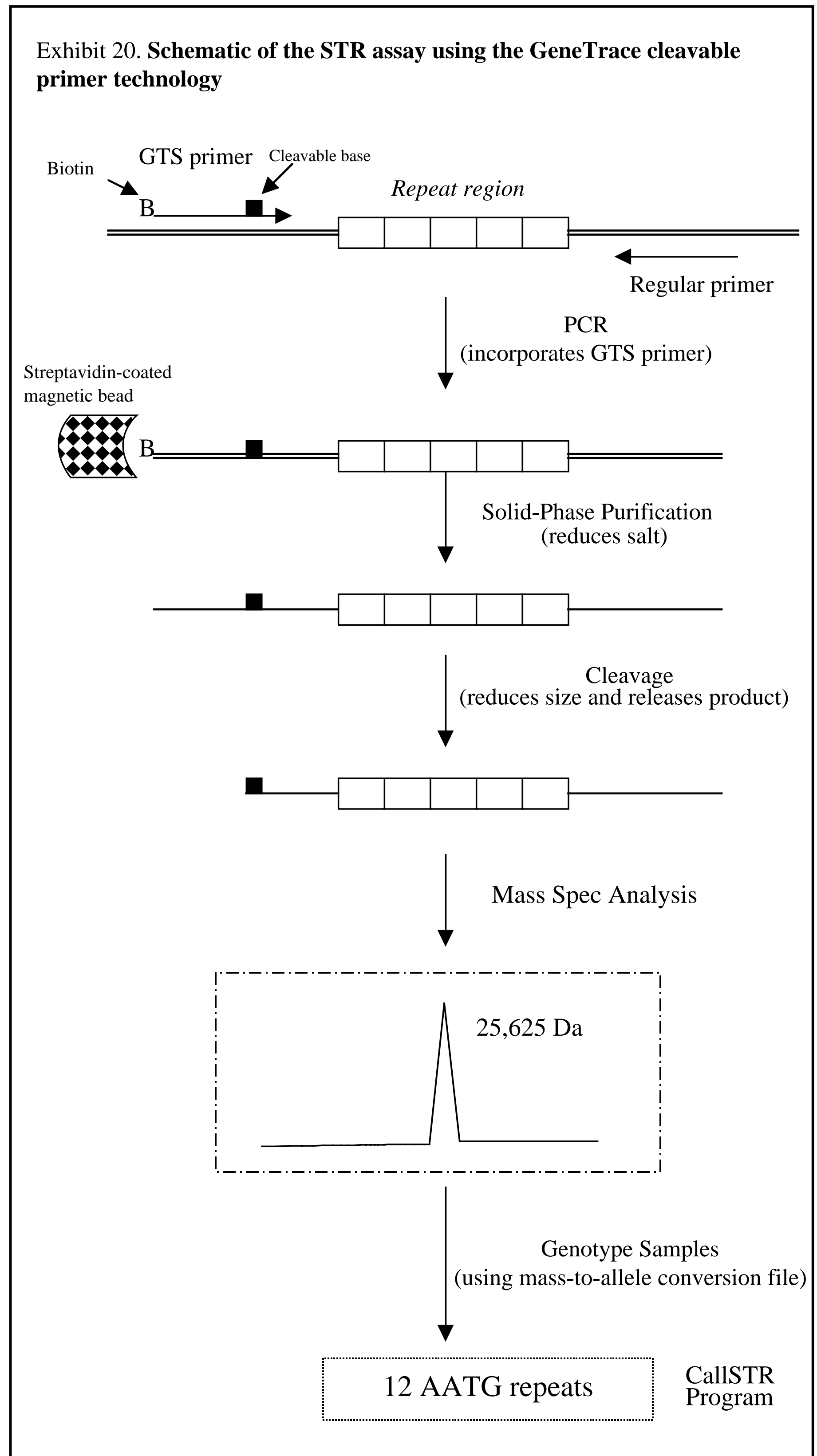
Scope and Methodology

Both the STR and the SNP genotyping assays used in this project involve the same fundamental (proprietary) sample preparation chemistry. This chemistry was important for salt reduction/removal prior to the mass spectrometric analysis and was automated on a 96-tip robotic workstation. A biotinylated, cleavable oligonucleotide was used as a primer in each assay and was incorporated through standard DNA amplification (i.e., PCR) methodologies into the final product, which was measured in the mass spectrometer. This process was covered by U.S. Patent 5,700,642, which was issued in December 1997, and is described in more detail in U.S. Patent 6,090,588 (Butler et al., 2000). The STR assay is schematically illustrated in exhibit 20 and involves a PCR amplification step where one of the primers is replaced by the GeneTrace cleavable primer. The biotinylated PCR product was then captured on streptavidin-coated magnetic beads for post-PCR sample cleanup and salt removal followed by mass spectrometry analysis.

The biology portion of the SNP assay, on the other hand, involves a three-step process: (1) PCR amplification, (2) phosphatase removal of nucleotides, and (3) primer extension, using the GeneTrace cleavable primer, with dideoxynucleotides for single-base addition of the nucleotide(s) complementary to the one(s) at the SNP site (Li et al., 1999). The SNP assay is illustrated in exhibit 21.

Simultaneous analysis of multiple SNP markers (i.e., multiplexing) is possible by simply putting the cleavage sites at different positions in the various primers so they do not overlap on a mass scale. Also important to both genotyping assays is proprietary calling software that was developed (and evolved) during the course of this work. A number of STR and SNP markers were developed and tested with a variety of human DNA samples as part of this project to demonstrate the feasibility of this mass spectrometry approach.

Exhibit 20. **Schematic of the STR assay using the GeneTrace cleavable primer technology**

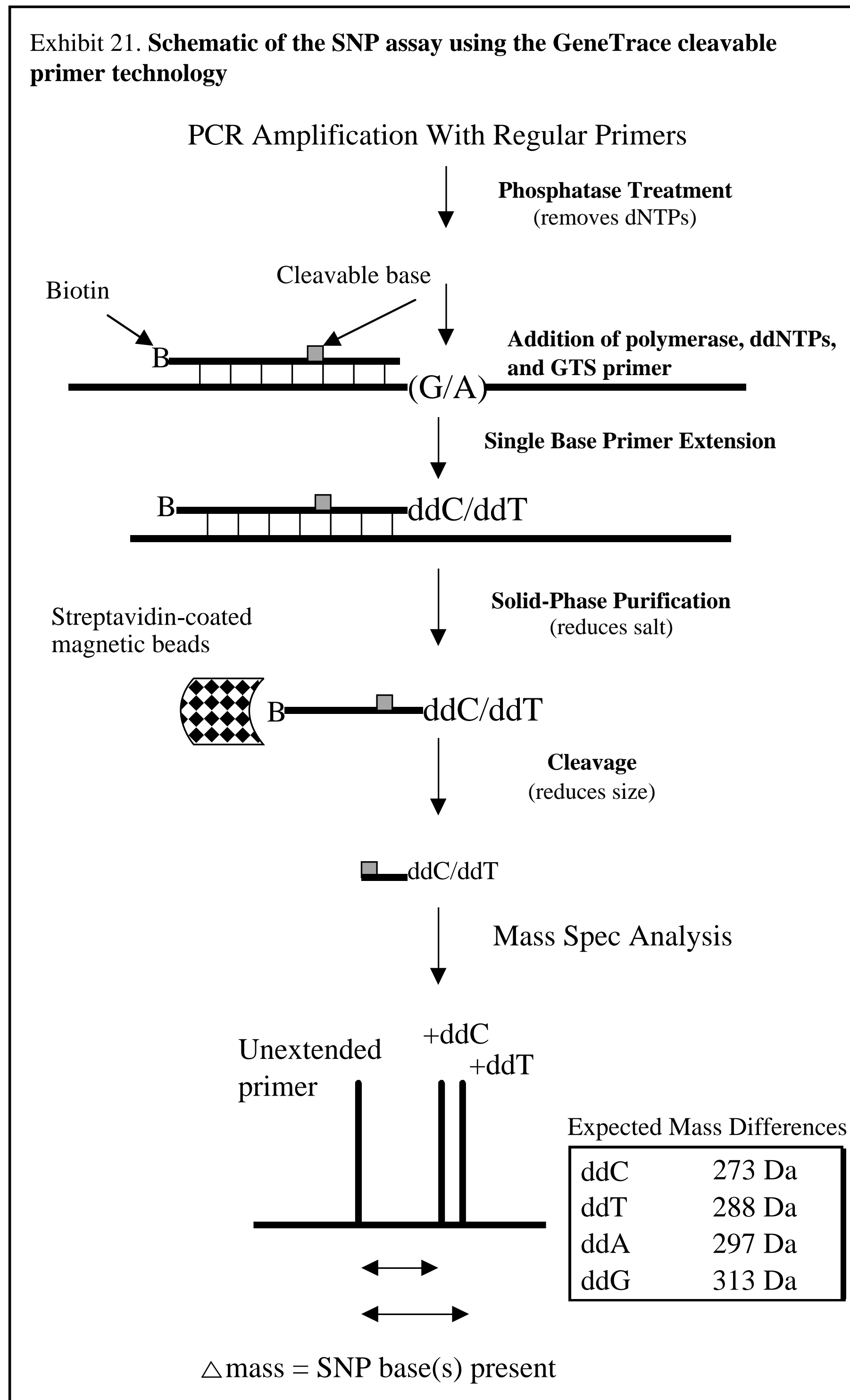


Assay Development and Primer Testing

Primer design

Primers were initially designed for each STR locus using Gene Runner software (Hastings Software, Inc., Hastings, NY) and then more recently with Primer 3 version 0.2 from the World Wide Web (*http://www-genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi*) (Rozen and Skaletsky, 1998). Multiplex PCR primers for the multiplex SNP work were designed with a UNIX version of Primer 3 (release 0.6) adapted at GeneTrace to utilize a mispriming library and Perl scripts for input of sequences and export of primer information.

