Final Report

Award No.: 2004-DN-BX-K214

Field Test of Current Technology Used in the Identification of Unidentified Remains

Principal Investigators:
Arthur J. Eisenberg, Ph.D.
Professor, Department of Pathology and Human Identification
Director, University of North Texas System Center for Human Identification
University of North Texas Health Science Center
3500 Camp Bowie Boulevard
Fort Worth, Texas 76107
Phone: (817) 735-0555
Fax: (817) 735-5016
e-mail: eisenber@hsc.unt.edu

John V. Planz, Ph.D.
Associate Professor, Department of Pathology and Human Identification
Associate Director, University of North Texas System Center for Human Identification
University of North Texas Health Science Center
3500 Camp Bowie Boulevard
Fort Worth, Texas 76107
Phone: (817) 735-2397
Fax: (817) 735-5016
e-mail: jplanz@hsc.unt.edu

Award: 2004-DN-BX-K214
Field Test of Current Technology Used in the Identification of Unidentified Remains

TABLE OF CONTENTS

Award: 2004-DN-BX-K214
Field Test of Current Technology Used in the Identification of Unidentified Remains

Figures and Tables

Award: 2004-DN-BX-K214
Field Test of Current Technology Used in the Identification of Unidentified Remains

Figures and Tables continued

Award: 2004-DN-BX-K214
Field Test of Current Technology Used in the Identification of Unidentified Remains

Figures and Tables continued

Figures and Tables continued

Award: 2004-DN-BX-K214

Field Test of Current Technology Used in the Identification of Unidentified Remains

ABSTRACT

The utilization of the National Missing Persons DNA Database component of the FBI's Combined DNA Index System (CODIS) has begun to increase. The identification of missing persons and skeletal remains has relied upon the comparison of both the STR and mitochondrial DNA (mtDNA) profiles generated from the remains and samples collected from families with a missing loved one. Several catastrophic events including the World Trade Center (WTC) disaster, Hurricane Katrina, and the Asian Tsunami have provided considerable knowledge regarding the effectiveness of the STR and mtDNA sequence analysis on highly degraded and compromised samples. The work from the WTC disaster has shown that the current mtDNA and STR technologies are often not sufficient to provide a level of statistical certainty required for a positive identification. The causes for this issue are twofold: 1) the current STR panels often yield incomplete or no results from skeletal remains that have been exposed to the elements for varying length of time due to the high state of sample degradation; and 2) although mtDNA typing is successful for greater than 95 percent of skeletal samples, common haplotypes found in Caucasian and African American populations result in numerous adventitious associations between remains and family reference samples that cannot be properly evaluated. Several newer technologies and applications have been described to aid in resolving both of the aforementioned issues. These include: reduced amplicon or miniSTRs to better interrogate polymorphisms in degraded material; expanded sequence analysis of the mtDNA genome to resolve common haplotypes; and the analysis of nuclear Single Nucleotide Polymorphisms (SNPs).

UNTHSC was in a unique position to develop a program to field test newer technologies that could provide a more effective means for the identification of human remains and missing persons. The availability of family reference samples and unidentified remains that have been already processed with current STR multiplex systems and mtDNA haplotypes data for HV1 and HV2 was highly advantageous. We

evaluated several of these newer technologies as a potential solution to the typing issues previously described. The utilization of one or more of the new miniSTR systems should provide a cost effective means of obtaining additional genetic information to augment the data generated using our conventional methodologies. This study has provided the UNT System Center for Human Identification with a strategy and additional technologies to maximize our ability to obtain genetic information for the identification of missing persons and unidentified human remains.

Field Test of Current Technology Used in the Identification of Unidentified Remains

EXECUTIVE SUMMARY

The analysis of skeletal remains and the identification of missing persons present a number of significant challenges not routinely encountered in typical forensic casework. Skeletal remains typically contain very limited amounts of nuclear DNA which is often highly degraded or contaminated with potent inhibitors of the polymerase chain reaction (PCR). The number of core STR loci that routinely provide a reliable profile is often limited. A review of data from our lab during 2004 and 2005 has shown that in 55% of the cases, 10 or fewer of the core CODIS loci gave an interpretable profile (n=133). In fact in approximately one third of the remains tested, profiles were obtained with 5 or fewer core loci. As a result, the evaluation of DNA from unidentified human remains relies heavily on the sequence analysis of the hypervariable regions 1 and 2 (HV1 and HV2) of the mitochondrial DNA (mtDNA) genome to glean whatever traces of genetic information that may be available from the sample. In recognition of these unique challenges, the FBI has established a separate version of CODIS to address the requirements of the identification of human remains and missing persons. This version, referred to as CODIS+mito consists of three indexes which contain the profiles from: Unidentified Human Remains; Missing Person Direct Reference Samples; and Biological Relatives of a Missing Person (Family Reference Samples). For these samples, CODIS+mito allows the simultaneous upload of the core STR loci along with the mitochondrial DNA (mtDNA) haplotype.

Unlike most forensic cases, the identification of skeletal remains is not typically accomplished through a direct comparison with a known reference sample, but rather through familial searching with both mtDNA and STR profiles from one or more close biological relatives. The quality and quantity of DNA recovered from a skeletal remain and the availability of the appropriate close biological relatives of the missing person will significantly affect the ability to make an identification. For a portion of human remains, the profiles generated for the core CODIS STR loci and current mtDNA sequence analysis of HV1 and HV2 are not sufficient to provide a level of statistical certainty

required for a positive identification. The causes for this issue are twofold: 1) the current STR panels often yield incomplete or no results from skeletal remains that have been exposed to the elements for varying length of time due to the high state of sample degradation; and 2) although mtDNA typing is successful for greater than 95 percent of skeletal samples, common haplotypes found in Caucasian and African American populations result in numerous adventitious associations between remains and family reference samples that cannot be properly evaluated.

Although mtDNA typing has proven to be highly successful on severely degraded or ancient remains, the nature of its inheritance and reduced genetic diversity within some subpopulations often does not provide the magnitude of statistical certainty required to support a positive identification. A major limiting factor in the utilization of mtDNA in forensic identification has been the large number of common haplotypes observed through the limited sequence analysis of only HV1 and HV2. The current SWGDAM /FBI mtDNA databases indicate that more than 50 percent of Caucasians and 20 percent of African Americans fall within two major haplogroups. Within the H haplogroup, 14% of Caucasians present one of two common haplotypes. Similarly, among the L1 haplogroup found among individuals of African origin, 6 % of African Americans present the 2 most common haplotypes. Evaluation of sequence data in the coding region of the mitochondrial genome has provided a good indication that assaying these regions may aid in subdividing the common haplotypes and reducing the number of adventitious matches seen in mtDNA searches.

The University of North Texas System Center for Human Identification (UNTSCHI) was in a unique position to develop a program to field test newer technologies that may provide a more effective means for the identification of human remains and missing persons. Funding provided by this cooperative agreement has allowed the evaluation of several newer technologies as a potential solution to the typing issues previously described. The availability of a large pool of both family reference samples and skeletal remains that have STR data for the core CODIS loci and sequence data for the HV1 and HV2 regions of the mtDNA genome was highly advantageous. Our goal was to develop a strategy based upon all available current technologies to maximize

our ability to derive useful genetic information for the identification of missing persons and unidentified human remains.

Several approaches have been suggested to enhance the level of genetic data obtained from challenged samples. Attempts to obtain profiles from nuclear DNA may prove to be the most informative for the kinship analysis used to ascertain familial relationships. Two methods attempted in the WTC victim identification process may provide the greatest likelihood of obtaining valuable additional data: the utilization of reduced size or "mini" Short Tandem Repeats (miniSTRs) marker systems; and the utilization of a panel of nuclear single nucleotide polymorphism (SNP) markers. The major tactic of both of these methods is the interrogation of smaller DNA target regions. Although the CODIS core STR loci have been previously adapted to a miniSTR format, none were commercially available. The use of additional, non-CODIS loci may be necessary to provide the increased resolving power required in missing persons and mass disaster scenarios. The nuclear SNP markers developed and utilized by Orchid Cellmark to aid in the WTC victim identification process proved useful especially when coupled with other forms of limited genetic data. Large panels of independent SNP markers in development by groups such as the European SNP*for*ID Consortium and Dr. Kenneth Kidd (Yale University) will be evaluated, in the near future, for their effectiveness in mass disaster/human identification scenarios.

The technologies evaluated through this cooperative agreement have been limited to the investigation of: (1) primer sets and multiplex systems for both the CODIS core STR loci as well as other non-CODIS loci which have been shown to work efficiently on limited amounts of degraded DNA; and (2) alternative methods to obtain genetic information from mtDNA. One or more of these new technologies may provide a reliable means of reducing the number of adventitious matches and increase the number of successful identifications.

Evaluation of non-CODIS miniSTR Loci

The NIST Human Identity Team supported through funding from NIJ has been instrumental in the development of miniSTR systems for the analysis of highly degraded and challenging DNA samples. With their support, we have tested two of their non-

CODIS miniSTR triplex systems. The NC01 (D10S1248, D14S1434, and D221045) and the NC02 (D1S1677, D2S441, and D4S2364) triplexes have proven very useful in obtaining additional genetic information from skeletal remains submitted to our laboratory. Databases for the African American, Caucasian, and Southwest Hispanic (Mexican American) population groups have been generated. Statistical analysis has shown that the non-CODIS loci conform to the same genetic requirements exhibited by the 13 Core CODIS STR loci.

DNA extracted from a set of bone samples that had been previously amplified with the Applied Biosystems Profiler Plus™ ID and COfiler™ STR multiplex systems were analyzed with the NC01 and NC02 miniSTR systems. The random match probability (RMP), power of discrimination (PD), and probability of exclusion (PE) were calculated using initially the data generated from the CODIS loci and then using the NC01 and NC02 miniSTR data. With the addition of the NC01 and NC02 miniSTRs, the power of discrimination was increased by several orders of magnitude for all samples. The random match probability was greatly reduced for all samples with the miniSTR data, and the probability of exclusion was increased for all bone samples. In almost all cases, the NC01 and NC02 systems provided complete profiles where as the CODIS loci had only provided partial profiles.

The European DNA community (EDNAP, ENFSI) has recently recommended the inclusion of three of the non-CODIS miniSTR loci (D10S1248, D2S441, D22S1045) to the core European STR markers. The availability of commercially manufactured kits containing these new miniSTR loci will be critical for validation and CODIS acceptance in the United States. Applied Biosystems has expressed an interest in the incorporation the NC01 and NC02 (non-CODIS) loci into their next generation miniSTR kits, and have further indicated that they would like the UNT System Center for Human Identification to participate in future development and validation studies. UNTSCHI will continue to utilize these two non-CODIS miniSTR systems in both casework and family reference samples where applicable.

Evaluation of the Applied Biosystems AmpFℓSTR® MiniFiler™ PCR Amplification System

At the same time that we began to work with the NC01 and NC02 mini STR's developed by NIST, we were approached by Applied Biosystems (AB) to collaborate with them on the development of a novel miniSTR multiplex system containing a number of the core CODIS loci. The loci chosen by AB coincided with the loci that were lost most often in the bones samples analyzed in our lab using the Profiler Plus® ID and COfiler® STR multiplex systems. The 9-plex miniSTR multiplex system which has been given the name MiniFiler™ utilizes AB's 5-dye chemistry and will simultaneously amplify the loci: D13S317, D7S820, Amelogenin, D2S1338, D21S11, D16S539, D18S51, CSF1PO, and FGA. With the exception of D2S1338 each of the STR loci plus Amelogenin are part of the core CODIS loci. The amplicons generated with this new miniSTR multiplex have been reduced significantly in comparison to their sizes in the Identifiler® multiplex system. The amplicons are less than 210 bp with the exception of the larger alleles detected at the FGA locus the amplicons sizes. AB has used their proprietary, non-nucleotide mobility modifiers to adjust the electrophoretic distribution of the fluorescently labeled amplicons on the CE instruments so that there is no overlap between the loci with the same fluorescent dye.

The reduced size amplicons have proven to be extremely advantageous in the analysis of DNA recovered from skeletal remains. AB has also demonstrated that a newly designed buffer system used in conjunction with their new miniSTR primers will significantly increase the amplification of DNA that contains a variety of inhibitors. Humic acid and other contaminants typically found in organic-rich soil samples complicate the isolation and PCR amplification of DNA from bones. Amplification of DNA samples with the new MiniFiler™ system is not inhibited at humic acid concentrations shown to adversely affect the current commercially available STR multiplex systems.

DNA extracts obtained from skeletal remains that were previously analyzed with the Profiler Plus® ID and COfiler® STR multiplex systems were amplified using the new MiniFiler™ System. The MiniFiler™ System was able to consistently provide a profile at the same loci in which the current systems failed. The MiniFiler™ System has been

utilized in several cases in which a putative match had been suggested by mtDNA and supported by the results from a limited number of STR loci. In each of theses cases, the MiniFiler™ System has provided additional STR data that has significantly increased the statistical certainty of the match, which has resulted in a successful identification.

As of April 2007, the Applied Biosystems MiniFiler™ kit has been released for sale to the forensic community. However, for use at the National level of CODIS (NDIS), the MiniFiler™ kit must be completely validated and the data must be reviewed and accepted by the NDIS board. The Chair of SWGDAM has asked that the Missing Persons, Mass Disaster subcommittee select laboratories to generate the data for NDIS approval. The UNTSCHI, along with the California Department of Justice, Missing Persons Laboratory, the Armed Forces DNA Identification Laboratory (AFDIL), the FBI DNA Unit 1 Laboratory, and the Office of the Chief Medial Examiner (OCME) of New York will assume the responsibility for conducting the appropriate validation and submittal of data for NDIS approval of the MiniFiler™ kit. The necessary validation experiments for NDIS approval are underway in our laboratory.

Field Test of New Technologies for Mitochondrial DNA (mtDNA) Analysis

Since the early 1990's, mtDNA analysis has been used to aid in the identification of human remains that defy analysis with more conventional nuclear DNA typing strategies. The convention adopted in the forensic community has been to sequence two hypervariable regions (HV1 and HV2) within the control region (D-loop). This constitutes approximately 610 base pairs (bp) of the 16569 bp genome. The sequence data obtained for HV1 and HV2 is then compared to a standard sequence referred to as the revised Cambridge Reference Sequence (rCRS; Andrews, et al. 1999; Anderson, et al. 1981). Differences relative to the rCRS provided a means for the development of a mtDNA database. As more extensive databases were developed for HV1 and HV2 it was shown that among Caucasians, several common haplotypes exist, with the two most common types accounting for approximately 14% of Caucasians. These two haplotypes contain only two or three differences from the rCRS. Although useful, the sequence analysis of only HV1 and HV2 provides a relatively low power of discrimination in comparison with nuclear DNA, and often fails to provide the statistical support required

by the medical legal community to make positive identifications. To augment the discriminatory power of HV1 and HV2 sequencing, the analysis of the entire mtDNA genome (coding regions) has been suggested.

UNTSCHI developed an initial mtDNA research database consisting of African American (N=105), Caucasian (N=98), Chinese (N=130), and Hispanic (N=98) mtDNA sequence data. All samples have complete sequence data for the HV1 (16024-16374) and HV2 (64-302) regions. Out of these samples, roughly 82% of the samples had unique haplotypes. The remaining 18% of samples shared the same mtDNA profile with at least one other individual. The most frequent haplotype was seen in 10 individuals (~9% Caucasian population). Two different systems were evaluated to increase the discrimination power of mtDNA analysis: the Applied Biosystems mitoSEQr™ Resequencing System, which uses fluorescence-based cycle sequencing reactions with BigDye Terminator v3.1 Cycle Sequencing Kit; and the Affymetrix GeneChip® Human Mitochondrial Resequencing Array 2.0 based on DNA hybridization to a chip. Whole mitochondrial genome sequence evaluations of 40 individuals (9 Caucasians, 12 Hispanics, and 19 African Americans) have been completed with both testing platforms. A significant amount of data has been generated for an additional 30 individuals. The data generated was utilized to determine if common haplotypes can be separated by expanding the search for polymorphisms into the whole mitochondrial genome.

Previous attempts to differentiate common haplotypes have been made through the development of SNP assays targeting polymorphic sites in the mtDNA coding region. We compared the utility of these assays relative to the whole genome sequencing performed in our lab by evaluating a total of 233 SNP positions described in the literature. These 233 SNP positions interrogate known polymorphisms in the coding region, specifically those sites differentiating sub-haplogroups. The evaluation of these 233 SNP positions would distinguish approximately 33% of the common, shared Caucasian and Hispanic haplotypes and 61% of the common, shared African American haplotypes. This approach, however valuable for anthropological and evolutionary studies, is generally ineffective for the task of differentiating individual mtDNA lineages that would fall within particular sub-haplogroups. The private polymorphisms that can be detected through full sequence analysis represent sites that fluctuate in the terminal

branches of mtDNA lineages. These polymorphisms are generally not selected for the development of SNP assays since their unpredictable occurrence makes the development of typing systems for these sites impractical.

Applied Biosystems mitoSEQr™ Resequencing System

The mitoSEQr™ Resequencing System was designed to discover variants in the human mitochondrial genome. The mitoSEQr™ Resequencing Set, mitoALL™, amplifies the complete mitochondrial genome using 46 primer pairs. Each primer pair generates a resequencing amplicon (RSAs). These resequencing amplicons are then cycle sequenced using Applied Biosystems BigDye® Terminator Ready Reaction Mix v3.1. The raw sequence data was collected using Sequencing Analysis 5.1.1 software (Applied Biosystems). The whole genome RSA sequences were aligned and analyzed using Sequencher software (Gene Codes Corp.). The Sequencher software generated a list of differences between assembled sample sequences and the Revised Cambridge Reference Sequence (rCRS).

The overall success in obtaining data with the Applied Biosystems mitoSEQr™ Resequencing System was >95%. Missing data was present in 3.2% of the Caucasian samples, 4.1% of the Hispanic samples, and 6.3% of the African American samples. Unsuccessful amplification/sequencing was almost twice as high in the African American population samples. A possible explanation for this occurrence is the presence of greater numbers of undocumented polymorphisms in the RSA primer regions in the African American population. This was not unexpected since the primer design was initially based on the sequence generated for the Caucasian (H2 haplotype) rCRS. The sequence analysis conducted in our laboratory was conducted using a beta version of the mitoSEQr™ Resequencing System. Applied Biosystems has made improvements to the kit by adding back degenerative primers for the African American population and is now commercially available. Further analysis of mtDNA genomes using the mitoSEQr™ Resequencing System is in progress to reamplify population samples with the additional degenerative primers in an attempt to obtain any missing sequence data.

The amplification and sequence analysis of an entire mitochondrial genome using the mitoSEQr™ Resequencing System is time consuming, labor intensive, and consumes

a significant amount of sample DNA. The integration of robotic systems for sample preparation and sequence set-up is critical for the implementation of the mitoSEQr™ Resequencing System. Several steps in the overall process have been automated in our laboratory. The extraction of DNA from buccal swab family reference samples using a robotic system has been validated and implemented in our laboratory. The DNA from an individual is aliquoted for each of the 46 amplicons and the amplification reactions are prepared using an additional robotic platform. However, to be useful, the analysis of the large volume of sequence data requires both better software and more seamless integration.

Affymetrix GeneChip® Human Mitochondrial Resequencing Array 2.0

The Affymetrix GeneChip® Human Mitochondrial Resequencing Array 2.0 interrogates the entire 16569 bp of the mitochondrial genome on a single array. The development of the mtDNA sequencing array was based upon the *in situ* synthesis of oligonucleotide probes using Affymetrix's standard resequencing array tiling strategy with eight unique 25-mer probes per base position. Each 25-mer probe is varied at the central position to incorporate each possible nucleotide—A, G, C, or T — allowing for the detection of both known and novel SNPs. The reference sequence used for the synthesis of the Oligonucleotide probes was selected from the public MITOMAP database (www.mitomap.org). The Affymetrix Array 2.0 also contains additional tiling for many of the common variants in the HV1 (nucleotide positions 16024-16365) and HV2 (nucleotide positions 73-340) regions selected from the FBI database.

