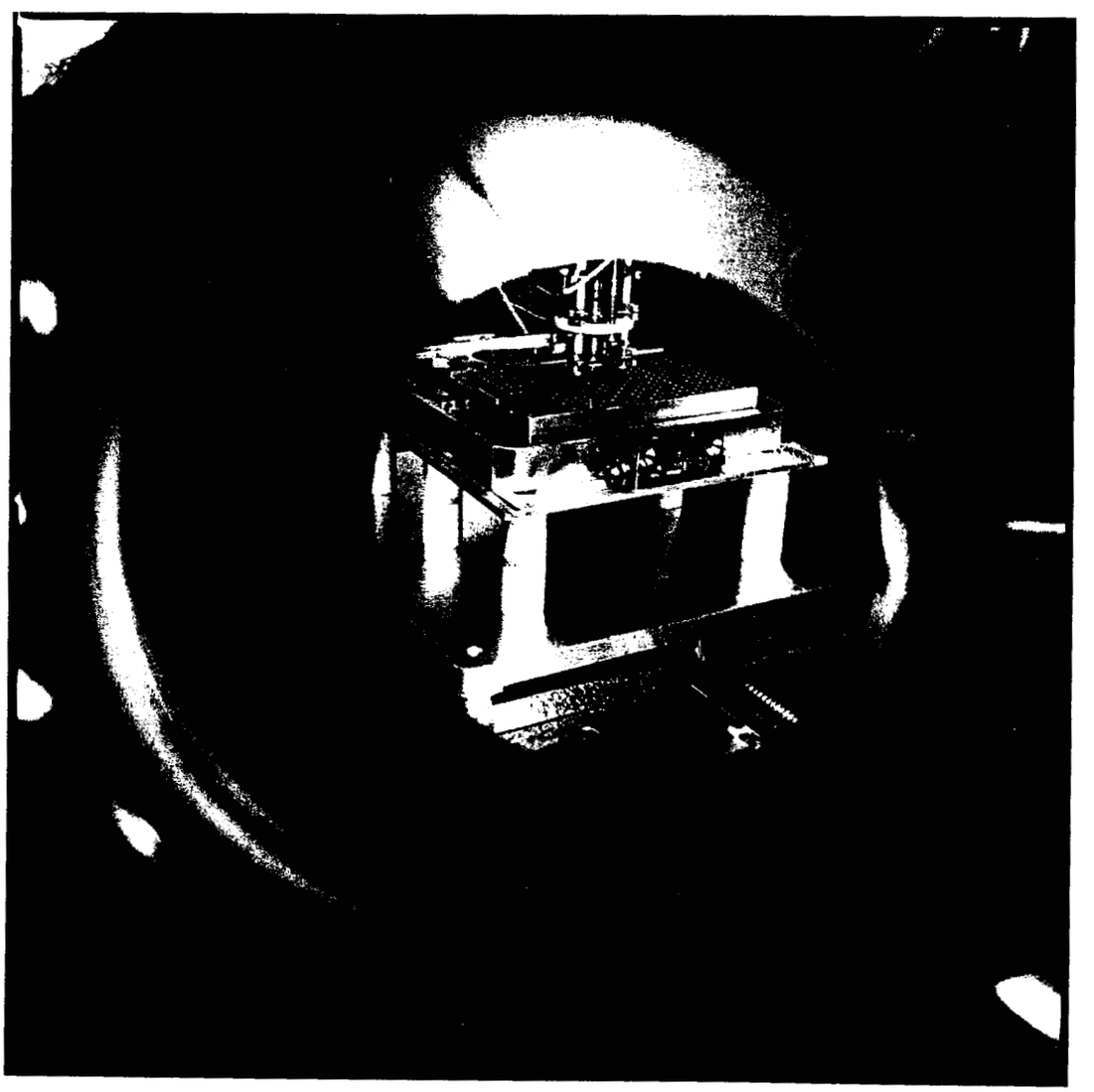
179419

FINAL REPORT
on NIJ Grant
97-LB-VX-0003

Improved Analysis of DNA Short Tandem Repeats with Time-of-Flight Mass Spectrometry

John M. Butler and Christopher H. Becker
GeneTrace Systems Inc.

Research performed from June 1997 to April 1999



Improved Analysis of DNA Short Tandem Repeats with Time-of-Flight Mass Spectrometry

John M. Butler and Christopher H. Becker

TABLE OF CONTENTS

ACKNOWLEDGMENTS

The project described in this final grant report could not have happened without the hard work and support of a number of people at GeneTrace Systems Inc. First and foremost was Jia Li, who did some of the early primer design and STR work to demonstrate that STRs could be effectively analyzed by mass spectrometry. Jia taught me a lot about PCR and was always encouraging of my work. Likewise, Tom Shaler was important in the early phases of this research with his expert advice in mass spectrometry and data processing. The first GeneTrace STR mass spectra were carefully collected by Tom and thus he along with Jia deserves credit for helping obtain the funding for this grant. Dan Pollart synthesized numerous cleavable primers for this project especially in the first year of our work. David Joo and Wendy Lam also prepared PCR and SNP primers for the later part of this work. A number of people assisted in robotic sample preparation and sample cleanup including Mike Abbott, Jon Marlowe, David Wexler, and Rebecca Turincio. Joanna Hunter, Vera Delgado, and Can Nhan ran many of the STR samples on the automated mass spectrometers. Their hard work made it possible to focus on experimental design and data analysis rather than routine sample handling. Having talented and supportive coworkers was a great blessing throughout the course of this project. Kathy Stephens, Jia Li, Tom Shaler, Yuping Tan, Christine Loehrlein, Joanna Hunter, Hua Lin, Gordy Haupt, and Nathan Hunt provided useful discussions on a number of issues and helped develop assay parameters and tackle automation issues among other things. Nathan Hunt was especially important to the success of this project as he developed the STR genotyping algorithm and CallSSR software as well as the multiplex SNP primer design software. Kevin Coopman developed the SNP genotyping algorithm and calling software and was always eager to analyze my multiplex SNP samples. Joe Monforte and Roger Walker served as my supervisors for the first year and second year of this project, respectively, allowed me the opportunity to devote sufficient time to doing the work described in this project. Last but not least, Debbie Krantz served as an able administrator of these two NIJ grants and took care of the financial end of the grants.

We were also supported with samples and sequence information from a number of scientific collaborators. Steve Lee and John Tonkyn from the California Department of Justice DNA Laboratory provided genomic DNA samples and STR allelic ladders. Dibang Liu from Northwestern University provided the D3S1358 DNA sequence used for improved primer design purposes. Peter Oefner and Peter Underhill from the Department of Genetics at Stanford University provided male population samples and Y chromosome SNP sequences. The encouragement and support of Lisa Forman and Richard Rao from the Office of Justice Programs at the National Institute of Justice propelled this work from an idea to a working product. In addition, Dennis Reeder from the National Institute of Standards and Technology was always a constant source of encouragement at scientific meetings.

This project was supported under award number 97-LB-VX-0003 from the National Institute of Justice, Office of Justice Programs, U.S. Department of Justice. Points of view in this document are those of the authors and do not necessarily represent the official position of the U.S. Department of Justice. For further information, contact: John Butler, GeneTrace Systems Inc., 1401 Harbor Bay Parkway, Alameda, CA 94025; (510) 748-6124; butler@genetrace.com.

EXECUTIVE SUMMARY

A. Introduction

The advent of DNA typing and its use for human identity testing has revolutionized law enforcement in recent years by allowing forensic laboratories to match suspects with miniscule amounts of biological evidence from a crime scene. Equally important is the use of DNA to exclude suspects who were not involved in a crime or to identify human remains in an accident. The last decade has seen numerous advances in the DNA testing procedures, most notably among them is the development of PCR (polymerase chain reaction)-based DNA typing methods. Technologies for measuring DNA variation, both length and sequence polymorphisms, have also advanced rapidly in the last decade. The time for the determination of a sample's DNA profile has dropped from 6-8 weeks to 1-2 days with the possibility of being able to process samples in only a few minutes or hours due to more recent advancements.

Simultaneous with the evolution of DNA markers and technologies embraced by the forensic community has been the acceptance and use of DNA typing information. The courtroom battles over statistical issues that were common in the late 1980s and early 1990s have subsided as DNA evidence has become more widely accepted. In the past five years, DNA databases have emerged as powerful tools for criminal investigations much like the fingerprint databases have been routinely used for decades. The United Kingdom launched a nationwide DNA database in 1995 that now contains more than 400,000 DNA profiles from convicted felons—profiles that have been used to aid over 50,000 criminal investigations. National DNA databases are springing up in countries all over the world as their value to law

enforcement is being recognized. In the United States, the FBI has developed the Combined DNA Index System (CODIS) with the anticipation that several million DNA profiles will be entered into this database in the next decade. All 50 states in the U.S. now have laws requiring DNA typing of convicted offenders, typically for violent crimes such as rape or homicide.

While the law enforcement community is gearing up to gather millions of samples from convicted felons, the DNA typing technology needs improvement. Large sample backlogs exist today due to the high cost of performing the DNA testing and limited capabilities in forensic laboratories. Several states, including California, Virginia, and Florida, have sample backlogs of over 50,000 samples. A need exists for more rapid and cost-effective methods for high-throughput DNA analysis to process samples currently being gathered for large criminal DNA databases around the world.

At the start of this project (June 1997), commercially available slab gel or capillary electrophoresis instruments could handle only a few dozen samples per day. While larger numbers of samples can be processed by increasing the number of laboratory personnel and instruments, the development of high-throughput DNA processing technologies promises to be more cost-effective in the long run especially for the generation of large DNA databases. GeneTrace Systems Inc., a small biotech company located in Alameda, California, has developed high-throughput DNA analysis capabilities using time-of-flight mass spectrometry coupled with parallel sample preparation on a robotic workstation. The GeneTrace technology allows several thousand samples to be processed daily with each instrument.

DNA samples may be analyzed in seconds rather than minutes or hours with improved accuracy compared to conventional electrophoresis methods.

This NIJ project was initiated to adapt the GeneTrace technology to human identity DNA markers commonly used by forensic DNA laboratories, specifically short tandem repeat (STR) markers. An extension of the original grant was submitted in December 1997 and funded to develop single nucleotide polymorphism (SNP) markers from mitochondrial DNA and the Y chromosome. Based on the results obtained in this study, we believe mass spectrometry can be a useful and effective means for high-throughput DNA analysis that has the capabilities to meet the needs of the forensic DNA community for offender DNA databasing. However, due to limited resources and a perceived difficulty to enter the forensic DNA market, GeneTrace has made a business decision to not pursue this market. While the STR milestones on the original grant were met, only the initial milestones were achieved on the SNP portion of our NIJ grant due to its premature termination on the part of GeneTrace.

B. GeneTrace Systems and Mass Spectrometry

GeneTrace Systems Inc. is a small biotech company located in Alameda, California, which employs approximately 65 biologists, chemists, engineers, computer programmers, and other support personnel. GeneTrace Systems was founded in 1994 as a spin-off of SRI International (formerly Stanford Research Institute) in Menlo Park, CA, to develop tools for genetic analysis using mass spectrometry. Dr. Christopher Becker, the President of GeneTrace and the Principal Investigator, has led research efforts on mass spectrometry of DNA beginning with a grant in 1990 from NIH's National Center for Human Genome Research. GeneTrace Systems Inc. has developed an integrated high-throughput DNA analysis system involving the use of proprietary chemistry, robotic sample manipulation, and time-of-flight mass spectrometry. The purpose of this grant was to apply the GeneTrace technology to improved analysis of short tandem repeat markers commonly used in forensic DNA laboratories.

Mass spectrometry is a versatile analytical technique that involves the detection of ions and the measurement of their mass-to-charge ratio. Due to the fact that these ions are separated in a vacuum environment, the analysis times can be extremely rapid, on the order of microseconds. Many advances have been made in the past decade for the analysis of biomolecules such as DNA, proteins, and carbohydrates since the introduction of a new ionization technique known as matrix-assisted laser desorption-ionization (MALDI) and the discovery of new matrixes that effectively ionize DNA without extensive fragmentation. When coupled with time-of-flight mass spectrometry, this method for measuring

biomolecules is commonly referred to as MALDI-TOF-MS. A schematic of MALDI-TOF mass spectrometry may be seen in **Exhibit 1**.

C. Short Tandem Repeats

Short tandem repeat DNA markers, also referred to as microsatellites or simple sequence repeats (SSRs), consist of tandemly repeated DNA sequences with a core repeat of 2-6 base pairs (bp). STR markers are readily amplified with the polymerase chain reaction (PCR) by using primers that bind in conserved regions of the genome flanking the repeat region. Forensic laboratories prefer tetranucleotide loci (i.e., 4 bp in the repeat) due to the lower amount of "stutter" produced during PCR. (Stutter products are additional peaks that appear in front of regular allele peaks that can complicate the interpretation of DNA mixtures.) The number of repeats can vary from 3 or 4 repeats to more than 50 repeats with extremely polymorphic markers. The number of repeats, and hence the size of the PCR product, may vary among samples in a population making STR markers useful in identity testing or genetic mapping studies.

Shortly after this project was initiated, the FBI designated 13 core STR loci for the nationwide Combined DNA Index System (CODIS) database. These STR loci are TH01, TPOX, CSF1PO, VWA, FGA, D3S1358, D5S818, D7S820, D13S317, D16S539, D8S1179, D18S51, and D21S11. The sex-typing marker, amelogenin, is also included in STR multiplexes that cover the 13 core STR loci. Each sample must therefore have these 14 markers tested in order to be entered into the national CODIS database. To illustrate the kinds of numbers involved to analyze the current national sample backlog of ~500,000 samples, more than 7

million genotypes must be generated. Using currently available technologies, an estimated cost of ~$25 million (~$50/sample) and more than 5 years for well-trained and well-funded laboratories would be required to determine those 7 million genotypes. With the high cost and effort required, most of these backlogged samples are being stored in anticipation of future analysis and inclusion in CODIS, pending the development of new, faster technology or the implementation of more instruments using the current electrophoretic technologies.

Time-of-flight mass spectrometry has the potential to bring DNA sample processing to a new level in terms of high-throughput analysis. However, there are several challenges for analysis of PCR products, such as short tandem repeat markers, using MALDI-TOF-MS. Mass spec resolution and sensitivity are diminished when either the DNA size or the salt content of the sample is too large. By redesigning the PCR primers to bind close to the repeat region, the STR allele sizes are reduced so that resolution and sensitivity of the PCR products are benefited. Therefore, much of this project involved designing and testing new PCR primers that produce smaller amplicon sizes for STR markers of forensic interest. We focused on STR loci that have been developed by commercial manufacturers and studied extensively by forensic scientists. These include all of the GenePrint™ tetranucleotide STR systems from Promega as well as the 13 CODIS STR loci that are covered by the Profiler Plus™ and COfiler™ kits from PE Applied Biosystems (**Exhibit 2**). Where possible we designed our primers to produce amplicons that are less than 100 bp although we have been able to resolve neighboring STR alleles that are as large as 140 bp in size. For example, TPOX alleles 6-14 range from 69-101 bp in size with GeneTrace-designed primers while with Promega's GenePrint™ primers the same TPOX alleles range in size from 224-256 bp. Unfortunately,

due to the long and complex repeat structures of several STR markers, we have been unable to obtain the necessary single base resolution with the following STR loci: D21S11, D18S51, and FGA (see Results section).

To verify the STR results we obtained from our mass spectrometry method, we collaborated with the California Department of Justice DNA Laboratory (CAL-DOJ) in Berkeley to generate a large data set. CAL-DOJ provided us with 88 samples that had been previously genotyped using validated fluorescent multiplex STR kits from PE Applied Biosystems. GeneTrace generated STR results using our primer sets for 9 STR loci (TH01, TPOX, CSF1PO, D3S1358, D16S539, D8S1179, FGA, DYS391, and D7S820) along with the sex-typing marker amelogenin. These experiments allowed us to compare over 700 genotypes (88 samples x 8 loci; data from D8S1179 and DYS391 were not available from CAL-DOJ) between the two methods. Although we did not obtain results for all possible samples on the mass spec, we observed almost 100% correlation with the genotypes obtained between the validated fluorescent STR method and our newly developed mass spectrometry technique demonstrating that our method is reliable (see Results section).

Multiplex STR Analysis

To reduce analysis cost and sample consumption and to meet the demands of higher sample throughputs, PCR amplification and detection of multiple markers (multiplex STR analysis) has become a standard technique in most forensic DNA laboratories. STR multiplexing is most commonly performed using spectrally distinguishable fluorescent tags and/or non-overlapping PCR product sizes. An example STR multiplex produced from a commercially available kit is shown in **Exhibit 3**. Multiplex STR amplification in one or two PCR

reactions with fluorescently labeled primers and measurement with gel or capillary electrophoretic separation and laser-induced fluorescence detection is becoming a standard method among forensic laboratories for analysis of the 13 CODIS STR loci. The STR alleles from these multiplexed PCR products typically range in size from 100-350 bp with commercially available kits.

Due to the limited DNA size constraints of mass spectrometry, we have adopted a different approach to multiplex analysis of multiple STR loci. Primers are designed such that the PCR product size ranges overlap between multiple loci but have alleles that interleave and are resolvable in the mass spectrometer (**Exhibit 4**). As described above, we have designed PCR primers that are closer to the STR repeat regions than those commonly used with electrophoretic systems. The high accuracy, precision, and resolution of our mass spec approach permit multiplexing STR loci in such a manner for a limited number of markers. During the course of the work funded by this grant, we developed a TH01-TPOX-CSF1PO STR triplex (**Exhibit 5**).

D. Single Nucleotide Polymorphisms (SNPs)

Single nucleotide polymorphisms represent another form of DNA variation that is useful for human identity testing. SNPs are the most frequent form of DNA sequence variation in the human genome and as such are becoming increasing popular genetic markers for genome mapping studies and medical diagnostics. SNPs are typically biallelic with two possible nucleotides (alleles) at a particular site in the genome. Because SNPs are less polymorphic (i.e., have fewer alleles) than the currently used STR markers, more SNP markers are required

to obtain the same level of discrimination between samples. Current estimates are that 30-50 unlinked SNPs will be required to obtain the matching probabilities of 1 in ~100 billion as seen with the 13 CODIS STRs.

The perceived value of SNPs for DNA typing in a forensic setting include more rapid analysis, cheaper cost, simpler interpretation of results (e.g., no stutter products), and improved ability to handle degraded DNA (e.g., smaller PCR product sizes are possible). While it is doubtful that autosomal SNPs will replace the current battery of STRs used in forensic laboratories in the near future, abundant mitochondrial and Y chromosome SNP markers exist and have already proven useful as screening tools. These maternal (mitochondrial) and paternal (Y chromosome) lineage markers are effective in identifying missing persons, war casualties, and helping answer historical questions such as the recent issue of whether or not Thomas Jefferson fathered a slave child. The forensic DNA community already has experience with applying SNP markers as a screening process, which can prove very helpful with excluding suspects from crime scenes. Many crime laboratories still use reverse dot blot technology for analyzing the SNPs from HLA-DQA1 and PolyMarker loci with kits from PE Applied Biosystems. In addition, mitochondrial DNA (mtDNA) sequencing is currently performed in some forensic laboratories.

In the work performed on multiplex SNP markers begun at GeneTrace, we examined 10 polymorphic sites within the mtDNA control region and 20 Y chromosome SNPs provided by Dr. Peter Oefner and Dr. Peter Underhill from Stanford University. A multiplex SNP assay was developed for 10 mtDNA SNP sites (**Exhibit 6**). Only limited work was performed on

the Y chromosome SNPs due to the premature termination of our work. However, we were able to demonstrate a male-specific 17-plex PCR of 17 different Y SNP markers (**Exhibit 70**).

E. Papers Published

Six publications resulted from the work funded by NIJ for the past two years with at least one more manuscript in preparation. All articles were published with journals or conference proceedings that are accessible and frequented by forensic DNA scientists to ensure proper dissemination of the information.

Butler, J.M., Li, J., Shaler, T.A., Monforte, J.A., Becker, C.H. (1998) Reliable genotyping of short tandem repeat loci without an allelic ladder using time-of-flight mass spectrometry. *Int. J. Legal Med.*, 112: 45-49.

Butler, J.M., Li, J., Monforte, J.A., Becker, C.H., Lee, S. (1998) Rapid and automated analysis of short tandem repeat loci using time-of-flight mass spectrometry. *Proceedings of the Eighth International Symposium on Human Identification 1997,* Promega Corporation, pp 94-101.

Butler, J.M., Stephens, K.M., Monforte, J.A., Becker, C.H. (1998) High-throughput STR analysis by time-of-flight mass spectrometry. *Proceedings of the Second European Symposium on Human Identification 1998*, Promega Corporation, in press.

Butler, J.M., Becker, C.H. (1998) High-throughput genotyping of forensic STR and SNP loci using time-of-flight mass spectrometry. *Proceedings of the Ninth International Symposium on Human Identification 1998*, Promega Corporation, in press.

Li, J., Butler, J.M., Tan, Y., Lin, H., Royer, S., Ohler, L., Shaler, T.A., Hunter, J.M., Pollart, D.J., Monforte, J.A., Becker, C.H. (1999) Single nucleotide polymorphism determination using primer extension and time-of-flight mass spectrometry. *Electrophoresis*, 20:1258-1265.

Butler, J.M. (1999) STR analysis by time-of-flight mass spectrometry. *Profiles in DNA* (Promega), 2(3): 3-6.

In addition, one patent was submitted based on our work funded by NIJ. This patent describes the PCR primer sequences used to generate smaller amplicons for 33 different STR loci along with representative mass spec results. The sequences for multiple cleavable primers are also described although this proprietary chemistry is the subject of U.S. Patent 5,700,642, which issued in December 1997. The process of multiplexing STR loci by interleaving the alleles on a compressed mass scale is also claimed.

> *DNA typing by mass spectrometry with polymorphic DNA repeat markers.* Provisional patent filed September 19, 1997. The authors are John M. Butler, Jia Li, Joseph A. Monforte, and Christopher H. Becker.

F. Presentations Made

We have done our best to disseminate our research findings to the forensic DNA community through both presentations at forensic meetings and publications in scientific journals read by the forensic scientists. During the course of this NIJ grant, we presented our STR research results to the scientific community at the following meetings:

Eighth International Symposium on Human Identification (September 20, 1997)
San Diego Conference (November 7, 1997)
NIJ Research Committee (February 8, 1998)
American Academy of Forensic Sciences (February 13, 1998)
Southwest Association of Forensic Scientists DNA training workshop (April 23, 1998)
California Association of Criminalists DNA training workshop (May 6, 1998)
National Conference on the Future of DNA (May 22, 1998)
Florida DNA Training Session (May 22, 1998)
American Society of Mass Spectrometry (June 4, 1998)
Second European Symposium on Human Identification (June 12, 1998)
IBC DNA Forensics Meeting (July 31, 1998)
Ninth International Symposium on Human Identification (October 8, 1998)
Fourth Annual CODIS User's Group Meeting (November 20, 1998)
NIJ Research Committee (February 15, 1999)

In addition, we participated in the NIJ "Technology Saves Lives" Technology Fair held on Capitol Hill in Washington, D.C., March 30-31, 1998, an event that provided excellent exposure for NIJ to Congress. At this technology fair, we demonstrated one of our DNA sample preparation robots in the lobby of the Rayburn Building. In September 1998, we prepared an 11-minute video to illustrate some of the advantages of mass spectrometry for high-throughput DNA typing.

II. PROJECT DESCRIPTION

A. STR Grant

This project was focused on the development of a powerful new technology for rapid and accurate analysis of DNA short tandem repeat (STR) markers using time-of-flight mass spectrometry. We collaborated with the California Department of Justice (CAL-DOJ) DNA Laboratory in Berkeley, California, primarily through Dr. Steve Lee. This collaboration provided us with samples that were used to verify the new GeneTrace technology by comparing the mass spec results to genotypes obtained using established and validated methods run at CAL-DOJ.

In order to accomplish the task of developing a new mass spec technology for STR typing, five milestones were proposed in the original grant application. Our original milestones for the STR grant included the following:

a) redesign of PCR primers for a number of commonly used STR markers to produce smaller PCR products that could be tested in the mass spectrometer (*see Exhibit 2*)

b) demonstrate multiplexing capabilities to a level of 2 or 3 for detection with our method (*see Exhibits 4-5 and 20-22*)

c) transfer the sample preparation protocols from manual to a highly parallel and automated pipetting robot

d) develop a large data set to confirm the accuracy and reliability of this method (*see Exhibits 26 and 40-47*)

e) automate and incorporate DNA extraction techniques onto the GeneTrace robots

As will be described in the results section, all of these milestones were met on time except the final one regarding DNA extraction. Two other companies, Rosys and Qiagen, produced robotic systems for DNA extraction after this project began. We felt that the focus of GeneTrace should remain in developing other steps in DNA sample processing as commercially available solutions had already been developed and thus eliminated the DNA extraction portion of our studies.

The original goal of the funded project was to produce and test a highly automated mass spectrometer plus multi-tip pipetting robots with reagent kits that could be used by forensic DNA laboratories. However, over the course of this project, the business focus of GeneTrace changed from wanting to sell instruments and reagents to wanting to provide a DNA typing service to finally not pursuing the human identity market at all. During this same time, the company grew from a dozen researchers in three basement laboratories with approximately 2,000 square feet at SRI International to a multi-disciplinary group of over 65 employees in Alameda that occupy 24,000 square feet of modern lab space. A $17 million deal with Monsanto Company to develop agricultural genomics products, that was announced in March 1998, fueled the growth of GeneTrace and ultimately the change in business focus.

Since this project began in June 1997, a number of advances that impact the ability to perform high-throughput DNA typing have occurred in the biotech field. In early 1998, PE Applied Biosystems (ABI) released their dual 384-well PE9700 ("Viper") thermal cycler, which makes it possible to prepare 768 PCR samples simultaneously. Beckman Instruments also came to market with a 96-tip Multimek pipetting robot. At the beginning of this project,

GeneTrace used funds from this NIJ grant to purchase an MJ Research 384-well thermal cycler and a custom-built 96-tip robotic pipetter on a CyberLab x-y-z gantry. Both of these pieces of equipment were the state-of-the-art at the time but are now obsolete at GeneTrace for routine operations and have been replaced by the newer and more reliable products from ABI and Beckman.

B. SNP Grant

Our grant extension, which began in August 1998, focused on development of multiplexed single nucleotide polymorphisms (SNP) markers from mitochondrial DNA and the Y-chromosome. Although the grant extension was terminated prematurely by GeneTrace management in April 1999, portions of the first four milestones were accomplished. The five milestones described in our original grant extension included:

a) produce and test a set of 10 or more SNP probes for mitochondrial DNA control region "hot spots" (*see Exhibit 5*)

b) develop software for multiplex SNP analysis and data interpretation

c) examine individual Y-chromosome SNP markers

d) develop multiplex PCR and multiplex SNP probes for Y-chromosome SNP loci (*see Exhibit 70*)

e) determine the discriminatory power for a set of Y-chromosome markers by running ~300 samples across 50 Y SNP markers

The goal of our grant extension project was to develop highly multiplexed SNP assays that worked in a robust manner with mass spectrometry and could be genotyped in an automated fashion. We planned to select markers with a high degree of discrimination to aid in rapid screening of mitochondrial DNA and Y-chromosome polymorphisms with the capability to

handle analysis of large offender DNA databases. At the time this proposal was written (December 1997), GeneTrace still intended to provide reagents and instruments to large DNA service laboratories or to provide a DNA typing service to the forensic DNA community. The grant extension was prematurely terminated due to a change in business focus and a need to consolidate the research efforts at GeneTrace.

III. 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 is important for salt reduction/removal prior to the mass spectrometric analysis and has been automated on a 96-tip robotic workstation. A biotinylated, cleavable oligonucleotide is used as a primer in each assay and is incorporated through standard DNA amplification (e.g., PCR) methodologies into the final product, which is measured in the mass spectrometer. This process is covered by U.S. Patent 5,700,642, which issued in December 1997. The STR assay is schematically illustrated in **Exhibit 7** and involves a PCR amplification step where one of the primers is replaced by the GeneTrace cleavable primer. The biotinylated PCR product is then captured on streptavidin-coated magnetic beads for post-PCR sample cleanup and salt removal followed by mass spec analysis. The biology portion of the SNP assay on the other hand involves a three-step process: (a) PCR amplification, (b) phosphatase removal of nucleotides, and (c) 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. The SNP assay is illustrated in **Exhibit 8**. 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.

A. 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 over 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.

DNA sequence information was obtained from GenBank (http://www.ncbi.nlm.nih.gov) and STRBase (http://ibm4.carb.nist.gov:8800/dna/home.htm) for the STR loci and mitochondrial DNA 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). Primers were placed close to the repeat region to make the PCR product size ranges under 120 bp in size (when possible; *see Exhibit 2*) to improve the sensitivity and resolution in the mass spectrometer. 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 spec 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 13**).

(Monforte et al. 1995) or (b) performing a nested linear amplification with a terminating nucleotide (Braun et al. 1997a). These methods only work for particular situations as will be described in the results and discussion section. 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 involve 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 Da in mass due to the improved sensitivity and resolution that is obtainable in the mass spectrometer. As described above 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 makes the choice of possible marker combinations more limited. 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 ddC terminator and primer extension along the AATG strand (*see Exhibits 4 and 5*).

SNP multiplexes were designed by calculating possible post-cleavage primer and extension product masses. Multiply-charged ions were abundant in the mass range of 1,500-7,000 Da where we performed our SNP multiplex analyses and were avoided for the most part by

Primer Synthesis

Unmodified primers were purchased from Biosource/Keystone (Foster City, CA), Operon Technologies (Alameda, CA), or synthesized in-house using standard solid-phase phorphoramidite 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 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

We discovered during early attempts with primer design 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: (a) using a restriction enzyme

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 (**Exhibit 62**).

Human DNA Samples Used

Human genomic DNA samples representing several ethnic groups (African American, European, and Oriental) were purchased from Bios Laboratories (New Haven, CT) for our initial studies as primers were designed. K562 cell line DNA (Promega Corporation, Madison, WI) was used as a control in our experiments since the genotypes have been reported for this cell line with most of the STR loci we examined (*GenePrint*™ 1995).

Allelic ladders were reamplified from a 1:1000 dilution of each of the allelic ladders supplied in fluorescent STR kits from PE Applied Biosystems (ABI; Foster City, CA) using the PCR conditions listed below and the primers shown in **Exhibit 12**. The ABI kits include allelic ladders for the following STR loci: AmpF*ℓ*STR® Green I (CSF1PO, TPOX, TH01, amelogenin), AmpF*ℓ*STR® Blue (D3S1358, VWA, FGA), AmpF*ℓ*STR® Green II (amelogenin, D8S1179, D21S11, D18S51), AmpF*ℓ*STR® Yellow (D5S818, D13S317, D7S820), and AmpF*ℓ*STR® COfiler™ (amelogenin, TH01, TPOX, CSF1PO, D3S1358, D16S539, D7S820).

While most of our PCR amplifications were performed with quantitated genomic DNA in liquid form, we did a few tests with blood stained FTA™ paper (Life Technologies, Rockville, Maryland). 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 in place of the TE (10 mM Tris-EDTA) solution.

Two studies were performed with larger numbers of DNA samples. The California Department of Justice DNA Laboratory (CAL-DOJ; Berkeley, CA) provided a plate of 88 samples, which was used repeatedly for multiple STR markers. This collaboration was with Dr. Steve Lee and Dr. John Tonkyn from CAL-DOJ's DNA research laboratory. These anonymous samples had been previously genotyped by CAL-DOJ using the PE Applied Biosystems AmpF*ℓ*STR® Profiler™ kit, which consists of AmpF*ℓ*STR® Blue, Green I, and Yellow markers and thus amplifies 9 STRs and the sex-typing marker amelogenin. The work by CAL-DOJ permitted an independent verification of our results. STR allelic ladders were also provided by CAL-DOJ and were used to illustrate that the common alleles for each STR locus could be detected with our primer sets. We retyped the samples using the AmpF*ℓ*STR® 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, we examined a set of 92 human DNA templates containing 3 different CEPH families (**Exhibit 39**) and 44 unrelated individuals from the NIH Polymorphism Discovery Resource (Collins et al. 1998). These samples were typed on the ABI 310 Genetic Analyzer using both the AmpF*ℓ*STR® Profiler Plus™ and AmpF*ℓ*STR® COfiler™ kits so that all 13 CODIS STRs were covered.

PCR Reaction

In order to speed the development of new STR markers, we worked towards the development of universal PCR conditions, both in terms of 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. Our early PCR reaction mix contained 1 U Taq polymerase (Promega Corporation, Madison, WI), 1X STR buffer with dNTPs (Promega), and typically 5, 10, or 25 ng of human genomic DNA. Later we developed our own PCR mix containing 200 μM dNTPs, 50 mM KCl, 10 mM Tris-HCl, 5% gycerol, and 2 mM $MgCl_2$. Typically a locus-specific master mix was prepared by addition of 12.8 μL of PCR mix times the number of samples (+ ~10% overfill) with 0.2 μL AmpliTaq Gold™ DNA polymerase (PE Applied Biosystems) 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 setup 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 (PE Applied Biosystems) 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 non-templated 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 non-templated 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 annealing temperature around 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 61**.

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 PCR buffer II (PE Applied Biosystems), and 20 pmol each universal primer with 0.2 pmol each locus-specific primer. The universal primer sequences were 5'-ATTTAGGTGACACTATAGAATAC-3' (attached on 5'-end of locus specific forward primers) and 5'-TAATACGACTCACTATAG GGAGAC-3' (attached on 5'-end of locus specific reverse primers). **Exhibit 65** 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 69**). 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 has been tested multiple times and a level of confidence has 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. Typically, 1 U 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 (PE Applied Biosystems), 12.5 μM 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, 70 °C for 10 sec. An annealing temperature of 52 °C was used for the mtDNA 10plex SNP assay.

B. 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. 1997b) to

remove salts which 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 Dynal MPC®-E (Magnetic Particle Concentrator for Microtubes of Eppendorf Type (1.5 mL); Dynal A.S., Oslo, Norway). Larger scale experiments performed towards the end of this project utilized a robotic workstation fitted with a 96-tip pipetter that mimicked the manual method. This sample cleanup method involved washing the DNA with a series of chemical solutions intended 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 from 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 spec analysis, samples were evaporated to dryness using a speed vac and then 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 (Aldrich) in 25 mM ammonium citrate (Sigma) and 25% acetonitrile. For SNP analysis, only saturated (~0.5 M) 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 was 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 rep 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 9** shows a sample plate on the X-Y table under the custom GeneTrace ion optics.

An important feature of this automated mass spectrometer is "peak picking" software that enables the user to define "good" verses "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, a fact that 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., 1000) 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 involved several multipoint Savitzky-Golay averages along with a baseline subtraction algorithm (Carroll and Beavis 1996) to improve data quality. A set of

samples is collected under a single "header" file with identical peak picking parameters. Each header file records the mass calibration constants, peak picking parameters, and lists all of the samples analyzed with the number of good shots collected verses 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 (10998 Da) and a 55-mer (16911 Da) were typically used when examining STRs in the size range of 10,000 to 40,000 Da. On the other hand, a 15-mer (4507 Da) and its doubly-charged ion (2253.5 Da) were used to cover SNPs in the size range of 1500 Da to 7500 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 as well as instrumental challenge. The calibration was typically performed only once per day due to the fact that 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. 1998).

Delayed extraction (Vestal et al. 1995) and mass gating ("blanking") were used to improve peak resolution and sensitivity, respectively. We typically used a delay of 500-1000 nanoseconds and eliminated ions below ~8,000 Da for STRs while the delay was 250-500 nanoseconds with a signal blanking below ~1,000 Da for SNPs.

C. Sample Genotyping

Automated STR Genotyping Program (CallSSR)

During the time period of this project, GeneTrace developed an automated sample genotyping program, which was 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 at GeneTrace by a scientific programmer named Nathan Hunt in C++ 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 14**). In the case of the forensic STR loci examined in this project, the GenBank sequences were used as the reference DNA sequences.

CallSSR accepts as input smoothed, baseline-subtracted data (.sat) 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 less than one sample per second so that a plate of 384 samples can be genotyped in less than five minutes. This high rate of processing speed is necessary in a high-throughput environment where thousands of samples would need to 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 51**. Plots are generated in an artificial repeat space in order to aid visual inspection of the mass spec data compared to allele bins. The CallSSR algorithm has been written to ignore stutter peaks and doubly charged peaks, which are artifacts of the DNA amplification step and mass spec ionization process, respectively.

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 by a scientific programmer named Kevin Coopman in C++ 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 as the fitting line. In this way, peak adducts from the ionization process are distinguished from true heterozygotes. The fit coefficients for the four possible nucleotides are then compared to one another to determine the appropriate SNP base. The one 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 SSR software, a layout file, a mass file, and mass spec 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.

D. Comparison Tests with ABI 310 Genetic Analyzer

For comparison purposes, over 200 genomic DNA samples were genotyped using the PE Applied Biosystems 310 Genetic Analyzer and the AmpF*ℓ*STR® Profiler Plus™ or AmpF*ℓ*STR® COfiler™ fluorescent STR kit (PE Applied Biosystems). **Exhibit 34** 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 a 47-cm (50 μm i.d.) capillary with the "GS STR POP4 (1 mL) F" separation module. 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 1μL PCR product to 20μL 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 2.1 software programs (PE Applied Biosystems).

While standard CE conditions were used, we developed new PCR conditions to dramatically reduce the cost of using the ABI STR kits. We were able to reduce the PCR volume from the standard 50 μL described in the ABI protocol (ProfilerPlus™ and Cofiler™ user's manuals) to 5 μL, which corresponds to a cost reduction of 90% per DNA amplification. The kit reagents were mixed in their ABI-specified proportions (i.e., 11μL primer mix, 1μL TaqGold polymerase (5 U/μL), and 21μL PCR mix). A 3μL aliquot of this master mix was then added to each tube along with 2μL 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 200μL PCR tubes were sealed well to prevent evaporation. We found 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 1μL is needed for CE sample preparation and that sample can be re-injected multiple times if needed. Thus, with a 50μL PCR reaction, 49μL are never used to produce a result under the standard ABI protocol. A 5μL PCR produces less "waste" in addition to being less expensive. More importantly, the multiplex STR amplicons are more concentrated in a lower volume and thus produce 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 cycles with some DNA templates. This 5μL PCR also works 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 AmpFℓSTR® Profiler™ kit on an ABI 310 Genetic Analyzer and provided those genotypes to us for comparison purposes. These results provided an independent verification of our work.

IV. RESULTS AND DISCUSSION FOR STR ANALYSIS BY MASS SPECTROMETRY

In the course of this work, as will be described below, thousands of data points were collected using STR markers of forensic interest verifying that our mass spec technology works. During this same time, we gathered tens of thousands of data points across hundreds of different microsatellite markers from corn and soybean as part of an ongoing plant genomics partnership with Monsanto Company. Whether the DNA markers used come from humans or plants, the characteristics described below apply when analyzing polymorphic repeat loci.

A. Marker Selection and Feasibility Studies with STR Loci

Prior to receiving this grant funding, feasibility work had been completed using the STR markers TH01, CSF1PO, FES/FPS, and F13A1 in the summer and fall of 1996 (Becker et al. 1997). At the start of this project, a number of STR loci were considered as possible candidates to expand upon the initial four STR markers and to develop a set of markers that would work well in the mass spectrometer and would be acceptable to the forensic DNA community. Searches were made of publicly available databases including the Cooperative Human Linkage Center (CHLC; http://lpg.nci.nih.gov/CHLC/), the Genome Database (GDB; http://gdbwww.gdb.org/), and the Marshfield Medical Research Foundation's Center for Medical Genetics (http://www.marshmed.org/genetics/; Weber set 8 examined). The literature was also searched for possible tetranucleotide markers with PCR product sizes below 140 bp in size to avoid having to redesign the PCR primers to meet our limited size range needs (Hammond et al. 1994; Lindqvist et al. 1996). The desired characteristics also included high heterozygosity, moderate number of alleles (<7 or 8 to maintain a narrow mass

range) with no known microvariants (to avoid the need for a high degree of resolution), and balanced allele frequencies (most commonly allele <40% and least common allele >5%). We found this type of marker screen to be rather inefficient due to the fact that the original primer sets reported in public STR databases were designed for gel based separations, which are optimal over a size range of 100-400 bp. In fact, most of the PCR product sizes were in the 200-300 bp range. From a set of several thousand publicly available STRs, only a set of eight candidate tetranucleotide STRs were initially identified and three of them were tested using the original reported primers (**Exhibit 10**). Our initial goal was to identify ~25 markers that spanned all 22 autosomal chromosomes as well as the X and Y sex chromosomes.

We also quickly realized that population data was not available on these "new" markers and they would not be readily accepted without extensive testing and validation on our part. Since we wanted to produce STR marker sets that would be of value to the forensic DNA community, we decided next to examine markers already in use. We selected STR markers used by the Promega Corporation, PE Applied Biosystems, and the Forensic Science Service and redesigned primer pairs for each STR locus in order to produce smaller PCR product sizes. These STR markers included TPOX, D5S818, D7S820, D13S317, D16S539, LPL, F13B, HPRTB, D3S1358, VWA, FGA, CD4, D8S1179, D18S51, and D21S11. The primers for TH01 and CSF1PO were also redesigned to improve PCR efficiencies and to reduce the amplicon sizes. Primers for amelogenin, a commonly used sex-typing marker, were also tested (Sullivan et al. 1993). In addition, two Y-chromosome STRs, DYS19 and DYS391, were examined briefly. **Exhibit 11** summaries the STR primer sets that we developed and tested over the course of this project. With the announcement of the 13 CODIS core loci in

the fall of 1997, we switched our emphasis to these markers which are CSF1PO, TPOX, TH01, D3S1358, VWA, FGA, D5S818, D7S820, D13S317, D16S539, D8S1179, D18S51, and D21S11 as well as the sex-typing marker amelogenin.

The newly designed GeneTrace primers produce PCR products that are smaller than those commercially available from PE Applied Biosystems or Promega (*see Exhibit 2*), yet result in identical genotypes almost all samples tested (see below). For example, we obtained the correct genotypes on the human cell line K562, a commonly used control for PCR amplification success. **Exhibit 16** shows the K562 results for CSF1PO, TPOX, TH01, and amelogenin. These results were included as part of a publication demonstrating that time-of-flight mass spectrometry could perform accurate genotyping of STRs without allelic ladders (Butler et al. 1998b).

Caveats for STR Analysis by Mass Spec

While our mass spec results worked well for a majority of the STR markers tested, there are limitations that exclude some STRs from working effectively by mass spectrometry. Two important issues that impact mass spectrometer results are DNA size and sample salts. Mass spectrometer resolution and sensitivity are diminished when either the DNA size or the salt content of the sample is too large. By designing the PCR primers to bind close to the repeat region, the STR allele sizes are reduced so that resolution and sensitivity of the PCR products are benefited. In addition, the GeneTrace-patented cleavage step reduces the measured DNA size even further. Where possible, we currently design our primers to produce amplicons that are less than 120 bp although we sometimes work with STR alleles that are as large as 140 bp in size. This limitation in size prevents reliable analysis of STR markers with samples

containing a large number of repeats such as most of the FGA, D21S11, and D18S51 alleles (*see Exhibit 2*).

To overcome the sample salt problem, we use a patented solid-phase purification procedure that reduces the concentration of magnesium, potassium, and sodium salts in the PCR products prior to their being introduced to the mass spectrometer (Monforte et al 1997b). Without the reduction of the salts, resolution is diminished by the presence of adducts. Salt molecules bind to the DNA during the MALDI ionization process and give rise to peaks that have a mass of the DNA molecule plus the salt molecule. Adducts broaden peaks and thus reduce peak resolution. Our sample purification procedure, which has been entirely automated on a 96-tip robotic workstation, reduces the PCR buffer salts and yields "clean" DNA for the mass spectrometer. Appropriate care must be taken to prevent samples from being contaminated with salts both during and after the sample purification procedure.

Size Reduction Methods

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: (a) using a restriction enzyme (Monforte et al. 1995) or (b) performing a nested linear amplification with a terminating nucleotide (Braun et al. 1997a). Both methods have pros and cons. A restriction enzyme, *Dpn*II, which recognizes the sequence 5'...^GATC...3', was used with VWA samples to remove 45 bp from each PCR product. For example, the GenBank allele which contains 18 repeat units and is 154 bp following PCR amplification may be reduced to 126 nucleotides following primer cleavage but can be shortened to 81 nucleotides if primer cleavage is combined with *Dpn*II

digestion. At 81 nucleotides or 25482 Da, the STR product size is much more manageable in the mass spectrometer. This approach works nicely provided the restriction enzyme recognition site remains unchanged. The *Dpn*II digestion of VWA amplicons worked on all samples tested including a reamplication of an allelic ladder from PE Applied Biosystems (*see Exhibit 17*). However, the cost and time of analysis are increased with the addition of a restriction enzyme step.

The second approach that we explored for reducing the overall size of the DNA molecule examined in the mass spectrometer involved using a single dideoxynucleotide triphosphate (ddNTP) with three regular deoxynucleotide triphosphates (dNTPs). A linear amplification extension reaction was performed with the ddNTP terminating the reaction on the opposite side of the repeat from the cleavable primer. However, there are several limitations with this "single base sequencing" approach. First, it will only work if the repeat does not contain all four nucleotides. For example, a nucleotide mixture of ddC with dA, dT, and dG will allow extension through an AATG repeat (as occurs in the bottom strand of TH01) but will terminate at the first C nucleotide in a TCAT repeat (the top strand of TH01). Thus one is limited with the DNA strand that can be used for a given combination of dideoxynucleotide and corresponding deoxynucleotides. In addition, primer position and STR sequence content is important. If a ddC mix is used, the DNA sample cannot contain any C nucleotides prior to the repeat region as well as no Cs within the repeat or the extension will prematurely halt and the information content of the full repeat will not be accurately captured. In most cases, this requires that the extension primer should be immediately adjacent to the STR repeat, a situation that is not universally available due to the flanking sequences around the repeat

region. For example, this approach will work with TH01 (AATG) but not VWA, which has three different repeat structures: AGAT, AGAC, and AGGT. Thus, with VWA a ddC would extend through the AGAT repeat but be prematurely terminated at the C in the AGAC repeat—and valuable polymorphic information would be lost.

The use of a terminating nucleotide also provides a sharper peak for an amplified allele compared to the split peaks or wider peaks (if resolution is poor) that can result from partially adenylated amplicons (i.e., -A/+A). **Exhibit 19** illustrates the advantage of a ddG termination on a D8S1179 heterozygous sample containing 11 and 13 TATC repeats. In the bottom panel, 23 nucleotides have been removed as compared to the top panel, which corresponds to a mass reduction of almost 8,000 Da. The peaks are sharper in the lower panel as the products are blunt ended. Identical genotypes were obtained with both approaches illustrating that the ddG termination is occurring at the same point on the two different sized alleles.

To summarize, STR sample sizes were reduced using primers that have been designed to bind close to the repeat region or even partially on the repeat itself. A cleavable primer was incorporated into the PCR product to allow post-PCR chemical cleavage and subsequent mass reduction. Two additional post-PCR methods were also explored to further reduce the measured DNA size. These methods included restriction enzyme digestion in the flanking region on the other side of the repeat region from the cleavable primer and a primer extension through the repeat region with a single dideoxynucleotide terminator (single base sequencing approach). As an illustration of the advantages of these approaches to reduce the overall DNA product mass, we will examine the STR locus TPOX. Using a conventional primer set,

a sample containing 11 repeats would measure 232 bp or ~66,000 Da. By redesigning the primers to anneal close to the repeat region, we obtain a PCR product of 89 bp. With the cleavable primer, the size is reduced to 69 nucleotides (nt) or 21351 Da. By incorporating a ddC termination reaction, another 20 nt are removed leaving only 49 nt or ~12,000 Da (primarily only the repeat region). The repeat region alone is 44 nt (4 nt x 11 repeats) or ~10,500 Da. The ddC termination was also used in multiplex STR analysis to produce a CSF1PO-TPOX-TH01 triplex (*see Exhibits 4 and 5*). The repeat sequences used for these STR loci were AGAT for CSF1PO, AATG for TPOX, and AATG for TH01. The level of sequence clipping by ddC is as follows: CSF1PO (-14 nt), TPOX (-20 nt), and TH01 (-4 nt).

B. Multiplex STR Work

Due to the limited size range of DNA molecules that may be analyzed by this technique, we developed a new approach to multiplexing that involved interleaving alleles from different loci rather than producing non-overlapping multiplexes. If the amplicons can be kept under ~25,000 Da then high the high degree of mass accuracy and resolution can be used to distinguish alleles from multiple loci that may differ by only a fraction of a single nucleotide. (*see Exhibit 22*). Allelic ladders are useful to demonstrate that all alleles in a multiplex are distinguishable (*see Exhibit 21*).

