J Forensic Sci, Jan. 2005, Vol. 50, No. 1
Paper ID JFS2004216
Available online at: www.astm.org

Michael D. Coble,[1] *Ph.D. and John M. Butler,*[1] *Ph.D.*

Characterization of New MiniSTR Loci to Aid Analysis of Degraded DNA*

ABSTRACT: A number of studies have demonstrated that successful analysis of degraded DNA specimens from mass disasters or forensic evidence improves with smaller sized polymerase chain reaction (PCR) products. We have scanned the literature for new STR loci, unlinked from the CODIS markers, which can generate amplicons less than 125 bp in size and would therefore be helpful in testing degraded DNA samples. New PCR primers were designed and tested for the STR loci D1S1677, D2S441, D4S2364, D10S1248, D14S1434, and D22S1045, arranged into two miniSTR triplexes. All loci show a moderate degree of polymorphism among 474 U.S. population samples tested and were reliable and sensitive to at least 100 pg of DNA template under controlled laboratory conditions and pristine DNA samples. The utility of these new loci were confirmed in comparing the success of the miniSTR assays for typing degraded bone samples while partial profiles were observed with the majority of the samples using a commercial STR kit.

KEYWORDS: forensic science, DNA typing, degraded DNA, STR, miniSTR, D1S1677, D2S441, D4S2364, D10S1248, D14S1434, D22S1045

Short Tandem Repeat (STR) markers have become a valuable tool for forensic DNA typing (1–3). Multiplex amplification of STR loci using the polymerase chain reaction (PCR) can generate typing results from meager amounts of starting material. However, in many forensic cases, DNA samples are highly degraded (e.g., mass disasters such as the World Trade Center collapse of September 11, 2001). If DNA is exposed to the elements or to fire for any length of time, degradation can occur due to bacterial, biochemical or oxidative processes (4). Additionally, environmental contaminants can also be commingled with the forensic evidence. In such specimens a loss of signal is typically observed with larger-sized STR products. This loss of signal may be the result of either particular PCR inhibitors present in the forensic evidence or a DNA template that has been fragmented into small sizes.

PCR inhibitors or DNA degradation typically produce a partial genetic profile with allele and/or locus dropout (5). The problem is further exacerbated when large multiplex PCR reactions are used due to the wide size range of PCR products generated. The current commercially available multiplex STR kits used in forensic DNA typing can generate amplicons in the size range of 100 bp to 450 bp (6,7). In situations where the samples are so badly degraded that STR analysis is not possible, the analysis of the mitochondrial DNA (mtDNA) hypervariable regions is typically used (8,9). However, mtDNA testing is a laborious and cost-prohibitive procedure for most forensic laboratories to utilize. Additionally, given the haploid and maternal-only transmission of mtDNA, the power of discrimination is not as powerful for identification as a full multilocus STR match.

An additional approach to trying to recover information from degraded DNA samples is to reduce the size of the PCR products by moving primers in as close as possible to the STR repeat region (10,11). The observation that smaller-sized PCR products from STR markers produced a higher success rate with degraded DNA samples was first reported in 1995 with analysis of victims of the Waco Branch Davidian fire (5). These results have been confirmed by a number of other publications in recent years (10–15).

One major advantage of these smaller STRs, or "miniSTRs," is that database compatibility can be maintained with convicted offender samples processed using commercial STR multiplexes. In addition, these smaller product sizes are more favorably analyzed by alternative technologies such as time-of-flight mass spectrometry (16,17) and rapid microchannel electrophoretic separations (18). Within the forensic community, a core set of STR markers have been developed for utilization in forensic casework, and large databases such as the Combined DNA Index System (CODIS) have been developed incorporating these markers (3,19). Recently, an effort to reduce the STR amplicon size for CODIS loci has resulted in a set of miniplexes to analyze degraded DNA (20). However, a few of the CODIS loci cannot be made into smaller amplicons due to repeat flanking regions that are not amenable for redesigned primers (e.g. D7S820) or due to the fact that some of the CODIS loci have large allele ranges (e.g., FGA).

Gill et al. (21) have highlighted instances where autosomal single nucleotide polymorphisms (SNPs) or novel miniSTRs would be used to supplement the current battery of forensic STRs in the cases of mass disasters where the DNA may be highly degraded. Additional miniSTR markers can increase the likelihood of identifying remains when only a partial profile is obtained with the current CODIS markers. Even if a full STR profile is determined from a mass disaster, there are occasions where victims have insufficient family members available to achieve a high level of confidence for an association (21). The use of SNPs or miniSTR markers can both

[1] Biotechnology Division, National Institute of Standards and Technology, Gaithersburg, MD 20899-8311.

Portions of this work were presented at the Annual American Academy of Forensic Sciences meeting in Feb. 2004, Dallas, TX.

† National Institute of Justice (NIJ) funded the work described here through an interagency agreement with the NIST Office of Law Enforcement Standards.

Received 21 May 2004; and in revised form 14 Aug. 2004; accepted 14 Aug. 2004; published 8 Dec. 2004.

increase the probability that a degraded sample can be typed, and provide added statistical support for an association. Additionally, there may be instances in routine paternity analyses where additional markers are required to improve the power of exclusion (22) or in complex paternity cases (e.g., incest).

This paper outlines the development of novel miniSTR loci that can be useful for the analysis of degraded DNA. These loci are not meant to replace the current CODIS loci, but complement them in cases where additional loci are needed to develop a genetic profile. Therefore, we chose to focus on loci unlinked to the current CODIS markers. Our goal is to create multiplexed sets, referred to as "miniplexes," usually with a single locus in each dye color. We focused on STRs that had high heterozygosities, small allele size ranges, and were capable of being made into short amplicons by having "clean" flanking regions where primers could be placed immediately adjacent to the STR repeat region.

Materials and Methods

Selection of Loci

Potential miniSTR markers were screened from the STR marker sets published by the Marshfield Clinic Center for Medical Genetics (23) (http://research.marshfieldclinic.org/genetics/). The original amplification primer sequence for loci selected in this study (described in greater detail below) were those present in the Genome Database (http://www.gdb.org/). Approximately 100 bp of sequence information flanking either side of the target repeat was downloaded as a text file from a BLAST-nucleotide search on the Internet (http://www.ncbi.nlm.nih.gov/BLAST/Blast.cgi). The chromosomal position of each STR repeat was determined by the web-based program, BLAT (http://genome.ucsc.edu/cgi-bin/hgBlat) and the July 2003 assembly of the human genome. The GenBank® (http://www.ncbi.nlm.nih.gov) accession numbers for each marker are listed in Table 1.

Primer Design

PCR primers were designed against each reference sequence using web-based Primer3 (24). Typically the default parameters were used: primer Tm values from 57°C–63°C, with 60°C as the optimum, minimum primer size of 18 nucleotide bases, and the primer GC% range from 20–80. The PCR product size was made as small as possible around the target region that consisted of the STR repeat (Table 2).

Resulting primers (Table 2) were tested for potential interactions with each other using an in-house program described elsewhere (25). For convenience, all forward primers were labeled with a fluorescent dye. An additional guanine base was added to each reverse

TABLE 1—*Information on six STRs examined in this study. Chromosomal location and base pair (bp) position of each marker was determined by using BLAT (http://genome.ucsc.edu/cgi-bin/hgBlat) and the July 2003 assembly of the human genome. The observed allele ranges and heterozygosities are from this study. The "old" size range (bp) values are based on the expected amplicon size with the primers used for the screening sets of the Marshfield Clinic's Center for Medical Genetics (http://research.marshfieldclinic.org/genetics.html). The "new" size range (bp) values are based on the miniSTR amplicon sizes that were observed in this study relative to the GS500 LIZ size standard. The GenBank Accession numbers and reference alleles were obtained from nucleotide searches through NCBI (http://www.ncbi.nlm.nih.gov/).*

STR Locus	Chromosome Location	Chromosome bp Position	Sequence Motif	Allele Range	New Size Range* (bp)	Old Size Range** (bp)	GenBank Accession	GenBank Allele	Observed Heterozygosity
D1S1677	1q23.3	160,747,193	$(GGAA)_n$	9–18	81–117	179–215	G09926	15	0.75
D2S441	2p14	68,213,613	$(TCTA)_n$	9–17	78–110	148–180	G08184	12	0.76
D4S2364	4q22.3	93,975,767	$(GAAT)(GGAT)(GAAT)_n$	8–12	67–83	120–136	G08326	9	0.53
D10S1248	10q26.3	130,566,908	$(GGAA)_n$	10–20	83–123	231–271	G08820	13	0.78
D14S1434	14q32.13	93,298,432	$(GATA)_n(GACA)_n$	13–20	70–98	222–250	G27275	13	0.68
D22S1045	22q12.3	35,779,368	$(TAA)_n$	5–16	76–109	132–165	G08085	13	0.77

* Apparent size observed relative to GS500 LIZ size standard.
** Expected amplicon size based upon original PCR primers.