Following the extraction of mtDNA from a sample, the GeneChip® Human Mitochondrial Resequencing assay requires the amplification of the entire genome in three reactions. However, traditional primer strategies may be employed for highly degraded samples. The three large amplicons were fragmented and then labeled with Terminal deoxynucleotidyl Transferase (TdT). Samples were then hybridized to the GeneChip® Human Mitochondrial Resequencing Array 2.0 according to Affymetrix's recommendations for the 169-array format. The arrays were washed and stained using the GeneChip® Fluidics Station 450. The GeneChip® Scanner 3000 was used to scan the Array in conjunction with the GeneChip® Operating Software.

Multiple samples were initially run with the GeneChip® Arrays in order to evaluate and train the data interrogation routines of the Affymetrix GeneChip® system. The whole genome sequence data was used as a concordance check to the overall base calling and SNP detection of the Affymetrix system. The evaluation of the Affymetrix GeneChip® Human Mitochondrial Resequencing Array 2.0 included 10 mtDNA samples (9 Caucasian, and 1 self-declared African American) with the same shared or common Caucasian haplotypes. Length polymorphisms in the homopolymeric regions of the control region could potentially differentiate some of these samples, however, these sequence differences would be considered inconclusive by current forensic interpretation policies. Sequence analysis of the entire mitochondrial genome successfully discriminated all but two of these ten samples.

The Affymetrix GeneChip® Human Mitochondrial Resequencing Array 2.0 requires significant optimization before it can be considered for casework. The Affymetrix system's analytical software was not fully optimized and currently does not apply the correct nomenclature and base position assignments as compared to the rCRS. Some of the limitations that we have observed to date include: the recommended amplification of large PCR products which are not likely to work with evidentiary material; an unacceptable number of ambiguous base calls; difficulties in the analysis of regions with polycytosine and other stretches of the same bases; current software analytical tools are cumbersome and require excessive data manipulation; and instrument parameters that are not sufficiently optimized. Although the manufacturer claims that the GeneChip® Human Mitochondrial Resequencing Array 2.0 can detect heteroplasmic substitutions, our data indicates the majority of heterozygous calls were false as compared to the sequence data which was confirmed in both the forward and reverse directions. Additionally, the technique can tile for insertions and deletions but no software tools exist to interpret these. The advantages of the technique include: a reduction in both time and labor, a lower cost than currently utilized sequencing methodologies, and once the software tools are optimized, it should be much easier to process and analyze data. With a Quality Threshold setting of 3.0, the comparison of the Affymetrix data with direct confirmed sequencing results demonstrated good concordance in base calling with the lowest number of ambiguities. As with other new

technologies, we believe that this product could be considered first for databasing of reference samples. Further discussions and testing are scheduled in an attempt to optimize the most effective analysis parameters that will yield reliable mtDNA results.

The Affymetrix GeneChip® Human Mitochondrial Resequencing Array 2.0 has the potential to significantly increase the utility of mitochondrial DNA testing for Human Identification and forensic casework. The GeneChip could streamline the labor-intensive process of mtDNA sequencing that is used today. The GeneChip generates sequence information for the entire mitochondrial Genome and therefore provides much more information than the 610 bases of the Control Region routinely sequenced in casework today. Data from the entire mitochondrial genome is extremely valuable in differentiating a significant portion of individuals with common haplotypes.

A major limiting factor of both the Applied Biosystems mitoSEQr™ Resequencing System and Affymetrix GeneChip® Human Mitochondrial Resequencing Array 2.0 stems from the amplicon size targeted in both assays. In each system the entire mitochondrial genome is amplified using amplicons much larger than what would be practical for challenged forensic samples (i.e. approximately 600 bp amplicons for the Applied Biosystems assay and between 4,000 and 7,000 bp for the Affymetrix assay). Testing the assays with reference sample DNA facilitated the evaluation of the degree of information gained from interrogating the entire mitochondrial genome for forensic identification purposes. An amplification strategy must be developed to be able to harness this information in evidentiary samples and unidentified remains. Neither assay is appropriate at this time for use on degraded or limited DNA samples. We have discussed the potential for the development of smaller amplicon primer sets with both kit manufacturers as one possible way to address this issue.

Award: 2004-DN-BX-K214

Field Test of Current Technology Used in the Identification of Unidentified Remains

Introduction

The Combined DNA Index System (CODIS) has become one of the most important tools available to law enforcement to both solve and deter future crimes. CODIS consists of two primary indexes utilized by approximately 180 accredited public sector crime laboratories throughout the United States. As of February 2007 the Convicted Offender index has over 4,200,000 profiles and the Forensic Crime Scene index has over 167,000 samples typed. In order for samples to be uploaded into the National DNA Index System (NDIS) component of CODIS the profiles must be generated utilizing NDIS accepted Short Tandem Repeat (STR) multiples kits such as the Applied Biosystems AmpFℓSTR: Profiler Plus™, COfiler™ and Identifiler™; and the Promega PowerPlex™ 16 systems. Each of these kits analyzes all or a subset of the 13 CODIS core STR loci. For a convicted offender profile to be entered into NDIS, all 13 core STR loci must be successfully typed. For forensic crime scene evidentiary samples all 13 core STR loci must be attempted and a minimum of 10 of the core STR are required for upload and searching. In a typical forensic case, the identification of the source of a crime scene sample is made by a direct comparison with a sample from a known individual or through a search of the convicted offender index. The frequency of occurrence of an evidentiary profile when all 13 core loci have been typed is typically in the range of 10^{-15} to 10^{-18}. The corresponding match probabilities are so large that the chance of observing anyone other than an identical twin is so remote, that the identification of the source of the evidentiary sample for all practical purposes is a certainty.

The analysis of skeletal remains and the identification of missing persons present a number of significant challenges not routinely encountered in typical forensic casework. Skeletal remains typically contain very limited amounts of nuclear DNA which is often highly degraded or contaminated with potent inhibitors of the polymerase chain reaction (PCR). The number of core STR loci that routinely provide a reliable profile is often limited. The evaluation of DNA profiles from unidentified human

remains relies heavily on the typing of the mitochondrial DNA (mtDNA) hypervariable regions 1 and 2 (HV1 and HV2) to glean whatever traces of DNA may still be available from the sample. In recognition of these unique challenges, the FBI has established a separate version of CODIS to address the requirements of the identification of human remains and missing persons. This version, referred to as CODIS+mito consists of three indexes which contain the profiles from: Unidentified Human Remains; Missing Person Direct Reference Samples; and Biological Relatives of a Missing Person (Family Reference Samples). For these samples, CODIS+mito allows the simultaneous upload of the core STR loci along with the mitochondrial DNA (mtDNA) haplotype. The identification process for missing persons and skeletal remains has relied primarily upon the comparison of both the STR profiles and mitochondrial DNA (mtDNA) haplotypes.

Unlike most forensic cases, the identification of skeletal remains is not typically accomplished through a direct comparison with a known reference sample, but rather through familial searching with both mtDNA and STR profiles from one or more close biological relatives. The quality and quantity of DNA recovered from a skeletal remain and the availability of the appropriate close biological relatives of the missing person will significantly affect the ability to make an identification. For a significant portion of human remains, the profiles generated for the core CODIS STR loci and current mtDNA sequence analysis of HV1 and HV2 are not sufficient to provide a level of statistical certainty required for a positive identification. The causes for this issue are twofold: 1) the current STR panels often yield incomplete or no results from skeletal remains that have been exposed to the elements for varying length of time due to the high state of sample degradation; and 2) although mtDNA typing is successful for greater than 95 percent of skeletal samples, common haplotypes found in Caucasian and African American populations result in numerous adventitious associations between remains and family reference samples that cannot be properly evaluated.

All of the NDIS acceptable multiplex systems produce amplicons greater than 200 base pairs (bp) for the majority of the 13 core CODIS STR loci. However, the analysis of DNA isolated from skeletal remains demonstrates that in many cases only a small percentage of the recovered DNA is of this size or greater. The large scale DNA typing efforts used in the identification of the victims from the World Trade Center (WTC)

disaster and other "Mass Disasters" has generated considerable knowledge regarding the effectiveness of the current commercially available STR loci used for the analysis of highly degraded and compromised samples. The data suggest that the STR kits in their current configurations are clearly not optimal for the analysis of these types of samples and those routinely encountered in the analysis of skeletal remains.

Although mtDNA typing has proven to be highly successful on severely degraded or ancient remains, the nature of its inheritance and reduced genetic diversity within some subpopulations often does not provide the magnitude of statistical certainty required in suggesting a positive identification. A major limiting feature in the utilization of mtDNA in forensic identification has been the large number of common haplotypes. The current SWGDAM /FBI mtDNA databases indicate that more than 50 percent of Caucasians and 20 percent of African Americans fall within two major haplogroups. Within the H haplogroup, 16% of Caucasians present one of two common haplotypes. Similarly, among the L1 haplogroup found among African origin individuals, 6 % of African Americans present the 2 most common haplotypes. These data are derived from sequence databases from the control region of the molecule, specifically HV1 and HV2. Evaluation of sequence data in the coding region of the mitochondrial genome has provided a good indication that assaying these regions may aid in subdividing the common haplotypes and reducing the number of adventitious matches seen in mtDNA searches.

Additional genetic data, either through the subdivision of common mtDNA haplotypes or through the successful analysis of one or two additional STR loci, can often mean the difference between samples that have a high degree of certainty in their association or adventitious matches that, if reported, could lead to a misidentification. The University of North Texas (UNT) System, Center for Human Identification was in a unique position to develop a program to field test newer technologies that may provide a more effective means for the identification of human remains and missing persons. Alternatively, one or more of these technologies may provide a cost effective means of obtaining additional genetic information to augment the data generated using our conventional methodologies. We proposed to evaluate these newer technologies as a potential solution to the typing issues previously described. The availability of family

reference samples and unidentified remains that have been previously processed with commercially available, NDIS approved, multiplex systems for the 13 core CODIS STR loci and sequence data for the HV1 and HV2 regions of the mtDNA genome was highly advantageous. Our goal was to develop a strategy based upon all available current technologies to maximize our ability to derive useful genetic information for the identification of missing persons and unidentified human remains.

Several approaches have been suggested to enhance the level of genetic data obtained from challenged samples. Attempts to obtain profiles from nuclear DNA may prove to be the most beneficial for the kinship analysis used to ascertain familial relationships. Two methods attempted in the WTC victim identification process may provide the greatest likelihood of obtaining valuable additional data are the utilization of reduced size or "mini" Short Tandem Repeats (miniSTRs) marker systems and the utilization of nuclear single nucleotide polymorphisms (SNPs) panels. The major tactic of both of these methods is the interrogation of smaller DNA target regions as compared to the standard STR kits in use by the forensic community. Although several of the CODIS core STR loci have been previously adapted to a miniSTR format, none were commercially available. In addition, the use of non-CODIS loci may be necessary to provide the increase the resolving power required in missing persons and mass disaster scenarios. The nuclear SNP markers developed and utilized by Orchid Cellmark to aid in the WTC victim identification process proved useful especially when coupled with other forms of limited genetic data. Panels of 50-70 independent SNP markers are in development by other groups such as the European SNP*for*ID Consortium and will be evaluated in mass disaster/human identification scenarios. In addition, several high throughput typing methods to implement nuclear SNP panels have been developed, however, in some cases the instrumentation required is specialized and outside of the reach of most crime laboratories. SNP typing methods using currently available platforms such as capillary electrophoresis (CE) instrumentation may provide the simplest means for implementing these markers in forensic laboratories.

The decision as to which technologies to evaluate during the period of this cooperative agreement have been limited to: 1. the utilization of miniSTR primer sets for both the CODIS core STR loci as well as other non CODIS loci which work efficiently

on limited amounts of degraded DNA to generate reduce amplicons products; and 2. investigation of alternative methods to obtain genetic information from mtDNA.

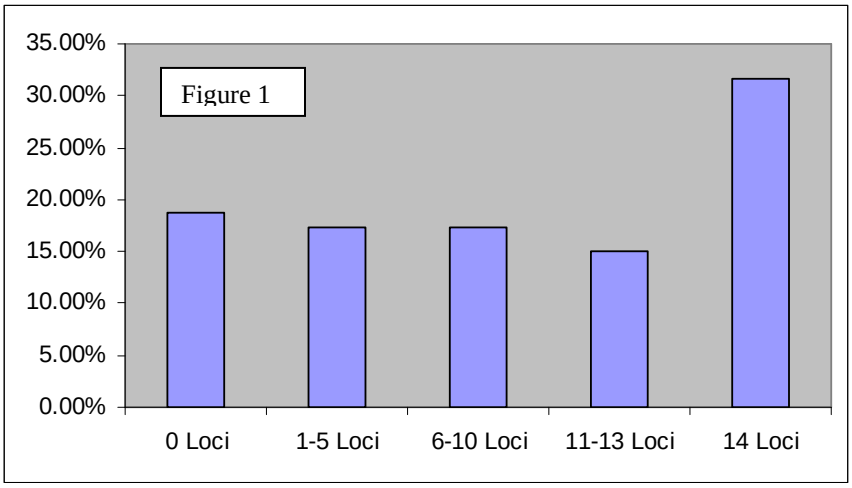
With the increased utilization of the missing persons databases at the state and national levels, the percentage of cases that cannot be conclusively resolved will also increase due to the limited systems currently validated and in use by the forensic community. Concurrently, as the number of samples (both human remains and family reference samples) submitted to laboratories for analysis increases, the number of specimens identified with a common haplotype will increase due to their inherent frequencies. This will result in a large number of adventitious matches that must be reviewed and very often ignored due to a lack of sufficient genetic data. The implementation of one or more of these new technologies may have provide a reliable means of reducing the number of adventitious matches and increase the number of successful identifications.

The focus of this cooperative agreement with NIJ was to address the aforementioned issues by conducting a thorough evaluation of the various technologies and genetic loci which may be used to augment the investigations of unidentified remains samples and the identification of missing persons. The technologies chosen were evaluated on previously processed bone samples for which varying levels of success in obtaining nuclear DNA data has been obtained, as well as samples in which the only genetic data obtained were common mtDNA haplotypes. Family reference samples exhibiting the common mtDNA haplotypes were analyzed in order to better quantify the effect that additional genetic data will have on individualizing such samples if only mtDNA data is first presented.

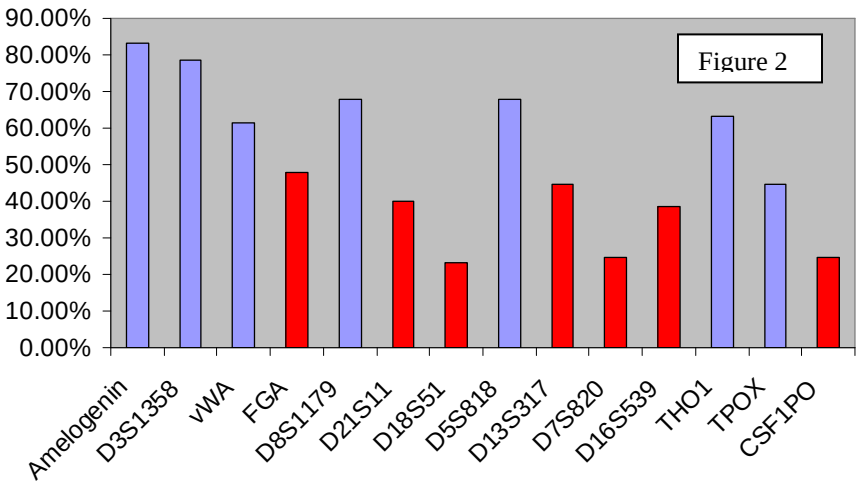
Section I Field Test of New MiniSTR Technology

The Problem Associated with the Analysis of DNA Recovered from Skeletal Remains with Commercially-Available STR Multiplex Systems

The majority of bones originally submitted to the UNT System, Center for Human Identification were from skeletal remains that had been retained by medical examiners, coroners, or law enforcement agencies typically between 10 and 30 years. Unable to make an identification using methodologies routinely available, these remains just sat untested for many years. The DNA recovered from these skeletal remains was often highly degraded and present in very limited quantity. Commercially-available NDIS accepted multiplex kits that have been utilized in our lab for traditional forensic casework samples often fail to obtain full profiles from DNA obtained from these bone samples and routinely show both allelic and locus dropout. The STR profiles generated from DNA recovered from 133 sets of skeletal remains which had been analyzed in 2004 and 2005 were reviewed. These DNA samples had been amplified with Applied Biosystems Profiler Plus™ ID and COfiler™ STR multiplex systems. The data shown in Figure 1 demonstrated that approximately 55% of the DNA profiles generated yielded 10 or fewer of the core STR loci.



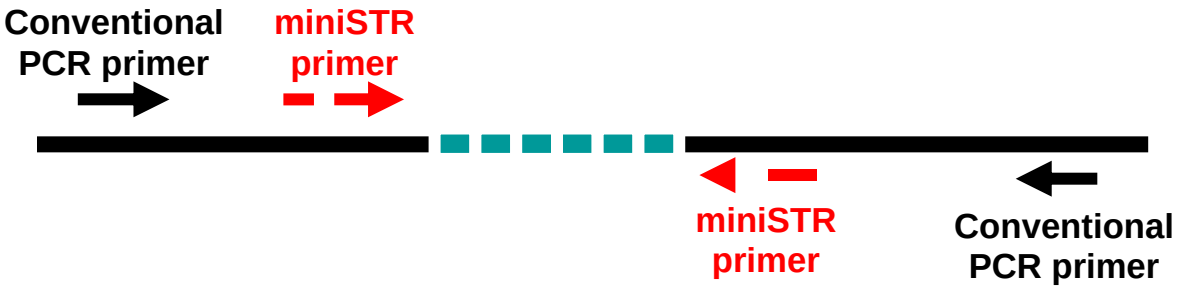
We further characterized these DNA profiles in order to determine the frequency at which each of the 13 core STR loci provided an interpretable result (Figure 2).



These results clearly demonstrate that the loci which generate larger amplicons are the ones that most often fail to produce an interpretable profile. Therefore, amplification of DNA recovered from these skeletal samples with currently available STR kits often results in partial DNA profiles, with limited statistical value, or worse no genetic data. The necessity for additional and/or modified STR typing systems for the analysis of skeletal remains is obvious.

The ability to obtain genetic information from degraded samples is often enhanced through the use of smaller PCR products. Smaller STR amplicons have been created by moving the forward and reverse PCR primers in close to the STR repeat region (Figure 3).

Figure 3



The use of these "miniSTR" primers can help provide information from degraded DNA samples that typically produce partial profiles and a total loss of information from larger STR amplicons. Butler, Shen, and McCord (2003) published the sequences for miniSTR

primers that reduce the amplicons for all 13 CODIS STR loci along with the D2S1338, Penta D, and Penta E loci (research supported by an interagency agreement with the NIST Office of Law Enforcement Standards and a research grant to Bruce McCord). Jim Schumm and his colleagues at the Bode Technology Group utilizing information provided by John Butler and NIST to develop the BodePlex 1 (D13S317, D21S11, D7S820, D16S539, and CSF1PO) and BodePlex 2 (TPOX, FGA, D7S820, and D18S51) miniSTR multiplexes. BodePlex 1 and BodePlex 2 were used to analyze bone and tissue samples that were part of the WTC investigation (Schumm, J.W., Wingrove, R.S., Douglas, E.K., 2004). In addition to miniSTR primer sets for the 13 CODIS core STR loci, the NIST Human Identity Genetic team has developed miniSTR primer sets for a large number of non-CODIS loci. Investigations of miniSTR's by the UNT System, Center for Human Identification has been focused on two projects: 1. Two triplex non-CODIS (NC) mini STR marker sets referred to as NC01 and NC02 developed by John Butler and Michael Coble and their collaborators at NIST, and 2. A new miniSTR multiplex system containing CODIS accepted STR loci developed by Applied Biosystems.