The expected masses for a triplex involving the STR loci CSF1PO, TPOX, and TH01 (commonly referred to as a CTT multiplex) are schematically displayed in **Exhibit 4**. All known alleles for these STR loci, as defined by STRBase (Butler et al 1998a), are fully resolvable and far enough apart to be accurately determined. For example, TH01 alleles 9.3

and 10 fall between CSF1PO alleles 10 and 11. For all three STR systems in this CTT multiplex, the AATG repeat strand is measured, which means that the alleles *within* the same STR system differ by 1260 Da. The smallest spread between alleles *across* multiple STR systems in this particular multiplex exists between the TPOX and TH01 alleles where the expected mass difference is 285 Da. TPOX and CSF1PO alleles differ by 314 Da while TH01 and CSF1PO alleles differ by 599 Da. By using the same repeat strand in the multiplex, the allele masses between STR systems all stay the same distance apart. Each STR has a unique flanking region and it is these sequence differences between STR systems that permit multiplexing in such a fashion as described here. An actual result with this CTT multiplex is shown in **Exhibit 5**. This particular sample is homozygous for both TPOX (8,8) and CSF1PO (12,12) and heterozygous at the TH01 locus (6,9.3).

It is also worth noting that this particular CTT multiplex was designed to account for possible, unexpected microvariants. For example, a CSF1PO allele 10.3 that appears to be a single base shorter than CSF1PO allele 11 was recently reported (Lazaruk et al 1998). With the CTT multiplex primer set described here, a CSF1PO 10.3 allele would have an expected mass of 21402 Da, which should be fully distinguishable from the nearest possible allele (i.e., TH01 allele 10) as these alleles would be 286 Da apart. Using our mass window of 100 Da as defined by our previous precision studies (Butler et al 1998b), all possible alleles including microvariants should be fully distinguishable. STR multiplexes are designed so that expected allele masses between STR systems are offset in a manner that possible microvariants, which are most commonly insertions or deletions of a partial repeat unit, may be distinguished from all other possible alleles. The larger the allele mass range, the more difficult it becomes to

maintain a high degree of mass accuracy. For example, **Exhibit 20** shows the observed mass for TH01 allele 9.3 is –52 Da from its expected mass while TPOX allele 9 is only 3 Da from its expected mass. In this particular case, the mass calibrants used were 4507 Da and 10998 Da. Thus, the TPOX allele mass measurement is more accurate as it is closer to the calibration standard. The ability to design multiplexes that have a relatively compact mass range is important to maintaining the high level of mass accuracy needed for closely spaced alleles from different, overlapping STR loci. The mass calibration standards should also span the entire region of expected measurement in order to guarantee the highest degree of mass accuracy.

Two possible multiplexing strategies for STR genotyping are illustrated in **Exhibit 23**. Starting with a single punch of blood stained FTA paper, it is possible to perform a multiplex PCR (simultaneously amplifying all STRs of interest) followed by a second-round PCR with primer sets that are closer to the repeat region. With this approach, single or multiplexed STR products can be produced that are small enough for mass spec analysis. Alternatively, multiple punches could be made from a single bloodstain on the FTA paper followed by singleplex or multiplex PCR with mass spec primers. After the genotype is determined for each STR locus in a sample, the information would be combined to form a single sample genotype for inclusion in CODIS or some other DNA database. This multiplexing approach permits flexibility for adding new STR loci or only processing a few STR markers across a large number of samples at a lower cost than processing extensive and inflexible STR multiplexes.

C. Comparison Tests Between ABI 310 and Mass Spec Results

A plate of 88 samples from the California Department of Justice DNA Laboratory was tested with 10 different STR markers and compared to results obtained using the ABI 310 Genetic Analyzer and commercially available STR kits. The samples were supplied as a 200 µL aliquot of extracted genomic DNA in a 96 well tray with each sample at a concentration of 1 ng/µL. A 5 µL aliquot was used for each PCR reaction, or 5 ng total per reaction. Since each marker was amplified and examined individually, approximately 35 ng of extracted genomic DNA was required to obtain genotypes on the same 7 markers as are amplified in a single AmpF*ℓ*STR® COfiler™ STR multiplex. Only 2 ng of genomic DNA were used per reaction with the AmpF*ℓ*STR® COfiler™ kit. Thus, a multiplex PCR reaction is much better suited for situations where the quantity of DNA is limited (e.g., crime scene sample). However, in most cases involving high-throughput DNA typing (e.g., offender database work), hundreds of nanograms of extracted DNA would be easily available.

A major advantage for the mass spec approach is speed of the technique and the high-throughput capabilities when combined with robotic sample preparation. The data collection times required for these 88 CAL DOJ samples using the ABI 310 Genetic Analyzer and our mass spec method are compared in **Exhibit 26**. While it took the ABI 310 almost 3 days to collect the data for the 88 samples, the same genotypes were obtained on the mass spec in less than two hours. Even the ability to analyze multiple STR loci simultaneously with different fluorescent tags on the ABI 310 could not match the speed of our mass spec data collection with each marker run individually.

To verify that the mass spec approach produces accurate results, we have performed comparison studies on the genotypes obtained from the two different methods across 8 different STR loci. **Exhibits 40-47** contain a direct comparison with 1408 possible data points (2 methods x 88 samples x 8 loci). With a few minor exceptions as will be described below, there was almost 100% correlation between the two methods. In addition to the 8 loci where we had data from both the ABI 310 and the mass spec, we measured two additional markers by mass spec, namely D8S1179 and DYS391, across these same 88 samples where we did not obtain results from a fluorescent STR primer set (**Exhibits 48-49**). Both of the D8S1179 and the DYS391 primer sets worked extremely well in the mass spectrometer. We believe that if results were made available on these samples with fluorescent STR primer sets (e.g., D8S1179 is in the AmpF*ℓ*STR® Profiler Plus™ kit), there would also be a further correlation between the two methods.

D. PCR Issues

Null Alleles

When making comparisons between two methods that use different PCR primer sets, the issue at hand is whether or not a different primer set for a given STR locus will result in different allele calls through possible sequence polymorphisms in the primer binding sites. In other words, do primers used for mass spectrometry that are closer to the repeat region than those primers used in fluorescent STR typing, yield the same genotype? Differences between primer sets are possible if there are sequence differences outside the repeat region that occur in the primer binding region of either set of primers (**Exhibit 74**). This phenomenon produces

what is known as a "null" allele, or in other words, the DNA template exists for a particular allele but fails to amplify during PCR due to primer hybridization problems. In all cases, except the STR locus D7S820, there was excellent correlation in genotype calls between the two methods (where mass spec and CE results were obtained) signifying that the mass spec primers did not produce any null alleles.

For the STR locus D7S820, a total of 17 out of 88 samples did not agree between the two methods (**Exhibit 46**). The bottom two panels in **Exhibit 53** illustrate that there is more micro heterogeneity at this locus than has previously been reported. On the lower left plot, only the allele 10 peak can be seen; allele 8, which was seen with PCR amplification using a fluorescent primer set, is missing (*see position of red arrow in Exhibit 53*). On the lower right plot, both allele 8 and allele 10 are amplified and detected in the mass spectrometer confirming that the problem is with the PCR amplification not the mass spec data collection. In this particular case, there is a difference between those two allele 8's, meaning that our mass spec primer set identified a new, previously unreported allele. When using fluorescent primer sets that anneal 50 to 100 bases or more from the repeat region are used, a single base change (e.g., T-to-C) out of a 300 bp PCR product is difficult to detect. Upon examining the mass spec data where there were missing alleles compared to the ABI 310 results, we noted that the situation only occurred with some allele 8's, 9's, and 10's (*see underlined alleles in the ABI 310 column of Exhibit 46*). Thus, these null alleles were variants of alleles with 8, 9, or 10 repeats. Most likely, a sequence microvariant occurs within the repeat region near the 3'-end of our reverse primer, which anneals to two full repeats. Unfortunately, we did not have the time to gather sequence information for these samples to confirm the observed

variation. Interestingly enough, the D7S820 locus has been reported to cause similar null allele problems with other primer sets (Schumm et al. 1997).

Microvariants

Sequence variation between alleles can take the form of insertions, deletions, or nucleotide changes. Alleles containing some form of sequence variation compared to more commonly observed alleles are often referred to as microvariants because they are slightly different from full repeat alleles. For example, the STR locus TH01 contains a 9.3 allele, which has 9 full repeats (AATG) and a partial repeat of 3 bases (ATG). In this particular example, the 9.3 allele differs from the 10 allele by a single base deletion of adenine. Microvariants exist for most STR loci and are being identified in greater numbers as more samples are being examined around the world. In the work performed here, we discovered three previously unreported STR microvariants (**Exhibit 37**) during the analysis of 38 genomic DNA samples from a male population data set provided by Dr. Peter Oefner of Stanford University (**Exhibit 38**). These microvariants occurred in the three most polymorphic STR loci, namely FGA, D21S11, and D18S51, that possess the largest and most complex repeat structures.

The ability to make accurate mass measurements with mass spectrometry is a potential advantage when locating new microvariants. If the mass precision is good, then any peaks that have large offsets from the expected full repeat alleles could be suspect microvariants in the form of insertions or deletions because their masses would fall outside the expected variance due to instrument variation. We found this possibility to be true especially when working with heterozygous samples. Microvariants can be detected by using the mass

difference between the two alleles and comparing this value to that expected for full repeats or by comparing the allele peak mass offsets. If the peak mass offsets shift together, then both alleles are full repeats, but if one of the peak mass offsets is significantly different (e.g., ~300 Da) than the other one, a possible insertion or deletion exists in one of the alleles. **Exhibit 29** illustrates this concept by plotting the mass offset (from a calculated allele mass) of allele 1 verses the mass offset (from a calculated allele mass) of allele 2. Note that the 9.3 alleles for TH01 cluster away from the comparison of full repeat verses full repeat alleles while the results for the other three STR loci, which have no known microvariants in this data set, have mass offsets that shift together for the heterozygous alleles. **Exhibit 28** compares the peak mass offsets for the amelogenin X allele verse the Y allele and demonstrates that full "repeats" shift together during mass spec measurements.

Non-Template Addition

DNA polymerases, particularly the *Taq* polymerase used in PCR, often add an extra nucleotide to the 3'-end of a PCR product as they are copying the template strand. This non-template addition, which is most often an adenine hence the term "adenylation", can be favored by adding a final incubation step at 60 °C or 72 °C after the temperature cycling steps in PCR (Clark 1988, Kimpton et al. 1993). However, the degree of adenylation is dependent on the sequence of the template strand, which in the case of PCR results from the 5'-end of the reverse primer. Thus, every locus will have different adenylation properties because the primer sequences are different.

Now why is all of this important? From a measurement standpoint, it is better to have all of the molecules as similar as possible for a particular allele. Partial adenylation, where some of the PCR products do not have the extra adenine (i.e., -A peaks) and some do (i.e., +A peaks), can contribute to peak broadness (*see Exhibit 19 top panel*) if the separation system's resolution is poor. Sharper peaks improve the likelihood that a system's genotyping software can make accurate calls. Variation in the adenylation status of an allele across multiple samples can have an impact on accurate sizing and genotyping potential microvariants. For example, a non-adenylated TH01 10 allele would look the same as a fully adenylated TH01 9.3 allele in the mass spectrometer because their masses are identical. Therefore, it is beneficial if all PCR products for a particular amplification are either +A or –A rather than a mixture (e.g., ±A). By using the temperature soak at the end of thermal cycling, most of the STR loci we examined were fully adenylated with the notable exception of TPOX, which was typically non-adenylated, and TH01, which under some PCR conditions produced partially adenylated amplicons. For making correct genotype calls, the STR mass ladder file (**Exhibit 14**) was altered according to the empirically determined adenylation status.

During the course of this project, a new DNA polymerase became available from Life Technologies that exhibits little-to-no non-template nucleotide addition. We decided to test this PLATINUM® GENOTYPE™ *Tsp* DNA polymerase with STR loci that had been shown to produce partial adenylation to see if the +A peak could be eliminated. **Exhibit 31** compares mass spec results obtained using AmpliTaq Gold (commonly used) polymerase with the new *Tsp* polymerase. The *Tsp* polymerase produced amplicons with only the –A peaks while TaqGold showed partial adenylation with these TH01 primers. Thus, this new polymerase

much smaller than the fluorescently labeled primer sets used by most forensic DNA laboratories (*see Exhibit 2*). However, this hypothesis needs to be studied more extensively with multiple primer sets on a particular STR locus that generate various sized amplicons. For example, the primer sets described in Exhibit 12 could be fluorescently labeled and analyzed on the ABI 310 where the stutter product peak heights could be quantitatively compared to the allele peak heights.

The more likely reason that less stutter is observed by mass spectrometry is that the signal-to-noise ratio is much lower in mass spec compared to fluorescence measurements. Fluorescence techniques have a much lower background and because of this are more sensitive than mass spectrometry for the detection of DNA. Thus, the stutter products may be present at similar ratios compared to those observed with fluorescence measurements but they are in the baseline noise of the mass spec data and therefore not seen in the mass spectrum. This later explanation is probably more likely as we have seen very strong stutter peaks for some dinucleotide repeat markers (**Exhibit 33**). Whether stutter products are present or not, our current STR genotyping software has been designed to recognize them and not call them as alleles.

Primer Sequence Determinations from Commercial STR Kits

Two commercial manufacturers supply STR kits to the forensic DNA community: Promega Corporation and PE Applied Biosystems. Those kits come with PCR primer sequences that permit multiplex PCR amplification of up to 10 STR loci simultaneously. One of the primers for each STR locus is labeled with a fluorescent dye to permit fluorescent detection of the

has the potential to produce sharper peaks (i.e., no partial adenylation) and allele masses that can be more easily predicted (i.e., all PCR products would be non-adenylated).

Stutter Products

During PCR amplification of STR loci, repeat slippage can occur and result in the loss of a repeat unit as DNA strand synthesis occurs through a repeated sequence. These stutter products are typically 4 bases, or one tetranucleotide repeat, shorter than the true allele PCR product. The amount of stutter product compared to the allele product is variable depending on the STR locus and the length of the repeat, but typically stutter peaks are 2-10% of the allele peak height (Walsh et al. 1996). Forensic DNA scientists are concerned about stutter products because their presence can interfere in the interpretation of DNA mixture profiles.

When forensic scientists see plots of our mass spec results for STR loci, one of the first comments is that there appears to be a reduced level of stutter product detection (*see Exhibit 16*). There are several possibilities of why we see less stutter in our STR results. The fact that we are amplifying smaller PCR products because our primers are closer to the repeat region means that the DNA polymerase does not have to hold on to the extending strand as long for synthesis purposes. It is possible that lower stutter could result because the polymerase reads through the repeat region "faster" and the template strands do not have as much of an opportunity to slip and reanneal out of register on the repeat region. For example, *Taq* polymerase has a processivity rate of ~60 bases before it falls off the extending DNA strand and therefore the closer the PCR product size is to 60 bases the better for the extension portion of the PCR cycle. Our PCR product sizes, which are typically less than 100 bp, are

labeled PCR products. Since the primer sequences have not been disclosed by the manufacturers and we wanted to know where they annealed to the STR sequences compared to our primers (see discussion on null alleles above), we decided to use mass spectrometry as a tool for the determination of these STR primer sequences.

First, the primer mixtures were spotted and analyzed to determine each primer's mass (**Exhibit 54**, top panel). Then a 5'→3' exonuclease was added to the primer mix and heated to 37 °C for several minutes to digest the primer one base at a time. An aliquot was removed every 5-10 minutes to obtain a time course on the digestion reaction. Each aliquot was spotted in 3-hydroxypicolinic acid matrix solution, allowed to dry, and then analyzed in the mass spectrometer. The digestion reaction produces a series of products that differ by one nucleotide. By measuring the mass difference between each peak, the original primer sequence may be determined (**Exhibit 54**, bottom panel). Only the unlabeled primers will be digested because the covalently attached fluorescent dye blocks the 5'-end of the dye-labeled primer. Using only a few bases of sequence (e.g., 4-5 bases), it is possible to make a match on the appropriate STR sequence obtained from GenBank and therefore determine the 5'-end of the primer without the fluorescent label. With the full-length primer mass obtained from the first experiment, the remainder of the unlabeled primer can be identified. The position of the 5'-end of the other primer can be determined using the GenBank sequence and the PCR product length for the appropriate STR allele listed in GenBank (*see Exhibit 2*). The sequence of the labeled primer can be ascertained by using the appropriate primer mass determined from the first experiment and subtracting the mass of the fluorescent dye. The primer mass is then used to obtain the correct length of the primer on the GenBank sequence and thereby the

primer's sequence. Finally, an entire STR multiplex primer set can be measured together in the mass spec to observe the primer balance (**Exhibit 55**). HPLC fraction collection can be used to pull primers apart from complex multiplex mixtures and each primer identified as described above. The primer sequences from both Promega and PE Applied Biosystems for the STR loci TH01 (**Exhibit 56**), TPOX (**Exhibit 57**), and CSF1PO (**Exhibit 58**) were identified using this procedure. A comparison of the primer sequences from the two manufacturers found that they were very similar. The 3'-ends of the primer sets, which are the most critical portions for annealing during PCR, are almost identical between the different kits. The PE Applied Biosystem primers were typically shorter at the 5'-end and therefore produced PCR products that were ~10 bases shorter than those produced by the corresponding Promega primers. In all three STR loci, the primers annealed further away from the repeat region than the GeneTrace primer sets.

E. Analytical Capabilities of This Mass Spectrometry Method

Using our current primer design strategy, most STR alleles range in size from ~10,000 Da to ~40,000 Da. In mass spectrometry, the smaller the molecule, the easier it is to ionize and detect (all other things being equal). Resolution, sensitivity, and accuracy are usually better the smaller the DNA molecule being measured. Because the possible STR alleles are relatively far apart, reliable genotyping is readily attainable even with DNA molecules at the higher mass region of the spectrum. For example, neighboring full-length alleles for a tetranucleotide repeat such as AATG differ in mass by 1260 Da.

Resolution

Dinucleotide repeats, such as CA repeats, require a resolution of at least 2 bp in order to resolve stutter products from the true allele or heterozygotes that differ by a single repeat. Trinucleotide and tetranucleotide repeats, with their larger repeat structure, are more easily resolved as there is a larger mass difference between adjacent alleles. However, the overall mass of the PCR product increases more rapidly with tri- or tetranucleotide repeats. For example, the repeat region for 40 GA repeats is 25,680 Da while the mass of the repeat region quickly increases to 37,200 Da for 40 AAT repeats and 50,400 Da for 40 AATG repeats. GeneTrace has demonstrated that a resolution of a single dinucleotide repeat (~600 Da) may be obtained for DNA molecules up to a mass of ~35,000 Da. This reduced resolution at higher mass presents a problem for polymorphic STR loci such as D18S51, D21S11, and FGA as single base resolution is often required to accurately call closely spaced alleles or to distinguish a microvariant containing a partial repeat from a full length allele. These three STR loci also contain long alleles. For example, D21S11 has reported alleles of up to 38 repeats (mixture of TCTA and TCTG) in length, D18S51 has alleles up to 27 AGAA repeats, and FGA has alleles up to 50 repeats (mixture of CTTT and CTTC). Heterozygous FGA alleles that differed by only a single repeat were more difficult to genotype accurately than smaller sized STR loci due to poor resolution at masses greater than ~35,000 Da (**Exhibit 47**; *see samples marked in red*).

We have found analysis of STR allelic ladders to be useful for demonstrating that all alleles can be resolved for an STR locus. Allelic ladders from commercial kits were typically diluted 1:1000 with deionized water and then reamplified with the GeneTrace primers that bound

closer to the repeat region than the primers from the commercial kits. This reamplification provided PCR products for demonstrating that the needed level of resolution (i.e., distinguishing adjacent alleles) is capable at that mass range in the mass spectrometer as well as demonstrating that the GeneTrace primers amplify all alleles (i.e., no allele dropout from a null allele). A number of STR allelic ladders were tested in this fashion including TH01 (**Exhibit 15**), CSF1PO (**Exhibit 17**), TPOX (**Exhibit 17**), VWA (**Exhibit 17**), and D5S818 (**Exhibit 18**). All tetranucleotide repeat alleles were resolvable in these examples demonstrating 4 bp resolution and with TH01 single base pair resolution was seen between alleles 9.3 and 10.

Sensitivity

To determine the sensitivity of our STR typing assay, we tested the TPOX primers with a dilution series of K562 genomic DNA (20 ng, 10 ng, 5 ng, 2 ng, 1 ng, 0.5 ng, 0.2 ng, and 0 ng). Promega Taq polymerase and STR buffer were used with 35 PCR cycles as described in the scope and methods section. Peaks for the correct genotype (heterozygote 8,9) could be seen down to the lowest level tested (0.2 ng or 200 picograms) while the negative control was blank. **Exhibit 25** contains a plot with the mass spectra for 20 ng, 5 ng, 0.5 ng, and 0 ng. While each PCR primer pair can exhibit a slightly different efficiency, we believe that human DNA down to a level of ~1 ng can be reliably PCR amplified and detected using mass spectrometry. Our more recent protocol involved 40 cycle PCR and the use of TaqGold™ DNA polymerase, a fact that should improve overall yield for STR amplicons. All of the samples tested from CAL DOJ were amplified with only 5 ng of DNA template and yielded excellent results (*see Exhibits 40-53*). In terms of absolute sensitivity in the mass

spectrometer, we found that typically several hundred femtomoles of relatively salt free DNA molecules were necessary for detection. Our PCR amplifications normally produced several picomoles of PCR product, or approximately an order of magnitude more material than is actually needed for detection.

Mass Accuracy and Precision

Mass accuracy is an important issue for this mass spectrometry approach to STR genotyping as a measured mass for a particular allele is compared to an ideal mass for that allele. Due to the excellent accuracy of mass spectrometry, internal standards are not required to obtain accurate DNA sizing results as in gel or capillary electrophoresis measurements (Butler et al. 1998b). It should be noted that to make an inaccurate genotype call for a tetranucleotide repeat, the mass offset from an expected allele mass would have to be larger than 600 Da (half the mass of a ~1200 Da repeat). GeneTrace has observed mass accuracies on the order of 0.01 nucleotides (<3 Da) for STR allele measurements. However, under routine operation with our automated mass spectrometers, some resolution, sensitivity, and accuracy are sacrificed compared to a research grade instrument to deliver data at a high rate of speed. Almost all STR allele size measurements should be within ± 200 Da, or a fraction of a single nucleotide, of the expected mass. **Exhibit 27** illustrates that the precision and accuracy for STR measurements is good enough to make accurate genotyping calls with only a routine mass calibration even when comparing data from the same samples collected months apart in time.

Precision is important for STR allele measurements in mass spectrometry because no internal standards are being run with each sample to make adjustments for slight variations in

instrument conditions between runs. To demonstrate the excellent reproducibility of mass spectrometry, we collected 15 mass spectra of a TPOX allelic ladder. A table of the obtained masses for alleles 6,7,8, 9, 10, 11, 12, and 13 show that all alleles were easily segregated and distinguishable (**Exhibit 24**). Statistical analysis of the data found that the standard deviation about the mean for each allele ranged from 20 to 27 Da or approximately 0.1% relative standard deviation (RSD). The mass between alleles is equal to the repeat unit, which in the case of TPOX is 1260 Da for an AATG repeat (*see also Exhibit 32*). Thus, each allele is easily distinguishable.

We have made measurements of the same DNA samples over a fairly wide time range and found the masses to be remarkably similar even when data points were recollected months later. **Exhibit 27** compares 57 allele measurements for six different TPOX alleles collected six months apart. The first data set was collected on October 1, 1998, and the second data set on March 26, 1999. Amazingly enough some of the alleles had identical measured masses even though different mass calibration constants (and even different instruments) were used. However, the bottom line is whether or not a correct genotype can be obtained using this new technology. As can be seen in **Exhibit 73**, which compares the genotypes obtained using a conventional CE separation method and this mass spec technique across 3 STR markers, excellent agreement is obtained between the methods. With the CAL DOJ samples tested, there was complete agreement on all observed genotypes for the STR loci CSF1PO, TH01, and D3S1358 as well as the sex-typing marker amelogenin (*see* **Exhibits 40,42-44**). Some "gas-phase" dimers and trimers fell into the allele mass range and confused the calling for TPOX (**Exhibit 41**) and D16S539 (**Exhibit 45**) on several samples. Gas-phase dimers or

trimers are assay artifacts that result from multiple excess primer molecules colliding in the gas-phase and being ionized during the MALDI process. A mass offset plot like that shown in **Exhibit 29** can be used to detect these assay artifacts as they fall outside the tight grouping inside the 300 Da window. With the CAL DOJ samples, D7S820 exhibited null alleles (**Exhibit 46**) and FGA has some unique challenges due to its larger size and therefore problems with resolution of closely spaced heterozygotes and poorer mass calibration since the measured alleles are further away from the calibration standards (**Exhibit 47**). Thus, when the PCR situations such as null alleles are accounted for and smaller loci are used, this mass spec method produces comparable results to traditional methods for STR genotyping.

Data Collection Speed

The tremendous speed advantage of mass spectrometry can be seen in **Exhibit 26**. Over the course of this project our data collection speed increased by a factor of 10 from ~50 seconds/sample to less than 5 seconds/sample. This speed increase resulted from improved software and hardware on the automated mass spectrometers and from improved sample quality (better PCR conditions that yielded more product and improved sample cleanup that yielded "cleaner" DNA). With data collection times around 5 seconds per sample, achieving sample throughputs of almost 1,000 samples per hour is possible and 3,000-4,000 samples per system per day when operating at full capacity is reasonable. Sample backlogs could be erased rather rapidly with this kind of throughput. By way of comparison it takes on average 5 minutes to obtain each genotype (assuming a multiplex level of 6 or 7 STRs) with conventional capillary electrophoresis methods (*see Exhibit 26*). Thus, the mass spectrometric method described in this work is two orders of magnitude faster in sample processing time than conventional techniques!

V. RESULTS AND DISCUSSION WITH MULTIPLEX SNPs

We began work on development of multiplexed single nucleotide polymorphism assays in the summer of 1998 after notice that a second NIJ grant entitled "Development of Multiplexed Single Nucleotide Polymorphism Assays from Mitochondrial and Y-Chromosome DNA for Human Identity Testing Using Time-of-Flight Mass Spectrometry" had been funded. Excellent progress was made towards the milestones on this grant but the work was not finished when this grant was prematurely terminated on the part of GeneTrace in the spring of 1999. The completed work was focused in two areas: the development of a 10-plex SNP assay from the mitochondrial DNA control region using a single amplicon and the development of a multiplex PCR assay from Y chromosome SNP markers that involved as many as 18 loci amplified simultaneously. In this section, we will describe the design aspects of multiplex PCR and SNP assays along with the progress that we made towards the goal of producing assays that would be useful for high-throughput screening of mitochondrial and Y chromosome SNP markers.

The approach to SNP determination described here has essentially three steps: (1) PCR amplification, (2) phosphatase digestion, and (3) SNP primer extension. Either strand of DNA may be probed simultaneously in this SNP primer extension assay. PCR primers are designed to generate an amplicon that includes one or more SNP sites. The initial PCR reaction is performed with standard (unlabeled) primers. A phosphatase is then added following PCR to remove all remaining dNTPs so that they will not interfere with the single base extension reaction involving dideoxynucleotides. These reactions can all be performed

in the same tube or well in a sample tray. A portion of the phosphatase-treated PCR product is then used for the primer extension assay.

In our SNP primer extension assay, a special primer containing a biotin moiety at the 5'-end and a cleavable nucleotide near the 3'-end hybridizes upstream of the SNP site with the 3'-end immediately adjacent to the SNP polymorphic site (*see Exhibit 8*). The biotin permits solid-phase capture for sample purification prior to mass spec analysis, and the cleavable nucleotide allows the 3'-end of the primer to be released from the immobilized portion as well as reducing the overall mass of the measured DNA molecule (Li et al. 1999). The complementary nucleotide(s) to the nucleotide(s) present at the SNP site is inserted during the extension reaction. In the case of a heterozygote, two extension products result. Only a single base is added to the primer during this process because only ddNTPs are used and the dNTPs left over from PCR are hydrolyzed with the phosphatase digestion step. If the extension reaction is not driven to completion (where the primer would be totally consumed), then both primer and extension product (i.e., primer plus single nucleotide) are present after the primer extension reaction. The mass difference between these two DNA oligomers is used to determine the nucleotide present at the SNP site. In our primer extension SNP assay, the primer acts as an internal standard and helps make the measurement more precise and more accurate. A histogram of mass difference measurements across 200 samples (50 per nucleotide) is shown in **Exhibit 64**. The ddT and ddA differ by only 9 Da and are the most difficult to resolve as heterozygotes or distinguish from one another in terms of mass. As reported in a recently published paper (Li et al. 1999), we have used this approach to reliably determine all four possible SNP homozygotes and all six possible heterozygotes.

A. Mitochondrial DNA Work

The control region of mitochondrial DNA, commonly referred to as the D-loop, is highly polymorphic and contains a number of possible SNP sites for analysis. MITOMAP, an internet database containing fairly comprehensive information on mtDNA, lists 408 polymorphisms over 1121 nucleotides of the control region (positions 16020-576) that have been reported in the literature (MITOMAP 1999). However, many of these polymorphisms are rare and population specific. We decided to focus on selecting a marker set from a few dozen well-studied potential SNP sites. Special Agent Mark Wilson from the FBI Laboratory in Washington D.C., who has been analyzing mtDNA for over 7 years, recommended a set of 27 SNPs that he felt would give a reasonable degree of discrimination and make our assay about half as informative as full sequencing. His recommended mtDNA sites were positions 16069, 16114, 16126, 16129, 16189, 16223, 16224, 16278, 16290, 16294, 16296, 16304, 16309, 16311, 16319, 16362, 73, 146, 150, 152, 182, 185, 189, 195, 198, 247, and 309. The underlined sites are those reported in a minisequencing assay developed by the Forensic Science Service (Tully et al. 1996). We decided to start with ten of the SNPs used in the Forensic Science Service (FSS) minisequencing assay since those primer sequences had already been reported and studied together.

The reported FSS sequences were modified slightly by removing the poly(T) tail and converting the degenerate bases into the most common sequence variant (identified by examination of MITOMAP information at the appropriate mtDNA position). A cleavable base was also incorporated at varying positions in different primers so that the cleaved primers would be resolvable on a mass scale. **Exhibit 61** lists the final primer set chosen for a

10-plex SNP reaction. Eight of the primers detect SNPs on the "heavy" GC-rich strand and two of them identify SNPs on the "light" AT-rich strand of the mtDNA control region (**Exhibit 60**). Five of the primers anneal within hypervariable region I (HV1) and five anneal within hypervariable region II (HV2). All of the 10 chosen SNP sites are transitions of either A to G (purine-to-purine) or C to T (pyrimidine-to-pyrimidine) rather than transversions (purine-to-pyrimidine).

Another important aspect of multiplex SNP primer design besides primer compatibility (i.e., lack of primer dimer formation or hairpins) is the avoidance of multiple-charged ions. Doubly-charged and triply-charged ions of larger mass primers can fall within the mass-to-charge range of smaller primers. Depending on the laser energy used and matrix crystallization the multiple-charged ions can be in fairly significant abundance (**Exhibit 63**). Primer impurities, such as n-1 failure products, can also impact how close together primers can be squeezed on a mass scale. These primer synthesis failure products will be ~300 Da smaller in mass than the full-length primer. Since an extension product ranges from 273 Da (ddC) up to 313 Da (ddG) larger than the primer itself, a minimum of 650-700 Da is needed between adjacent primers (post-cleavage mass) if primer synthesis failures exist to avoid any confusion in making the correct SNP genotype call.

We observed that primer synthesis failure products become more prevalent for larger mass primers. Because resolution and sensitivity in the mass spectrometer decrease at higher masses, it is advantageous to keep the multiplexed primers in a fairly narrow mass window and as small as possible. The primers we studied here ranged from 1580 Da up to 6500 Da.

Exhibit 62 displays the expected primer masses for the mtDNA SNP 10-plex along with their doubly- and triply-charged ions. The smallest four primers, in the mass range of 1580 Da to 3179 Da, have primer and extension masses that are similar to multiple-charged ions of larger primers. For example, in the **Exhibit 6** bottom panel, which shows the 10plex primers, the doubly-charged ion from MT4e (3250 Da) falls very close to the singly-charged ion from MT3' (3179 Da). The impact of primer impurity products can also be seen in **Exhibit 6**. An examination of the extension product region from primer MT7/H00073 (~6200 Da) shows two peaks where only one is expected (top panel). The lower mass peak in the doublet is labeled as "+ddC" (6192 Da) but the larger peak in the doublet is a primer impurity of MT4e/L00195 (6232 Da). The mass difference between these two peaks is 40 Da or exactly what one would expect for a C/G heterozygote extension of primer MT7/H00073. Thus, to avoid a false positive, it is important to run the 10 primers alone as a negative control to verify any primer impurities.

To aid development of this multiplex SNP assay, we produced a large quantity of PCR product from K562 genomic DNA (enough for ~320 reactions) and pooled it together so that multiple experiments would have the same starting material. With the K562 amplicon pool, we examined the impact of primer concentration variation without worrying about the DNA template as a variable. The K562 amplicon pool was generated using the PCR primers noted in **Exhibit 61**, which produce a 1021 bp PCR product that spans the entire D-loop region (Wilson et al. 1995). Thus, all 10 SNP sites could be examined from a single DNA template. Using standard ABI sequencing procedures and a dRhodamine dye-terminator sequencing kit (PE Applied Biosystems), we sequenced this PCR pool to verify the identity of the nucleotide

at each of the SNP sites in the 10-plex. The sequencing primers were the same as those reported previously (Wilson et al. 1995). We obtained identical results between the sequencing and the mass spec (*see Exhibit 6*), which helped verify our method.

A variety of primer combinations and primer concentrations were tested on the way to obtaining results with the 10-plex. For example, a 4-plex and a 6-plex were developed first with primers that were further apart in terms of mass and therefore could be more easily distinguished. An early 6-plex was published in *Electrophoresis* (Li et al. 1999). Primer concentrations were balanced empirically by first running all primers at 10 pmol and then raising or lowering the amount of primer in the next set to obtain a good balance between those in the multiplex primer mix. In general, a higher amount of primer was required for primers of higher mass. However, this trend did not always hold true probably because ionization efficiencies in MALDI mass spectrometry differ depending on DNA sequence content. The primer concentrations in the final "optimized" 10-plex ranged from 10 pmol for MT3' (3179 Da) to 35 pmol for MT7 (5891 Da). Primer extension efficiencies also vary between primers making optimization of these multiplexes rather challenging.

Due to the early termination of this project, we were unable to run this multiplex SNP assay across a panel of samples to verify that it worked with more than one sample. Future work could include examination of a set of population samples and correlation to DNA sequencing results. We originally proposed testing at least 100 samples. Examination of the impact of SNPs that are close to the one being tested and that might impact primer annealing also needs to be done. In addition, more SNP sites can be developed and the multiplex could be

expanded to include a larger number of loci. In short, we only began a very interesting project.

B. Y Chromosome Work

While the mitochondrial DNA work gave us a chance to examine the mass spec factors in developing an SNP multiplex, this work only involved a single DNA template with multiple SNP probes. A more common situation for multiplex SNP development is multiple DNA templates with one or more SNP per template. SNP sites may not be closely spaced along the genome and could thus require unique primer pairs to amplify each section of DNA. To test this multiplex SNP situation, we investigated multiple SNPs scattered across the Y chromosome. Through a collaboration with Stanford University scientists Dr. Peter Oefner and Dr. Peter Underhill, we obtained access to 20 Y chromosome SNP markers that we examined in this study. Dr. Oefner and Dr. Underhill have identified almost 150 SNP loci on the Y chromosome, some of which have been reported in the literature (Underhill et al. 1997). We decided to start examining an initial set of 20 Y SNPs and then add additional markers as needed with the eventual goal of ~50 Y SNP loci from which a final multiplex set would be developed. Our collaboration provided us with detailed sequence information around the SNP sites (typically several hundred bases on either side of the SNP site), which is important for multiplex PCR primer design. Dr. Oefner also provided a set of 38 male genomic DNA samples from various populations around the world for testing purposes.

The sequences were provided in two batches of 10 sequences each. We first attempted a 9-plex PCR with the initial set of 10 sequences and later when the second set of 10 arrived we

designed primers for a 17-plex PCR. Due to primer incompatibilities, we were unable to incorporate all possible SNPs into each multiplex set. However, with a larger set of sequences to choose from, it is conceivable that much larger PCR multiplexes could be developed. The first set of Y SNP markers included the following loci (according to Dr. Underhill's nomenclature): M9 (C→G), M17 (1 bp deletion, 4G's→3G's), M35 (G→C), M42 (A→T), M45 (G→A), M89 (C→T), M96 (G→C), M122 (T→C), M130 (C→T), and M145 (G→A). The second set of Y SNP markers contained these loci: M119 (A→C), M60 (1 bp insertion, a "T"), M55 (T→C), M20 (A→G), M69 (T→C), M67 (A→T), M3 (C→T), M13 (G→C), M2 (A→G), and M26 (G→A).

Multiplex PCR primers were designed with a UNIX version of Primer 3 (release 0.6; Rozen et al. 1998) adapted at GeneTrace by Nathan Hunt to utilize a mispriming library and Perl scripts for input and export of SNP sequences and primer information, respectively. The PCR primer sequences produced by Nathan's program are listed in **Exhibit 65**. The universal tags attached to each primer sequence aid in multiplex compatibility (Shuber et al. 1995). This tag added 23 bases to the 5'-end of the forward primers and 24 bases to the 5'-end of the reverse primers and therefore increased the overall length of PCR products by 47 bp. The addition of the universal tag makes multiplex PCR development much easier and reduces the need to empirically adjust primer concentrations to balance PCR product quantities obtained from multiple loci (Ross et al. 1998b).

In order to compare the amplicon yields from various loci amplified in the multiplex PCR, the product sizes were selected to make them resolvable by capillary electrophoretic separation.

Thus, the PCR product sizes range from 148 bp up to 333 bp (**Exhibit 66**) using the primers listed in **Exhibit 65**. To make sure that each primer pair worked, each marker was amplified individually as well as in the multiplex set using the same concentration of PCR primers. If a substantial amount of primers remained after PCR, then we had a pretty good idea that the PCR efficiency was lower for that particular marker (**Exhibit 71**). We were able to demonstrate male-specific PCR with a 17-plex set of PCR primers. The male test sample AM209 (from Amish CEPH family; *see Exhibit 39*) produced amplicons for 17 Y SNP loci while K562 genomic DNA yielded no detectable PCR product due to the fact that it is female DNA and therefore does not contain a Y chromosome (**Exhibit 70**).

SNP primers were designed and synthesized for probing the SNP sites either in a singleplex (**Exhibit 68**) or a multiplex (**Exhibit 67**) format. Unfortunately work on this project was halted before SNP sample testing could be performed in earnest. Some additional SNP primers and multiplex PCR primers were also designed for testing 12 autosomal SNPs throughout the human genome (**Exhibit 72**) with the hope of comparing the informativeness of SNPs to STRs. Analysis of these same 12 SNPs was recently demonstrated in a multiplex PCR and SNP assay by PerSeptive Biosystems (Ross et al. 1998b).

Optimal SNP markers for identity testing typically have allele frequencies of 30-70% in a particular human population. By way of comparison, highly polymorphic STRs can have 10-15 or more alleles with allele frequencies below 15% (i.e., more alleles and thus lower allele frequencies). Unless the cost becomes sufficiently low for obtaining genotypes that the large DNA databases could be regenerated (all samples would have to be rerun because the DNA

markers are different). The characteristics of STR and SNP markers are compared in **Exhibit 59**. SNPs have the capability of being multiplexed to a much higher level than STRs but more SNP markers are required for the same level of discrimination compared to STRs. Only time will tell what role new SNP markers will have in human identity testing.

VI. CONCLUSIONS AND IMPLICATIONS OF FINDINGS

Time-of-flight mass spectrometry offers a rapid, cost-effective alternative for genotyping large numbers of samples. Each DNA sample can be accurately measured in a few seconds. Due to the increased accuracy with mass spectrometry, short tandem repeat (STR) alleles may be reliably typed without comparison to allelic ladders. Mass spectrometry holds significant promise as a technology for high-throughput DNA processing that will be valuable for large-scale DNA database work.

In summary the positive features of mass spectrometry for STR analysis include:

- **Rapid** results – STR typing at a rate of seconds per sample
- **Accurate** without allelic ladders
- **Direct** DNA measurement – no fluorescent or radioactive labels
- **Automated** sample preparation and data collection
- **High-throughput** capabilities of thousands of samples daily per system
- **Flexible** – single nucleotide polymorphism (SNP) assays can be run on the same instrument platform

Over the course of this project, we demonstrated that both STR and SNP analysis are reliably performed with our mass spectrometric technology. We tested a large number of human DNA markers that are of forensic interest. We have developed new primer sets for the 13 CODIS STR loci that may prove useful in the future for situations where degraded DNA is present and thus requires smaller amplicons to obtain successful results. We also explored the possibility of developing multiplexed SNP markers, and constructed a mtDNA 10-plex assay

and a Y chromosome 17-plex PCR. Both STR and SNP areas appear promising for future research. In another project at GeneTrace, we recently demonstrated a sample throughput of approximately 4,000 STR samples in a single day with a single automated mass spectrometer. Clearly, we have improved the analysis of DNA short tandem repeat markers using time-of-flight mass spectrometry.

Commercializing this new technology

For several months in late 1998, we actively pursued the possibility of a business partnership to introduce our technology to the forensic DNA typing market. We contacted the leading companies involved with contract DNA typing work including the Bode Technology Group, Fairfax Identity Laboratory, and the Laboratory Corporation of America. In 1998, Dr. Butler gave over a dozen presentations at scientific conferences and to individual laboratories working in the DNA typing arena. Unfortunately, we found that interest in new technology appeared to be limited by the fact that contract laboratories are being required to use established commercially available fluorescent STR typing kits and conventional technology for CODIS DNA database contracts as dictated by the contracting agencies. We concluded that the desire for stability and continuity throughout the DNA typing community would make it difficult for a small company such as GeneTrace, with limited resources, to enter this market against established technology backed by a large corporation. We intensively studied the situation from a business perspective and determined that GeneTrace should not attempt to introduce our technology to the forensic DNA community at this time even though it is apparent that the technology has true potential.

VII. ENDNOTES

A. Works Cited

ABI Prism 310 Genetic Analyzer User's Manual. PE Applied Biosystems, 1998.

AmpF*ℓ*STR® Profiler Plus™ PCR Amplification Kit User's Manual. PE Applied Biosystems, 1998.

AmpF*ℓ*STR® COfiler™ PCR Amplification Kit User Bulletin. PE Applied Biosystems, 1998.

Becker, C.H., Li, J., Shaler, T.A., Hunter, J.M., Lin, H., Monforte, J.A. (1997) Genetic analysis of short tandem repeat loci by time-of-flight mass spectrometry. *Proceedings from the Seventh International Symposium on Human Identification 1996*, Promega Corporation, pp. 158-162.

Becker, C.H. and Young, S.E. (1999) Mass spectrometer. U.S. Patent 5,864,137

Braun, A., Little, D.P., Koster, H. (1997a) Detecting CFTR gene mutations by using primer oligo base extension and mass spectrometry. *Clin Chem* 43:1151-1158.

Braun, A., Little, D.P., Reuter, D., Muller-Mysok, B., Koster, H. (1997b) Improved analysis of microsatellites using mass spectrometry. *Genomics* 46:18-23.

Butler, J.M., McCord, B.R., Jung, J.M., Lee, J.A., Budowle, B., Allen, R.O. (1995) Application of dual internal standards for precise sizing of polymerase chain reaction products using capillary electrophoresis. *Electrophoresis* 16: 974-980.

Butler, J.M., Ruitberg, C.M., Reeder, D.J. (1998a) STRBase: a short tandem repeat DNA internet-accessible database. *Proceedings of the Eighth International Symposium on Human Identification 1997*, Promega Corporation, pp. 38-47.

Butler, J.M., Li, J., Shaler, T.A., Monforte, J.A., Becker, C.H. (1998b) Reliable genotyping of short tandem repeat loci without an allelic ladder using time-of-flight mass spectrometry. *Int. J. Legal Med.* 112: 45-49.

Carroll, J.A. and Beavis, R.C. (1996) Using matrix convolution filters to extract information from time-of-flight mass spectra. *Rapid Commun. Mass Spectrom.* 10: 1683-1687.

Clark, J.M. (1988) Novel non-templated nucleotide addition reactions catalysed by procaryotic and eucaryotic DNA polymerases. *Nucleic Acids Res.* 16: 9677-9686.

Collins, F.S., Brooks, L.D., Chakravarti, A. (1998) A DNA polymorphism discovery resource for research on human genetic variation. *Genome Res.* 8: 1229-1231.

GenePrint™ STR Systems Technical Manual (1995) Part# TMD004, Promega Corporation.

Haff, L.A. and Smirnov, I.P. (1997a) Single-nucleotide polymorphism identification assays using a thermostable DNA polymerase and delayed extraction MALDI-TOF mass spectrometry. *Genome Res. 7*: 378-388.

Haff, L.A. and Smirnov, I.P. (1997b) Multiplex genotyping of PCR products with MassTag-labeled primers. *Nucleic Acids Res. 25*: 3749-3750.

Hammond, H.A., Jin, L., Zhong, Y., Caskey, C.T. and Chakraborty, R. (1994) Evaluation of 13 short tandem repeat loci for use in personal identification applications. *Am. J. Hum. Genet.* 55: 175-189.

Kimpton, C.P., Gill, P., Walton, A., Urquhart, A., Millican, E.S., Adams, M. (1993) Automated DNA profiling employing multiplex amplification of short tandem repeat loci. *PCR Meth. Appl.* 3: 13-22.

Lazaruk, K, Walsh, PS, Oaks, F, Gilbert, D, Rosenblum, BB, Menchen, S, Scheibler, D, Wenz, HM, Holt, C, Wallin, J. (1998) Genotyping of forensic short tandem repeat (STR) systems based on sizing precision in a capillary electrophoresis instrument. *Electrophoresis* 19: 86-93.

Li, H., Schmidt, L., Wei, M.-H., Hustad, T., Lerman, M.I., Zbar, B. and Tory, K. (1993) Three tetranucleotide polymorphisms for loci: D3S1352, D3S1358, D3S1359. *Hum. Mol. Genet.* 2: 1327.

Li, J., Butler, J.M., Tan, Y., Lin, H., Royer, S., Ohler, L., Shaler, T.A., Hunter, J.M., Pollart, D.J., Monforte, J.A., Becker, C.H. (1999) Single nucleotide polymorphism determination using primer extension and time-of-flight mass spectrometry. *Electrophoresis* 20: 1258-1265.

Lindqvist, A.-K.B., Magnusson, P.K.E., Balciuniene, J., Wadelius, C., Lindholm, E., Alarcon-Riquelme, M.E. and Gyllensten, U.B. (1996) Chromosome-specific panels of tri- and tetranucleotide microsatellite markers for multiplex fluorescent detection and automated genotyping: evaluation of their utility in pathology and forensics. *Genome Res.* 6: 1170-1176.

MITOMAP: A Human Mitochondrial Genome Database. Center for Molecular Medicine, Emory University, Atlanta, GA, USA, http://www.gen.emory.edu/mitomap.html, 1999.

Monforte, J.A., Shaler, T.A., Tan, Y., Becker, C.H. (1995) Methods of preparing nucleic acids for mass spectrometric analysis. Provisional U.S. patent submitted December 2, 1995.

Monforte, J.A., Becker, C.H. (1997a) High-throughput DNA analysis by time-of-flight mass spectrometry. *Nature Med. 3*: 360-362.

Monforte, J.A., Becker, C.H., Shaler, T.A., Pollart, D.J. (1997b) Oligonucleotide sizing using immobilized cleavable primers. US Patent 5,700,642

Ross, P.L., Belgrader, P. (1997) Analysis of short tandem repeat polymorphisms in human DNA by matrix-assisted laser desorption/ionization mass spectrometry. *Anal Chem* 69: 3966-3972.

Ross, P.L., Davis, P.A., Belgrader, P. (1998a) Analysis of DNA fragments from conventional and microfabricated PCR devices using delayed extraction MALDI-TOF mass spectrometry. *Anal Chem* 70: 2067-2073.

Ross, P., Hall, L., Smirnov, I., Haff, L. (1998b) High level multiplex genotyping of MALDI-TOF mass spectrometry. Nature Biotech. 16: 1347-1351.

Shaler, T.A., Wickham, J.N., Sannes, K.A., Wu, K.J., Becker, C.H. (1996) Effect of impurities on the matrix-assisted laser desorption mass spectra of single-stranded oligonucleotides. *Anal Chem* 68: 576-579.

Rozen, S. and Skaletsky, H.J. (1998) Primer3 (primer3_www.cgi v 0.2); Code available at http://www-genome.wi.mit.edu/genome_software/other/primer3.html.

Schumm, J.W., Lins, A.W., Micka, K.A., Sprecher, C.J., Rabbach, D.R., Bacher, J.W. (1997) Automated fluorescent detection of STR multiplexes—development of the GenePrint™ PowerPlex™ and FFFL multiplexes for forensic and paternity applications. *Proceedings from the Seventh International Symposium on Human Identification 1996*, Promega Corporation, pp. 70-88.

Shuber, A.P., Grondin, V.J., Klinger, K.W. (1995) A simplified procedure for developing multiplex PCRs. *Genome Res.* 5: 488-493.

Sullivan, K.M., Mannucci, A., Kimpton, C.P. and Gill, P. (1993) A rapid and quantitative DNA sex test: fluorescence-based PCR analysis of X-Y homologous gene amelogenin. *BioTechniques* 15: 637-641.