TABLE 2—*MiniSTR primers used in this study. The "Oligo distance from repeat" column refers to the distance of the 3′ end of the primer to the STR repeat region (in base pairs). A negative number indicates that the end of the primer is in the repeat region (with the nucleotide bases underlined at the 3′ end of the primer). The 5′ Guanine residue in each reverse primer was added to promote adenylation (20). The GenBank reference allele in terms of repeat number is listed in Table 1.*

Marker Name	Primer Sequence (5′-3′)	Primer Concentration (μM)	Oligo Distance from Repeat (bp)	Amplicon Size (bp) with GenBank Allele
D1S1677	for [NED]-TTCTGTTGGTATAGAGCAGTGTTT	1.3	0	103
	rev **G**TGACAGGAAGGACGGAATG	1.3	0	
D2S441	for [VIC]-CTGTGGCTCATCTATGAAAACTT	0.7	0	92
	rev **G**AAGTGGCTGTGGTGTTATGAT	0.7	0	
D4S2364	for [FAM]-CTAGGAGATCATGTGGGTATGATT	1.1	2	78
	rev **G**CAGTGAATAAATGAAC<u>GAATGGA</u>	1.1	−7	
D10S1248	for [FAM]-TTAATGAATTGAACAAATGAGTGAG	1.3	1	102
	rev **G**CAACTCTGGTTGTATTGTCTTCAT	1.3	0	
D14S1434	for [VIC]-TGTAATAACTCTACGA<u>CTGTCTGTCTG</u>	1.3	−11	88
	rev **G**AATAGGAGGTGGATGGATGG	1.3	0	
D22S1045	for [NED]-ATTTTCCCCGATGATAGTAGTCT	0.8	3	105
	rev **G**CGAATGTATGATTGGCAATATTTTT	0.8	6	

primer (noted in bold and underlined—see Table 2) to promote fully adenylated PCR products (+A form) (20).

PCR Primers and Other Reagents

Fluorescently labeled primers, AmpliTaq Gold® DNA polymerase, and associated buffers were obtained from Applied Biosystems (Foster City, CA). All forward miniSTR primers were labeled with 6FAM™, VIC™, and NED™ dyes (Applied Biosystems), which enabled the use of available matrix standards and reliable color separations on the ABI 3100 Genetic Analyzer. Unlabeled primers were purchased from Qiagen Operon (Alameda, CA). Commercial multiplexed genotyping kits utilized in these studies included Powerplex® 16 from Promega Corporation (Madison, WI).

Source of DNA Samples for Population Testing

Anonymous liquid blood samples with self-identified ethnicities were purchased from Interstate Blood Bank (Memphis, TN) and Millennium Biotech, Inc. (Ft. Lauderdale, FL) and extracted and quantified as previously described (26). A subset of 474 samples from 700 individuals representing three U.S. populations (164 African Americans, 170 Caucasians, and 140 Hispanics) were used in this study (26). Sixteen samples extracted from degraded bones were provided by Bruce McCord, Ohio University.

PCR Amplification

Amplifications were performed for general sample (population) processing in reaction volumes of 10 μL using a master mix containing 1X GeneAmp® PCR Gold buffer (Applied Biosystems), 1.5 mmol/L $MgCl_2$, 200 μmol/L deoxynucleotide triphosphates (USB Corporation, Cleveland, OH; dNTPs: dATP, dCTP, dGTP, dTTP), 1 μL of template DNA at 1 ng/μL concentration, and 1 unit of AmpliTaq Gold DNA polymerase. Primer concentrations were adjusted empirically to balance peak heights within and between dye colors usually starting with 1 μmol/L (final primer concentrations are listed in Table 2). Thermal cycling was performed with the GeneAmp 9700 (Applied Biosystems) using the following conditions in 9600-emulation mode (i.e., ramp speeds of 1°C/s):

95°C for 10 min
28 cycles: 94°C for 1 min, 55°C for 1 min, 72°C for 1 min
60°C for 45 min
25°C forever

For degraded samples and sensitivity studies, 2 units of AmpliTaq Gold DNA polymerase were used and the reaction volume was increased to 25 μL. Additionally, the number of cycles was increased to 32.

Analysis on ABI 310 (single capillary) and ABI 3100 (16-capillary) Genetic Analyzer

The allelic ladders were separated electrophoretically using an ABI Prism® 310 Genetic Analyzer (Applied Biosystems) using the filter set G5 in order to process the data from the 5 dyes 6FAM (blue), VIC (green), NED (yellow), PET (red) and LIZ (orange). An appropriate matrix was created using materials from matrix standard sets DS-32 and DS-33 (Applied Biosystems). Each sample for analysis on the 310 was prepared by adding 1 μL of each PCR product to 19 μL of Hi-Di™ formamide (Applied Biosystems) containing 0.75 μL GS500 LIZ size standard (Applied Biosystems). Samples were then placed immediately into the instrument for analysis without heat denaturing or snap cooling the samples prior to running them (26). Samples were injected for 5 s at 15,000 V and separated at 15,000 V for 50 min with a run temperature of 60°C. Standard electrophoretic conditions were used including 310 Genetic Analyzer POP™-6, 1X A.C.E. Buffer (Amersco, Solon, OH), and a 47 cm × 50 μm capillary (Applied Biosystems).

Prior to running any miniplex samples on the ABI Prism 3100 Genetic Analyzer, a 5 dye matrix was established under the "G5 filter" with the dyes 6FAM, VIC, NED, PET and LIZ. Samples were typically prepared with 15 μL Hi-Di formamide (Applied Biosystems), 0.6 μL GS500 LIZ (Applied Biosystems), and with 1 μL PCR product. Samples were prepared as described above. The miniSTR samples were run using a modified version of the module GeneScan36_POP4DefaultModule, which performs an electrokinetic injection onto the 16-capillary array for 10 s at 3,000 V. The STR alleles were then separated at 15,000 V for approximately 20 min with a run temperature of 60°C using the 3100 POP-6 sieving polymer, 1X A.C.E. buffer (Amresco), and a 36 cm array (Applied Biosystems). Data from the ABI 3100 was analyzed using GeneScan 3.7 and Genotyper 3.7 programs (Applied Biosystems) for a Windows NT platform.

Generation of Allelic Ladders and Genotyper Macros

Allelic ladders were created with the miniSTRs using a combination of individual templates that represent the range of alleles observed in the populations. The samples were amplified separately for each marker according to the protocol above with the 72°C extension time increased to 2 min during each cycle. The final extension is lengthened to 60 or 90 min since there are more PCR products to create.

Genotyper macros were constructed for each of the miniplex combinations using fixed bin allele sizes to perform allele designations. Bin width ranged from ±0.5 to 1.0 bp for the six STRs analyzed in this study. The number of repeats was calibrated to allele size by sequencing (described below).

DNA Sequencing

At least one homozygote sample for each miniSTR locus was sequenced to calibrate observed size to repeat number. To fully characterize the number of repeating units, a set of sequencing primers located outside of the target amplicon was designed as outlined above. The advantages of using the sequencing primers, rather than the amplification primers, include the ability to (1) fully sequence the repeat and (2) sequence the primer binding region to detect nucleotide mutations that can create null alleles (3). The sequencing primers used for each locus are listed in Table 3.

