Final Report

Mitochondrial DNA Genome Sequencing and SNP Assay Development for Increased Power of Discrimination

NIJ Grant
2000-IJ-CX-K010

Thomas J. Parsons, Principal Investigator
American Registry of Pathology*†

*Research conducted primarily at the Armed Forces DNA Identification Laboratory (AFDIL), with collaborative interaction with National Institutes of Standards and Technology (NIST), and the Institute of Legal Medicine, University of Innsbruck, Austria.

†The following individuals made critical contributions of time, effort, and expertise to this project: Rebecca Just, Michael Coble, Jodi Irwin, Jessica Saunier, Toni Diegoli, Kimberly Sturk, Katharine Strouss, Jennifer O'Callaghan, Carla Paintner, Christine Peterson, Ilona Letmanyi, Peter Vallone, Walther Parson, Harrald Niederstätter, and Anita Brandstätter.

We acknowledge the following individuals for particular assistance or discussion: James Ross, Jon Norris, Suzanne Barritt-Ross, Timothy McMahon, Brion Smith, James Canik, Michael Parry, and Constance Fisher.

Technical assistance was provided by: Heather Williams, Kimberly Watson, William Ivory, Devon Pierce, Victoria Glynn, Rachel Berry, Trina Bersola, Serena Filosa, and Carrie Guyan.

Abstract

This report details the work performed over a period of five years in an NIJ-funded project intended primarily to develop a practical means for accessing the large reserve of genetic variation in the ~15,000 bp mtDNA coding region, as a means for augmenting the forensic discrimination provided by sequencing the hypervariable (HV) regions of the mtDNA control region. Our approach was to focus on the relatively small number of particularly common types present in US Caucasian, African American, and Hispanic populations, where the problem of limited discrimination is concentrated in practice. We sequenced 506 entire mtDNA genomes corresponding to 56 common HV types present at 0.5% or more in the respective populations. Full genome sequencing resolved the 56 HV types into 423 haplotypes, and permitted the identification of 123 SNP sites suitable for practical assay development. Selection criteria for target SNPs included that they be present in multiple individuals, not be redundant with variation in the HV regions, and that they not occur at positions that affect either amino acid sequence or changes in structural RNA genes. Selected discriminatory SNPs were organized into panels that are specific for particular common HV types, or their near relatives. The intent was for a forensic scientist, encountering a common HV type, to be able to turn to one or two multiplex assays that represent the best chance for detecting additional discriminatory variation for the HV type in question. This was demonstrated to be a particularly effective strategy based on the observed distribution of variation in the mtDNA coding region. For Caucasian common HV types, eight multiplex allele-specific primer extension (ASPE) assays were designed, optimized, and tested. These proved to be highly suitable with regard to characteristics important for forensic mtDNA testing. Developmental validation for sensitivity, mixture detection, and degraded samples was completed for seven of the eight panels (one still in development). One multiplex has been applied in numerous case investigations, establishing a clear practical utility. Population databases were established for the completed multiplexes. Additional work on this project included generation of new control region databases, including four regional Hispanic population samples that were significantly differentiated at the haplogroup and haplotype level. Also developed, in a collaborative project, was a multiplex ASPE SNP assay for haplogroup assignment among mtDNAs of W. European origin.

Table of Contents

Structure of This Report:

In order to maximize the efficiency of reporting the large body of data and analysis that has resulted from this project, we will convey the information in three main sections:

1. An Executive Summary that can serve as a stand-alone overview of the project, its methods, and findings; the summary points the reader to more detailed information/discussion in one of the other two following sections.

2) An Experimental Details and Main Findings section that provides

a) an abbreviated review of background, results, protocols, and discussion relating to portions of this project that have already been published, with appropriate reference to the published work;

b) a detailed presentation of results, protocols, and discussion that has not been previously published;

3) Several appendices containing technical information.

Note on Ethnic/Racial Categories: Following a convention that is common in forensics in the United States, this report will refer to the following ethnic/racial categories: US Caucasian, African American, and Hispanic. These essentially reflect categories of self-identification when individuals are given a small number of pre-set choices for ethnic/racial association (on forms, or in the national census, etc…). It is not the case that any of these categories behave as distinct biological populations, or can be clearly defined in genetic terms. All three are of heterogeneous origin and experience significant degrees of admixture, both historical and present-day. It may be the use of these terms in forensics ought to be revised, but this issue is beyond the scope of this project/report, and the standard usage is retained solely for convenience, and to relate to

current practices. Nevertheless, the three categories as referred to in this report represent significantly different distributions of mtDNA sequences and the project was usefully structured around separate studies relating to each; overlap between the categories is evident in the results of this study and will be noted.

Executive Summary:

This is the final report of a two stage, five-year grant project funded by the National Institute of Justice (award number: 2000-IJ-CX-K010). The first stage of this award, spanning 9/30/2000 to 9/30/2002 was entitled: *Homogeneous Fluorescent PCR Assays for Forensically-Informative Sites over the Entire mtDNA Genome: Increasing the Power of Discrimination, Ease, and Speed of Forensic mtDNA Testing.* A supplementary award permitted continuation of this work from 10/1/2002 to 10/1/2005 with the same award number, but a revised title: *Mitochondrial DNA Genome Sequencing and SNP Assay Development for Increased Power of Discrimination.* The latter will be chosen for the overall project title in this report. A no-cost extension of the grant activity period has been awarded till 12/31/05, to permit completion of some work and reporting of results.

Background and Introduction:

Forensic testing of mtDNA sequence variation has become well established to assist in identification casework, primarily in criminalistic and missing persons applications (Holland and Parsons, 1999). MtDNA typing is performed mostly under two circumstances: 1) when degradation of DNA or sample type makes recovery of nuclear DNA difficult or impossible, and 2) when comparison to matrilineal relatives is desired. Critically important features of mtDNA that underlie its application are the following: 1) it is present in high copy number per cell (on

the order of 1000 depending on cell type), 2) it is maternally inherited, and 3) is does not recombine. In its current practice, mtDNA typing usually targets complete sequence determination of 610 bp of the control region (or D-Loop), from two portions referred to as hypervariable region I (HV1, 16024-16365) and hypervariable region II (HV2, 73-340). Results from this sequencing are reported as a list of nucleotide positions at which there are differences from a standard reference sequence referred to as the "Anderson" sequence, or more commonly now the rCRS (the revised Cambridge reference sequence, Anderson *et al*. 1989; Andrews *et al*., 1999). [*In this report, HV-type will be used as shorthand to mean a particular haplotype with defined by the full sequence of HV1 and HV2*.]

Because mtDNA is a non-recombining locus, sequence polymorphisms must be treated as wholly linked within the same molecule. Therefore, in order to determine the evidentiary significance of an mtDNA sequence match, one cannot consider the frequency of individual polymorphisms, but rather the frequency with which the entire haplotype occurs within relevant populations. This requires reference to large population databases of mtDNA sequences. Both forensic and academic studies have generated vast amounts of mtDNA control region sequence variation from populations worldwide, although only a small proportion of this is available for forensic applications due to the requirement for a searchable compilation of uniform sequence ranges and consistent quality control criteria. The basic structure of mtDNA variation in all populations is such that a majority of HV-types are rare, with a small number of more common types. The frequency distribution of any population will reflect its own history regarding age, migration patterns, founder effect, and population expansion.

Figure 1 illustrates the frequency distribution of HV-types in the US Caucasian population. The most common HV-type is present in > 7% of the population while ~50% of the

population is "unique in the database" (Fig 1). While most sequences are rare and of correspondingly high evidentiary value, only a small portion of HV-types are present at ~0.5% or greater in the population. In US Caucasians, about 20% of individuals have an HV-type that has a frequency of 0.5% or more of the population, and it is in these relatively common sequences where mtDNA manifests its greatest limitation: a low power of forensic discrimination. A similar skewed mtDNA frequency distribution is characteristic of African American and US Hispanic populations, as well.