Evaluation of Non-CODIS miniSTR Loci

Coble and Butler (2005) published data on two new miniplex systems, Miniplex01 and Miniplex02. Each of these miniplex systems are composed of 3 Non-CODIS loci. Miniplex01, referred to as NC01, contains the loci D10S1248, D14S1434, and D221045. Miniplex02, now referred to as NC02, contains the loci D1S1677, D2S441, and D4S2364. With the exception of D221045, which is a tri-nucleotide repeat, all are tetra-nucleotide repeats. Drs. John Butler and Michael Coble and their associates at NIST were very gracious in providing us with both sequence information for their primers as well as aliquots of reagents to get us started. We purchased fluorescently-labeled and unlabeled oligonucleotide primers for NC01 and NC02 loci. We optimized the primer concentrations and amplification conditions based upon the recommendations of NIST. Allelic ladders for each of the loci in NC01 and NC02 multiplexes were developed. These allelic ladders were extremely important in order to make reliable allele calls with these new STR loci. Figure 4 shows the development of allelic ladders for the 3 miniSTR

loci comprising the NC01 triplex. The alleles shown comprise the most common alleles detected in the three major population groups (African American, Caucasian, and Mexican American (SW Hispanic)).

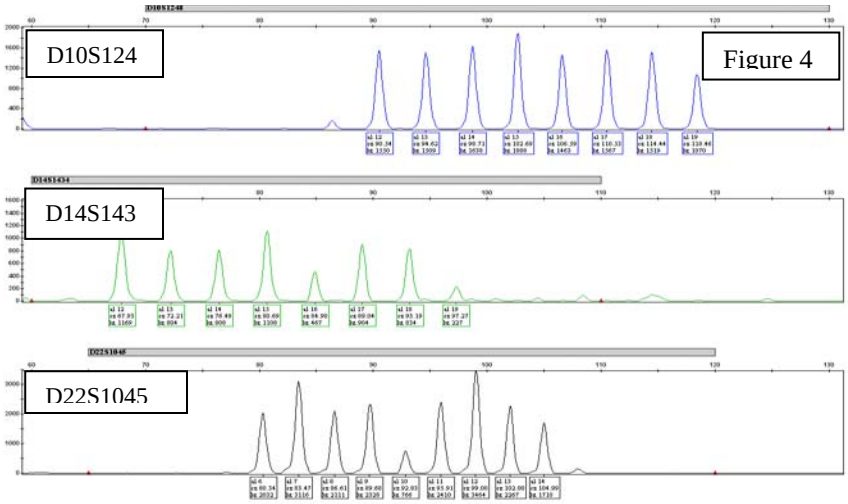
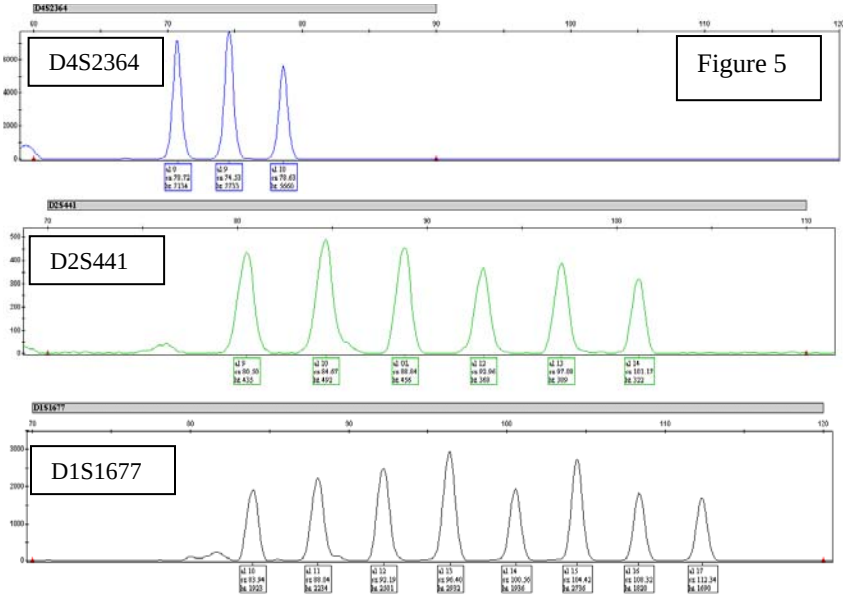


Figure 5 shows the development of allelic ladders for the 3 miniSTR loci comprising the NC02 triplex. To date we have only been able to detect 3 alleles at the D4S2364 locus. Coble and Butler have reported two other alleles that have only been seen once in their population groups. Databases with approximately 200 individuals from each population group (African American, Caucasian, and Mexican American (SW Hispanic)) were developed. Each of these individuals had been previously typed with the core CODIS STR loci.



The population samples were evaluated for their allele distribution, heterozygosity, and Hardy-Weinberg Equilibrium (HWE). Allele frequencies for each miniSTR locus in each population are summarized in Table 1. Butler et al. on their website (http://www.cstl.nist.gov/div831/strbase/miniSTR.htm) and at a poster by Becky Hill at the Promega 17th International Symposium on Human Identification (Nashville, TN), October 10-12, 2006 have changed the nomenclature for the allele designations for several of the loci within the NC01 and NC02 triplexes. The allele designation nomenclature for the D22S1045 have increased by +3 repeats from what was originally published by Coble and Butler (2005); the allele designation nomenclature for the D10S1248 has been reduced by -1 repeat; the allele designation nomenclature for the D14S1434 has been reduced by -4 repeats; the allele designation nomenclature for the D1S1677 has been increased by +1 repeat; and the allele designation nomenclature for the D4S2364 has been reduced by -1 repeat. The corrected population frequencies for the NC01 and NC02 loci are shown in Table 1. The macro for making the automated allele calls on the CE instruments will be modified to reflect the corrected allele designations for each of these loci.

Most markers showed similar levels of variation, except the D4S2364 locus, which showed only four different alleles across all population samples tested. These results are similar to those obtained by Coble and Butler (2005), who found only five alleles across the three major U.S. populations for D4S2364. There are no CODIS loci that have this low level of allele variation. The lowest number of alleles observed in a population for the CODIS loci was eight, observed at TPOX, THO1, and D16S539 (Budowle et al., 1999). All other miniSTR loci exhibited at least eight different alleles within a given population, which is consistent with observations for the CODIS loci.

Each population displayed Hardy-Weinberg Equilibrium at all miniSTR loci, except the African American population at D14S1434 with a p value of 0.0306, based on the Exact Test with 3200 shufflings (data not shown). This is consistent with the results obtained by Coble and Butler who found that only the African American population deviated from HWE at D14S1434 in their population database. After applying the Bonferroni correction, these departures were no longer significant, with the corrected significance level of 0.0003. The highest level of heterozygosity was observed in the

African American population at D22S1045, with a value of 0.8227 (Table 2). The lowest level of heterozygosity was also observed in the African American population at D4S2364, with a value of 0.4752 (Table 2). This is also consistent with previously published findings of Coble and Butler.

Allele frequencies, heterozygosity values, and departure from HWE were also calculated for the 13 CODIS loci in all populations (data not shown). The lowest observed heterozygosity in this dataset was in the Caucasian population at TPOX with a value of 0.6197, while the highest was observed in the African American population at FGA with a value of 0.8676. These heterozygosity values were generally consistent with those observed by Budowle et al. who found a value of 0.670 in the Caucasian population at TPOX and a value of 0.872 in the African American population at FGA. Most loci were in HWE for all populations except the Hispanic population at CSF1PO and the Caucasian population at vWA with values of 0.0172 and 0.0284 respectively. After applying the Bonferroni correction, all populations were found to be in HWE at all loci, which is consistent with previously published data.

An interclass correlation analysis was conducted to detect any potential linkage disequilibria between alleles for pair-wise comparisons between all loci (CODIS and miniSTR) for the three populations using the Exact Test with 3200 shufflings. With a significance level of 0.05, multiple miniSTR and CODIS loci appear to be in linkage disequilibria for all populations (Table 3). The Caucasian population showed nine significant deviations from expected. Three of these involved pair wise comparison between miniSTR loci and CODIS loci, while the remaining six were comparisons between CODIS loci. The African American population exhibited seven deviations. Of these, one deviation occurred between miniSTR loci, four occurred between miniSTR loci and CODIS loci, and only two significant departures occurred between the CODIS loci. The Hispanic population had eight significant departures, with one between miniSTR loci, one between miniSTR and CODIS loci, and the remaining six deviations were between CODIS loci (Table 3).

The levels of departure seen in the CODIS loci for the African American and Caucasian samples are comparable to previously published data (Budowle et al.).

Table 1- Allele Frequency distributions for six MiniSTR loci in the three major U.S. population groups (199 Caucasians, 205 African Americans, and 195 Hispanics).

NC01

D10S1248				D14S1434				D22S1045			
Allele	Cauc.	Afr. Am.	Hisp.	Allele	Cauc.	Afr. Am.	Hisp.	Allele	Cauc.	Afr. Am.	Hisp.
9	0.0000	0.0000	0.0026	9	0.0000	0.0074	0.0000	8	0.0000	0.0025	0.0000
10	0.0000	0.0049	0.0000	10	0.1206	0.2079	0.1000	10	0.0025	0.0369	0.0026
11	0.0025	0.0345	0.0051	11	0.0553	0.0272	0.0205	11	0.1005	0.1256	0.0434
12	0.0503	0.1305	0.0332	12	0.0327	0.0792	0.0410	12	0.0101	0.0517	0.0026
13	0.2915	0.2438	0.2194	13	0.4447	0.2351	0.3410	13	0.0000	0.0074	0.0026
14	0.3065	0.2586	0.4031	14	0.3090	0.4233	0.4744	14	0.0402	0.0764	0.0230
15	0.1859	0.1970	0.2398	15	0.0226	0.0173	0.0154	15	0.3693	0.2759	0.4668
16	0.1307	0.0911	0.0765	16	0.0025	0.0025	0.0077	16	0.3844	0.1626	0.3699
17	0.0276	0.0394	0.0204					17	0.0804	0.2291	0.0714
18	0.0050	0.0000	0.0000					18	0.0101	0.0320	0.0128
19	0.0000	0.0000	0.0000					19	0.0025	0.0000	0.0051

NC02

D4S2364				D2S441				D1S1677			
Allele	Cauc.	Afr. Am.	Hisp.	Allele	Cauc.	Afr. Am.	Hisp.	Allele	Cauc.	Afr. Am.	Hisp.
7	0.0000	0.0000	0.0000	7.2	0.0025	0.0000	0.0000	10	0.0052	0.0025	0.0000
8	0.1802	0.1559	0.2359	9	0.0051	0.0050	0.0000	11	0.0130	0.0025	0.0026
9	0.5533	0.7054	0.5077	10	0.2146	0.0965	0.4149	12	0.0725	0.0916	0.1094
10	0.2640	0.1386	0.2538	11	0.3308	0.3738	0.2887	13	0.2332	0.2129	0.1745
11	0.0025	0.0000	0.0026	11.3	0.0025	0.0000	0.0000	14	0.3679	0.2847	0.3828
				12	0.0732	0.1782	0.0696	15	0.2306	0.2921	0.2630
				12.3	0.0000	0.0025	0.0000	16	0.0570	0.0668	0.0521
				13	0.0328	0.0545	0.0155	17	0.0130	0.0297	0.0156
				14	0.2753	0.2525	0.1881	18	0.0078	0.0099	0.0000
				15	0.0581	0.0322	0.0232	19	0.0000	0.0074	0.0000
				16	0.0025	0.0025	0.0000				
				19	0.0025	0.0025	0.0000				

Table 2 - Population statistics for the six miniSTR loci.
The significant deviations from Hardy-Weinberg Equilibrium are highlighted in bold.
Results are based on 199 Caucasians, 205 African Americans, and 195 Hispanics.

NC01									
	D10S1248			D14S1434			D22S1045		
	Cauc.	Afr. Am.	Hisp.	Cauc.	Afr. Am.	Hisp.	Cauc.	Afr. Am.	Hisp.
Observed Heterozygosity	0.7387	0.8177	0.7143	0.6583	0.6733	0.7077	0.7085	0.8227	0.6582
Expected Heterozygosity	0.7681	0.8092	0.7248	0.6823	0.7178	0.6459	0.6998	0.8197	0.6408
Power of Discrimination	0.9090	0.9330	0.8789	0.8623	0.8751	0.7893	0.8592	0.9384	0.7973
Mean Power of Exclusion	0.4906	0.6324	0.4507	0.3667	0.3881	0.4403	0.4416	0.6418	0.3665
Power of Exclusion	0.9062	0.9339	0.8769	0.8451	0.8735	0.8094	0.8565	0.9424	0.8949
HWE p-values	0.7891	0.5866	0.2050	0.4369	**0.0306**	0.2581	0.6797	0.2659	0.7763

NC02									
	D4S2364			D2S441			D1S1677		
	Cauc.	Afr. Am.	Hisp.	Cauc.	Afr. Am.	Hisp.	Cauc.	Afr. Am.	Hisp.
Observed Heterozygosity	0.6061	0.4752	0.5949	0.7980	0.7277	0.7113	0.7202	0.7574	0.7083
Expected Heterozyogosity	0.5912	0.4588	0.6223	0.7615	0.7533	0.7048	0.7474	0.7781	0.7406
Power of Discrimination	0.8821	0.8307	0.8920	0.8924	0.8975	0.8585	0.8994	0.9159	0.8952
Mean Power of Exclusion	0.2982	0.1667	0.2848	0.5953	0.4724	0.4460	0.4602	0.5226	0.4413
Power of Exclusion	0.7653	0.6561	0.7838	0.9016	0.8991	0.8588	0.8945	0.9125	0.8882
HWE p-values	0.9747	0.8744	0.0978	0.1794	0.1413	0.7366	0.4709	0.6056	0.1534

Table 3 - Interclass correlation test analysis that yielded significant departures from expected. Results are based on Fisher exact test analysis with 3200 shufflings.

Population	Locus Pair	P value
Caucasian	D10S1248/D21S11	0.0178
	D1S1677/D21S11	0.0494
	D1S1677/TH01	0.0109
	D13S317/D18S51	0.0406
	D13S317/vWA	0.0169
	D16S539/TH01	0.0294
	D21S11/D8S1179	0.0003
	D21S11/FGA	0.0213
	D5S818/vWA	0.0319
African American	D14S1434/D2S441	0.0209
	D14S1434/D16S539	0.0469
	D14S1434/TPOX	0.0047
	D2S441/D8S1179	0.0406
	D2S441/D7S820	0.0266
	D13S317/D7S820	0.0038
	D3S1358/D5S818	0.0356
Hispanic	D10S1248/TH01	0.0203
	D14S1434/D2S441	0.0200
	D13S317/D7S820	0.0481
	D13S317/TPOX	0.0259
	D16S539/D7S820	0.0213
	D18S51/D5S818	0.0278
	D5S818/D7S820	0.0488
	FGA/TH01	0.0216

The Hispanic population deviations found here between the CODIS loci exceed that observed by Budowle et al. who found only two. However, all of the pair wise comparisons calculated here are above the corrected level of significance using the Bonferroni correction. Therefore, it would be accurate to conclude that all of the loci are unlinked in the populations tested.

The mean power of exclusion and power of discrimination were evaluated for each locus in each population using PowerStatsV12 by Promega. The power of exclusion was also calculated using the formula recommended by the National Research Council II. The power of discrimination was high for all populations and all miniSTR loci (Table 2). The lowest level observed was in the Hispanic population at D14S1434 (0.7893), while the highest level was observed in the African American population at D22S1045 (0.9384) (Table 2). The same descriptive statistics were calculated for the 13 CODIS loci for the populations tested (data not shown). The summary statistics were comparable to previously published data (Budowle et al.).

DNA extracts from 5 sets skeletal remains that were previously amplified with Applied Biosystems Profiler Plus® ID and COfiler® STR multiplex systems were used to evaluate the ability of the non-CODIS, miniSTR loci to increase the power of discrimination and probability of exclusion for an individual sample. The random match probability, power of discrimination, and probability of exclusion were calculated using only CODIS profile data and then again with the miniSTR data added to the profile (Table 4). With the addition of the NC01 and NC02 miniSTRs, the power of discrimination was increased by several orders of magnitude for all samples (Table 4). The random match probability (RMP) was greatly reduced for all samples when the miniSTR data was used in the calculations (Table 4). For example, DNA extracted from bone sample 2432.1A had previously yielded only six of the thirteen CODIS loci. With these six loci, the random match probability was 1 in several million for the three major population groups. With the addition of the NC01 and NC02 miniSTR loci, the random match probability for the various population groups was reduced to 1 in 1 trillion (Table 4). The probability of exclusion was increased for all bone samples. An example of this increase can be seen with bone sample 2719.1BC. Using only the partial profile obtained with the CODIS loci, the probability of exclusion was 0.99. Therefore, only

approximately 99% of the population could be excluded as having contributed the observed profile. With the addition of the non-CODIS miniSTR loci, the probability of exclusion was increased two orders of magnitude to 0.9999, meaning that greater than 99.99% of the population could be excluded from having contributed the profile observed in this bone sample.

Table 4 - Random match probability, probability of exclusion, and probability of discrimination for five bone samples typed with CODIS loci and NC01 and NC02 miniSTR triplexes.

	2355.1D			2423.1A		
	Cauc.	Afr. Am.	Hisp.	Cauc.	Afr. Am.	Hisp.
RMP without miniSTR loci	2.3458E+19	5.3488E+19	4.5991E+17	6.9268E+06	4.4607E+11	1.6316E+07
RMP with miniSTR loci	1.9508E+24	1.0827E+25	8.2881E+21	4.4235E+11	2.4293E+11	3.4431E+12
PE without miniSTR loci	0.999999999998	>0.999999999999	0.99999999995	0.99996	0.99995	0.99996
PE with miniSTR loci	>0.999999999999	>0.999999999999	0.999999999999	0.99999995	0.9999997	0.99999998
PD without miniSTR loci	>0.999999999999	>0.999999999999	>0.999999999999	0.99999986	0.999999999998	0.99999994
PD with miniSTR loci	>0.999999999999	>0.999999999999	>0.999999999999	0.999999999998	0.999999999996	0.9999999999997

	2432.1A			2719.1BC		
	Cauc.	Afr. Am.	Hisp.	Cauc.	Afr. Am.	Hisp.
RMP without miniSTR loci	8.9285E+06	2.9955E+08	5.5100E+06	1.4953E+11	3.2069E+11	8.6963E+06
RMP with miniSTR loci	1.6362E+12	3.7964E+13	2.8877E+12	9.3733E+15	5.6266E+16	2.1594E+15
PE without miniSTR loci	0.999992	0.9999995	0.99998	0.99999996	0.99999999994	0.994
PE with miniSTR loci	0.99999996	0.999999998	0.99999995	0.99999999999988	0.99999999999989	0.99998
PD without miniSTR loci	0.99999989	0.999999997	0.99999982	0.999999999993	0.999999999997	0.99999989
PD with miniSTR loci	0.9999999999994	0.99999999999997	0.9999999999997	>0.999999999999	>0.999999999999	>0.999999999999

	2908.3ABC		
	Cauc.	Afr. Am.	Hisp.
RMP without miniSTR loci	1.1689E+06	1.1152E+06	2.4403E+06
RMP with miniSTR loci	3.2383E+12	9.4211E+09	1.0121E+10
PE without miniSTR loci	0.9991	0.99994	0.9993
PE with miniSTR loci	0.99998	0.99999984	0.999998
PD without miniSTR loci	0.99999914	0.99999910	0.99999959
PD with miniSTR loci	0.9999999999997	0.99999999989	0.9999999999

Utilizing the amplification conditions recommended by Coble and Butler (ie 32 cycles), we were able to obtain extremely good sensitivity with both the NC01 and NC02 triplex systems. However, we did see considerable peak height imbalance. Shown in Figure 6 (NC01) and Figure 7 (NC02) are the profiles with 125 pg of 9947A control DNA at 30 cycles. The amplification reactions at 30 cycles provide excellent sensitivity with more consistent peak height ratios.