Template DNA was amplified as outlined above for 32 cycles of PCR. Cycle sequencing was performed with the ABI PRISM® BigDye® Terminator (Version 3.0) cycle sequencing kit (Applied Biosystems). The sequencing mixture consisted of: 9 μL dH_2O, 4 μL of 2.5X Sequencing Dilution Buffer (Edge Biosystems, Gaithersburg, MD), 4 μL BigDye Terminator kit mix (Applied Biosystems), 1 μL of forward or reverse primer (10 μM each), and 2 μL of PCR product for a total volume of 20 μL. Thermal cycling was conducted in a GeneAmp 9700 thermocycler with the

TABLE 3—*Unlabeled primers used in this study for sequencing purposes. These primers bind outside the miniSTR typing primers and will enable analysis of any miniSTR primer binding site mutations should they be suspected of causing allele dropout or extreme heterozygote peak imbalance.*

Marker Name	Primer Sequence (5′-3′)	PCR Product Size (with Reference Allele)
D1S1677	F-GTAGTGCTGGTGCAGCGTAA R-TGCAATAGCAAATATCAGAATGTGT	261 bp
D2S441	F-CTGTTCCTGAGCCCTAATGC R-CACCACACCCAGCCATAAAT	290 bp
D4S2364	F-TGTTGTCTGTAGGAGCTGAGAAA R-GGTGTTTGGAGATGGCTGTT	258 bp
D10S1248	F-AGCAAACCTGAGCATTAGCC R-AGTGCTTGGCAAAGAGCAGA	257 bp
D14S1434	F-TTCCCAGCCTCCATAATCAG R-TGCAAATGCACACAGATTTC	262 bp
D22S1045	F-CCCACTATGGGCAAACCTTA R-TGTGCTTCAGTCTCCTCAGC	342 bp

following conditions: an initial 1 min denaturation at 96°C; followed by 25 cycles of 15 s at 94°C (denaturation), 5 s at 50°C (annealing), and 2 min at 60°C (extension). The DNA product was purified by filtration though a single tube filtration cartridge (Edge BioSystems).

Electrophoresis and sequencing were performed on the 3100 Genetic Analyzer using POP-6polymer on a 36 cm capillary array. Sequences were aligned with the GenBank reference allele (accession numbers listed in Table 1) and edited using Sequencher Plus 4.0.5b11 (GeneCodes, Ann Arbor, MI).

Data Analysis

The resulting data were evaluated using program H (exact test, 2000 shuffling) of the DNATYPE program (27). Allele frequencies, heterozygosity values, and p-values from the Hardy-Weinberg test for the three U.S. populations are listed in Tables 4 and 5. Other statistical measures such as power of discrimination, PIC, power of exclusion, and typical paternity index were calculated using the PowerStatsV12 spreadsheet (Promega Corporation). Heterozygosity values from STR loci typed with the Identifiler kit (Applied Biosystems) using the same 474 individuals were also determined using DNATYPE (27).

Results and Discussion

Steps in Screening Candidate Loci

We sought to identify STR loci that would make an "ideal" miniSTR marker. As an example, the CODIS locus TH01 has many characteristics that make it an ideal marker for the analysis of degraded DNA. First, this marker has a small number of alleles observed in the population that range from 3–14 repeats (see http://www.cstl.nist.gov/biotech/strbase/str_th01.htm). The allele spread from the smallest allele to the largest allele, therefore, is 44 bases. Second, the flanking regions adjacent to the repeat are considered "clean." That is, PCR primers can hybridize immediately adjacent to the repeats, making the PCR amplicon as small as possible. Previous research (20) has demonstrated the ability to design a miniSTR TH01 forward primer immediately adjacent to the repeat, and a reverse primer one-nucleotide base from the repeat. The amplicon size for the reduced size TH01 allele ranges from 55–98 base pairs. Third, TH01 is known to have a heterozygosity value greater than 0.70 in many populations. And finally, TH01 has a tetranucleotide repeating unit, which is important for reduced stutter to enable easier identification of mixed source samples (3).

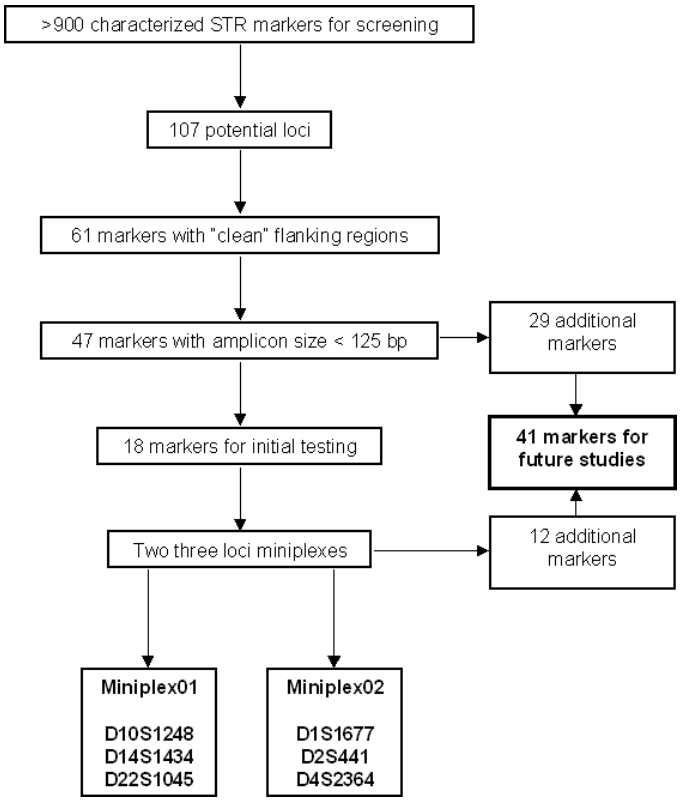

FIG. 1—*Schematic of the screening process to identify miniSTR loci. Published STR markers (23, 28–30) were used to screen for "ideal" markers to make small amplicons. In general, potential loci were selected for allele ranges less than 24 bp and heterozygosity values greater than 0.70.*

Potential miniSTR markers were selected from previously well-characterized STR loci including the whole genome screening sets of STRs developed by the Marshfield Clinic Center for Medical Genetics (23) (http://research.marshfieldclinic.org/genetics/). These screening sets have been used extensively in genetic linkage studies across the entire nuclear genome to associate specific genes that cause or are linked to human diseases (an extensive list of studies that have used the screening sets can be found on the Marshfield website). A subset of the screening markers has also been used to characterize the levels of genetic diversity in global populations (28–30).

A schematic diagram outlining the screening process for identifying markers can be found in Fig. 1. In order to identify novel loci for additional forensic discrimination of degraded material, we chose to avoid loci that were linked to the current core 13 CODIS loci (3). Using the information within the references outlined above, we screened 920 STR loci. From these markers, we identified 107 potential loci that met our criteria for creating miniSTRs—in general, those loci had an allele spread of <24 bp, observed heterozygosities of >0.70, and contained tri or tetranucleotide repeat units. Of the 107 potential loci, 61 markers had "clean" flanking regions, permitting PCR primers with the characteristics listed in the Materials and Methods section to be close to the target repeat and thus produce the smallest possible PCR products. From these 61 candidates, 47 markers were further selected for having amplicon sizes less than 125 bp according to the GenBank reference allele size when primers were designed as close as possible to the repeat

TABLE 4—*Allele frequency distributions of six miniSTR loci among 474 individuals from three U.S. groups. This information comes from analysis of 164 African Americans, 170 Caucasians, and 140 Hispanics. The most common allele is shown in bold font for each group.*