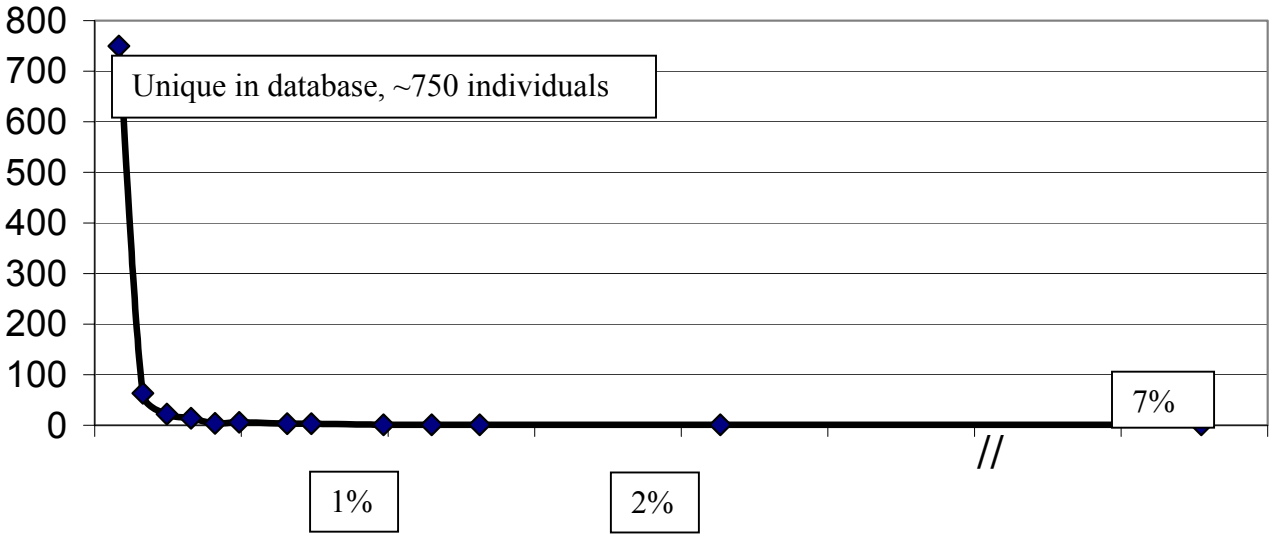


Figure 1: Frequency Distribution of mtDNA Types in a US Caucasian Database.
Frequencies based on FBI database of 1219 Caucasian individuals.

The practical consequence of the mtDNA frequency distribution listed above is that about 20% of the cases that are encountered will involve an HV-type with decidedly limited evidentiary value in comparison to the rarer types (for most of these, the evidentiary significance is limited by the size of the database). This has its greatest effect in cases involving multiple

sequence comparisons, involving multiple individuals; extreme examples are mass fatality incidents, or searches of missing person's databases. Under those conditions it is not uncommon, and in large cases it is assured, that reference samples for unrelated missing individuals will be indistinguishable in HV-type, therefore negating any ability to unambiguously associate a questioned sample with the proper family reference. This is complicated further by the standard requirement that at least two mtDNA sequence differences need to be present to constitute a practical exclusion. What is needed is some means for further distinguishing between individuals who match in HV-type.

Sequence variation in mtDNA is concentrated in the non-coding control region, and especially the HV regions, with an evolutionary nucleotide substitution rate in the control region approximately ten-fold higher than in the coding region. This makes the HV regions the obvious first choice for seeking mtDNA variation, but Figure 2 shows a scale map of the mtDNA genome that makes it clear that even with a 10-fold lower rate of variation, a majority of the sequence variation in mtDNA occurs in the coding region (as there is ~15 times more coding region). Given that regions outside of HV1/HV2 also contain large amounts of variation, our logic at the outset of this project was that by extending typing beyond HV/HV2 it ought to be possible to substantially alleviate the low power of discrimination of mtDNA testing. However, it was also anticipated that sites that provide additional discrimination, in relation to HV1/HV2 sequences, would be relatively rare, widely scattered, and dependent on the sequence background reflecting the history of particular mtDNA haplogroups. Unfortunately, the complete linkage of mutations over the mtDNA genome results in the fact that most randomly encountered coding region mutations are likely to be redundant, in terms of forensic discrimination, either with each other, or with variation already accessed by sequencing HV1/HV2. Because of the lower evolutionary

rate in the coding region, most mutational variants in the coding region are older than those in the D-loop, with the anticipated results that fast evolving D-loop mutations tend to discriminate within coding region lineage markers, rather than *vice versa*. Based on these considerations, which in the end proved to be correct, it was decided to identify additional discriminatory information by a brute force approach: for the most common HV-types in a population, multiple individuals matching that type would be sequenced for the entire mtDNA genome (mtGenome) to identify the specific single nucleotide polymorphisms (SNPs) that provide maximal increased discrimination.

This undertaking, applied separately to US Caucasian, African American, and Hispanic populations represents the bulk of the work performed in this project, together with the development of appropriate SNP assays that permit recovery of the discriminatory markers from real world forensic samples. Associated with these overarching goals were a number of component subprojects, and related investigations which are summarized below.

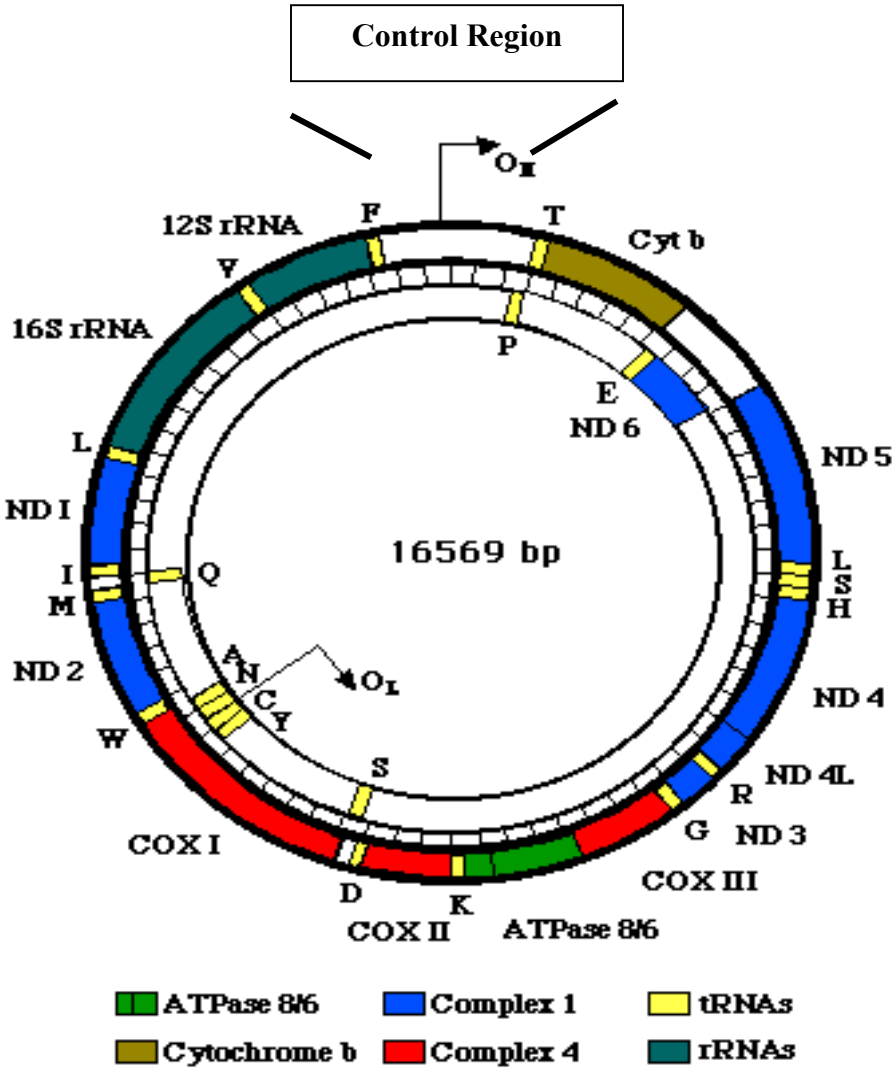


Figure 2. Physical Map of the mtDNA Genome
From www.mitomap.org. Colored regions indicate protein-coding, ribosomal RNA, or tRNA genes.

Overview of Specific Goals:

Specific stated goals of this project, over its five-year duration, were the following:

1) Sequence the entire mtDNA genome of multiple individuals corresponding to the most common mtDNA types (within hypervariable regions I and II) in the US Caucasian, African

American, and US Hispanic populations, and identify nucleotide sites that provide increased forensic discrimination among these common types.

2) Create new databases of mtDNA control region sequences for African-American and U.S. Hispanic individuals so that physical samples corresponding to common HV1/HV2 types would be available for mtGenome sequencing.

3) Evaluate the potential for forensically significant regional variation in mtDNA distribution among the populations studied under goal 2) above.

4) Develop SNP assays for selected informative sites that perform suitably for forensic mtDNA applications and establish developmental validation with relation to a) sensitivity, b) mixture and heteroplasmy characterization, and c) performance on degraded DNA.

5) Establish population databases of SNP loci on samples for which other primary forensic markers have also been typed (mtDNA control region sequences, autosomal STRs, Y-STRs, and Y-SNPs).

6) Develop multiplex coding region SNP assays for determining world mtDNA haplogroup attribution of mtDNAs from unknown individuals.

Summary of Progress on Goals:

1) Entire mtDNA Genome Sequencing and identification of Discriminatory SNPs.