Taranenko, N.I., Golovlev, V.V., Allman, S.L., Taranenko, N.V., Chen, C.H., Hong, J., and Chang, L.Y. (1998) Matrix-assisted laser desorption/ionization for short tandem repeat loci. *Rapid Commun.Mass Spectrom.* 12: 413-418.

Tully, G., Sullivan, K.M., Nixon, P., Stones, R.E., Gill, P. (1996) Rapid detection of mitochondrial sequence polymorphisms using multiplex solid-phase fluorescent minisequencing. *Genomics* 34: 107-113.

Underhill, P.A., Jin, L., Lin, A.A., Mehdi, S.Q., Jenkins, T., Vollrath, D., Davis, R.W., Cavalli-Sforza, L.L., Oefner, P.J. (1997) Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography. *Genome Res.* 7: 996-1005.

Vestal, M.L., Juhasz, P., Martin, S.A. (1995) Delayed extraction matrix-assisted laser desorption time-of-flight mass spectrometry. *Rapid Commun. Mass Spectrom.* 9: 1044-1050.

Walsh, P.S., Fildes, N.J. and Reynolds, R. (1996) Sequence analysis and characterization of stutter products at the tetranucleotide repeat locus vWA. *Nucleic Acids Res.* 24: 2807-2812.

Wilson, M.R., Polanskey, D., Butler, J.M., DiZinno, J.A., Replogle, J., Budowle, B. (1995) Extraction, PCR amplification, and sequencing of mitochondrial DNA from human hair shafts. *BioTechniques* 18: 662-669.

Wu, K.J., Steding, A., Becker, C.H. (1993) Matrix-assisted laser desorption time-of-flight mass spectrometry of oligonucleotides using 3-hydroxypicolinic acid as an ultraviolet-sensitive matrix. *Rapid Commun Mass Spectrom* 7: 142-146.

Wu, K.J., Shaler, T.A., Becker, C.H. (1994) Time-of-flight mass spectrometry of underivatized single-stranded DNA oligomers by matrix-assisted laser desorption. *Anal Chem* 66: 1637-1645.

B. Biographical Note

John M. Butler

Dr. Butler did his graduate work in the FBI Laboratory's Forensic Science Research Unit under the direction of Dr. Bruce McCord and Dr. Bruce Budowle and received his Ph.D. from the University of Virginia in 1995. As a pioneer in the use of capillary electrophoresis for DNA analysis, he was the first to demonstrate that STR typing could be performed with CE. Dr. Butler also developed a method for quantitating mitochondrial DNA that is used today in forensic casework at the FBI Laboratory. Dr. Butler went on to do postdoctoral research at the National Institute of Standards and Technology with Dr. Dennis Reeder. While at NIST, he designed and built STRBase—a comprehensive, internet-accessible database on STRs used in human identity testing. He also worked on a project, consisting of scientists from the Armed Forces DNA Identification Laboratory and American University, to develop peptide nucleic acid probes for rapid DNA typing using time-of-flight mass spectrometry. In June of 1997, Dr. Butler joined GeneTrace Systems as a staff scientist, where he is doing STR and single nucleotide polymorphism (SNP) analysis by mass spectrometry, most recently as a project leader leading SNP genotyping efforts at GeneTrace.

Christopher H. Becker

Dr. Becker is the President and Chief Operation Officer of GeneTrace Systems Inc., which he co-founded with Joe Monforte in 1994. In addition to providing leadership for the Company and overseeing operations, Dr. Becker guides the instrumentation and software development. Before founding GeneTrace, Dr. Becker was a Group Leader (Surface Chemistry) at SRI International. He had been with SRI since 1980. He is a physical chemist by training (Ph.D. from the University of California at Berkeley, 1979). Dr. Becker's research group was the first in the world to analyze DNA oligomers by laser-based, time-of-flight mass spectrometry. He has also previously developed a new surface analytical method using laser ionization, and has designed high-performance commercial time-of-flight mass spectrometers. Dr. Becker is author of over 100 scientific and technical articles and holds 5 patents.

VIII. EXHIBITS

Lists of Exhibits included with this report:

Exhibit 1. Schematic of GeneTrace automated time-of-flight mass spectrometer
Exhibit 2. Table of PCR product sizes with newly designed primers
Exhibit 3. Fluorescent multiplex STR result with COfiler™ kit and ABI 310 CE
Exhibit 4. Schematic of expected allele masses for a CTT multiplex
Exhibit 5. Mass spectrum of a CTT multiplex sample
Exhibit 6. Mass spectrum of SNP 10plex from mtDNA control region
Exhibit 7. Schematic of the STR assay using GeneTrace cleavable primer technology
Exhibit 8. Schematic of the SNP assay using GeneTrace cleavable primer technology
Exhibit 9. Photo showing the automated GeneTrace mass spec ion optics
Exhibit 10. Tetranucleotide markers identified during public database searches
Exhibit 11. STR markers examined at GeneTrace over the course of this project
Exhibit 12. Primer sequences designed and used for STR markers
Exhibit 13. D3S1358 sequence with PCR primer locations
Exhibit 14. Mass ladder file for STR loci analyzed in this study
Exhibit 15. Mass spec detection of TH01 allelic ladder
Exhibit 16. STR results from K562 DNA for CSF1PO, TPOX, TH01, and AMEL
Exhibit 17. STR allelic ladders: CSF1PO, TPOX, TH01, and VWA
Exhibit 18. STR allelic ladders: D5S818 from ABI and Promega ladder reamplification
Exhibit 19. D8S1179 result with ddG approach
Exhibit 20. Multiplex STR with TH01 and TPOX non-overlapping alleles
Exhibit 21. TH01 and TPOX allelic ladder multiplex
Exhibit 22. Multiplex STR analysis with 85 Da difference between TH01 and TPOX
Exhibit 23. Multiplex strategies for STR genotyping
Exhibit 24. Table with 15 replicate analyses of TH01 allelic ladder
Exhibit 25. TPOX sensitivity study: 20 ng, 5 ng, 0.5 ng, 0 ng K562 DNA template
Exhibit 26. CAL DOJ STR data collection times comparing mass spec to ABI 310 CE
Exhibit 27. Comparison of allele masses over time with TH01 and TPOX data
Exhibit 28. Plot of Amelogenin X allele mass offset vs. Y allele mass offset
Exhibit 29. Heterozygotes from 4 different loci and mass offset1 vs. mass offset2
Exhibit 30. Plot of measured masses for 4 different STRs
Exhibit 31. *Tsp* vs *TaqGold* polymerase data on TH01 STR sample
Exhibit 32. Table with top and bottom strand mass differences for TH01 allelic ladder
Exhibit 33. Stutter products produced from a dinucleotide repeat marker
Exhibit 34. Sample sets run on ABI 310 with ABI STR kits
Exhibit 35. Genotypes for K562, AM209, and UP006 standard samples
Exhibit 36. ABI 310 result for a D3S1358 15, 15.2 heterozygous sample
Exhibit 37. ABI 310 results for new FGA, D21S11, and D18S51 alleles (Stanford)
Exhibit 38. Table with Stanford male sample results for 9 STRs
Exhibit 39. CEPH family pedigrees for samples examined
Exhibit 40. Comparison table of CSF1PO results with ABI 310 and mass spec
Exhibit 41. Comparison table of TPOX results with ABI 310 and mass spec
Exhibit 42. Comparison table of TH01 results with ABI 310 and mass spec

Exhibit 43. Comparison table of AMEL results with ABI 310 and mass spec
Exhibit 44. Comparison table of D3S1358 results with ABI 310 and mass spec
Exhibit 45. Comparison table of D16S539 results with ABI 310 and mass spec
Exhibit 46. Comparison table of D7S820 results with ABI 310 and mass spec
Exhibit 47. Comparison table of FGA results with ABI 310 and mass spec
Exhibit 48. CAL DOJ results for D8S1179 with mass spec
Exhibit 49. CAL DOJ results for DYS391 with mass spec
Exhibit 50. Mass spec results for TPOX samples from CAL DOJ study
Exhibit 51. Mass spec results for CSF1PO samples from CAL DOJ study
Exhibit 52. Mass spec results for TH01 samples from CAL DOJ study
Exhibit 53. Mass spec results for D7S820 samples from CAL DOJ study
Exhibit 54. Primer digestion with exonuclease to determine unknown sequences
Exhibit 55. Mass spectrum of Green I primer mix
Exhibit 56. TH01 STR primer positions for commercial kits
Exhibit 57. TPOX STR primer positions for commercial kits
Exhibit 58. CSF1PO STR primer positions for commercial kits
Exhibit 59. Types of genetic markers—characteristics of STRs vs. SNPs
Exhibit 60. Schematic of mtDNA control region 10plex SNP assay
Exhibit 61. Mitochondrial DNA primers used for 10plex SNP reaction
Exhibit 62. Expected masse-to-charge ratios of various ions for the mtDNA 10plex assay
Exhibit 63. SNP ions impacting multiplex design
Exhibit 64. SNP precision studies
Exhibit 65. Multiplex PCR primers used for Y SNP markers.
Exhibit 66. Multiplex PCR information for 17plex PCR reaction
Exhibit 67. Y SNP multiplex primer information (9plex with non-overlapping masses)
Exhibit 68. Y SNP singleplex primer information
Exhibit 69. Drop-out experiments from 9plex to balanced 5plex PCR
Exhibit 70. Male specific amplification of Y SNP markers involving 17plex PCR
Exhibit 71. Multiplex PCR (17plex) compared to individual PCR reactions
Exhibit 72. Human autosomal SNP markers (multiplex PCR and singleplex SNP)
Exhibit 73. CEPH/Diversity results compared between methods
Exhibit 74. Effects of sequence variation around STR repeat region on PCR amplification

ADDITIONAL EXHIBITS

Reprints from 2 peer-reviewed papers published during this project:

Butler, J.M., Li, J., Shaler, T.A., Monforte, J.A., Becker, C.H. (1998) Reliable genotyping of short tandem repeat loci without an allelic ladder using time-of-flight mass spectrometry. *Int. J. Legal Med.* 112: 45-49.

Li, J., Butler, J.M., Tan, Y., Lin, H., Royer, S., Ohler, L., Shaler, T.A., Hunter, J.M., Pollart, D.J., Monforte, J.A., Becker, C.H. (1999) Single nucleotide polymorphism determination using primer extension and time-of-flight mass spectrometry. *Electrophoresis* 20:1258-1265.

Exhibit 1. **Schematic of GeneTrace automated time-of-flight mass spectrometer**. DNA molecules are liberated from a solid-phase matrix environment with a laser pulse. The DNA reaction products are separated by size (mass) in a matter of microseconds as opposed to hours by conventional methods used in molecular biology labs today. For each run, hundreds of samples are prepared in parallel using a robotic workstation and spotted on a sample plate that is introduced to the vacuum environment of the mass spectrometer. The sample plate moves under the fixed laser beam to allow sequential sample analysis.

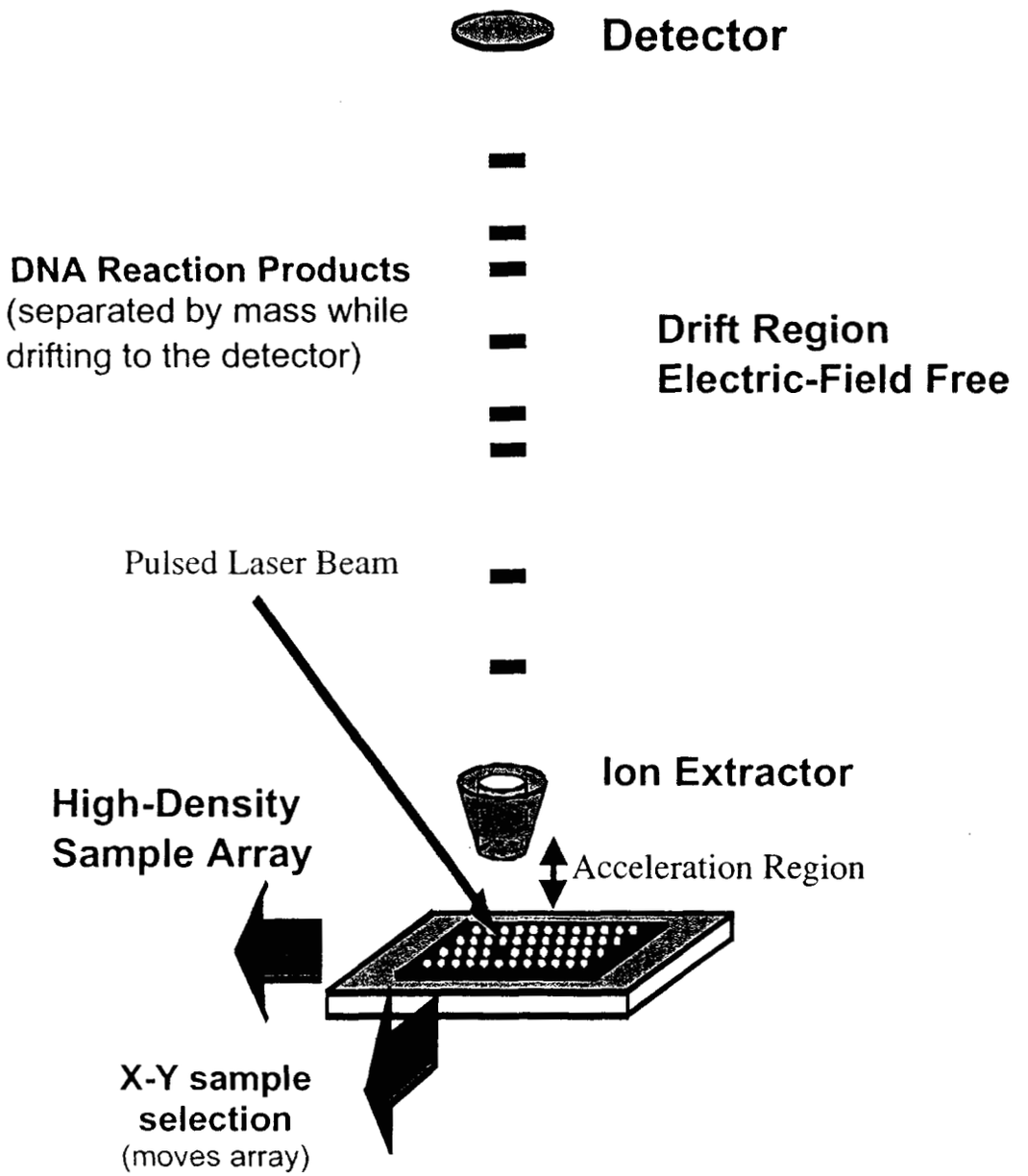


Exhibit 2. **Table of PCR product sizes with newly designed primers** for commonly used STR loci compared with commercially available primers used in multiplex sets for fluorescence-based assays. Some primer sets are not available from one or both of the manufacturers (n.a.: not available).

STR Locus	Known Alleles	**GeneTrace Sizes**	Commerically Available Sizes*	
		Newly designed primers with this study	PE Applied Biosystems	Promega
Amelogenin	X, Y	106, 112 bp	106,112 bp	212,218 bp
CD4	4—15	81-136 bp	Not available	
CSF1PO	6—15	87-123 bp	280-316 bp	291-327 bp
F13A1	3—17	112-168 bp	n.a.	279-335 bp
F13B	6—12	110-134 bp	n.a.	169-193 bp
FES/FPS	7—15	76-108 bp	n.a.	222-254 bp
FGA	15—30	118-180 bp	206-266 bp	n.a.
D3S1358	9—20	76-120 bp	101-145 bp	n.a.
D5S818	7—15	89-121 bp	134-166 bp	119-151 bp
D7S820	6—14	66-98 bp	257-289 bp	215-247 bp
D8S1179	8—18	92-130 bp	127-167 bp	n.a.
D13S317	7—15	98-130 bp	201-233 bp	165-197 bp
D16S539	5,8—15	81-121 bp	233-273 bp	264-304 bp
D18S51	9-27	120-192 bp	272-344 bp	n.a.
D21S11	24-38	150-190 bp	186-242 bp	n.a.
DYS19	8—16	76-108 bp	Not available	
DYS391	9—12	99-111 bp	Not available	
HPRTB	6—17	84-128 bp	n.a.	259-303 bp
LPL	7—14	105-133 bp	n.a.	105-133 bp
TH01	3—13.3	55-98 bp	160-203 bp	171-214 bp
TPOX	6—14	69-101 bp	217-249 bp	224-256 bp
VWA	11—22	126-170 bp	156-200 bp	127-171 bp
other STRs				
GATA132B04	10—14	99-115 bp	Not available	
D22S445	10—16	110-130 bp	Not available	
D16S2622	4—8	71-87 bp	Not available	

*sizes are listed without adenylation (add 1 base for +A form)

Exhibit 3. **Fluorescent multiplex STR result with PE Applied Biosystems AmpF*ℓ*STR® COfiler™ kit and ABI 310 Genetic Analyzer.** The red peaks are an internal sizing standard (GS350-ROX). The DNA size window shown here is from 90-350 bp. The allele calls for the 7 loci in this multiplex are amelogenin: X,Y; D3S1358: 15, 17; TH01: 6, 7; TPOX: 8, 12; D16S539: 8, 12; D7S820: 9, 11; and CSF1PO: 12, 12.

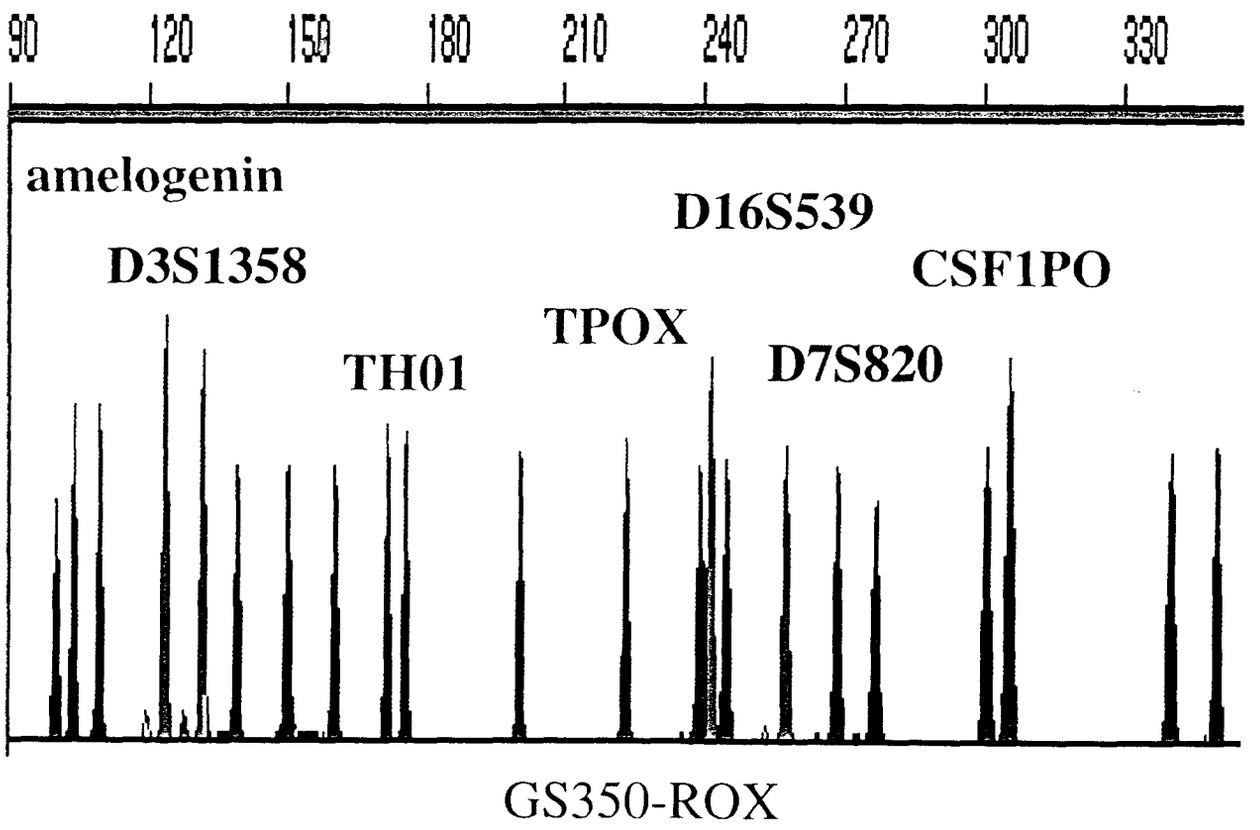


Exhibit 4. **Schematic view of expected allele masses for a CSF1PO-TPOX-TH01 (CTT) multiplex** involving overlapping allele size ranges. All known alleles are fully distinguishable by mass with this interleaving approach.

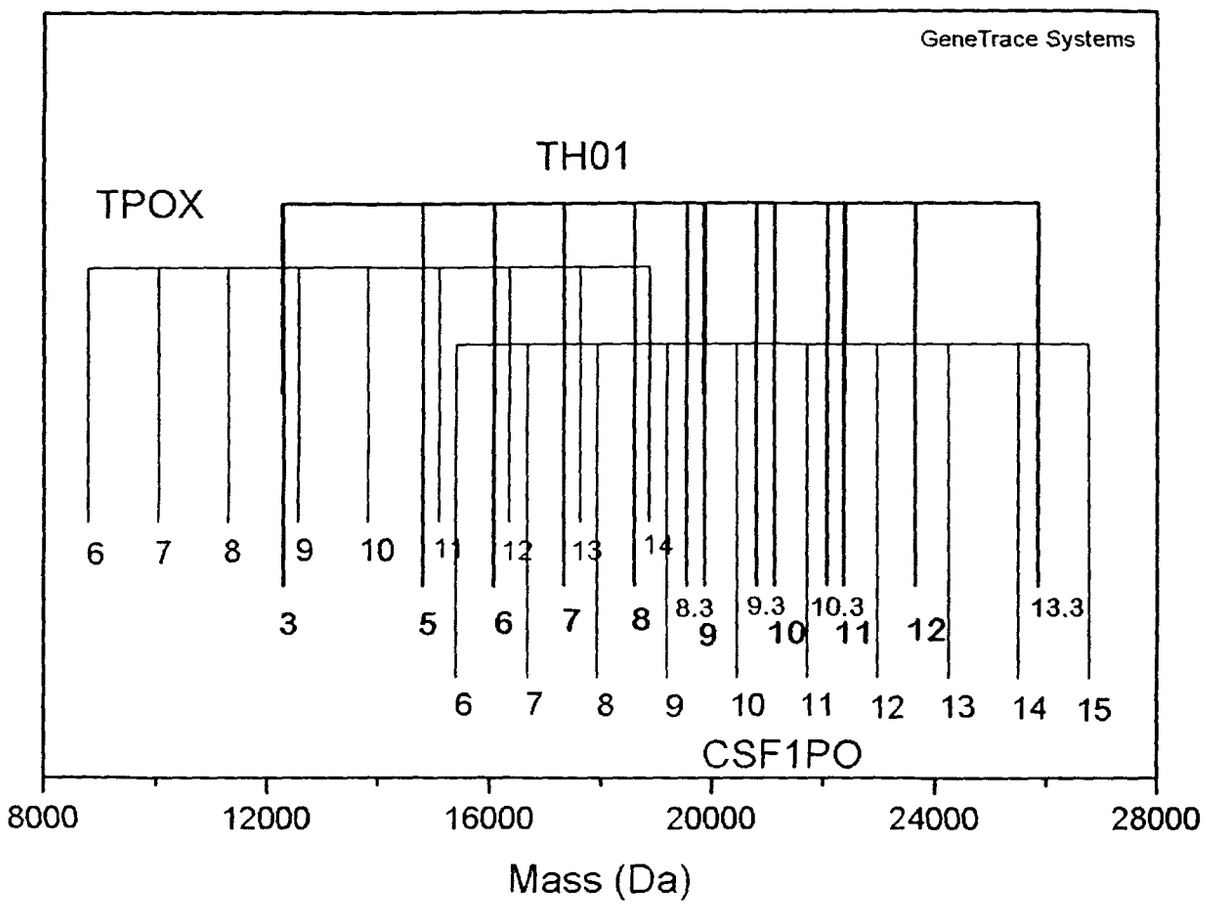


Exhibit 5. **Mass spectrum of an STR triplex involving TPOX, TH01, and CSF1PO**. This sample is a mass spec result using the interleaving allele approach schematically illustrated in Exhibit 4. Multiplex PCR as well as multiplex primer extension with ddC termination was used to obtain this result.

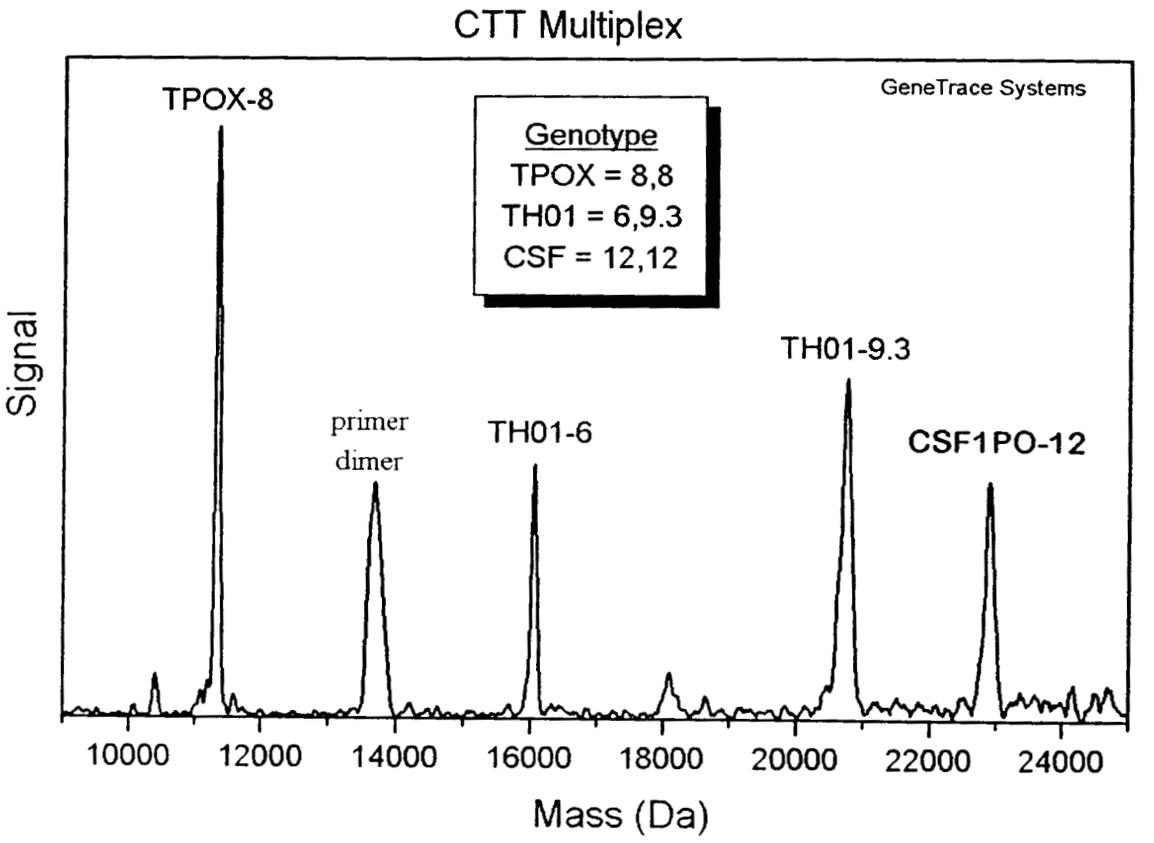


Exhibit 6. **Mass spectrum of SNP 10plex assay for screening polymorphic sites in the mtDNA control region.** The bottom panel shows the 10 SNP primers prior to the primer extension reaction. The top panel contains the multiplexed reaction products each labeled with the observed extension product. The results for this K562 PCR product are (in order across HV1 and HV2): H16069 (G), H16129 (C), H16189 (A), H16224 (G), H16311 (A), H00073 (C), L00146 (T), H00152 (A), L00195 (T), and H00247 (C). SNP nucleotide results have been confirmed by sequencing. The primer sequences are listed in Exhibit 61. The primer concentrations were 25 pmol MT5, 15 pmol MT8', 15 pmol MT10, 10 pmol MT3', 20 pmol MT9, 25 pmol MT6, 20 pmol MT2g, 27 pmol MT1, 35 pmol MT7, and 20 pmol MT4e.

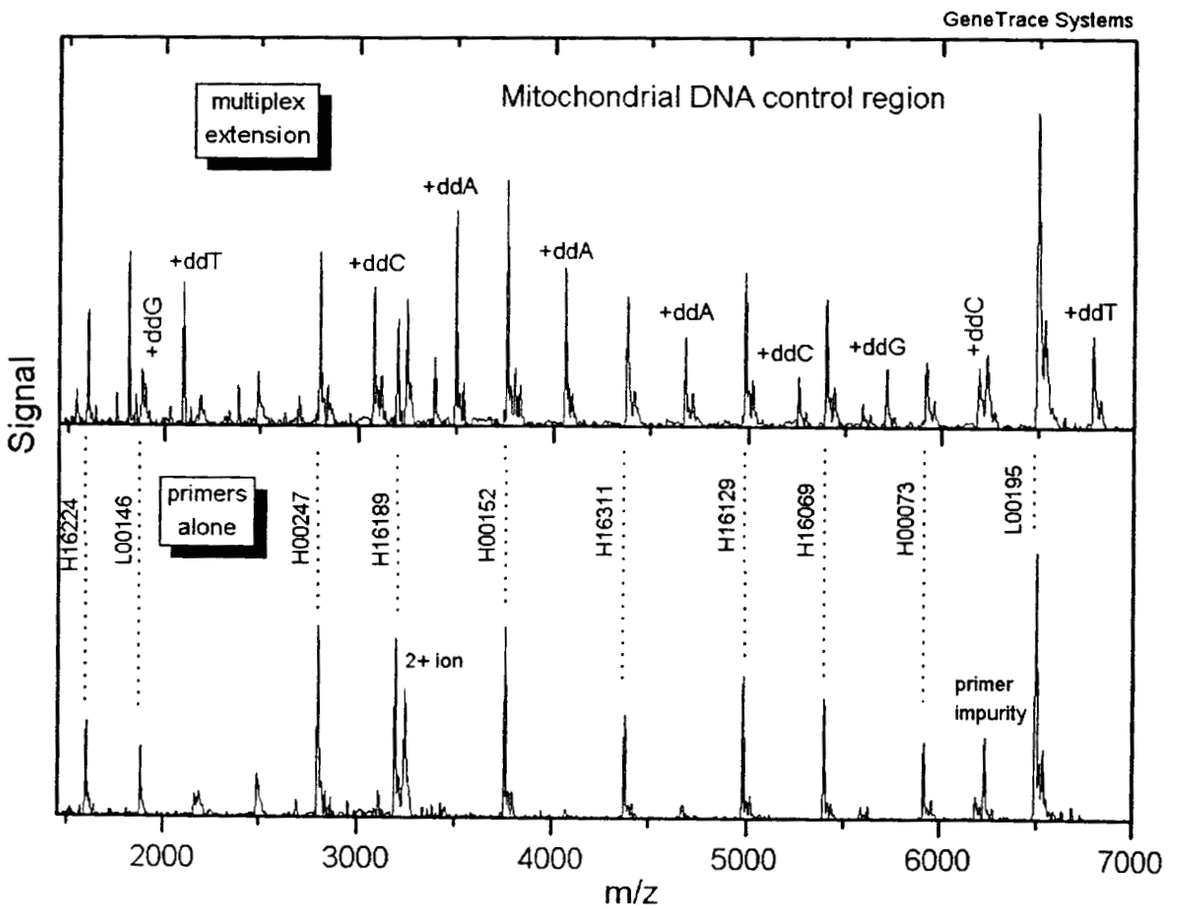


Exhibit 7. **Schematic of the STR assay involving the GeneTrace cleavable primer.**

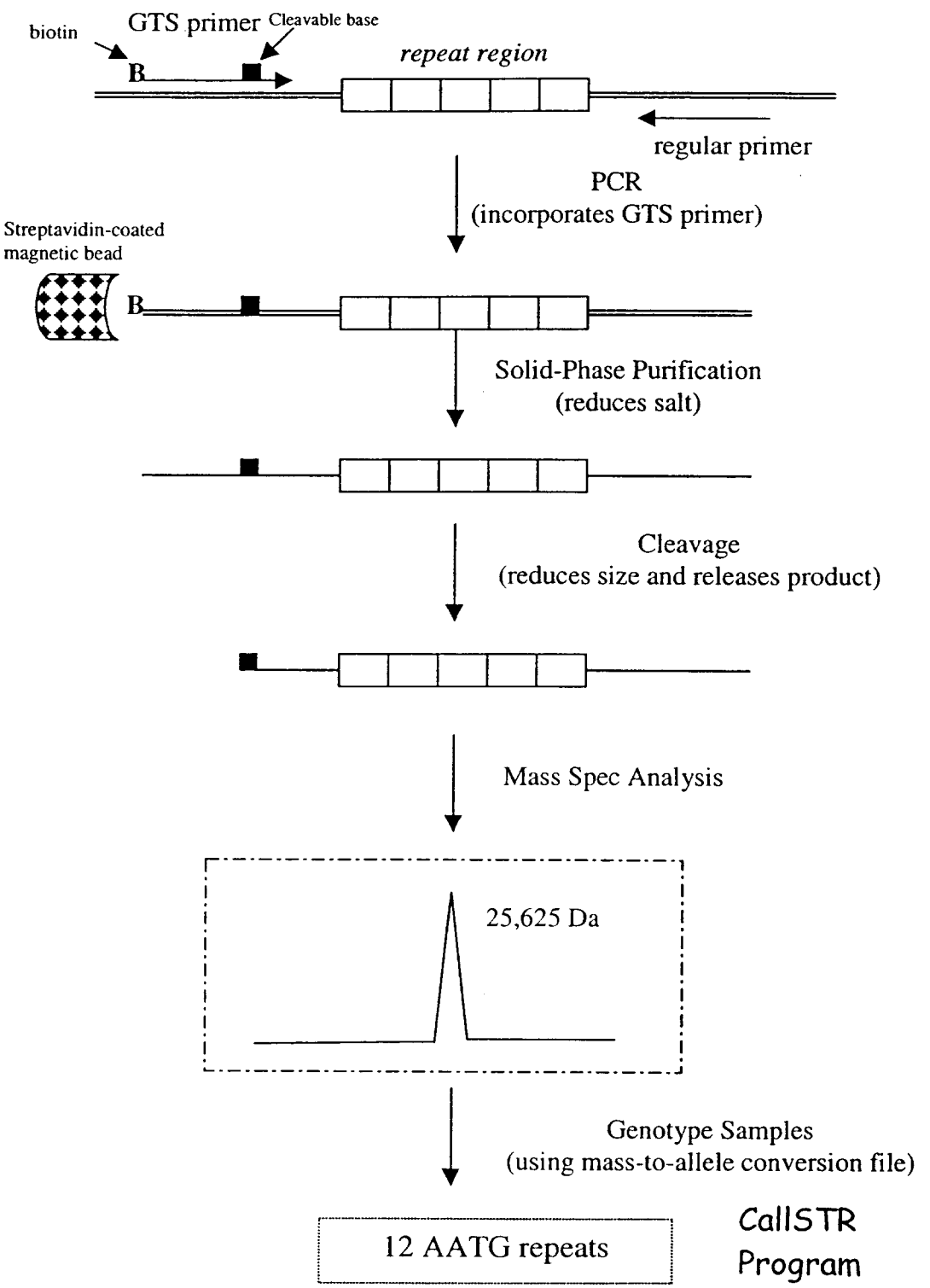


Exhibit 8. **Schematic of the SNP assay using GeneTrace cleavable primer technology.**

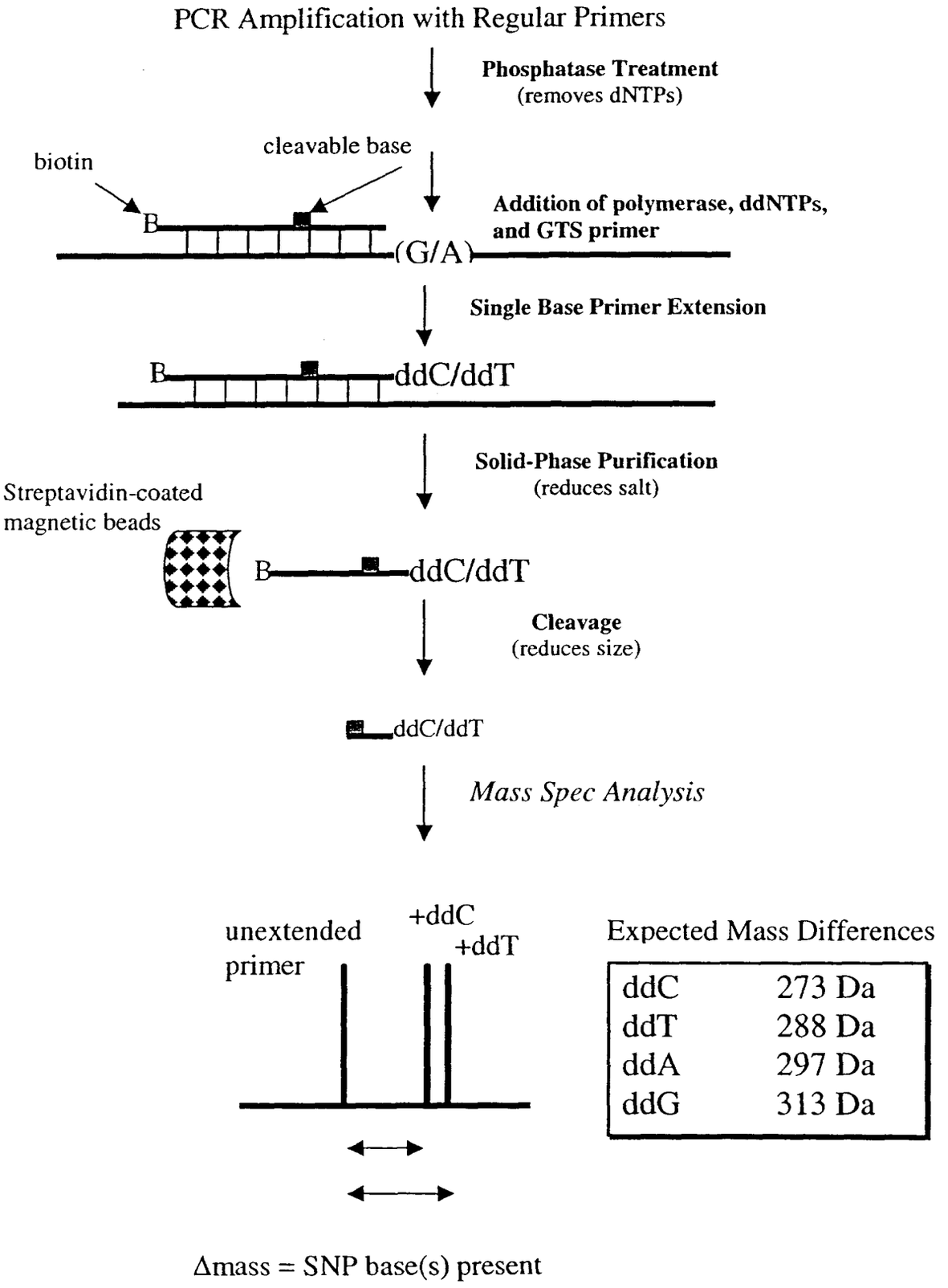


Exhibit 9. **Photo showing the automated GeneTrace mass spec ion optics** over a sample plate containing 384 different DNA samples. A sample in the center of the plate may be seen illuminated by pulsing laser light.

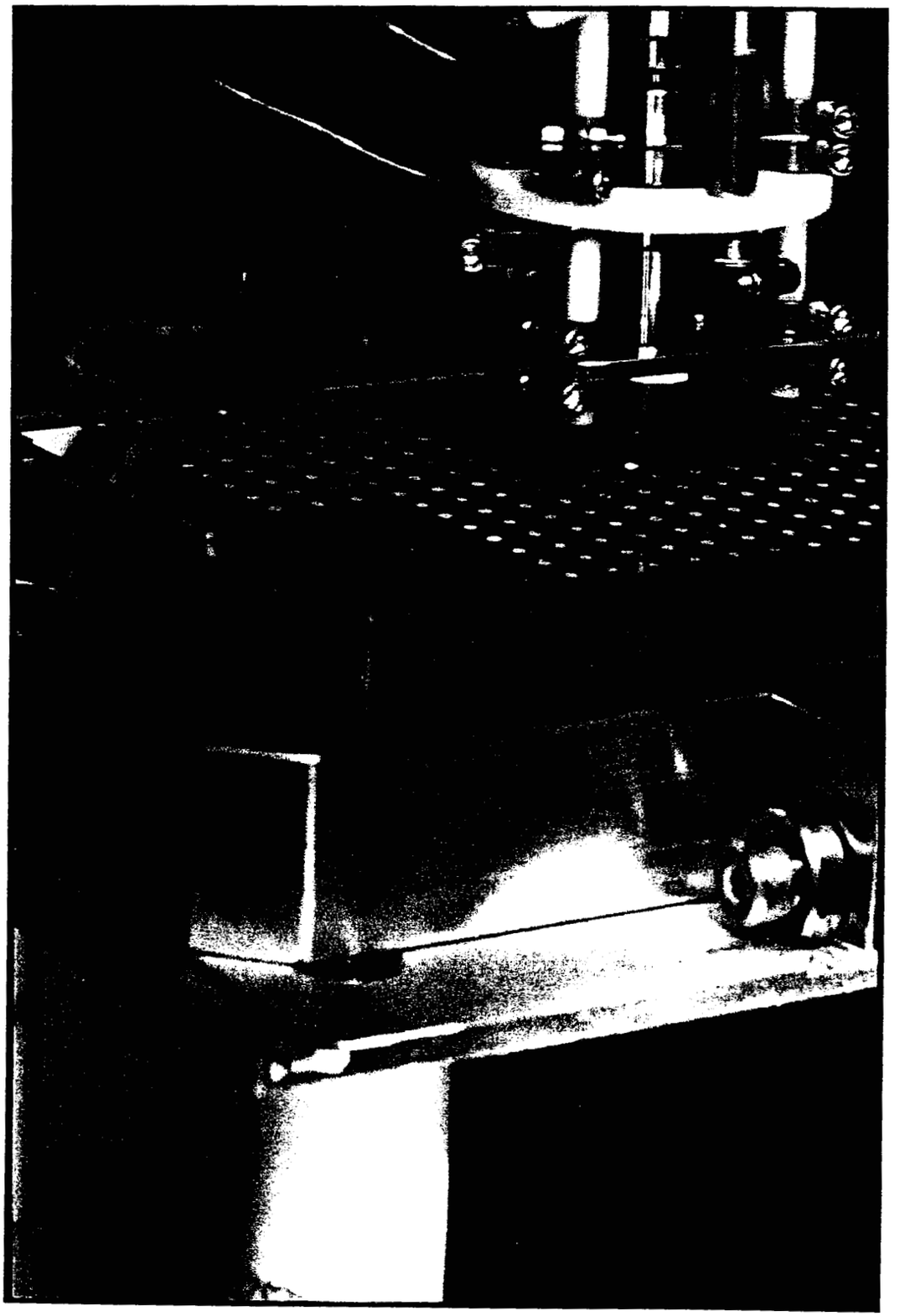


Exhibit 10. **Tetranucleotide markers identified during literature and public database searches** to be possible candidates for early STR marker development. The databases searched included the Cooperative Human Linkage Center (CHLC), Marshfield Clinic Weber set 8, and the Genome Database (GDB). Primers were synthesized and tested for the markers in **bold**.

Marker Name	Heterozygosity	Size Range	Allele Frequencies	Number of Alleles	GenBank Sequence
D1S1612	0.83	94-134 bp	0.9-26%	10 alleles	G07863
D2S1391	0.79	109-137 bp	5-35%	8 alleles	G08168
D5S1457	0.74	97-127 bp	1-32%	8 alleles	G08431
GATA132B04 (D5S2843)	not reported	98-114 bp	12-35%	5 alleles	G10407
D16S2622	not reported	71-91 bp	3-54%	5 alleles	G07934
D16S764	0.70	96-116 bp	4-38%	5 alleles	G07928
D19S591	0.74	96-112 bp	6-36%	5 alleles	G09745
D22S445	0.65	110-130 bp	3-37%	6 alleles	G08096

Exhibit 11. **STR markers examined at GeneTrace over the course of this project as sorted by their chromosomal position.** Primers were designed, synthesized, and tested for each of these markers. The most extensive testing was performed with the markers highlighted in **green**. Amelogenin, which is a gender identification marker rather than an STR, is listed twice because it occurs on both the X and Y chromosomes. The *italicized* STRs are ones not commonly used by the forensic DNA community.

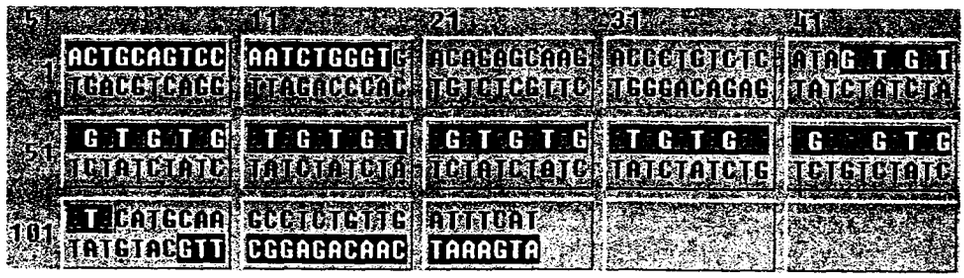
Human Chromosome	STR Marker	Human Chromosome	STR Marker
1	F13B	13	D13S317
2	**TPOX**	14	
3	**D3S1358**	15	FES/FPS
4	**FGA**	16	**D16S539**, *D16S2622*
5	**CSF1PO**, D5S818, *GATA132B04*	17	
6	F13A1	18	D18S51
7	**D7S820**	19	
8	LPL, **D8S1179**	20	
9		21	D21S11
10		22	*D22S445*
11	**TH01**	X	HPRTB, **Amelogenin**
12	VWA, CD4	Y	DYS19, **DYS391**, **Amelogenin**

Exhibit 12. **Primer sequences designed for STR markers** tested with mass spectrometry. For the PCR product size produced with these primers, see Exhibit 2. "b" is listed for biotin; ( ) brackets the cleavable base

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

Exhibit 13. **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 to the repeat region for a primer pair originally published by Li, H. et al. (1993) *Hum. Mol. Genet.* 2: 1327. 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 (over 8,000 Da) compared to 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

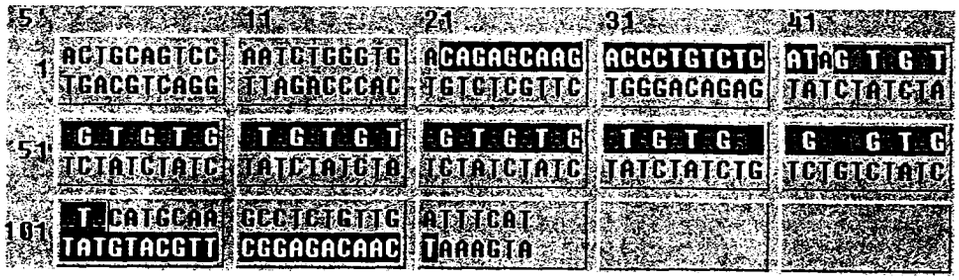


Exhibit 14. **Mass ladder file for STR loci analyzed in this study**. Different masses exist for a locus due to different primer positions. The most commonly used primer sets are highlighted. The reference mass does not include non-template addition by the polymerase. The genotyping program automatically adds 313 Da to the reference mass for compensation of adenylation.

locus name	repeat mass	ref allele	ref mass (in Da)	min repeat	max repeat	forward primer	reverse primer
TH01	1260	9	21133	3	12	TH01-F2	TH01-R2-GTS
TH01b	1260	9	18685	3	12	TH01-F	TH01-R-GTS
TH01b2	1260	9	18372	3	12	TH01-F (w/o+A)	TH01-R-GTS
TH01c	1260	9	19856	3	12	ddC	TH01-R2-GTS
TPOX	1260	11	21351	5	14	TPOX-F-GTS	TPOX-R
TPOXc	1260	11	15119	5	14	TPOX-F-GTS	ddC
CSF	1260	12	28890	5	16	CSF-F3-GTS	CSF-R2
CSFb	1211	12	27989	5	16	CSF-F3	CSF-R3-GTS
CSFc	1260	12	22993	5	16	CSF-F3-GTS	ddC
AMEL	951	6	27333	3	7	AMEL-F-GTS	AMEL-R
AMELb	924	6	18239	4	7	AMEL-F2	AMEL-R2-GTS
D7S820	1260	12	21639	5	16	D7S820-F-GTS	D7-R
D7S820c	1260	12	16949	5	16	D7S820-F-GTS	ddC
D3S1358	1211	16	34300	9	21	D3-F	D3-R-GTS
D3S1358b	1260	15	25624	9	21	D3-F2-GTS	D3-R2
D3S1358c	1260	15	15994	9	21	D3-F2-GTS	ddC
D16S539	1211	11	18180	4	16	D16-F	D16-R-GTS
D16S539b	1260	11	26668	4	16	D16-F4-GTS	D16-R4
D16S539c	1260	11	14765	4	16	D16-F4-GTS	ddC
D16S539d	1211	11	25572	4	16	D16-F4	D16-R4-GTS
D16S539e	1211	11	19127	4	16	D16-F	D16-R2-GTS
FGA	1202	21	37180	15	30	FGA-F2-GTS	FGA-R2
D8S1179	1211	12	24417	8	18	D8S1179-F-GTS	D8S1179-R
DYS391	1211	9	23004	7	14	DYS391-F-GTS	DYS391-R

Exhibit 15. **Mass spectrum of a TH01 allelic ladder** reamplified from AmpF*ℓ*STR® Green I allelic ladders. The PCR product size of allele 10 is only 83 bp with a measured mass of 20,280 Da and separation time of 204 μsec. The allele 9.3 and allele 10 peaks, which are only a single nucleotide apart, differ by only 1.5 μsec on a separation time scale and can be fully resolved with this method.