Exhibit 21. **Schematic of the SNP assay using the GeneTrace cleavable primer technology**



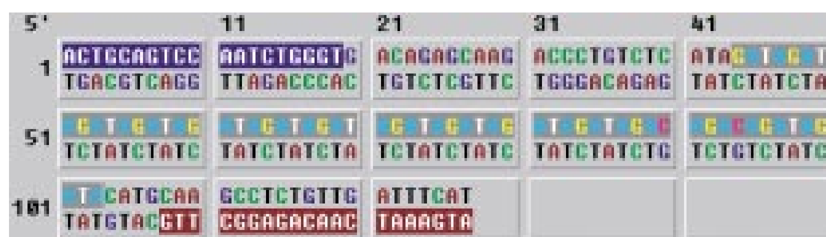
DNA sequence information was obtained from GenBank (*http://www.ncbi.nlm.nih.gov*) and STRBase (*http://www.cstl.nist.gov/biotech/strbase*) for the STR loci and mtDNA and from Dr. Peter Underhill of Stanford University for the Y-chromosome SNPs. These sequences served as the reference sequence for primer design and, in the case of STRs, the calibrating mass for the genotyping software (see below). When possible, primers were placed close to the repeat region to make the PCR product size ranges under 120 bp to improve the sensitivity and resolution in the mass spectrometer (exhibit 2). Previously published primers were used in the case of amelogenin (Sullivan et al., 1993), D3S1358 (Li et al., 1993), CD4 (Hammond et al., 1994), and VWA (Fregeau and Fourney, 1993) because their PCR product sizes were analyzable in the mass spectrometer or the amplicons could be reduced in size following the PCR step (see below). Later, D3S1358 experiments were performed with primers that produced smaller products after sequence information became available for that particular STR locus (exhibit 22).

Primer synthesis

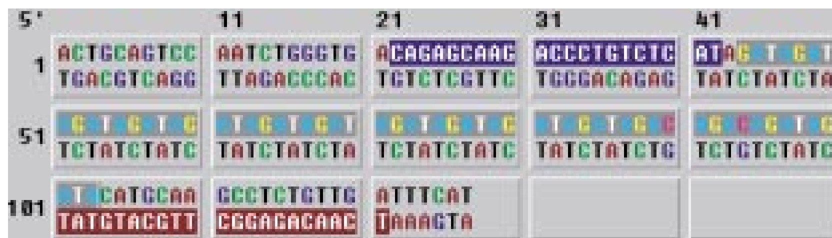
Unmodified primers were purchased from Biosource/Keystone (Foster City, CA) or Operon Technologies (Alameda, CA) or synthesized in-house using standard solid-phase phosphoramidite chemistry. The GeneTrace cleavable primers were synthesized in-house using a proprietary phosphoramidite that was incorporated near the 3' end of the oligonucleotide along with a biotin attached at the 5' end. Primers were quality control tested via mass spectrometry prior

Exhibit 22. **D3S1358 sequence with PCR primer locations.** This STR sequence was not publicly available in GenBank at the start of this project but was obtained as part of this work. The forward primer is shown in blue and the reverse primer in red, with the GATA repeat region shaded grey. Section (A) shows the primer locations compared with the repeat region for a primer pair originally published by Li and colleagues (1993). Section (B) shows the primers designed as part of this project and their positions relative to the repeat region. The overall size of PCR products was reduced by 27 bp (more than 8,000 Da) compared with an amplicon generated from the previously published primer set.

(A) Previously Published Primers: PCR product size = 127 bp for 15 repeats



(B) Newly Designed Primers: PCR product size = 100 bp for 15 repeats



to further testing to confirm proper synthesis and to determine the presence or absence of failure products. Synthesis failure products (i.e., n-1, n-2, etc.) can especially interfere with multiplex SNP analysis. The cleavable base is stable during primer synthesis and PCR amplification. Comparisons of regular primers with cleavable primers containing the same base sequence showed no significant difference, indicating that the primer annealing is not compromised by the cleavable base.

Methods for STR product size reduction

It was discovered early in the study that the PCR primer opposite the biotinylated cleavable primer could be moved into the repeat region as much as two full repeat units to reduce the overall size without severely compromising the PCR reaction. For the cleavable primer, the cleavable base was typically placed in the second or third position from the 3' end of the primer in order to remove as much of the modified primer as possible. Thus, the cleavage step reduces the overall PCR product size by the length of the cleavable primer minus two or three nucleotides. Typically, this size reduction is approximately 20 bases.

The portion of the DNA product on the other side of the repeat region from the cleavable primer was removed in one of two possible ways: using a restriction enzyme (Monforte et al., 1999) or performing a nested linear amplification with a terminating nucleotide (Braun et al., 1997a and 1997b), such as dideoxynucleotide (ddN). These methods work only for particular situations (see Results and Discussion). Almost all singleplex STR work was performed without either of these product size reduction methods. However, these size reduction methods played a role in the multiplex STR work.

Multiplex design

STR multiplexes were designed by construction of virtual allelic ladders or "mass simuplexes" that involved the predicted mass of all known alleles for a particular locus. STR markers were then interleaved based on mass with all alleles between loci being distinguishable (exhibit 4). STR multiplexes work best if alleles are below 20,000–25,000 Daltons (Da) in mass due to the improved sensitivity and resolution that is obtainable in the mass spectrometer. As previously described in the section on size reduction, a restriction enzyme or a ddN terminator may be used to shorten the STR allele sizes. For multiplex design, locating a restriction enzyme with cut sites common to all STR loci involved in the multiplex complicates the design process and limits the choice of possible marker combinations. The use of a common dideoxynucleotide terminator is much easier. For example, with the STR loci CSF1PO, TPOX, and TH01, a multiplex was developed using a dideoxycytosine (ddC) terminator and primer extension along the AATG strand (exhibits 4 and 5).

SNP multiplexes were designed by calculating possible postcleavage primer and extension product masses. Multiply charged ions were abundant in the mass range of 1,500–7,000 Da in SNP multiplex analyses, which were avoided for the most part by calculating interfering doubly charged and triply charged ions. The cleavage sites for candidate multiplex SNP primers were chosen for the least amount of overlap between singly and multiply charged ions (exhibits 23).

Human DNA samples used

Human genomic DNA samples representing several ethnic groups (African-American, European, and Asian) were purchased from Bios Laboratories (New Haven, CT) for the initial studies. K562 cell line DNA (Promega) was used as a control sample since the genotypes for this cell line were reported in most of the STR loci (*GenePrint*™ STR Systems Technical Manual, 1995).

Allelic ladders were reamplified from a 1:1000 dilution of each of the allelic ladders supplied in fluorescent STR kits from ABI using the PCR conditions listed below and the primers shown in exhibit 24. The ABI kits included allelic

Exhibit 23. **Expected mass-to-charge ratios for various ions in the mtDNA 10-plex assay**

Name	Primer Mass	Singly Charged Ions				Primer Mass	Doubly Charged Ions				Primer Mass	Triply Charged Ions			
		ddC	ddT	ddA	ddG		ddC	ddT	ddA	ddG		ddC	ddT	ddA	ddG
MT5	**1,580**			1,877	1,893	790			939	947	527			626	631
MT8′	**1,790**	2,063	2,078			895	1,032	1,039			597	688	693		
MT10	**2,785**	3,058	3,073			1,393	1,529	1,537			928	1,019	1,024		
MT3′	**3,179**			3,476	3,492	1,590			1,738	1,746	1,060			1,159	1,164
MT9	**3,740**			4,037	4,053	1,870			2,019	2,027	1,247			1,346	1,351
MT6	**4,355**			4,652	4,668	2,178			2,326	2,334	1,452			1,551	1,556
MT2g	**4,957**	5,230	5,245			2,479	2,615	2,623			1,652	1,743	1,748		
MT1	**5,375**			5,672	5,688	2,688			2,836	2,844	1,792			1,891	1,896
MT7	**5,891**	6,164	6,179			2,946	3,082	3,090			1,964	2,055	2,060		
MT4e	**6,500**	6,773	6,788			3,250	3,387	3,394			2,167	2,258	2,263		

Exhibit 24. **Primer sequences designed for STR markers tested by mass spectrometry.** For the PCR product size produced with these primers, see exhibit 2. Note: "b" is listed for biotin; parentheses ( ) indicate the cleavable base.