In order to meet the high throughput sequencing goals of this project, a system of automated instrumentation was developed together with informatic tools for automated data transfer and analysis. Robotic instruments were employed in the following process stages, processed in 96-well plates: blood or buccal swab punching; DNA extraction; PCR amplification; post-PCR clean-up; dideoxy-terminator cycle sequencing reaction set up; and

post-sequencing clean-up. Automated fluorescent sequencing was performed on either ABI 377 polyacrylamide gel instruments (early on in the project), or the with the multi-capillary array instruments, the ABI 3100 or ABI 3730. Sequence data analysis was performed using the assembly program Sequencher (GeneCodes Corporation, Ann Arbor, MI), and unambiguous consensus sequences were finalized through a information management system that eliminated manual data transcription, facilitated efficient review and confirmatory checking, and exported data into a final locked database from which additional analysis could be performed without fear of corruption of the original data.

For the project, the following numbers of mtGenomes were sequenced: 241 Caucasian individuals, representing 18 common HV-types; 140 African American individuals, representing 20 common HV-types; and 125 Hispanic individuals representing 18 common HV-types. In all three groups, large numbers of candidate SNPs were discovered that, in general, provided a large degree of additional forensic resolution in comparison to the HV-types alone. The data were scrutinized for the identification of SNPs useful for practical forensic assays using the following criteria: i) the SNP variants should occur in more than a single individual, ii) the SNPs should not be redundant in providing discrimination, and iii) the SNPs should occur at either non-coding positions, or the variants should involve synonymous substitutions within protein-coding genes. Details of the increased forensic discrimination provided by both the entire mtGenome sequences and panels of select SNP sites are presented below for all three ethnic groups, and for US Caucasians were published in Coble *et al*. (2004), and discussed in Coble (2004). There were 506 mtGenomes sequenced for this study, representing only 56 different HV-types; when the complete mtGenome was taken into account, these resolved to 423 distinct mtDNA sequences.

377 of the sequences were unique, so that 74% of the original 506 samples were found to be distinct from all other mtDNA types upon full sequence analysis.

2) mtDNA Control Region Databases

In order to accomplish this study, it was necessary to have DNA samples matching common HV-types to sequence for the mtGenome. The common HV-types to target were identified through pairwise comparisons of the SWGDAM mtDNA population database and internal AFDIL databases. AFDIL contributed a majority of the Caucasian sequences in the SWGDAM database, and for Caucasians was able to identify a suitable number of in-house samples matching the common types for completion of the study. Use of AFDIL samples for this study was approved by the Armed Forces Institute of Pathology Institutional Review Board, and all samples were rigorously anonymized to avoid any genetic information identifiable to a named individual. For the Hispanic and African American groups some additional samples matching common types had to be obtained. For this purpose, new control region databases were established with samples obtained from collaborators. Data collection, review, and archiving was achieved using the same robust, automated system as was developed for the mtGenome sequencing. The new databases were: 249 African Americans, and four regional Hispanic samples: Florida (n=96), Texas (n=200), Midwest (n=128) and New York (n=151). The African American samples and the Midwest Hispanics were from the NIST population collection for which many other forensic markers have also been typed. The new mtDNA control region population databases have been submitted to EMPOP (www.empop.org), a new online searchable international mtDNA population database. Pending final crosschecking of data between EMPOP and AFDIL, the data will be submitted to GenBank, donated to the SWGDAM database, and prepared for publication.

3) **Regional Substructure in US mtDNA Population Databases**

Few studies have investigated the level of heterogeneity existing within regional population subsamples from the major "ethnic" groups as traditionally considered in forensic databases. "Hispanics" are poorly defined, and can represent individuals of Native American, Caribbean, Central American, South American, and European heritage, with mtDNA contributions of Native American, European, and African origin. We obtained four regionally distinct Hispanic population samples for the control region databasing portion of this grant, permitting an initial investigation into the degree to which regional Hispanic samples may differ from one another. The results were rather striking, but in line with expectation: the different samples reflect divergent mtDNA distributions, reflecting a different history of migration and admixture within each of the samples. Figure 3 shows components of each population sample, broken down by categorization of the mtDNA types into continental origin, by virtue of identifiable sequence motifs associated with mtDNA haplogroups.

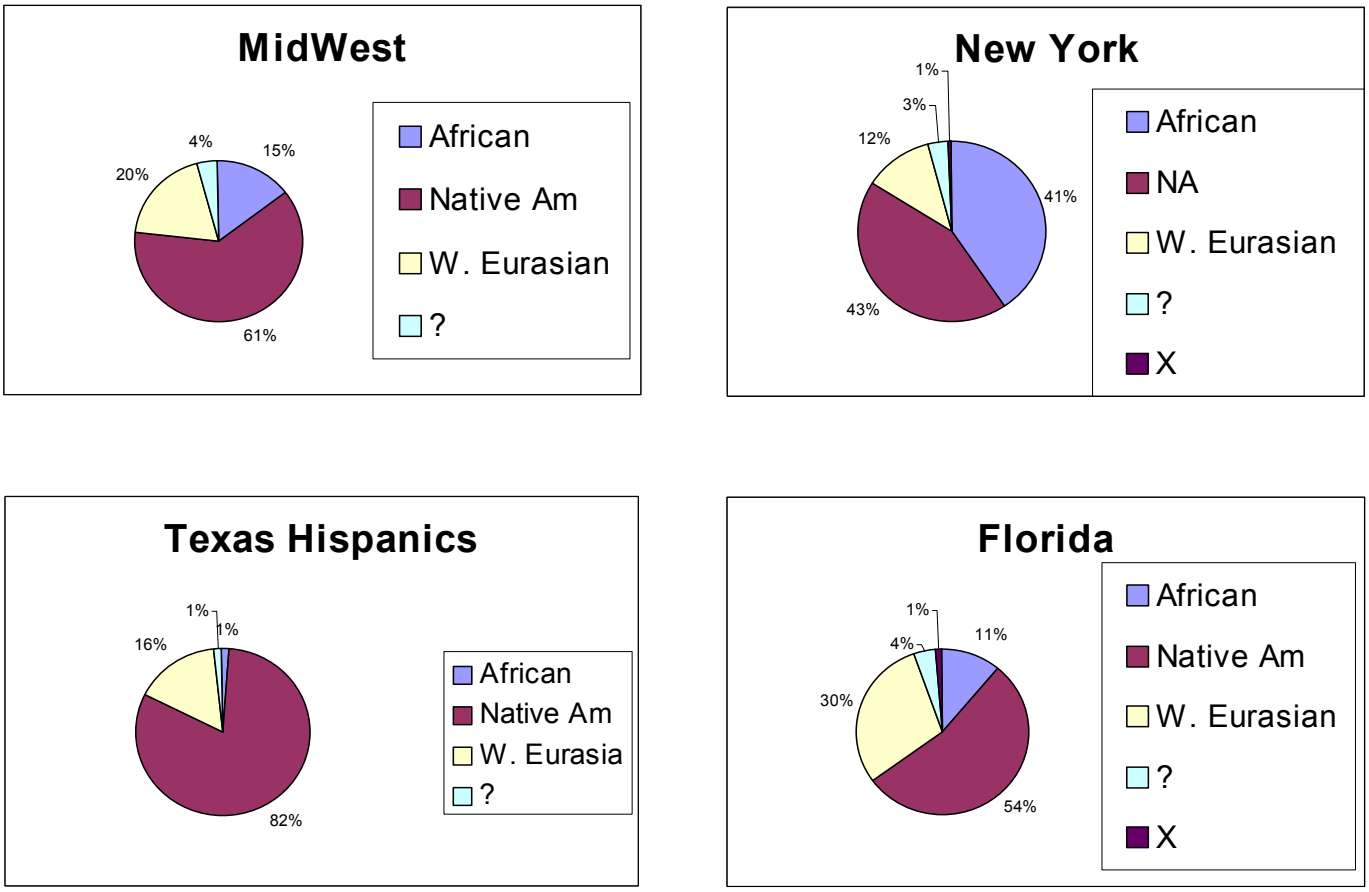


Figure 3. Major haplogroup affiliations for regional Hispanic population samples.
Database sizes were Midwest (128), New York (151), Texas (200), and Florida (96).

Each of the samples reflect differential composition of haplogroups of Native American, European, and African origin that seem consistent, broadly, with trends one might expect. For example, the New York population sample is comprised of more than 40% African haplotypes probably reflecting a stronger Afro-Caribbean contribution, while the Texas sample shows almost no African mtDNA types, with a strong predominance of Native American mtDNAs, probably reflecting a primarily Mexican/Central American contribution.

4) SNP Assays Suitable for Forensic Applications

Having identified many sites over the entire mtGenome capable of providing additional forensic discrimination, the next project goal was to develop an assay system and conceptual approach allowing the information to be accessed in useful manner, bearing in mind the real world constraints of forensic casework samples. Since mtDNA testing is normally dictated by the advanced degradation of the sample, with low amounts of fragmented DNA, there is a clear requirement that the testing method be highly sensitive and use short amplicons. Additionally, a multiplex amplification approach is highly desirable given the typical limitations of extract volume from degraded samples; and, because of the prospects for heteroplasmy and/or mixture, the assay must be able to quantitatively represent these conditions to some practical extent.