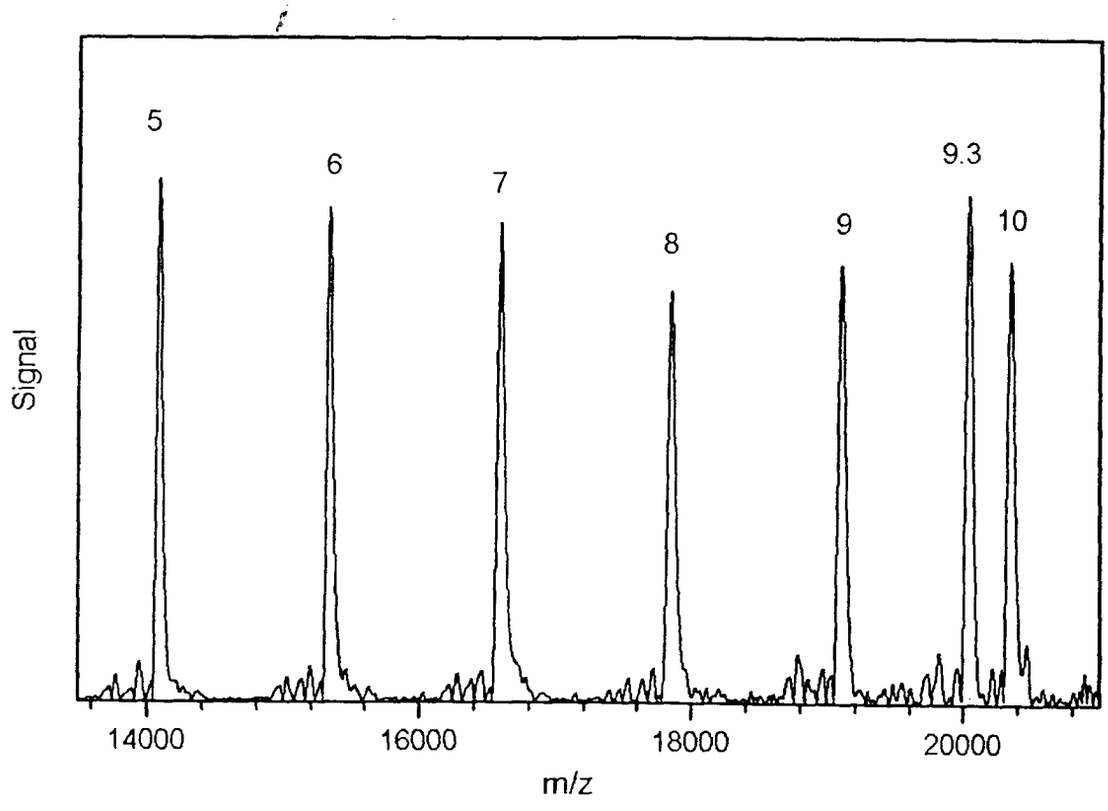


Exhibit 16. **Mass spectra for CSF1PO, TPOX, TH01, and amelogenin** using K562 DNA. Genotypes agree with results reported by the manufacturer (Promega Corporation). The numbers above the peaks represent the allele calls based upon the observed mass. The allele imbalance on the heterozygous samples is due to the fact that the K562 strain is known to contain an unusual number of chromosomes and some of them are represented more than twice per cell. The TH01 peak is split because it is not fully adenylated. See *Int. J. Legal Med.* article for more details (Butler et al. 1998b).

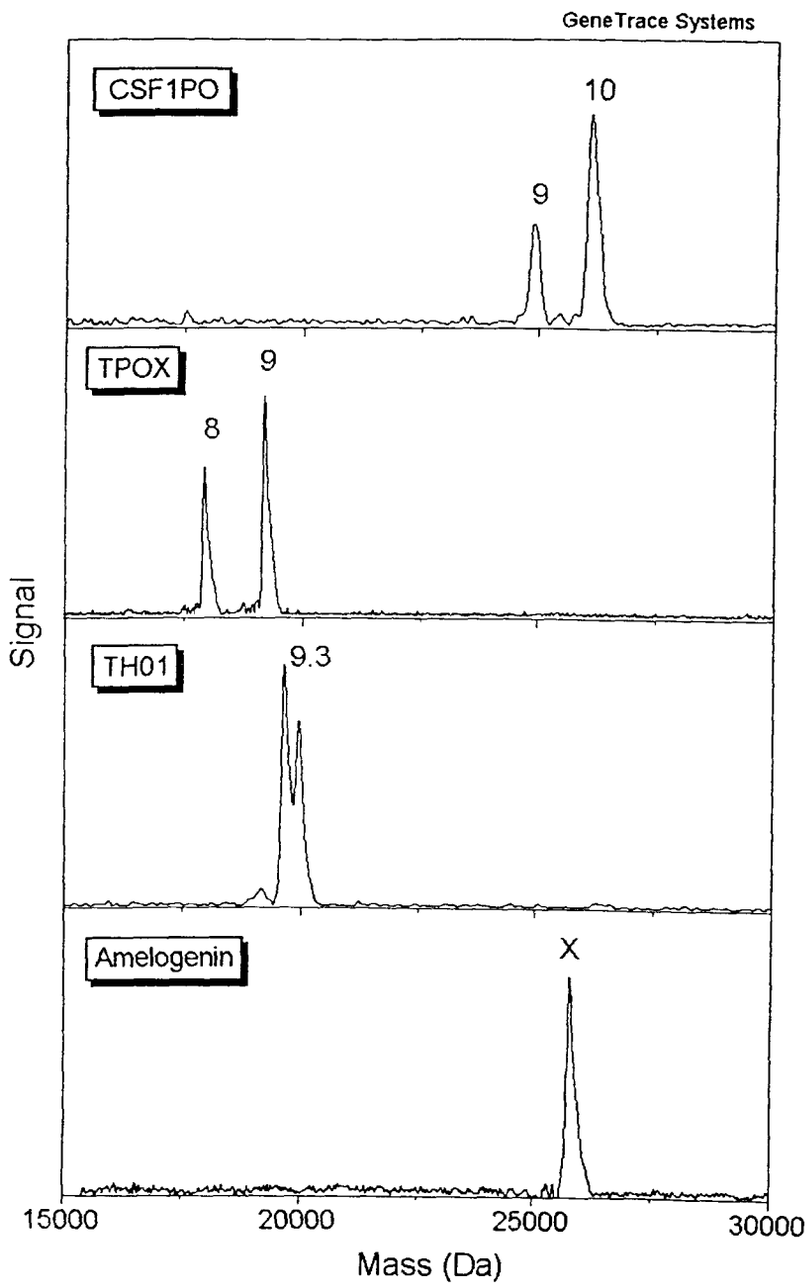


Exhibit 17. **Mass spectra of STR allelic ladders** from CSF1PO, TPOX, TH01, and VWA. The numbers above each peak designate the allele name (number of repeats). Peak widths vary between samples based on DNA size and salt content. Smaller sizes (e.g., TH01) give sharper peaks than larger sizes (e.g., CSF1PO). On a mass scale as shown here, each nucleotide is approximately 300 Daltons (Da). The VWA ladder was digested with *Dpn*II restriction enzyme following PCR to reduce the overall size of the amplicons.

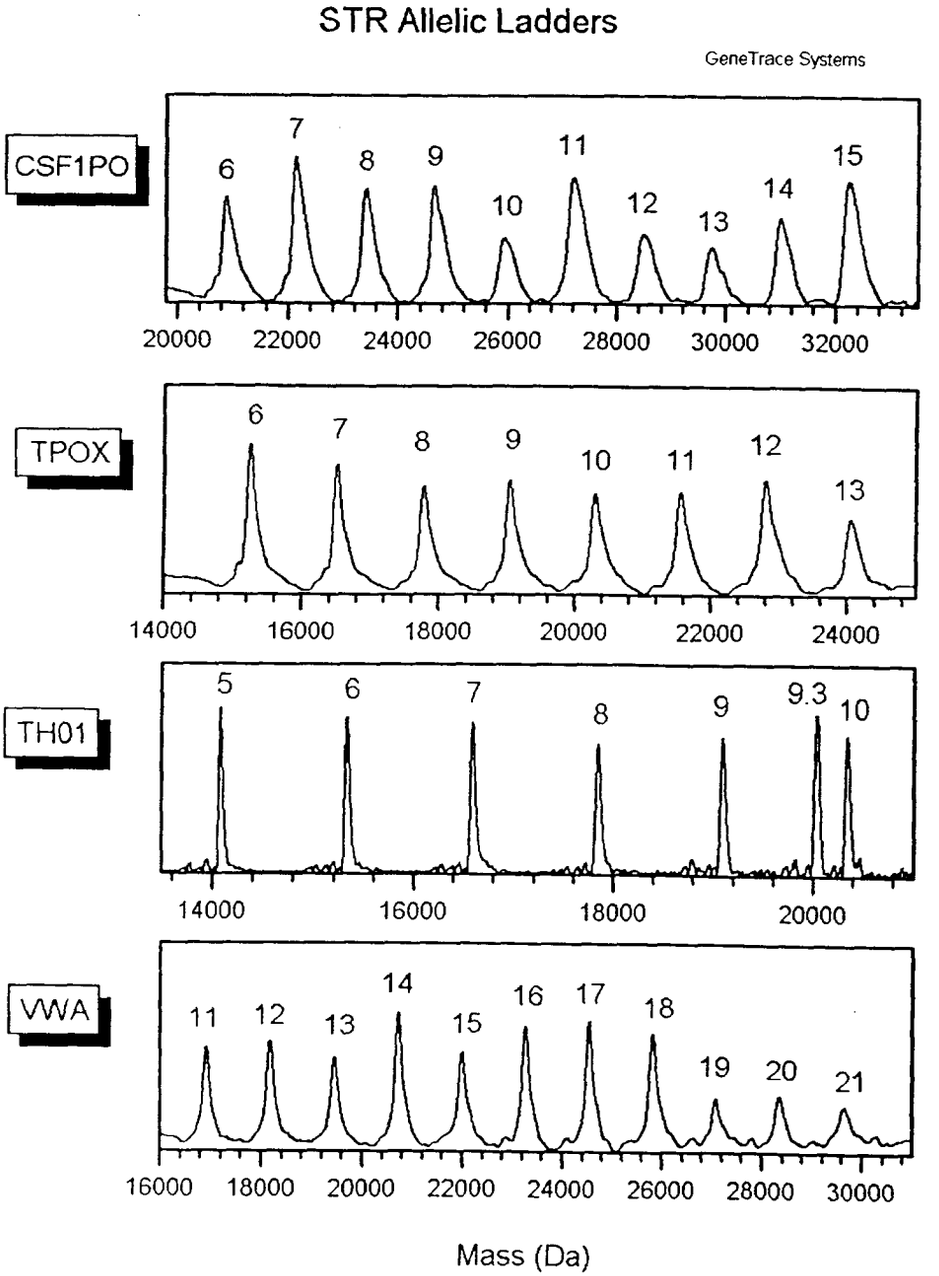


Exhibit 18. **Mass spectra of D5S818 allelic ladders** from two different manufacturers demonstrating that the GeneTrace primers can amplify all common alleles for this particular STR locus. The PE Applied Biosystems D5S818 ladder contains 10 alleles (top panel) while the Promega ladder contains only 8 alleles (bottom panel). The GeneTrace primers bind internal to commerically available multiplex primers and all alleles in the commercial allelic ladders are therefore amplified.

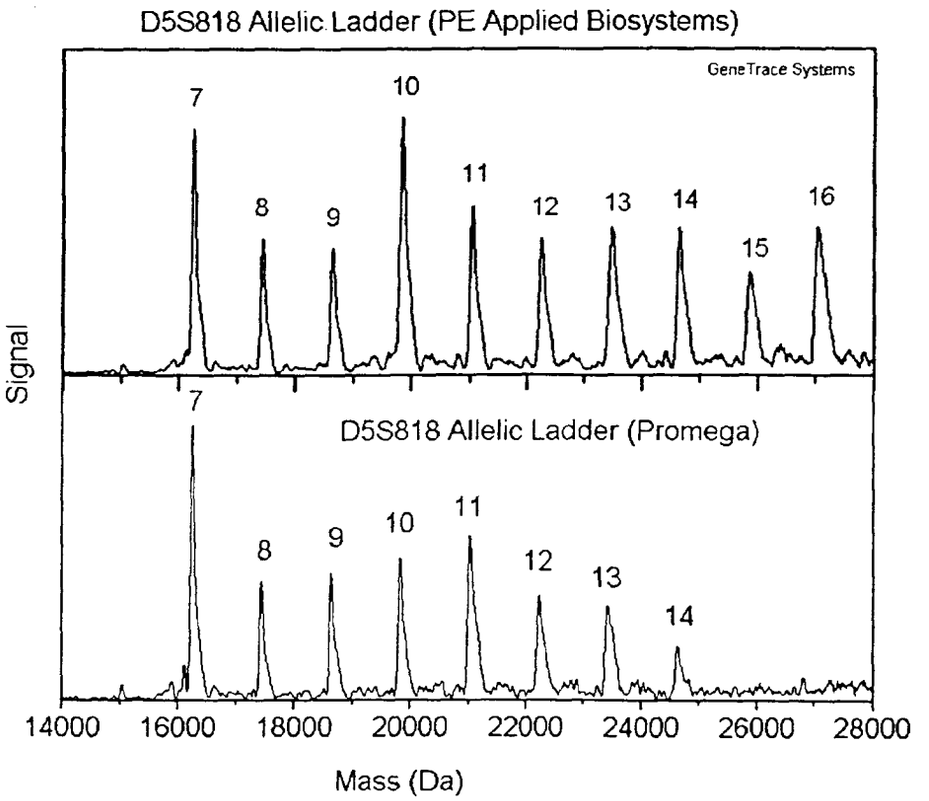


Exhibit 19. **Mass spectra of a D8S1179 sample illustrating the benefit of a dideoxynucleotide termination approach.** The top panel displays a result from a regular PCR product while the bottom panel contains the same sample treated with a linear amplification mix containing a ddG terminator with dA, dT, and dC deoxynucleotides. With the ddG approach, the problem of incomplete adenylation (both –A and +A forms of a PCR product) is eliminated and the amplicons are smaller which improves their sensitivity and resolution in the mass spectrometer. Note that the genotype (i.e., 11 and 13 repeats) is identical between the two approaches, even though almost 8,000 Da are removed with the ddG termination.

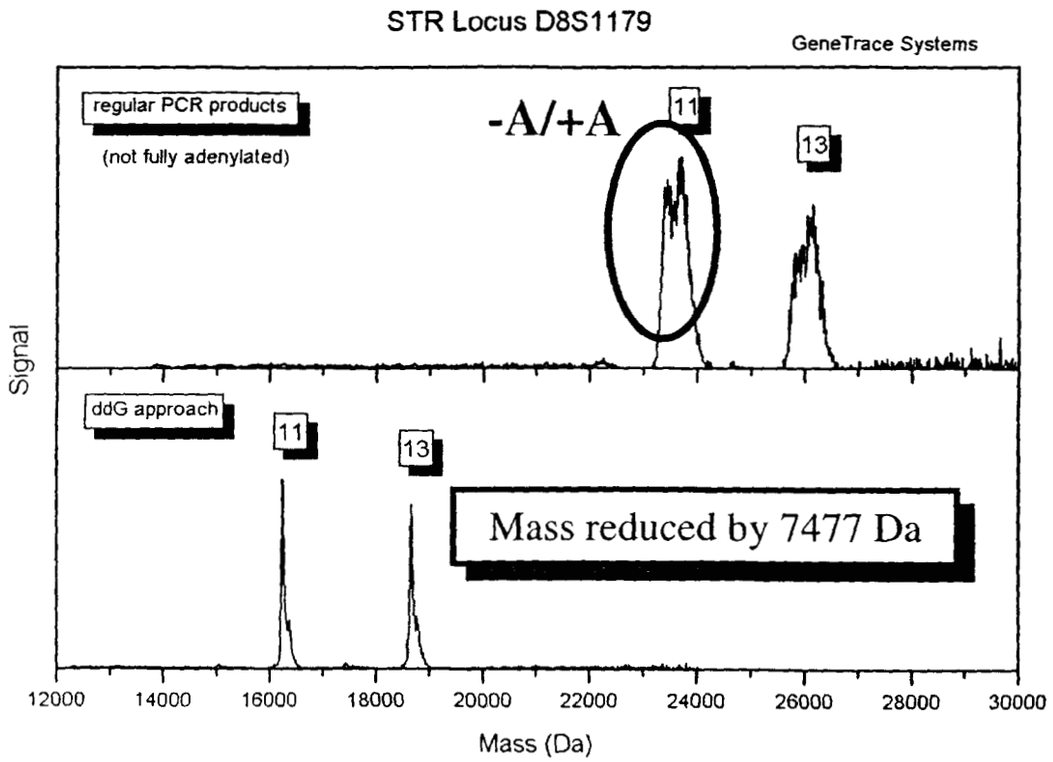
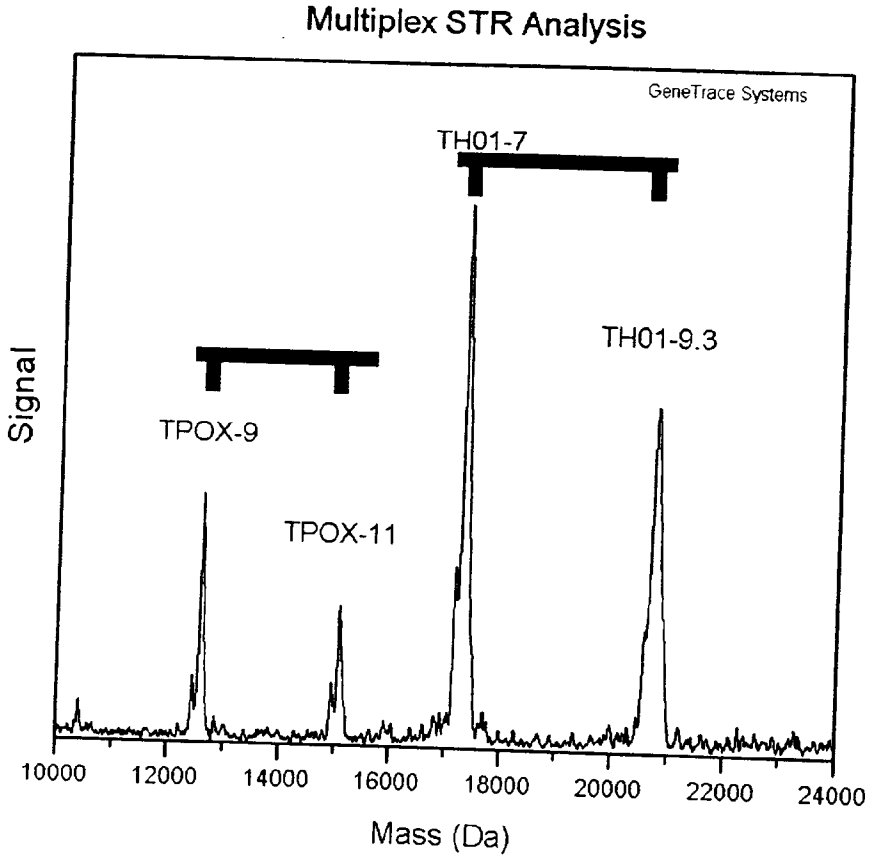


Exhibit 20. **Mass spectrum of an STR multiplex sample with non-overlapping alleles.** The two loci, TH01 and TPOX, were co-amplified using the ddC termination approach. Note that the mass accuracy is improved for peaks closer to the calibration standard of 10998 Da.



Expected Mass	Observed Mass	ΔMass
TPOX-9 = 12581 Da	12584 Da	+3 Da
TPOX-11 = 15101 Da	15080 Da	-21 Da
TH01-7 = 17336 Da	17305 Da	-31 Da
TH01-9.3 = 20803 Da	20751 Da	-52 Da

Exhibit 21. **Mass spectrum of STR multiplex mixture of TH01 and TPOX allelic ladders.** The PCR products from the two loci differ by 120 Da. The ladders were reamplified from AmpF*ℓ*STR® Green I kit materials. The TH01 ladder ranges from 5-10 repeats with 9.3 included while the TPOX ladder ranges from 6-13 repeats with 13 not shown here.

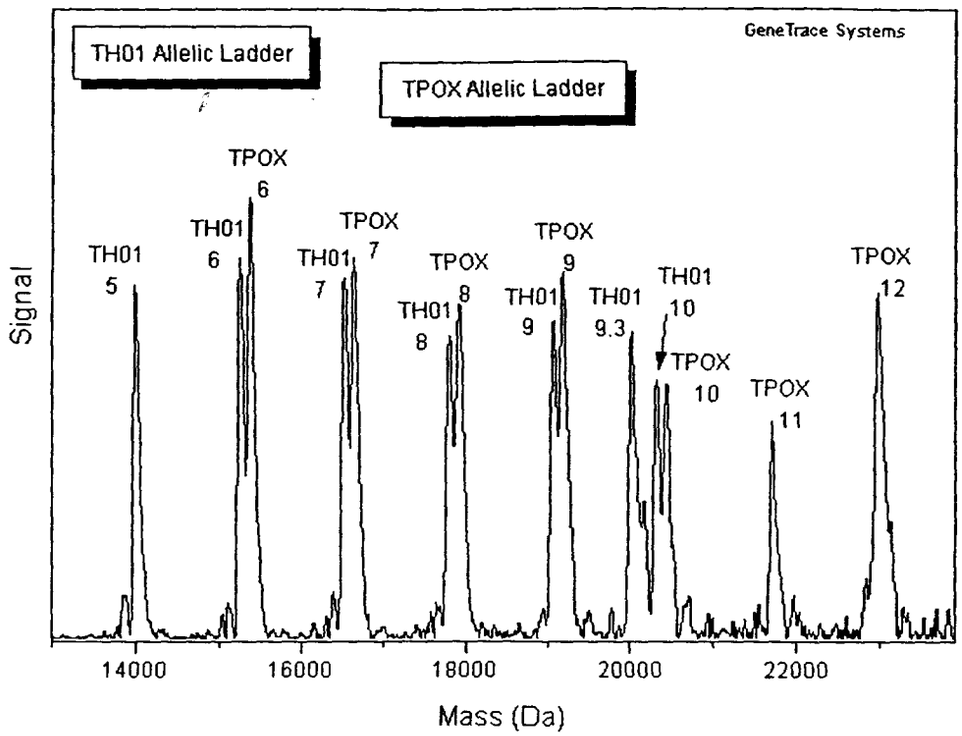


Exhibit 22. **Mass spectrum of a multiplexed TH01 and TPOX sample** showing interleaved amplicons that are only 85 Da apart.

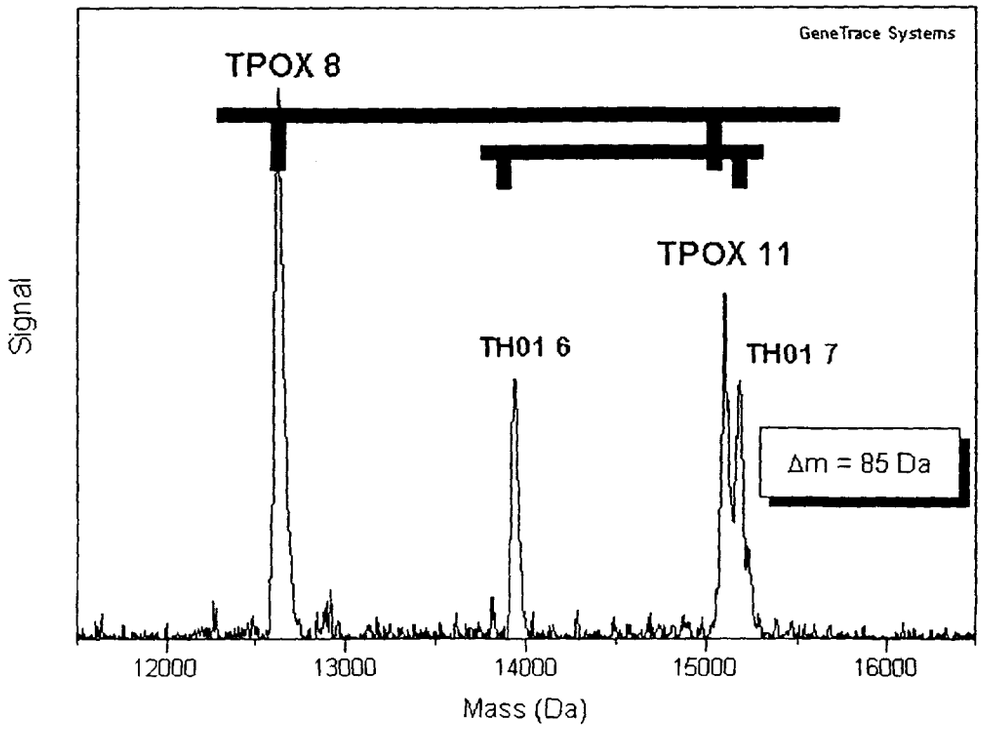


Exhibit 23. **Multiplexing strategies for STR genotyping using FTA™ paper.**

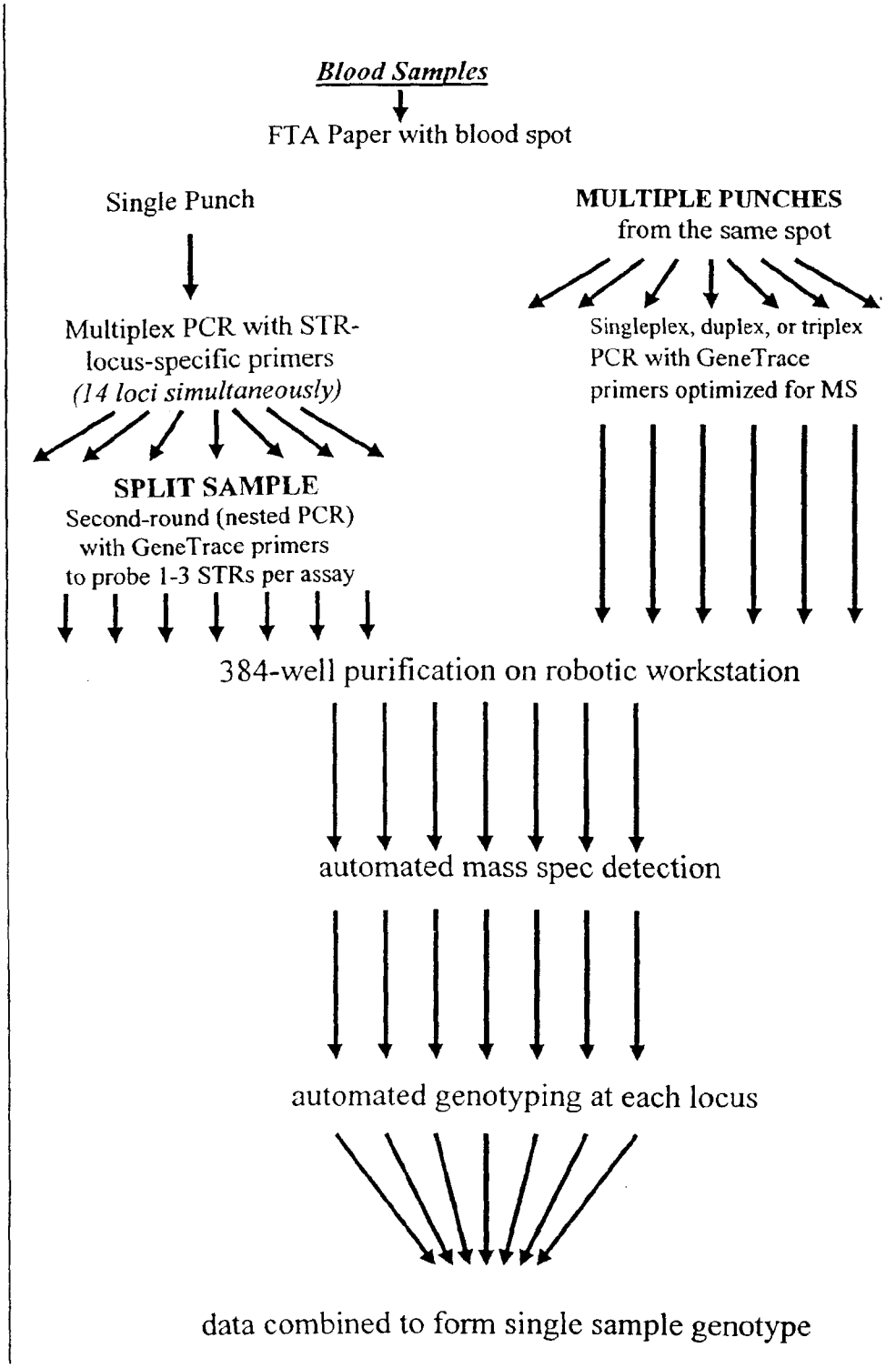


Exhibit 24. **Table of 15 replicate analyses of a TPOX allelic ladder** to measure mass precision and accuracy. The precision was less than 30 Da for a single standard deviation, which corresponds to less than 0.1 nucleotide, and the measured mass accuracy from the calculated expected allele masses was on average ~30 Da. Across the 8 alleles in the ladder, there are 120 data points used to make this determination. All numbers are in Daltons (Da). Percentage error was calculated as (observed-expected)/expected. This same data was presented in histogram format in our *Int. J. Legal Med.* paper (see Fig. 1; Butler *et al.* 1998).

	allele 6	**allele 7**	**allele 8**	**allele 9**	**allele 10**	**allele 11**	**allele 12**	**allele 13**
Expected Mass (Da)	15345	16605	17865	19125	20385	21644	22904	24164
1	15346	16623	17903	19130	20388	21623	22860	24074
2	15387	16667	17901	19129	20387	21639	22893	24114
3	15372	16629	17887	19143	20400	21615	22877	24091
4	15385	16653	17903	19163	20384	21642	22903	24076
5	15388	16642	17898	19155	20337	21654	22870	24111
6	15336	16600	17857	19105	20362	21599	22832	24064
7	15388	16637	17894	19131	20383	21635	22904	24110
8	15363	16618	17872	19129	20368	21604	22853	24087
9	15365	16628	17891	19150	20385	21620	22892	24087
10	15373	16638	17892	19136	20394	21631	22878	24085
11	15383	16640	17896	19152	20387	21621	22884	24129
12	15388	16648	17912	19172	20388	21623	22850	24149
13	15407	16660	17941	19208	20425	21674	22944	24148
14	15410	16659	17930	19174	20425	21666	22893	24132
15	15390	16648	17915	19157	20423	21636	22897	24126
average	15379	16639	17899	19149	20389	21632	22882	24106
std dev	20.2	17.8	20.6	24.7	23.7	21.0	27.2	27.3
%RSD	0.13	0.11	0.12	0.13	0.12	0.10	0.12	0.11
% error	0.22	0.21	0.19	0.13	0.02	-0.06	-0.10	-0.24
obs-exp	33.7	34.3	34.5	23.9	4.1	-11.9	-22.0	-58.5

Exhibit 25. **Mass spectra of TPOX PCR products from various amounts of K562 DNA template material.** This sensitivity test demonstrates that DNA templates in the quantity range of 0.5 ng to 20 ng may be effectively amplified and detected by mass spectrometry.

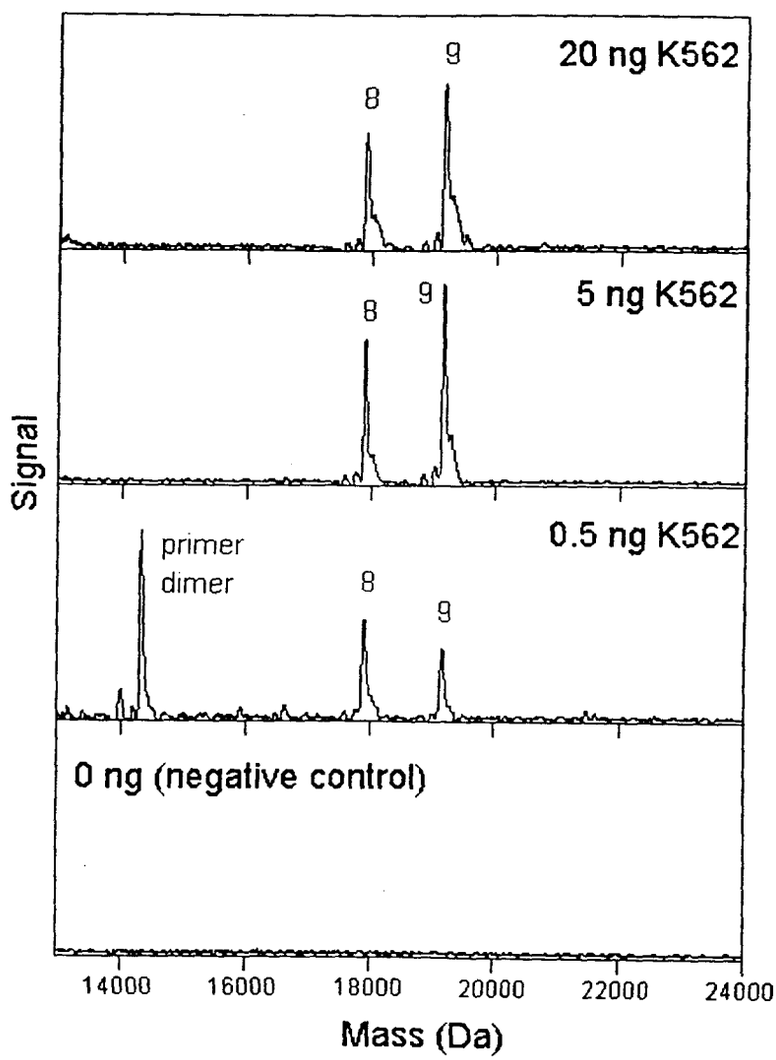


Exhibit 26. **STR Data Collection Times for CAL DOJ Samples**. Comparison of ABI 310 run times with multiple mass spec runs. Note that the data collection speed increased almost 10-fold during 1998 due to a number of improvements in the GeneTrace process. Data collection speed is a combination of laser rate, total number of shots taken, and sample cleanliness among other factors. Overall the mass spec data collection speed is approximately 2 orders of magnitude faster than the ABI 310. Exhibits 40-49 list the observed masses for these samples measured by our mass spec method.

ABI 310 Runs

January 5, 1999 – January 8, 1999

- Multiplex STR detection and analysis (COfiler™ kit amplifies 6 STRs + amelogenin simultaneously: TH01, TPOX, CSF1PO, D3S1358, D16S539, D7S820)
- ~54 hours of continuous operation (107 injections that included 88 samples, 9 allelic ladders, and reinjections of 10 samples with ~30 min per injection)
- ~2209 sec/sample or 316 sec/genotype (**~5 min/genotype**)

Mass Spec Runs

Date	STR Loci Tested	Number of samples	Time Required for Data Collection	Average Time per Sample
Feb 11, 1998 (160921.hed)	TH01	88 + 8 controls	76 min for 96 samples	~50 sec/sample
May 6, 1998 (174405.hed)	TH01	6 samples	2.5 min for 6 samples	~25 sec/sample
October 1, 1998 (195607.hed)	TPOX, CSF1PO, D3S1358, D7S820	72 + 8 controls for each locus	60 min for 320 samples	11.3 sec/sample
Jan 12, 1999 (102808.hed)	TH01, TPOX, CSF1PO, D7S820	88 + 8 controls for each locus	30 min for 384 samples	**4.7 sec/sample**
Jan 14, 1999 (140329.hed)	TH01, D16S539, D3S1358, Amelogenin	88 + 8 controls for each locus	37 min for 384 samples	5.8 sec/sample
Feb 12, 1999 (154954.hed)	D3S1358, FGA, D8S1179, DYS391	88 + 8 controls for each locus	41 min for 384 samples	6.4 sec/sample
March 26, 1999 (120423.hed)	TH01, TPOX, CSF1PO, D16S539	88 + 8 controls for each locus	51 min for 384 samples	8.0 sec/sample

Exhibit 27. **Comparison of observed allele masses collected six months apart**. This plot compares 57 allele measurements for six different TPOX alleles. The ideal line is shown on the same plot to demonstrate how reproducible the masses are over time. The average standard deviation of allele mass measurements between these two data sets was 47 Da. This result further confirms that no allelic ladders or other internal DNA standards are needed to obtain accurate measurements with mass spectrometry.

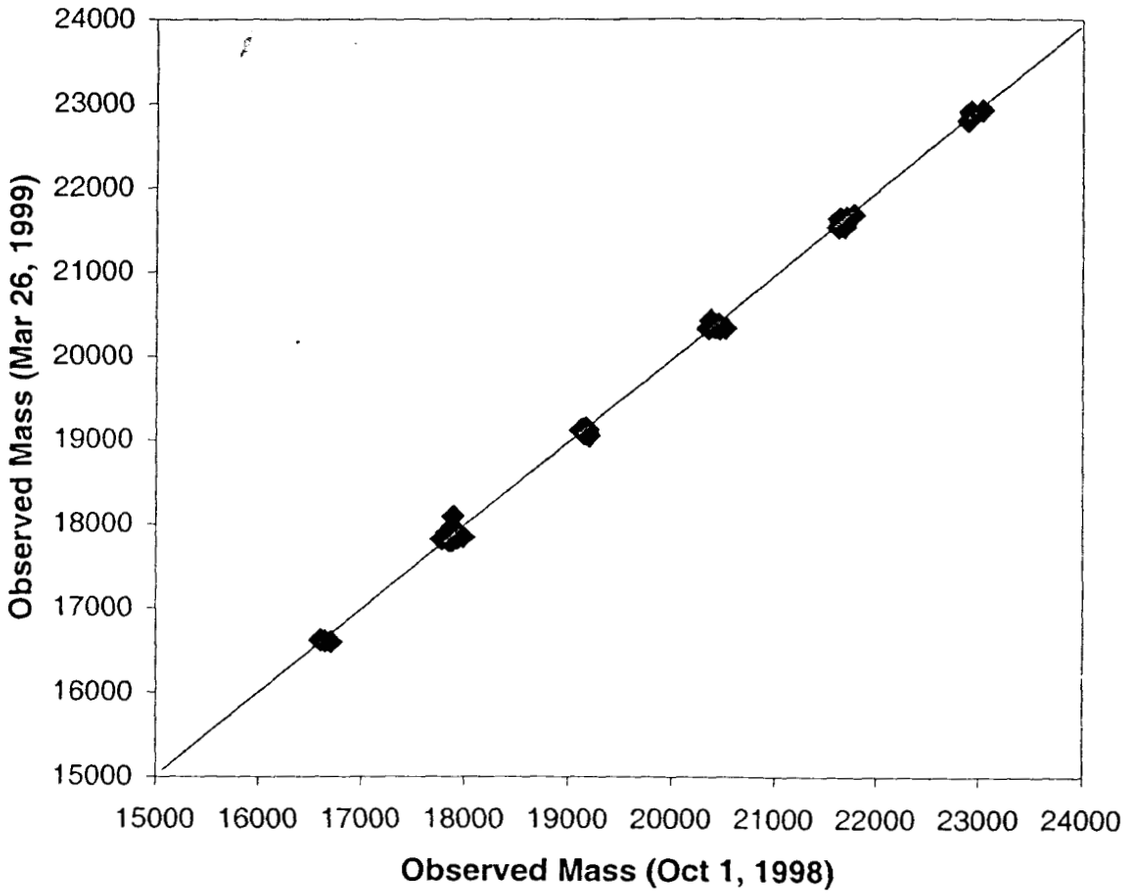


Exhibit 28. **Plot of X allele mass offset verses Y allele mass offset for 88 amelogenin samples.** The red box shows +/- 100 Da around the expected values and the green box shows +/- 300 Da. The blue line is the ideal situation where "heterozygous" peaks would shift in unison compared to the expected masses.

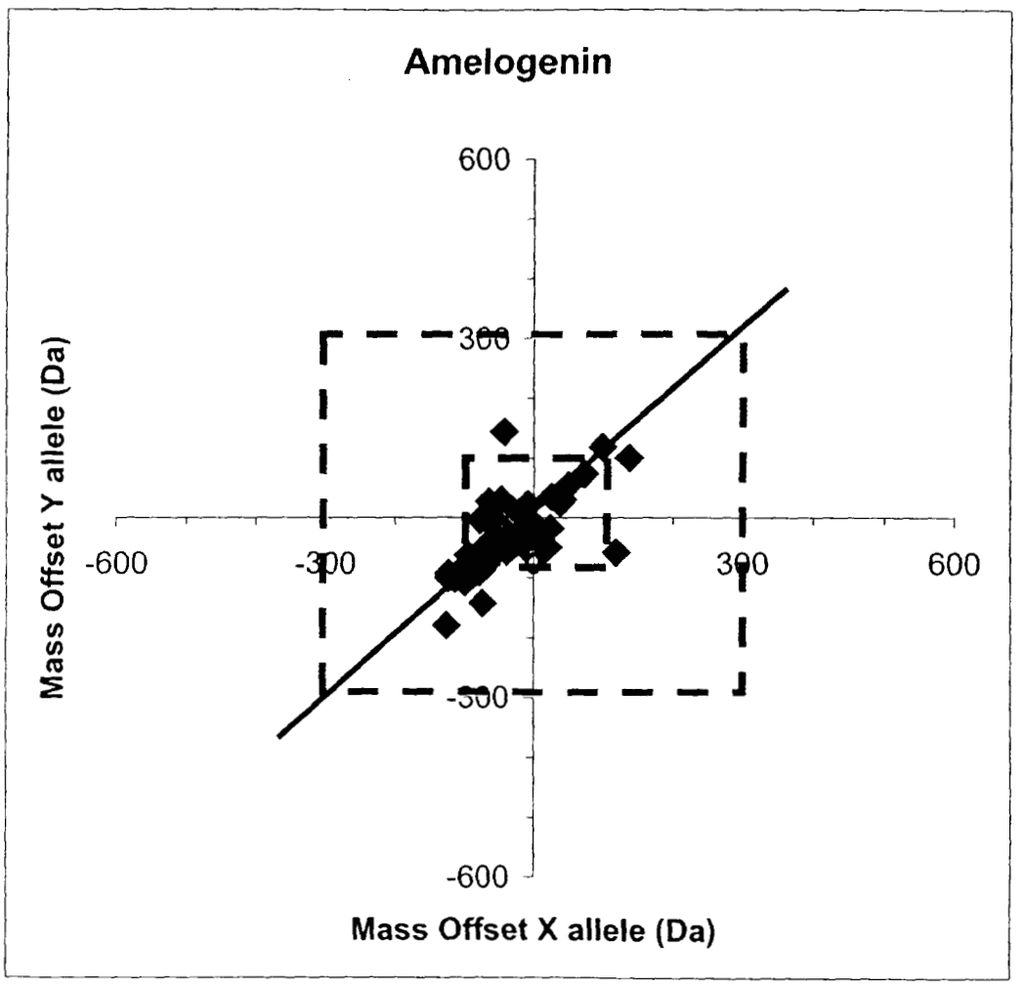


Exhibit 29. **Plot of allele mass offsets (allele 1 verse allele 2) for heterozygous samples from 4 different loci.** These are 88 CAL DOJ samples for the STR loci TH01, TPOX, CSF1PO, and D16S539.

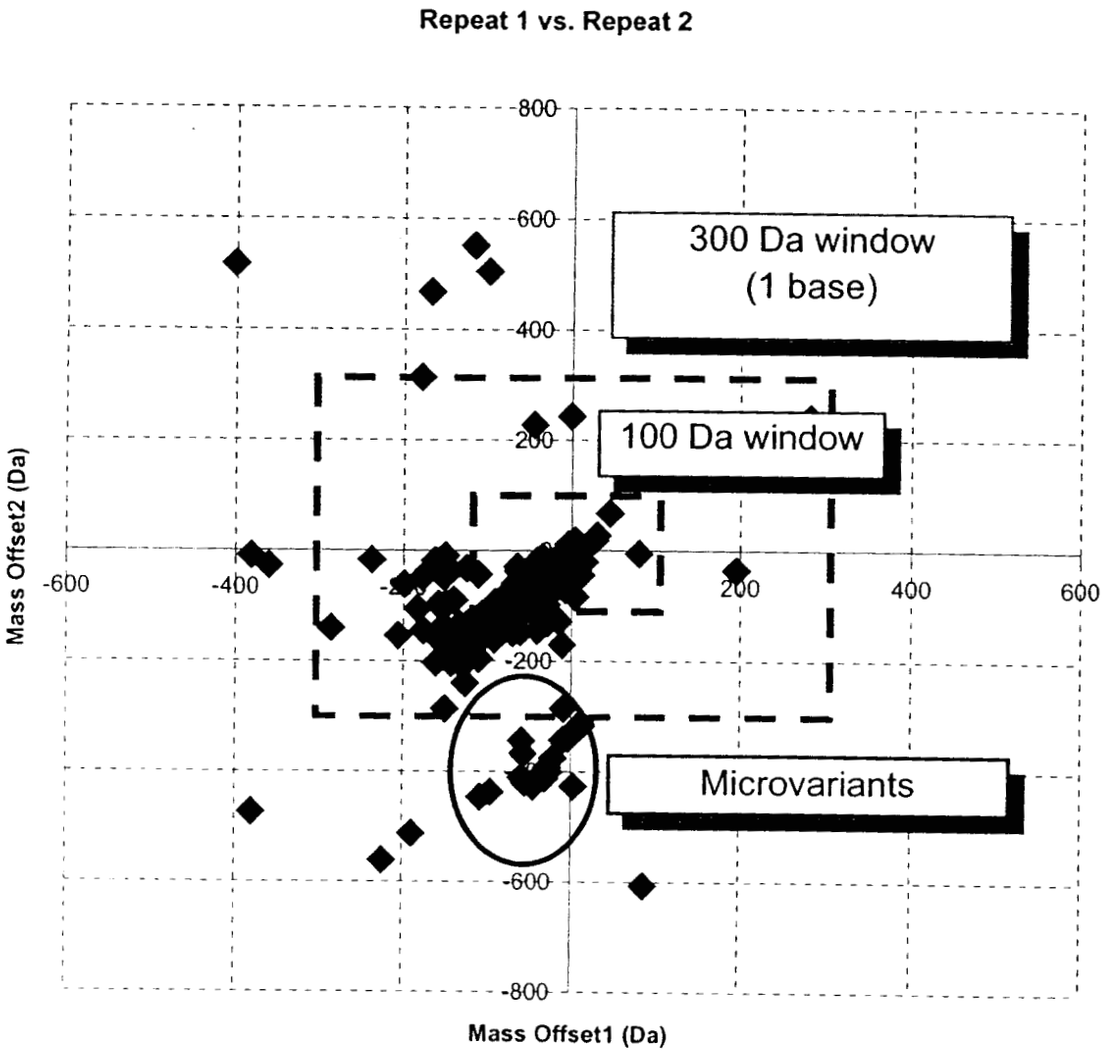


Exhibit 30. **Plot of measured masses verses sample number for 4 different STR loci.** FGA, which has a high degree of scatter, is the most polymorphic marker and has the largest mass alleles. The highest and lowest alleles observed for each STR locus in this study are shown on the plot.