Figure 6

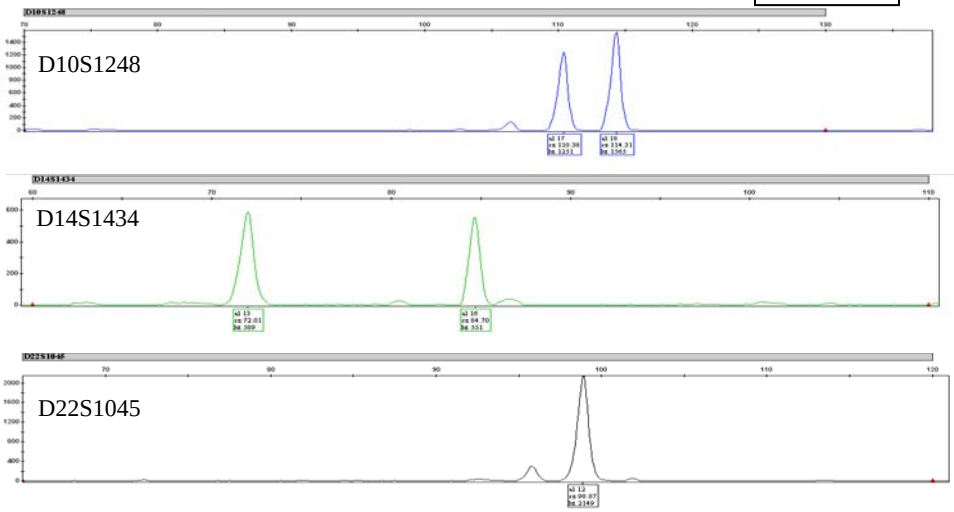
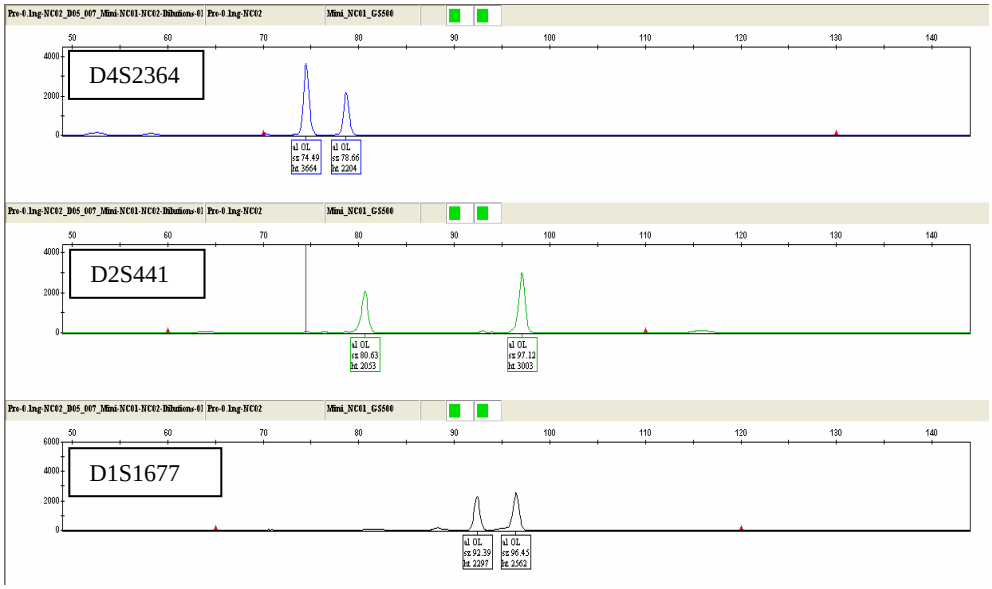


Figure 7



The UNT System, Center for Human Identification has standardized the number of cycles in the amplification reaction for the NC01 and NC02 systems at 30 cycles. The NC01

and NC02 triplexes using the amplification parameters with 30 cycles allow us to consistently detect profiles with 62.5 pg of input DNA with more consistent peak height ratios. Profiles with as little as 15 pg of DNA generated a limited number of the miniSTR allelic data.

The fluorescently-labeled primers initially purchased for the NC01 and NC02 triplexes revealed the presence of dye blobs. The interpretation of electropherograms containing dye blobs is usually not problematic since these artifact peaks typically have a characteristic flat and wide appearance. There were three dye blobs found in two of the six loci. The D14S1434 locus had two dye blobs that were consistent in all samples tested. One artifact occurs at approximately 85 bp and coincides with an allele. It was observed in every sample however, it could be distinguished from a true allele based upon its very wide and flat appearance and its peak height was only 10-15% of the height of a true allele. In samples with reduced RFU values, the peak height of the dye blob did increase up to 30% of the true allele; however, its appearance was distinguishable from a true allele. The other dye blob found at D14S1434 was approximately 108 bp. This dye blob falls outside of normal allele range and did not interfere with typing samples. The D1S1677 locus exhibited a dye blob at approximately 92 bp, which corresponds with allele 12. As with the other two dye blobs, it had a wide characteristic appearance that easily distinguished it from a true allele. For most samples it was less than 10% of the true allele peak height, but in samples with low RFU values it increased to as much as 30% of the true allele peak height, but could still be distinguished from the true alleles. Additional purification will be required to eliminate dye blobs from these fluorescently labeled primers. The purification of fluorescently labeled primers is a costly and laborious process that is required during the manufacture of commercially available kits.

The NIST Human Identity Team supported through funding from NIJ have been instrumental in conducting research in the development of miniSTR systems for the analysis of highly degraded and challenging DNA samples. With their support, we have tested two of their non-CODIS miniSTR triplex systems. The NC01 (D10S1248, D14S1434, and D221045) and the NC02 (D1S1677, D2S441, and D4S2364) triplexes have proven very useful in obtaining increased amount of genetic information from

skeletal remains submitted to our laboratory. While generating the population databases we observed a few individuals that may have potential primer binding site mutations at the D14S1434 and D10S1248 loci. Sequencing of the DNA from these individuals is needed to confirm these results. In the event that these are true primer-binding site mutations, the primers at these two loci may have to be redesigned or supplemented with additional redundant primers.

We will continue the evaluation and the use of these two miniSTR systems in both casework and family reference samples where applicable. The results of our field tests with these non-CODIS loci have been presented at numerous meetings including the Promega 17th International Symposium on Human Identification (Nashville, TN), October 10-12, 2006 and most recently on February 20, 2007 at the 59th Annual Meeting of the American Academy of Forensic Science; Workshop #18: Missing Persons: Resources. Techniques, and Identification; The UNT System, Center for Human Identification: A Resource for the Identification of Human Remains and Missing Persons; San Antonio, Texas.

In discussions with Applied Biosystems, they have indicated that they are pursuing the development of additional miniSTR multiplex systems that will incorporate most of the NC01 and NC02 (non-CODIS) miniSTR loci. The European DNA community (EDNAP) has done extensive studies on the utility of miniSTR systems for the analysis of degraded samples (Dixon, L.A. et.al. 2006). Gill et al (2006) have recommended the inclusion of three of the non-CODIS miniSTR loci (ENFSI and EDNAP have selected the 3 loci D10S1248, D2S441, D22S1045) to increase the number of core European STR markers. The availability of commercially manufactured kits containing these new miniSTR loci will be critical for the validation and CODIS acceptance. We will continue to work with NIST and AFDIL and others to generate the necessary documentation for submittal to the NDIS board to have these loci accepted for inclusion in CODIS. Their inclusion in CODIS could be instrumental in making future identifications.

Evaluation of the Applied Biosystems AmpFℓSTR® MiniFiler™ PCR Amplification System

Approximately the same time that we began to work with the NC01 and NC02 mini STR's developed by NIST, we were approached by Applied Biosystems (AB) to collaborate with them on the development of a novel miniSTR multiplex system containing a number of the core CODIS loci. The loci chosen by AB coincide with the loci that were lost most often in the bones samples analyzed in our lab using the Profiler Plus® ID and COfiler® STR multiplex systems. (Figure 8). The 9-plex miniSTR multiplex system which has been given the name MiniFiler™, utilizes AB's 5-dye chemistry and will simultaneously amplify the loci: D13S317, D7S820, Amelogenin, D2S1338, D21S11, D16S539, D18S51, CSF1PO, and FGA (Figure 9, ©2006, Applied Biosystems). With the exception of D2S1338 each of the STR loci plus Amelogenin are part of the core CODIS loci. The D2S1338 locus is part of the SGM and Identifiler® multiplex systems. The amplicons generated with this new miniSTR multiplex have been reduced significantly in comparison to their sizes in the Identifiler® multiplex system (Figure 10, ©2006, Applied Biosystems). The amplicons are less than 210 bp with the exception of the larger alleles detected at the FGA locus the amplicons sizes. AB has used their proprietary, non-nucleotide mobility modifiers to adjust the electrophoretic distribution of the fluorescently labeled amplicons on the CE instruments so that there is no overlap between the loci with the same fluorescent dye (Figure 11, ©2006, Applied Biosystems). The reduced size amplicons have proven to be extremely advantageous in the analysis of DNA recovered from skeletal remains. AB has also demonstrated that a newly designed buffer system in conjunction with the new miniSTR primers will significantly increase the amplification of DNA extracted from highly degraded samples that contain a variety of inhibitors. Humic acid and other contaminants typically found in organic-rich soil samples complicate the isolation and PCR amplification of DNA from bones. Figure 12 (©2006, Applied Biosystems) shows that at a concentration of 30 ng/µl of humic acid, the Identifiler® multiplex is completely inhibited, whereas, the new MiniFiler™ system will reproducibly amplify each of the 9-loci. Amplification of DNA samples with the new miniSTR system is not inhibited at humic acid concentrations

Figure 8 Amplification of 500pg of 007 Control DNA with the MiniFiler™ System ©2006, Applied Biosystems

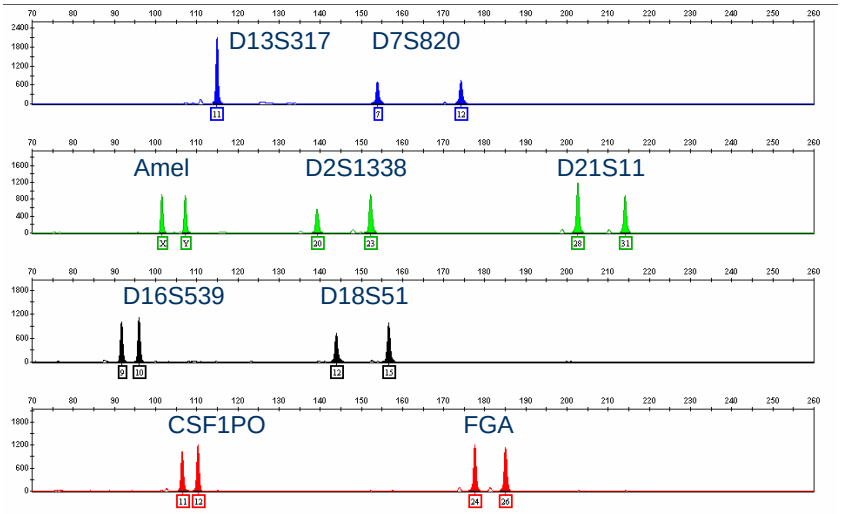


Figure 9 Amplicon Size Range of MiniFiler™ Multiplex in Comparison with Size Range with Identifiler® ©2006, Applied Biosystems

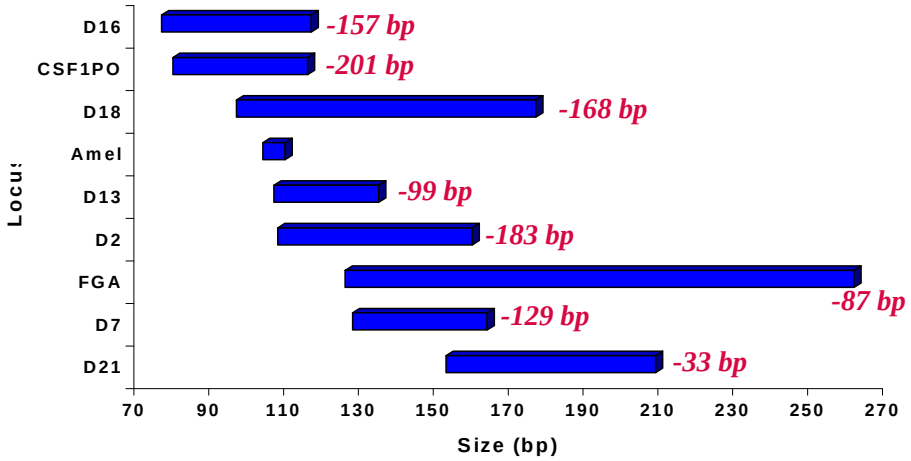
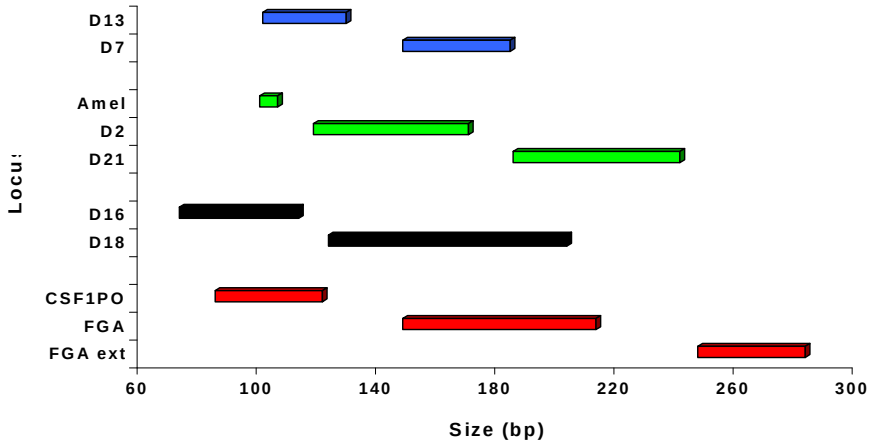
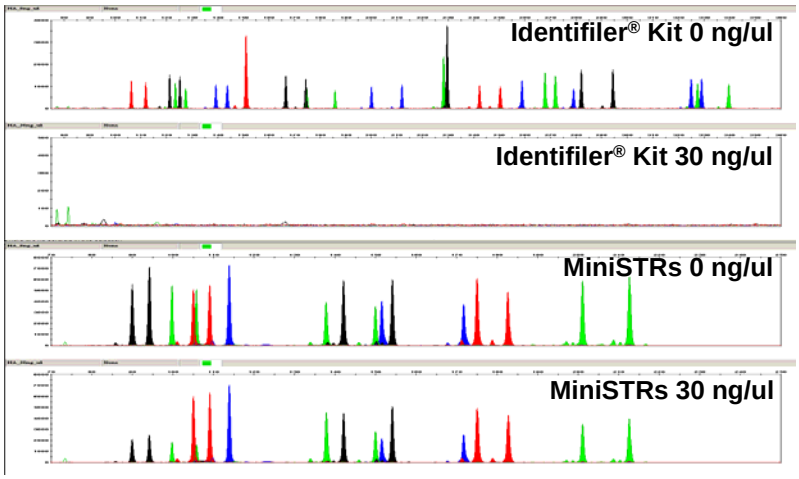


Figure 10 MiniFiler™ Multiplex Configuration using 5-dye Chemistry and Mobility Modifiers, ©2006, Applied Biosystems



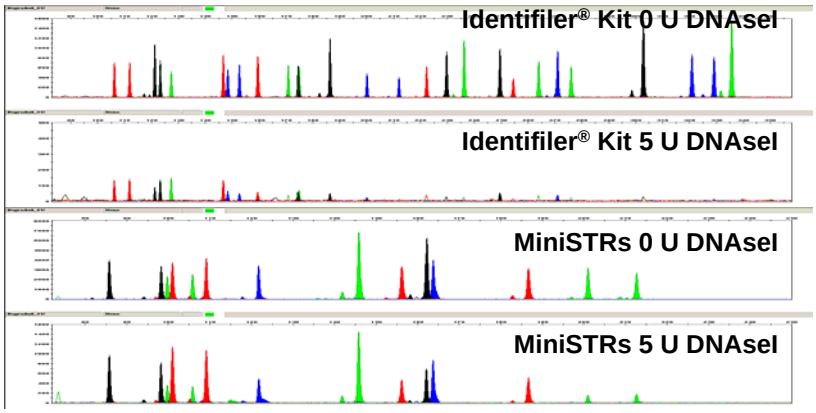
as high as 50 ng/µl (data not shown). The MiniFiler™ system is also much more resistant to inhibition with Heme, another potent inhibitor of *Taq* Polymerase (data not shown).

Figure 11 MiniFiler™ vs Identifiler® Kit Simulated Inhibition with Humic Acid, ©2006, Applied Biosystems



Applied Biosystems simulated the degradation of DNA with various concentrations of DNAse I. Figure 6 (©2006, Applied Biosystems) demonstrates that as a result of reduced size amplicons genetic profiles are possible even at the larger loci.

Figure 12 MiniSTRs vs Identifiler® Kit Simulated Degradation with DNAse I, ©2006, Applied Biosystems

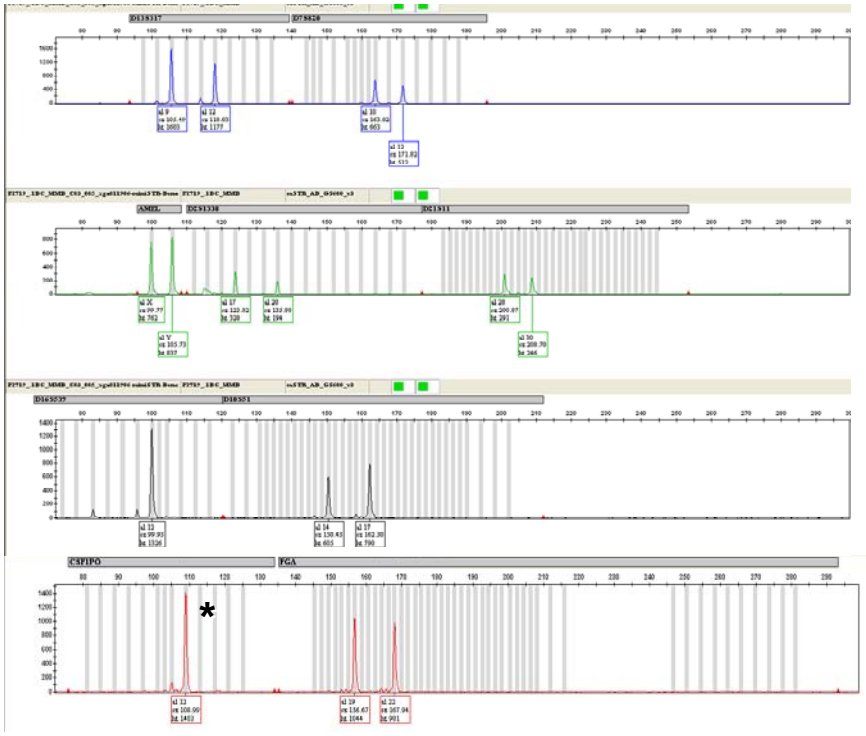


Applied Biosystems has recommended that 30 cycles be used for the amplification of DNA with the new MiniFiler™ System. At this cycle number, the MiniFiler™ System

appears about 2-3 fold more sensitive as compared to Profiler Plus® ID and COfiler® STR multiplex systems (28 cycles recommended with these kits). A complete profile can be obtained with as little as 62.5 pg of DNA. With lower amounts of DNA partial profiles can be detected with the MiniFiler™ System (data not shown). Our laboratory has validated amplification of LCN samples (such as those routinely encountered in extracts from skeletal remains) at 32 cycles with the Profiler Plus® ID and COfiler® STR multiplex systems.

Initially we selected 25 DNA extracts obtained from skeletal remains that were previously analyzed with the Profiler Plus® ID and COfiler® STR multiplex systems were amplified with the new MiniFiler™ System. The MiniFiler™ System consistently provided more genetic information at the common loci. DNA extracted from bone sample F-2719 only yielded 8 of the 13 core STR loci with the Profiler Plus® ID and COfiler® STR multiplex systems. The loci D21S11, D13S317, D7S820, D16S539, CSF1PO, and TPOX failed to amplify. The same extract generated profiles for each of these loci with the exception of TPOX (not contained within the MiniFiler™ System) when amplified with the new MiniFiler™ System (Figure 13). The loci marked with an asterisk are the loci that failed to amplify with conventional systems.