The distribution of discriminatory sites, and their specificity to particular HV-types, indicated that the best approach for a typing assay would be to target particular strategic SNPs, rather than sequence broad sections of the mtDNA coding region (considerations relating to this are presented in more detail in Coble et al., 2005). Moreover, we were guided throughout by a desire to usefully complement the current state of the art of mtDNA testing (sequence analysis in HV1-HV2, or sometimes in adjacent areas of the control region), rather than to redesign mtDNA typing in its entirety (as would be conceptually possible using a dense SNP typing approach, for example). The intended goal was to identify a suitable, off-the-shelf SNP typing platform that would be relatively easily accessible to wide variety of practitioners, and to adapt it to a number of assays that practitioners could turn to in particular circumstances where common HV-types are encountered to maximize the chance for increased discrimination with just one or a few additional multiplex amplifications.

The initial focus for a readily available SNP assay platform was on the TaqMan® fluorogenic system by ABI. This system was investigated in collaboration with Harrald Niederstätter and Walther Parson (University of Innsbruck) regarding sensitivity and mixture detection, and found to be quite suitable (results of these analyses were published in Niederstätter et al., 2005). However, the system was not amenable to multiplexing, and the data analysis and manipulation was found to be cumbersome. In collaboration with Peter Vallone (NIST) we began working with the SNaPShot™ allele specific primer extension (ASPE) assay, and this became the platform for further development.

We designed, optimized, and performed developmental validation of the ASPE assay for eight multiplex panels that target common HV types in Europeans. Each multiplex was tested for sensitivity, mixture/heteroplasmy detection, and use on degraded skeletal remains extracts. The details of the work for Multiplex "A," targeting the most common Caucasian HV type were published in Vallone *et al*. (2004). Details of the stages of design, optimization, and testing for the other multiplex panels, are included in *Experimental Details and Main Findings* below. The SNP assay panels are intended to be used when needed for particular HV types or their near relatives, when encountered in casework. Because of variation in the number of discriminatory SNPs identified for particular HV types, and some overlap in SNPs that discriminate within multiple common HV types, some HV types are targeted by more than one SNP panel, and some SNP panels are designed for more than one HV type (details below). Optimization for one of the Caucasian panels, D, has not been completed as of the expiration of the grant period, but will be completed and included in future publications. Following the model described above for Caucasian HV types, similar panels have been devised for common African American and Hispanic HV types, although SNP assays have not been designed or tested.

While conceptually straightforward, and—once optimized—simple in practice, multiplex SNaPShot ASPE assay development for practical forensic use can be problematic and lengthy. While it is comparatively simple to devise a "quick and dirty" ASPE assay where an investigator can infer results while ignoring many and various artifact peaks, bringing the assay to a stage where standardized interpretation guidelines typical of forensic testing can be applied is substantially more challenging. There are a variety of reasons for this, such as primer interactions and PCR competitive effects that are only evident in later stages in the development process. The single largest difficulty in our experience was in obtaining high quality preparations of the many, sometimes quite large, primers used in the assay. The assay is rather prone to artifact due to minor problems or variation in primer synthesis. Sometimes the causes are identifiable (minor components of incompletely synthesized extension primers; inaccurate quantitation by manufacturer), sometimes not (unusual artifact peaks or blobs resulting from a new lot of primer).

The ASPE assays required additional minor modifications (increasing Taq and amplification volume) to become suitable for typing of degraded sample extracts. However, once optimized, the assays were highly sensitive and performed well with even highly degraded samples, as the amplicon sizes are small and the detection platform inherently sensitive. AFDIL has used Multiplex A successfully in quite a few investigatory cases where additional discrimination was required. The utility in these cases was to exclude multiple matching reference families, or to confirm association among multiple elements. The utility of these assays, and of the approach of selectively augmenting forensic resolution in a directed manner with specific SNP panels, has therefore been clearly established.

5) **SNP Panel Population Databases**

For seven of the eight Caucasian multiplex panels, databasing was performed on a random population sample of 284 US Caucasian individuals. These samples were from the population reference collection held by NIST, for which a large number of other forensic loci have been typed (http://www.cstl.nist.gov/div831/strbase/NISTpop.htm). The purpose of this databasing was to assess the effect of the various Caucasian ASPE assay panels on random population samples (see results below in *Experimental Details and Main Findings*). These panels do not provide particularly powerful stand-alone resolution, as was expected since they were not designed for this purpose. For example, multiplex A by itself resolves 281 individuals into only 17 different types. However, when combined with HV sequence data, as intended, our SNP database results indicate a significant utility for the assays (detailed information for all the panels is presented in *Experimental Details and Main Findings,* below). 46 individuals in the NIST Caucasian database match either the most common Caucasian type (H:1, present in 6% of this database), or differ from it by a single base, with the 46 individuals representing 10 different HV types. As designed, one might turn to multiplex A for additional resolution for any of these, which would resolving the 10 HV types into 22 HV/SNP types, 14 of which are now unique in the database. Four of the H:1 sequences would be resolved to "unique in the database" (i.e. $\geq$ ~ 0.4%), for a ~15-fold increase in evidentiary significance.

6) SNP Assays for Haplogroup Attribution

One goal of the second, supplementary portion of this project was the development of SNP panels based on the mtDNA coding region that would permit haplogroups to be unambiguously assigned. Work was performed on this in collaboration with Anita Brandstätter and Walther Parson (University of Innsbruck), resulting in an ASPE assay for 16 SNPs that

permit the classification of major W. European haplogroups and sub-haplogroups (Brandstätter, *et al*. [2003] and [2004]). As our project progressed, however, no further effort was applied to this goal for two reasons: 1) greater than anticipated effort in SNP assay development for the discriminatory panels, and 2) a number of other groups became active with similar projects, such as the SNPforID Consortium (www.snpforid.org, see also, e.g., Grignani *et al*. [2005]), and there seemed little benefit in competing or performing redundant work.

Executive Summary, Conclusions and Future Directions

This project has resulted in the development of a series of multiplex SNP assays that forensic scientists can use for significantly enhancing the discrimination when common Caucasian HV types are encountered in casework. It is in such cases that the need for increased discrimination will occur most frequently and will be the most pressing. 506 entire mtDNA genomes were sequenced for this study, and useful panels of discriminatory SNPs were also identified for common HV types with African Americans and Hispanics. Control region databases of four regionally distinct Hispanic samples indicate that there is substantial regional differentiation within individuals ascribed to this heterogeneous classification. This indicates the desirability of increasing the sample size and regional representation within US forensic mtDNA databases.

The data generated from this project will likely be of primary significance as the field progresses toward routine utilization of coding region variation in forensic mtDNA testing. Maximum utility would be achieved with commercialization of robust SNP typing kits targeting the sites we have identified, as well as other sites identified through further characterization.

Experimental Details and Main Findings

Control Region Sequencing.

Partly as a result of the system developed in this project for robotic sequencing of the mtGenome, the AFDIL research section has a highly refined system for high throughput control region sequencing. This employs robotic automation in chemistry steps, and an effective bioinformatic data management system that facilitates a rapid but robust data review process, and preserves data integrity by removing any manual transcription of data. In the end, data is stored in a locked database to prevent corruption and inadvertent over-writing. Primary sequence alignment and editing is performed only a single time by a single scientist, but the resulting Sequencher layout and consensus sequence export is independently reviewed at various stages by two additional scientists. One reason this system works so effectively is that the sequence coverage was designed to be highly redundant, with a good chance of obtaining full double strand coverage even in cases where C-stretch length heteroplasmy is present in HV1 and/or HV2 (see figure 4). This results in larger than necessary reagent costs for "easy" samples, but more than pays for itself by reducing effort in the rate limiting step (data analysis is simplified by very strong support for base calls), and decreasing the amount of specific re-do reactions which are time consuming and error-prone. When human handling is required, as in blood stain punching to plates using a Wallac hole puncher (figure 5 shows the robotic processing steps in the CR sequencing process), or particular wells of plates are accessed for re-do's, three scientists are involved for a redundant check that no errors are introduced. Amplification, cycle sequencing reaction conditions, and clean-up were as described in the section below on *Whole Genome Sequencing*. Amplification and sequencing primer sequences are listed in Appendix I.

The following steps outline the data analysis and review of CR sequence data: 1) Scientist 1 performs alignment in Sequencher (GeneCodes, Ann Arbor, MI, version 4.1.4Fb19), checking for unambiguous double strand coverage, and generating a list of differences ("polymorphisms" by colloquial reference) from the Cambridge Reference Sequence (CRS; Andrews, 1992); 2) Scientist 2 reviews strand coverage, independently deduces the polymorphisms of the consensus sequence, and generates a text file of the polymorphisms; 3) Scientist 1 exports the text file to the database and compares this to the list he/she independently generated; once confirmed, Scientist 1 clicks the "Initial Review" button on the database GUI; 4) Scientist 3 confirms the data as entered into the database with reference to the origin polymorphism lists, and elevates the status to "Final Review" whereby the sequence becomes locked in the database and can only be modified by an administrator.