STR Locus	Primer Sequences for GTS Mass Spec Analysis	Primer Name	STR Locus	Primer Sequences for GTS Mass Spec Analysis	Primer Name
Amelogenin	5'-b-CCCTGGGCTCTGTAAAGAATAG(T)G-3' 5'-ATCAGAGCTTAAACTGGGAAGCTG-3'	AMEL-F AMEL-R	D21S11	5'-b-CCCAAGTGAATTGCCTTC(T)A-3' 5'-GTAGATAGACTGGATAGATAGACGATAGA-3'	D21-F D21-R
CD4	5'-b-TTGGAGTCGCAAGCTGAAC(T)AGC-3' 5'-GCCTGAGTGACAGAGTGAGAACC-3'	CD4-F CD4-R	F13A1	5'-b-CAGAGCAAGACTTCATC(T)G-3' 5'-TCATTTTAGTGCATGTTC-3'	F13A1-F F13A1-R
CSF1PO	5'-ACAGTAACTGCCTTCATAGA(T)AG-3' 5'-b-GTGTCAGACCCTGTTCTAAGTA-3'	CSF-F3 CSF-R3	FES/FPS	5'-b-TTAGGAGACAAGGATAGCAGT(T)C-3' 5'-GCGAAAGAATGAGACTACATCT-3'	FES-F2 FES-R2
D3S1358	5'-b-CAGAGCAAGACCCTGTC(T)CAT-3' 5'-TCAACAGAGGCTTGCATGTAT-3'	D3-F2 D3-R2	FGA	5'-b-AAAATTAGGCATATTTACAAGCTAG(T)T-3' 5'-TCTGTAATTGCCAGCAAAAAAGAAA-3'	FGA-F FGA-R
D5S818	5'-b-CTCTTTGGTATCCTTATGTAATA(T)T-3' 5'-ATCTGTATCCTTATTTATACCTCTATCTA-3'	D5-F D5-R	HPRTB	5'-b-GTCTCCATCTTTGTCTCTATCTCTATC(T)G-3' 5'-GAGAAGGGCATGAATTTGCTTT-3'	HPRTB-F HPRTB-R
D7S820	5'-b-TGTCATAGTTTAGAACGAAC(T)AAC-3' 5'-AAAAACTATCAATCTGTCTATCTATC-3'	D7-F D7-R	LPL	5'-b-CTGACCAAGGATAGTGGGATA(T)AG-3' 5'-GGTAACTGAGCGAGACTGTGTCT-3'	LPL-F LPL-R
D8S1179	5'-b-TTTGTATTTCATGTGTACATTCGTA(T)C-3' 5'-ACCTATCCTGTAGATTATTTTCACTGTG-3'	D8-F D8-R	TH01	5'-CCTGTTCCTCCCTTATTTCCC-3' 5'-b-GGGAACACAGACTCCATGG(T)G-3'	TH01-F TH01-R
D13S317	5'-b-CCCATCTAACGCCTATCTGTA(T)T-3' 5'-GCCCAAAAAGACAGACAGAAAG –3'	D13-F D13-R2	TPOX	5'-b-CTTAGGGAACCCTCACTGAA(T)G-3' 5'-b-GTCCTTGTCAGCGTTTATTTGC-3'	TPOX-F TPOX-R
D16S539	5'-b-ATACAGACAGACAGACAGG(T)G-3' 5'-GCATGTATCTATCATCCATCTCT-3'	D16-F4 D16-R4	VWA	5'-b-CCCTAGTGGATGATAAGAATAATCAGTATG-3' 5'-b-GGACAGATGATAAATACATAGGATGGA(T)GG-3'	VWA-F VWA-R
D18S51	5'-TGAGTGACAAATTGAGACCTT-3' 5'-b-GTCTTACAATAACAGTTGCTACTA(T)T-3'	D18-F D18-R			

ladders for the following STR loci: AmpF1STR® Green I (CSF1PO, TPOX, TH01, amelogenin), AmpF1STR® Blue (D3S1358, VWA, FGA), AmpF1STR® Green II (amelogenin, D8S1179, D21S11, D18S51), AmpF1STR® Yellow (D5S818, D13S317, D7S820), and AmpF1STR® COfiler™ (amelogenin, TH01, TPOX, CSF1PO, D3S1358, D16S539, D7S820).

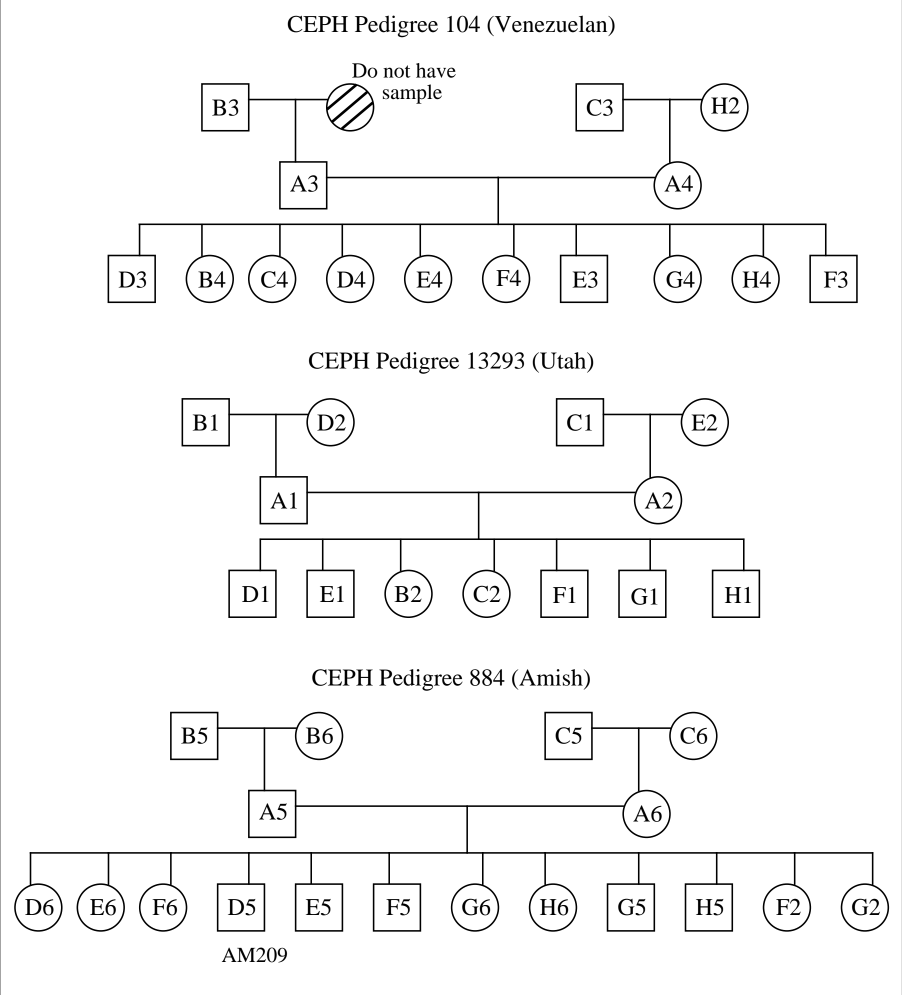
While most PCR amplifications were performed with quantitated genomic DNA in liquid form, a few were tested with blood-stained FTA™ paper (Life Technologies, Rockville, MD). Sample punches were removed from the dried FTA™ paper card with a 1.2 mm Harris MICRO-PUNCH™ (Life Technologies). The recommended washing protocol of 200 µL was reduced to 25 or 50 µL in order to reduce reagent costs and to work with volumes that are compatible with 96- or 384-well sample plates. The number of washes was kept the same as recommended by the manufacturer, but deionized water was used instead of a 10 mM Tris-EDTA solution.