Figure 13 DNA Extracted from Bone Sample F-2719 Amplified with Applied Biosystems MiniFiler™ System



Similar results were obtained with the majority of the other bone samples tested. Four of the core 13 loci failed to amplify (FGA, D21S11, D18S51, and D7S820) with the Profiler Plus® ID and COfiler® STR multiplex systems (Figure 14a). Profiles for these 4 loci were obtained using the new MiniFiler™ System (figure 14b)

Figure 14a Loci Circled Failed to Amplify with Profiler Plus® ID and COfiler® STR Multiplex Systems

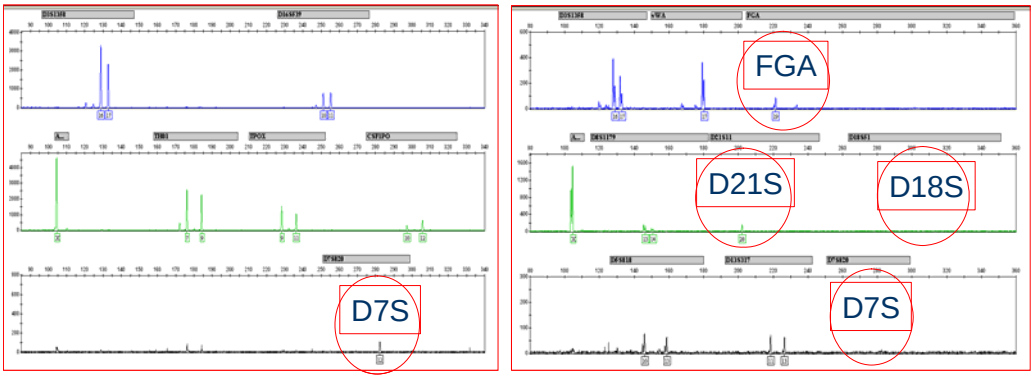
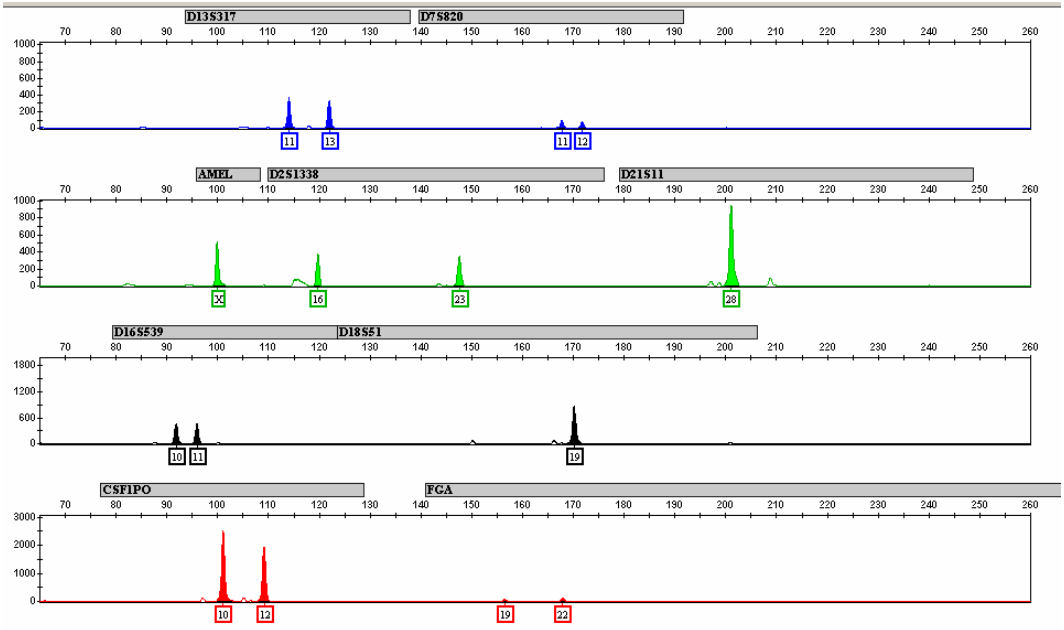


Figure 14b Same Bone sample as in 14a which now produces an interpretable profile with the MiniFiler™ system



The decision to work with Applied Biosystems and not to utilize the primer sequences originally developed by other Butler, Shen, and McCord (2003) or by Schumm, J.W.,

Wingrove, R.S., Douglas, E.K., (2004) was based upon our need as well as the entire forensic community for the availability of a commercially produced kit. A kit that includes miniSTR primers for the CODIS loci could greatly increase their utilization within the forensic community. It is clearly advantageous to have a commercial product developed, validated, and produced under rigorous manufacturing standards, available for use with forensic samples such as the skeletal remains.

The Success of our Field Tests with These New MiniSTR Systems

The success of our field tests with these new miniSTR systems can be demonstrated by their utilization in actual casework. The unidentified body of a young girl who had been shot 3 times was found in a shallow gave outside of Missoula, Montana on Christmas Eve, 1984. The remains were given the name of Debbie Deer Creek, and for more than twenty years the bones of Debbie Deer Creek were kept in a Tupperware box on a shelf. On August 25th, 2004 the National Center for Missing and Exploited Children, working with the Missoula County Sheriff's Office, submitted a femur from the unidentified body of Debbie Deer Creek to the UNT System Center for Human Identification. DNA was extracted from a femoral window and amplified with the Profiler Plus® *ID* and COfiler® AmpF*ℓ*STR® Systems. Sequence analysis of HVI and HV2 within the mtDNA was also performed, and the mtDNA haplotype determined. The mtDNA haplotype and only 8 of the 13 Core STR loci plus Amelogenin could be uploaded into CODIS+mito. The loci D21S11, D18S51, D7S820, D16S539 and CSF1PO failed to amplify.

In March of 2006 a cold hit and potential match was identified in CODIS+mito. The mtDNA profile from the remains of Debbie Deer Creek matched that of the mother of 16 year-old runaway from Vancouver, Washington named Marcella Bachmann. A family reference sample from the mother of Marcella Bachmann was collected in August of 2005 by the Green River Task Force out of Washington State. In November of 2003, Gary Ridgway the "Green River Killer" was convicted of the murder of 48 women, mainly runaways and prostitutes. Ridgway had indicated that there were many other women that he had killed, but could not provide an exact number. Although Marci's description did not match any of the known victims of Ridgway, her file was associated

with the Green River investigation because the time of her disappearance coincided with the Green River murders (1982-1998). Beverly Chilton (Marcella's mother) voluntarily provided a sample with the hope of finding her daughter, who she had not heard from for over twenty years. In addition to the mtDNA match, the 8 loci uploaded from the remains of Debbie Deer Creek shared at least one allele in common with the haplotype of the mother of Marcella Bachmann. The maternity index with only the original 8 loci amplified with Profiler Plus® *ID* and COfiler® Systems was **520**. A maternity index of 520 is considered suggestive of a familial relationship; however, statistically we were not satisfied. The same DNA extract, isolated from the femur, was amplified with the NC01 and NC02 systems developed by NIST and the MiniFiler™ system developed by Applied Biosystems. The loci in NC01 and NC02 systems all provided genetic results, and the loci D21S11, D18S51, D7S820, D16S539 and CSF1PO all amplified with the MiniFiler® system. The results from these additional systems are shown in Figure 15a and 15b. The maternity index with the addition of the miniSTR loci (MiniFiler™ System, NC01, and NC02) was **22,700.** This likelihood ratio was much more suggestive of a strong investigative lead and provided us with the justification to request additional family reference samples. A reference sample from Marcella Bachmann's father and brother were then provided. Based upon the original STR results and those from NC01, NC02 and the Minifiler® systems, the genetic results were approximately 130 million times more likely under the scenario that the unidentified remains of "Debbie Deer Creek" was the child of Beverly Chilton (Marcella's mother) and Mark Bachmann (Marcella's father) as opposed to the remains originating from an unknown, unrelated individual. The body of Debbie Deer Creek was positively identified as that of Marcella Bachmann. Law enforcement officials from Montana are convinced that Debbie Deer Creek was the victim of Wayne Nance, an 8-time serial killer. This cold hit exemplifies a national effort involving several agencies, including the King County Sheriff's Office, the National Center for Missing and Exploited Children, the Missoula County Sheriff's Office and the UNT System Center for Human Identification. The identification of the remains from Marcella Bachmann in April of 2006 demonstrates the potential of the CODIS+mito database, the utility of the miniSTR's and the value of the funding provided to the UNT System Center for Human Identification by NIJ.

Figure 15a NC01 and NC02 miniSTR Profile from the Unidentified Human Remain of Debbie Deer Creek

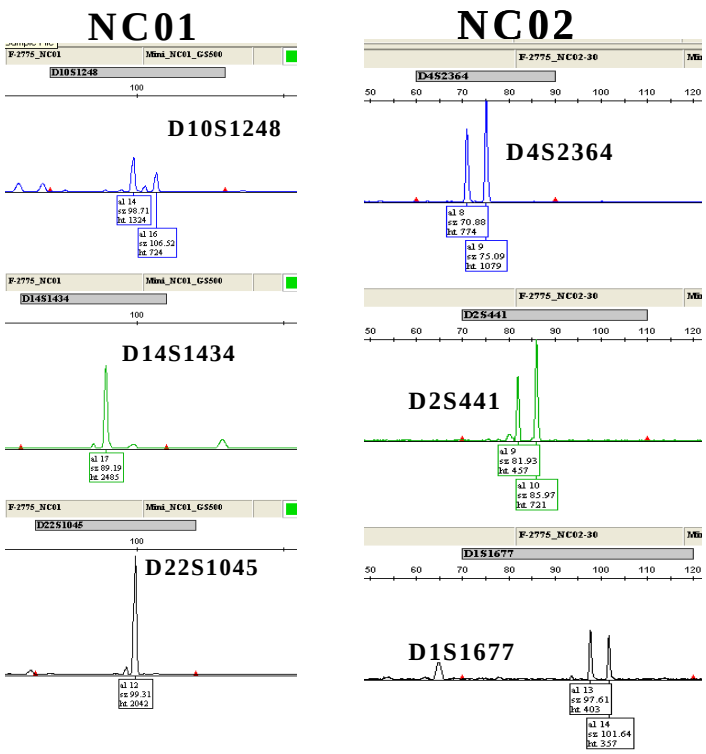


Figure 15b MiniFiler™ Profile from the Unidentified Human Remain of Debbie Deer Creek

Summary of the Field Test of the MiniFiler™ System for the Analysis of Unidentified Human Remains and the Identification of Missing Persons

In mid December, 2006 we received the first production kits manufactured by Applied Biosystems. During the initial manufacturing phase at Applied Biosystems they were only been able to provide us with limited amounts of the MiniFiler™ kit components. We have continued to analyze bone samples that have only yielded limited amounts of genetic data with the conventional Profiler Plus® ID and COfiler® STR multiplex systems. In each of these cases, the MiniFiler™ system has provided additional genetic information. However, for use at the National level of CODIS (NDIS), the MiniFiler™ kit must be completely validated and the data must be reviewed and accepted by the NDIS board. Dave Kaufman, Chairman of SWGDAM has asked that Missing Persons, Mass Disaster subcommittee select laboratories to generate the data for NDIS approval. Our Laboratory (UNT System Center for Human Identification), the California Department of Justice, Missing Persons Laboratory, the Armed Forces DNA Identification Laboratory (AFDIL), the FBI DNA Unit 1 Laboratory, and the Office of the Chief Medial Examiner (OCME) of New York will assume the responsibility for conducting the appropriate validation and submittal of data to NDIS for the approval of the MiniFiler™ kit. Once this has been accomplished the profiles generated with the MiniFiler™ kit can be uploaded into NDIS. Ultimately, the availability of both CODIS and non-CODIS miniSTR systems will increase the likelihood of identifying the skeletal remains. As of April 2007, the Applied Biosystems MiniFiler™ kit has been officially released for sale to the forensic community. The necessary validation experiments for NDIS approval are underway in our laboratory.

Section II Field Test of New Technologies for Mitochondrial DNA (mtDNA) Analysis

Introduction

Mitochondrial DNA (mtDNA) has been in use for over 20 years in the evaluation of population structure, human migration and admixture. Since the early 1990's, mtDNA analysis has been used to aid in the identification of human remains that defy analysis with more conventional nuclear DNA typing strategies. There are several characteristics of the mtDNA genome that have made it useful for human identification: high copy number per cell; circular structure and its location within the mitochondria offers protection from degradation; maternal mode of inheritance; lack of recombination; and a high mutation rate as compared to single copy nuclear DNA. Because of its high copy number and its resistance to degradation, it is often the only source of genetic material remaining in highly challenged samples or limited quantity samples. The maternal mode of inheritance provides the opportunity for a larger pool of reference samples. The high mutation rate provides a high degree genetic of diversity especially within the non-coding, control region of the molecule. These characteristics make the analysis of mtDNA a valuable tool in the identification of skeletal remains and missing persons through a comparison with maternal reference samples. The convention adopted in the forensic community has been to sequence two hypervariable regions (HV1 and HV2) within the control region (D-loop) which constitutes approximately 610 base pairs (bp) of the 16569 bp genome. The sequence data obtained for HV1 and HV2 is then compared to a standard sequence referred to as the revised Cambridge Reference Sequence (rCRS; Andrews, et al. 1999; Anderson, et al. 1981). Differences relative to the rCRS are recorded and provided a means for the development of a mtDNA database. On average 13-18 nucleotide differences are seen between unrelated African Americans, 9-12 differences between Hispanics, but typically only 5-7 nucleotide differences are found between unrelated Caucasians. Fewer observed differences are seen in Caucasians as a result of the selection of a predominantly Caucasian sample for the reference sequence (rCRS). However, a significant portion of the reduced variability can be attributed to the relatively recent evolution of Caucasian populations.

As more extensive databases were developed for HV1 and HV2 it was shown that among Caucasians, several common haplotypes exist with the two most common types accounting for approximately 14% of Caucasians (Allard, et al., 2002). These two haplotypes display only two or three differences from the rCRS (263G, 315.1C and 263G, 309.1C, 315.1C, respectively). Although useful, the sequence analysis of HV1 and HV2 provides a marker system with a relatively low power of discrimination. To augment the discriminatory power of HV1 and HV2 sequencing, the analysis of the entire mtDNA genome has been suggested. The remaining portion of the mtDNA genome, referred to as the Control Region, codes for 13 proteins, 22 tRNAs, and 2 rRNAs. The sequence analysis of the entire mtDNA genome has proven difficult in forensic case work as a result of the significant sample requirement, and the lack of an available, efficient, accurate, and cost effective sequencing technology. Current sequence analysis methodologies utilized for the identification of human remains are very labor intensive and time consuming. This is further complicated by the vary nature of the samples needing analysis. In the absence of useable nuclear DNA, and our in ability to routinely distinguish common haplotypes, the identification of skeletal remains is often inconclusive. There is a clear need for newer technologies to provide the additional discriminatory power to assist in the correct association of unidentified human remains with available family reference samples.

Project Design and Methodology

The University of North Texas System, Center for Human Identification has developed a mtDNA research database consisting of African American (N=105), Caucasian (N=98), Chinese (N=130), and Hispanic (N=98) mtDNA sequence data. All samples have complete sequence data for the HV1 (16024-16374) and HV2 (64-302) regions. Out of these samples, roughly 82% of the samples had unique haplotypes. The remaining 18% of samples shared the same mtDNA profile with at least one other individual. The most frequent haplotype was seen in 10 individuals (~9% Caucasian population). Other haplotypes were seen between 5 individuals (1 haplotype), 4 individuals (3 haplotypes), 3 individuals (7 haplotypes), and 2 individuals (15 haplotypes). Although the analysis HV1 and HV2 is routinely used for identification

purposes, it is clear that these regions do not have enough discriminating power to separate all individuals. Preliminary investigations have been performed to determine if common haplotypes can be separated by expanding the search for polymorphisms into the whole mitochondrial genome. The evaluation of two different systems has been initiated as a part of our field test of newer technologies for mtDNA analysis funded through this cooperative agreement: the Affymetrix GeneChip® Human Mitochondrial Resequencing Array 2.0 based on DNA hybridization to a chip, and the Applied Biosystems mitoSEQr™ Resequencing System, which uses fluorescence-based cycle sequencing reactions with BigDye Terminator v3.1 Cycle Sequencing Kit.

Affymetrix GeneChip® Human Mitochondrial Resequencing Array 2.0

The Affymetrix GeneChip® Human Mitochondrial Resequencing Array 2.0 interrogates the entire 16569 bp of the mitochondrial genome on a single array. The development of the mtDNA sequencing array was based upon the *in situ* synthesis of oligonucleotide probes using Affymetrix's standard resequencing array tiling strategy with eight unique 25-mer probes per base position (Figure 1). Each 25-mer probe is varied at the central position to incorporate each possible nucleotide—A, G, C, or T —allowing for the detection of both known and novel SNPs. The unique design of the probes enables the detection of heteroplasmic mutations (Figure below). The reference sequence used for the synthesis of the Oligonucleotide probes was selected from the public MITOMAP database (www.mitomap.org). The database uses the universally accepted revised Cambridge Reference Sequence (rCRS). As previously described, the profiles (haplotypes) are designated as nucleotide differences from the rCRS. The Affymetrix Array 2.0 also contains additional tiling for many of the common variants in the HV1 (nucleotide positions 16024-16365) and HV2 (nucleotide positions 73-340) regions selected from the FBI database (www.fbi.gov/hq/lab/fsc/backissu/april2002/miller1.htm).

Following the extraction of mtDNA from a sample, the GeneChip® Human Mitochondrial Resequencing assay requires the amplification of the entire genome in only three reactions. However, traditional primer strategies may be employed for highly degraded samples.

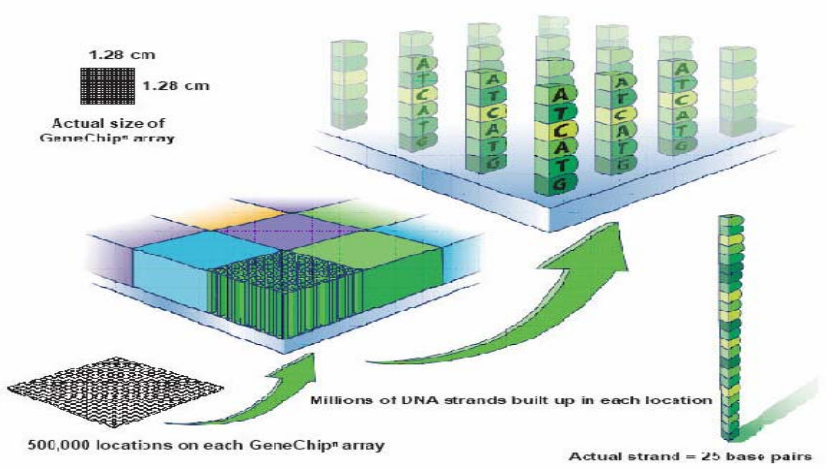


Figure 1 Diagrammatic representation of Affymetrix chip-probe strategy.

Sample Preparation

DNA was extracted from whole blood samples using Phenol-Chloroform Isoamyl Alcohol followed by ethanol precipitation. Samples were then extracted using the Promega DNA IQ system using 7µL of resin for binding and then 60µL of elution buffer to yield ~5ng/µL DNA. Three amplicons were generated using Long-Range PCR optimized for the 169 array format. Each amplification reaction consists of 1X LA-Taq buffer with Mg^{2+}, 0.05U HS LA-Taq (TaKaRa Bio Inc.), 400uM dNTPs, 3uM forward and reverse primer, and 25ng genomic DNA in a total reaction volume of 25µL. Thermal cycling conditions consisted of a 94ºC hotstart for 2 min, 30 cycles: 94ºC denaturation for 15 sec and 68ºC anneal/extension for 8 min followed by a 13 min final extension at 68ºC. PCR amplicons were verified and quantitated using the Agilent system. PCR reactions were pooled according to the *PCR Product Pooling Worksheet* supplied by Affymetrix. Affymetrix suggests that assay performance may be compromised during hybridization if the amplicon concentration varies by more than ten-fold from what is recommended**.** The calculated amount of each PCR amplicon was added to the assay for a final concentration of 111.7 picomolar. Pooled products were purified using 1µL of ExoSap-It per 5µL of pooled product. Samples were again quantitated using absorption spectrophotometry following manual instructions.

Fragmentation and Labeling

The fragmentation reaction consists of 628ng total pooled PCR products in 1X fragmentation buffer and 0.04U fragmentation reagent in a total volume of 20.9µL. Samples were digested on a thermal cycler at 37°C for 35 minutes followed by enzyme deactivation at 95°C for 15 minutes. Proper fragmentation was confirmed by running 2 µL of fragmented product on a 20% TBE PAGE gel stained with SYBR Gold (diluted 1:10,000 in TE) for 15 minutes and visualized on an FMBIO II. Properly fragmented DNA will run between 20-200 bp. Fragmented samples are then labeled with Terminal deoxynucleotidyl Transferase (TdT) using the GeneChip® DNA Labeling Reagent in 1X TdT buffer for a final reaction volume of 26.9µL. Fragment labeling was conducted in a AB 9700 thermalcycler for 2 hours at 37°C, then 15 minutes at 95°C, followed by a 4°C hold.