Our system of data analysis has been validated through complete independent reanalysis of multiple databases, performed by the University of Innsbruck, Institute of Legal Medicine, when our samples were submitted for the EMPOP database (www.empop.org). 100% concordance in results were obtained. Additional quality control analyses using phylogenetic methods have been applied to one of our databases, published in Brandstätter *et al*. (2004). The rigor we have applied to ensuring the high quality or our final mtDNA sequence data is motivated by the huge amount of attention that mtDNA database errors have received in recent high profile publications (e.g. Bandelt et al., 2002, 2004a, 2004b).

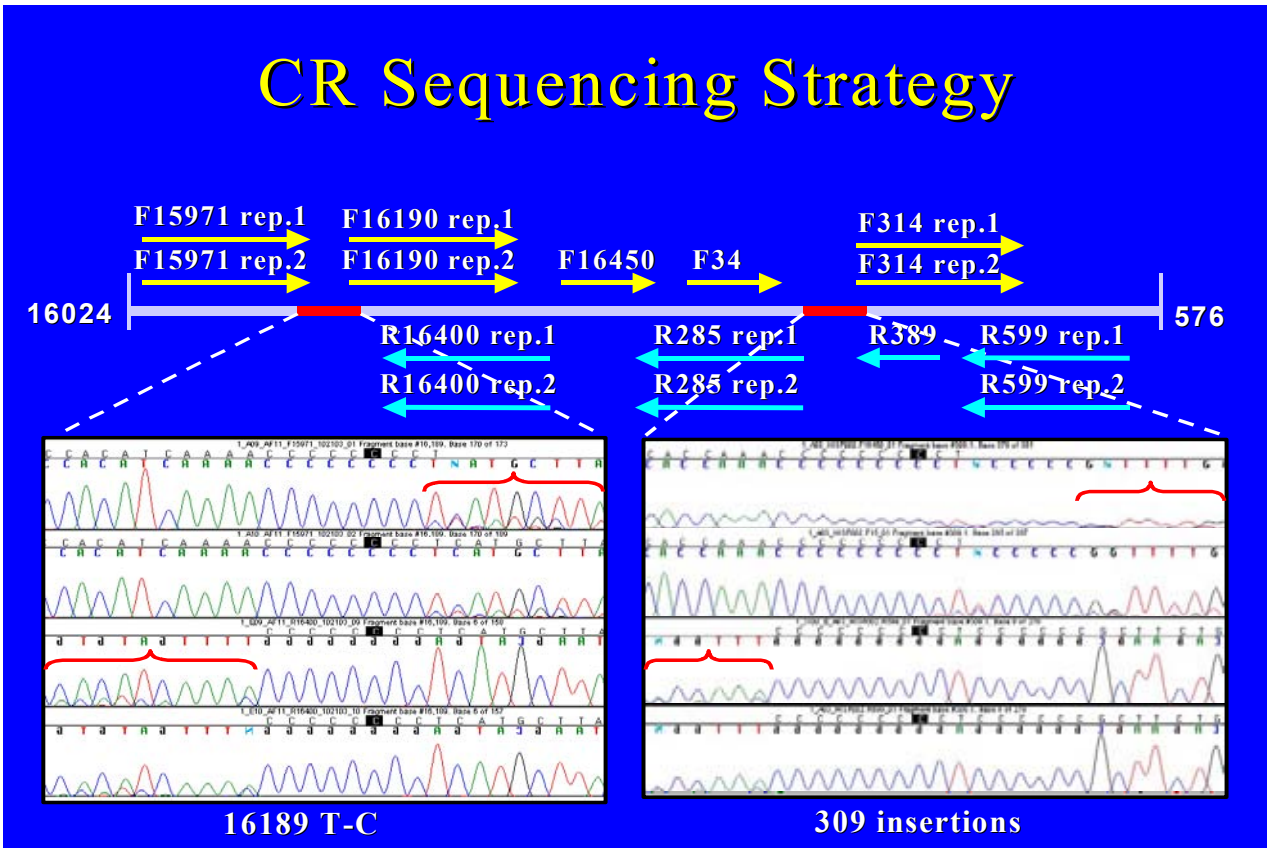


Figure 4. Control Region Sequencing Strategy.
Sequencing primer placement strategy for full double stranded confirmation of CR sequence even when C-stretch length heteroplasmy is present (HV1 and HV2 C-stretches shown). Some primers are used twice (rep.1 and rep.2).

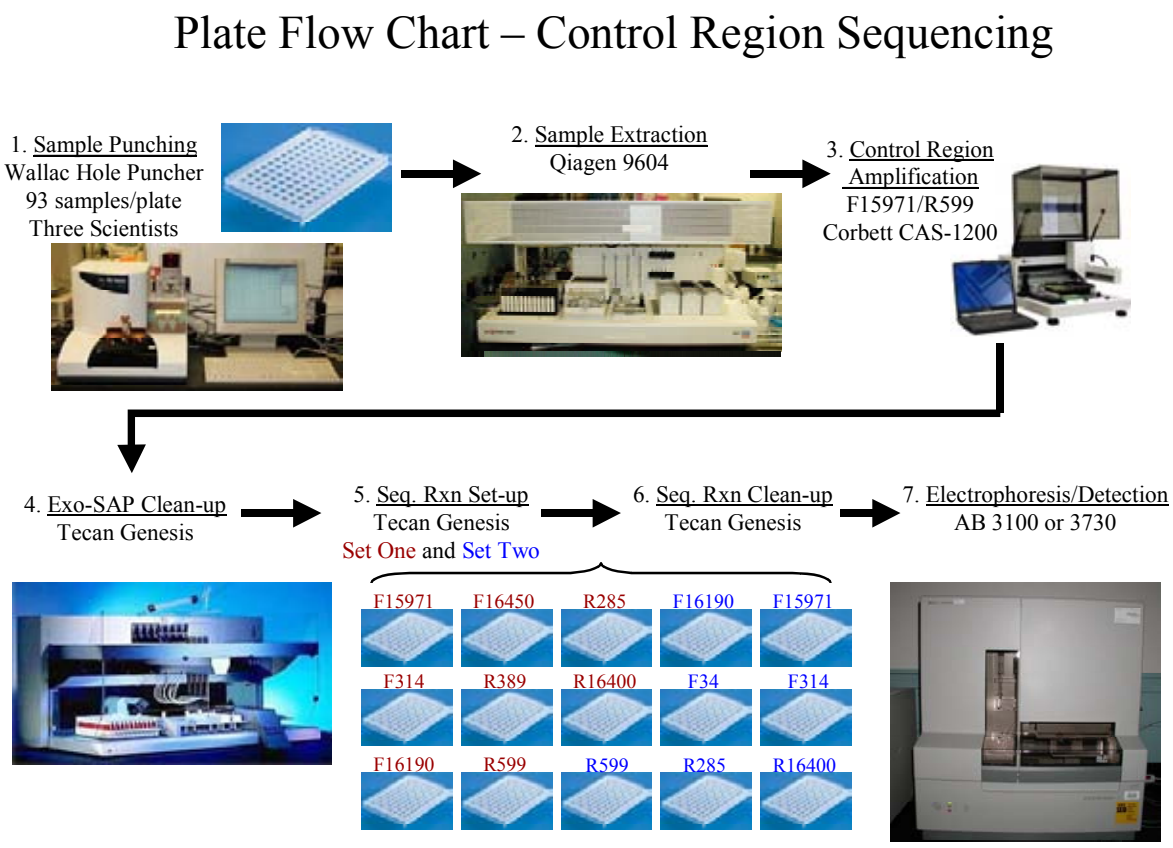


Figure 5. Control Region Sequencing Overview
Overview of robotic laboratory processing for high throughput CR sequencing.

Regional Substructure in US mtDNA Population Databases

Figure 3 of the *Executive Summary* (above) depicts the results of four regional "Hispanic" population samples, at the level of haplogroups ascribed to continental origin. It is clear at a glance that these samples do not reflect subsets of a single homogeneous population, due to pronounced deviation in representation at the haplogroup level. This suggests that the issue of regional variation should be addressed with further sampling amongst various groups that are currently maintained as single population designations. However, the practical significance of such subdivision in forensic applications is something that should be examined

carefully. Given large enough sample sizes, we can anticipate with certainty that significant differentiation will be the rule among any regional subsamples, but the practical effect may be low when examined on the haplotype (rather than haplogroup) level. As an extreme example for the purposes of discussion, if each of the regional Hispanic samples were comprised of sequences that were unique within and between the databases, the fact that these are "significantly differentiated subpopulations" would by itself have no effect on forensic reporting. However, it does appear that there would be practical difficulties associated with simply combining these separate databases together. For example, the most common type in the Texas Hispanic dataset is found in 8/200 (4%) of that database, but is not present at all in either the New York or Florida Hispanic databases. Applying confidence limits to reporting these database comparisons (following Holland and Parsons [1999]), for the most common Texas Hispanic sequence one would report an 95% upper bound for the Texas sample of 4.4%, while the upper bound value on the four databases pooled together (10/575) would be reported as 2.5%. It could be debated whether these two values would be significantly different from one another with respect to any meaningful effect on a juror's consideration. However, this does not really address the larger question: when, and under what criteria should databases be pooled together or retained as distinct?