Two studies were performed with larger numbers of DNA samples. In collaboration with Dr. Steve Lee and Dr. John Tonkyn from CDOJ's DNA research laboratory, CDOJ provided a plate of 88 samples, which was used repeatedly for multiple STR markers. These anonymous samples had been previously genotyped by CDOJ using ABI's AmpF1STR® Profiler™ kit, which consists of AmpF1STR® Blue, Green I, and Yellow markers and amplifies 9 STRs and the sex-typing marker amelogenin. STR allelic ladders were also provided by CDOJ and were used to illustrate that the common alleles for each STR locus could be detected with GeneTrace primer sets. Researchers retyped the samples using the AmpF1STR® COfiler™ fluorescent STR kit, which contains 6 STRs and amelogenin (5 of the 6 STR loci overlap with Profiler loci), thereby providing a further validation of each sample's true genotype. More recently, a set of 92 human DNA templates containing 3 different Centre d'Etude du Polymorphisme Humain (CEPH) families (exhibit 25) and 44 unrelated individuals from the NIH Polymorphism Discovery Resource were examined (Collins et al., 1998). These samples were typed on the ABI 310 Genetic Analyzer using both the AmpF1STR® Profiler Plus™ and AmpF1STR® COfiler™ kits so that all 13 CODIS STRs were covered.

PCR reaction

To speed the development of new STR markers, researchers worked toward the development of universal PCR conditions, in terms of both thermal cycling parameters and reagents used. Since almost all amplifications were singleplex PCRs, development effort was much simpler than multiplex PCR development. Generally, all PCR reactions were performed in 20 µL volumes with 20 pmol (1 µM) both forward and reverse primers and a PCR reaction mix containing everything else. The

Exhibit 25. **CEPH family pedigrees for samples examined.** The 96 well positions are indicated for each sample. These samples were typed at 13 STRs and amelogenin using Profiler Plus™ and COfiler™ fluorescent STR kits. The samples were also typed at 8 STRs and amelogenin using the GeneTrace mass spectrometry primers.



early PCR reaction mix contained 1 U Taq polymerase (Promega); 1X STR buffer with deoxynucleotide triphosphates (dNTPs) (Promega); and typically 5, 10, or 25 ng of human genomic DNA. Later in the study, a PCR mix containing 200 µM dNTPs, 50 mM KCl, 10 mM Tris-HCl, 5% glycerol, and 2 mM $MgCl_2$ was used. Typically, a locus-specific master mix was prepared by the addition of 12.8 µL of PCR mix times the number of samples (+ ~10% overfill) with 0.2 µL AmpliTaq Gold™ DNA polymerase (ABI) and the appropriate volume and quantity of forward and reverse primers to bring them to a concentration of 1 µM in each reaction. PCR reactions in a 96- or 384-well format were set up manually with an 8-channel pipettor or robotically with a Hamilton 16-tip robot.

Thermal cycling was performed in 96- or 384-well MJ Research DNA Engine (MJ Research, Watertown, MA) or 96 or dual block 384 PE9700 (ABI) thermal cyclers.

Initial thermal cycling conditions with Taq polymerase (Promega) were as follows:

94 °C for 2 min
35 cycles:
  94 °C for 30 sec
  50, 55, or 60 °C for 30 sec
  72 °C for 30 sec
72 °C for 5 or 15 min
4 °C hold

The final incubation at 72 °C favors nontemplated nucleotide addition (Clark 1988, Kimpton et al., 1993). This final incubation temperature was dropped to 60 °C for some experiments in an effort to drive the nontemplated addition even further. Later experiments, including all of the larger sample sets, involved using the following thermal cycling program with TaqGold DNA polymerase:

95 °C for 11 min (to activate the TaqGold DNA polymerase)
40 cycles:
  94 °C for 30 sec
  55 °C for 30 sec
  72 °C for 30 sec
60 °C for 15 min
4 °C hold

Primers were typically designed to have an approximate annealing temperature of 57–63 °C and thus worked well with a 55 °C anneal step under this "universal" thermal cycling protocol. The need for extensive optimization of primer sets, reaction components, or cycling parameters was greatly reduced or eliminated with this approach for primer development on STR markers. Mitochondrial DNA samples were amplified with 35 cycles and an annealing temperature of 60 °C using the PCR primers listed in exhibit 26.

Multiplex PCR

Multiplex PCR was performed using a universal primer tagging approach (Shuber et al., 1995; Ross et al., 1998b) and the following cycling program:

95 °C for 10 min
50 cycles:
  94 °C for 30 sec
  55 °C for 30 sec
  68 °C for 60 sec
72 °C for 5 min
4 °C hold

The PCR master mix contained 5 mM $MgCl_2$, 2 U AmpliTaq Gold with 1X

Exhibit 26. **Mitochondrial DNA primers used for 10-plex SNP reaction**

Primer Name	SNP Site Position	SNP Base*	Sequence (5'→3')	Cleaved Mass (Da)
MT5	H16224	A/G	b-GGAGTTGCAGTTGATGTGTGA(T)AGTTG	1,580
MT8'	L00146	T/C	b-GTCGCAGTATCTGTCTTTGAT(T)CCTGCC	1,790
MT10	H00247	C/T	b-CTGTGTGGAAAGTGGCTC(T)GCAGACATT	2,785
MT3'	H16189	A/G	b-GGTTGATTGCTGTACTTGCTTG(T)AAGCATGGGG	3,179
MT9	H00152	A/G	b-CTGTAATATTGAACGTAGG(T)GCGATAAATAAT	3,740
MT6	H16311	A/G	b-GTGCTATGTACGG(T)AAATGGCTTTATGT	4,355
MT2g	H16129	C/T	b-GTACTACAGGTGG(T)CAAGTATTTATGGTAC	4,957
MT1	H16069	G/A	b-AAATACA(T)AGCGGTTGTTGATGGGT	5,375
MT7	H00073	T/C	b-CCAGCGTC(T)GCGAATGCTATCGCGTGCA	5,891
MT4e	L00195	T/C	b-CTACGT(T)CAATATTACAGGCGAACATAC	6,500
DLOOP-F1**	L15997		CACCATTAGCACCCAAAGCT	
DLOOP-R1**	H00401		CTGTTAAAAGTGCATACCGCCA	

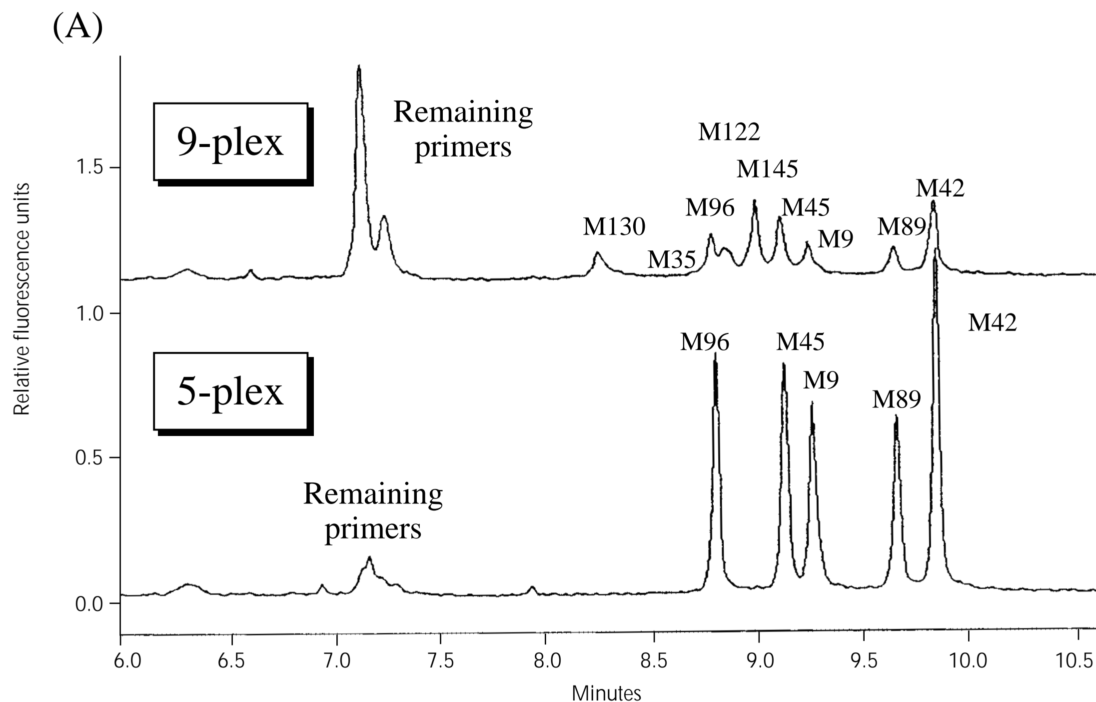
Note: "b" is listed for biotin; parentheses ( ) indicate the cleavable base.
*Anderson reference sequence base listed first.
**PCR Primers produced a 1,021 bp PCR product spanning positions 15977–00422.