Hybridization and Detection

Samples were hybridized to the GeneChip® Human Mitochondrial Resequencing Array 2.0 according to Affymetrix's recommendations for the 169-array format. The arrays were washed and stained using the GeneChip® Fluidics Station 450. The Affymetrix staining protocol consisted of a three step process:

1. A Streptavidin Phycoerythrin (SAPE) stain.

2. An antibody amplification step.

3. A final stain with Streptavidin Phycoerythrin (SAPE).

Once stained, the array was filled with Array Holding Buffer prior to scanning. The GeneChip® Scanner 3000 is used to scan the Array in conjunction with the GeneChip® Operating Software.

GeneChip Analysis

The Affymetrix GeneChip® Sequence Analysis Software (GSEQ) 4.0 (Affymetrix, Inc.) was used to generate and evaluate sequence data from the GeneChip® Human Mitochondrial Resequencing Array. GSEQ 4.0 enabled the comparative sequence analysis between questioned samples and the Revised Cambridge Reference Sequence (rCRS).

Applied Biosystems mitoSEQr™ Resequencing System

The mitoSEQr™ Resequencing System is designed to discover variants in the human mitochondrial genome. The mitoSEQr™ Resequencing Set, mitoALL™ (RSS000056015_01), amplifies the complete mitochondrial genome using 46 primer pairs. Each primer pair generates a resequencing amplicon (RSA) (Figure 2). Applied Biosystems has developed an advanced primer design based on a proprietary algorithm that provides primer sets for respective RSAs. The algorithm predicts PCR primer performance while taking into consideration low-complexity repeats, GC content, GC count in localized regions, and the tendency to form triple helices. Primers are optimized for high quality, reproducible resequencing. The primer sequences for each of the RSAs are available from the NCBI Probe database: **http://www.ncbi.nih.gov/genome/probe**

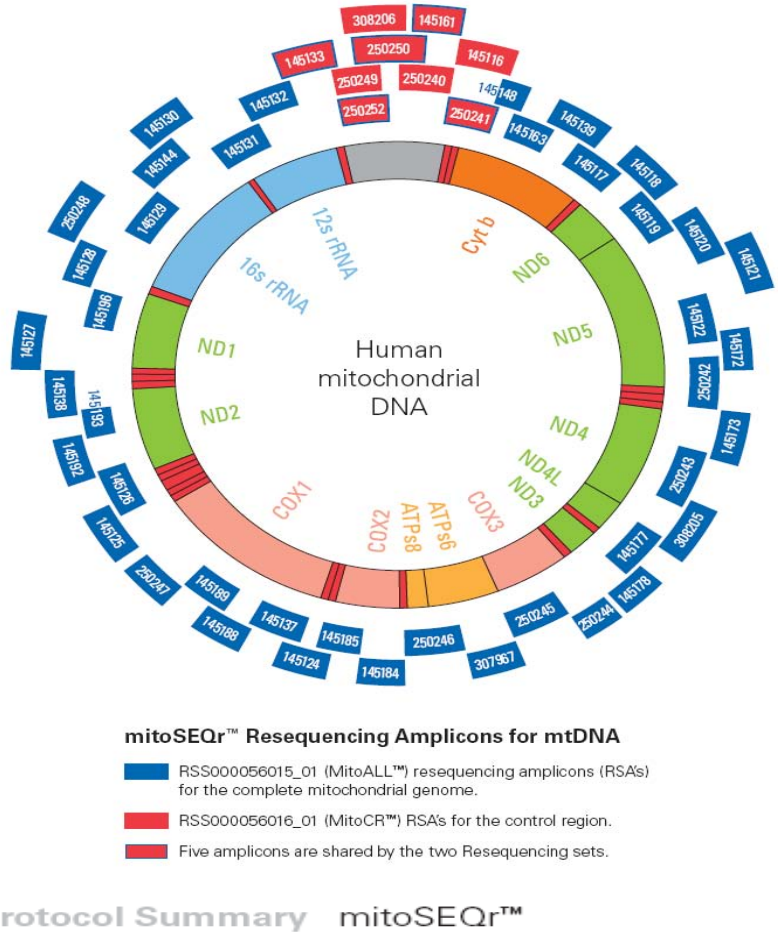
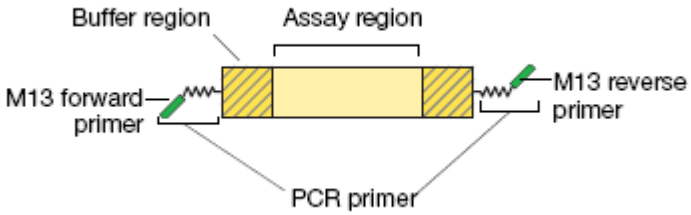


Figure 2 Amplification strategy used for the Applied Biosystems mitoSEQr™ resequencing system

Each primer is tailed with a universal M13 sequence which has been designed for universal PCR and Sequencing conditions (Figure 3).



VariantSEQr and mitoSEQr Resequencing Systems Protocol

Figure 3 Amplicon design incorporating M13 universal sequencing primers.

Sample Preparation and PCR Amplification

DNA was extracted from whole blood samples using Phenol-Chloroform Isoamyl Alcohol followed by ethanol precipitation. Samples were then extracted using the Promega DNA IQ system using 7µL of resin for binding and then 60µL of elution buffer to yield ~5ng/µL DNA. 10 µL PCR reactions consisted of 5.0 µL AmpliTaq Gold® PCR Master Mix (2X); 1.6µL 50% of UltraPure™ glycerol, forward and reverse mitoSEQr™ RSA primers (0.6µM each), and 0.5-1.0ng genomic DNA 1.0 µL (0.5–1.0 ng/µL). Thermocycling conditions with the AB 9700 were: Heat activation 96ºC for 5 min, followed by 40 cycles of 94ºC for 30 sec, 60ºC for 45 sec, and 72ºC for 45 sec. A final extension at 72ºC for 10 min, followed by a 4ºC hold. The PCR products were purified enzymatically by adding 2 µL of ExoSAP-IT® (USB Corporation) with a 37 ºC incubation for 30 min, followed by heat inactivation of the enzyme at 80 ºC for 15 min.

RSA Sequencing

Forward and reverse sequencing reaction master mixes were made with the M13 forward primer and M13 reverse primer respectively (3.2pmol), 0.8 µL of BigDye® Terminator Ready Reaction Mix v3.1, 1.6 µL 5X sequencing buffer, 2µL of the PCR product in a final volume of 10 µL. Thermocycling conditions consisted of heat activation of 96 ºC for 1 min, followed by 25 cycles of 96 ºC for 10 sec, 50ºC for 5 sec, and 60 ºC for 4 min and 4 ºC hold. Performa DTR Ultra 96-Well Plates

(EdgeBioSystems) were used to remove dye terminators, dNTPs, salts and other low molecular weight materials from sequencing reactions.

Data Collection and Analysis

The cycle sequence reactions were electrophoresed on an AB 3130*xl* (POP-6™) Genetic Analyzer with a 36-cm array under the following program: RapidSeq36_POP6_1. Run file; KB_3130_POP6_BDTv3.mob; KB.bcp. Raw data was collected using Sequencing Analysis 5.1.1 software (Applied Biosystems). Sequences were aligned and analyzed using Sequencher software (Gene Codes Corp.). The Sequencher software generates a list of differences between assembled sample sequences and the Revised Cambridge Reference Sequence (rCRS).

Summary of the Two New Technologies for mtDNA Sequence Analysis

Applied Biosystems mitoSEQr™ Resequencing System

Whole mitochondrial genome sequence evaluations of 40 individuals (9 Caucasians, 12 Hispanics, and 19 African Americans) have been completed with both testing platforms. The overall success of the genome sequencing is >95% by the more traditional dideoxynucleotide termination method. Missing data was present in 3.2% of the Caucasian samples, 4.1% of the Hispanic samples, and 6.3% of the African American samples. Unsuccessful amplification/sequencing was almost twice as high in the African American population samples. A possible explanation for this occurrence is the presence of greater numbers of undocumented polymorphisms in the RSA primer regions in the African American population. This is not unexpected since the primer design was based on the sequence generated for the Caucasian (H2 haplotype) rCRS. The sequence analysis conducted in our laboratory was conducted using a beta version of the mitoSEQr™ Resequencing System. Applied Biosystems has made improvements to the kit by adding back degenerative primers for the African American population. Work is in progress to reamplify population samples with the additional degenerative primers in an attempt to obtain any missing sequence data.

In the evaluation of shared or common Caucasian haplotypes, a total of 10 sequences (9 Caucasian, and 1 self-declared African American) from the data set were reviewed for discriminatory capability. Control Region (HV1 and HV2) sequencing for these samples provided little resolving power with the following two haplotypes: 263G, 309.1C, 315.1C (7 individuals); 263G, 315.1C (3 individuals). Although there is a sequence difference dividing this set of samples, length polymorphisms in homopolymeric regions of the control region would be considered inconclusive by current forensic interpretation policies. Sequence analysis of the remainder of the mitochondrial genome successfully discriminated all but two of these ten samples (Table 1). Three individuals were initially sorted into the H1 sub-haplogroup due to the presence of a diagnostic G→A transition at base 3010. These samples could then be further differentiated with 1-3 additional SNPs. The remaining seven individuals fell into the H haplogroup which is defined by polymorphisms of 7028C and 14766C and lacking any diagnostic polymorphisms that define sub-haplogroups. The remaining seven individuals displayed 6 distinct haplotypes, differing at 1 to 7 SNPs. Included in this group is an individual who is self-declared as African American, indicating that the origin of this individual's mtDNA lineage is Caucasian. Additionally, one novel insertion and two positions exhibiting heteroplasmy were documented in these data. An A/G heteroplasmy was detected at position 5031 fell in the ND2 (NADH Dehydrogenase subunit 2) in which A or G result in synonymous codon usage for Leucine. Another A/G heteroplasmy was detected at position 6890 (Cytochrome C Oxidase subunit I) in which A or G result in synonymous codon usage for Glycine.

Attempts to deal with the issue of differentiating common haplotypes have been made through the development of SNP assays targeting polymorphic sites in the mtDNA coding region (Brandstetter et al., 2003, 2006; Quintans, et al., 2004; Grignani et al., 2006; Pereira et al., 2006; Sigurdsson, et al., 2006; Coble et al., 2004; Vallone et al., 2004, Table 7). We compared the utility of these assays to our whole genome sequence results by evaluating all of the targeted SNP positions in these publications (N=233) in order to see if we could individualize our samples based on these SNPs alone. The overall success that could be expected, if we were able to assay all 233 SNPs utilized in the combined assays results in approximately 33% success in distinguishing common,

shared Caucasian and Hispanic haplotypes (Table 8). Success in differentiating African American samples is somewhat higher (61%), however that would be expected considering the greater SNP diversity present in African origin populations. A detailed evaluation, by population group for the listed SNP panels is found in Tables 9-11. The SNP panels, especially that described by Sigurdsson, et al., have effectively captured sites of known polymorphism in the coding region, specifically those sites differentiating sub-haplogroups. This approach, however valuable for anthropological and evolutionary studies, is generally ineffective for the task of differentiating individual mtDNA lineages that would fall within particular sub-haplogroups. These private polymorphisms are generally not selected for the development of SNP assays since their unpredictable occurrence makes the development of typing systems for these sites impractical.

Affymetrix GeneChip® Human Mitochondrial Resequencing Array 2.0

Multiple samples were initially run with the GeneChip® Arrays in order to evaluate and train the data interrogation routines of the Affymetrix GeneChip® system. The whole genome sequence data was used as a concordance check to the overall base calling and SNP detection of the Affymetrix system. Sequencing comparisons between both platforms were evaluated for coverage and concordance and the results of these comparisons are described in Tables 2 - 4. Initial evaluations of the Affymetrix GeneChip® system indicate that the system is not sensitive to the detection of insertions or deletions and commonly incorporated multiple ambiguous calls in homopolymeric stretches. In addition, the Affymetrix system's analytical software is not fully optimized and currently does not apply the correct nomenclature and base position assignments as compared to the rCRS. Although the manufacturer claims that the GeneChip® Human Mitochondrial Resequencing Array 2.0 can detect heteroplasmic substitutions, our data indicates the majority of heterozygous calls were false as compared to the sequence data which was confirmed in both the forward and reverse directions. The number of ambiguous base calls ranged from 2.9 – 23.9% (Table 5). There are numerous options and thresholds for which filtering can be applied in the analysis software. Varying settings, such as Quality Threshold or whether the software should interpret the data as diploid or haploid data yields different levels of resolution, ranging from low resolution

or accuracy to a higher degree of base call confidence (Table 6). At this time, no recommendation can be made regarding an appropriate interpretation setting or range for these variables. From the comparison of the Affymetrix data with direct confirmed sequencing results, a Quality Threshold setting of 3.0 produced the most concordance in base calling with the lowest number of ambiguities. The use of the Diploid detection mode, at this time is unreliable for the detection of mtDNA heteroplasmy. We are working extensively with Affymetrix to improve resolution and base calling efficiency. Further discussions and testing are scheduled in an attempt to optimize the most effective analysis parameters that will yield reliable mtDNA results.

The Affymetrix GeneChip® Human Mitochondrial Resequencing Array 2.0 has the potential to revolutionize mitochondrial DNA testing for Human Identification and forensic casework. It shows promise as an exciting new tool for the forensic DNA analyst. This GeneChip could streamline the labor-intensive process of mtDNA sequencing that is used today. Additionally, since the GeneChip is designed to produce sequence information for the entire mitochondrial Genome – it provides much more information than the 610 to 1,000 bases of the Control Region that are sequenced today. The GeneChip was designed with redundancy within the Control Region since this is the area: with the greatest number of polymorphisms; that has been highly characterized by population geneticists: used by forensic laboratories worldwide. However, the Affymetrix GeneChip® Human Mitochondrial Resequencing Array 2.0 requires optimization of the entire process before it can be considered for casework processing. Some of the limitations that we have observed to date include: extremely large amplicons which are not likely to work with evidentiary material; an unacceptable number of ambiguous base calls; difficulty in results with polycytosine and other stretches of the same bases; software analysis tools that are cumbersome and require excessive manipulation in the data; and instrument parameters that are not yet optimized. We have observed inconsistencies of the reported base positions with minor variations in the analysis parameters used. This technique must be optimized so the accuracy of base calls is improved. Additionally, the technique can tile for insertions and deletions but no software tools exist to interpret these. The advantages of the technique include: a

reduction in both time and labor, lower cost than currently utilized sequencing methodologies in forensic laboratories, and it should ultimately be easier to process and analyze data once the software tools are optimized.

As with other new technologies, we believe that this product could be considered first for databasing of reference samples. At this time, the assay requires an excessive amount of DNA (approximately 10 ng). This quantity of DNA is usually available from reference samples but not typically from skeletal remains or hair samples (typically analyzed with mtDNA) often recovered from forensic crime scenes. Moreover, one major reason that mtDNA is the assay chosen for evidentiary material is due to the high copy number of mitochondria in a single cell and the ability to obtain mtDNA sequence information where nuclear DNA cannot be visualized. It is our opinion that the required amount of input DNA for this assay must be decreased in order to be considered for evidentiary material.

A major limiting factor of both the Applied Biosystems mitoSEQr™ Resequencing System and Affymetrix GeneChip® Human Mitochondrial Resequencing Array 2.0 stems from the amplicon size targeted in both assays. In each system the entire mitochondrial genome is amplified using amplicons much larger than what would be practical for challenged forensic samples (i.e. approximately 600 bp amplicons for the Applied Biosystems assay and between 4,000 and 7,000 bp for the Affymetrix assay). Testing the assays with reference sample DNA facilitated the evaluation of the degree of information gained from interrogating the entire mitochondrial genome for forensic identification purposes. Next, a strategy must be developed to be able to harness this information in evidentiary samples and unidentified remains. Neither assay is appropriate at this time for use on degraded or limited DNA samples. We have discussed the potential for the development of smaller amplicon primer sets with both kit manufacturers as one possible way to address this issue. The use of mtDNA-targeted whole genome amplification from challenged samples is also a possible approach to reducing the limitations of the high template amount required by both assays. Another approach to typing challenged samples may be a stepwise approach, wherein an initial association through traditional mtDNA control region sequencing can be followed up with targeted interrogation of coding region SNPs that have been identified in the

putative family reference sample using one of the assays. A combination of all of these methods may be warranted to provide an overall typing strategy for missing persons samples.

Table 1. mtDNA sequence polymorpism in 10 samples indistinguishable by HV1 and HV2 sequencing alone using mitoSEr resequencing kit

Polymorphic sites

Sample	private SNPs	16	263	309.1	315.1	520	710	750
Rcrs		A	A	:	:	C	T	A
00011C	7	A	G	C	C	C	T	G
00047C	1	A	G	C	C	C	T	G
00053C	2	A	G	C	C	C	T	A
00070C	4	A	G	C	C	T	T	G
00109C	2	A	G	C	C	C	T	G
00112C	1*	A	G	C	C	C	T	G
00115C	1*	A	G	C	C	C	T	G
00143C	5	T	G	:	C	C	T	G
00160C	3	A	G	:	C	C	T	G
00082B	6	A	G	:	C	C	C	G

Polymorphic sites

Sample	1438		2259	2393	2483		3010	3107
Rcrs	A		C	C	T		G	C
00011C	G		T	C	T	2798-2943	G	:
00047C	G		C	C	T	2798-2926	A	:
00053C	1074-1487	2063-2094	C	C	T		G	:
00070C	G		C	C	T		A	:
00109C	G		C	C	T		A	:
00112C	G		C	C	T		G	:
00115C	G	2074-2200	C	C	T		G	:
00143C	1420-1508		C	T	T		G	:
00160C	G		C	C	C		G	:
00082B	G		C	C	T		A	:
	H2 A						G H1 A	

Polymorphic sites

Sample	3338		4688	4745	4769		5351	5460
Rcrs	T		T	A	A		A	G
00011C	T		T	G	G		A	G
00047C	T		T	A	G		A	G
00053C	T			4531-4818			A	5470-6420
00070C	T	4168-4413	C	A	G		A	G
00109C	T		C	A	G		R	G
00112C	T		T	A	G	4978-5282	A	G
00115C	T		T	A	G		A	G
00143C	T	4168-4444	T	A	G		A	G
00160C	T		T	A	G		A	G
00082B	C		T	A	G	4961-5290	A	A
					H2 A		ND2	

Table 1. (continued)

Polymorphic sites

Sample	6890		7337			8512	8860
Rcrs	A		G			A	A
00011C	A		A			A	G
00047C	R		G			A	G
00053C	A		G	7382-7660		8299-8782	G
00070C	A		G		7805-8160	8671-8791	G
00109C	A		G			A	G
00112C	A		G			A	G
00115C	A		G			A	G
00143C	A		G			A	G
00160C	A	6893-6918	G			A	G
00082B	A		G			G	G

CO1

Polymorphic sites

Sample	9129	9438		10394	11016	12906	
Rcrs	C	G		C	G	C	
00011C	9048-9058			C	G	A	12993-13206
00047C	9050-9058		10005-10122	C	G	C	12995-13198
00053C	C	G		C	G	C	
00070C	C	A		C	G	C	
00109C	C	G		C	G	C	
00112C	C	G		C	A	C	
00115C	C	G		C	A	C	
00143C	C	G		C	G	C	
00160C	T	G		T	G	C	
00082B	C	G		C	G	C	

Polymorphic sites

Sample	13326	13680	13708	14053	14319.1	14364	14574
Rcrs	T	C	G	A	:	G	C
00011C	C	T	G	A	:	G	C
00047C	T	C	G	A	:	G	C
00053C	T	C	G	G	:	G	C
00070C	T	C	G	A	:	G	C
00109C	T	C	G	A	:	G	C
00112C	T	C	G	A	:	G	C
00115C	T	C	G	A	:	G	C
00143C	T	C	A	A	:	A	T
00160C	T	C	G	A	:	G	C
00082B	T	C	G	A	A	G	C

Table 1. (continued)

Polymorphic sites

Sample	14574	14872	14901	15326	16519
Rcrs	C	C	C	A	T
00011C	C	T	C	G	C
00047C	C	C	C	G	C
00053C	C	C	C	G	C
00070C	C	C	C	G	C
00109C	C	C	C	G	C
00112C	C	C	C	G	C
00115C	C	C	C	G	C
00143C	T	C	C	G	C
00160C	C	C	C	G	C
00082B	C	C	T	G	C

Polymorphisms from rCRS are in red

cells represent region assayed in HV1 and HV2

cells represent individuals assignable to H1 haplogroup

Revised Cambridge Reference sequence H2 haplogroup

Regions where sequence resultse were not obtained

3 ind H1 haplogroup

7 ind H haplogroup * 2 unresolved 1 sample a self declared AA

24 private SNPs

3 shared SNPs (1 is H1 defining)

2 Hets observed

1 novel insertion

Note: shared deletion at 3107 represents an error in the rCRS that as been maintained to avoid nomenclature issues among mtDNA studies (Andrews, et al. 1999)

Table 2. Summary of developed mtDNA coding region SNP panels

Publication	# SNPs Assayed
(Brandstatter, Parsons et al. 2003)	16
(Quintans, Alvarez-Iglesias et al. 2004)	17
(Brandstatter, Salas et al. 2006)	45
(Grignani, Peloso et al. 2006)	9
(Pereira, Richards et al. 2006)	8
(Sigurdsson, Hedman et al. 2006)	150
(Coble, Just et al. 2004)	59
(Vallone, Just et al. 2004)	11

Table 3. Summary of SNP panel effectiveness compared to full sequence evaluation.