It is beyond the scope and purpose of this report to deal with these issues in depth, although a more detailed analysis is planned for a future publication. Resolution of these issues will likely require both dialog among practitioners and population geneticists, as well as a great deal more data to frame the question.

Whole Genome Sequencing

Pairwise Comparisons and Samples.

Note on HV type naming for this project: HV types were given names indicating their major haplogroup (and in some cases, subhaplogroup) of origin, followed by a number distinguishing that type from others of the same haplogroup. The haplogroup designation and sequential number (e.g. L3e:1) are separated by a colon in an attempt to avoid confusion with other subhaplogroup identifiers.

Pairwise comparisons of the relevant SWGDAM mtDNA forensic databases were performed to identify the common HV types for each of the Caucasian, African-American, and Hispanic populations. For each pairwise comparison, the regions were restricted to positions 16024-16365 and 73-340. Insertions at positions 16183 and 309 were ignored. HV types that occurred at a frequency of greater than 0.5% in each population were selected as the target "common" HV-types. With reference to diagnostic polymorphisms associated with continental origin, in Hispanics, 13 of the common HV types were of Native American origin, four were of European origin, and two were of African origin. The common types within African Americans were all of African continental origin, and the common types within US Caucasians were all of European origin.

Samples from each of the common HV-types for whole genome sequencing were obtained from one of three sources: 1) AFDIL family reference samples, 2) CR population databasing samples, and 3) samples provided by the FBI from the SWGDAM database. The AFDIL family reference databases were searched for samples matching the common HV-types. When samples were identified, the bloodstain cards were obtained and the samples anonymized prior to processing. The control region population databases generated from the NIST African-

American and Hispanic samples, as well as Hispanic samples from other sources, were combed for samples that matched the common HV-types. Thus the number of samples sequenced for each of the common HV-types was limited by sample availability from the family reference and CR population databases. Tables 1-3 list the HV sequences, population frequency (within the SWGDAM database for each ethnic group), and number of individuals sequenced for the entire mtGenome.

HVI/HVII Types	% of Population Matching HVI/HVII Type	# of Samples Sequenced	HV-type Defining Polymorphisms
H:1	7.92%	31	CRS
H:2	2.05%	25	152 C
H:3	1.03%	11	16129 A
H:4	0.54%	8	16263 C
H:5	1.03%	12	16304 C
H:6	0.91%	11	73 G
H:7	0.66%	7	73 G 16162 G 16209 C
J:1	1.21%	15	73 G 185 A 228 A 295 T 16069 T 16126 C
J:2	0.66%	8	73 G 228 A 295 T 16069 T 16126 C
J:3	0.79%	13	73 G 185 A 188G 228 A 295 T 16069 T 16126 C
J:4	0.36%	8	73 G 242 T 295 T 16069 T 16126 C 16145 A 16172 C 16222 T 16261 T
T:1	1.33%	21	73 G 16126 C 16294 T 16296 T 16304 C
T:2	0.60%	10	73 G 152 C 195 C 16126 C 16163 G 16186 T 16189 C 16294 T
T:3	0.66%	8	73 G 16126 C 16294 T 16296 T
V:1	1.21%	25	16298 C
K:1	1.09%	14	73 G 146 C 152 C 16224 C 16311 C
K:2	0.60%	7	73 G 16093 C 16224 C 16311 C
K:3	0.60%	7	73 G 16224 C 16311 C
	23.25%	241	

Table 1. Common US Caucasian HV-types sequenced for the entire mtDNA genome.
Bottom row lists the total number of common HV types, the cumulative percentage of the population these types represent, and the total number of mtGenomes sequenced.

HV-types	% of Population Matching HVI/HVII Type	# of Samples Sequenced	HV-type Defining Polymorphisms
A2:1	3.57%	15	73G, 146C, 153G, 235G, 16111T, 16223T, 16290T, 16319A, 16362C
A2:2	1.35%	7	73G, 146C, 153G, 235G, 16111T, 16129A, 16223T, 16290T, 16319A, 16362C
A2:3	1.35%	15	73G, 146C, 153G, 235G, 16223T, 16290T, 16319A, 16362C
A2:4	1.11%	4	73G, 146C, 152C, 153G, 214G, 235G, 16083T, 16111T, 16223T, 16256T, 16290T, 16319A, 16362C
A2:5	0.56%	1	73G, 146C, 152C, 153G, 235G, 16111T, 16223T, 16290T, 16319A, 16362C
A2:6	0.56%	5	73G, 146C, 153G, 235G, 16111T, 16223T, 16274A, 16290T, 16319A, 16362C
B2:1	1.11%	7	73G, 16182C, 16183C, 16189C, 16217C
B2:2	0.63%	6	73G, 16111T, 16183C, 16189C, 16217C
C1:1	2.62%	12	73G, 249D, 290D, 291D, 16223T, 16298C, 16325C, 16327T (-263G)
C1:2	1.74%	13	73G, 249D, 290D, 291D, 16223T, 16298C, 16325C, 16327T
C1:3	0.63%	3	73G, 249D, 290D, 291D, 16051G, 16188T, 16204A, 16223T, 16325C, 16327T, 16362C
C1:4	0.56%	2	73G, 143A, 249D, 290D, 291D, 16086C, 16183C, 16189C, 16223T, 16278T, 16298C, 16325C, 16327T
D1:1	1.98%	11	73G, 16223T, 16325C, 16362C
U5:1	0.56%	4	73G, 150T, 16189C, 16192T, 16270T, 16320T
K:4	2.06%	12	73G, 146C, 152C, 207A, 16182C, 16183C, 16189C, 16224C, 16311C
J:5	0.79%	5	73G, 295T, 16069T, 16126C, 16145A, 16222T, 16261T
L1b:3	0.63%	2	73G, 152C, 182T, 185T, 195C, 228A, 247A, 16126C, 16187T, 16189C, 16223T, 16264T, 16270T, 16278T, 16293G, 16311C
L3e:5	0.56%	1	90A, 97A, 106D, 107D, 108D, 109D, 110D, 111D, 150T, 189G, 200G, 16223T, 16327T
18	22.37%	125	
H:1	1.03%	31	rCRS
19	23.40%	156	

Table 2. Common US Hispanic HV types sequenced for the entire mtDNA genome.
Bottom row lists the total number of common HV types, the cumulative percentage of the population these types represent, and the total number of mtGenomes sequenced. The W. European H:1 type was one of the common HV types in Hispanics, but no additional sequencing was performed beyond the 31 individuals listed Table 1 (for Caucasians).

HVI/HVII Types	% of Population Matching HVI/HVII Type	# of Samples Sequenced	HV-type Defining Polymorphisms
L0a:1	0.29%	4	93G, 95C, 185A, 189G, 236C, 247A, 16093C, 16129A, 16148T, 16168T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16278T, 16293G, 16311C, 16320T
L0a:2	0.64%	4	93G, 185A, 189G, 200G, 236C, 247A, 16129A, 16148T, 16168T, 16172C, 16187T, 16188G, 16189C, 16223T, 16230G, 16311C, 16320T
L1b:1	2.22%	14	73G, 152C, 182T, 185T, 189G, 195C, 247A, 16126C, 16187T, 16189C, 16223T, 16264T, 16270T, 16278T, 16293G, 16311C
L1b:2	1.07%	5	73G, 152C, 182T, 185T, 195C, 247A, 16126C, 16187T, 16189C, 16223T, 16264T, 16270T, 16278T, 16311C
L2a:1	2.29%	10	73G, 146C, 152C, 195C, 16223T, 16278T, 16294T, 16309G
L2a:2	0.72%	4	73G, 146C, 152C, 195C, 16189C, 16223T, 16278T, 16294T, 16309G
L2a:3	0.79%	4	73G, 143A, 146C, 152C, 195C, 16189C, 16192T, 16223T, 16278T, 16294T, 16309G
L2a:4	0.93%	10	73G, 143A, 146C, 152C, 195C, 16223T, 16278T, 16294T, 16309G
L2a:5	1.15%	16	73G, 146C, 152C, 195C, 16189C, 16192T, 16223T, 16278T, 16294T, 16309G
L2a:6	1.15%	10	73G, 146C, 152C, 195C, 16223T, 16278T, 16286T, 16294T, 16309G
L2b:1	0.93%	8	73G, 150T, 152C, 182T, 195C, 198T, 204C, 16114A, 16129A, 16213A, 16223T, 16278T, 16355T, 16362C
L2c:1	0.50%	4	73G, 93G, 146C, 150T, 152C, 182T, 195C, 198T, 325T, 16223T, 16264T, 16278T
L3b:1	0.43%	3	73G, 16124C, 16223T, 16278T, 16311C, 16362C
L3b:2	1.00%	5	73G, 16124C, 16223T, 16278T, 16362C
L3e:1	1.22%	12	73G, 150T, 152C, 195C, 16172C, 16183C, 16189C, 16223T, 16320T
L3e:2	1.22%	8	73G, 150T, 195C, 16223T, 16265T
L3e:3	0.72%	3	73G, 150T, 195C, 16172C, 16183C, 16189C, 16223T, 16320T
L3e:4	0.43%	3	73G, 150T, 195C, 16172C, 16189C, 16223T, 16320T
L3e:6	1.00%	8	73G, 150T, 195C, 198T, 16223T, 16320T
L3f:1	0.72%	5	73G, 189G, 200G, 16129A, 16209C, 16223T, 16292T, 16295T, 16311C
20	19.42%	140	

Table 3. Common African American HV types sequenced for the entire mtDNA genome. Bottom row lists the total number of common HV types, the cumulative percentage of the population these types represent, and the total number of mtGenomes sequenced.