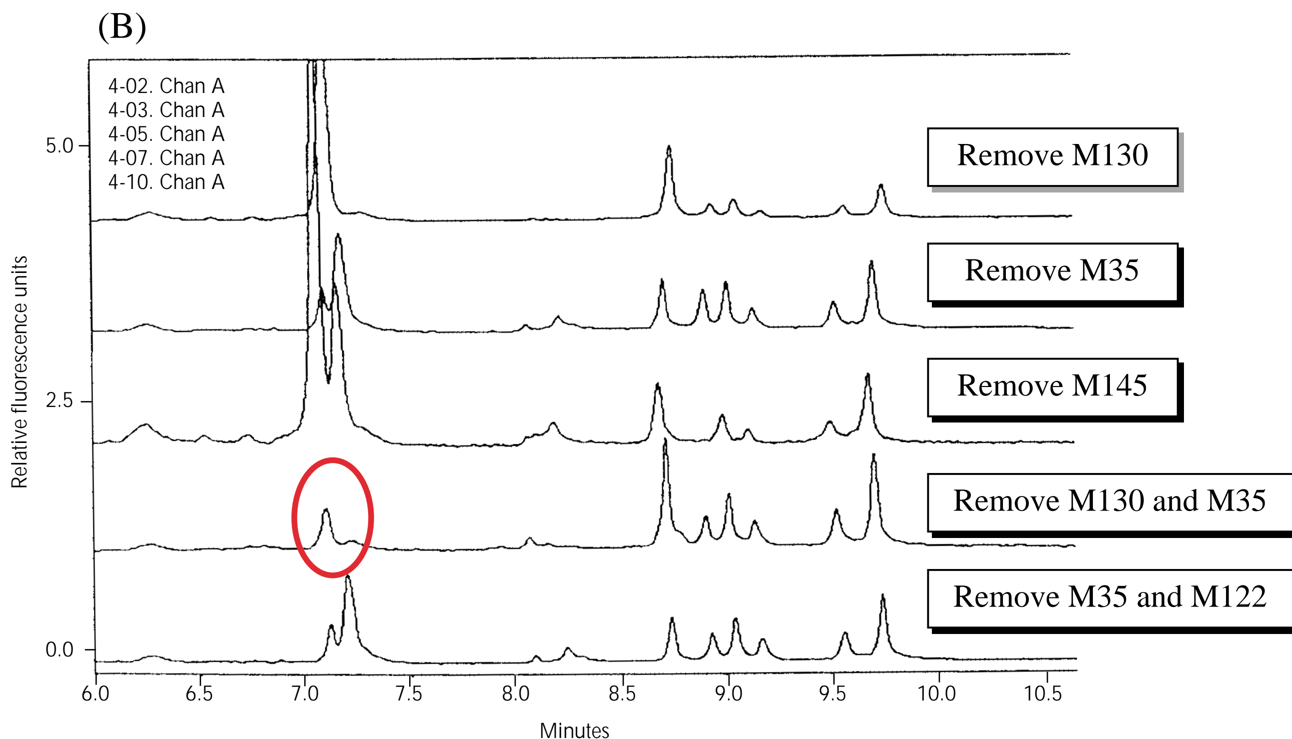
Exhibit 27. **Multiplex PCR primers used for Y SNP markers.** Universal sequences have been attached to the 5'-end of the primers.

Primer Name	Primer Sequence (5' → 3')
M2-F3u	ATT TAG GTG ACA CTA TAG AAT ACG ACC CAG GAA GGT CCA GTA A
M2-R3u	TAA TAC GAC TCA CTA TAG GGA GAC CCC CTT TAT CCT CCA CAG AT
M3-F1u	ATT TAG GTG ACA CTA TAG AAT ACC TGC CAG GGC TTT CAA ATA G
M3-R1u	TAA TAC GAC TCA CTA TAG GGA GAC TGA AAT TTA AGG GCA TCT TTC A
M13-F1u	ATT TAG GTG ACA CTA TAG AAT ACT TAT GCC CAG GAA TGA ACA AG
M13-R1u	TAA TAC GAC TCA CTA TAG GGA GAC CCA TGA TTT TAT CCA ACC ACA TT
M119-F1u	ATT TAG GTG ACA CTA TAG AAT ACG GAA GTC ACG AAG TGC AAG T
M119-R1u	TAA TAC GAC TCA CTA TAG GGA GAC GGG TTA TTC CAA TTC AGC ATA CAG
M35-F1u	ATT TAG GTG ACA CTA TAG AAT ACA GGG CAT GGT CCC TTT CTA T
M35-R5u	TAA TAC GAC TCA CTA TAG GGA GAC TGG GTT CAA GTT TCC CTG TC
M55-F1u	ATT TAG GTG ACA CTA TAG AAT ACC AAA TAG GTG GGG CAA GAG A
M55-R1u	TAA TAC GAC TCA CTA TAG GGA GAC CCT GGG ATT GCA TTT GTA CTT
M60-F1u	ATT TAG GTG ACA CTA TAG AAT ACC CAA CAC TGA GCC CTG ATG
M60-R1u	TAA TAC GAC TCA CTA TAG GGA GAC GAG AAG GTG GGT GGT CAA GA
M42-F1u	ATT TAG GTG ACA CTA TAG AAT ACA GAT CAC CCA GAG ACA CAC AAA
M42-R1u	TAA TAC GAC TCA CTA TAG GGA GAC GCA AGT TAA GTC ACC AGC TCT C
M67-F1u	ATT TAG GTG ACA CTA TAG AAT ACG ACA AAC TCC CCT GCA CAC T
M67-R1u	TAA TAC GAC TCA CTA TAG GGA GAC CTT GTT CGT GGA CCC CTC TA
M69-F1u	ATT TAG GTG ACA CTA TAG AAT ACA CTC CTG GGT AGC CTG TTC A
M69-R1u	TAA TAC GAC TCA CTA TAG GGA GAC GAA CCA GAG GCA AGG GAC TA
M26-F1u	ATT TAG GTG ACA CTA TAG AAT ACC ACA GCA GAA GAG ACC AAG ACA
M26-R1u	TAA TAC GAC TCA CTA TAG GGA GAC TGG GGC TGT ATT TGA CAT GA
M96-F1u	ATT TAG GTG ACA CTA TAG AAT ACT GCC CTC TCA CAG AGC ACT T
M96-R1u	TAA TAC GAC TCA CTA TAG GGA GAC AGA TTC ACC CAC CCA CTT TG
M122-F2u	ATT TAG GTG ACA CTA TAG AAT ACA GTT GCC TTT TGG AAA TGA AT
M122-R2u	TAA TAC GAC TCA CTA TAG GGA GAC GGT ATT CAG GCG ATG CTG AT
M145-F1u	ATT TAG GTG ACA CTA TAG AAT ACG CTG GAG TCT GCA CAT TGA T
M145-R1u	TAA TAC GAC TCA CTA TAG GGA GAC TGG ATC ATG GTT CTT GAT TAG G
M45-F2u	ATT TAG GTG ACA CTA TAG AAT ACC ATC GGG GTG TGG ACT TTA C
M45-R2u	TAA TAC GAC TCA CTA TAG GGA GAC ACA GTG GCA CCA AAG GTC AT
M9-F1u	ATT TAG GTG ACA CTA TAG AAT ACA CTG CAA AGA AAC GGC CTA A
M9-R1u	TAA TAC GAC TCA CTA TAG GGA GAC TTT TGA AGC TCG TGA AAC AGA
M89-F4u	ATT TAG GTG ACA CTA TAG AAT ACC CAA ACA GCA AGG ATG ACA A
M89-R4u	TAA TAC GAC TCA CTA TAG GGA GAC TGC AAC TCA GGC AAA GTG AG
M17-F1u	ATT TAG GTG ACA CTA TAG AAT ACC TGG TCA TAA CAC TGG AAA TCA G
M17-R1u	TAA TAC GAC TCA CTA TAG GGA GAC CCA CTT AAC AAA CCC CAA AAT
M130-F1u	ATT TAG GTG ACA CTA TAG AAT ACG GGC AAT AAA CCT TGG ATT TC
M130-R1u	TAA TAC GAC TCA CTA TAG GGA GAC GCA ATT TAG CCA CTG CTC TG

Univ-F: ATT TAG GTG ACA CTA TAG AAT AC
Univ-R: TAA TAC GAC TCA CTA TAG GGA GAC

Exhibit 28. **CE electropherograms of "dropout" experiments conducted on a 9-plex PCR primer set used in developing Y-chromosome SNP markers.** Section (A) demonstrates that by simply removing 4 primer pairs (M130, M35, M122, and M145), the multiplex PCR yield improves (i.e., there are fewer remaining primers and the amplicon yields are more balanced). Panel (B) shows that by removing only M130 and M35, the remaining primers are reduced to the greatest extent (see red oval).