	INDIVIDUALIZED HAPLOTYPES	
Ethnicity	**Combined SNP Panels**	**Whole Genome Sequencing**
Caucasians (N=9)	33.3%	77.8%
Hispanics (N=12)	33.3%	83.3%
Blacks (N=21)	61.9%	100%

Table 4. Caucasian Haplotype Breakdown by conventional SNP panels: Combining all published SNP panels would only pick up 59 out of a total of 113 SNPs, leaving 66.67% of our Caucasian individuals unresolved (compared to 32.22% with whole genome sequencing).

Position	*rCRS*	*109-C*	*70-C*	*47-C*	*11-C*	*160-C*	*112-C*	*115-C*	*143-C*	*53-C*	*# of Diff*
750	**A**	G	G	G	G	G	G	G	G	G	9
1,438	**A**	G	G	G	G	G	G	G	X	X	7
3,010	**G**	A	A	A							3
4,688	**T**	C	C							X	2
4,745	**A**				G					X	1
4,769	**A**	G	G	G	G	G	G	G	G	X	8
7,337	**G**				A						1
8,860	**A**	G	G	G	G	G	G	G	G	G	9
10,394	**C**					T					1
15,326	**A**	G	G	G	G	G	G	G	G	G	9
16,519	**T**	C	C	C	C	C	C	C	C	C	9
Cannot Resolve		*	*				*	*	*	*	**66.67%**

Table 5. Hispanic Haplotype Breakdown by conventional SNP panels: Combining all published SNP panels would only pick up 178 out of a total of 446 SNPs, leaving 66.67% of our Hispanic individuals unresolved (compared to 16.67% of samples unresolved by whole genome sequencing).

Position	rCRS	111-H	5-H	57-H	7-H	81-H	179-H	186-H	70-H	92-H	173-H	76-H	176-H	# of Diff
64	C	T	T	T	T	T								5
489	T					---	C	C						2
750	A	G	G	G	G	G	G	G	G	G	G	G	G	12
1,438	A	G	G	G	G	G	G	G	G	G	G	G	G	12
2,706	A	G	G	G	G	G	G	G	G	G	G	G	G	12
3,552	T				---	---	A	A	---	---				2
4,715	A						G	G	G					3
4,769	A	G	G	G	G	G	G	G	G	G	G	G	G	12
5,460	G					---				A				1
7,028	C	T	T	T	T	T	T	T	T	T	T	T	T	12
8,027	G	A	A	A	A	A								5
8,584	G						A	A	A					3
8,860	A	G	G	G	G	G	G	G	G	G	G	G	G	12
9,540	T						C	C	C					3
10,398	A						G	G	G					3
10,400	C						T	T	T					3
10,685	G	---	---							A				1
10,873	T					---	C	C	C					3
11,719	G	A	A	A	A	A	A	A	A	A	A	A	A	12
11,914	G						A	A	A					3
12,007	G	A	A	A	A	A								5
12,705	C	T	T	T	T	T	T	T	T					8
13,263	A						G	G	G					3
13,590	G									A	A	A	A	4
14,318	T					---	C	C	C					3
14,766	C	T	T	T	T	---	T	T	T	T	T	T	T	11
14,783	T					---	C	C	C					3
15,043	G						A	A	A					3
15,326	A	G	G	G	G	G	G	G	G	G	G	G	G	12
16,519	T							C		C	C	C	C	5
Cannot Resolve		*	*	*	*	*					*	*	*	66.67%

Table 6a. African Origin Haplotype Breakdown by conventional SNP panels: Combining all published SNP panels would only pick up 393 out of a total of 915 SNPs, leaving 38.1% of our Black individuals unresolved (compared to all samples resolved by whole genome sequencing).

Position	rCRS	339-B	290-B	309-B	34-B	67-B	273-B	329-B	64-B	148-B	151-B	299-B	# of Diff
64	C	T											1
709	G		A	A	A	---							3
750	A	G	G	G	G	G	G	G	G	G	G	G	19
769	G	A	A	A		A	A	A					6
1,018	G	A	A	A		A	A	A					6
1,048	C	T											1
1,438	A	G			G		G	---	G	G	G	G	17
1,719	G					A			A				2
2,245	A	C	---								---	---	1
2,352	T		C	C		C			C				6
2,416	T						C	C					2
2,706	A	G	G	G	G	G	G	G	G	G	G	G	18
2,758	G	A	A	A		A							4
2,789	C						T	T					2
2,885	T	C	C	C		C							4
3,010	G												2
3,516	C	A			---					---			1
3,594	C	T	T	T	---	---	T	T		---			5
3,666	G		A	A		A							3
3,915	G						---						2
4,104	A	G	G	G		G	G	G					6
4,216	T				C	---			---				1
4,580	G									A			1
4,586	T	C											1
4,769	A	G	G	G	G	---	G	G	G	G	G	G	20
4,917	A				G	---							1
5,147	G			---		---					A	A	2
5,442	T	C											1
5,460	G	A											2
5,603	C	T				---							1
6,185	T	C											1
6,293	T												1
7,028	C	T	T	T	T	T	T	T	T	T	T	T	19
7,055	A		G	G		G							3
7,146	A	G	G	G		---							3
7,256	C	T	T	T		---	T	T					5
8,602	T												1
8,697	G	---			A								1
8,860	A	G	G	G	G	G	G	G	G	G	G	G	21
9,042	C	T											1
9,055	G												1
9,221	A						G	G			---		2
9,347	A	G								---	---		1
9,540	T	C	C	C		C	C	C	C	---	---	C	13

Table 6a. continued

9,899	**T**				C							1
10,398	**A**	G	G	G		G	G	G	G		G	G	17
10,463	**T**				C	---							1
10,589	**G**	A				---							3
10,664	**C**	T			---	---							1
10,810	**T**	C	C	C		C							4
10,819	**A**								G				3
10,873	**T**	C	C	C		C	C	C	C		C	C	16
10,915	**T**	C											1
11,251	**A**				G								1
11,641	**A**	G											1
11,719	**G**	A	A	A	A	A	A	A	A		A	A	18
11,914	**G**	A					A	A					3
12,007	**G**	A											1
12,519	**T**		C	C		C							3
12,705	**C**	T	T	T		T	T	T	T		T	T	16
12,720	**A**	G											1
13,105	**A**	G	G	G		G			G		G	G	8
13,276	**A**	G											1
13,368	**G**				A								1
13,590	**G**						A	A					2
13,789	**T**		C	C		C							3
14,178	**T**		C	C		C							3
14,212	**T**					---			C				4
14,560	**G**		A	A		A							3
14,766	**C**	T	T	T	T	T	T	T	T		T	T	18
14,869	**G**					---	---		A				1
14,905	**G**				A				A				2
15,110	**G**										---		1
15,326	**A**	G	G	G	G	G	G	G	G	G	G	G	20
15,607	**A**				G								1
15,784	**T**						C	C					2
15,904	**C**									T			1
15,928	**G**				A								1
16,368	**T**							C					1
16,390	**G**						A	A					2
16,519	**T**		C	C	C	C		C	C	C			13
Cannot Resolve			*	*							*	*	

Table 6b. Black Haplotype Breakdown by conventional SNP panels (Cont. of sample data).

Position	rCRS	83-B	22-B	143-B	86-B	16-B	49-B	348-B	82-B	125-B	153-B	Number of Variants
64	C			---								1
709	G											3
750	A	G	G	G	G	G	G	G	G			19
769	G											6
1,018	G											6
1,048	C											1
1,438	A	G	G	G	G	G	G	G	G	G	G	17
1,719	G											2
2,245	A											1
2,352	T									C	C	6
2,416	T											2
2,706	A	G	G	G	G	G	G			---	G	18
2,758	G									---		4
2,789	C									---		2
2,885	T									---		4
3,010	G							A	A			2
3,516	C	---					---					1
3,594	C	---	---	---	---		---			---		5
3,666	G	---		---	---							3
3,915	G	---			---					A	A	2
4,104	A	---			---							6
4,216	T	---	---		---		---			---		1
4,580	G											1
4,586	T											1
4,769	A	G	G	G	G	G	G	G	G	G	G	20
4,917	A											1
5,147	G	---	---		---	---			---	---		2
5,442	T	---	---	---	---					---		1
5,460	G	---	---	---	---				A	---		2
5,603	C	---	---		---					---		1
6,185	T											1
6,293	T	C										1
7,028	C	T	T	T	T	T	T			T	T	19
7,055	A											3
7,146	A											3
7,256	C											5
8,602	T			---				C				1
8,697	G			---				---				1
8,860	A	G	G	G	G	G	G	G	G	G	G	21
9,042	C											1
9,055	G		A									1
9,221	A											2
9,347	A	---		---	---						---	1
9,540	T	C	C	---	---	C	C			C	---	13
9,899	T											1
10,398	A	G	G	G	G	G	G			G	G	17
10,463	T											1

Table 6b. continued

10,589	**G**			A	A		---					3
10,664	**C**		---				---					1
10,810	**T**	---										4
10,819	**A**	---								G	G	3
10,873	**T**	---	C	C	C	C	C			C	C	16
10,915	**T**	---										1
11,251	**A**	---	---	---	---							1
11,641	**A**											1
11,719	**G**	A	A	A	A	A	A			A	A	18
11,914	**G**											3
12,007	**G**											1
12,519	**T**											3
12,705	**C**	---	T	T	T	T	T			T	T	16
12,720	**A**	---										1
13,105	**A**	---	---		---	---	G			---		8
13,276	**A**	---										1
13,368	**G**	---										1
13,590	**G**											2
13,789	**T**		---									3
14,178	**T**	---										3
14,212	**T**	---						C		C	C	4
14,560	**G**											3
14,766	**C**	T	T	T	T	T	T			T	T	18
14,869	**G**	---			---							1
14,905	**G**	---			---							2
15,110	**G**	---			---	A						1
15,326	**A**	---	G	G	G	G	G	G	G	G	G	20
15,607	**A**	---										1
15,784	**T**	---										2
15,904	**C**	---	---							---		1
15,928	**G**	---	---							---		1
16,368	**T**											1
16,390	**G**											2
16,519	**T**		---	---	C		C	C	C	C	C	13
Cannot Resolve				*	*					*	*	**38.1%**

Table 7. Comparison of typing effectiveness between the Affymetrix Mito2.0 chip and mitoSEQr ™ Sequencing Approach - Caucasians

RCRS		00011-C		00047-C		00053-C		00070-C		00109-C		00112-C		00115-C		00143-C		00160-C	
pos	Ref	Mito2.0	BD3.1	Mito2.0	BD3.1	Mito2.0	BD3.1	Mito2.0	BD3.1	Mito2.0	BD3.1	Mito2.0		Mito2.0	BD3.1	Mito2.0	BD3.1	Mito2.0	BD3.1
16	A															T	T		
263	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
309.1	:	???	insC	???	insC	???	insC	???	insC	???	insC	???	insC	???	insC				
315.1	:	???	insC	???	insC	???	insC	???	insC	???	insC	???	insC	???	insC	???	insC	???	insC
520	C							T	T										
750	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
980	T									C	C								
1,438	A	G	G	G	G	G	???	G	G	G	G	G	G	G	G	G	???	G	G
2,259	C	N	T																
2,393	T															T	T		
2,483	T																	C	C
3,010	G			A	A			A	A	A	A								
3,107	C	???	delC		delC	???	delC	???	delC	???	delC	???	delC	???	delC	???	delC	???	delC
4,688	T							C	C	C	C								
4,745	A	G	G																
4,769	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
4793	A					G	???												
5,351	A									N	R								
6,890	A			N	R														
7,337	G	A	A																
8,860	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
9,129	C																	N	T
9438	G							A	A										
10,394	C																	T	T
11,016	G											A	A	A	A				
12,906	C	N	A																
13,326	T	C	C																
13,680	C	N	T																
13,708	G															A	A		
14,053	A					G	G												
14,364	G															A	A		
14,574	C															T	T		
14,872	C	T	T																
15,326	A	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G	G
16,519	T	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C

Table 7. Continued

RCRS		00012-C		00019-C	
pos	Ref	Mito2.0	BD3.1	Mito2.0	BD3.1
64	C	T	T		
73	A	G	G		
146	T	N	C		
150	C			T	T
152	T	N	C		
153	A	N	G		
235	A	G	G		
263	A	G	G	G	G
309.1	:	??	insC	??	insC
315.1	:	??	insC	??	insC
522	C	??	delC		
523	A	??	delA		
663	A	G	G		
750	A	G	G	G	G
1,438	A	G	G	G	G
1,736	A	G	G		
2,706	A	G	G		
3,010	G			A	A
3,107	C	???	delC	???	delC
4,248	T	C	C		
4,769	A	G	G	G	G
4,824	A	G	G		
6,491	C	A	A		
7,028	C	T	T		
8,027	G	A	A		
8,794	C	T	T		
8,860	A	G	G	G	G
9,599	C			T	T
10,003	T			C	C
11,719	G	A	A		
12,007	G	N	A		
12,705	C	T	T		
14,766	C	T	T		
15,326	A	G	G	G	G
16,111	C	T	T		
16,223	C	T	T		
16,239	C	T	T		
16,290	C	T	T		
16,319	G	A	A		
16,362	T	T	C		
16,519	T	C	C	C	??

Represents the areas of the genome when indels were detected through traditional sequencing; however were not able to be determined through the MitoChip 2.0

Represents the discripencies between the two methods

Represents where heteroplasmy was detected through the mitoSEQr® resequencing kit; however the MitoChip 2.0 was unable to detect the two base calls

Repreesnts the individualizing polymorphisms detected by way of both methods

Table 8. Comparison of typing effectiveness between the Affymetrix Mito2.0 chip and mitoSEQr™ Sequencing Approach - Hispanic

RCRS		00057-H		00070-H		00076-H	
pos	Ref	Mito2.0	BD3.1	Mito2.0	BD3.1	Mito2.0	BD3.1
60	T					C	C
64	C	C	T				
73	A	G	G	G	G	G	G
146	T	T	C				
153	A	A	G				
185	G					C	??
186.1	:					??	insC
214	A			G	G		
235	A	A	G				
249	A			??	delA		
263	A	G	G	G	G	G	G
290	A			??	delA		
291	A			??	delA		
302.1	:	??	insC				
302.2	:	??	insC				
309.1	:					??	insC
310.1	:	??	insC				
315.1	:			??	insC	??	insC
499	G					A	A
522	C	N	delC				
523	A	N	delA				
663	A	A	G				
750	A	A	G	G	G	G	G
827	A					G	G
1,438	A	G	G	G	G	G	G
1,736	A	G	G				
2,706	A	G	G	G	G	G	G
3,107	C	??	delC	??	delC	??	delC
3,906	T			N	C		
3,547	A					G	G
4,248	T	C	C				
4,715	A			G	G		
4,769	A	G	G	G	G	G	G
4,820	G					A	A
4,824	A	G	G				
4,977	T					C	C
6,308	C	T	T				
6,284	A					G	G
6,473	C					T	T
7,028	C	T	T	T	T	T	T
7,196	C			A	A		
8,027	G	A	A				
8,584	G			A	A		
8,701	A			G	G		
8,705	T	C	C				
8,751	A					G	G
8,794	C	T	T				
8,860	A	G	G	G	G	G	G
9,097	A					G	G
9,497	T					C	C
9,540	T			N	C		
9,545	A			N	G		
9,950	T					C	C
10,398	A			N	G		
10,400	C			N	T		
10,873	T			N	C		
11,177	C					T	T
11,719	G	A	A	A	A	A	A
11,914	G			A	A		
12,007	G	N	A				
12,705	C	N	T	T	T		
13,263	A			G	G		
13,590	G					A	A
14,318	T			C	C		
14,766	C	T	T	T	T	T	T
14,783	T			C	C		
14,962	C			C	N		
15,043	G			A	A		
15,301	G			A	A		
15,326	A	G	G	G	G	G	G
15,487	A			T	T		
15,535	C					T	T
15,930	G			A	A		
16,111	C	T	T			T	T
16,129	G			N	A		
16,183	A					C	C
16,189	T					C	C
16,217	T					C	C
16,223	C	T	T	T	T		
16,290	C	T	T				
16,298	T			C	C		
16,319	G	N	A				
16,325	T			N	C		
16,327	C			N	T		
16,362	T	T	C				
16,468	T	C	C				
16,483	G					A	A
16,519	T					C	C

RCRS		00173-H		00176-H		00186-H	
pos	Ref	Mito2.0	BD3.1	Mito2.0	BD3.1	Mito2.0	BD3.1
73	A	G	G	G	G	G	G
146	T	N	C				
249	A					N	delA
263	A	G	G	G	G	G	G
290	A					N	delA
291	A					N	delA
309.1	:	??	insC			??	insC
309.2	:	??	insC				
315.1	:	??	insC	??	insC	??	insC
337	A			T	T		
489	T					C	C
493	A					G	G
499	G	A	A	A	A		
522	C					N	delC
523	A					N	delC
750	A	G	G	G	G	G	G
827	A	G	G	G	G		
1,438	A	G	G	G	G	G	G
2,706	A	G	G	G	G	G	G
3,107	C	??	delC	??	delC	??	delC
3,547	A	G	G	G	G		
3,552	T					A	A
4,715	A					G	G
4,769	A	G	G	G	G	G	G
4,820	G	A	A	A	A		
4,977	T	C	C	C	C		
6,407	T			C	C		
6,473	C	T	T	T	T		
7,028	C	T	T	T	T	T	T
7,196	C					A	A
7,278	T	C	C				
8,281	C			??	delC		
8,282	C			??	delC		
8,283	C			??	delC		
8,284	C			??	delC		
8,285	C			??	delC		
8,286	T			??	delT		
8,287	C			??	delC		
8,288	T			??	delT		
8,289	A			??	delA		
8,466	A	N	C				
8,584	G					A	A
8,701	A					G	G
8,860	A	G	G	G	G	G	G
9,540	T					N	C
9,545	A					N	G
9,950	T	C	C	C	C		
10,398	C					G	G
10,400	C					N	T
10,837	A					N	G
11,177	C	T	T	T	T		
11,719	G	A	A	A	A	A	A
11,914	G					A	A
11,950	A	G	G				
12,705	C					T	T
12,727	T			C	C		
13,263	A					G	G
13,590	G	N	A	A	A		
13,635	T					C	C
13,845	C	C	Y				
14,097	C	T	T				
14,318	T					C	C
14,569	G	A	A				
14,766	C	T	T	T	T	T	T
14,783	T					C	C
15,043	G					A	A
15,301	G					A	A
15,326	A	G	G	G	G	G	G
15,470	T					C	C
15,487	A					T	T
15,535	C	T	T	T	T		
16,183	A	C	C	C	C		
16,189	T	C	C	C	C		
16,217	T	C	C	C	C		
16,223	C					T	T
16,270	C	N	G				
16,278	C	T	T	T	T		
16,298	T					C	C
16,311	T			C	C	C	C
16,325	T					C	C
16,327	C					T	T
16,519	T	C	C	C	C	C	C