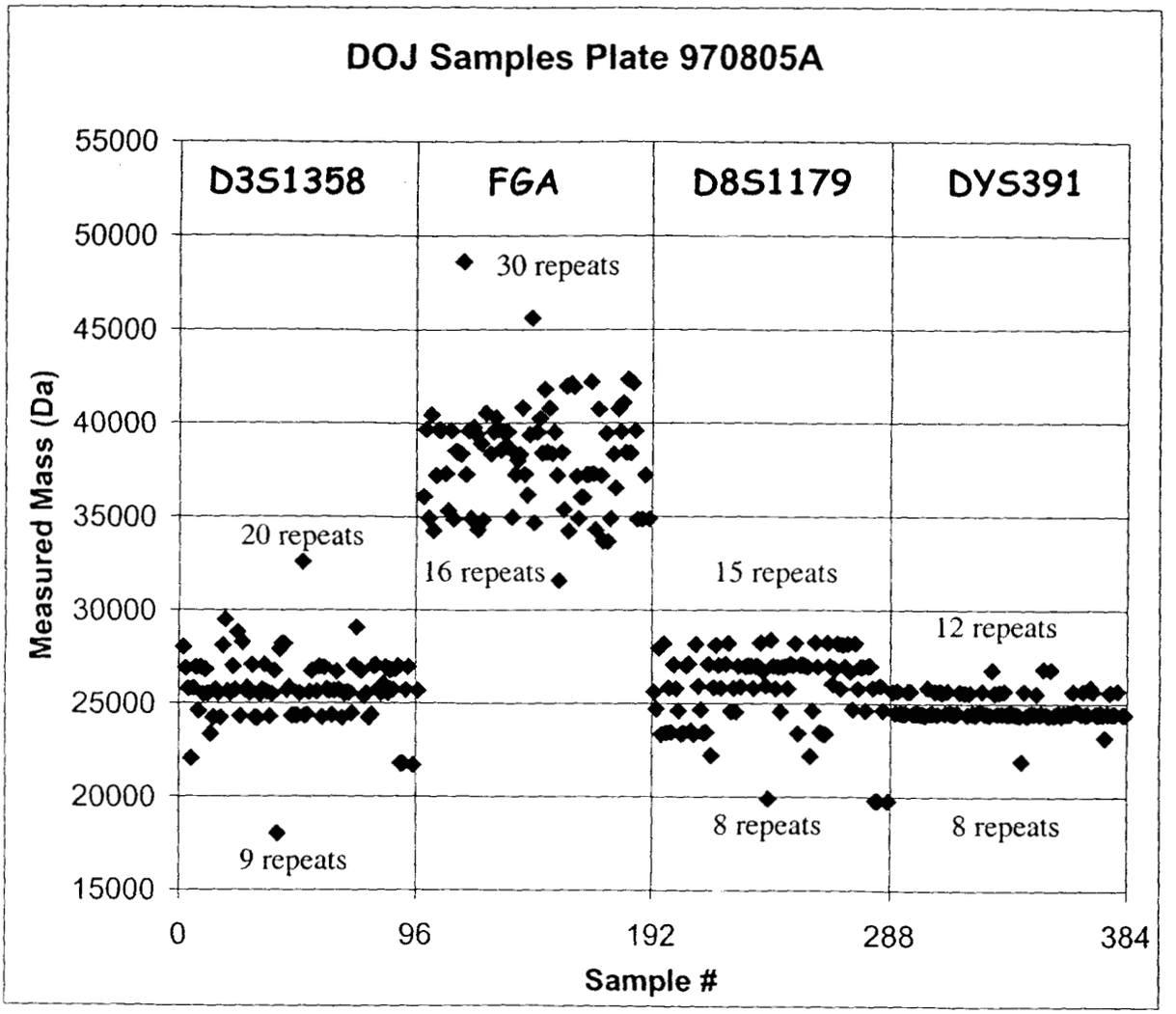


Exhibit 31. **Mass spectra comparing an STR sample amplified with TaqGold polymerase (PE Applied Biosystems) and Tsp polymerase (Life Technologies).** The Tsp polymerase favors production of the non-adenylated form of PCR products, which results in a single peak for each allele (bottom panel) while TaqGold produces a mixture of –A and +A peaks, which leads to two peaks for each allele (top panel). The peak masses in Daltons are indicated next to each peak. Mass difference measurements between the +A and –A peaks reveal that a "T" is added by TaqGold instead of the expected "A" since the mass difference is 308 Da and 305 Da (expected masses: T = 304 Da and A = 313 Da). The sample's genotype was TH01 6,8.

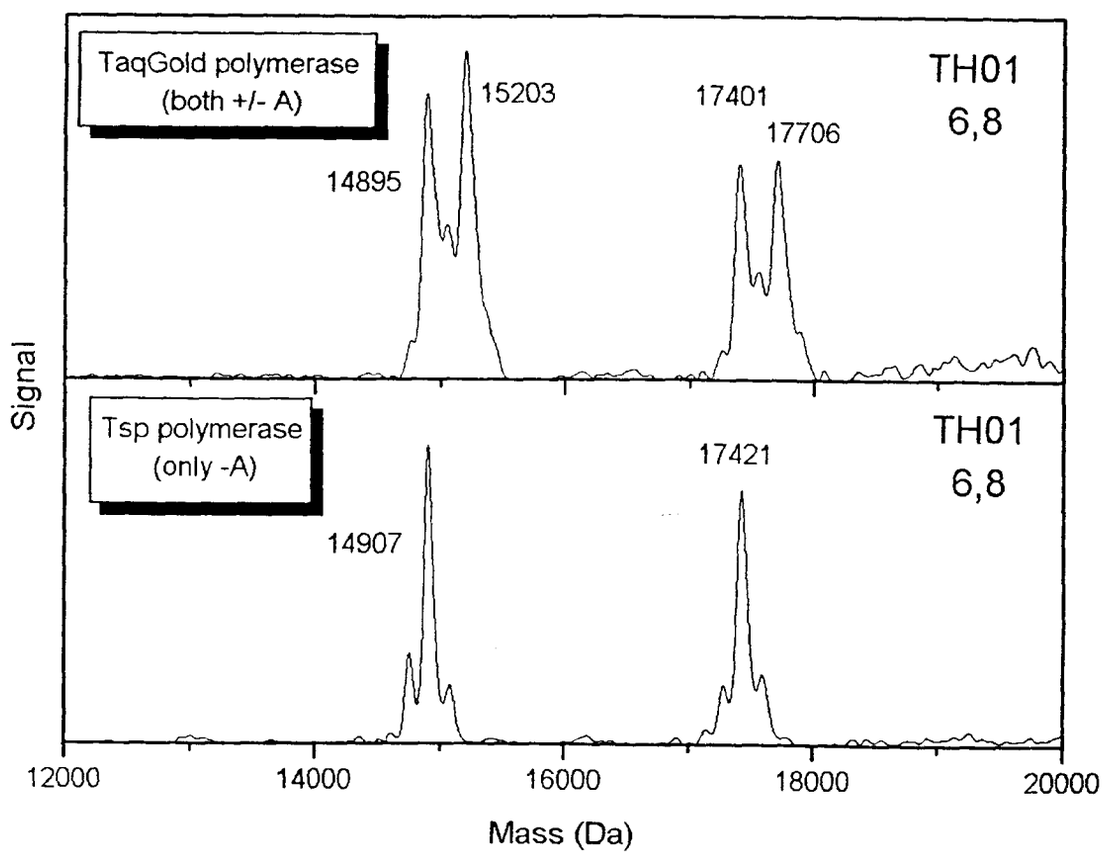


Exhibit 32. **Mass differences between TH01 alleles** using either upper strand (TCAT repeat) or lower strand (AATG repeat). The upper strand was discernible from the lower strand due to the different sequence contents of the repeats on the two strands. The STR repeat structure and nucleotide content thus can be seen using mass spectrometry. Note that the upper strand mass difference between alleles 9.3 and 10 is 306 Da, or a "T", while the lower strand mass difference between these same two alleles is 315 Da, or an "A". For more details, see our *Int. J. Legal Med.* article (Butler et al. 1998b).

Upper Strand

	Expected (Da)	Observed (Da)
Allele 5-6	1211	1210
Allele 6-7	1211	1211
Allele 7-8	1211	1215
Allele 8-9	1211	1215
Allele 9-9.3	907	915
Allele 9.3-10	304	306
Allele 9-10	1211	1221

Repeat = TCAT = 1210.8 Da
= --CAT = 906.6 Da

Lower Strand

	Expected (Da)	Observed (Da)
Allele 5-6	1260	1259
Allele 6-7	1260	1262
Allele 7-8	1260	1269
Allele 8-9	1260	1267
Allele 9-9.3	947	948
Allele 9.3-10	313	315
Allele 9-10	1260	1263

Repeat = AATG = 1259.8 Da
= --ATG = 946.6 Da

Exhibit 33. **Mass spectrum demonstrating detection of stutter products** with a particularly stutter prone dinucleotide repeat locus. The mass differences between the stutter product peaks and the allele peaks can be used to determine the repeat sequence that is present on the measured DNA strand. Note that the amount of stutter is larger in the longer repeat allele compared to the shorter allele.

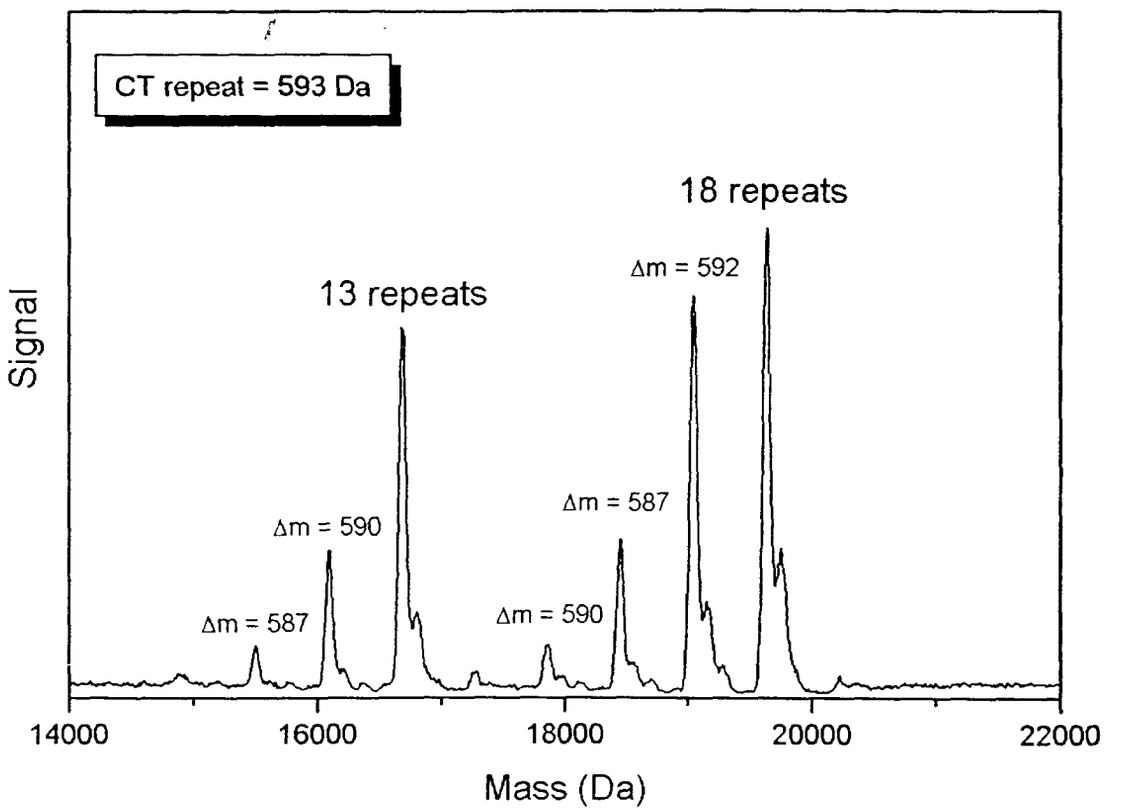


Exhibit 34. **Sample sets run on ABI 310 Genetic Analyzer** with AmpF*ℓ*STR® Profiler Plus™ or AmpF*ℓ*STR® COfiler™ fluorescent STR kits (PE Applied Biosystems).

Sample Set	# of samples	Kit Used	MS Data Compared	Comments
DOJ plate	88	COfiler	TH01, TPOX, CSF1PO, D7S820, D3S1358, D16S539, D8S1179, FGA, DYS391, amelogenin	*See Exhibits 40-49*
CEPH/Diversity plate	92	COfiler + ProfilerPlus	TH01, TPOX, CSF1PO, D7S820, D3S1358, D16S539, D8S1179, amelogenin	*See Exhibit 73*
Stanford male samples	37	ProfilerPlus	For Y SNP testing (not completed)	*See Exhibits 37-38* new microvariants seen in D18, D21, FGA
Butler family samples	34	COfiler		
JMB family samples	4	COfiler + ProfilerPlus		
Standard templates	3	COfiler + ProfilerPlus	TH01, TPOX, CSF1PO, D7S820, D3S1358, D16S539, D8S1179, FGA, DYS391, amelogenin	*See Exhibit 35* K562, AM209, and UP006
TOTAL TESTED	**258**	221 COfiler 136 ProfilerPlus	**2907 total genotypes measured**	

CAL-DOJ results	88	Profiler (under CAL-DOJ validated method)	*See above*	B11, D3S1358: 15,15.2 incorrectly called 14,15; *See Exhibit 36*

Exhibit 35. **STR genotypes for standard DNA templates K562, AM209, and UP006** obtained using AmpF*ℓ*STR® Profiler Plus™ and AmpF*ℓ*STR® COfiler™ fluorescent STR kits (PE Applied Biosystems). These samples were the primary controls used for testing of newly developed primer sets. Alleles were not always well balanced for the K562 cell line DNA due to possible chromosome imbalances in the original sample. For example, the STR marker D21S11 produced three balanced alleles most likely because three chromosome 21's are present in the K562 genomic DNA. In addition, the allele 24 peak ("24w" = "weak") for the FGA locus in K562 is only one-third the height of the allele 21 peak probably because there are more copies of the chromosome 4 on which the allele with 21 repeats resides. Our K562 genomic DNA came from Promega Corporation. AM209 came from CEPH pedigree 884 (Amish family) and UP006 is an anonymous population sample of European origin purchased from Bios Laboratory.

STR locus	K562	AM209	UP006
TH01	9.3,9.3	6,7	7,9.3
TPOX	8,9	8,9	8,8
CSF1PO	9,10	10,12	11,14
D8S1179	12,12	11,13	13,13
D21S11	29,30,31	30,30	28,31
D18S51	15,16	12,12	13,13
D3S1358	16,16	15,18	15,15
VWA	16,16	18,18	16,17
FGA	21,24w	20,25	19,22
D16S539	11,12	11,12	12,13
D5S818	11,12	11,12	10,11
D13S317	8,8	14,14	11,14
D7S820	9,11	9,10	10,12
Amelogenin	X,X	X,Y	X,Y

Exhibit 36. **An ABI 310 CE electropherogram showing a heterozygous sample with a 15.2 microvariant at STR locus D3S1358**. A high degree of resolution is often necessary to resolve two closely spaced alleles in which one of them is a microvariant. In this case, the two alleles differ by only 2 bp even though the core repeat is a tetranucleotide. In the comparison between STR data collected by CAL DOJ, this was the only sample where a different result was obtained between our analysis and that performed by CAL DOJ on the same STR markers. CAL DOJ reported 14,15 for this sample.

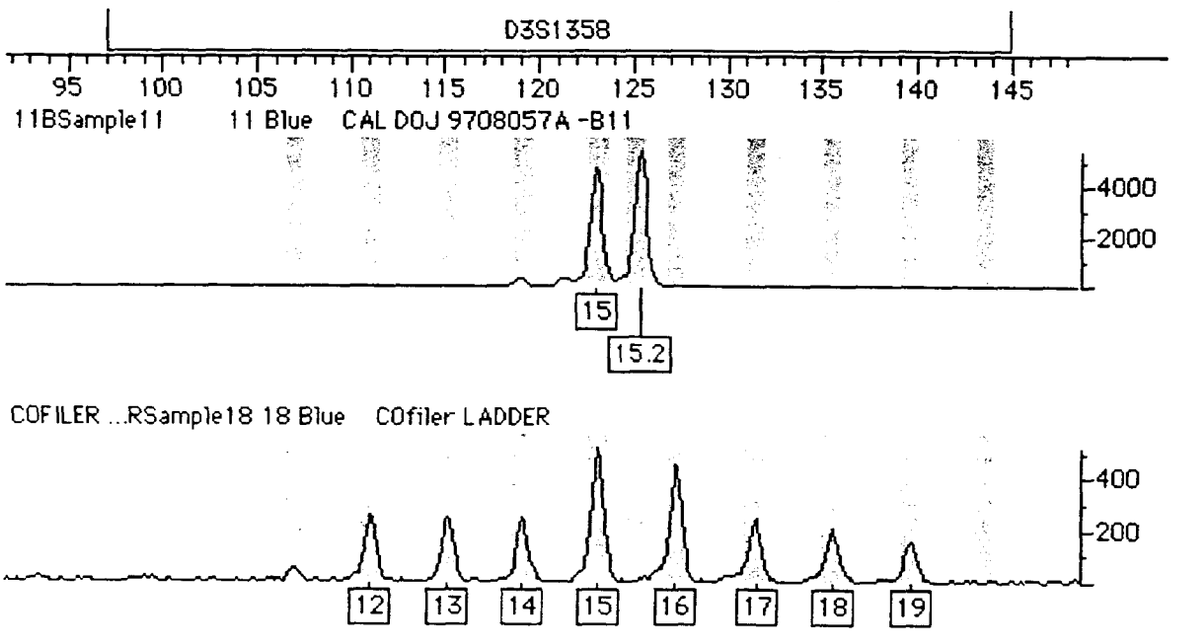


Exhibit 37. **Electropherograms of new STR microvariants seen in the Stanford male population samples.** The D18S51 16.2 allele, D21S11 30.3 allele, and FGA 28.1 allele have not been reported previously in the literature. These plots are views from Genotyper 2.0 with results overlaid on shaded allele bins. Note that the microvariants fall between the shaded bins but that the other allele in the heterozygote set contains complete repeats and falls directly on the shaded (expected) allele bin. The base pair size range is indicated at the top of each plot.

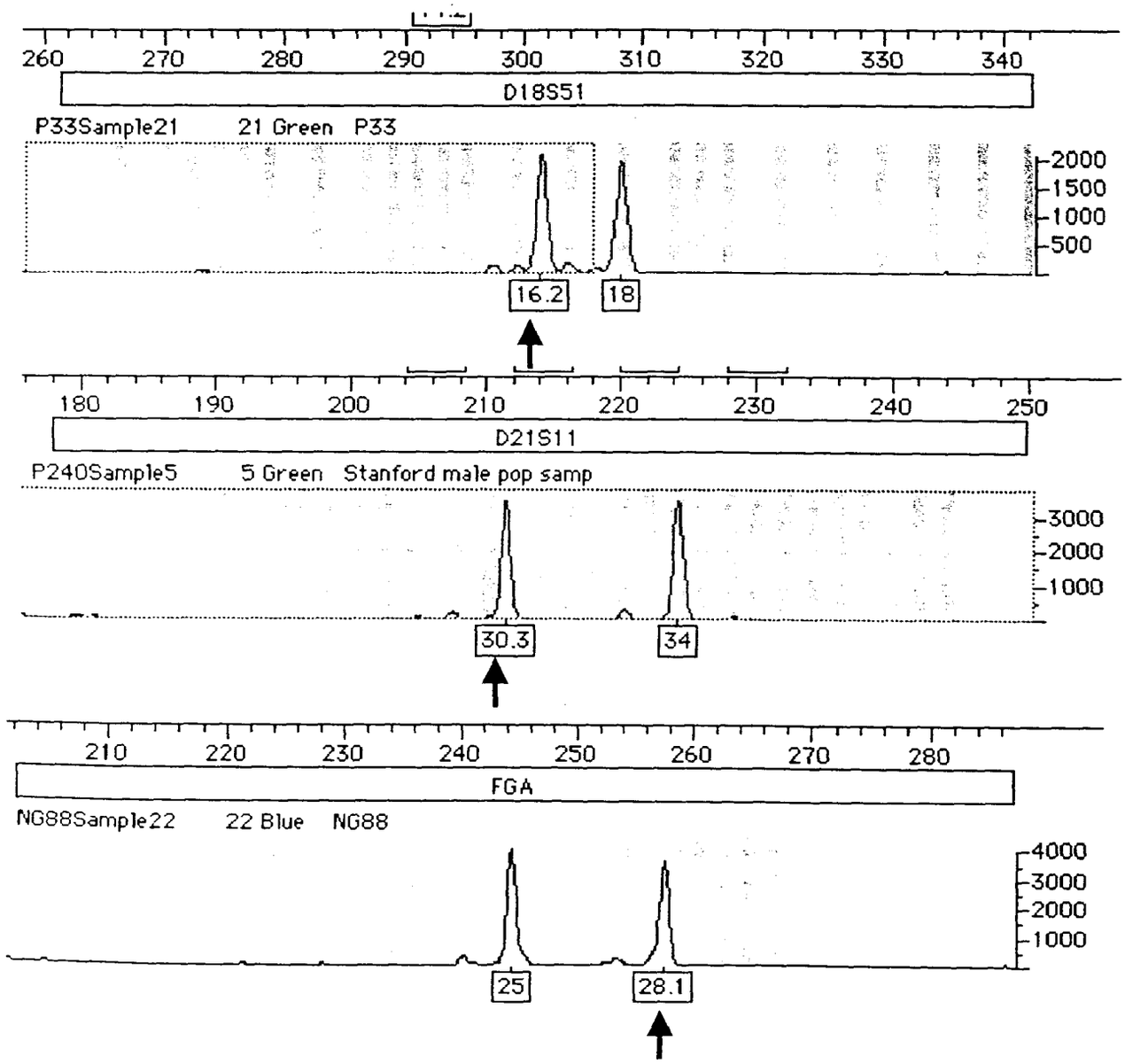
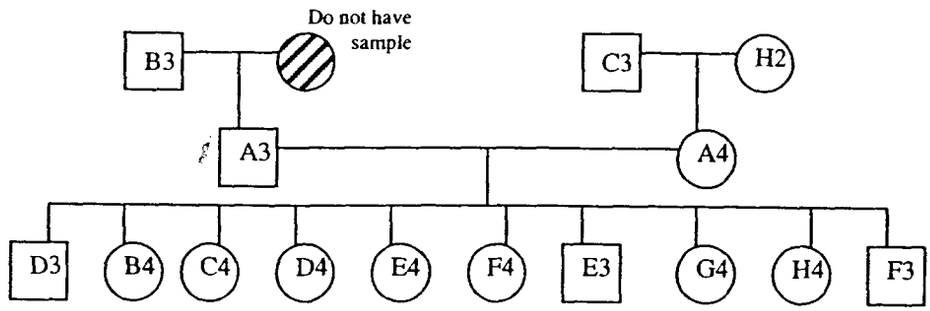


Exhibit 38. **ProfilerPlus™ results from Stanford male population samples.**

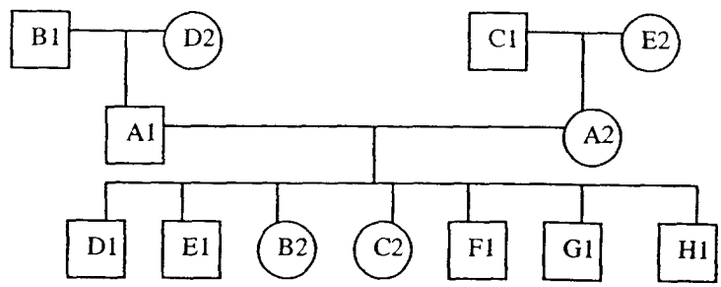
Sample Name	AMEL A1	AMEL A2	CSF A1	CSF A2	D2 A1	D2 A2	D13 A1	D13 A2	D3 A1	D3 A2	VWA A1	VWA A2	FGA A1	FGA A2	D5-A1	D5-A2	D13-A1	D13-A2	D7-A1	D7-A2
Aus21	X	Y	8	12	29	31.2	12	18	15	17	14	17	18	22	12	12	11	12	9	11
Aus28	X	Y	14	15	29	31.2	15	19	16	17	15	17	19	20	12	12	10	12	8	10
Berg15	X	Y	10	12,13	28,30.2	31.2	12,17	14	15,17	18,19	16	17,18	20,21	24	11	13w	8,14	11	7w	10
Berg19	X	Y	14	14	32.2	32.2	12	18	16	16	14	18	22	22	12	13	12	12	9	13
BI12	X	Y	12	12	30	31.2	17	17	16	18	17	19	24	27	12	12	11	11	9	11
Bsk092	X	Y	14	14	29	33.2	13	16	17	17	17	17	23	23	11	12	12	12	11	11
Bsk111	X	Y	12	13	30	32.2	16	19	14	14	16	16	18	19	12	12	8	10	8	9
Bsk118	X	Y	13	14	30.2	30.2	11	14	16	16	17	17	23	25	10	12	8	10	8	8
CH17	X	Y	12	14	32.2	34.2	14	23	15	15	16	18	20	26	9	12	8	12	11	12
CH23	X	Y	12	13	29	31	16	16	15	15	14	17	24	24	11	12	9	12	10	12
CH42	X	Y	12	14	28.2	31.2	13	15	16	16	17	17	21	22	12	12	10	12	11	11
F18	X	Y	14	14	31	32.2	19	19	15	15	14	16	21	23	13	13	11	12	8	12
F21	X	Y	14	15	27	31	14	19	15	15	17	18	22	26	12	13	11	13	8	8
J13	X	Y	14	14	32.2	33.2	17	19	15	17	14	16	20	22	10	11	8	11	10	10
J3	X	Y	13	14	30	30	13	15	16	16	17	17	22	24	10	14	8	11	10	10
J37	X	Y	10	13	31	31.2	13	20	15	15	16	18	22	25	10	12	8	11	10	10
J39	X	Y	13	14	30	31.2	13	13	15	18	17	17	23	24	11	13	12	13	8	12
JK2921	X	Y	10	10	29	31.2	14	15	15	16	14	19w	19	22	10	11	11	11	9	9
JK2979	X	Y	10	12	32.2	33.2	14	14	15	16	14	18	21	24.2	11	14	11	12	11	11
MDK204	X	Y	14	17	28	30.2	12	19	14	15	16	16	23	25	12	13	11	12	8	10
Mel12	X	Y	13	14	30	32.2	17	20	15	16	17	18	21	22	13, 14,15		11	12	12	12
Mel15	X	Y	14	15	28	31.2	14,15	22	16	18	16	18	23	24	10	10	12	13	11	11
Mel18	X	Y	15	15	28	32.2	14	15	15	16	17	18	24	24	10	11	12	13	10	11
NG83	X	Y	13	13	30	31	13	18	15	16	15	17	23	24	11	11	8	8	8	10
NG85	X	Y	15	15	29	31.2	15	21	16	17	17	19	23	26	13	13	8	8	11	12
NG88	X	Y	12	15	33.2	38.2	13	17	15	15	16	16	25	28.1	11	11	8	11	8	10
OM135	X	Y	12	13	28	39	15	16	16	16	16	18	21	25	12	12	12	12	11	11
P103G	X	X	12	17	29	30	19	20	17	17	15	19	20	24	13	13	10	12	8	8
P109	X	Y	11	14	28	34	17	21	15	16	15	19	22	27	11	12	12	12	9	10
P205	X	Y	11	15	28	36	15	15	14	15	15	16	21	24	8	10	12	12	9	10
P240	X	Y	14	14	30.3	34	16	16	16	17	15	21	22	22.2	10	11	8	13	11	12
P33	X	Y	13	14	29	30	16.2	18	14	17	16	17	20	24	8	10	11	12	10	10
P37G	X	Y	14	15	29	32.2	14	20	16	17	17	18	24	24	12	13	9	12	8	11
P37G?	X	Y	14	15	29	32.2	14	20	16	17	17	18	24	24	12	13	9	12	8	11
P73	X	Y	14	15	29	30	14	16	15	16	14	19	22.2	24	12	12	11	13	10	11
PG1162	X	Y	8	14	31.2	32.2	14	15	15	16	16	18	18	23	11	12	9	11	8	10
PKH062	X	Y	12	15	29	31	13	16	16	18	16	18	21	22	11	11	9	13	10	12
SDH053	X	Y	10	12	30	33.2	16	16	18	18	17	17	21	25	11	12	8	8	8	8

Exhibit 39. **CEPH family pedigrees for samples examined as part of this study**. The 96well 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 spec primers.

CEPH Pedigree 104 (Venezuelan)



CEPH Pedigree 13293 (Utah)



CEPH Pedigree 884 (Amish)

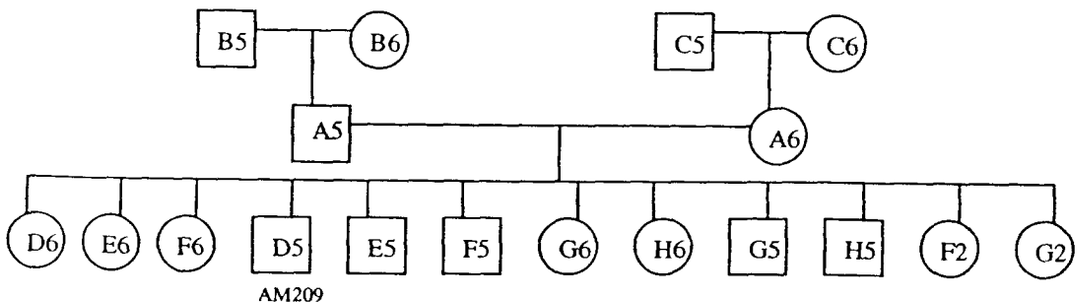


Exhibit 40. CAL DOJ CSF1PO results compared between ABI 310 and mass spec methods.

Position	ABI 310	Mass Spec	Allele 1	Allele 2	Position	ABI 310	Mass Spec	Allele 1	Allele 2
A1	11,12	11,12	27823	29060	A7	11,12	11,12	27867	29000
B1	10,12	10,12	26520	29033	B7	7,13	7,13	22809	30271
C1	7,10	7,10	22917	26684	C7	10,13	10,13	26529	30253
D1	10,13	10,13	26586	30344	D7	10,11	10,11	26621	27794
E1	8,12	8,12	24077	29075	E7	10,11	10,11	26609	27834
F1	8,10	8,10	24107	26639	F7	7,12	7,12	22923	29202
G1	12,13	12,13	29144	30373	G7	11,12	11,12	27958	29244
H1	11,12	11,12	27812	29018	H7	10,12	10,12	26497	29005
A2	12,12	12,12	29033		A8	12,12	12,12	29129	
B2	10,10	10,10	26594		B8	8,11	8,11	24117	27862
C2	10,10	10,10	26550		C8	7,12	7,12	22913	29071
D2	10,11	10,11	26609	27823	D8	10,10	10,10	26575	
E2	10,12	10,12	26620	29129	E8	10,12	10,12	26552	29007
F2	10,10	10,10	26584		F8	8,10	8,10	24061	26572
G2	11,12	11,12	27862	29080	G8	12,12	12,12	29064	
H2	10,10	10,10	26605		H8	11,12	11,12	27796	29026
A3	12,12	12,12	28987		A9	8,12	8,12	24085	29091
B3	10,13	10,13	26614	30380	B9	11,12	11,12	27819	29022
C3	10,10	10,10	26618		C9	11,12	11,12	27810	29000
D3	11,11	11,11	27823		D9	8,13	8,13	24093	30348
E3	11,12	11,12	27817	28998	E9	11,12	11,12	27878	29098
F3	11,12	11,12	27871	29093	F9	7,12	7,12	22836	29104
G3	10,11	10,11	26616	27812	G9	9,13	9,13	25387	30389
H3	11,12	11,12	27849	29024	H9	7,11	7,11	22866	27836
A4	8,10	8,10	24048	26499	A10	9,10	9,10	25391	26607
B4	11,13	11,13	27941	30412	B10	11,12	11,12	27844	29044
C4	11,11	11,11	27873		C10	10,10	10,10	26586	
D4	10,11	10,11	26638	27794	D10	8,11	8,11	24091	27838
E4	7,10	7,10	22835	26553	E10	11,12	11,12	27843	29040
F4	10,14	10,14	26592	31565	F10	10,12	10,12	26607	29111
G4	12,14	12,14	29113	31625	G10	8,11	8,11	24029	27807
H4	9,10	9,10	25360	26582	H10	8,10	8,10	24099	26609
A5	10,10	10,10	26590		A11	12,12	12,12	29011	
B5	7,11	7,11	22818	27838	B11	7,12	7,12	22836	29095
C5	10,10	10,10	26628		C11	7,11	7,11	22826	27834
D5	10,11	10,11	26647	27925	D11	10,11	10,11	26567	27784
E5	11,12	11,12	27847	29031	E11	10,12	10,12	26577	29078
F5	10,10	10,10	26601		F11	10,11	10,11	26611	27825
G5	10,12	10,12	26603	29113	G11	10,10	10,10	26605	
H5	11,13	11,13	27856	30389	H11	10,12	10,12	26584	29078
A6	11,11	11,11	27832						
B6	no data	10,12	26654	29180					
C6	11,11	11,11	27862						
D6	11,12	11,12	27818	29079					
E6	11,12	11,12	27838	29049					
F6	8,10	8,10	24206	26739					
G6	11,12	11,12	27843	29062					
H6	11,11	11,11	27773						

Complete agreement between all observed genotypes

Exhibit 41. CAL DOJ TPOX results compared between ABI 310 and mass spec methods. Allele masses are in Daltons.

Position	ABI 310	Mass Spec	Allele 1	Allele 2	Position	ABI 310	Mass Spec	Allele 1	Allele 2
A1	8,10	8,10	17914	20431	A7	9,9	**5,5**	gas-phase dimer	
B1	6,8	6,8	15376	17897	B7	8,8	8,8	17820	
C1	9,11	9,11	19137	21663	C7	10,12	10,12	20361	22830
D1	11,12	11,12	21562	22801	D7	11,12	**6,6**	gas-phase dimer	
E1	9,10	9,10	19131	20374	E7	6,8	6,8	15366	17850
F1	9,12	9,12	19106	22889	F7	10,10	10,10	20372	
G1	8,8	8,8	17838		G7	8,9	8,9	17820	19070
H1	8,8	8,8	17825		H7	9,9	9,9	19098	
A2	8,8	8,8	17837		A8	9,10	**5,5**	gas-phase dimer	
B2	8,8	8,8	17841		B8	9,9	**5,5**	gas-phase dimer	
C2	8,9	8,9	17860	19101	C8	8,10	8,10	17848	20366
D2	8,9	8,9	17785	19049	D8	8,11	8,11	17843	21596
E2	8,11	8,11	17829	21577	E8	9,9	9,9	19117	
F2	8,12	8,12	17834	22816	F8	9,12	9,12	19108	22793
G2	7,10	7,10	16525	20257	G8	8,8	8,8	17905	
H2	8,9	8,9	17870	19171	H8	8,10	8,10	17850	20366
A3	8,10	8,10	17844	20366	A9	8,11	8,11	17820	21563
B3	9,10	9,10	19063	20320	B9	11,11	11,11	21575	
C3	9,11	9,11	19178	21676	C9	8,8	8,8	17841	
D3	7,10	7,10	16626	20370	D9	10,11	10,11	20379	21600
E3	9,11	9,11	19146	21665	E9	8,11	8,11	17829	21569
F3	8,10	8,10	17839	20357	F9	7,8	7,8	16607	17837
G3	8,8	8,8	17853		G9	8,10	8,10	17904	20381
H3	6,9	6,9	15313	19070	H9	8,11	**5,5**	gas-phase dimer	
A4	6,8	6,8	15359	17895	A10	9,10	9,10	19076	20342
B4	11,11	11,11	21579		B10	8,11	8,11	17827	21569
C4	8,9	8,9	17865	19126	C10	8,12	8,12	17855	22850
D4	8,9	8,9	17839	19088	D10	8,9	8,9	17846	19093
E4	6,9	6,9	15289	19070	E10	9,10	9,10	19086	20346
F4	9,11	**5,5**	gas-phase dimer		F10	9,12	9,12	19065	22801
G4	8,8	8,8	17827		G10	11,12	11,12	21554	22795
H4	8,8	8,8	17798		H10	8,11	8,11	17841	21584
A5	9,11	9,11	19126	21663	A11	6,9	6,9	15299	19076
B5	8,11	8,11	17846	21586	B11	8,9	8,9	17851	19095
C5	10,11	10,11	20264	21550	C11	9,11	9,11	19120	21592
D5	9,12	9,12	19137	22905	D11	8,8	8,8	17829	
E5	9,10	9,10	19076	20337	E11	11,11	11,11	21569	
F5	9,11	9,11	19081	21571	F11	11,11	11,11	21548	
G5	8,8	8,8	17831		G11	11,11	11,11	21579	
H5	6,11	**5,5**	gas-phase dimer		H11	8,9	8,9	17870	19095
A6	8,11	**5,5**	gas-phase dimer						
B6	no data	11,11	21573						
C6	8,8	8,8	17844						
D6	8,12	**5,10**	gas-phase dimer						
E6	8,12	8,12	17831	22816					
F6	8,11	8,11	17846	21600					
G6	9,11	9,11	19070	21565					
H6	9,10	9,10	19151	20368					

SUMMARY	Allele	Number	Ave. mass	Std. Dev
	6	5	15343	34.3
	7	3	16586	53.7
	8	43	17846	26.5
	9	31	19102	33.1
	10	18	20353	40.7
	11	24	21590	37.9
	12	10	22830	39.7

Exhibit 42. CAL DOJ TH01 STR results compared between ABI 310 and mass spec methods. The shaded samples were run with a different primer set and thus have different masses.

Position	ABI 310	Mass Spec	Allele 1 (Da)	Allele 2 (Da)	Position	ABI 310	Mass Spec	Allele 1 (Da)	Allele 2 (Da)
A1	7,7	7,7	18854		A7	7,8	7,8	18877	20104
B1	7,9.3	7,9.3	17404	20854	B7	8,9.3	8,9.3	20128	22337
C1	7,8	7,8	18894	20108	C7	8,9	8,9	20193	21415
D1	6,9	6,9	17574	21294	D7	6,9.3	6,9.3	16167	20876
E1	7,9.3	7,9.3	18938	22387	E7	9.3,9.3	9.3,9.3	22349	
F1	9,9	9,9	21415		F7	9,9.3	9,9.3	21417	22345
G1	7,9	7,9	18919	21408	G7	7,8	7,8	18911	20147
H1	5,7	5,7	16403	18926	H7	9,9.3	9,9.3	21400	22277
A2	6,10	6,10	16145	21170	A8	8,8	8,8	20091	
B2	6,6	6,6	17662		B8	7,9.3	7,9.3	18938	22391
C2	7,8	7,8	18895	20115	C8	7,9	7,9	17436	19957
D2	7,9	7,9	18838	21309	D8	7,7	7,7	18944	
E2	7,9.3	7,9.3	18904	22329	E8	7,7	7,7	17376	
F2	7,9	7,9	18831	21302	F8	7,9	7,9	17387	19891
G2	7,7	7,7	17391		G8	9.3,9.3	9.3,9.3	22263	
H2	9,9.3	9,9.3	19871	20685	H8	7,9.3	7,9.3	18829	22267
A3	7,7	7,7	18856		A9	9,9	9,9	19907	
B3	7,8	7,8	18917	20114	B9	7,7	7,7	18938	
C3	7,9.3	7,9.3	17380	20828	C9	9.3,9.3	9.3,9.3	22351	
D3	7,9	7,9	17427	19944	D9	7,8	7,8	18845	20097
E3	7,9.3	7,9.3	18922	22364	E9	8,9.3	8,9.3	18684	20790
F3	9,9.3	9,9.3	21377	22321	F9	6,9.3	6,9.3	17622	22287
G3	8,9.3	8,9.3	20193	22384	G9	7,9	7,9	17380	19885
H3	7,9.3	7,9.3	18894	22292	H9	6,7	6,7	17677	18926
A4	8,8	8,8	20126		A10	7,8	7,8	17389	18652
B4	6,9.3	6,9.3	17606	22294	B10	9,9	9,9	21377	
C4	6,7	6,7	17586	18838	C10	7,9	7,9	17294	19812
D4	7,9.3	7,9.3	18897	22312	D10	7,8	7,8	17277	18547
E4	7,9	7,9	17401	19917	E10	6,7	6,7	17570	18819
F4	7,9	7,9	18836	21334	F10	8,8	8,8	18640	
G4	8,9.3	8,9.3	20077	22258	G10	6,7	6,7	17644	18906
H4	7,9.3	7,9.3	18899	22304	H10	8,9	8,9	18643	19902
A5	7,7	7,7	17205		A11	6,7	6,7	17540	18782
B5	6,9	6,9	16121	19874	B11	6,8	6,8	17682	20191
C5	8,9.3	8,9.3	20126	22360	C11	7,7	7,7	18927	
D5	6,8	6,8	17679	20184	D11	7,7	7,7	17477	
E5	8,10	8,10	20155	22658	E11	6,6	6,6	17596	
F5	7,9.3	7,9.3	18870	22285	F11	7,8	7,8	17400	18656
G5	6,7	6,7	17624	18906	G11	8,8	8,8	18710	
H5	7,9	7,9	17404	19921	H11	6,7	6,7	17600	18895
A6	6,8	6,8	17648	20139					
B6	7,7	7,7	18910						
C6	8,9.3	8,9.3	18626	20823					
D6	8,8	8,8	20147						
E6	8,9.3	8,9.3	19961	22143					
F6	8,8	8,8	20195						
G6	7,9	7,9	18858	21383					
H6	8,9	8,9	18704	19969					

SUMMARY

All 88 samples in agreement

Exhibit 43. CAL DOJ amelogenin results compared between ABI 310 and mass spec methods.

Position	ABI 310	Mass Spec	Allele 1 (Da)	Allele 2 (Da)	Position	ABI 310	Mass Spec	Allele 1 (Da)	Allele 2 (Da)
A1	X,Y	X,Y	25676	27566	A7	X,Y	X,Y	25774	27687
B1	X,Y	X,Y	25730	27596	B7	X,Y	X,Y	25659	27555
C1	X,Y	X,Y	25693	27667	C7	X,Y	X,Y	25797	27695
D1	X,Y	X,Y	25686	27607	D7	X,Y	X,Y	25641	27533
E1	X,Y	X,Y	25684	27765	E7	X,Y	X,Y	25674	27583
F1	X,Y	X,Y	25693	27592	F7	X,Y	X,Y	25676	27585
G1	X,Y	X,Y	25674	27574	G7	X,Y	X,Y	25668	27568
H1	X,Y	X,Y	25653	27551	H7	X,Y	X,Y	25666	27564
A2	X,Y	X,Y	25676	27574	A8	X,Y	X,Y	25659	27555
B2	X,Y	X,Y	25674	27579	B8	X,Y	X,Y	25653	27544
C2	X,Y	X,Y	25666	27559	C8	X,Y	X,Y	25668	27566
D2	X,Y	X,Y	25768	27626	D8	X,Y	X,Y	25674	27566
E2	X,Y	X,Y	25670	27568	E8	X,Y	X,Y	25843	27763
F2	X,Y	X,Y	25657	27544	F8	X,Y	X,Y	25649	27581
G2	X,Y	X,Y	25672	27646	G8	X,Y	X,Y	25661	27561
H2	X,Y	X,Y	25641	27523	H8	X,Y	X,Y	25757	27585
A3	X,Y	X,Y	25663	27559	A9	X,Y	X,Y	25641	27531
B3	X,Y	X,Y	25770	27581	B9	X,Y	X,Y	25684	27579
C3	X,Y	X,Y	25691	27635	C9	X,X	X,X	25645	
D3	X,Y	X,Y	25722	27594	D9	X,Y	X,Y	25653	27553
E3	X,Y	X,Y	25749	27628	E9	X,Y	X,Y	25674	27594
F3	X,Y	X,Y	25682	27577	F9	X,Y	X,Y	25720	27646
G3	X,Y	X,Y	25645	27536	G9	X,Y	X,Y	25766	27637
H3	X,Y	no data			H9	X,Y	no data		
A4	X,Y	X,Y	25666	27553	A10	X,Y	X,Y	25697	27618
B4	X,Y	X,Y	25726	27637	B10	X,Y	X,Y	25684	27590
C4	X,Y	X,Y	25736	27583	C10	X,Y	no data		
D4	X,Y	X,Y	25782	27670	D10	X,Y	X,Y	25657	27564
E4	X,Y	X,Y	25701	27789	E10	X,Y	X,Y	25657	27549
F4	X,Y	X,Y	25674	27585	F10	X,Y	X,Y	25653	27551
G4	X,Y	X,Y	25599	27572	G10	X,Y	no data		
H4	X,Y	X,Y	25682	27579	H10	X,Y	X,Y	25722	27654
A5	X,Y	X,Y	25816	27719	A11	X,Y	X,Y	25676	27572
B5	X,Y	X,Y	25672	27570	B11	X,Y	X,Y	25655	27549
C5	X,Y	X,Y	25680	27583	C11	X,Y	X,Y	25666	27577
D5	X,Y	X,Y	25688	27661	D11	X,Y	X,Y	25618	27464
E5	X,Y	X,Y	25643	27523	E11	X,Y	X,Y	25697	27676
F5	X,Y	X,Y	25666	27572	F11	X,Y	X,Y	25653	27557
G5	X,Y	X,Y	25778	27672	G11	X,Y	X,Y	25661	27553
H5	X,Y	X,Y	25738	27568	H11	X,Y	X,Y	25643	27540
A6	X,Y	X,Y	25695	27587					
B6	X,Y	X,Y	25649	27523					
C6	X,Y	X,Y	25655	27551					
D6	X,Y	no data							
E6	X,Y	X,Y	25720	27605					
F6	X,Y	X,Y	25674	27568					
G6	X,Y	X,Y	25881	27745					
H6	X,Y	no data							

SUMMARY	X allele	Y allele
Count	82	81
Average mass	25690 Da	27594 Da
Std dev	50.4 Da	59.6 Da

82 samples agreed
6 "no data" by mass spec

Exhibit 44. CAL DOJ D3S1358 STR results compared between ABI 310 and mass spec methods.

Position	ABI 310	Mass Spec	Allele 1 (Da)	Allele 2 (Da)	Position	ABI 310	Mass Spec	Allele 1 (Da)	Allele 2 (Da)
A1	17,17	17,17	28182		A7	16,17	16,17	27027	28112
B1	16,16	16,16	26967		B7	14,19	14,19	24305	30503
C1	15,17	15,17	25668	28139	C7	14,14	14,14	24254	
D1	12,17	12,17	21820	27991	D7	15,16	15,16	25524	26713
E1	15,15	15,15	25741		E7	16,16	16,16	26822	
F1	16,16	16,16	26762		F7	18,18	18,18	29491	
G1	14,15	14,15	24342	25501	G7	15,16	15,16	25772	26830
H1	15,16	15,16	25802	26941	H7	16,17	16,17	27018	28273
A2	15,17	15,17	25638	28134	A8	14,16	14,16	24331	26856
B2	16,17	16,17	26795	27889	B8	16,17	16,17	26843	28056
C2	15,16	15,16	25661	26869	C8	15,16	15,16	25638	26766
D2	13,16	13,16	23171	27091	D8	15,15	15,15	25724	
E2	14,17	14,17	24352	28126	E8	14,15	14,15	24343	25477
F2	15,17	15,17	25549	28091	F8	15,16	15,16	25662	26703
G2	15,17	15,17	25641	28119	G8	16,16	16,16	26907	
H2	14,15	14,15	24311	25507	H8	15,15	15,15	25618	
A3	17,17	17,17	28184		A9	14,16	14,16	24323	26832
B3	18,18	18,18	29384		B9	15,16	15,16	25537	26739
C3	15,15	15,15	25645		C9	15,15	15,15	25499	
D3	15,15	15,15	25645		D9	15,15	15,15	25522	
E3	16,17	16,17	26967	27945	E9	14,15	14,15	24440	25459
F3	15,16	15,16	25573	26790	F9	16,17	16,17	26745	28021
G3	17,18	17,18	28025	29302	G9	17,18	17,18	28372	29315
H3	14,15	14,15	24360	25530	H9	15,16	15,16	25761	26777
A4	17,17	17,17	28001		A10	16,16	16,16	26913	
B4	15,16	15,16	25518	26811	B10	15,18	15,18	25539	29347
C4	15,16	15,16	25705	26722	C10	15,16	15,16	25477	26741
D4	15,17	15,17	25589	28110	D10	14,15	14,15	24214	25341
E4	16,17	16,17	27035	28023	E10	14,15	14,15	24252	25507
F4	14,16	14,16	24374	26886	F10	16,17	16,17	26871	28139
G4	14,17	14,17	24342	28115	G10	16,17	16,17	27033	27934
H4	15,18	15,18	25655	29378	H10	15,16	15,16	25716	26715
A5	15,16	15,16	25727	27110	A11	15,18	15,18	25545	29293
B5	16,17	16,17	26988	28300	B11	15,15.2	15,15.2	25734	26379
C5	15,15	15,15	25684		C11	16,17	16,17	26937	27845
D5	14,16	14,16	24520	26941	D11	15,16	15,16	25544	26672
E5	15,17	15,17	25678	28132	E11	16,16	16,16	26899	
F5	16,16	16,16	26813		F11	15,16	15,16	25706	26678
G5	9,17	**9,17**	18092	28285	G11	16,16	16,16	26841	
H5	17,17	17,17	27893		H11	16,17	16,17	26881	28246
A6	16,17	16,17	26999	28178					
B6	15,17	15,17	25645	28182					
C6	15,16	15,16	25681	26710					
D6	15,15	15,15	25833						
E6	14,16	14,16	24301	26816					
F6	14,14	14,14	24437						
G6	14,15	14,15	24425	25524					
H6	15,15	15,15	25523						

SUMMARY

All 88 samples in agreement

Exhibit 45. CAL DOJ D16S539 STR results compared between ABI 310 and mass spec methods.