PCR buffer II (ABI), 20 pmol of each universal primer, and 0.2 pmol of each locus-specific primer. The universal primer sequences were 5'-ATTTAGGT-GACACTATAGAATAC-3' (attached on 5' end of locus specific forward primers) and 5'-TAATACGACTCAC-TATAGGGAGAC-3' (attached on 5' end of locus specific reverse primers). Exhibit 27 shows the primer sequences used for multiplex amplification of up to 18 Y SNP markers. During multiplex PCR development studies, each primer set was tested individually as well as in the multiplex set. Primer sets that were less efficient exhibited a higher amount of remaining primers or primer dimers in CE electropherograms of the PCR products. "Drop-out" experiments, where one or more primers were removed from the multiplex set, were then conducted to see which primer sets interfered with one another (exhibit 28). Finally, primer concentrations were adjusted to try and improve the multiplex PCR product balance between amplicons.

Verification of PCR amplification

Following PCR, a 1 µL aliquot of the PCR product was typically checked on a 2% agarose gel stained with ethidium bromide to verify amplification success. After a set of primers had been tested multiple times and a level of confidence had been gained for amplifying a particular STR locus, the gel PCR confirmation step was no longer used.

Later in this project, a Beckman P/ACE 5500 capillary electrophoresis (CE) instrument was used to check samples after PCR. The quantitative capabilities of CE are especially important when optimizing a multiplex PCR reaction. As long as the products are resolvable, their relative peak area or heights can be used to estimate amplification efficiency and balance during the multiplex PCR reaction. The CE separations were all performed using an intercalating dye and sieving polymer solution as previously described (Butler et al., 1995) to avoid having to fluorescently label the PCR products. Samples were prepared for CE analysis by simply diluting a 1 µL aliquot of the amplicon in 49 µL of deionized water.

SNP reaction and phosphatase treatment

For SNP samples, the amplicons were treated with shrimp-alkaline phosphatase (SAP) (Amersham Pharmacia Biotech, Inc., Piscataway, NJ) to hydrolyze the unincorporated dNTPs following PCR (Haff and Smirnov, 1997). Typically, 1 U of SAP was added to each 20 µL PCR reaction and then incubated at 37 °C for 60 minutes followed by heating at 75 °C for 15 minutes. The SNP extension reaction consisted of a 5 µL aliquot of the SAP-treated PCR product, 1X TaqFS buffer, 1.2–2.4 U TaqFS (ABI), 12.5 µM dideoxynucleotide triphosphate (ddNTP) mix, and 0.5 µM biotinylated, cleavable SNP primer in a 20 µL volume. For multiplex analysis, SNP primer concentrations were balanced empirically, typically in the range of 0.3–1.5 µM, and polymerase and ddNTP concentrations were also doubled from the singleplex conditions to facilitate extension from multiple primers. The SNP extension reaction was performed in a thermal cycler with the following conditions: 94 °C for 1 min and 25–35 cycles at 94 °C for 10 sec, 45–60 °C (depending on the annealing temperature of the SNP primer) for 10 sec, and 70 °C for 10 sec. An annealing temperature of 52 °C was used for the mtDNA 10plex SNP assay.

Sample Cleanup and Mass Spectrometry

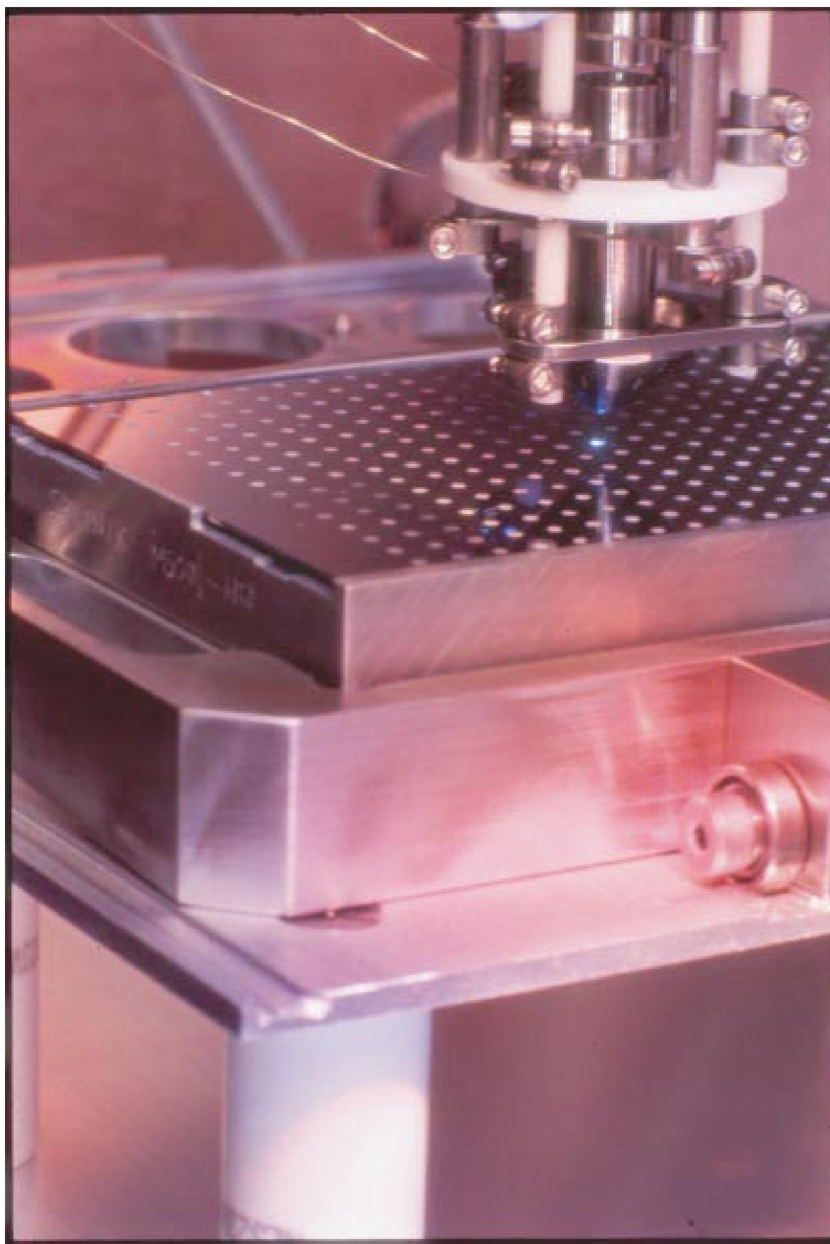
Following PCR amplification, a purification procedure involving solid-phase capture and release from streptavidin-coated magnetic beads was utilized (Monforte et al., 1997) to remove salts that interfere with the MALDI ionization process (Shaler et al., 1996). At the start of this project, most of the sample purification was performed manually in 0.6 mL tubes with a 1.5 mL Dynal MPC®-E (Magnetic Particle Concentrator for Microtubes of Eppendorf Type) (Dynal A.S., Oslo, Norway). Larger scale experiments performed toward the end of this project utilized a robotic workstation fitted with a 96-tip pipettor that mimicked the manual method. This sample cleanup method involved washing the DNA with a series of chemical solutions to remove or reduce the high levels of sodium, potassium, and magnesium present from the PCR reaction. The PCR products were then released from the bead with a chemical cleavage step that breaks the covalent bond between the 5'-biotinylated portion of the DNA product and the remainder of the extension product, which contains the STR repeat region or the dideoxynucleotide added during the SNP reaction. In the final step prior to mass spectrometry analysis, samples were evaporated to dryness using a speed vac, reconstituted in 0.5 µL of matrix (manual protocol) or 2 µL of matrix (robotic protocol), and spotted on the sample plate.

The matrix typically used for STR analysis was a 5:1 molar ratio of 3-hydroxypicolinic acid (3-HPA) (Lancaster Synthesis, Inc., Windham, NH) with picolinic acid in 25 mM ammonium citrate (Sigma-Aldrich, St. Louis, MO) and 25% acetonitrile. For SNP analysis, ~0.5 M saturated 3-HPA was used with the same solvent of 25 mM ammonium citrate and 25% acetonitrile. A GeneTrace-designed and built linear time-of-flight mass spectrometer was used as previously described (Wu et al., 1994). Much of the early data were collected manually on a research mass spectrometer. During the time period of this project, GeneTrace also built multiple high-throughput instruments.