Whole Genome Sequencing

In total, 506 entire mtDNA genomes were sequenced for this project: 241 Caucasian individuals, representing 18 common HV-types; 140 African American individuals, representing 20 common HV-types; and 125 Hispanic individuals representing 18 common HV-types. These

sequences have been submitted to GenBank with the following accession numbers: Caucasians, AY495090-AY495330; African Americans, DQ304897-DQ30506; Hispanics, DQ282387-DQ282511.

Most samples identified for whole genome analysis were cut from blood cards or bloodstains on cloth using a Wallac Hole Puncher and extracted on the Qiagen 9604 robot. Three scientists confirmed the identity of each sample and its placement in the 96-well plate. Alternatively, extracts were provided to us from NIST. In this case, the Corbett-CAS 1200 robot was used to aliquot these extracts into 96-well plates used for processing. At this initial sampling step, each sample was given a unique whole genome identifier based on the sample's HVI/HVII type, as determined through previous control region sequencing. After extraction, samples were amplified using the Corbett CAS-1200 robot. Figure 6 depicts the location of the individual amplicons used to amplify the entire mtGenome for each sample, seven per 96-well plate, and one amplification negative.

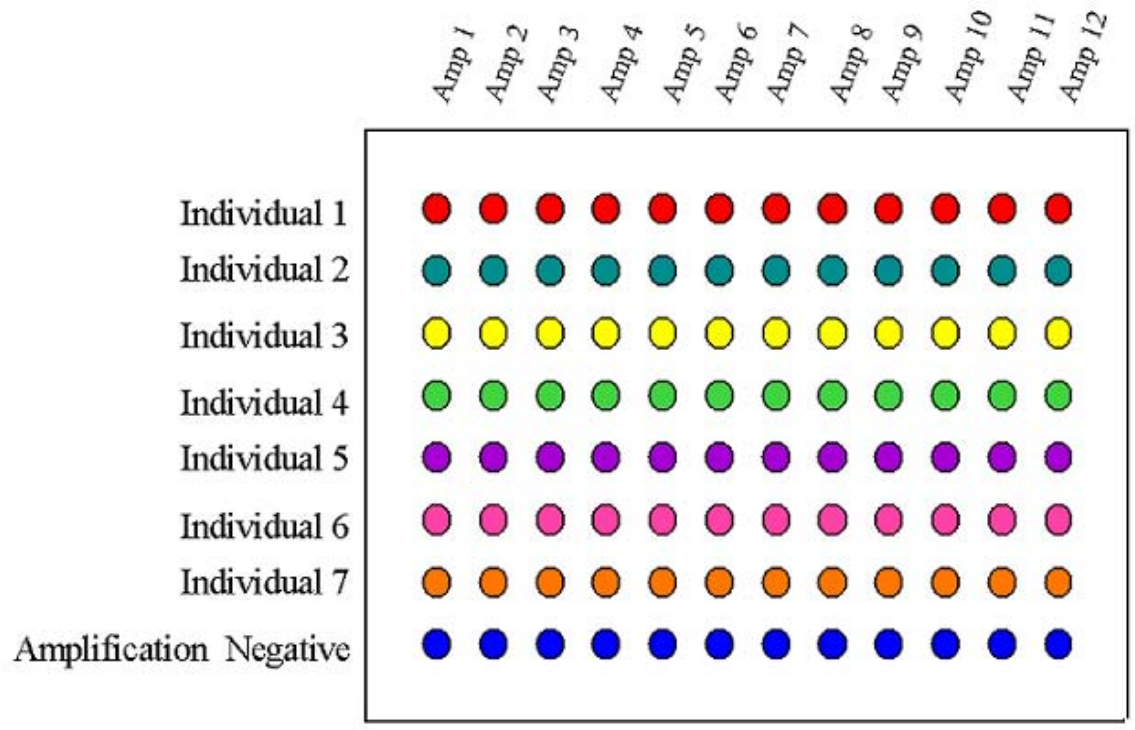


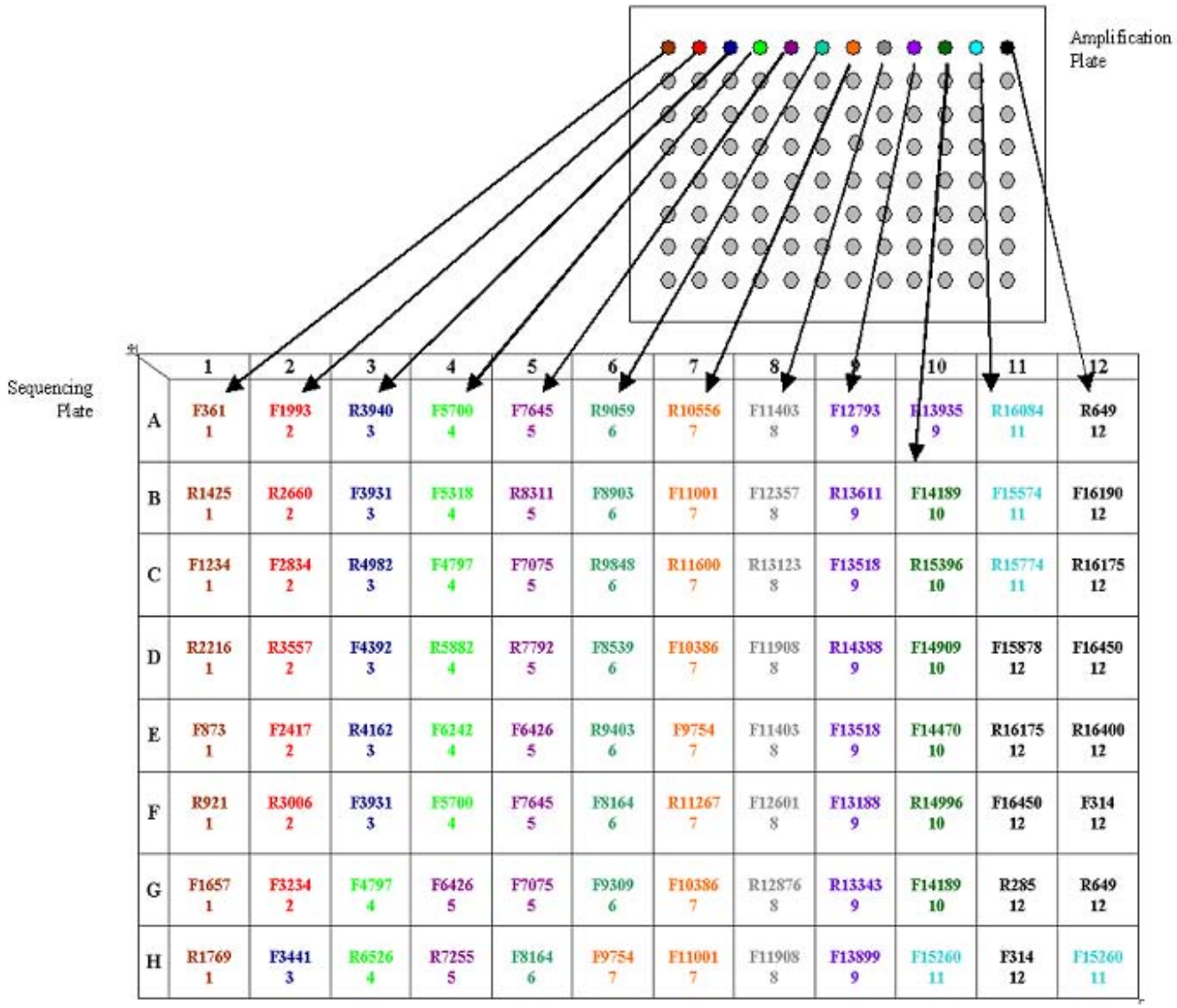
Figure 6. Whole Genome Amplification Strategy.
From a 96-well plate of extracts, seven individuals (and one amplification negative) are each divided into twelve amplicons of varying sizes that span the 16569 base pairs of the entire mtDNA genome.