Miniplex 01

	D10S1248					D14S1434					D22S1045				
Allele	Afr. Am.	Cauc.	Hisp.	Total		Allele	Afr. Am.	Cauc.	Hisp.	Total	Allele	Afr Am.	Cauc.	Hisp.	Total
10	0.0061			0.0021		13	0.0061		0.0036	0.0032	5	0.0061			0.0021
11	0.0031			0.0011		14	0.2561	0.1471	0.1857	0.1966	6				
12	0.0335			0.0116		15	0.0213	0.0294	0.0357	0.0285	7	0.0457		0.0179	0.0211
13	0.1189	0.0382	0.0607	0.0729		16	0.0945	0.0177	0.0429	0.0518	8	0.1128	0.1441	0.0607	0.1089
14	0.2592	**0.3441**	0.2643	0.2918		17	0.2317	**0.3971**	0.3321	0.3203	9	0.0579	0.0177	0.0179	0.0317
15	**0.2652**	0.2941	**0.3179**	0.2918		18	**0.3628**	0.3912	**0.3857**	0.3795	10	0.0061	0.0088	0.0107	0.0085
16	0.2256	0.1824	0.2357	0.2125		19	0.0244	0.0088	0.0143	0.0159	11	0.0610	0.0500	0.0250	0.0465
17	0.0640	0.1147	0.1036	0.0930		20	0.0031	0.0088		0.0042	12	**0.2683**	0.3382	**0.4536**	0.3478
18	0.0213	0.0235	0.0143	0.0201							13	0.2012	**0.3588**	0.3107	0.2907
19	0.0031			0.0011							14	0.2195	0.0794	0.0964	0.1321
20		0.0029	0.0036	0.0021							15	0.0183	0.0029	0.0071	0.0095
											16	0.0031			0.0011

Miniplex 02

	D1S1677					D2S441					D4S2364				
Allele	Afr. Am	Cauc.	Hisp.	Total		Allele	Afr. Am.	Cauc.	Hisp.	Total	Allele	Afr. Am.	Cauc.	Hisp.	Total
9	0.0061	0.0059		0.0042		9		0.0029		0.0011	8		0.0029		0.0011
10	0.0061	0.0088	0.0107	0.0085		10	0.0854	0.2088	**0.3357**	0.2040	9	0.1463	0.1794	0.2393	0.1850
11	0.1189	0.0853	0.1214	0.1078		11	**0.3567**	**0.3529**	0.3143	0.3425	10	**0.7683**	**0.5500**	**0.5179**	0.6163
12	0.1372	0.2294	0.2714	0.2104		11.3	0.0488	0.0618	0.0357	0.0497	11	0.0854	0.2677	0.2393	0.1966
13	0.2713	**0.3559**	**0.3179**	0.3140		12	0.1829	0.0441	0.0250	0.0856	12			0.0036	0.0011
14	**0.3201**	0.2265	0.2036	0.2526		12.3	0.0061	0.0029	0.0036	0.0042					
15	0.1067	0.0618	0.0500	0.0740		13	0.0427	0.0177	0.0179	0.0264					
16	0.0305	0.0206	0.0250	0.0254		13.3	0.0031			0.0011					
17						14	0.2531	0.2412	0.2143	0.2378					
18	0.0031	0.0059		0.0032		14.3	0.0031			0.0011					
						15	0.0183	0.0677	0.0464	0.0444					
						16			0.0036	0.0011					
						17			0.0036	0.0011					

TABLE 5—*Population statistics of six miniSTR loci among 474 individuals from three U.S. populations (number of samples: 164 African Americans, 170 Caucasians, and 140 Hispanics). The significant p-value for the Hardy-Weinberg Equilibrium (HWE) in African Americans for the D10S1248 marker is highlighted in bold. Heterozygosity and HWE values were calculated using 2,000 shufflings of the exact test in the DNATYPE program (27). All other statistics were calculated using PowerStatsV12 (Promega).*

	Miniplex 01								
	D10S1248			D14S1434			D22S1045		
	Afr. Am.	Cauc.	Hisp.	Afr. Am.	Cauc.	Hisp.	Afr. Am.	Cauc.	Hisp.
Observed Heterozygosity	0.829	0.753	0.743	0.646	0.712	0.646	0.817	0.759	0.721
Power of Discrimination	0.914	0.891	0.903	0.890	0.812	0.866	0.940	0.881	0.839
Polymorphism Information Content	0.760	0.710	0.720	0.700	0.600	0.650	0.790	0.680	0.630
Power of Exclusion	0.654	0.515	0.498	0.350	0.447	0.385	0.631	0.525	0.462
Typical Paternity Index	2.930	2.020	1.940	1.410	1.730	1.520	2.730	2.070	1.790
HWE p-values (unbiased)	0.266	0.902	0.602	**0.005**	0.229	0.375	0.935	0.382	0.363
	Miniplex 02								
	D1S1677			D2S441			D4S2364		
	Afr. Am.	Cauc.	Hisp.	Afr. Am.	Cauc.	Hisp.	Afr. Am.	Cauc.	Hisp.
Observed Heterozygosity	0.750	0.741	0.743	0.793	0.753	0.729	0.366	0.582	0.664
Power of Discrimination	0.918	0.897	0.908	0.899	0.912	0.880	0.580	0.769	0.775
Polymorphism Information Content	0.750	0.720	0.730	0.730	0.730	0.690	0.350	0.530	0.550
Power of Exclusion	0.510	0.495	0.498	0.586	0.515	0.474	0.094	0.270	0.375
Typical Paternity Index	2.000	1.930	1.940	2.410	2.020	1.840	0.790	1.200	1.490
HWE p-values (unbiased)	0.338	0.565	0.470	0.415	0.705	0.741	0.667	0.729	0.275

region. We selected 18 markers for initial testing, all with amplicon sizes less than 110 bp (based upon the GenBank reference allele).

Of the 18 candidates selected for initial testing, we focused on six markers (D1S1677, D2S441, D4S2364, D10S1248, D14S1434, and D22S1045) for further characterization (Table 1). One marker, D22S1045, consisted of a trinucleotide repeat while the remaining five loci consist of tetranucleotide repeats (Table 1). Two of these markers are found on chromosomes that have CODIS loci—TPOX is on chromosome 2 and FGA is on chromosome 4. To determine the relative location of these markers to the CODIS loci, each marker was mapped onto each respective chromosome using BLAT (http://genome.ucsc.edu/cgi-bin/hgBlat) and the July 2003 assembly of the human genome. D2S441 and D4S2364 were found to be greater than 50 centimorgans from TPOX and FGA respectively, suggesting that these markers are probably unlinked.

Although the markers used in this study are routinely used in genetic association studies for linkage to human disease, we do not anticipate difficulties in using these loci to characterize genetic differences for forensic purposes. In fact five of the thirteen CODIS loci (D7S820, D8S1179, D13S317, D16S5539, and TPOX) were serendipitously included within the Marshfield Set 12 screening STRs (23). Therefore, we would find it no more likely to see a genetic disease association between these novel miniSTR loci than any of the CODIS loci. However, it would be imprudent to assert that such an association would never occur in the future. Multiple studies may be required in order to confirm clinical studies associating forensic markers with human diseases. For example, Klintschar et al. (31) recently found no genetic association between HumTH01 alleles 9.3 and 10 to arterial hypertension, a clinical finding proposed by Sharma et al. (32).

The six markers that were characterized in this study were arranged into two multiplexes (triplexes): "mini01" contained D10S1248, D14S1434, and D22S1045 while "mini02" contained D1S1677, D2S441, and D4S2364. The dye labels associated with each locus are listed in Table 2. All of these loci have forward and reverse primers that are very close to the repeat unit. In two of the primers (D4S2364 and D14S1434), it was possible to achieve successful amplification and include two complete repeats in the 3′ end of the primer, a possibility discovered while working with new primers for time-of-flight mass spectrometry (33).