Position	ABI 310	Mass Spec	Allele 1 (Da)	Allele 2 (Da)	Position	ABI 310	Mass Spec	Allele 1 (Da)	Allele 2 (Da)
A1	11,12	11,12	26862	28095	A7	11,11	11,11	26873	
B1	9,14	9,14	24421	30665	B7	9,10	9,10	24338	25595
C1	9,11	9,11	24360	26888	C7	11,13	11,13	26894	29471
D1	10,11	10,11	25622	26831	D7	11,11	11,11	26879	
E1	11,13	11,13	26903	29367	E7	9,12	9,12	24425	28200
F1	10,11	10,11	25543	26830	F7	9,12	9,12	24368	28160
G1	11,12	11,12	26852	28070	G7	11,12	11,12	26997	28145
H1	8,9	8,9	23157	24360	H7	12,12	12,12	28104	
A2	9,11	9,11	24342	26805	A8	9,13	9,13	24360	29362
B2	12,13	12,13	28191	29323	B8	9,9	9,9	24394	
C2	11,11	11,11	26907		C8	9,13	9,13	24352	29369
D2	13,13	13,13	29445		D8	11,11	11,11	26933	
E2	12,12	12,12	28128		E8	8,12	8,12	23332	28486
F2	12,13	12,13	28202	29436	F8	11,12	11,12	26937	28128
G2	10,11	10,11	25638	26869	G8	9,13	9,13	24336	29351
H2	10,11	10,11	25576	26864	H8	11,14	11,14	26864	30573
A3	9,12	9,12	24340	28106	A9	11,13	11,13	26869	29378
B3	9,12	9,12	24356	28117	B9	11,12	11,12	26888	28078
C3	12,13	12,13	28141	29483	C9	12,12	12,12	28106	
D3	9,9	9,9	24417		D9	11,13	11,13	26849	29342
E3	11,13	11,13	27014	29603	E9	11,13	11,13	26899	29396
F3	11,11	11,11	26899		F9	12,15	12,15	28115	31845
G3	9,11	9,11	24376	26931	G9	11,13	11,13	26886	29389
H3	8,10	8,10	23161	25649	H9	11,13	11,13	26888	29387
A4	9,11	9,11	24427	26935	A10	9,11	9,11	24423	26920
B4	9,12	9,12	24480	28248	B10	9,12	9,12	24449	28126
C4	13,13	13,13	29318		C10	12,13	12,13	28302	29498
D4	11,12	no data			D10	8,9	8,9	23072	24317
E4	13,13	13,13	29389		E10	9,13	9,13	24376	29402
F4	9,11	9,11	24374	26894	F10	11,13	11,13	26856	29436
G4	13,14	no data			G10	9,9	9,9	24358	
H4	11,13	11,13	26837	29336	H10	10,13	10,13	25653	29441
A5	12,12	12,12	28117		A11	9,12	9,12	24427	28128
B5	8,9	8,9	23181	24417	B11	11,12	11,12	26862	28043
C5	10,11	10,11	25670	27007	C11	11,12	no data		
D5	12,12	12,12	28265		D11	11,11	11,11	26890	
E5	9,11	9,11	24482	26984	E11	11,13	11,13	26928	29619
F5	9,11	9,11	24329	26854	F11	9,10	no data		
G5	10,12	10,12	25688	28222	G11	10,12	no data		
H5	11,12	11,12	26918	28041	H11	11,11	11,11	26894	
A6	9,11	9,11	24372	26888					
B6	no data	9,12	24405	28208					
C6	9,13	9,13	24380	29398					
D6	9,13	9,13	24465	29416					
E6	12,12	12,12	28170						
F6	13,13	13,13	29329						
G6	11,13	11,13	26867	29371					
H6	11,11	**9,9**	gas-phase trimer						

SUMMARY

81 samples in agreement
5 "no data" from mass spec
1 "no data" from ABI 310
1 error due to gas-phase trimer

Exhibit 46. CAL DOJ D7S820 STR results compared between ABI 310 and mass spec methods.

Position	ABI 310	Mass Spec	Allele 1 (Da)	Allele 2 (Da)	Position	ABI 310	Mass Spec	Allele 1 (Da)	Allele 2 (Da)
A1	11,13	11,13	20322	22807	A7	10,12	10,12	19097	21621
B1	10,11	10,11	19081	20333	B7	11,12	11,12	20545	21592
C1	8,11	8,11	16641	20379	C7	**8**,10	**10,10**	19360	
D1	9,10	9,10	17857	19131	D7	11,12	11,12	20383	21646
E1	**8**,10	**10,10**	19072		E7	9,12	9,12	17841	21583
F1	10,12	10,12	19088	21594	F7	8,**9**	**8,8**	16631	
G1	8,10	8,10	16643	19174	G7	11,12	11,12	20374	21575
H1	8,10	8,10	16616	19088	H7	11,12	11,12	20366	21581
A2	8,11	8,11	16646	20448	A8	10,12	10,12	19189	21703
B2	10,11	10,11	19169	20381	B8	10,11	10,11	19265	20400
C2	8,8	8,8	16534		C8	10,10	10,10	19095	
D2	10,10	10,10	19189		D8	9,10	9,10	17848	19093
E2	8,8	8,8	16641		E8	**8**,10	**10,10**	19169	
F2	8,8	8,8	16616		F8	8,9	no data	*Double null alleles?*	
G2	**8**,10	**10,10**	19101		G8	9,12	9,12	17949	21745
H2	8,9	8,9	16550	17834	H8	**8**,10	**10,10**	19187	
A3	11,12	11,12	20372	21569	A9	**8**,11	**11,11**	20571	
B3	10,11	10,11	19104	20366	B9	8,8	8,8	16534	
C3	8,12	8,12	16532	21548	C9	10,10	10,10	19169	
D3	12,13	12,13	21560	22789	D9	7,8	7,8	15383	16638
E3	10,12	10,12	19187	21695	E9	9,10	9,10	17844	19090
F3	10,11	10,11	19198	20381	F9	10,12	10,12	19395	21866
G3	**8**,11	**11,11**	20374		G9	8,10	8,10	16651	19201
H3	**8**,10	**10,10**	19097		H9	8,11	8,11	16826	20363
A4	9,9	9,9	17848	17848	A10	12,13	12,13	21586	22820
B4	11,11	11,11	20528		B10	11,12	11,12	20387	21586
C4	8,12	8,12	16639	21667	C10	**8**,10	**10,10**	19330	
D4	8,9	8,9	16636	17860	D10	8,11	8,11	16554	20342
E4	**9**,11	**11,11**	20392		E10	10,11	10,11	19086	20348
F4	10,11	10,11	19185	20396	F10	9,10	9,10	17851	19097
G4	10,10	10,10	19101		G10	**9**,10	**10,10**	19398	
H4	**10**,12	**12,12**	21579		H10	10,11	10,11	19382	20377
A5	9,10	9,10	17836	19083	A11	10,10	10,10	19254	
B5	11,12	11,12	20631	21678	B11	10,11	10,11	19093	20353
C5	11,12	11,12	20363	21573	C11	9,10	9,10	17886	19187
D5	10,10	10,10	19088		D11	8,11	8,11	16643	20519
E5	9,11	9,11	17851	20372	E11	10,10	10,10	19252	
F5	8,11	8,11	16634	20377	F11	9,12	9,12	17831	21569
G5	11,12	11,12	20590	21787	G11	10,12	10,12	19142	21653
H5	10,10	10,10	19322		H11	10,12	10,12	19160	21883
A6	8,10	8,10	16556	19074					
B6	8,**9**	**8,8**	16643						
C6	**8**,11	**11,11**	20374						
D6	8,10	8,10	16614	19090					
E6	**8**,9	**9,9**	17851						
F6	8,8	8,8	16831						
G6	**8**,10	**10,10**	19155						
H6	9,9	9,9	18169						

SUMMARY

70 samples in agreement

1 "no data" from mass spec (null alleles?)

17 samples disagreed due to null alleles

Exhibit 47. CAL DOJ FGA STR results compared between ABI 310 and mass spec methods.

Position	ABI 310	Mass Spec	Allele 1 (Da)	Allele 2 (Da)	Position	ABI 310	Mass Spec	Allele 1 (Da)	Allele 2 (Da)
A1	20,21	no data			A7	23,24	**24,24**	40518	not resolved
B1	20,23	20,23	35987	39547	B7	22,25	22,25	38290	41811
C1	23,25	23,25	39390	41917	C7	24,25	**24,24**	41079	not resolved
D1	19,23	19,23	34858	39547	D7	22,22	**21,21**	37930	
E1	23,24	**23,23**	40317	not resolved	E7	24,27	24,27	40680	44130
F1	18.2,26	18.2,26	34066	43221	F7	22,22	22,22	38295	
G1	21,21	21,21	37011		G7	22,23	**22,22**	38582	not resolved
H1	no data	23,23	39555		H7	21,26	21,26	37148	43116
A2	22,23	**23,23**	39446	not resolved	A8	OL?,24	16,24	31415	40688
B2	21,22	**21,21**	37570	not resolved	B8	22,24	22,24	38226	40874
C2	21,25	21,25	37146	41944	C8	19.2,25	19.2,25	35193	41718
D2	19,19.2	**19,19**	34699	not resolved	D8	24,25	**24,24**	41184	not resolved
E2	23,27	**23,26**	39310	43958	E8	18.2,24	**18,23**	33960	40247
F2	19,25	19,25	34667	41575	F8	23,25	23,25	39482	41914
G2	22,22	22,22	38554		G8	25,25	**24,24**	41668	
H2	22,25	22,25	38244	41861	H8	21,23	21,23	37113	39632
A3	22,24	22,24	38110	40604	A9	19,25	19,25	34631	41827
B3	30.2,OL?	**30,30**	48114		B9	20,24	20,24	35888	40688
C3	21,24	21,24	37158	40777	C9	20,24	**19,23**	35667	40312
D3	23,23	23,23	39475		D9	21,24	21,24	36973	40573
E3	19,23	19,23	34769	39547	E9	21,25	**21,24**	36921	41549
F3	23,24	**23,23**	39852	not resolved	F9	25,26	**25,25**	42218	not resolved
G3	22,23	**22,22**	38582	not resolved	G9	20,21	**21,21**	36901	not resolved
H3	18.2,23	18.2,23	34138	39423	H9	18.2,25	18.2,25	34111	41774
A4	22,23	**22,22**	39056	not resolved	A10	24,28	24,28	40706	45467
B4	19,26	19,26	34578	42761	B10	21,24	21,24	37018	40636
C4	24,24	24,24	40649		C10	18,23	18,23	33560	39490
D4	26,27	no data			D10	22,23	**22,22**	38537	not resolved
E4	22,22	22,22	38442		E10	18,23	18,23	33292	39289
F4	22,23	**22,22**	39081	not resolved	F10	19,22	19,22	34829	38402
G4	23,24	**23,23**	39741	not resolved	G10	22,24	22,24	38216	40581
H4	23,23	23,23	39562		H10	20,26	20,26	36127	43407
A5	20,22	20,22	35856	38364	A11	24,26	24,26	40659	43043
B5	23,23	**22,22**	39017		B11	23,28	23,28	39493	45376
C5	22,23	**22,22**	38659	not resolved	C11	24,25	**24,24**	40869	not resolved
D5	23,26	**22,25**	39233	42879	D11	22,25	22,25	38361	41989
E5	22,23	**22,22**	38554	not resolved	E11	25,26	**25,25**	42333	not resolved
F5	19,25	19,25	34795	41994	F11	22,24	22,24	38209	40560
G5	21,24	21,24	37366	40853	G11	23,25	23,25	39511	41909
H5	21,22	**21,21**	37565	not resolved	H11	23,26	23,26	39408	42965
A6	22,24	22,24	38155	40570					
B6	24,27	24,27	40612	44288					
C6	21,25	21,25	37136	41997					
D6	20,23	20,23	35972	39477					
E6	22,23	**22,22**	38644	not resolved					
F6	25,28	**25,27**	41737	45299					
G6	18.2,24	18.2,24	34387	40869					
H6	21,23	21,23	37113	39534					

SUMMARY

53 samples in agreement
2 "no data" from mass spec
1 "no data" from ABI 310
1 too large for detection
9 wrong calls from poor calibration
22 wrong calls from poor resolution

Exhibit 49. **CAL DOJ DYS391 STR results with mass spec method.**

Position	Mass Spec	Allele 1 (Da)	Allele 2 (Da)	Position	Mass Spec	Allele 1 (Da)	Allele 2 (Da)
A1	10	24489		A7	10	24416	
B1	11	25672		B7	10	24406	
C1	10	24487		C7	10	24353	
D1	10	24455		D7	8	21905	
E1	10	24471		E7	11	25662	
F1	11	25641		F7	10	24353	
G1	11	25637		G7	10	24422	
H1	10	24436		H7	10	24451	
A2	10	24459		A8	10	24455	
B2	10	24465		B8	11	25529	
C2	10	24440		C8	10	24410	
D2	10	24455		D8	10	24473	
E2	10	24359		E8	12	26841	
F2	11	25837		F8	10	24438	
G2	10	24444		G8	10	24359	
H2	10	24451		H8	12	26805	
A3	11	25639		A9	10	24367	
B3	10	24414		B9	10	24416	
C3	11	25654		C9	10	24463	
D3	11	25591		D9	10	24343	
E3	10	24463		E9	10	24471	
F3	11	25648		F9	10	24457	
G3	10	24475		G9	10	24465	
H3	10	24436		H9	10	24479	
A4	10	24408		A10	11	25631	
B4	10	24475		B10	10	24617	
C4	11	25625		C10	10	24560	
D4	no data			D10	10	24444	
E4	11	25581		E10	11	25650	
F4	10	24396		F10	10	24414	
G4	11	25562		G10	10	24453	
H4	10	24463		H10	11	25866	
A5	10	24390		A11	11	25662	
B5	10	24446		B11	10	24432	
C5	10	24599		C11	10	24400	
D5	11	25652		D11	10	24463	
E5	10	24451		E11	10	24420	
F5	10	24436		F11	9	23175	
G5	10	24380		G11	10	24459	
H5	12	26775		H11	11	25585	
A6	11	25550					
B6	10	24428					
C6	11	25583					
D6	10	24457					
E6	11	25658					
F6	10	24436					
G6	10	24473					
H6	10	24463					

Exhibit 48. **CAL DOJ D8S1179 STR results with mass spec method.**

Position	Mass Spec	Allele 1 (Da)	Allele 2 (Da)	Position	Mass Spec	Allele 1 (Da)	Allele 2 (Da)
A1	12,15	24693	28228	A7	13,13	25791	
B1	14,15	27159	28369	B7	13,14	25845	26954
C1	11,14	23342	26933	C7	12,15	24540	28101
D1	15,15	28199		D7	13,14	25772	26931
E1	11,14	23421	26988	E7	14,14	26960	
F1	13,17	25822	30529	F7	13,16	25785	29381
G1	11,14	23449	27093	G7	14,14	27097	
H1	14,14	27067		H7	14,14	27039	
A2	13,13	25777		A8	13,15	25770	28204
B2	12,14	24597	27007	B8	11,13	23383	25816
C2	11,11	23361		C8	14,15	27033	28043
D2	11,13	23399	25768	D8	14,14	27076	
E2	13,14	25904	27012	E8	14,15	26984	28221
F2	14,14	27125		F8	13,14	25781	26946
G2	no data			G8	10,12	22183	24581
H2	11,15	23344	28151	H8	12,15	24587	28197
A3	15,15	28160		A9	15,16	28265	29390
B3	13,15	25904	28256	B9	13,14	25835	26963
C3	12,14	24640	27022	C9	11,14	23447	27084
D3	11,15	23407	28201	D9	11,14	23391	26988
E3	11,14	23461	27093	E9	11,14	23361	26931
F3	14,15	27086	28186	F9	14,15	27031	28225
G3	10,15	22218	28230	G9	14,14	26978	
H3	13,16	25849	29355	H9	11,13	23451	26008
A4	14,15	27018	28112	A10	13,14	25789	26841
B4	12,14	24605	27018	B10	15,15	28206	
C4	13,13	25797		C10	13,13	25791	
D4	no data			D10	14,15	26999	28123
E4	14,14	27108		E10	13,14	25895	26980
F4	14,15	27080	28215	F10	13,15	25777	28197
G4	12,13	24311	25519	G10	14,15	26733	27909
H4	13,14	25824	27044	H10	12,14	24688	27093
A5	12,16	24548	29321	A11	12,15	24574	28230
B5	13,14	25893	26997	B11	13,14	25822	27054
C5	12,13	24727	25885	C11	14,16	26963	29324
D5	no data			D11	13,14	25812	26950
E5	14,14	26960		E11	12,15	24617	28232
F5	13,14	25916	27014	F11	14,14	26982	
G5	14,15	27009	28106	G11	14,14	27005	
H5	13,14	25804	26708	H11	12,13	24631	25818
A6	14,14	26988					
B6	14,16	26982	29364				
C6	15,15	28249					
D6	14,14	26716					
E6	13,16	25933	29584				
F6	8,14	19884	27041				
G6	15,15	28394					
H6	14,14	27009					

Exhibit 50. **Mass spectra of CAL DOJ samples amplified with TPOX primers.** From left side top-to-bottom followed by right side top-to-bottom, sample genotypes are (8,9), (9,10), (10,11), (11,12), (8,8), (8,11), (11,11), and allelic ladder. The mass range shown here is 15,000-25,000 Da.

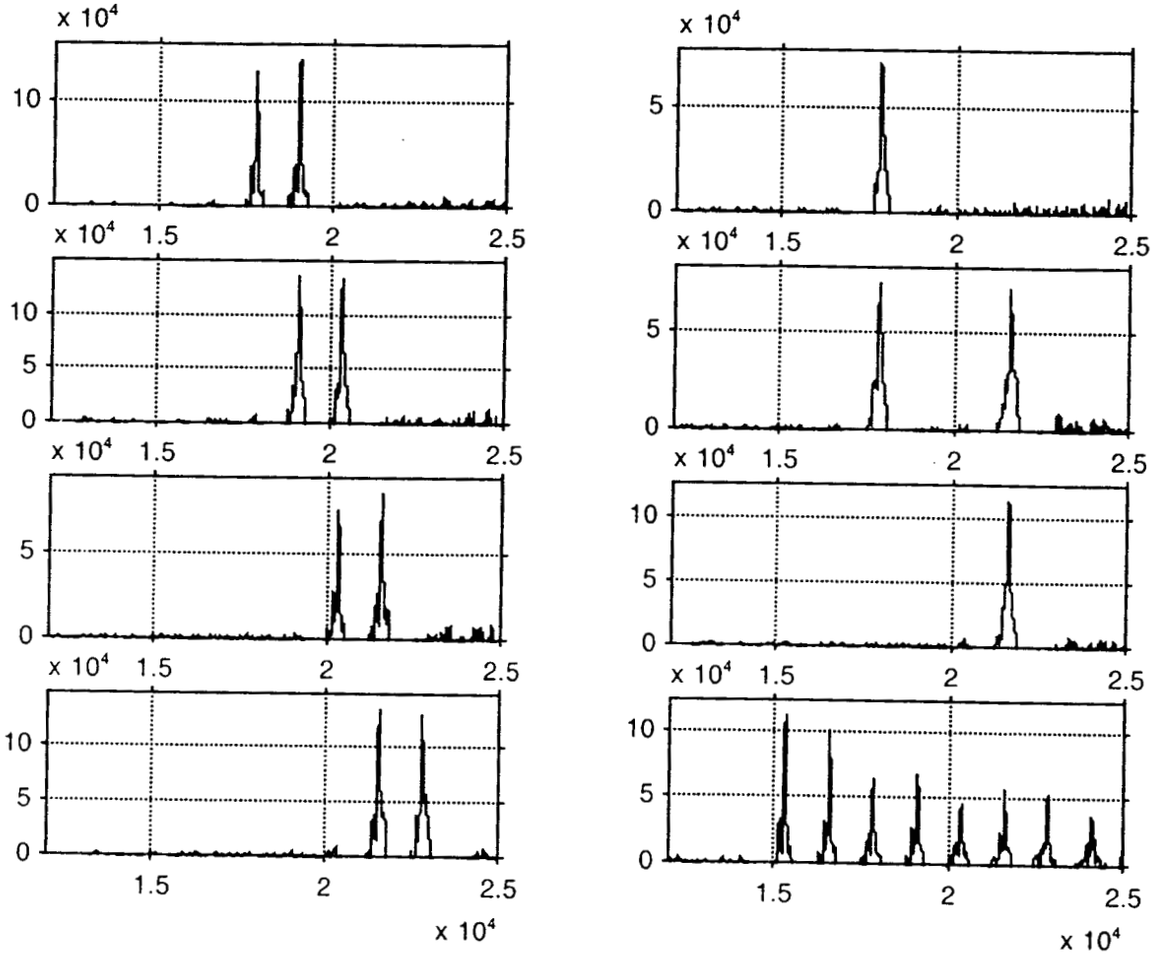


Exhibit 51. **Mass spectra of CAL DOJ samples amplified with CSF1PO primers**. From left side top-to-bottom followed by right side top-to-bottom, sample genotypes are (7,11), (8,10), (10,10), (9,10), (12,12), (11,12), (12,14), and allelic ladder. The mass range shown here is 15,000-35,000 Da.

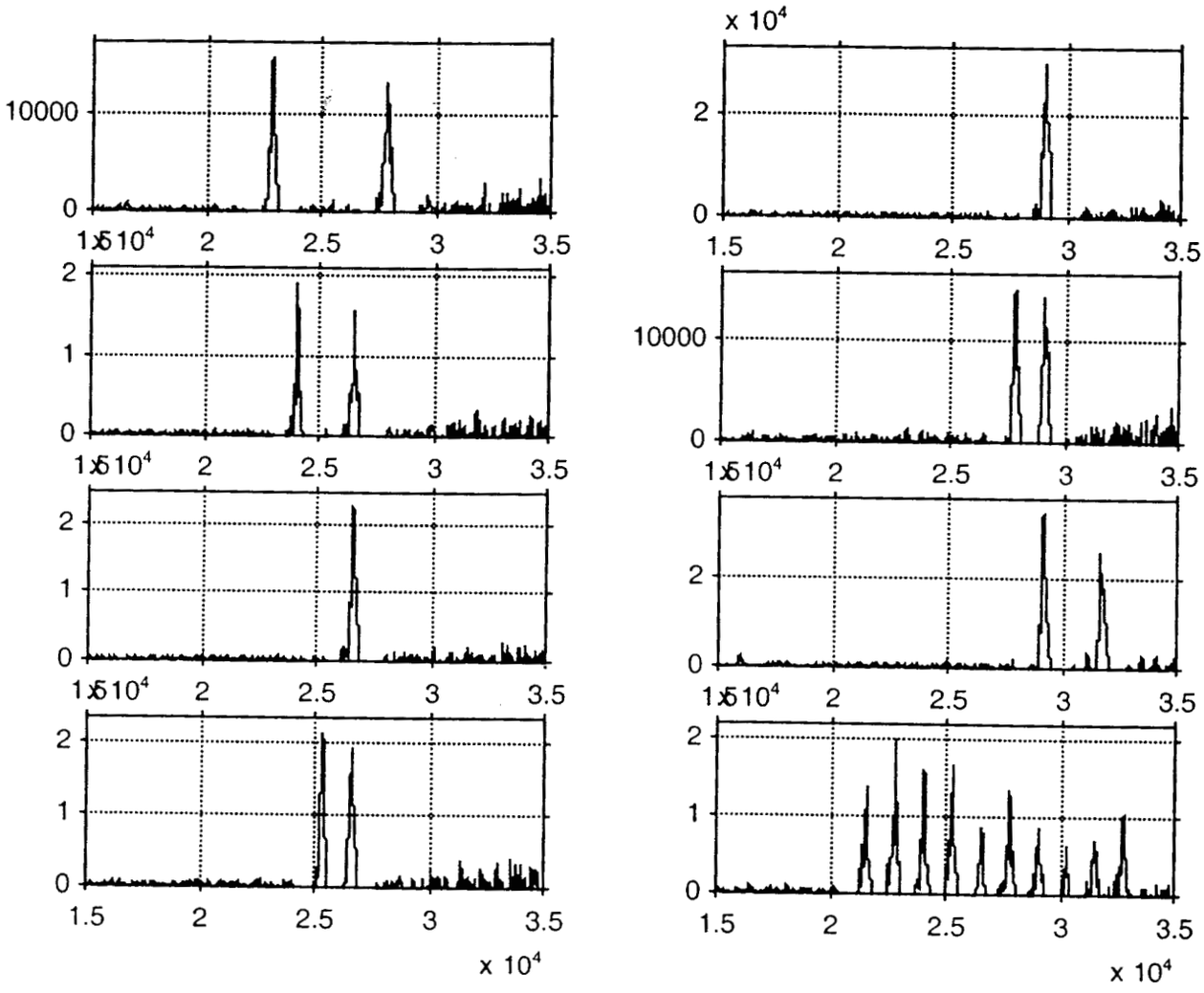


Exhibit 52. **Mass spectra of CAL DOJ samples amplified with TH01 primers**. From left side top-to-bottom followed by right side top-to-bottom, sample genotypes are (6,7), (6,8), (6,9.3), (6,10), (7,7), (7,7), (7,8), and (7,9.3). The split peaks for each allele result from partial adenylation—i.e., both ±A peaks are present. The mass range shown here is 12,000-22,000 Da.

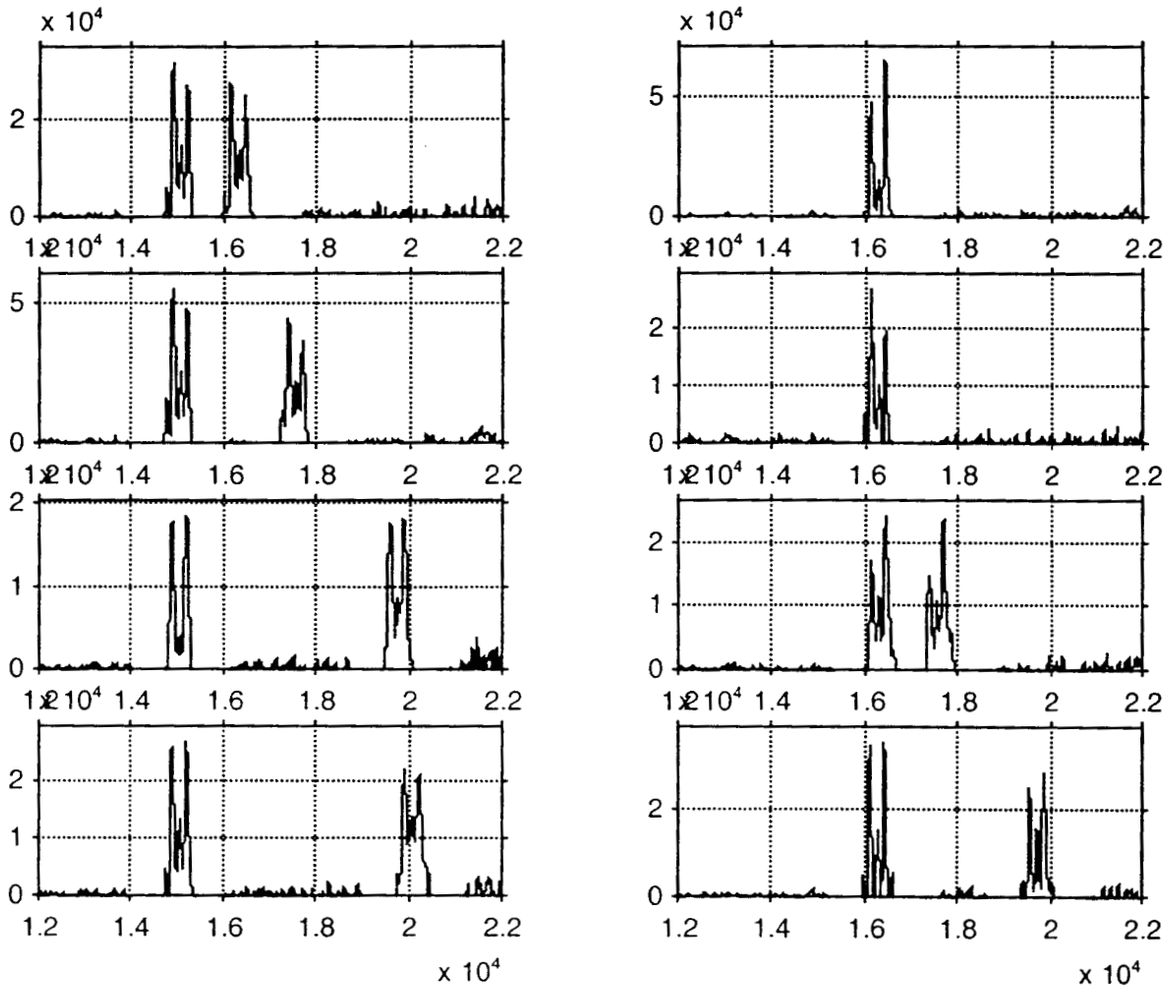


Exhibit 53. **Mass spectra of CAL DOJ samples amplified with D7S820 primers**. From left side top-to-bottom followed by right side top-to-bottom, sample genotypes are (10,11), (9,10), (10,10), (null allele 8, 10), (8,11), (11,13), (10,12), and (8,10). The arrow indicates the position where allele 8 should be present in the sample but is missing due to a primer annealing binding site sequence polymorphism that results in a "null allele" (see text for discussion). The mass range shown here is 12,000-25,000 Da.

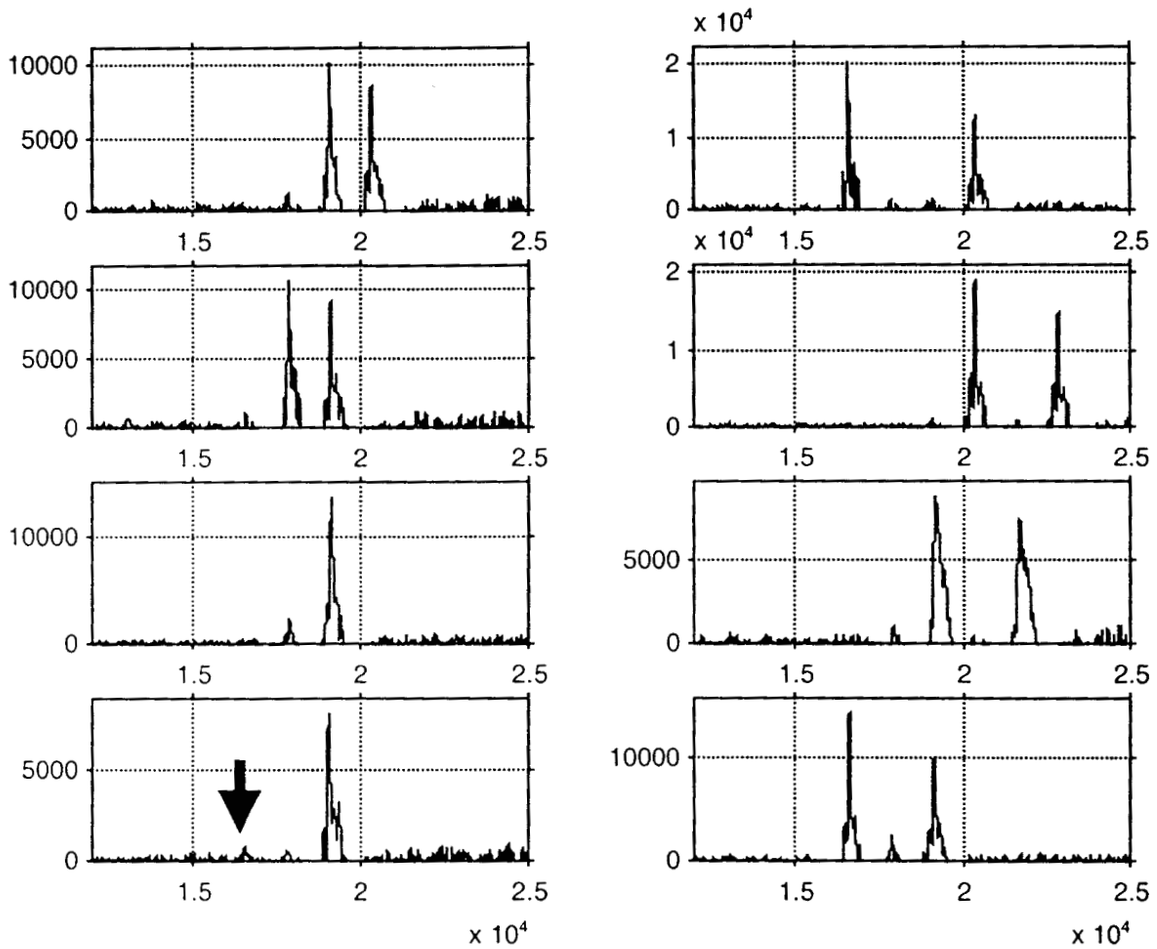


Exhibit 54. **Primer sequence determination with exonuclease digestion and mass difference measurements.** This example is a D5S818 primer pair purchased from Promega Corporation and used in their PowerPlex™ STR kit. The top panel shows a mass spectrum of the original primer pair prior to digestion. The bottom panel is the mass spectrum of the same primers following a 6 minute digestion at 37 °C with calf spleen phosphodiesterase which is a 5'→ 3' exonuclease. The dye labeled primer is not digested because the dye protects the 5'-end of the primer. Mass difference measurements between the digestion peaks leads to the sequence determination of the 5'-end of the unlabeled primer (see underlined portion of forward sequence). The determined sequences are 5'-GGTGATTTTCCTCTTTGGTATCC-3' (forward) and 5'-fluorescein dye—TTTACAACATTTGTATCTATATCTGT-3' (reverse).

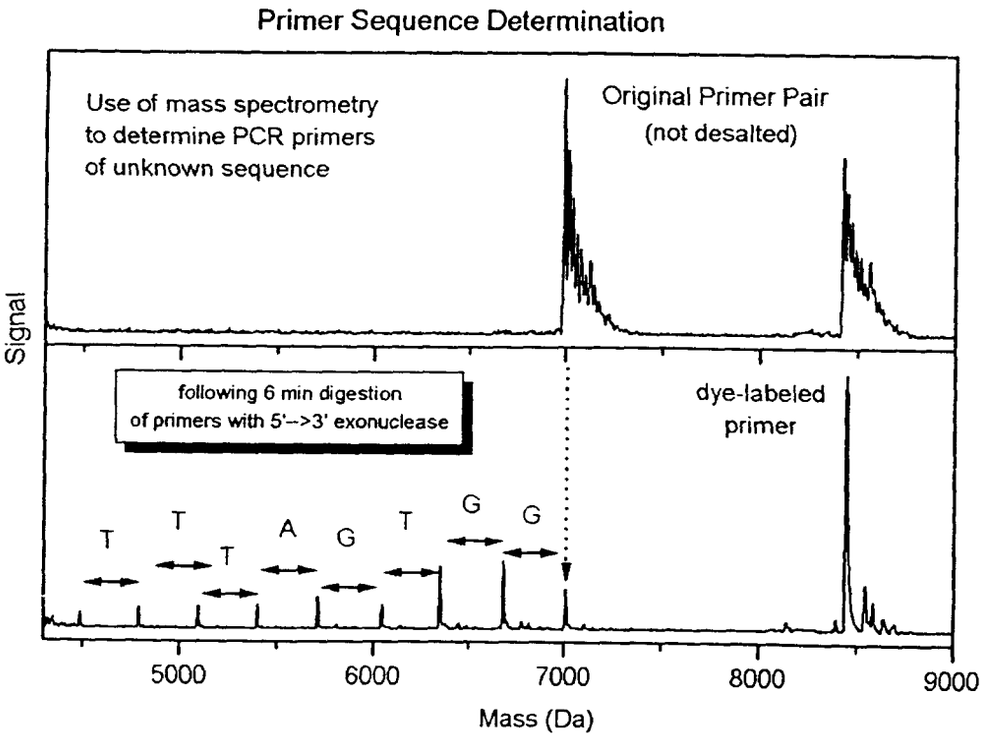


Exhibit 55. **Mass spectrum of AmpF*ℓ*STR® Green I primer mix**. Each peak has been identified to its corresponding primer. Peaks containing the fluorescent dye (JOE) are underlined.

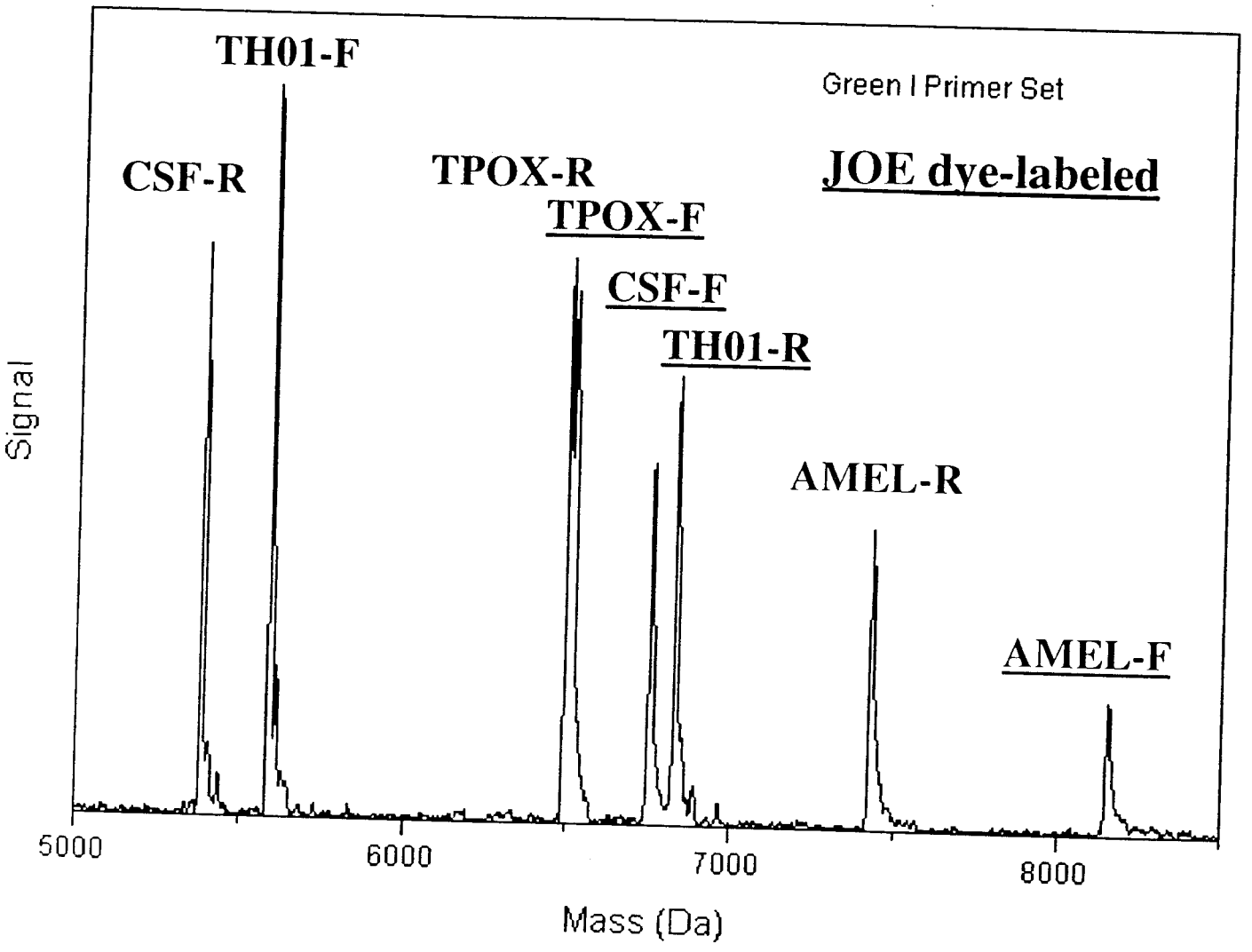
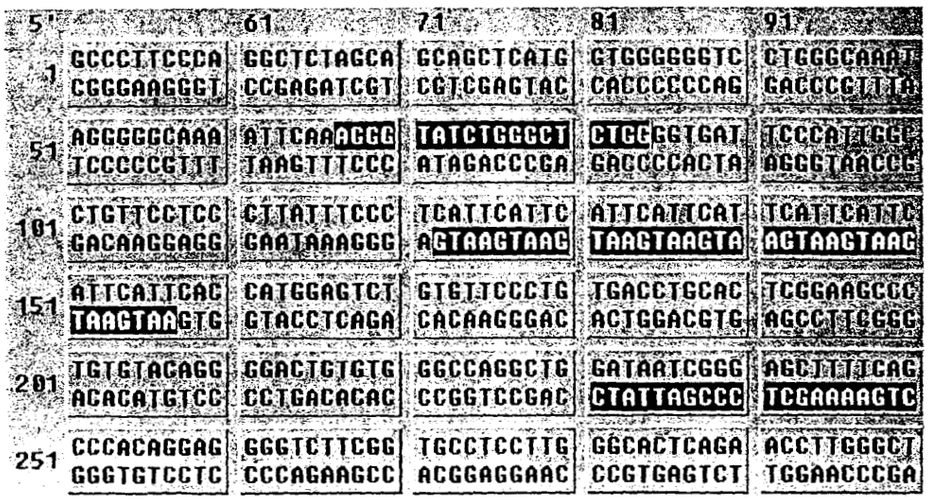


Exhibit 56. **TH01 STR primer positions for commercially available primers highlighted on the GenBank sequence.** The forward primer is shown in blue with the reverse primer in brown. The repeat is highlighted in green on the strand that contains the fluorescent dye. The 3'-positions of the forward primers are identical but differ by a single base for the reverse primers. Promega primers are longer at the 5'-end, which produces a larger PCR product by 11 bp (6 bases on forward and 5 bases on reverse).

AmpFlSTR® Green I Kit (PE Applied Biosystems)

Reverse primer is labeled with JOE dye (fluorescein derivative)



PCR product =184 bp (9 repeats)

PowerPlex™ Kit (Promega Corporation)

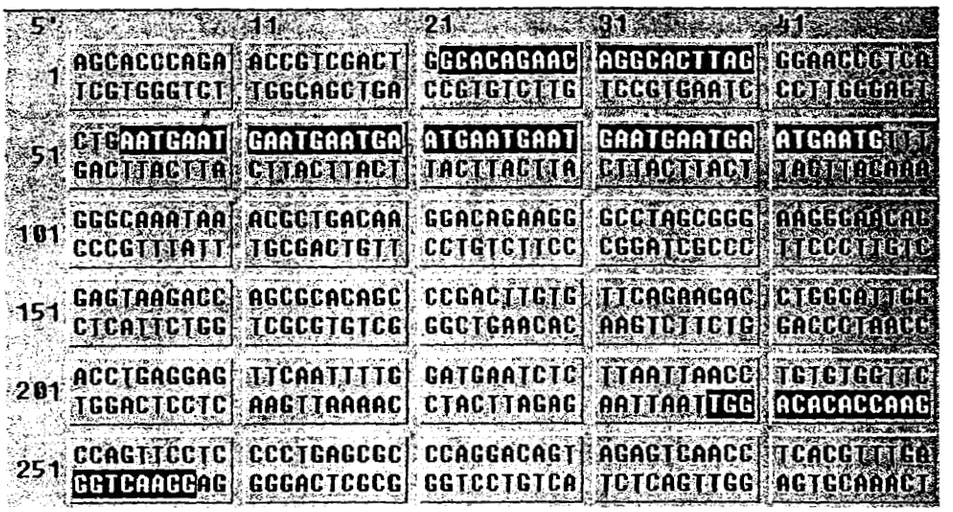
Forward primer is labeled with TMR dye (tetramethylrhodamine)

PCR product = 195 bp (9 repeats)

Exhibit 57. **TPOX STR primer positions for commercially available primers highlighted on the GenBank sequence.** The forward primer is shown in blue with the reverse primer in brown. The repeat is highlighted in green on the strand that contains the fluorescent dye. The 3'-positions of the reverse primers are identical but differ by a single base for the forward primers. Promega primers are longer at the 5'-end, which produces a larger PCR product by 7 bp (4 bases on forward and 3 bases on reverse).

AmpFlSTR® Green I Kit (PE Applied Biosystems)

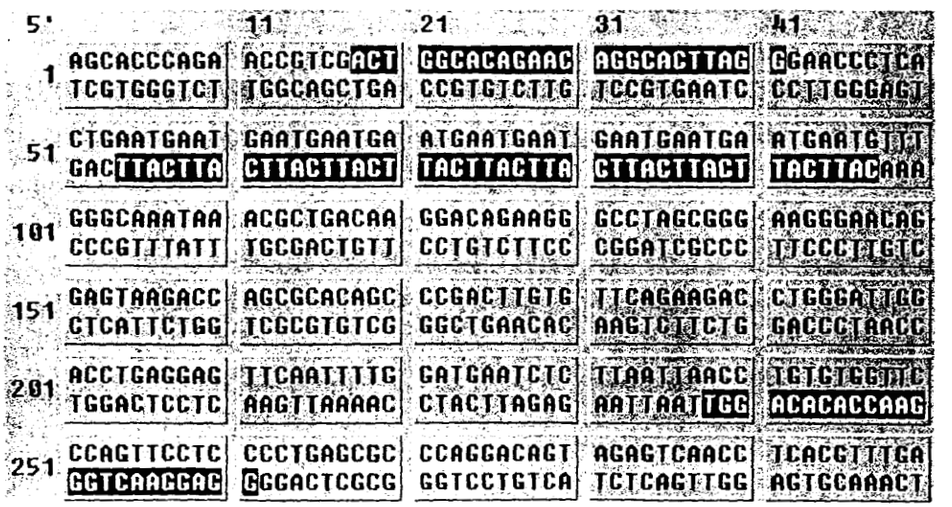
Forward primer is labeled with JOE dye (fluorescein derivative)



PCR product = 237 bp (11 repeats)

PowerPlex™ Kit (Promega Corporation)

Reverse primer is labeled with TMR dye (tetramethylrhodamine)

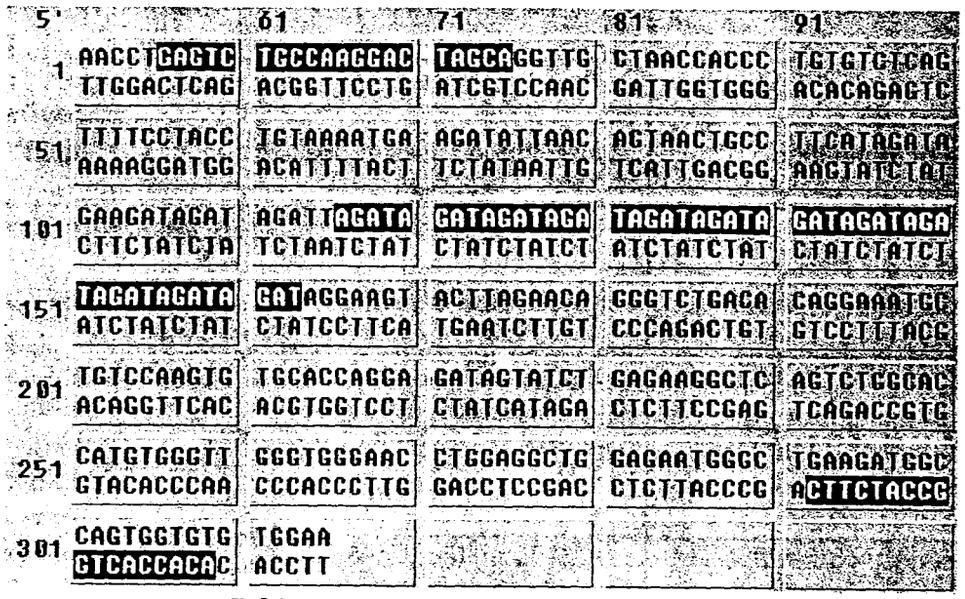


PCR product = 244 bp (11 repeats)

Exhibit 58. **CSF1PO STR primer positions for commercially available primers highlighted on the GenBank sequence.** The forward primer is shown in blue with the reverse primer in brown. The repeat is highlighted in green on the strand that contains the fluorescent dye. The 3'-positions of the reverse primers are identical but differ by a single base for the forward primers. Promega primers are longer at the 5'-end, which produces a larger PCR product by 11 bp (5 bases on forward and 6 bases on reverse).