Represents the areas of the genome when indels were detected through traditional sequencing; however were not able to be determined through the MitoChip 2.0

Represents the discripencies between the two methods

Represents where heteroplasmy was detected through the mitoSEQr® resequencing kit; however the MitoChip 2.0 was unable to detect the two base calls

Represents the individualizing polymorphisms detected by way of both methods

Table 9. Comparison of typing effectiveness between the Affymetrix Mito2.0 chip and mitoSEQ™ Sequencing Approach

Hispanic

RCRS		00005-H		00081-H		00111-H	
pos	Ref	Mito2.0	BD3.1	Mito2.0	BD3.1	Mito2.0	BD3.1
64	C	T	T	T	T	T	T
73	A	G	G	G	G	G	G
146	T	C	C	C	C	C	C
153	A	G	G	G	G	G	G
214	A						
235	A	G	G	G	G	G	G
263	A	G	G	G	G	G	G
290	A						
291	A						
302.1	:	??	insC				
309.1	:			??	insC		
309.2	:			??	insC		
310.1	:	??	insC	??	insC	??	insC
315.1	:						
514	C	??	delC			??	delC
515	A	??	delA			??	delA
518	C						
520	C						
663	A	G	G	G	??	G	G
750	A	G	G	G	G	G	G
1,438	A	G	G	G	G	G	G
1,736	A	G	G	G	G	G	G
2,706	A	G	G	G	G	G	G
3,010	G						
3,107	C	??	delC	??	delC	??	delC
3,202	T	C	C				
3,450	C						
3,866	T						
3,972	A			G	G		
4,248	T	C	C	C	??	C	C
4,688	T						
4,715	A						
4,769	A	G	G	G	G	G	G
4,820	G						
4,824	A	G	G	G	G	G	G
5,375	C						
5,773	G						
6,221	T						
7,028	C	T	T	T	T	T	T
7,961	T	C	C				
8,027	G	A	A	A	A	A	A
8,460	A	G	G				
8,794	C	T	T	T	T	T	T
8,860	A	G	G	G	G	G	G
11,719	G	A	A	A	A	A	A
12,007	G	A	A	A	A	N	A
12,705	C	T	T	T	T	T	T
13,978	C			C	N		
14,766	C	T	T	T	??	T	T
15,326	A	G	G	G	G	G	G
15,629	T					C	C
15,924	A	G	G				
16,111	C	T	T				
16,223	C	T	T	T	T	T	T
16,290	C	T	T	T	T	T	T
16,319	G	A	A	A	A	A	A
16,362	T	T	C	T	C	T	C

African

RCRS		00153-B		00022-B	
pos	Ref	Mito2.0	BD3.1	Mito2.0	BD3.1
73	A	G	G	G	G
150	C	T	T		
249	A				
263	A	G	G	G	G
315.1	:	??	insC	??	insC
518	C			T	??
520	C			T	??
522	C	??	delC		
523	A	??	delA		
750	A			G	G
1,438	A	G	G	G	G
2,352	T	C	C		
2,706	A	G	G	G	G
3,107	C	??	delC	??	delC
3,450	C			T	T
3,866	T			C	C
3,915	G	A	A		
3,552	T				
4,715	A				
4,769	A	G	G	G	G
5,262	G	A	??		
5,375	C			T	??
5,584	A	G	G		
5,773	G			A	A
6,221	T			C	C
7,028	C	T	T	T	T
8,688	A	C	C		
8,584	G				
8,701	A	G	G	G	G
8,860	A	G	G	G	G
9,055	G			A	A
9,449	C			T	T
9,540	T	C	??	C	C
10,086	A			G	G
10,373	G			A	A
10,398	C	G	G	G	G
10,819	A	G	G		
10,873	T	C	C	N	C
11,002	A			G	G
11,257	C	T	T		
11,719	G	A	A	A	A
11,914	G				
11,950	A				
12,705	C	T	T	T	T
13,105	A			G	??
13,749	C	T	T		
13,914	C			A	??
14,212	T	C	C		
14,251	A			A	N
14,252	C			C	N
14,522	A			A	N
14,760	G			G	K
14,766	C	T	T	T	T
14,791	C			T	T
15,301	G	A	A	A	A
15,311	A			G	G
15,326	A	G	G	G	G
15,824	A			G	??
16,051	A	G	G		
16,124	T			C	C
16,223	C	T	T	T	T
16,264	C	T	T		
16,278	C			T	T
16,519	T	C	C	T	??
16,362	T			T	C

Represents the areas of the genome when indels were detected through traditional sequencing; however were not able to be determined through the MitoChip 2.0

Represents the discripencies between the two methods

Represents where heteroplasmy was detected through the mitoSEQr® resequencing kit; however the MitoChip 2.0 was unable to detect the two base calls

Represents the individualizing polymorphisms detected by way of both methods

Table 10. Quality Threshold Comparison Evaluated through % N calls for three settings (3.0. 6.0. & 12.0)

Sample	3.0	6.0	12.0
00011-C	0.7	1.4	3.5
00012-C	1	1.8	3.1
00019-C	1	1.8	3.3
00028-C	1	1.7	3.4
00031-C	0.9	1.6	3.1
00042-C	0.9	1.5	2.9
00045-C	1	1.8	3.1
00047-C	1	1.6	3.3
00048-C	1.1	1.9	3.8
00050-C	1.4	2.3	4.1
00053-C	0.7	1.4	2.7
00067-C	0.8	1.4	2.8
00070-C	0.8	1.8	4.0
00075-C	1.1	1.8	3.5
00081-C	16.1	20.7	23.7
00102-C	1	1.8	3.2
00107-C	1	1.8	3.2
00109-C	1.4	2.2	3.7
00112-C	1.3	2.4	4.7
00115-C	1.1	2	4.0
00120-C	0.9	1.7	3.1
00126-C	1.4	2.4	3.8
00132-C	1	1.7	3.2
00143-C	1	1.8	3.2
00160-C	0.9	1.7	4.1
00016-B	1	2.1	6.3
00022-B	1.2	1.9	6.1
00034-B	0.9	1.7	3.3
00049-B	0.7	1.5	3.9
00064-B	0.7	1.6	4.2
00067-B	1.1	1.9	4.6
00083-B	0.7	1.5	3.9
00086-B	0.6	1.5	3.9
00125-B	0.6	1.2	4.2
00151-B	0.8	1.5	3.7
00153-B	1.2	1.8	3.3
00005-H	2.6	3.5	5.1
00007-H	0.9	1.7	4.1
00057-H	1	1.7	3.6
00070-H	1.7	2.9	4.4
00076-H	1	2	5.2
00081-H	3.5	5.1	6.8
00092-H	1	2.2	4.5
00111-H	2.7	3.8	4.9
00173-H	1	1.6	3.5
00176-H	2	3.1	4.7
00179-H	1.1	1.9	3.2
00186-H	0.9	1.9	3.5

Table 11. Examples of basecalling success, based on number of ambiguous calls (N) for 3 samples obtained using 3 quality thresholds (12, 6.0, and 3.0) for haploid detection, and Quality Threshold options set at 3.0 for the diploid detection mode necessary to call heterplasmic sites

00011-C

RCRS			Haploid			Diploid
pos	Ref	mitoSEQr	12.0	6.0	3.0	3.0
263	A	G	G	G	G	G
309.1	:	insC	???	???	???	???
315.1	:	insC	???	???	???	???
750	A	G	G	G	G	G
1,438	A	G	G	G	G	G
2,259	C	T	N	N	T	N
3,107	C	delC	???	???	???	???
4,745	A	G	G	G	G	G
4,769	A	G	G	G	G	G
7,337	G	A	A	A	A	A
8,860	A	G	G	G	G	G
9,909	T	T	T	T	T	W
12,906	C	A	N	A	A	N
13,326	T	C	C	C	C	C
13,680	C	T	N	T	T	N
14,872	C	T	T	T	T	T
15,326	A	G	G	G	G	G
16,519	T	C	C	C	C	C
16, 551	T	T	T	T	T	Y
16,555	T	T	T	T	T	W
		% N calls	3.5	1.4	0.7	11.1

00016-b

RCRS			Haploid			Diploid
Pos	Ref	BD3.1	12.0	6.0	3.0	3.0
73	A	G	N	G	G	G
146	T	C	C	C	C	N
152	T	C	C	C	C	C
263	A	G	G	G	G	G
315.1	:	insC	???	???	???	???
522	C	delC	???	???	???	???
523	A	delA	???	???	???	???
750	A	G	G	G	G	G
921	T	C	C	C	C	C
1,438	A	G	G	G	G	G
2,706	A	G	N	G	G	G
3,107	C	delC	???	???	???	???
4,769	A	G	G	G	G	G
5,046	G	N	A	A	A	A
5,147	G	N	A	A	A	A
6,025	T	T	T	T	T	W
6,446	G	A	A	A	A	A
6,641	T	C	C	C	C	C
6,680	T	C	C	C	C	C
7,028	C	T	T	T	T	T
7,424	A	G	G	G	G	G
8,618	T	C	C	C	C	C
8,701	A	G	G	G	G	G
8,860	A	G	N	G	G	N
8,899	C	C	C	C	C	M
9,179	T	T	T	T	T	Y
9,539	A	A	A	A	A	M
9,540	T	C	C	C	C	C
9,541	T	T	T	T	T	Y
10,393	A	A	A	A	A	M
10,398	A	G	G	G	G	G
10,873	T	C	C	C	C	N
11,719	G	A	A	A	A	A
12,705	C	T	T	T	T	T
13,105	A	N	G	G	G	G
13,886	T	C	C	C	C	C
14,284	C	T	T	T	T	T
14,634	T	C	C	C	C	C
14,766	C	T	T	T	T	T
15,110	G	A	A	A	A	N
15,301	G	A	A	A	A	A
15,326	A	G	G	G	G	G
15,748	T	C	C	C	C	C
16,124	T	C	C	C	C	C
16,223	C	T	T	T		T
16,256	C	T	N	T	T	T
16,556	A	A	A	A	A	W
		% N calls	6.3	2.1	1	22.2

00176-H

RCRS			Haploid			Diploid
pos	Ref	mitoSEQr	12.0	6.0	3.0	3.0
68	G	G	G	G	G	K
73	A	G	G	G	G	G
263	A	G	G	G	G	G
315.1	:	insC	???	???	???	???
337	A	T	T	T	T	T
499	G	A	A	A	A	A
750	A	G	G	G	G	G
827	A	G	G	G	G	G
1,438	A	G	G	G	G	G
2,706	A	G	G	G	G	G
2,824	G	G	G	G	G	S
3,107	C	delC	???	???	???	???
3,547	A	G	G	G	G	G
4,769	A	G	G	G	G	G
4,820	G	A	A	A	A	A
4,977	T	C	C	C	C	C
6,407	T	C	C	C	C	C
6,473	C	T	T	T	T	T
7,028	C	T	T	T	T	T
8,281	C	delC	???	???	???	???
8,282	C	delC	???	???	???	???
8,283	C	delC	???	???	???	???
8,284	C	delC	???	???	???	???
8,285	C	delC	???	???	???	???
8,286	T	delT	???	???	???	???
8,287	C	delC	???	???	???	???
8,288	T	delT	???	???	???	???
8,289	A	delA	???	???	???	???
8,860	A	G	G	G	G	G
8,899	C	C	C	C	C	M
9,950	T	C	C	C	C	C
11,177	C	T	T	T	T	N
11,719	G	A	A	A	A	A
12,727	T	C	C	C	C	C
13,590	G	A	A	A	A	A
14,766	C	T	T	T	T	T
14,769	A	A	A	A	A	R
15,326	A	G	G	G	G	G
15,535	C	T	T	T	T	N
16,183	A	C	C	C	C	N
16,189	T	C	C	C	C	N
16,217	T	C	C	C	C	C
16,278	C	T	T	T	T	T
16,311	T	C	C	C	C	C
16,519	T	C	C	C	C	C
		% N calls	3.5	1.6	1.0	7.4

Reference List

1. Allard, et all 2002. Characterization of the Caucasian haplogroups present in the SWGDAM forensic mtDNA dataset for 1771 human control region sequences. *J. Forensic Sci.* **47**:1215-1223

2. Andrews, R. M. et al. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA, *Nature Genetics* (1999) 23:147.

3. Anderson, S. *et al*. Sequence and organization of the mitochondrial genome, *Nature* (1981) **290**:457-465.

4. Brandstatter, A., T. J. Parsons, et al. (2003). "Rapid screening of mtDNA coding region SNPs for the identification of west European Caucasian haplogroups." *Int. J. Legal Med.* **117**(5): 291-8.

5. Brandstatter, A., A. Salas, et al. (2006). "Dissection of mitochondrial superhaplogroup H using coding region SNPs." *Electrophoresis* **27**(13): 2541-50.

6. Budowle, B., Moretti, T.R., Baumstark, A.L., Defenbaugh, D.A., Keys, K.M. (1999): Population data on the thirteen CODIS core short tandem repeat loci in African Americans, U.S. Caucasians, Hispanics, Bahamians, Jamaicans, and Trinidadians. *J. Forensic Sci.* **44**:1277-1286.

7. Butler, Shen and McCord (2003) The development of reduced size STR amplicons as tools for analysis of degraded DNA. *J. Forensic Sci.* **48**, 1054-1064.

8. Coble, M.D. and Butler, J.M. (2005) Characterization of new miniSTR loci to aid analysis of degraded DNA. *J. Forensic Sci.* **50**(1):43-53.

9. Coble, M. D., R. S. Just, et al. (2004). "Single nucleotide polymorphisms over the entire mtDNA genome that increase the power of forensic testing in Caucasians." *Int. J. Legal Med.* **118**(3): 137-46.

10. Dixon, L.A., Dobbins, A.E., Pulker, H., Butler, J.M., Vallone, P.M., Coble, M.D., Parson, W., Berger, B., Grubweiser, P., Mogensen, H.S., Morling, N., Nielsen, K., Sanchez, J.J., Petkovski, E., Carracedo, A., Sanchez-Diz, P., Brion, M., Irwin, J.A., Just, R.S., Loreille, O., Parsons, T.J., Syndercombe-Court, D., Schmitter, H., Gill, P. (2006) Analysis of artificially degraded DNA using STRs and SNPs--results of a collaborative European (EDNAP) exercise. *Forensic Sci. Int.* ***164****: 33-44.*

11. Drabek, J., Chung, D.T., Butler, J.M., McCord, B.R. (2004) Concordance study between miniplex STR assays and a commercial STR typing kit, *J. Forensic Sci.* ***49****(4): 859-860.*

12. Erratum on NC01 and NC02 (D1S1677, D4S2364, D10S1248, D22S1045) nomenclature: http://www.cstl.nist.gov/div831/strbase/miniSTR.htm#Nomenclature_Errata

13. Gill, P., Fereday, L., Morling, N., Schneider, P.M. (2006) The evolution of DNA databases--recommendations for new European loci. *Forensic Sci. Int.* ***156****:242-244*.

14. Gill, P., Fereday, L., Morling, N., Schneider, P.M. (2006) Letter to editor -- New multiplexes for Europe-amendments and clarification of strategic development. *Forensic Sci. Int.* **163**:155-157.

15. Grignani, P., G. Peloso, et al. (2006). "Subtyping mtDNA haplogroup H by SNaPshot minisequencing and its application in forensic individual identification." *Int J. Legal Med.* **120**(3): 151-6.

16. Hill, Becky poster at 58th Annual Meeting of the American Academy of Forensic Sciences (Seattle, WA), February 24, 2006, "Development of 27 New miniSTR Loci for Improved Analysis of Degraded DNA Samples" [.pdf]

17. Hill, Becky poster at 17th International Symposium on Human Identification (Nashville, TN), October 10-12, 2006, "Characterization of 26 New miniSTR Loci" [.pdf]

18. October 2006 Forensic News: http://marketing.appliedbiosystems.com/images/enews/ForensicNews_Vol7/PDF/00_ForensicNews.pdf

19. Pereira, L., M. Richards, et al. (2006). "Evaluating the forensic informativeness of mtDNA haplogroup H sub-typing on a Eurasian scale." *Forensic Sci. Int.* **159**(1): 43-50.

20. Quintans, B., V. Alvarez-Iglesias, et al. (2004). "Typing of mitochondrial DNA coding region SNPs of forensic and anthropological interest using SNaPshot minisequencing." *Forensic Sci. Int.* **140**(2-3): 251-7.

21. Schumm, Wingrove and Douglas (2004) Robust STR multiplexes for challenging casework samples. *Progress in Forensic Genetics ICS*, 1261, 547-549.

22. Sigurdsson, S., M. Hedman, et al. (2006). "A microarray system for genotyping 150 single nucleotide polymorphisms in the coding region of human mitochondrial DNA." *Genomics* **87**(4): 534-42.

23. Vallone, P. M., R. S. Just, et al. (2004). "A multiplex allele-specific primer extension assay for forensically informative SNPs distributed throughout the mitochondrial genome." *Int J. Legal Me.d* **118**(3): 147-57.

24. Wiegand and Kleiber (2001) Less is more – length reduction of STR amplicons using redesigned primers. *Int. J. Legal Med*. **114**, 285-287.

Presentations Based Upon Award: 2004-DN-BX-K214
Field Test of Current Technology Used in the Identification of Unidentified Remains

September 29, 2005	Promega 16th International Symposium on Human Identification; Workshop on Missing Persons and Unidentified Human Remains; Grapevine, Texas
November 7-9, 2005	U.S. Department of Justice, Missing Persons Regional Training; Advancing Justice Through DNA Technology, Missing Persons Program, Denver, Colorado
January 12-13, 2006	Association of Forensic Analysts and Administrators (AFDAA); Missing Persons Training; Austin, Texas
Jan 30-Feb 1, 2006	2006 NIJ Applied Technologies and Partnerships Conference; Advancing Justice Through DNA Technology, Missing Persons Program, Hilton Head, South Carolina
February 21, 2006	Applied Biosystems 11th Annual User Forum, 58th Annual Meeting of the American Academy of Forensic Sciences, Evaluation of a Novel Mini-STR Multiplex System for the Analysis of Unidentified Human Remains; Seattle, Washington
June 26, 2006	NIJ Seventh Annual DNA Grantees Workshop; Putting it All Together: Field Tests of Tools to Help Solve Missing Persons Cases; Washington, DC
August 29, 2006	INPALMS-AICEF Symposium on Forensic DNA, Application of DNA Analysis for Human Identification; New Delhi, India
October 11, 2006	Promega 17th International Symposium on Human Identification; Development of New Tools to Aid in the Identification of Missing Persons and Unidentified Human Remains; Nashville, Tennessee
November 15, 2006	Honolulu Police Department Criminal Investigation Division; Application of DNA Analysis for Human Identification; Honolulu, Hawaii
November 20, 2006	First International Congress of Forensic and Legal Medicine in Andalusia; Solving Cold Cases Through the Identification of Missing Persons and Unidentified Human Remains; Institutes of Legal Medicine, Andalusia, Granada, Spain

February 20, 2007	59th Annual Meeting of the American Academy of Forensic Science; Workshop #18: Missing Persons: Resources. Techniques, and Identification; The UNT System, Center for Human Identification: A Resource for the Identification of Human Remains and Missing Persons; San Antonio, Texas
April 3, 2007	NIJ Applied Technology Conference; What Do I Do with These Bones in My Office? ; Orange County, California

Publications Based Upon Award: 2004-DN-BX-K214
Field Test of Current Technology Used in the Identification of Unidentified Remains

October, 2006 Eisenberg, AJ, Aranda, XG, and Planz, JP. Evaluation of the AmpF*l*STR® MiniFiler™ PCR Amplification Kit for Use Compromised DNA Samples. Applied Biosystems Forensic News; http://marketing.appliedbiosystems.com/images/enews/ForensicNews_Vol7/PDF/00_ForensicNews.pdf