In order to calibrate measured allele size to repeat number, at least one homozygous sample from each locus was sequenced. Previous studies by the Marshfield Clinic with these loci relied solely on PCR product size rather than calibrated repeat numbers. For consistency, the total number of repeats was included in characterizing the number of repeats for genotyping according to guidelines outlined by the International Society of Forensic Haemogenetics (3,34). For example, the sequenced D4S2364 allele had the following repeat motif: (GAAT) (GGAT) $(GAAT)_7$ and was designated allele 9. Note that 7 of the 8 nucleotides in the two repeats adjacent to the core repeating unit of seven repeats were included in the D4S2364 reverse primer (see Table 2).

An example of one sample characterized for both miniplexes can be found in Fig. 2. This particular individual, an African American, was observed to be heterozygous at all six miniSTR markers. All markers showed good balance between alleles (Fig. 2). The D1S1677 miniSTR required additional primer optimization to boost the overall peak height compared to the other two miniplex02 markers. Allelic ladders (Fig. 3) were constructed from combining DNA templates from various individual samples to illustrate the allele spread for each marker. Additional quantities of each miniSTR ladder can be amplified to create an allelic ladder for future genotyping macro development.

Evaluation of MiniSTR Loci

We analyzed mini01 and mini02 across a subset of U.S. population samples comprised of Caucasian ($n = 170$), African American

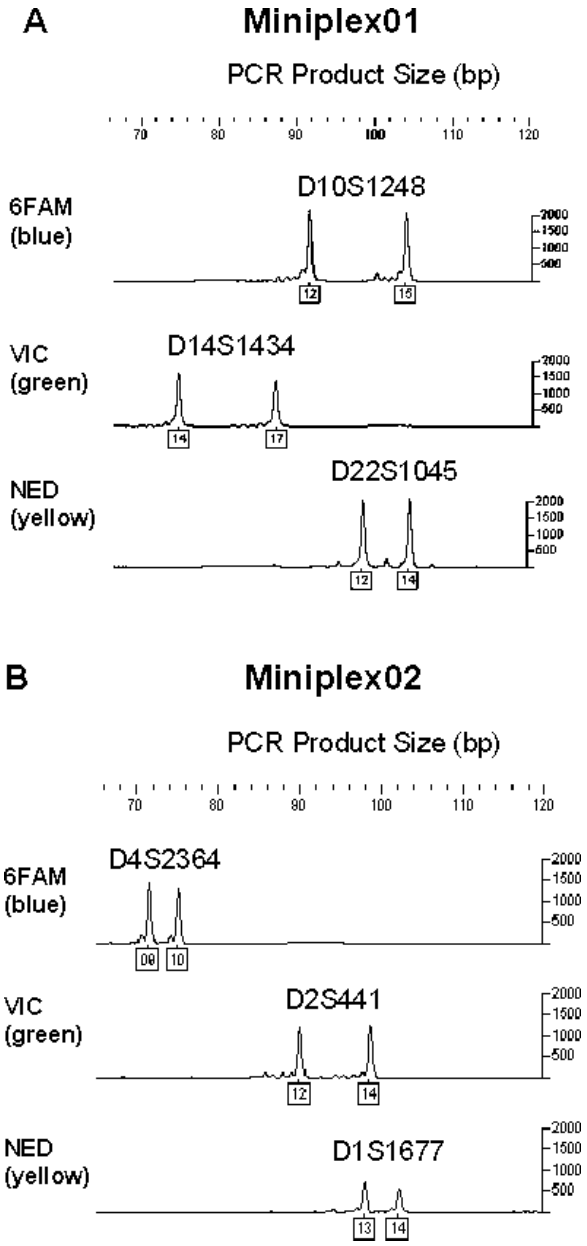

FIG. 2—*Representative sample from an individual analyzed with both (A) miniplex01 and (B) miniplex02 using standard conditions for amplification (28 cycles, 1U AmpliTaq Gold polymerase, 10 μL PCR).*

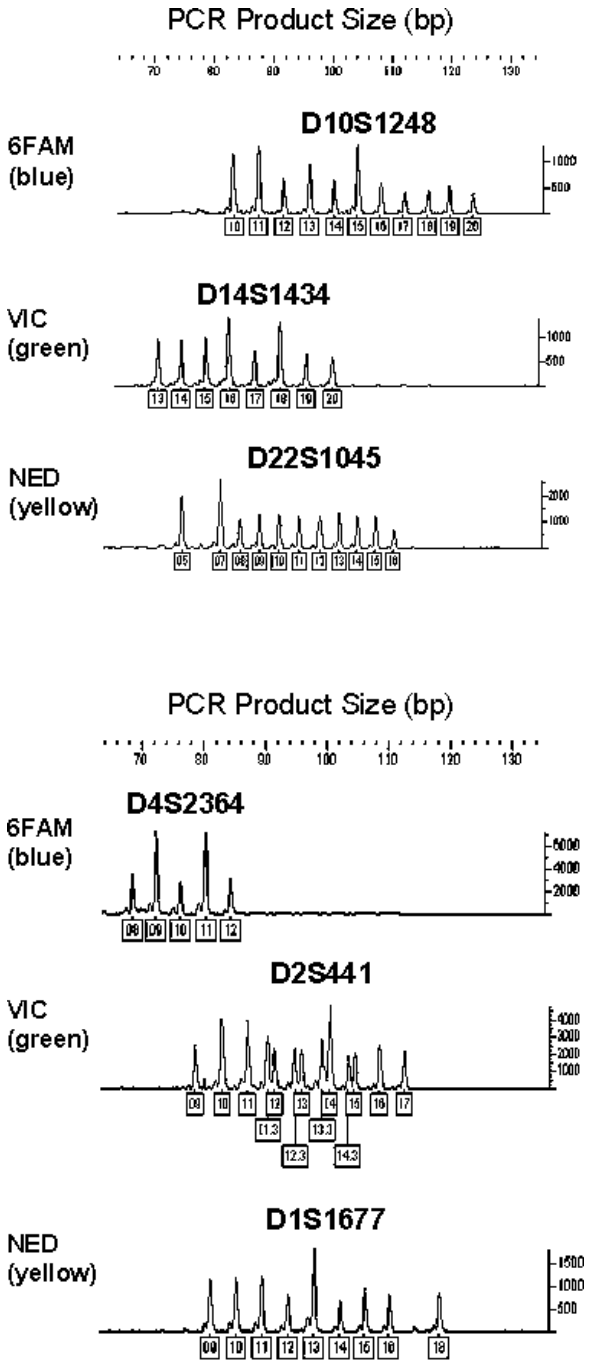

FIG. 3—*MiniSTR allelic ladders for miniplex01 (top panel) and miniplex02 (bottom panel). Samples were amplified individually and electrophoretically separated on an ABI 310.*

($n = 164$) and Hispanic ($n = 140$) individuals (26). Allele frequencies for each locus among the 474 samples are listed in Table 4. The D4S2364 marker showed the lowest amount of variation with only five alleles observed among our population samples. The D2S441 locus showed the most variation with 12 alleles observed among the populations. The D22S1045 marker was the only STR having a trinucleotide repeat structure between the two miniplexes (Table 1). The stutter products for the D22S1045 locus were not observed to be any higher than the five other tetranucleotide markers in this study (data not shown).

All loci were observed to be in Hardy-Weinberg equilibrium with the exception of D14S1434 among African Americans (Table 5). Further characterization of the primer binding regions, using amplification and sequencing primers outside of the miniSTR amplicon (Table 3), will be necessary to identify potential mutations that create null alleles (3). Heterozygosity values were calculated for each locus and are listed in Table 5. The D10S441 marker showed the highest heterozygosity values among all of the population samples, while the D4S2364 marker revealed low levels of heterozygosity among the population samples (Table 5). Within the African American samples, we observed a heterozygosity value of 0.366 for the D4S2364 loci (Table 5).