AmpFlSTR® Green I Kit (PE Applied Biosystems)

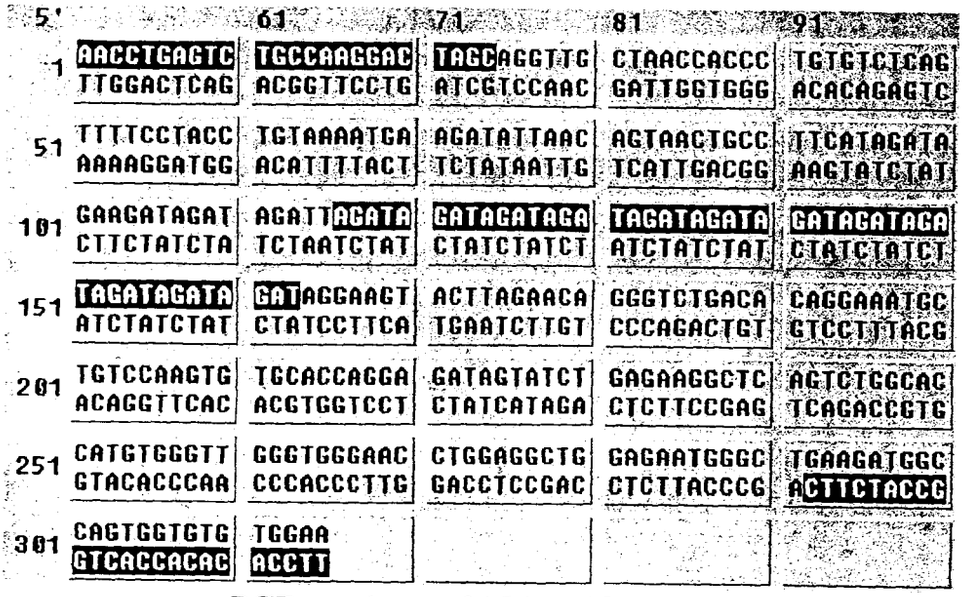
Forward primer is labeled with JOE dye (fluorescein derivative)



PCR product = 304 bp (12 repeats)

PowerPlex™ Kit (Promega Corporation)

Forward primer is labeled with TMR dye (tetramethylrhodamine)

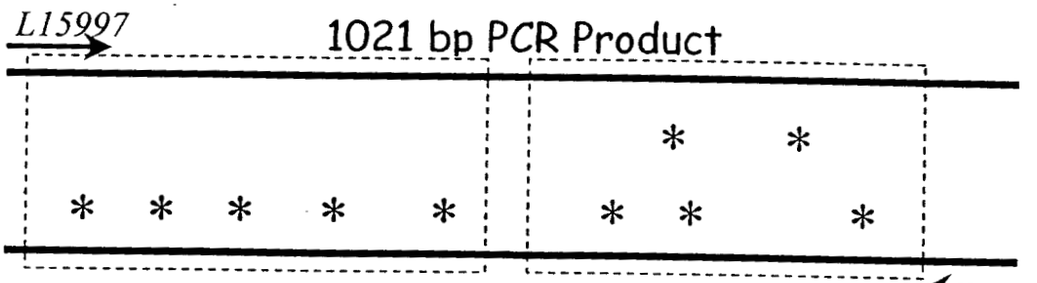


PCR product = 315 bp (12 repeats)

Exhibit 59. **Comparison of several characteristics for STR and SNP markers.** SNPs are more common in the human genome than STRs but are not as polymorphic.

Characteristics	**Short Tandem Repeats (STRs)**	**Single Nucleotide Polymorphisisms (SNPs)**
Occurrence in Human Genome	~1 in every 15 kb	~1 in every 1 kb
General Informativeness	high	low (20-30% as informative as STRs)
Marker Type	di-, tri-, tetranucleotide repeat markers	biallelic markers
Number of Alleles per Marker	typically >5	typically 2
Current Detection Methods	gel/capillary electrophoresis	microchip hybridization
Multiplex Capability (fluorescence)	>10 markers with multiple spectral channels	potential of 1000s on microchip
Heterozygote Resolution (fluorescence)	mobility differences between alleles	spectral differences between labeled nucleotides
Mass Spectrometry Measurement	Mass measurement of PCR-amplified allele(s)	Mass difference between primer and extension product

Exhibit 60. **Schematic representation of the mtDNA control region 10plex SNP assay.** The asterisks represent the relative positions of the SNP sites and the strand that is probed.



H16069	G or A	H00073	T or C
H16129	C or T	L00146	T or C
H16189	A or G	H00152	A or G
H16224	A or G	L00195	T or C
H16311	A or G	H00247	C or T

Exhibit 61. **Mitochondrial DNA primers used for 10plex SNP reaction**

Primer Name	SNP site position	SNP base*	Sequence (5'→3') "b": biotin; ( ): cleavable base	Cleaved mass (Da)
MT5	H16224	A/G	b-GGAGTTGCAGTTGATGTGTGA(T)AGTTG	1580
MT8'	L00146	T/C	b-GTCGCAGTATCTGTCTTTGAT(T)CCTGCC	1790
MT10	H00247	C/T	b-CTGTGTGGAAAGTGGCTC(T)GCAGACATT	2785
MT3'	H16189	A/G	b-GGTTGATTGCTGTACTTGCTTG(T)AAGCATGGGG	3179
MT9	H00152	A/G	b-CTGTAATATTGAACGTAGG(T)GCGATAAATAAT	3740
MT6	H16311	A/G	b-GTGCTATGTACGG(T)AAATGGCTTTATGT	4355
MT2g	H16129	C/T	b-GTACTACAGGTGG(T)CAAGTATTTATGGTAC	4957
MT1	H16069	G/A	b-AAATACA(T)AGCGGTTGTTGATGGGT	5375
MT7	H00073	T/C	b-CCAGCGTC(T)GCGAATGCTATCGCGTGCA	5891
MT4e	L00195	T/C	b-CTACGT(T)CAATATTACAGGCGAACATAC	6500
PCR Primers (produce a 1021 bp PCR product spanning positions *15977-00422*)				
DLOOP-F1	L15997	CACCATTAGCACCCAAAGCT		
DLOOP-R1	H00401	CTGTTAAAAGTGCATACCGCCA		

*Anderson reference sequence base listed first

Exhibit 62. **Expected mass-to-charge ratios for various ions in the mtDNA 10-plex assay**

Name	Primer Mass	singly-charged ions				doubly-charged ions					triply-charged ions				
	primer	ddC	ddT	ddA	ddG	primer	ddC	ddT	ddA	ddG	primer	ddC	ddT	ddA	ddG
MT5	1580			1877	1893	790			939	947	527			626	631
MT8'	1790	2063	2078			895	1032	1039			597	688	693		
MT10	2785	3058	3073			1393	1529	1537			928	1019	1024		
MT3'	3179			3476	3492	1590			1738	1746	1060			1159	1164
MT9	3740			4037	4053	1870			2019	2027	1247			1346	1351
MT6	4355			4652	4668	2178			2326	2334	1452			1551	1556
MT2g	4957	5230	5245			2479	2615	2623			1652	1743	1748		
MT1	5375			5672	5688	2688			2836	2844	1792			1891	1896
MT7	5891	6164	6179			2946	3082	3090			1964	2055	2060		
MT4e	6500	6773	6788			3250	3387	3394			2167	2258	2263		

Exhibit 63. **SNP ions impacting multiplex design.** Doubly-charged and triply-charged ions from higher mass primers can interfere with singly-charged ions of smaller mass primers if the multiplex is not well designed.

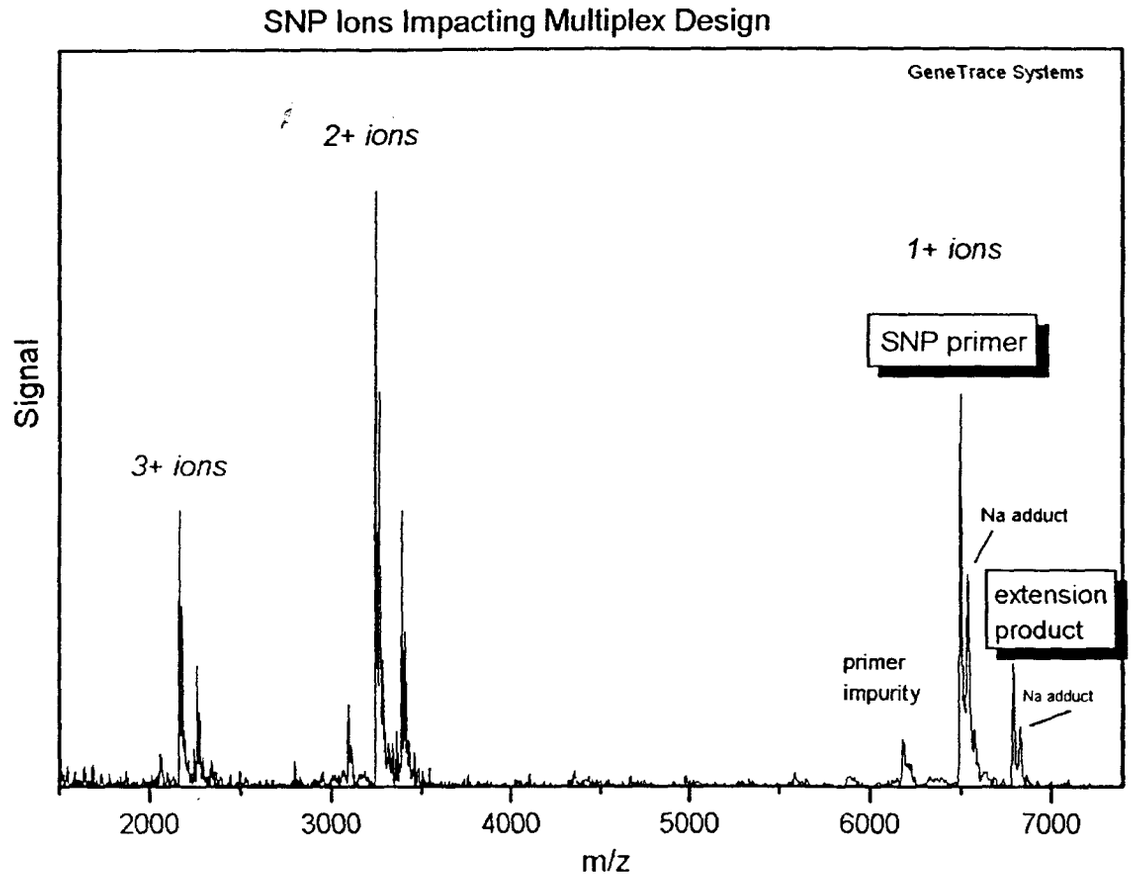


Exhibit 64. **Histogram of mass difference measurements for 200 samples** (50 for each ddN). Expected masses for the dideoxynucleotides are 273.2 Da for ddC, 288.2 Da for ddT, 297.2 Da for ddA, and 313.2 Da for ddG. The overall mass precision with this set of samples was less than 2 Da. See Li *et al.* (1999) *Electrophoresis* 20: 1258-1265 for more details.

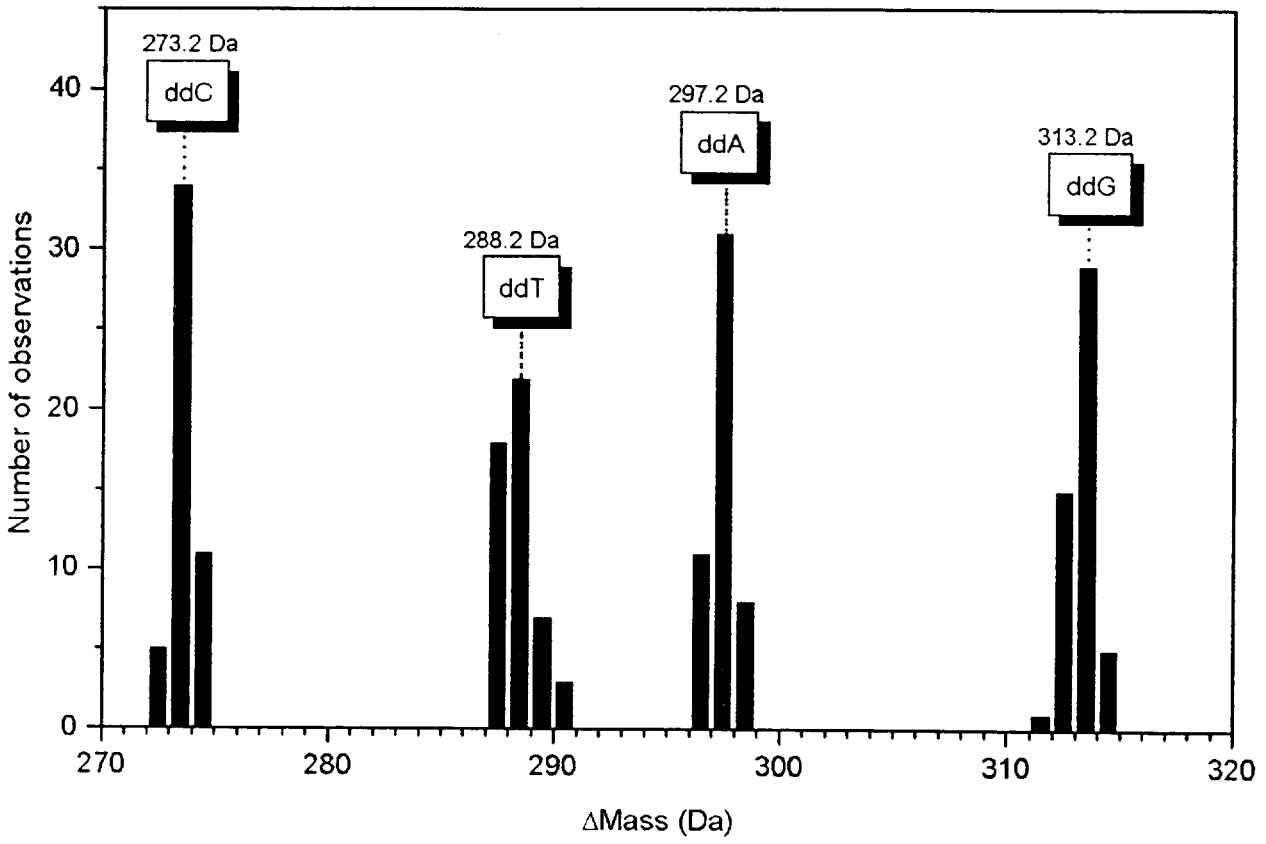


EXHIBIT 65. Multiplex PCR primers used for Y SNP markers. Universal sequences have been attached to the 5'-end of the primers.

Primer Name	Primer Sequence (5'→3')
M2-F3u	*ATT TAG GTG ACA CTA TAG AAT AC*G 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 AC*C 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 AC*T 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 AC*G 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 AC*A 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 AC*C 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 AC*C 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 AC*A 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 AC*G 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 AC*A 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 AC*C 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 AC*T 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 AC*A 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 AC*G 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 AC*C 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 AC*A 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 AC*C 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 AC*C 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 AC*G 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 66. **Multiplex PCR information for 17plex PCR reaction.**

Locus Name	Expected SNP	Primers Used	Primer Amounts	PCR Product Size (bp)
M3	C/T	F1u/R1u	0.4 pmole each	148 bp
M17	1 bp del	F1u/R1u	0.4 pmole each	149 bp
M13	G/C	F1u/R1u	0.4 pmole each	172 bp
M119	A/C	F1u/R1u	0.4 pmole each	188 bp
M2	A/G	F3u/R3u	0.4 pmole each	194 bp
M96	G/C	F1u/R1u	0.4 pmole each	198 bp
M122	T/C	F2u/R2u	0.4 pmole each	208 bp
M145	G/A	F1u/R1u	0.4 pmole each	218 bp
M45	G/A	F2u/R2u	0.4 pmole each	230 bp
M35	G/C	F1u/R5u	0.4 pmole each	233 bp
M9	C/G	F1u/R1u	0.4 pmole each	243 bp
M55	T/C	F1u/R1u	0.4 pmole each	247 bp
M60	1 bp ins	F1u/R1u	0.4 pmole each	263 bp
M89	C/T	F4u/R4u	0.4 pmole each	275 bp
M42	A/T	F1u/R1u	0.4 pmole each	300 bp
M67	A/T	F1u/R1u	0.4 pmole each	314 bp
M69	T/C	F1u/R1u	0.4 pmole each	326 bp
M26	G/C	F1u/R1u	0.4 pmole each	333 bp
M130	C/T	F1u/R1u	Included in 9plex but not 17/18plex PCR	155 bp

M20 primers not compatible with the other loci in the multiplex

PCR Mix: 5 mM $MgCl_2$, 1X PCR buffer II, 250 µM dNTPs, 2 U TaqGold, 40 pmol univ-F primer, 40 pmol univ-R primer, and 0.4 pmol each locus-specific primer pair in 20 µL volume

Thermal Cycling: 95 °C – 10 min; 50 cycles: 94 °C – 30 sec, 55 °C – 30 sec, 68 °C – 60 sec; 72 °C – 5 min; 4 °C forever (~3 hr total)

Exhibit 67. **Y SNP Multiplex Primer Information**. Designed manually for multiplex SNP assay (9plex) with non-overlapping masses.

Primer Name	Primer Sequence (5'→3')	Cleaved Mass (Da)	Expected SNP
Y1 (M42b-a)	b-CCAGCTCTCTTTTTCATTA(T)GTAGT	1580	T/A
Y2 (M96t-b)	b-CTTGGAAAACAGGTCTC(T)CATAATA	2150	C/G
Y3 (M35b-c)	b-TTCGGAGTCTC(T)GCCTGTGTC	2768	C/G
Y4 (M130b-b)	b-CCTT(T)CCCCTGGGCAG	3380	C/T
Y5 (M145b-b)	b-GATTAGGC(T)AAGGCTGGCTCT	3723	A/G
Y6 (M122t-c)	b-TAGAAAAGCAAT(T)GAGATACTAATTCA	4333	C/T
Y7 (M45t-d)	b-AAATTGGCAG(T)GAAAAATTATAGATA	4694	A/G
Y8 (M9b-c)	b-ACATGTCTAAA(T)TAAAGAAAAATAAAGAG	5354	C/G
Y9 (M89t-e)	b-CTTCC(T)AAGGTTATGTACAAAAATCT	6210	C/T

Exhibit 68. **Primers for Testing Y-Chromosome SNP Markers in Singleplexes**. Primers were designed automatically with UNIX SNP probe scripts written by Nathan Hunt but have not yet been tested.

Primer Name	Primer Sequence (5'→3')	Cleavage Mass	Expected SNP
M145-P1	b-CTTGCCTCCACGAC(T)TTCCT	1491	A/G
M35-P1	b-CGGAGTCTCTGCC(T)GTGTC	1556	C/G
M9-P1	b-AACGGCCTAAGATGG(T)TGAAT	1564	C/G
M26-P1	b-AGGCCATTCAGTG(T)TCTCTG	1820	C/G
M67-P1	b-TTGTTCGTGGACCCC(T)CTATAT (overlaps PCR reverse primer)	1828	A/T
M45-P1	b-CCTCAGAAGGAGC(T)TTTTGC	1835	C/T
M145-P2	b-GATTAGGCTAAGGC(T)GGCTCT	1845	C/T
M119-P1	b-TTCCAATTCAGCA(T)ACAGGC (overlaps PCR reverse primer)	1863	T/G
M55-P1	b-GCCCCTGGATGGTT(T)AAGTTA	1877	C/T
M20-P1	b-ACCAACTGTGGAT(T)GAAAAT (no PCR primers designed)	1886	A/G
M122-P1	b-TCAGATTTTCCCC(T)GAGAGC	1903	A/G
M96-P1	b-TTGGAAAACAGGTCTC(T)CATAATA	2150	C/G
M69-P1	b-GAGGCTGTTTACAC(T)CCTGAAA	2151	C/T
M130-P1	b-GGGCAATAAACCT(T)GGATTTC (overlaps PCR forward primer)	2173	C/T
M42-P1	b-CACCAGCTCTCTTTTTCAT(T)ATGTAGT	2198	A/T
M89-P1	b-CAACTCAGGCAAAG(T)GAGAGAT (overlaps PCR reverse primer)	2232	A/G
M3-P1	b-GGGTCACCTC(T)GGGACTGA	2537	A/G
M2-P1	b-CCTTTATCC(T)CCACAGATCTCA (overlaps PCR reverse primer)	3636	C/T
M13	Not designed		C/G
M60	Not designed (1bp insertion)		ins
M17	Not designed (1bp deletion)		del

Exhibit 69. **CE electropherograms of "drop out" 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 less 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.

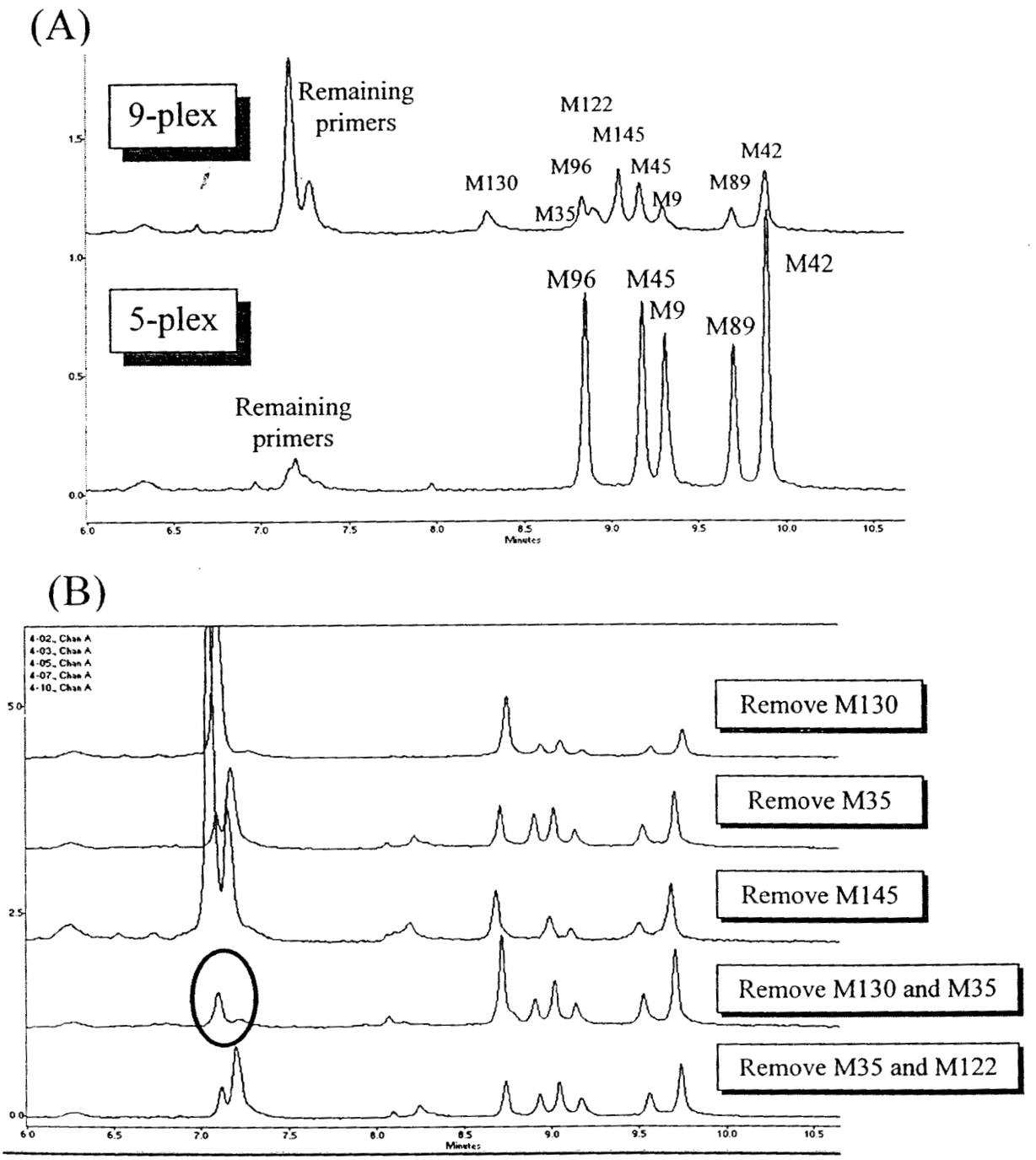


Exhibit 70. **Overlay of CE electropherograms demonstrating male specific amplification with the 17-plex set of PCR primers**. Note that a female sample (K562) failed to yield any peaks illustrating that the PCR reaction is Y chromosome-specific. The PCR primers used are described in Exhibit 65 with each locus-specific primer set at 0.4 pmole each. For the PCR reaction conditions, see Exhibit 66 and details in the text. The 17 amplicons may be seen more clearly in Exhibit 71.

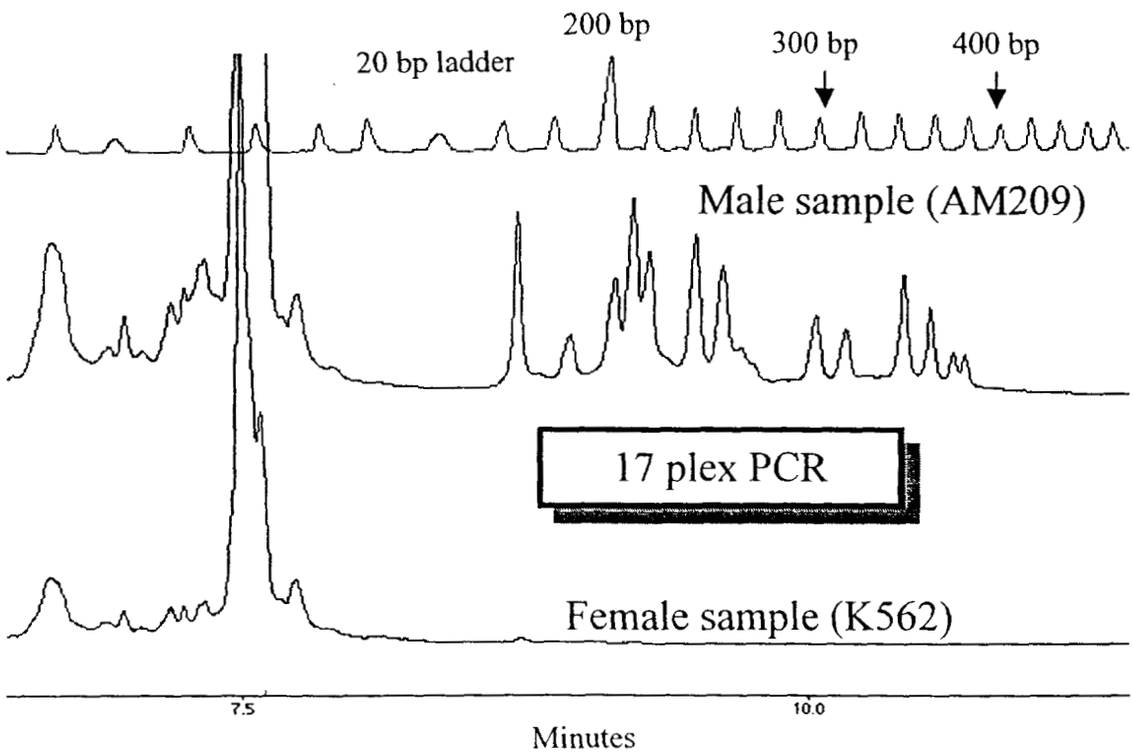


Exhibit 71. **CE electropherograms showing multiplex PCR sample compared to individual PCR reactions.** Notice that M67 primers are more efficient as there is a lower amount of remaining primers following the PCR reaction.

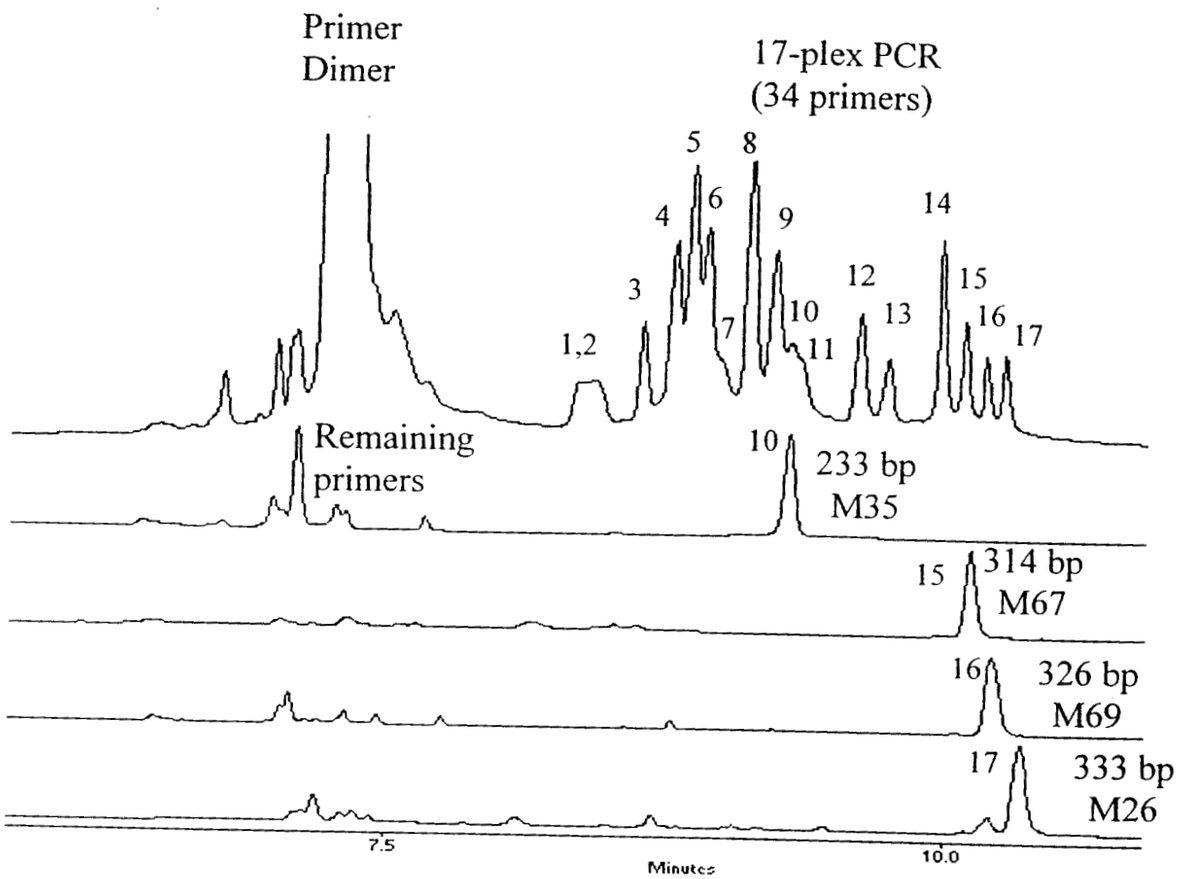


Exhibit 72. **Human Autosomal SNP Markers Designed for Testing NIH Diversity Panel**

SNP markers are not designed for multiplexing

SNP Primer Name	Primer Sequence (5'→3')	Cleavage Mass (Da)	Expected SNP
C6-P1	b-GGGGACAGCCA(T)GCACTG	1854	A/C
A2M-P1	b-GAAACACAGCAGCTTAC(T)CCAGAG	1863	A/G
LDLR-P1	b-CCTATGACACCGTCA(T)CAGCAG	1863	A/G
IL1A-P4	b-TTTTAGAAATCATCAAGCC(T)AGGTCA	1878	T/G
CD18-P2	b-GGACATAGTGACCG(T)GCAGGT	1894	C/T
IGF2-P1	b-CCACCTGTGATT(T)CTGGGG	1910	C/T
ALDOB-P1	b-CGGGCCAAGAAGG(T)ATCTACC	2102	A/G
PROS1-P1	b-CATAATGATATTAGAGCTCAC(T)CATGTCC	2118	A/G
NF1-P1	b-CGATGGTTGTATTTGTCACCA(T)ATTAATT	2156	A/G
AT3-P0	b-ATCTCCA(T)GGGCCCAGC	2787	C/T
CYP2D6-P1	b-GCAGCTTCAATGA(T)GAGAACCTG	2810	C/T
LIPC-P0	b-AACATGGCT(T)CGAGAGAGTTG	3483	A/C

Multiplex PCR Primers for 12plex PCR (not all amplicons are resolvable by CE). Universal primer sequences are in color.

Primer Name	Primer Sequence (5'→3')	PCR product size
AT3-F1u	*ATT TAG GTG ACA CTA TAG AAT ACT* GAG ACC TCA GTT TCC TCT TCT G	159 bp
AT3-R1u	*TAA TAC GAC TCA CTA TAG GGA GAC* CCT GGT CCC ATC TCC TCT AC	
C6-F1u	*ATT TAG GTG ACA CTA TAG AAT ACA* TCT GTC TTG CGT CCC AGT C	160 bp
C6-R1u	*TAA TAC GAC TCA CTA TAG GGA GAC* TCT TGC AGT CAG CCT CTT CA	
IGF2-F1u	*ATT TAG GTG ACA CTA TAG AAT ACA* GTC CCT GAA CCA GCA AAG A	163 bp
IGF2-R1u	*TAA TAC GAC TCA CTA TAG GGA GAC* TTT TCG GAT GGC CAG TTT AC	
LIPC-F1u	*ATT TAG GTG ACA CTA TAG AAT ACA* ACA CAC TGG ACC GCA AAA G	173 bp
LIPC-R1u	*TAA TAC GAC TCA CTA TAG GGA GAC* ACC CAG GCT GTA CCC AAT TA	
NF1-F1u	*ATT TAG GTG ACA CTA TAG AAT ACA* AGG AGC AAA CGA TGG TTG TA	181 bp
NF1-R1u	*TAA TAC GAC TCA CTA TAG GGA GAC* TAG GTG GCT GCA AGG TAT CC	
LDLR-F5u	*ATT TAG GTG ACA CTA TAG AAT ACC* CAC GGC GTC TCT TCC TAT	181 bp
LDLR-R5u	*TAA TAC GAC TCA CTA TAG GGA GAC* TGG TAT CCG CAA CAG AGA CA	
CYP2D6-F1u	*ATT TAG GTG ACA CTA TAG AAT ACG* GTG CAG AAT TGG AGG TCA T	182 bp
CYP2D6-R1u	*TAA TAC GAC TCA CTA TAG GGA GAC* AGA ACA GGT CAG CCA CCA CTA	
CD18-F2u	*ATT TAG GTG ACA CTA TAG AAT ACA* TCC AGG AGC AGT CGT TTG T	193 bp
CD18-R2u	*TAA TAC GAC TCA CTA TAG GGA GAC* ATG CCG CAC TCC AAG AAG	
ALDOB-F1u	*ATT TAG GTG ACA CTA TAG AAT ACC* ACA TTT GGG GCT TGA CTT T	231 bp
ALDOB-R1u	*TAA TAC GAC TCA CTA TAG GGA GAC* TCC TTC AGT CTC CTG TCA TCA A	
A2M-F1u	*ATT TAG GTG ACA CTA TAG AAT ACC* TCT GCC ATG CAA AAC ACA C	247 bp
A2M-R1u	*TAA TAC GAC TCA CTA TAG GGA GAC* AAC ATT CAA GTT TCC CTT ACT CAA	
PROS1-F1u	*ATT TAG GTG ACA CTA TAG AAT ACT* AAT GGC TGC ATG GAA GTG A	292 bp
PROS1-R1u	*TAA TAC GAC TCA CTA TAG GGA GAC* CCA GGA AAG GAC CAC AAA AT	
IL1A-F1u	*ATT TAG GTG ACA CTA TAG AAT ACT* TTG CTT CCT CAT CTG GAT TG	324 bp
IL1A-R1u	*TAA TAC GAC TCA CTA TAG GGA GAC* AGC AGC CGT GAG GTA CTG AT	

Exhibit 73. **Comparison of ABI 310 and mass spec allele calls for 90 CEPH/Diversity samples.** Out of 1080 possible allele calls with these 3 STR loci, there were 100 with no data collected (indicated as a zero "0" on the allele call axis) and only 12 calls differed between the two methods, or ~98% correlation.

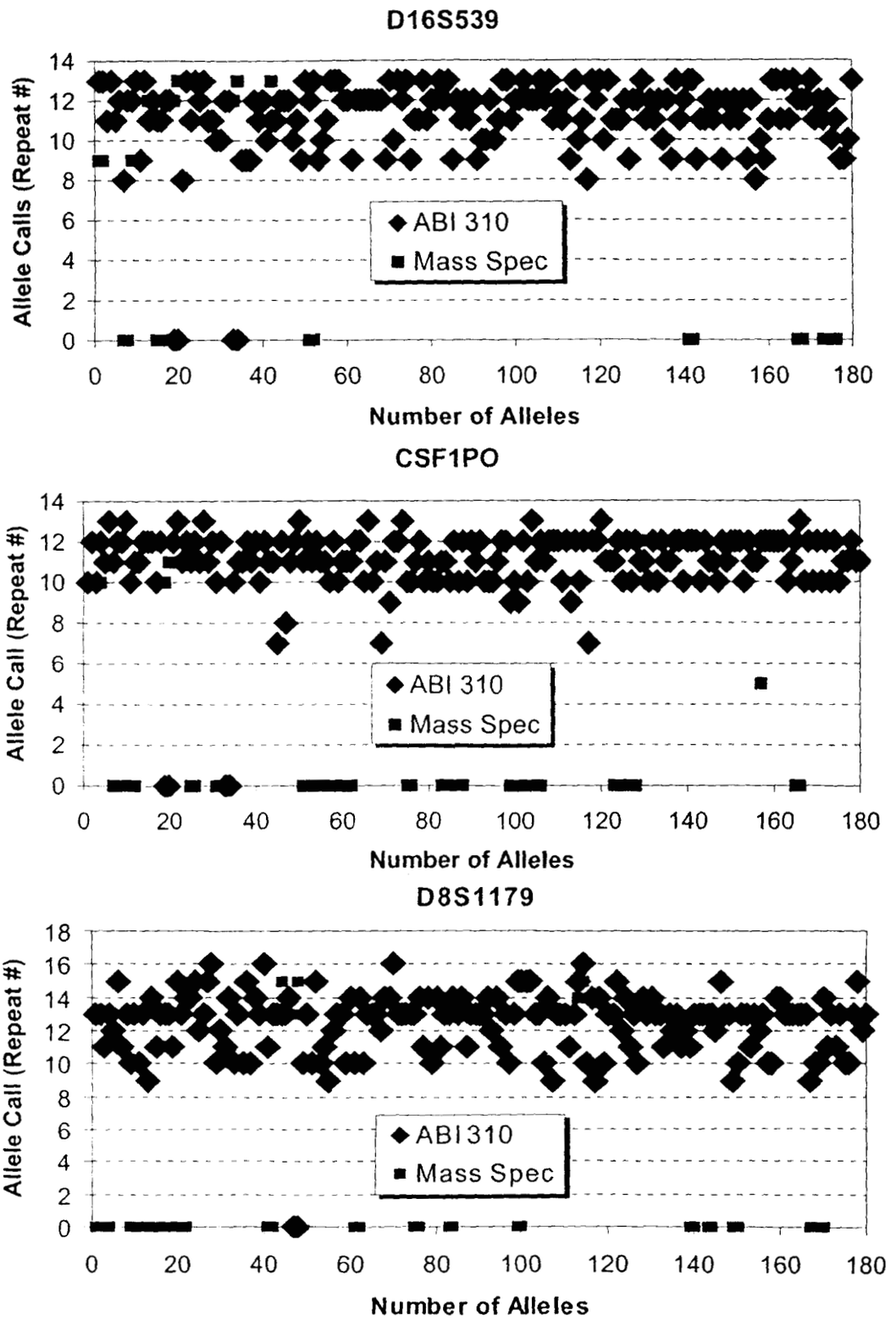
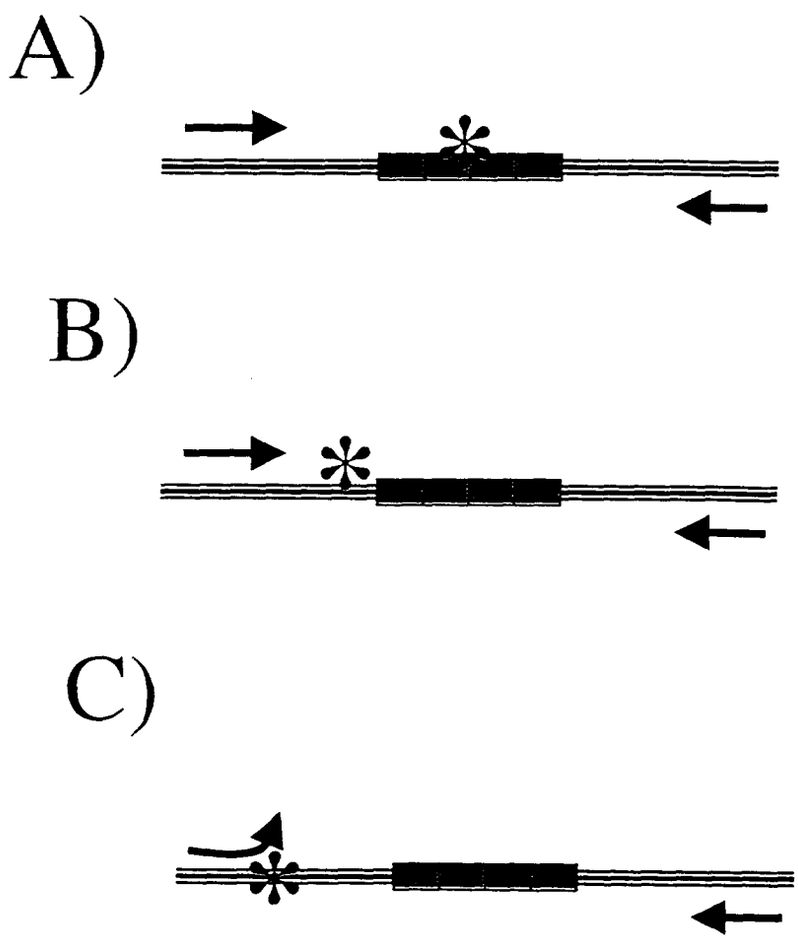


Exhibit 74. **Effects of sequence variation in or around STR repeat regions**. The asterisk symbolizes a DNA difference (base change, insertion or deletion of a nucleotide) from a typical allele for a STR locus. In situation (A), the variation occurs within the repeat region and should have no impact on the primer binding and the subsequent PCR amplification (although the overall amplicon size may vary slightly). In situation (B), the sequence variation occurs just outside the repeat in the flanking region but interior to the primer annealing sites. Again, PCR should not be effected although the size of the PCR product may vary slightly. However, in situation (C) the PCR can fail due to a disruption in the annealing a primer because the primer no longer perfectly matches the DNA template sequence. Therefore, if sequence variation occurs in the flanking region for a particular locus, one set of primers may work while another would fail to amplify the template. The template would therefore be a "null" allele.



Int J Legal Med (1998) 112:45–49

ORIGINAL ARTICLE

J. M. Butler · J. Li · T. A. Shaler · J. A. Monforte
C. H. Becker

Reliable genotyping of short tandem repeat loci without an allelic ladder using time-of-flight mass spectrometry

Received: 16 January 1998 / Received in revised form: 17 March 1998

Abstract DNA separations which traditionally have been performed by slab gel or capillary electrophoresis, may now be conducted via time-of-flight mass spectrometry (TOF-MS). The advantages of using a mass spectrometry approach for short tandem repeat (STR) characterization include a dramatic increase in both the speed of analysis and the accuracy of mass measurements. We report here typing of the STR loci TH01, TPOX, and CSF1PO as well as the sex-typing marker amelogenin using TOF-MS. Allelic ladders, which are typically used with electrophoretic separation systems to correct for mobility differences of DNA fragments under various conditions, are not needed for accurate genotyping with TOF-MS. A mass precision of 0.1% RSD, which corresponds to approximately 0.1 nucleotide, was routinely observed. Mass accuracies were better than a fraction of a single nucleotide when a daily mass calibration was used. STR microvariants, such as the TH01 allele 9.3, could be detected and resolved from alleles which differ by as little as a single base. In addition, the smaller PCR product sizes (55–125 bp) examined in this study have the potential advantage of being more successful when amplifying forensic samples with degraded DNA.

Key words Short tandem repeats (STRs) · Allelic ladders · Mass spectrometry · DNA fragment analysis · Microvariants

Introduction

Short tandem repeat (STR) loci are useful DNA markers for human identity testing and genetic mapping (Edwards et al. 1991; Fregeau and Fourney 1993; Kimpton et al. 1993). Dozens of tetranucleotide STRs have been examined by the forensic DNA community and validated for use in paternity testing and forensic identification (Hammond et al. 1994; Kimpton et al. 1996; Sprecher et al. 1996). The polymorphic variation in allele length has been previously detected by slab gel electrophoresis with silver staining (Sprecher et al. 1996) or multiple color fluorescent detection (Fregeau and Fourney 1993; Kimpton et al. 1993). More recently, capillary electrophoresis (Butler et al. 1994; Wang et al. 1995; Mansfield et al. 1996) has been used to resolve and type STR alleles. We introduce here the application of time-of-flight mass spectrometry to STR typing (see also Ross and Belgrader 1997).

Mass spectrometry offers unprecedented analysis times – on the order of seconds per sample – with excellent accuracy in measuring DNA fragment size. Substantial improvements have been made in recent years with the development of an effective ionization procedure, known as matrix-assisted laser desorption ionization (MALDI) and the discovery of new matrices, particularly 3-hydroxypicolinic acid (Wu et al. 1993). In MALDI, DNA samples are mixed with an organic matrix and allowed to co-crystallize on a sample plate. Multiple samples are typically spotted in an array on the sample plate and then sequentially analyzed. After the sample plate is placed in the mass spectrometer, which is under vacuum, a pulse of laser energy liberates a small portion of the DNA sample. While the generated ions travel to the detector in a matter of microseconds, multiple spectra are averaged for signal processing, which extends the measurement time to a few seconds. The DNA size is calculated by the time-of-flight to the detector in comparison to mass standards. A daily calibration (or alternatively when instrument conditions are varied) is usually sufficient to maintain a high degree of mass accuracy. We demonstrate here that the STR results from time-of-flight mass spectrometry analysis may be obtained more accurately than gel electrophoresis and orders of magnitude faster. In addition, STR alleles may be reliably typed without comparison to allelic ladders.

Materials and methods

Human genomic DNA samples representing several ethnic groups (African American, European, and Oriental) were purchased from

J. M. Butler (✉) · J. Li · T. A. Shaler · J. A. Monforte
C. H. Becker
Gene Trace Systems Inc., 1401 Harbor Bay Parkway, Alameda, CA 94502
e-mail: butler@genetrace.com
Tel.: +1-510-748-6124; Fax: +1-510-748-6001

Table 1 PCR product sizes and primer sequences used in this study

Locus Name (GenBank accession)	Known alleles*	PCR product size ranges	Primer sequences for this study	Locus reference
Amelogenin (M55418, M55419)	X, Y	106, 112 bp	5′-**B**-CCCTGGGCTCTGTAAAGAATAGTG-3′ 5′-ATCAGAGCTTAAACTGGGAAGCTG-3′	Sullivan et al. 1993
TH01 (D00269)	3,5,6,7,8,8.3,9,9.3, 10,10.3,11,12,13.3	55–98 bp	5′-CCTGTTCCTCCCTTATTTCCC-3′ 5′-**B**-GGGAACACAGACTCCATGGTG-3′	Edwards et al. 1991
TPOX (M68651)	6,7,8,9,10,11,12,13,14	69–101 bp	5′-**B**-CTTAGGGAACCCTCACTGAATG-3′ 5′-GTCCTTGTCAGCGTTTATTTGC-3′	Anker et al. 1992
CSF1PO (X14720)	6,7,8,9,10,11,12,13, 14,15	89–125 bp	5′-ACAGTAACTGCCTTCATAGATAG-3′ 5′-**B**-GTGTCAGACCCTGTTCTAAGTA-3′	Hammond et al. 1994

*: as reported in the literature; see STR Fact Sheets in STRBase (http://ibm4.carb.nist.gov:8800/dna/home.htm)
B: biotinylated strand

Bios Laboratories (New Haven, Conn.). K562 cell line DNA (Promega Corporation, Madison, Wis.) was used as a control in our experiments since the genotypes have been reported for this cell line with the STR loci examined here (GenePrint STR manual Promega 1995). Allelic ladders were reamplified using the PCR conditions and primers listed below from a 1:1000 dilution of AmpFlSTR Green I (CSF1PO, TPOX, TH01, amelogenin) allelic ladders (PE Applied Biosystems, Foster City, Calif.).

Primers were designed for each STR locus using Gene Runner software (Hastings Software, Inc., Hastings, N.Y.) and sequence information from GenBank (http://www.ncbi.nlm.nih.gov). To improve the sensitivity and resolution in the mass spectrometer, primers were placed close to the repeat region to make the PCR product size ranges under 120 bp in size when possible (Table 1). In the case of amelogenin, previously published primers were used. Primers were purchased from Biosource/Keystone (Menlo Park, Calif.) or synthesized in-house. One primer in each locus-specific set was biotinylated at the 5′ end.

Each PCR reaction contained 20 pmol of both forward and reverse primers, 1 U Taq polymerase (Promega Corporation, Madison, Wis.), 1X STR buffer (Promega) and typically 1–10 ng human genomic DNA. PCR was performed in 20 µl reaction volumes using a MJ Research DNA Engine (MJ Research, Watertown, Mass.). The PCR thermal cycling conditions were 94 °C for 2 min, 35 cycles at 94 °C for 30 s, 60 °C for 30 s, 72 °C for 30 s and 72 °C for 5 min.

The final incubation at 72 °C favors non-templated nucleotide addition (Clark 1988; Kimpton et al. 1993). Following PCR, a 1 µl aliquot of the PCR product was typically checked on a 2% agarose gel to verify amplification success. A purification procedure involving solid-phase capture and release from streptavidin-coated magnetic beads was then utilized (Monforte et al. 1997) to remove salts which interfere with the mass spectrometry process (Shaler et al. 1996). In the final step, samples were evaporated to dryness using a speed vac and then reconstituted in 0.5 µl of matrix and spotted on the sample plate. The matrix typically used for STR analysis was 3-hydroxypicolinic acid (3-HPA; Lancaster Synthesis, Windham, N.H.) in 25 mM ammonium citrate (Sigma) and 25% acetonitrile. A GeneTrace-designed and built linear time-of-flight mass spectrometer was used as previously described (Wu et al. 1994).