Automated high-throughput mass spectrometer

GeneTrace has designed and custom-built unique, automated time-of-flight mass spectrometers for high-throughput DNA analysis. The basic instrument design is covered under U.S. Patent 5,864,137 (Becker and Young, 1999). A high repetition rate UV laser (e.g., 100 Hz) is used to enable collection of high quality mass spectra consisting of 100–200 summed shots in only a few seconds. The sample chamber can hold up to two sample plates at a time with each plate containing 384 spotted samples. Exhibit 29 shows a sample plate on the X–Y table under the custom GeneTrace ion optics.

Exhibit 29. **Photo of the automated GeneTrace mass spectrometry ion optics over a sample plate containing 384 different DNA samples.** A sample in the center of the plate is illuminated by a pulsing laser light.



An important feature of this automated mass spectrometer is "peak picking" software that enables the user to define "good" versus "bad" mass spectra. After each laser pulse, the "peak picker" algorithm checks for peaks above a user-defined signal-to-noise threshold in a user-defined mass range. Only "good" spectra are kept and summed into the final sample spectrum, which improves the overall signal quality. The X–Y table moves in a circular pattern around each sample spot until either the maximum number of good shots (e.g., 200) or the maximum number of total shots (e.g., 1,000) is reached. A mass spectrum's signal-to-noise level is related to the number of laser shots collected. In general, signal-to-noise improves as the square root of the number of shots. Thus, improving the signal by a factor of two would require increasing the number of good shots collected by a factor of four.

Raw data files (.dat) were converted to "smoothed" data files (.sat) using custom software developed at GeneTrace that improved data quality and involved several multipoint Savitzky-Golay averages along with a baseline subtraction algorithm (Carroll and Beavis, 1996). A set of samples was collected under a single "header" file with identical peak picking parameters. Each header file recorded the mass calibration constants and peak picking parameters and listed all of the samples analyzed with the number of good shots collected versus the number of total shots taken for each sample.

Data points in mass spectrometry are collected in spectral channels that must be converted from a time value to a mass value. This mass calibration is normally performed with two oligonucleotides that span the mass range being examined. For example, a 36-mer (10,998 Da) and a 55-mer

(16,911 Da) were typically used when examining STRs in the size range of 10,000–40,000 Da. On the other hand, a 15-mer (4,507 Da) and its doubly charged ion (2,253.5 Da) were used to cover SNPs in the size range of 1,500–7,500 Da. Ideally, larger mass oligonucleotides would be used for STR analysis to obtain more accurate masses, but producing a clean, well-resolved peak above 25–30 kDa is a synthetic and instrumental challenge. The calibration was typically performed only once per day because the calibration remains consistent over hundreds of samples. The mass accuracy and precision are such that no sizing standards or allelic ladders need to be run to determine a sample's size or genotype (Butler et al., 1998b).

Delayed extraction (Vestal et al., 1995) and mass gating ("blanking") were used to improve peak resolution and sensitivity, respectively. Typically, a delay of 500–1,000 nanoseconds was used to eliminate ions below ~8,000 Da for STRs, and a delay of 250–500 nanoseconds was used with a signal blanking below ~1,000 Da for SNPs.

Sample Genotyping

Automated STR genotyping program (CallSSR or CallSTR)

During the time period of this project, GeneTrace developed an automated sample genotyping program, named CallSSR. The data sets described in the Results section were processed either with CallSSR version 1.82 or a modified version of the program named CallSTR. The program was written in C++ at GeneTrace by a scientific programmer named Nathan Hunt and can run on a Windows® NT platform. A reference DNA sequence is used to establish the possible STR alleles and their expected masses based on an expected repeat mass and range of alleles. This mass information is recorded in a mass ladder file (exhibit 30). In the case of the forensic STR loci examined in this project, the GenBank sequences were used as the reference DNA sequences.

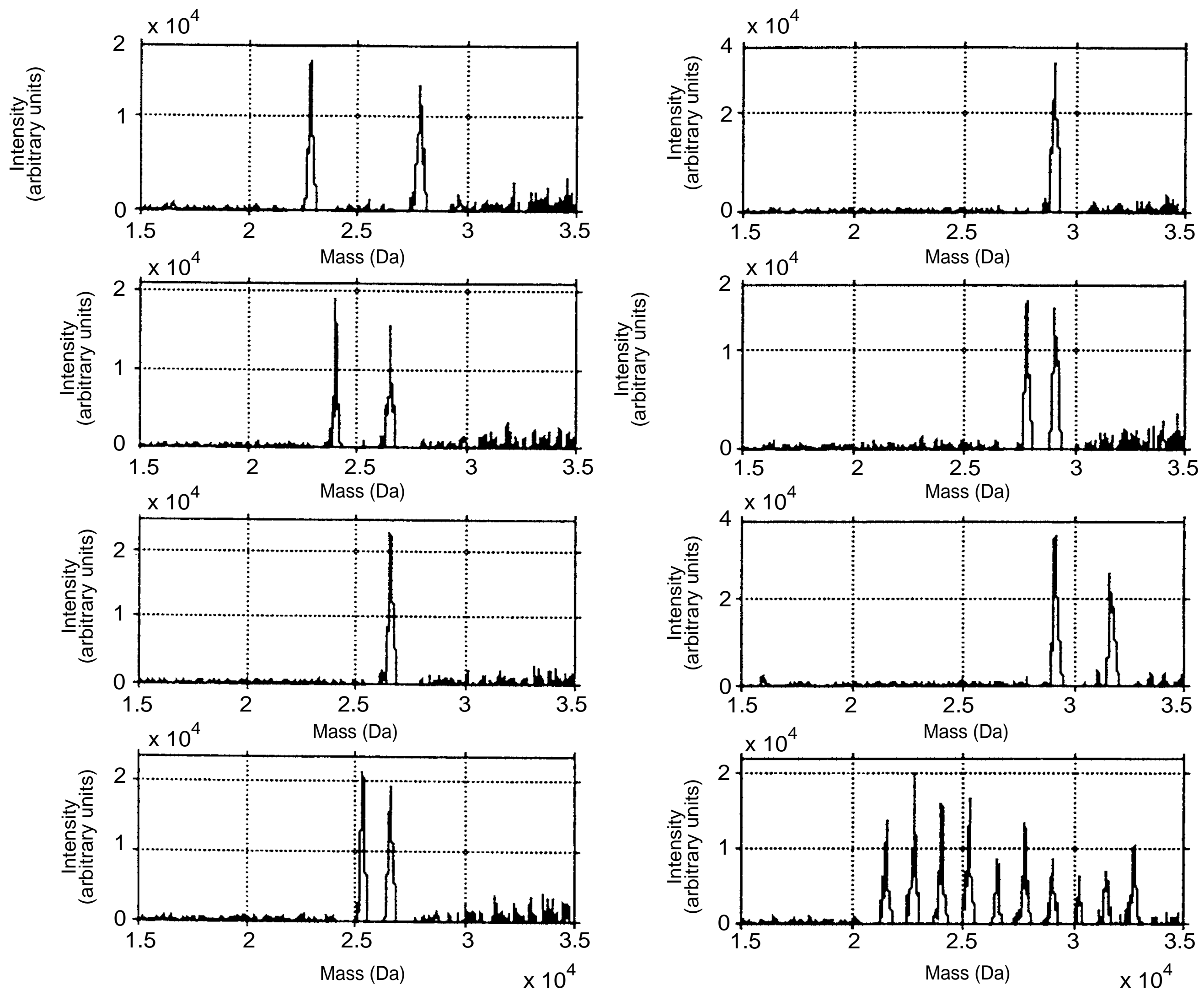
Exhibit 30. **Mass ladder file for STR loci analyzed in this study.** Different masses exist for a locus due to different primer positions. The most commonly used primer sets are highlighted. The reference mass does not include nontemplate addition by the polymerase. The genotyping program automatically adds 313 Da to the reference mass for compensation of adenylation.