A comparison of the heterozygosity values calculated by genotyping the same 474 samples using the commercially available Identifiler kit is listed in Table 6. Three of the miniplex loci (D10S1248, D22S1045, and D2S441) fell in the "middle" of the heterozygosity

TABLE 6—*Comparison of heterozygosity values from 474 individuals (164 African Americans, 170 Caucasians, and 140 Hispanics) using six miniSTR loci (denoted in bold and italics) and the 15 STR loci within the Identifiler™ kit (26). Heterozygosity values for the miniSTR loci are the average values from all of the groups presented separately in Table 5.*

Heterozygosity	Marker
0.8784	D2S1338
0.8753	D18S51
0.8710	FGA
0.8393	D21S11
0.8245	vWA
0.8076	D7S820
0.7970	D19S433
0.7759	***mD10S1248 - mini01***
0.7759	D16S539
0.7674	***mD22S1045 - mini01***
0.7674	D8S1179
0.7590	***mD2S441 - mini02***
0.7548	D3S1358
0.7526	D13S317
0.7463	***mD1S1677 - mini02***
0.7378	CSF1PO
0.7378	THO1
0.7294	D5S818
0.7146	TPOX
0.6765	***mD14S1434 - mini01***
0.5307	***mD4S2364 - mini02***

values of the Identifiler markers (Table 6). Two of the miniplex loci (D14S1434 and D4S2364) had the lowest observed heterozygosity values in the comparison (Table 6). However, it is important to note that one of the major benefits of these miniplexes are the small size amplicons that can be generated, increasing the likelihood of amplifying degraded material. Additionally, the heterozygosity values observed in D14S1434 and D4S2364 are greater than the heterozygosity values determined from a single biallelic single nucleotide polymorphism (SNP), which have a maximum attainable expected heterozygosity of 0.50 (35).

Sensitivity Studies

To evaluate the newly developed miniplexes, tests were conducted to assess the sensitivity of the markers. A template containing 1 ng/μL was used to create DNA dilutions of 500, 200, 100, 50, 20, 10, and 5 pg/μL. Each dilution was typed with miniplex01 and miniplex02 for ten replicates. Each replicate was compared to the "correct" genotype determined when the sample was analyzed with 1 ng/μL. Replicates were determined to have either: (1) the correct genotype, (2) a "partial" genotype which was observed when one allele in a heterozygote sample dropped below the interpretational threshold of 50 relative fluorescent units (RFUs), (3) an incorrect genotype which was observed whenever an allele "drop-in" creates a wrong genotype, or (4) a failure which was observed when no peaks were above the interpretational threshold of 50 RFU. Results from the marker D10S1248 are shown in Fig. 4.

Under "standard" conditions (28 cycles of PCR, 1 U AmpliTaq Gold polymerase, 25 uL PCR) all samples correctly typed to 100 pg (Fig. 4*A*). The average peak height RFU value at 100 pg was 165 RFUs for the 14 allele and 108 RFUs for the 17 allele in this particular sample (data not shown). Samples containing less than 100 pg of template started to show allele dropout or an incorrect genotype due to allele "drop-in" (Fig. 4*A*). Only 4 of 10 samples typed correctly at 20 pg, and none of the samples typed correctly at 10 pg or below (Fig. 4*A*).

Increasing the number of cycles (+4) and additional AmpliTaq Gold polymerase (2U) improved the percentage of "correct" allele calls in samples below 50 pg (Fig. 4*B*). One half of the samples containing 5 pg gave a partial profile. The mean peak height RFU values of the alleles were also greatly increased overall with the additional cycles and polymerase (Fig. 4*C*). For example, with the 100 pg samples the average RFU values for the 14 and 17 alleles were 2850 and 2300, respectively (Fig. 4*C*). Similar results were also obtained for the other miniplex loci (data not shown).

Increasing the number of PCR cycles and the addition of more DNA polymerase can allow more reliable testing of lower levels of DNA (36). However, it is important to be attentive of the potential interpretational problems associated with low copy number DNA (36–38).

Performance with Degraded DNA Samples

To evaluate the performance of the newly developed miniSTRs, tests were conducted on 16 bone samples provided from Bruce McCord's laboratory at Ohio University that were in various stages of decomposition from the University of Tennessee Forensic Anthropological Center and the Franklin County (Ohio) Coroner's Office. Typing was performed with the commercially available PowerPlex 16 kit and each miniSTR miniplex. Only one of the sixteen samples failed to type for each system due to a low amount of provided DNA template. For the remaining fifteen samples typed with the PowerPlex 16 kit, one sample produced a full profile. The remaining fourteen samples gave partial profiles, with the larger alleles, such as CSF1PO and the Penta loci missing or greatly reduced in signal (see Fig. 5*A* for a demonstration of the loss of the larger alleles using PowerPlex 16). All of the miniplex01 and miniplex02 loci were successfully typed for the fifteen bone samples (Figs. 5*B* and 5*C*).

Conclusions

This paper outlines the initial efforts to develop novel miniSTRs for increasing the power of discrimination of highly degraded DNA. We have developed two miniSTR triplexes with target amplicon sizes under 125 bp. In addition, 41 more candidate loci have been identified for further characterization in future studies. We have characterized the six loci D1S1677, D2S441, D4S2364, D10S1248, D14S1434, and D22S1045 among 474 U.S. population samples from African Americans, Caucasians, and Hispanics. All 474 samples were resolved from one another with the six loci. These new miniSTR markers had middle to low heterozygosity values when compared to commonly used STRs including the 13 core CODIS loci. The utility of these new loci were confirmed in comparing the success of the miniSTR assays for typing degraded bone samples while partial profiles were observed with the majority of the samples using the PowerPlex 16 kit at 32 cycles of PCR amplification. These loci were seen to be reliable and sensitive to at least 100 pg of DNA template under controlled laboratory conditions and pristine DNA samples. The careful evaluation of typing low copy number or degraded DNA should include an interpretation strategy for resulting data (36–39).

In addition to the usefulness of these miniSTRs for degraded DNA, we anticipate these new loci to be useful in a variety of scenarios involving closely related individuals where there is a need for additional genetic information (e.g., paternity cases and mass disaster reconstruction cases). MiniSTR assays may also provide