Results and discussion

Mass calibration and precision

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 done at the beginning of each day or whenever conditions are changed for the mass spectrometer. We typically use two oligonucleotides, a 15mer (4507.0 Da) and a 36mer (10998.2 Da) for mass calibration. Over the course of multiple analyses under the same conditions, the calibration drift is neglible. Reanalysis of the standards, even after several hours, results in a shift of only a few Daltons. While the use of these mass standards was not ideal for calibration (since all of the examined STR alleles fell outside the mass range defined by 15mer and 36mer), reliable results were still obtained. It should be pointed out that a two-point method of calibration is typically used in time-of-flight mass spectrometry and works very well (Whittal et al. 1997). Future work will examine improvements in mass determination using internal standards as well as external standards over various size ranges.

Sample purity can play an important role in mass accuracy, resolution and precision. In particular, sodium and potassium ions in a sample can interact with the DNA molecules. A salty sample will be evidenced by adducts which are multiples of 23 Da or 39 Da (or a combination of the two) larger in size than the DNA molecule of interest. In addition, photo adducts from the matrix, which are approximately 139 Da in the case of 3-HPA, may be seen as a result of the MALDI process. These adducts can reduce the spectral resolution and mass accuracy, especially with larger sized DNA molecules. If a sample is not well purified, peak masses may be shifted slightly higher due to unresolved adducts (Shaler et al. 1996). We carefully purified our DNA samples and removed contaminating salts using a solid-phase capture and release method (Monforte et al. 1997). This procedure has been fully automated on a robotic workstation and will be described in a future communication. Using our robotic workstation in combination with a single high-throughput mass spectrometer, we have been able to purify and analyze over 2,000 samples in a single day.

To demonstrate the excellent reproducibility of mass spectrometry, we collected 15 mass spectra of a TPOX allelic ladder. A histogram plot of the obtained masses for alleles 8, 9, 10, and 11 show that all alleles were easily segregated and distinguishable (Fig. 1). Statistical analysis of the data found that the standard deviation about the mean for each allele ranged from 20 to 27 Da or approximately 0.1% relative standard deviation (RSD). For each TPOX allele, all of the data fell within four 30 Da bins, which corresponds to 120 Da or approximately 0.4 nucleotides since each nucleotide is near 300 Da in mass (C = 289.2, T = 304.2, A = 313.2, and G = 329.2 Da). The mass between alleles is equal to the repeat unit, which in

Table 2 Mass differences between TH01 alleles using upper strand (TCAT repeat) or lower strand (AATG repeat)

	Upper Strand			Lower Strand	
	Expected (Da)	Observed (Da)		Expected (Da)	Observed (Da)
Allele 5–6	1211	1210	Allele 5–6	1260	1259
Allele 6–7	1211	1211	Allele 6–7	1260	1262
Allele 7–8	1211	1215	Allele 7–8	1260	1269
Allele 8–9	1211	1215	Allele 8–9	1260	1267
Allele 9–9.3	907	915	Allele 9–9.3	947	948
Allele 9.3–10	304	306	Allele 9.3–10	313	315
Allele 9–10	1211	1221	Allele 9–10	1260	1263
Repeat = TCAT = 1210.8 Da = –CAT = 906.6 Da			Repeat = AATG = 1259.8 Da = –ATG = 946.6 Da		

ther verification that the correct type was assigned to a sample, the mass difference between heterozygous alleles may be used to confirm the repeat sequence and number of repeats.

Advantages of smaller PCR products and resolution of STR alleles

By designing primers that are near the repeat region, smaller PCR products are amplified than are normally used with standard gel or capillary electrophoresis studies (see Table 1). The primary purpose in using smaller PCR products is that they may be detected with higher sensitivity and resolution in the mass spectrometer than larger amplicons. There are several advantages to using smaller PCR products besides an improvement in the mass sensitivity and resolution.

PCR efficiency typically improves with smaller amplicon size and shorter cycle times may be used. In a forensic context, smaller PCR product size ranges have the potential to handle highly degraded DNA samples. For example, in a recent validation study, the Forensic Science Service found that amelogenin and TH01, the two loci with the smallest allele sizes, were the last to drop out when extremely degraded DNA samples were used (Sparkes et al. 1996). Larger PCR products are typically examined with gel electrophoresis because the optimum resolution for separation of DNA fragments is in the size range of 100–350 bp. By examining smaller PCR products, we improve the detection of STR alleles in the mass spectrometer with added benefits to the molecular biology of the PCR reaction and to potential forensic situations with degraded DNA.

An excellent test for resolution with STR typing is the separation of the HUMTH01 allele 10, which contains 10 repeats of AATG and allele 9.3, a common variant of allele 10 with an adenine deletion in the seventh repeat (Puers et al. 1993). Using the mass spectrometry technique described here, a HUMTH01 allelic ladder containing alleles 9.3 and 10 can be fully resolved (Fig. 3). In addition, the loss of the adenine can be detected because a mass difference between the two alleles of 315 Da was observed, which is near the expected value for adenine (313.2 Da).

Comparison of allele sizing with and without allelic ladders

With the development of each new STR system, an allelic ladder is typically prepared from a mixture of known alleles to ensure reliable typing (Edwards et al. 1991; DNA recommendations 1994; Smith 1995). While multiple color fluorescent detection systems permit DNA fragment sizing using internal standards added to each sample (Mayrand et al. 1992), an allelic ladder is typically run for accurate determination of STR alleles (Edwards et al. 1991; Smith 1995; Kline et al. 1997). The variability of DNA mobilities under different electrophoretic conditions make an allelic ladder important to reliable genotyping of STR markers (Kline et al. 1997). However, the construction of an allelic ladder can be a time-consuming process. Also, whether an allelic ladder is prepared in a DNA typing laboratory or purchased from a commercial source, it increases the overall cost of typing each sample. Eliminating the need for allelic ladders to perform reliable STR genotyping would result in lower sample costs and lower costs for research and development of new STR loci.

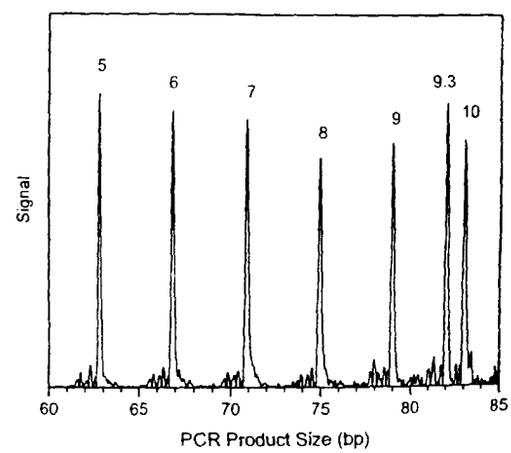


Fig. 3 TH01 allelic ladder demonstrating single-base resolution of alleles 9.3 and 10. The allele names, corresponding to the number of repeats, are included over each peak. The mass difference between alleles 9.3 and 10 was 315 Da, which corresponds to the difference of a single A (expected mass: 313.2 Da)

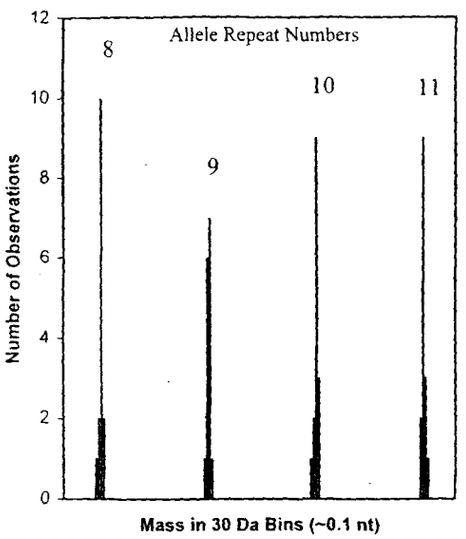


Fig. 1 Histogram showing mass precision in sizing TPOX alleles. The allele masses from 15 consecutive measurements of a TPOX allelic ladder were placed into mass bins of 30 Daltons (Da), which corresponds to approximately 0.1 nucleotide (nt). The standard deviation about each allele ranged from 20–27 Da

the case of TPOX is 1260 Da (AATG). Thus in Fig. 1, each allele differs by forty-two 30-Da bins and are easily distinguishable. Comparison to a similar histogram of data generated from fluorescent scanning of a polyacrylamide gel (Kimpton et al. 1993, Fig. 2) clearly shows the higher precision with mass spectrometry.

Mass accuracy compared to expected STR allele sizes

The GenBank sequence for each STR locus was downloaded into our Gene Runner software, and the number of repeats was counted. The mass of this reported GenBank sequence was then calculated, again using the Gene Runner software. The expected masses for other reported alleles in each STR system were obtained by adding or subtracting the mass of the repeat sequence. An additional mass of 313.2 Da was added to each allele to account for the nontemplate addition of adenine (Kimpton et al. 1993).

The two different strands of DNA can be examined in the mass spectrometer. For example, with the short tandem repeat TH01, if the bottom strand from GenBank is used, the expected repeat mass is 1259.8 Da for an AATG repeat. Alternatively, if the top strand is examined, the expected repeat mass is 1210.8 Da for a TCAT repeat. An examination of mass differences between alleles in a TH01 allelic ladder showed that the sequence content of the repeat could be differentiated (Table 2). The upper strand was discernible from the lower strand due to the different sequence contents of the repeats on the two strands. On the bottom strand with an AATG repeat, the mass difference observed between the 9.3 and 10 alleles was 315 Da, or the deletion of an adenine (expected mass = 313.2 Da), as previously reported (Puers et al. 1993). Alternatively, on the top strand with a TCAT repeat, the

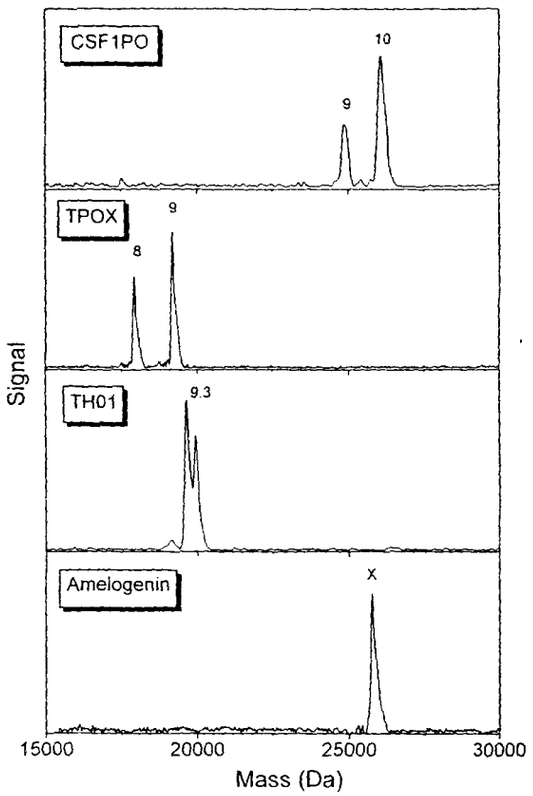


Fig. 2 Mass spectra from several STR loci and the sex-typing marker amelogenin using a K562 DNA template. The numbers above the peaks represent the allele calls based upon the observed mass. The expected types for K562 were obtained for all measured loci: CSF1PO (9, 10), TPOX (8, 9), TH01 (9.3, 9.3) and amelogenin (X, X). The TH01 peak is split because it is not fully adenylated

mass difference between the two alleles was 306 Da, or the deletion of a thymine (expected mass = 304.2 Da). Thus, even the STR repeat structure and nucleotide content of microvariants can be seen using our mass spectrometry technique.

When STR samples are analyzed using mass spectrometry, peaks that come within 100 Daltons of an expected allele mass are assigned to a particular allele. Since the measured standard deviation was approximately 30 Da, a mass window of plus or minus 100 Da from the expected allele mass should define 99.9% (3.3 standard deviations) of all measurements for a particular allele. This 100 Da mass window represents a fraction of a single nucleotide and has applied well to the STR data we have taken thus far from over a dozen different DNA templates obtained from Bios Laboratory. A mass window of ± 100 Da should also allow the flexibility to correctly determine PCR-amplified alleles that include other nucleotides besides adenine resulting from non-templated nucleotide addition (Clark 1988). We have correctly typed the standard K562 DNA template at each of the loci examined (Fig. 2). When a mass calibration has been performed at the instrument conditions used to analyze the sample, peak masses have fallen within 100 Da of expected allele masses. As a fur-

The high degree of accuracy for sizing STR alleles using mass spectrometry permits reliable typing of STR loci without the use of an allelic ladder. For example, in four different spectra of a heterozygote TPOX sample with alleles 8 and 11, the measured mass deviated from the expected mass by 41, 39, 29 and 38 Da for allele 8, and 16, 12, 28, and 53 Da for allele 11. On the other hand, the mass deviation for these same sample alleles from an allelic ladder, run just previous to their analysis, was 3, 1, 9, and 0 Da for allele 8 and 37, 9, 7, and 26 Da for allele 11. As might be expected, there was a slight improvement in mass accuracy when using an allelic ladder although the advantages are not significant. A mass difference of over 630 Da (half of a 1260 Da repeat) from the expected value would be required to incorrectly assign an adjacent allele with a tetranucleotide repeat. As part of another study, we reliably genotyped several thousand samples containing dinucleotide repeat loci as determined through obtaining comparable results by conventional gel electrophoresis methods (Butler et al., manuscript in preparation). With all samples tested to date, the correct alleles were accurately typed using the simple mass calibration procedure described here without need of an allelic ladder. In addition, further improvements in mass accuracy may be possible with mass calibrants that more closely reflect the mass range of measured STR alleles.

In conclusion time-of-flight mass spectrometry lends itself well to high-throughput STR typing with only a few seconds required for analysis of each sample. With mass spectrometry, the actual DNA oligomer resulting from PCR is detected intact rather than a fluorescent or radioactive label, as in gel electrophoresis. In addition, the high accuracy of mass spectrometry permits a measured peak mass to be correlated back to an STR allele without the use of an allelic ladder. Finally, the smaller PCR products (55–125 bp) described here for TH01, TPOX, and CSF1PO improve the sensitivity and resolution in the mass spectrometer as well as benefiting the molecular biology of the PCR reaction and potential forensic situations with degraded DNA.

Acknowledgements We thank Dr. Steve Lee of the California Department of Justice, Berkeley, California, for providing the allelic ladders and Joanna Hunter, Hua Lin, Dan Pollart, Mike Abbott, and Gordon Haupt for technical assistance. This research was supported in part by a grant from the National Institute of Justice (97-LB-VX-0003).

References

Anker R, Steinbrueck T, Donis-Keller H (1992) Tetranucleotide repeat polymorphism at the human thyroid peroxidase (hTPO) locus. Hum Mol Genet 1:137

Butler JM, McCord BR, Jung JM, Allen RO (1994) Rapid analysis of the short tandem repeat HUMTH01 by capillary electrophoresis. Biotechniques 17:1062–1068

Clark JM (1988) Novel non-templated nucleotide addition reactions catalysed by procaryotic and eucaryotic DNA polymerases. Nucleic Acids Res 16:9677–9686

DNA recommendations (1994) 1994 report concerning further recommendations of the DNA Commission of the ISFH regarding PCR-based polymorphisms in STR (short tandem repeat) systems. Int J Legal Med 107:159–160

Edwards A, Civitello A, Hammond HA, Caskey CT (1991) DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. Am J Hum Genet 49:746–756

Fregeau CJ, Fourney RM (1993) DNA typing with fluorescently tagged short tandem repeats: a sensitive and accurate approach to human identification. Biotechniques 15:100–119

Hammond HA, Jin L, Zhong Y, Caskey CT, Chakraborty R (1994) Evaluation of 13 short tandem repeat loci for use in personal identification applications. Am J Hum Genet 55:175–189

Kimpton CP, Gill P, Walton A, Urquhart A, Millican ES, Adams M (1993) Automated DNA profiling employing multiplex amplification of short tandem repeat loci. PCR Methods Appl 3:13–22

Kimpton CP, Oldroyd NJ, Watson SK, Frazier RRE, Johnson PE, Millican ES, Urquhart A, Sparkes BL, Gill P (1996) Validation of highly discriminating multiplex short tandem repeat amplification systems for individual identification. Electrophoresis 17:1283–1293

Kline MC, Duewer DL, Newall P, Redman JW, Reeder DJ, Richard M (1997) Interlaboratory evaluation of short tandem repeat triplex CTT. J Forensic Sci 42:897–906

Mansfield ES, Vainer M, Enad S, Barker DL, Harris D, Rappaport E, Fortina P (1996) Sensitivity, reproducibility, and accuracy in short tandem repeat genotyping using capillary array electrophoresis. Genome Res 6:893–903

Mayrand PE, Corcoran KP, Ziegle JS, Robertson JM, Hoff LB, Kronick MN (1992) The use of fluorescence detection and internal lane standards to size PCR products automatically. Appl Theor Electrophor 3:1–11

Monforte JA, Becker CH, Shaler TA, Pollart DJ (1997) Oligonucleotide sizing using immobilized cleavable primers. US Patent 5,700,642

Puers C, Hammond HA, Jin L, Caskey CT, Schumm JW (1993) Identification of repeat sequence heterogeneity at the polymorphic short tandem repeat locus HUMTH01[AATG]n and reassignment of alleles in population analysis by using a locus-specific allelic ladder. Am J Hum Genet 53:953–958

Ross PL, Belgrader P (1997) Analysis of short tandem repeat polymorphisms in human DNA by matrix-assisted laser desorption/ionization mass spectrometry. Anal Chem 69:3966–3972

Shaler TA, Wickham JN, Sannes KA, Wu KJ, Becker CH (1996) Effect of impurities on the matrix-assisted laser desorption mass spectra of single-stranded oligonucleotides. Anal Chem 68:576–579

Smith RN (1995) Accurate size comparison of short tandem repeat alleles amplified by PCR. Biotechniques 18:122–128

Sparkes R, Kimpton C, Watson S, Oldroyd N, Clayton T, Barnett L, Arnold J, Thompson C, Hale R, Chapman J, Urquhart A, Gill P (1996) 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 109:186–194

Sprecher CJ, Puers C, Lins AM, Schumm JW (1996) General approach to analysis of polymorphic short tandem repeat loci. BioTechniques 20:266–276

Sullivan KM, Mannucci A, Kimpton CP, Gill P (1993) A rapid and quantitative DNA sex test: fluorescence-based PCR analysis of X-Y homologous gene amelogenin. Biotechniques 15:637–641

Wang Y, Ju J, Carpenter BA, Atherton JM, Sensabaugh GF, Mathies RA (1995) Rapid sizing of short tandem repeat alleles using capillary array electrophoresis and energy-transfer fluorescent primers. Anal Chem 67:1197–1203

Whittal RM, Russon LM, Weinberger SR, Li L (1997) Functional wave time-lag focusing matrix-assisted laser desorption/ionization in a linear time-of-flight mass spectrometer: improved mass accuracy. Anal Chem 69:2147–2153

Wu KJ, Steding A, Becker CH (1993) Matrix-assisted laser desorption time-of-flight mass spectrometry of oligonucleotides using 3-hydroxypicolinic acid as an ultraviolet-sensitive matrix. Rapid Commun Mass Spectrom 7:142–146

Wu KJ, Shaler TA, Becker CH (1994) Time-of-flight mass spectrometry of underivatized single-stranded DNA oligomers by matrix-assisted laser desorption. Anal Chem 66:1637–1645

Electrophoresis 1999, *20*, 1258–1265

Jia Li
John M. Butler
Yuping Tan
Hua Lin
Stephanie Royer
Lynne Ohler
Thomas A. Shaler
Joanna M. Hunter
Daniel J. Pollart
Joseph A. Monforte
Christopher H. Becker

Gene Trace Systems,
Alameda, CA, USA

Single nucleotide polymorphism determination using primer extension and time-of-flight mass spectrometry

The high frequency of single nucleotide polymorphisms (SNPs) in the human genome makes them a valuable source of genetic markers for identity testing, genome mapping, and medical diagnostics. Conventional technologies for detecting SNPs are laborious and time-consuming, often prohibiting large-scale analysis. A rapid, accurate, and cost-effective method is needed to meet the demands of a high-throughput DNA assay. We demonstrate here that analysis of these genetic markers can now be performed routinely in a rapid, automated, and high-throughput fashion using time-of-flight mass spectrometry and a primer extension assay with a novel cleavable primer. SNP genotyping by mass spectrometry involves detection of single-base extension products of a primer immediately adjacent to the SNP site. Measurement of the mass difference between the SNP primer and the extension peak reveals which nucleotide is present at the polymorphic site. The primer is designed such that its extension products can be purified and chemically released from the primer in an automated format. The reduction in size of the products as a result of this chemical cleavage allows more accurate identification of the polymorphic base, especially in samples from a heterozygotic population. All six possible heterozygotes are resolved unambiguously, including an A/T heterozygote with extension products differing by only 9 Da. Multiplex SNP determination is demonstrated by simultaneously probing multiple SNP sites from a single polymerase chain reaction (PCR) product as well as from multiplexed PCR amplicons. Samples are processed in parallel on a robotic workstation, and analyzed serially in an automated mass spectrometer with analysis times of only a few seconds per sample, making it possible to process thousands of samples per day.

Keywords: Single nucleotide polymorphism / Time-of-flight mass spectrometry / Heterozygote determination / Mitochondrial DNA / High-throughput DNA analysis EL 3381

1 Introduction

Single nucleotide polymorphisms (SNPs) are the most frequent form of DNA sequence variation in the human genome and, as such, are becoming increasingly popular genetic markers for genome mapping studies, medical diagnostics, and identity testing [1–4]. SNPs are typically biallelic with two possible nucleotides (alleles) having frequencies of > 1% throughout the human population at a particular site in the genome. Although SNPs are less polymorphic than the currently used microsatellite markers, SNPs are more abundant – occurring approximately every kilobase [5]. Conventional technologies for detecting and scoring SNPs (*e.g.*, direct sequencing) are laborious and time-consuming, prohibiting large-scale analysis without great effort and expense [6, 7]. Detection of SNPs is commonly performed using single-strand conformational polymorphism assays or denaturing gradient gel electrophoresis [7]. More recently, denaturing high-performance liquid chromatography has been used for identification of candidate SNP sites [8]. After an SNP site has been identified and confirmed through sequence analysis, scoring assays are performed to determine the nucleotide(s) present in the sample population of interest. A number of PCR-based strategies for scoring SNP markers have been developed recently – including restriction digestion [7], ligation assays [4, 7], allele-specific PCR [2, 7], primer extension assays [9–13], and hybridization assays using fluorescent probes with solution-phase [14, 15] and solid-phase detection formats [3]. Scoring methods must be rapid, accurate, and cost-effective to meet the demands of a high-throughput DNA assay. This report will focus on a new, high-throughput DNA analysis approach to scoring SNP markers using the rapid capabilities of time-of-flight mass spectrometry (TOF-MS).

Correspondence: Dr. John M. Butler, 1401 Harbor Bay Parkway, Alameda, CA 94502, USA
E-mail: butler@genetrace.com
Fax: +510-748-6001

Abbreviations: ddN, dideoxynucleotide; **mtDNA**, mitochondrial DNA; **SAP**, shrimp-alkaline phosphatase; **SNP**, single nucleotide polymorphism

0173-0835/99/0606-1258 $17.50+.50/0

Int J Legal Med (1998) 112:50–54

ORIGINAL ARTICLE

M. Klintschar · B. Glock · E. M. Dauber · W. R. Mayr

Genetic variation and sequence studies of a highly variable short tandem repeat at the D17S976 locus

Received: 7 October 1997 / Received in revised form: 7 May 1998

Abstract The STR locus D17S976 was investigated by PCR amplification and native polyacrylamide gel electrophoresis in 158 unrelated Austrian Caucasians. No deviations from Hardy-Weinberg expectations were observed. The mean exclusion chance was 0.792, the discriminating power was 0.980 and the observed heterozygosity rate was 0.873. Moreover two alternative denaturing electrophoretic protocols are proposed. An allelic ladder consisting of 14 sequenced alleles (236–288 bp) was constructed. Sequence analysis revealed that the locus contained three different repeat motifs: ATCA, ATCT and ACCT, all of which vary in number between alleles. The aggregate number of the three tetrameric repeat types was used for allele designation. As a repeat with a single base deletion (ATC) was found in both the smallest and the largest alleles, a ".3" was added to the allele designation in those cases. Therefore the smallest allele is designated 19.3, and the largest allele is designated 32.3. To evaluate the exact extent of sequence variation more extensive sequence studies are necessary.

Key words Short tandem repeat · D17S976 · DNA Sequencing · Population data · Forensic DNA typing

Introduction

Short tandem repeat systems (STRs) are DNA polymorphisms consisting of di- to pentameric repeats [1, 2]. They are highly sensitive and allow typing of stains which are severely degraded as the amplified fragments are usually shorter than 300 bp [3]. On the other hand, most STRs are distinctly less polymorphous than RFLPs with only 3–6 common alleles [1]. Therefore a larger number of systems has to be typed for comparable results and some of the advantages of the STR approach, e.g. the saving of time and material [3], are affected. One possibility to overcome this problem is typing highly polymorphic STRs such as ACTBP2 [4]. But these STRs are more difficult to type and are less suitable for inter-laboratory controls [5]. As STR loci are abundant in the human genome [2] and only few of them have yet been evaluated for forensic purposes, it might be possible to find new, more suitable STRs. Therefore, the tetrameric repeat locus D17S976 (GDB entry number G00–250–233) was selected as the forensic utility has not yet been investigated. The aim of this study was to evaluate the genetic variation of this STR in an Austrian Caucasian population sample, to test its sensitivity and to construct a sequenced allelic ladder.

M. Klintschar (✉)
Institut für Gerichtliche Medizin, Karl Franzens Universität, Universitätsplatz 4, A-8010 Graz, Austria
e-mail: michael.klintschar@kfunigraz.ac.at
Fax +42 (316) 380 9655

B. Glock · E. M. Dauber · W. R. Mayr
Institut für Blutgruppenserologie und Transfusionsmedizin, Universität Wien, Währinger Gürtel 18–20/4i, A-1090 Wien, Austria

Materials and methods

Preparation of DNA

DNA was extracted from blood samples of 158 unrelated Austrian Caucasians as described [6]. For the sensitivity study K562 cell line DNA was stepwise diluted down to 10 pg per µl.

PCR protocol

Amplification was performed using 5 µl extracts without prior quantification in a 25 µl volume. The reaction mixture contained 0.5 U Dynazyme DNA Polymerase (Finnzymes, Espoo, Finland), 0.4 µM each primer, 100 µM each nucleotide, 2.5 µl 10× buffer (Finnzymes, Espoo, Finland).

Primer sequences

Primer 1: ATA TGC CAC CAC ACC TGG TT
Primer 2: TGG TAG CAT GCA TCT GTA GTC C
(GDB entry number G00–250–233)

In recent years, TOF-MS has been recognized as a technique with great potential for high-throughput DNA analysis [16, 17]. Speed, accuracy and capability for automation are appealing aspects of TOF-MS. Accurate data can be generated in only a few seconds per sample. In the past year, several groups have published reports using TOF-MS for SNP analysis. These reports include ligation analysis [28] and primer extension formats [13, 18–20] as well as hybridization approaches using peptide nucleic acids probes [21–23]. We describe here a novel approach to SNP analysis using mass spectrometry that is more cost-effective and flexible than previous approaches, as well as being fully automated, with a demonstrated capability for processing thousands of samples in a single day.

2 Materials and methods

2.1 Materials

Human genomic DNA from the K562 cell line (Promega, Madison, WI) was used as a standard template to test each new SNP marker. Additional DNA samples were provided by Dr. Peter Oefner (Standord University Department of Genetics) and the Monsanto Company to demonstrate and validate this technology. Oligonucleotides used for the PCR and the SNP extension reactions were obtained from Biosource/Keystone (Foster City, CA) or synthesized in-house. The SNP primer was biotinylated at the 5′-end and contained a cleavable nucleotide near the 3′-end. The polymerase and buffers were obtained from PE Applied Biosystems (Foster City, CA) and the dNTP and ddNTP mixtures and shrimp-alkaline phosphatase (SAP) were purchased from Amersham Life Sciences (Cleveland, OH).

2.2 Sample preparation

Each 20 μL PCR reaction contained 1 μM of both forward and reverse primers, 5–20 ng DNA template, 1 U AmpliTaq Gold, 1 × PCR buffer II, 1.5 –2.5 mM $MgCl_2$, and 200 μM each dNTP. PCR was performed using an MJ Research DNA Engine (Watertown, MA). The typical thermal cycling conditions were 95°C for 10 min, 35 cycles at 94°C for 10 s, 55°C for 30 s, 72°C for 30 s, and 72°C for 5 min. For the mitochondrial DNA work, the primer sequences were the same as those described by Piggee *et al.* [11] and Tully *et al.* [12]. Following PCR, the amplicons were treated with shrimp-alkaline phosphatase to hydrolyze the unincorporated dNTPs. Typically, 1 U SAP was added to each 20 μL PCR reaction and then incubated at 37°C for 60 min followed by heating at 75°C for 15 min. The SNP extension reaction consisted of a 5 μL aliquot of the SAP-treated PCR product, 1 × *Taq*FS buffer, 1.2–2.4 U *Taq*FS, 12.5 μM ddNTP mix, and 0.5 μM biotinylated 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. The SNP extension reaction was performed in a thermal cycler at 94°C for 1 min, and 25–35 cycles at 94°C for 10 s, 45–60 °C (depending on the annealing temperature of the SNP primer) for 10 s, and 70° for 10 s. These universal conditions worked well under a wide variety of SNP loci tested. A purification procedure involving solid-phase capture and release from streptavidin-coated magnetic beads was utilized to purify the DNA samples prior to mass spectral analysis [24, 27]. Parallel sample preparation was conducted on a robotic workdeck operated with a 96-tip pipet head developed at GeneTrace. The entire sample process from PCR setup to mass spectrometric sample preparation has been automated on multiple robotic workstations and will be described in a future communication.

2.3 Mass spectrometry

The DNA samples, consisting of the primer and the extension product(s), were spotted with 1–2 μL of a 3-hydroxypicolinic acid matrix solution [25]. A GeneTrace-designed linear TOF mass spectrometer was used as previously described with ultraviolet laser pulses [26]. Sample analysis has been fully automated with this mass spectrometer so that sample spots may be located and a particular mass range searched without operator intervention. For SNP analysis, mass calibration was performed with two oligonucleotide mass markers: a 4-mer (1272 Da) and a 15-mer (4507 Da).

2.4 Data analysis

In-house automated SNP analysis software was developed and used to determine the genotype for each SNP marker. The software accepts as input an expected primer mass and, after locating the pertinent primer, searches for the four possible extension products. 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.

3 Results

3.1 Design aspects for SNP assays

The approach to SNP determination described here has essentially three steps: (i) PCR amplification, (ii) phosphatase digestion, and (iii) SNP primer extension. Either strand of DNA may be probed in this SNP primer exten-

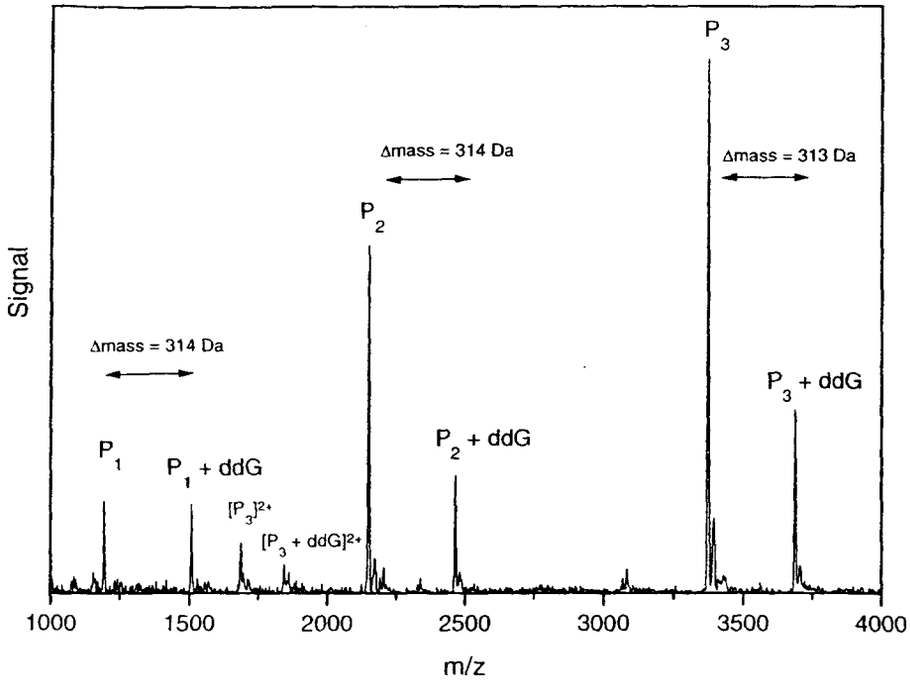


Figure 5. Mass spectrum of a multiplex SNP analysis involving three known mutations associated with Leber's hereditary optic neurophathy. The PCR products resulted from multiplex PCR of amplicons containing the three SNP sites. Some doubly charged ions from the larger DNA products may be seen between the first extension product (P_1+ddG) and the second primer (P_2). Primer masses and SNP sites are described in Table 2.

Table 2. Multiplexed SNP markers examined from human mitochondrial K562 cell line DNA

SNP marker (mtDNA[d] position)	Expected bases (wild-type listed first)	Primer mass (Da)	Extension mass (Da)	Mass difference (Da)	Calculated SNP base
Control region polymorphisms[a]			Fig. 4		
L00146[b]	T or C	P_1 = 1819.7	2109.8	290.1	T
H00247	C or T	P_2 = 2813.7	3087.2	273.5	C
H00152	A or G	P_3 = 3770.0	4.068.6	298.6	A
H16311	A or G	P_4 = 4387.9	4686.8	298.9	A
H16069	G or A	P_5 = 5411.3	5725.9	314.6	G
L00195	T or C	P_6 = 6509.4	6798.4	289.0	T
LHON point mutations[c]			Fig. 5		
L03460	G or A	P_1 = 1194.1	1508.2	314.1	G
L11778	G or A	P_2 = 2152.4	2466.5	314.1	G
L14459	G or A	P_3 = 3374.6	3688.3	313.7	G

a) SNP primers same as Tully *et al.* [12] without poly-T tail
b) L, light strand; H, heavy strand
c) PCR and SNP primers same as Piggee *et al.* [11]
d) mtDNA, mitochondrial DNA
Mass spectra may be seen in Figs. 4 and 5.

4 Discussion

The cleavable primer approach for SNP analysis described here has a number of advantages over other existing technologies in terms of the molecular biology and the ability to be automated as well as the MS detection. One feature of using a cleavable primer is the ability to probe both DNA strands simultaneously and still perform a solid-phase purification after a single PCR reaction (see Fig. 4). Because the 3′-end of the primer can be released following cleavage, the primer probe rather than the template strand may be biotinylated. Other primer extension methods, such as solid-phase minisequencing [12] and Genetic Bit Analysis [9], require that all of the primers used for probing SNP sites be on the same DNA strand because the template strand is captured and then

sion assay. PCR primers are designed to generate an amplicon that includes one or more SNP sites. The initial PCR reaction is performed with standard (unlabeled) primers. A phosphatase is added following PCR to remove all remaining dNTPs so that they will not interfere with the single-base extension reaction involving dideoxynucleotides. These reactions can all be performed in the same tube or well in a sample tray. A portion of the phosphatase-treated PCR product is then used for the primer extension assay. In our SNP primer extension assay, a special primer containing a biotin moiety at the 5′-end and cleavable nucleotide near the 3′-end hybridizes upstream of the SNP site with the 3′-end immediately adjacent to the SNP polymorphic site (Fig. 1). The biotin permits solid-phase capture for sample purification prior to MS analysis, and the cleavable nucleotide allows the 3′-end of the primer to be released from the immobilized portion as well as reducing the overall mass of the measured DNA molecule [27]. The complementary nucleotide(s) to the nucleotide(s) present at the SNP site are inserted during the extension reaction. In the case of a heterozygote, two extension products result. Only a single base is added to the primer during this process because only ddNTPs are used and the dNTPs left over from PCR are hydrolyzed with the phosphatase digestion step. If the extension reaction is not driven to completion (where the primer would be totally consumed), then both primer and extension product (*i.e.*, primer plus single nucleotide) are present after the primer extension reaction. The mass difference between these two DNA oligomers is used to determine the nucleotide present at the SNP site. As will be described below, we have used this approach to reliably determine all four possible SNP homozygotes and six possible heterozygotes.

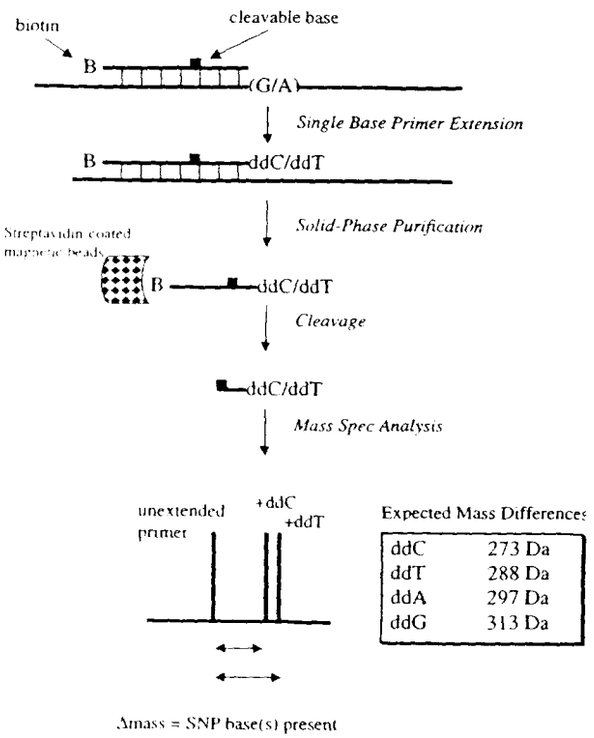


Figure 1. Schematic of SNP assay using a cleavable primer approach. An SNP primer hybridizes to the target DNA immediately upstream of the polymorphic site and a polymerase incorporates the complementary dideoxynucleotide in a single-base extension reaction. The sample then undergoes a solid-phase purification *via* a capture and release protocol to prepare it for mass spectrometry analysis. The mass difference between the primer and the extension product(s) indicates the nucleotide(s) present at the SNP site.

3.2 Analytical results

A high degree of precision and accuracy in making a mass difference measurement is required to correctly call the SNP nucleotide. We balance the primer extension reaction so that some of the primer remains to act as an internal standard. Thus, an absolute mass for the extension product(s) is not as crucial as the mass difference between the primer and the extension product(s). Figure 2 illustrates the precision and accuracy of 50 SNP meas-

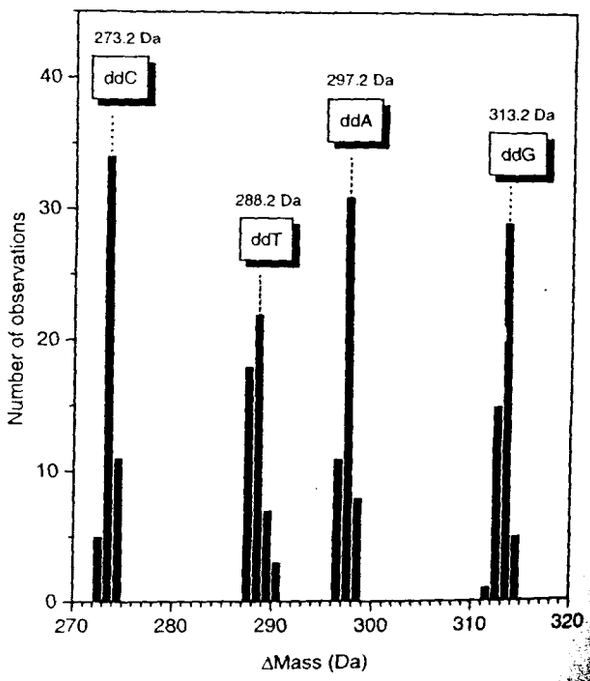


Figure 2. Histogram of mass difference measurements for 200 samples (50 for each ddN). Expected masses for the dideoxynucleotides are 273.2 Da for ddC, 288.2 Da for ddT, 297.2 Da for ddA, and 313.2 Da for ddG.

urements at each of the four possible nucleotides. The observed masses for each dideoxynucleotide (ddN) were 273.5 ± 0.5 Da for ddC, 288.3 ± 0.7 for ddT, 297.2 ± 0.5 Da for ddA, and 313.2 ± 0.5 Da for ddG. The average mass difference measurement for each ddN came within 0.3 Da of the expected masses, which are ddC = 273.2 Da, ddT = 288.2 Da, ddA = 297.2 Da, and ddG = 313.2 Da. In addition, the standard deviation for each ddN was less than 1 Da.

In terms of resolution, the most difficult bases to distinguish from one another are adenine and thymine, which differ by only 9 Da. As seen in Fig. 3, a T/A heterozygote can be resolved with our cleavable primer approach. A heterozygote containing C and G bases is the easiest to resolve as these two nucleotides differ by 40 Da. Table 1 describes the results from these heterozygous samples. The SNP scoring approach described here works well over a wide range of PCR product amounts. No prequantitation of PCR product is required to obtain an SNP result. We have reliably called the SNP nucleotide from as little as 10 fmol PCR product although we typically work in the high femtomole to low picomole range.

3.3 Multiplexing SNP markers

Examining multiple SNP markers in the same reaction reduces time, labor, and cost compared to single reactions. Multiplexing SNP markers may be achieved by using primers that are resolvable on a mass scale. With the approach described here, compatible primers with similar annealing temperatures may be used and cleavage sites may be placed at different positions in each primer. For example, one primer could have the cleavage site five bases from the 3′-end and another primer could have eight bases between the cleavage site and the 3′-

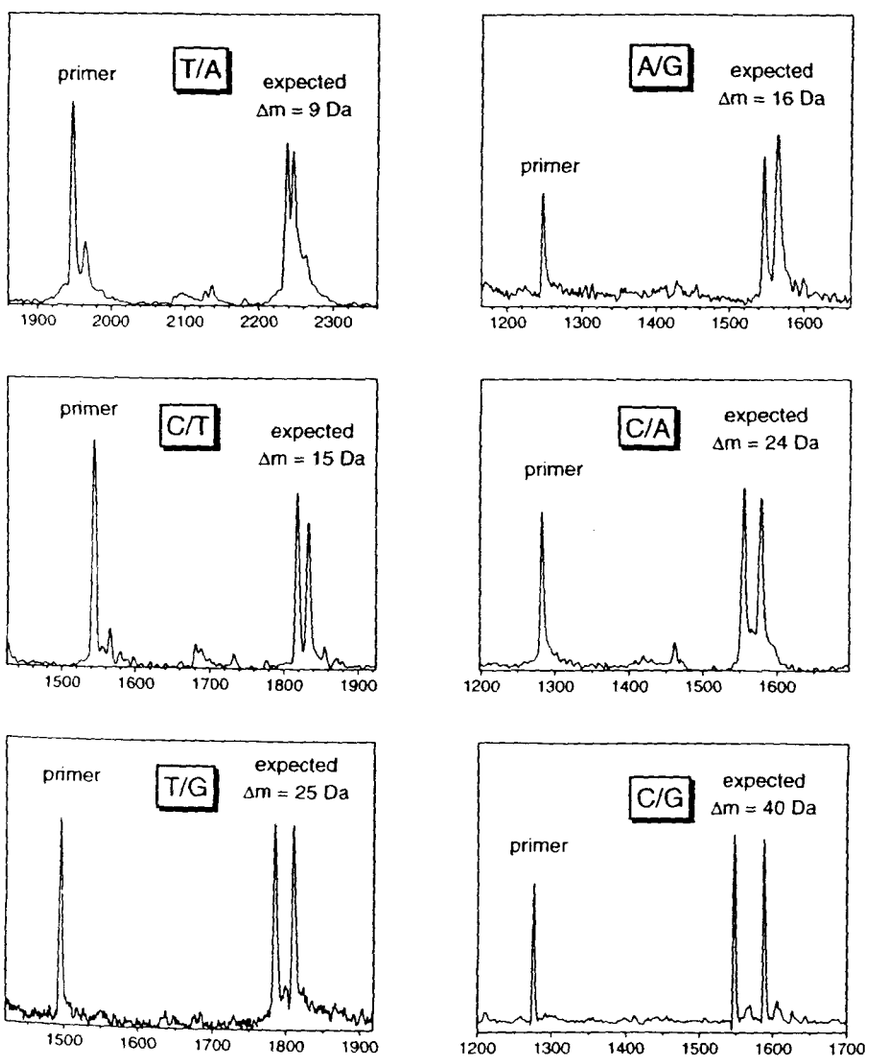


Figure 3. Mass spectra of heterozygous SNP samples. The six possible combinations of dideoxynucleotides are shown here (T/A, C/T, A/G, C/A, T/G, and C/G) with the expected mass differences between them. The SNP genotype is determined for each sample by measuring the mass difference between the primer and the extension products. Table 1 contains the measured mass information for these heterozygous samples.

Table 1. Heterozygote SNP detection results

Primer sequence[a)]	Expected primer mass (Da)	Primer mass (Da)	Extension masses (Da)	Mass difference (Da)	SNP site	Δ Mass between extension products (Da)
GACGGG	1937	1942.0	2231.0	289.0	T/A	9.1
			2240.1	298.1		
CAAA	1247	1244.7	1541.7	297.0	A/G	18.4
			1560.1	315.3		
ACACA	1536	1538.6	1812.9	274.3	C/T	15.1
			1828.0	289.4		
CGAG	1279	1277.7	1550.8	273.1	C/A	23.2
			1574.0	296.3		
TTGTC	1485	1490.5	1780.4	289.9	T/G	25.0
			1805.4	314.9		
GAAT	1278	1275.2	1547.8	272.6	C/G	39.9
			1587.7	312.5		

a) Sequence listed is from cleavage point to the 3′-end of the SNP primer
Mass spectra may be seen in Fig. 3

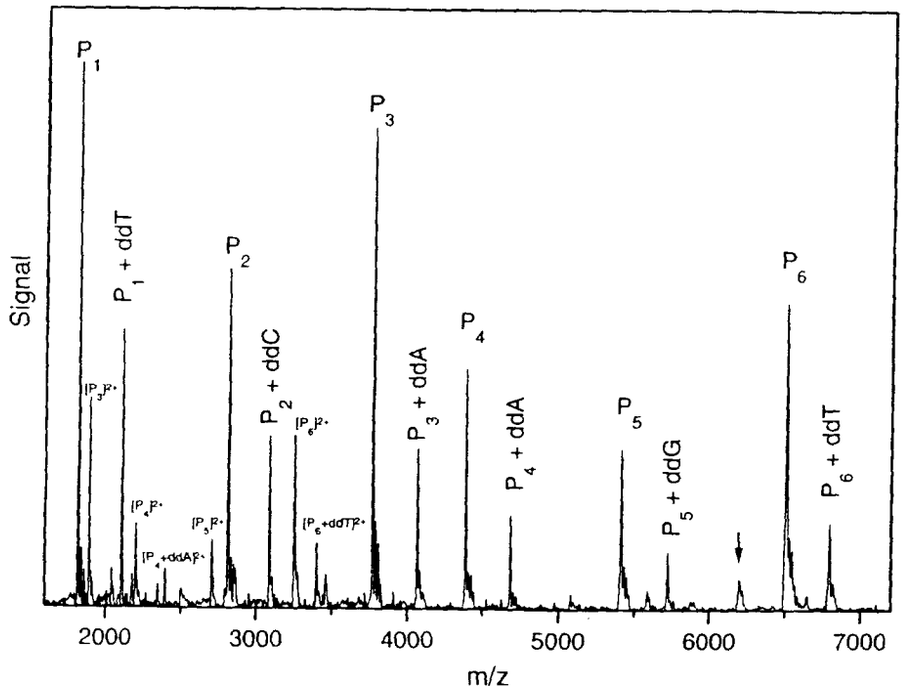


Figure 4. Mass spectrum of a multiplex SNP analysis involving six polymorphic sites within a single 1021 bp PCR product from the human mitochondrial DNA control region. All four bases are simultaneously determined from both strands of the PCR product. Primers were designed to avoid overlap between singly and doubly charged ions from the DNA products produced in this assay. The peak at 6200 Da indicated by the arrow is an impurity (n-1) from P_6 synthesis. Primer masses and SNP sites are described in Table 2.

end. Thus, after performing the SNP extension reaction and cleaving the primers, each primer and extension product can be easily resolved in the mass spectrometer.

Figures 4 and 5 illustrate multiplex SNP results from human mitochondrial DNA polymorphic sites used in human identity testing [12] and point mutations that are diagnositic for Leber's hereditary optic neuropathy [11]. Figure 4 is the mass spectrum of a 6-plex reaction from a single PCR product that demonstrates all four SNP bases being detected simultaneously. Additionally, both strands of the mitochondrial DNA PCR product are being probed in this example. Primers 2 through 5 probe the so-called heavy strand at positions H00247, H00152, H16311, and H16069, respectively, while primers 1 and 6 examine positions L00146 and L00195 on the light strand. In Fig. 5, three PCR products are multiplexed at the PCR level as well as in the SNP extension reaction portion. Table 2 describes the results from these human mitochondrial DNA SNP markers. DNA sequence analysis for these PCR templates confirmed the mitochondrial DNA control region SNP results listed in Table 2 (data not shown).