Locus Name	Repeat Mass (Da)	Reference Allele	Reference Mass (Da)	Minimum Repeat	Maximum Repeat	Forward Primer	Reverse Primer
TH01	1,260	9	21,133	3	12	TH01-F2	TH01-R2-GTS
TH01b	1,260	9	18,685	3	12	TH01-F	TH01-R-GTS
TH01b2	1,260	9	18,372	3	12	TH01-F (w/o+A)	TH01-R-GTS
TH01c	1,260	9	19,856	3	12	ddC	TH01-R2-GTS
TPOX	1,260	11	21,351	5	14	TPOX-F-GTS	TPOX-R
TPOXc	1,260	11	15,119	5	14	TPOX-F-GTS	ddC
CSF	1,260	12	28,890	5	16	CSF-F3-GTS	CSF-R3
CSFb	1,211	12	27,989	5	16	CSF-F3	CSF-R3-GTS
CSFc	1,260	12	22,993	5	16	CSF-F3-GTS	ddC
AMEL	951	6	27,333	3	7	AMEL-F-GTS	AMEL-R
AMELb	924	6	18,239	4	7	AMEL-F2	AMEL-R2-GTS
D7S820	1,260	12	21,639	5	16	D7S820-F-GTS	D7-R
D7S820c	1,260	12	16,949	5	16	D7S820-F-GTS	ddC
D3S1358	1,211	16	34,300	9	21	D3-F	D3-R-GTS
D3S1358b	1,260	15	25,624	9	21	D3-F2-GTS	D3-R2
D3S1358c	1,260	15	15,994	9	21	D3-F2-GTS	ddC
D16S539	1,211	11	18,180	4	16	D16-F	D16-R-GTS
D16S539b	1,260	11	26,668	4	16	D16-F4-GTS	D16-R4
D16S539c	1,260	11	14,765	4	16	D16-F4-GTS	ddC
D16S539d	1,211	11	25,572	4	16	D16-F4	D16-R4-GTS
D16S539e	1,211	11	19,127	4	16	D16-F	D16-R2-GTS
FGA	1,202	21	37,180	15	30	FGA-F2-GTS	FGA-R2
D8S1179	1,211	12	24,417	8	18	D8S1179-F-GTS	D8S1179-R
DYS391	1,211	9	23,004	7	14	DYS391-F-GTS	DYS391-R

CallSSR accepts as input smoothed, baseline-subtracted data files, a "layout" file, and the mass ladder file. The layout file describes each sample's position on the 384-well plate, the primer set used for PCR (i.e., the STR locus), and the DNA template name. The program processes samples at a rate of more than one sample per second so that a plate of 384 samples can be genotyped in less than 5 minutes. This high rate of processing speed is necessary in a high-throughput environment where thousands of samples must be genotyped every day.

Two files result from running the program: a "call file" and a "plot file." The call file may be imported into Microsoft® Excel for data examination and contains information like the allele mass and calculated sample genotype. The plot file generates plotting parameters that work with MATLAB (The Math Works, Inc., Natick, MA) scripts to plot 8 mass spectra per page, as seen in exhibit 31. Plots are generated in an artificial repeat space to aid visual inspection of the mass spectrometry data compared with allele bins. The CallSSR algorithm has been written to ignore stutter peaks and double-charged peaks, which are artifacts of the DNA amplification step and mass spectrometry ionization process, respectively.

Exhibit 31. **Mass spectra of CDOJ samples amplified with CSF1PO primers.** From left side, top-to-bottom followed by right side, top-to-bottom, sample genotypes are (7,11), (8,10), (10,10), (9,10), (12,12), (11,12), (12,14), and allelic ladder. The mass range shown here is 15,000–35,000 Da.



Automated SNP genotyping program (CallSNP)

In-house automated SNP analysis software was developed and used to determine the genotype for each SNP marker. This program, dubbed CallSNP, was written in C++ by a scientific programmer named Kevin Coopman and will run on a Windows NT or UNIX platform. The software searches for an expected primer mass and, after locating the pertinent primer, searches for the four possible extension products by using a linear least squares fit, with the primer peak shape serving as the fitting line. In this way, peak adducts from the ionization process are distinguished from true heterozygotes. The fit coefficients of the four possible nucleotides are then compared with one another to determine the appropriate SNP base. The base with the highest value (i.e., best fit) is the called base. The mass between the primer and the extension product can then be correlated to the incorporated nucleotide at the SNP site. In the case of a heterozygote at the SNP site, two extension products exist and are called by the software.

As with the CallSSR software, a layout file, a mass file, and mass spectrometry data files are required as input. Call files generate information regarding the closeness of the fit for each possible nucleotide with an error value associated for each call. The SNP mass information file includes the SNP marker name, expected primer mass (post-cleavage), and expected SNP bases. The current version of CallSNP works well for singleplex SNPs but needs modifications before it can work effectively on multiplex SNP samples. In principle, the program could be scaled to limited, widely spaced multiplexes where the doubly charged ions of larger mass peaks do not fall in the range of lower mass primer peaks.

Comparison Tests With ABI 310 Genetic Analyzer

For comparison purposes, more than 200 genomic DNA samples were genotyped using the Applied Biosystems 310 Genetic Analyzer and the AmpF1STR® Profiler Plus™ or AmpF1STR® COfiler™ fluorescent STR kit. Exhibit 32 lists the numbers of samples analyzed with each STR kit. STR samples were run in the ABI 310 CE system using the POP-4 polymer, 1X Genetic Analysis buffer, and

Exhibit 32. **Sample sets run on ABI 310 Genetic Analyzer with AmpF1STR® Profiler Plus™ or AmpF1STR® COfiler™ fluorescent STR kits**

Sample Set	Number of Samples	Kit Used	MS Data Compared	Comments
DOJ plate	88	COfiler	TH01, TPOX, CSF1PO, D7S820, D3S1358, D16S539, D8S1179, FGA, DYS391, and amelogenin	*See exhibits 12–19, 39, and 40*
CEPH/Diversity plate	92	COfiler and ProfilerPlus	TH01, TPOX, CSF1PO, D7S820, D3S1358, D16S539, D8S1179, and amelogenin	*See exhibit 63*
Stanford male samples	37	ProfilerPlus	For Y SNP testing (not completed)	*See exhibits 44–45;* new microvariants in D18, D21, FGA seen
Butler family samples	34	COfiler	No mass spectrometry data collected	
JMB family samples	4	COfiler and ProfilerPlus	No mass spectrometry data collected	
Standard templates	3	COfiler and ProfilerPlus	TH01, TPOX, CSF1PO, D7S820, D3S1358, D16S539, D8S1179, FGA, DYS391, and amelogenin	*See exhibit 73* K562, AM209, and UP006
TOTAL TESTED	**258**	221 COfiler 136 ProfilerPlus	**2,907 total genotypes measured**	

CDOJ results	88	Profiler (under CDOJ validated method)	*See above*	B11, D3S1358:15,15.2 incorrectly called 14,15 *(see exhibit 74)*

a 47-cm (50 µm i.d.) capillary with the GS STR POP4 (1 mL) F separation module (ABI Prism 310 Genetic Analyzer User's Manual, 1998). With this module, samples were electrokinetically injected for 5 seconds at 15,000 volts and separated at 15,000 volts for 24 minutes with a run temperature of 60 °C. DNA sizing was performed with ROX-labeled GS500 as the internal sizing standard. Samples were prepared by adding 1 µL PCR product to 20 µL deionized formamide containing the ROX-GS500 standard. The samples were heat-denatured at 95 °C for 3 minutes and then snap cooled on ice prior to being loaded into the autosampler tray. These separation conditions and sizing standards are commonly used in validated protocols by forensic DNA laboratories. Following data collection, samples were analyzed with Genescan 2.1 and Genotyper 2.0 software programs (ABI).

While standard CE conditions were used, new PCR conditions were developed to dramatically reduce the cost of using ABI's STR kits. The PCR volume was reduced from the standard 50 µL described in the ABI protocol (AmpF1STR® ProfilerPlus™, 1998, and AmpF1STR® Cofiler™, 1998) to 5 µL, which corresponded with a cost reduction of 90% per DNA amplification. The kit reagents were mixed in their ABI-specified proportions—11 µL primer mix, 1 µL TaqGold polymerase (5 U/µL), and 21 µL PCR mix. A 3 µL aliquot of this master mix was then added to each tube along with 2 µL of genomic DNA template (typically at 1–2 ng/µL). Both PE9700 and MJ Research thermal cyclers worked for this reduced PCR volume method provided that the 200 µL PCR tubes were sealed well to prevent evaporation. Results showed that an 8-strip of 0.2 mL thin wall PCR tubes from Out Patient Services, Inc., (OPS) (Petaluma, CA) worked best for the PE9700 thermal cycler. Only 1 µL is needed for CE sample preparation; a sample that can be re-injected multiple times if needed. Thus, with a 50 µL PCR reaction, 49 µL were never used to produce a result under the standard ABI protocol. A 5 µL PCR produces less waste in addition to being less expensive. More importantly, the multiplex STR amplicons were more concentrated in a lower volume and produced higher signals in the ABI 310 data collection. Peak signals were often off-scale, and the number of cycles in the cycling program could be reduced from 28 to 26, or even 25 with some DNA templates. This 5 µL PCR also worked well with FTA paper punches that have been washed with FTA purification reagent (Life Technologies).

The California Department of Justice ran one plate of 88 samples with the AmpF1STR® Profiler™ kit on an ABI 310 Genetic Analyzer and provided those genotypes for comparison purposes. These results provided an independent verification of our work.