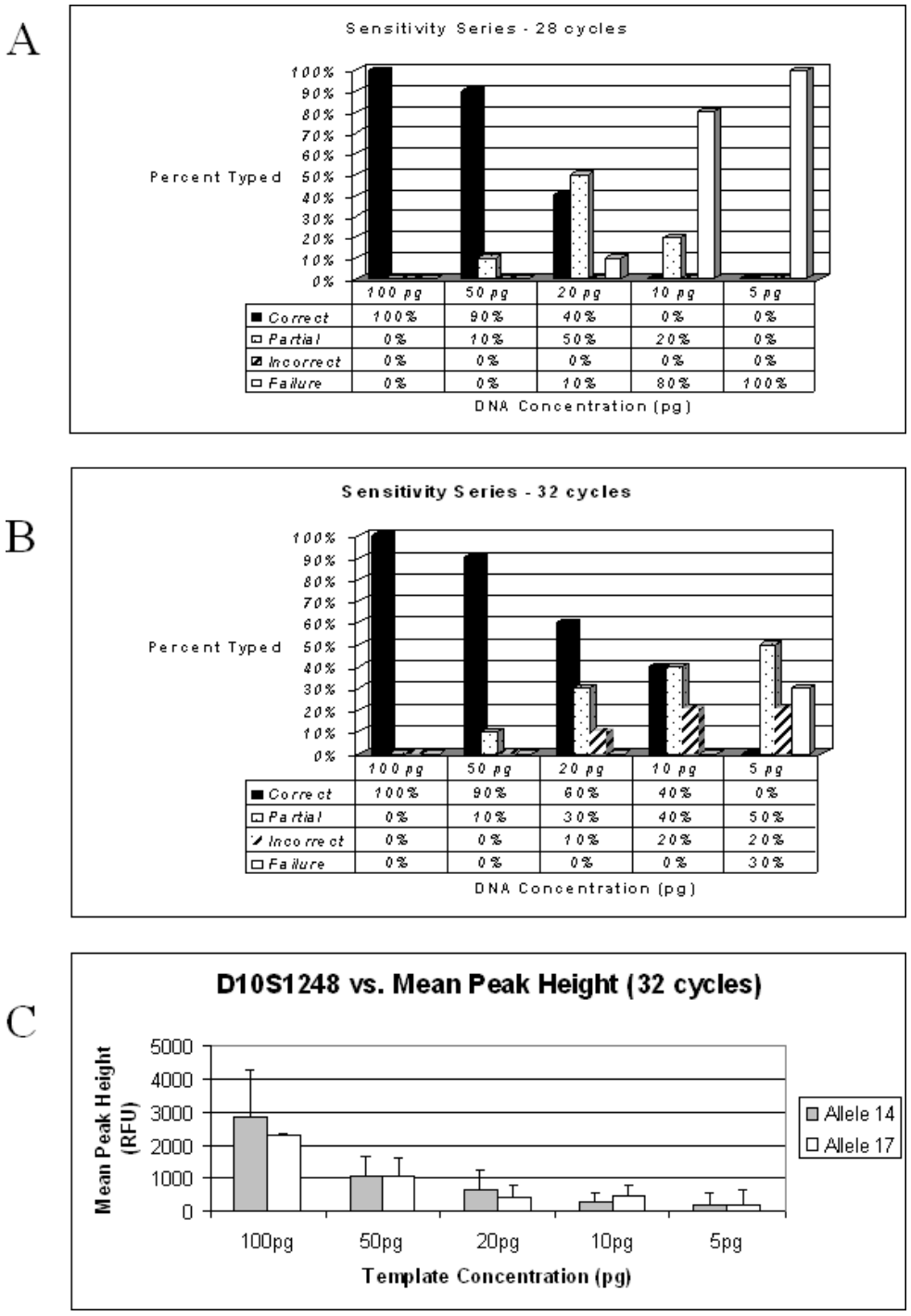

FIG. 4—*Sensitivity results for DNA dilutions using the D10S1248 marker. Results are from 10 replicate PCR amplifications and were analyzed on the ABI 3100. An interpretational threshold of 50 relative fluorescent units (RFUs) was used for genotyping each dilution. Each replicate was typed as being either correct (matching the genotype at 1 ng amplification), partial (one allele in a heterozygote drops below 50 RFUs), incorrect (allele drop-in that creates a wrong genotype), or failure (no peaks were observed above 50 RFUs). (A) Results with 28 cycles of PCR amplification and 1U of DNA polymerase. (B) Results with 32 cycles of PCR amplification and 2U of DNA polymerase. These results show an increase in the successful typing at low template levels. However, the number of partial profiles and profiles having the incorrect typing can be increased with extra samples and extra polymerase. (C) Mean Peak Height values from samples amplified at 32 cycles with 2U AmpliTaq Gold polymerase for the D10S1248 marker.*

a potential alternative to mitochondrial sequencing in the forensic analysis of degraded DNA. It is anticipated that these loci will be useful for the analysis of shed hair shafts, a common evidentiary material, as previous efforts to use miniSTRs to type telogen hairs have been successful (15). Finally, the small sizes of these miniplexes have great potential in application to high-speed microfluidic and mass spectrometric approaches. In general miniSTR assays, such as those described here, should offer a new potential tool for recovering useful information from samples that generated partial profiles with present STR multiplexes.

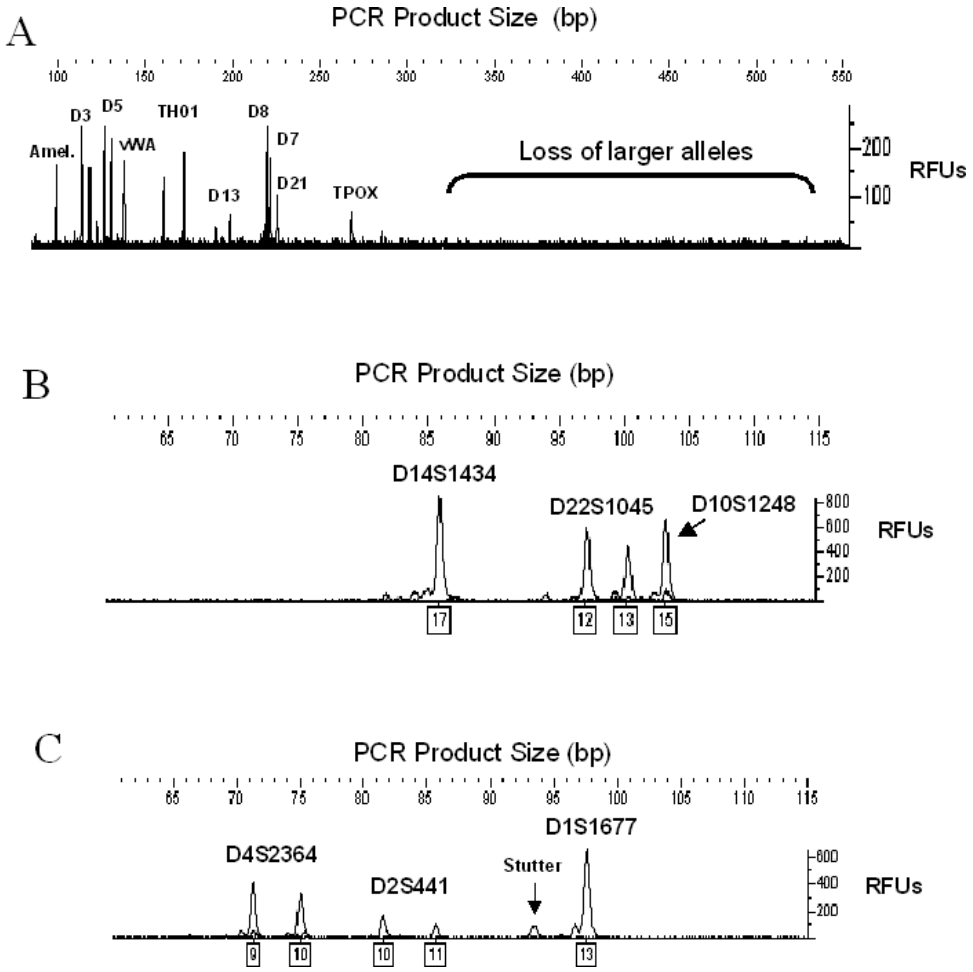

FIG. 5—*Comparison of STR allele peak heights obtained from (A) PowerPlex® 16, (B) miniplex01 and (C) miniplex02 loci with equivalent amounts of the same degraded bone sample (approximately 6 pg/μL quantified using real time PCR—Denise Chung, personal communication). All samples were amplified for 32 cycles of PCR with 2U of AmpliTaq Gold polymerase in a 25 μL volume. Peak heights are in relative fluorescence units (RFUs).*

Acknowledgments

The authors express appreciation for helpful discussions with Bruce McCord and the technical assistance of Peter Vallone, Margaret Kline, Jan Redman, Richard Schoske, and Jill Appleby in previous preparation of the DNA samples used in this study, and to Amy Decker for primer quality control. Denise Chung and Bruce McCord of Ohio University supplied the bone samples used. M. Coble was supported by a National Research Council, Research Associateship Program, Post-doctoral fellowship.

