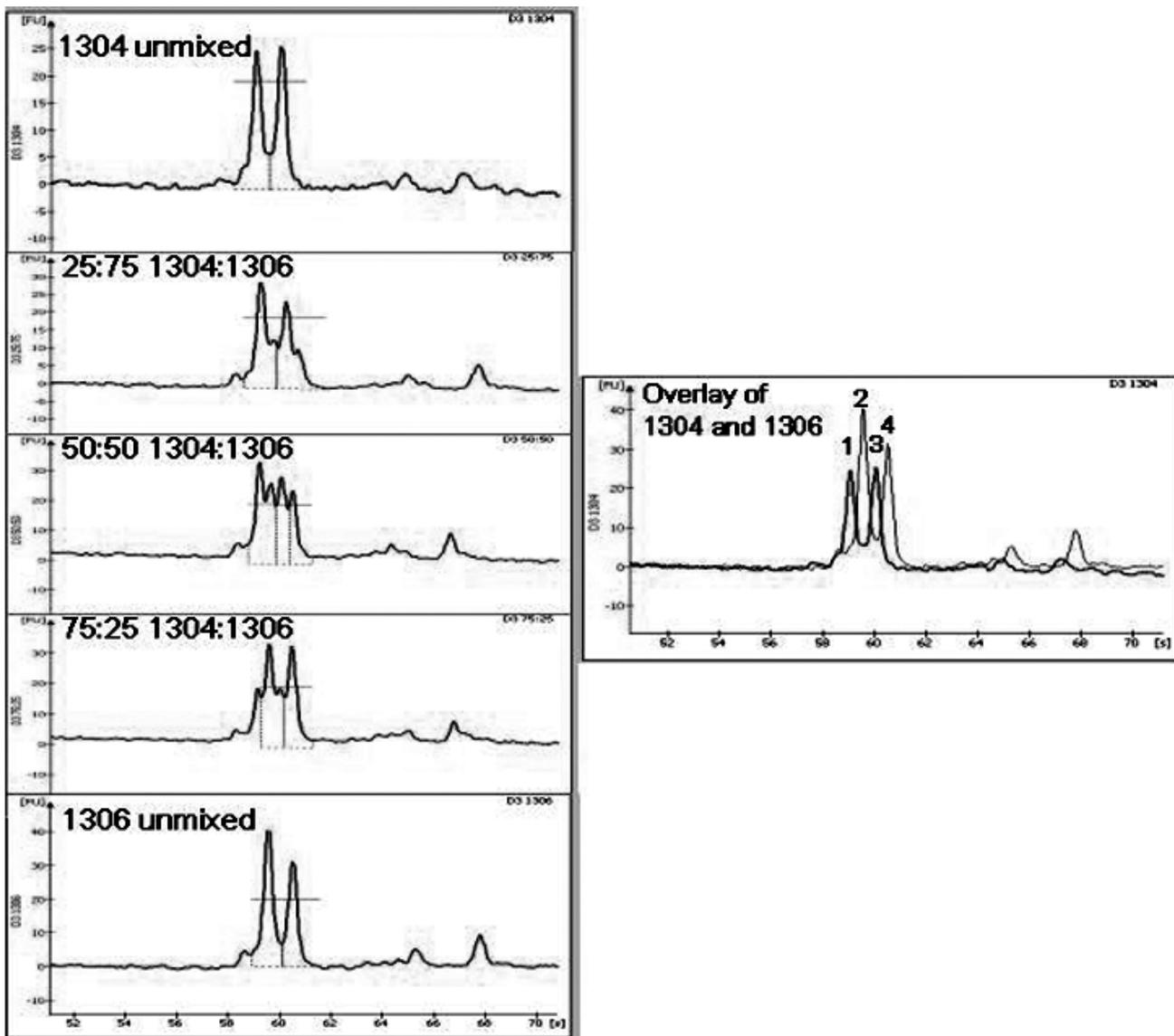


Figure 3. The appearance of PCR products from mixture of two sample templates in three proportions.

A. Appearance of D3S1358 PCR products from mixed template Paternity group samples 1304 and 1306 in three proportions where the presence of both templates is easily detected. Peaks 1 and 3 are from Sample 1304 and Peaks 2 and 4 are from Sample 1306, as indicated in the overlay panel.

B. Appearance of mixed samples following PCR with D7S820 using Control samples C1 and C14. The overlay shows the relative mobility of the peaks from each sample with peaks 1 and 2 from C14 and peak 3 from C1. See text for discussion.