References

1. Lygo JE, Johnson PE, Holdaway DJ, Woodroffe S, Whitaker JP, Clayton TM, et al. The validation of short tandem repeat (STR) loci for use in forensic casework. Int J Legal Med 1994;107:77–89.
2. Budowle B, Moretti TR, Baumstark AL, Defenbaugh DA, Keys KM. Population data on the thirteen CODIS core short tandem repeat loci in African Americans, U.S. Caucasians, Hispanics, Bahamians, Jamaicans, and Trinidadians. J Forensic Sci 1999;44(6):1277–86.
3. Butler JM. Forensic DNA typing: biology and technology behind STR markers. London: Academic Press, 2001.
4. Bar W, Kratzer A, Machler M, Schmid W. Postmortem stability of DNA. Forensic Sci Int 1988;39(1):59–70.
5. Whitaker JP, Clayton TM, Urquhart AJ, Millican ES, Downes TJ, Kimpton CP, Gill P. Short tandem repeat typing of bodies from a mass disaster: high success rate and characteristic amplification patterns in highly degraded samples. BioTechniques 1995;18(4):670–7.
6. Krenke BE, Tereba A, Anderson SJ, Buel E, Culhane S, Finis CJ, et al. Validation of a 16-locus fluorescent multiplex system. J Forensic Sci 2002;47(4):773–85.
7. Holt CL, Buoncristiani M, Wallin JM, Nguyen T, Lazaruk KD, Walsh PS. TWGDAM validation of AmpFlSTR PCR amplification kits for forensic DNA casework. J Forensic Sci 2002;47(1):66–96.
8. Butler JM, Levin BC. Forensic applications of mitochondrial DNA. Trends Biotech 1998;16:158–62.
9. Holland MM, Parsons TJ. Mitochondrial DNA sequence analysis—validation and use for forensic casework. Forensic Sci Rev 1999;11: 21–50.
10. Wiegand P, Kleiber M. Less is more—length reduction of STR amplicons using redesigned primers. Int J Legal Med 2001;114(4–5): 285–7.
11. Tsukada K, Takayanagi K, Asamura H, Ota M, Fukushima H. Multiplex short tandem repeat typing in degraded samples using newly designed primers for the TH01, TPOX, CSF1PO, and vWA loci. Legal Med 2002;4:239–45.
12. Sparkes R, Kimpton CP, Watson S, Oldroyd NJ, Clayton TM, Barnett L, et al. The validation of a 7-locus multiplex STR test for use in forensic casework: (I) mixtures, ageing, degradation and species studies. Int J Legal Med 1996;109:186–94.

13. Yoshida K, Sekiguchi K, Kasai K, Sato H, Seta S, Sensabaugh GF. Evaluation of new primers for CSF1PO. Int J Legal Med 1997;110(1): 36–8.
14. Szibor R, Kayser M, Roewer L. Identification of the human Y-chromosomal microsatellite locus DYS19 from degraded DNA. Am J Forensic Med Pathol 2000;21(3):252–4.
15. Hellmann A, Rohleder U, Schmitter H, Wittig M. STR typing of human telogen hairs–a new approach. Int J Legal Med 2001;114(4–5): 269–73.
16. Butler JM, Li J, Shaler TA, Monforte JA, Becker CH. Reliable genotyping of short tandem repeat loci without an allelic ladder using time-of-flight mass spectrometry. Int J Legal Med 1998;112(1):45–49.
17. Butler JM, Becker CH. Improved analysis of DNA short tandem repeats with time-of-flight mass spectrometry. Washington, DC: Science and Technology Research Report, National Institute of Justice. 2001 Oct.; see http://www.ojp.usdoj.gov/nij/pubs-sum/188292.htm
18. Schmalzing D, Koutny L, Adourian A, Belgrader P, Matsudaira P, Ehrlich D. DNA typing in thirty seconds with a microfabricated device. Proc Natl Acad Sci USA 1997;94:10273–8.
19. Budowle B, Moretti TR, Niezgoda SJ, Brown BL. CODIS and PCR-based short tandem repeat loci: law enforcement tools. Proceedings of the Second European Symposium on Human Identification, Innsbruck, Austria, June 10–12, 1998;73–88; http://www.promega.com/geneticidproc/eusymp2proc/17.pdf
20. Butler JM, Shen Y, McCord BR. The development of reduced size STR amplicons as tools for analysis of degraded DNA. J Forensic Sci 2003;48(5):1054–64.
21. Gill P, Werrett DJ, Budowle B, Gurrier R. An assessment of whether SNPs will replace STRs in national DNA databases–joint considerations of the DNA working group of the European Network of Forensic Science Institutes (ENFSI) and the Scientific Working Group on DNA Analysis Methods (SWGDAM). Sci Justice 2004;44(1):51–3.
22. Goodwin W, Ballard D, Simpson K, Thacker C, Syndercombe Court D, Gow J. Case study: paternity testing—when 21 loci are not enough. In: Doutremepuich C, Morling, N, editors. Progress in forensic genetics 10. Amsterdam, The Netherlands: International Congress Series. 2004;1261:460–2.
23. Ghebranious N, Vaske D, Yu A, Zhao C, Marth G, Weber JL. STRP Screening sets for the human genome at 5cM density. BMC Genomics 2003;4(6):1–10.
24. Rozen S, Skaletsky HJ. Primer3 (primer 3_www.cgi v 0.2); 1998. Code available at http://www.genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi.
25. Vallone PM, Butler JM. AutoDimer: A screening tool for primer-dimer and hairpin structures. Biotechniques 2004;37(2):226–31.
26. Butler JM, Schoske R, Vallone PM, Redman JW, Kline MC. Allele frequencies for 15 autosomal STR loci on U.S. Caucasian, African American, and Hispanic populations. J Forensic Sci 2003;48(4):908–11.
27. Chakraborty R, Stivers D, Zhong Y, CHG. DNATYPE, (Windows 95/NT ver.). University of Texas, Houston, TX.
28. Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, Zhivotovsky LA, Feldman MW. Genetic structure of human populations. Science 2002;298(5602):2381–5.
29. Rosenberg NA, Li LM, Ward R, Pritchard JK. Informativeness of genetic markers for inference of ancestry. Am J Hum Genet 2003;73(6):1402–22.
30. Ayub Q, Mansoor A, Ismail M, Khaliq S, Mohyuddin A, Hameed A, et al. Reconstruction of human evolutionary tree using polymorphic autosomal microsatellites. Am J Phys Anthropol 2003;122(3):259–68.
31. Klintschar M, Immel U, Stiller D, Kleiber M. HumTH01 and blood pressure. An obstacle for forensic application? In: Doutremepuich C, Morling, N, editors. Progress in forensic genetics 10. Amsterdam, The Netherlands. International Congress Series, 2004;1261:589–91.
32. Sharma P, Hingorani A, Jia H, Ashby M, Hopper R, Clayton D, Brown MJ. Positive association of tyrosine hydroxylase microsatellite marker to essential hypertension. Hypertension 1998;32(4):676–82.
33. Butler JM, Li J, Monforte JA, Becker CH. DNA typing by mass spectrometry with polymorphic DNA repeat markers. 2000. US Patent 6,090,558.
34. Bar W, Brinkmann B, Budowle B, Carracedo A, Gill P, Lincoln P, Mayr W, Olaisen B. DNA recommendations. Further report of the DNA commission of the ISFH regarding the use of short tandem repeat systems. International Society for Forensic Haemogenetics. Int J Legal Med 1997;110(4):175–6.
35. Stephens JC, Schneider JA, Tanguay DA, Choi J, Acharya T, Stanley SE, et al. Haplotype variation and linkage disequilibrium in 313 human genes. Science 2001;293(5529):489–93.
36. Gill P, Whitaker J, Flaxman C, Brown N, Buckleton J. An investigation of the rigor of interpretation rules for STRs derived from less than 100 pg of DNA. Forensic Sci Int 2000;112(1):17–40.
37. Taberlet P, Griffin S, Goossens B, Questiau S, Manceau V, Escaravage N, et al. Reliable genotyping of samples with very low DNA quantities using PCR. Nucleic Acids Res 1996;24(16):3189–94.
38. Whitaker JP, Cotton EA, Gill P. A comparison of the characteristics of profiles produced with the AmpF/STR SGM Plus multiplex system for both standard and low copy number (LCN) STR DNA analysis. Forensic Sci Int 2001;123(2–3):215–23.
39. Schneider PM, Bender K, Mayr WR, Parson W, Hoste B, Decorte R, et al. STR analysis of artificially degraded DNA-results of a collaborative European exercise. Forensic Sci Int 2004;139(2–3):123–34.

Additional information and reprint requests:
Michael D. Coble, Ph.D.
Biotechnology Division
National Institute of Standards and Technology
100 Bureau Dr., Mail Stop 8311
Gaithersburg, MD 20899-8311
Tel: 301-975-4330
Fax: 301-975-8505
E-mail: michael.coble@nist.gov