A final note, the DNA1000 series II kit, which will replace the DNA500 kit, was tested with both the Promega 25 bp ladder and the case study samples and found to match the DNA500 resolution and reproducibility. There was a roughly 3.5% increase in assigned sizes for all peaks using the newer kit (data not shown).

The protocol presented here is an inexpensive, rapid method using instrumentation available at a fraction of the cost of capillary electrophoresis instruments and shows great potential for labs screening small numbers of samples. DNA isolated from blood and tissue from control subjects established the feasibility of the Bioanalyzer for the separation of PCR products from STR loci using the well-established CODIS primers. Both an aviation case study and a paternity CE validation set were successfully analyzed with the protocol. Use of a microfluidics chip system precluded purification of PCR products or use of specialized equipment and reagents. This makes the assay more accessible to laboratories with basic molecular biology capabilities in instances where small numbers of samples are assayed. The analysis was done with the Bioanalyzer software, which determined a basepair size assignment following electrophoresis and for each PCR product generated an electropherogram that could be overlaid with other electropherograms for visual inspection of results due to the reliability of size determination and the comparison to a standard. While small discrepancies in size assignments are common, they can easily be addressed by viewing the electropherogram overlays of the size-separated PCR products. This analysis method worked well for both uniplex and multiplexed data. The study did reveal a shortcoming of the microfluidics chip platform to discriminate tetrameric n and $n+1$ CSF1PO alleles where products were greater than 300 bp. This suggested that this locus was less informative for genotyping on the Bioanalyzer, and loci that yield PCR products within the more sensitive range of the instrument will allow a determination of overall assay matching probability. The protocol takes advantage of the well-characterized CODIS primer sets so a large body of literature is available regarding characterization and use of the tetrameric STR loci for genotyping (for example 4, 10). The application utilizes a relative comparison of electrophoretically separated products to return results that are accurate within the limitation of the instrumentation. Absolute identification of the specific alleles is not necessary, adding to the simplicity of analysis. It should be noted that this is not a high throughput method. A large number of samples would be more appropriately analyzed by capillary electrophoresis because of the availability of multiplexed reactions and increased sample capacity. However, this protocol is appropriate for application to small sets of forensic samples as an independent assay to support toxicological and other biochemical tests.

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(3): 647-60.
2. Budowle B, Moretti TR, Niezgoda SJ, Brown BL. CODIS and PCR-based short tandem repeat loci: Law enforcement tools. From Second European Symposium on Human Identification, 1998; 73-88. Madison, WI: Promega Corp.
3. Masibay A, Mozer TJ, Sprecher C. Promega Corporation reveals primer sequences in its testing kits. *J Forensic Sci*, 2000; 45(6): 1360-62.
4. Ruitberg M, Reeder DJ, Butler JM. STRBase: A short tandem repeat DNA database for the human identity testing community. *Nucleic Acids Res*, 2001; 29(1): 320-2.
5. Sullivan KM, Mannucci A, Kimpton CP, Gill P. A rapid and quantitative DNA sex test: fluorescence-based PCR analysis of X-Y homologous gene amelogenin. *BioTechniques*, 1993; 15(4): 636-8, 640-1.
6. Budowle B, Moretti TR. Genotype profiles for six population groups at the 13 CODIS short tandem repeat core loci and other PCR-based loci. *Forensic Sci Commun*, 1999; 1(2) www.fbi.gov/hq/lab/fsc/backissu/july1999/budowle.htm (accessed January 23, 2006).
7. Promega Corporation. PowerPlex 2.1 System technical manual no. D011:11-12, 2001.
8. Panaro NJ, Yuen PK, Sakazume T, Fortina P, Kricka LJ, Wilding P. Evaluation of DNA fragment sizing and quantification by the Agilent 2100 Bioanalyzer. *Clin Chem*, 2000; 46(11):1851-3.
9. Einum DD, Scarpetta MA. Genetic analysis of large data sets of North American Black, Caucasian, and Hispanic populations at 13 CODIS STR loci. *J Forensic Sci* 2004; 49(6): 1381-5.
10. Butler J. *Forensic DNA typing. Biology, technology, and genetics of STR markers*. 2nd ed. St. Louis, MO: Academic Press, 2005.

11. Lins E, Sprecher, CJ, Puers C, Schumm JW. Multiplex sets for the amplification of polymorphic short tandem repeat loci-silver stain and fluorescence detection. *BioTechniques*, 1996; 20(5):882-9.
12. Buel E, Schwartz M, LaFountain MJ. Capillary electrophoresis STR analysis: Comparison to gel-based systems. *J Forensic Sci*, 1998; 43(1):164-70.