Amplification primer sets were designed to overlap adjacent amplicons in order to provide redundant sequence data and prevent amplicon switches ("phantom recombination") between samples (mtGenome sequencing is described in more detail in Coble et al [2004]). Amplicons ranged from 824 to 1885 bases in length, with a minimum overlap between adjacent amplicons of 94 bases (maximum 330 bases). Amplification reactions were performed with 3 ul of DNA in 50ul reactions containing components listed in Table 4. Thermal cycling conditions were 96^{o}C for 10 minutes, 40 cycles of 94^{o}C for 15 seconds, 56^{o}C for 30 seconds, and 72^{o}C for 1 minute, followed by 72^{o}C for 7 minutes and a 4^{o}C soak on the GeneAmp PCR System 9700 (Applied Biosystems). The presence of PCR product was confirmed on agarose gels, and post-PCR clean-up was performed using Exo and SAP.

Amplification Component	**Final Concentration (in 50ul reaction)**
$MgCl_2$	1.5uM
Forward and Reverse Primers	0.4uM each primer
dNTPs	0.2uM
TaqGold	2.5 units (0.0025uM)
10x PCR Buffer	1x

Table 4: Amplification Reaction Components and Final Concentrations.
Amplification components included $MgCl_2$, Primers, dNTPs, TaqGold polymerase, and a PCR buffer.

From the amplification plate containing amplicons from seven individuals, cycle sequencing plates were set up using the Tecan Genesis BioRobot with 96 sequencing primers targeting a single individual per plate; seven sequencing plates were required per amplification plate. Five to 12 sequencing primers were used per amplicon (Figure 7). This single sequencing plate was designed to produce the entire sequence for an individual, but as the project proceeded it was determined that full coverage was rarely obtained with the single plate, so a secondary sequencing plate was routinely added with additional primers for more robust coverage; in this case a single additional plate was used for all seven individuals on the original amplification plate, with 12 additional sequencing reactions per individual (Figure 8). In total, 108 sequencing primers were used per mtGenome, and it normally required a small number of additional reactions, set up by hand on a case-to-case basis, to fill in small gaps in the double strand coverage. For a complete list of primers and primer sequences used in the amplification reactions, see Appendix II.



	1	2	3	4	5	6	7	8	9	10	11	12
A	F361 1	F1993 2	R3940 3	F5700 4	F7645 5	R9059 6	R10556 7	F11403 8	F12793 9	F13935 9	R16084 11	R649 12
B	R1425 1	R2660 2	F3931 3	F5318 4	R8311 5	F8903 6	F11001 7	F12357 8	R13611 9	F14189 10	F15574 11	F16190 12
C	F1234 1	F2834 2	R4982 3	F4797 4	F7075 5	R9848 6	R11600 7	R13123 8	F13518 9	R15396 10	R15774 11	R16175 12
D	R2216 1	R3557 2	F4392 3	R5882 4	R7792 5	F8539 6	F10386 7	F11908 8	R14388 9	F14909 10	F15878 12	F16450 12
E	F873 1	F2417 2	R4162 3	F6242 4	F6426 5	R9403 6	F9754 7	F11403 8	F13518 9	F14470 10	R16175 12	R16400 12
F	R921 1	R3006 2	F3931 3	F5700 4	F7645 5	F8164 6	R11267 7	F12601 8	F13188 9	R14996 10	F16450 12	F314 12
G	F1657 1	F3234 2	F4797 4	F6426 5	F7075 5	F9309 6	F10386 7	R12876 8	R13343 9	F14189 10	R285 12	R649 12
H	R1769 1	F3441 3	R6526 4	R7255 5	F8164 6	F9754 7	F11001 7	F11908 8	F13899 9	F15260 11	F314 12	F15260 11

Figure 7. Whole Genome Sequencing Strategy.
In the amplification plate, each row contains one individual. For the sequencing reaction, each individual covers one plate. Amplification product from each of the twelve amplicons, located in the twelve columns of the amplification plate, are distributed across the sequencing plate by the Tecan Genesis according to the location of their corresponding sequencing primers. Here, each color represents a different amplicon from a single individual.

	1	2	3	4	5	6	7	8	9	10	11	12
A	F756 1	F10127 7	R6526 4	F756 1	F10127 7	R6526 4	F756 1	F10127 7	R6526 4	F756 1	F10127 7	
B	F756 1	R11166 7	R7255 5	F756 1	R11166 7	R7255 5	F756 1	R11166 7	R7255 5	F756 1	R11166 7	
C	F2105 2	R11166 7	R10275 7	F2105 2	R11166 7	R10275 7	F2105 2	R11166 7	R10275 7	F2105 2	R11166 7	
D	F2834 2	R16400 12	R10556 7	F2834 2	R16400 12	R10556 7	F2834 2	R16400 12	R10556 7	F2834 2	R16400 12	
E	R6526 4	F756 1	F10127 7	R6526 4	F756 1	F10127 7	R6526 4	F756 1	F10127 7	R6526 4		
F	R7255 5	F756 1	R11166 7	R7255 5	F756 1	R11166 7	R7255 5	F756 1	R11166 7	R7255 5		
G	R10275 7	F2105 2	R11166 7	R10275 7	F2105 2	R11166 7	R10275 7	F2105 2	R11166 7	R10275 7		
H	R10556 7	F2834 2	R16400 12	R10556 7	F2834 2	R16400 12	R10556 7	F2834 2	R16400 12	R10556 7		

Figure 8: Additional Whole G Sequencing Plate.
An eighth sequencing plate was added to the processing of each seven individuals after it was noticed that gaps remained in the sequence data when using only 96 sequences. Twelve additional primers in six amplicons were processed per individual, for a total of 84 additional sequencing reactions per seven individuals.

Sequencing reactions were performed as 20ul reactions (2ul of amplification product) as listed in Table 4. Thermal cycling conditions were 96°C for 1 minute, 25 cycles of 96°C for 15 seconds, 50°C for 5 seconds, and 60°C for 2 minutes, followed by a 4°C soak on the GeneAmp PCR System 9700 (Applied Biosystems).

Sequencing Component	Volume (20ul reaction total)
dH_2O	8ul
10uM Primer	2ul
Sequencing Buffer	6ul
Big Dye	1.5ul
dGTP	0.5ul

Table 4: Sequencing Reaction Components and Volumes.
Sequencing reaction components included dH_2O, Primer, Sequencing buffer, BigDye (ABI) and dGTP.

For sequencing reaction clean-up, the samples were run through a Performa V3 96-well short plate (Edge Biosystems), dried down, and resuspended in 10ul Hi-Di with the aid of the Tecan Genesis for sample transfer steps. The samples were then run on the ABI 3730 Genetic Sequencer using a 50-cm capillary for maximum sequence recovery. Sequences were analyzed in Sequencing Analysis 5.1.1. Sequence alignments were created using Sequencher 4.1.4Fb19. Data analysis, review, and archiving followed the process described in the section on *Control Region Sequencing*, above.

Selection of SNPs from Common HV-type mtDNA Entire Genome Data and SNP Panel Development

Following completion of the whole genome sequencing, the sequence data for each population was examined to select discriminatory SNPs appropriate for assay development. An excel file containing only the polymorphic positions for each sample was generated for each target population (W. European Caucasian, African-American, and US Hispanic). Each population was examined separately to identify potential discriminatory SNPs that met the criteria that: 1) Variation was observed in more than one individual in the population. The

observed variation was either intra-type variation (more than one sample within an HV-type varied at a particular site) or inter-type variation (the position was polymorphic and showed variation in sample from more than one HV-type). 2) Variation was synonymous. These included SNPs in the Control Region outside of HVI/HVII, coding region intergenic spacers, and silent mutations in protein coding genes (first and third codon positions). One exception was made to this criterion, for position 3010, mutation at which results in a non-synonymous change in the 16S rRNA gene. 3010 is highly polymorphic in several Caucasian HV-types, and mutation at this position is not associated with any known medical disorder. Once the SNPs with potential for assay development were identified for each population, the specific samples that exhibited variation for each site was evaluated for each HV-type. In this manner, SNPs that provided redundant discrimination were identified and removed from consideration. Additional details relating to SNP selection considerations can be found in Coble (2004), Coble *et al*. (2004), Just *et al*. (2004), and Coble *et al*. (2005).

In selecting SNPs for assay development, we included the criterion that the SNPs not have the potential for phenotypic effect, i.e. amino acid changes in protein or structural changes in ribosomal or tRNAs (however, just above we described an exception for the particularly useful SNP, 3010, in the 16S rRNA gene). The basis for this criterion is that there are many diseases directly caused by mtDNA mutations or hypothesized to be correlated to variation in amino acid or structural RNA sequences (see www.mitomap.org). Detailed discussion relating to this criterion are beyond the scope of this report, but are dealt with in detail in Budowle *et al*. (2005) and Coble *et al*. (2005). AFDIL has made the practical determination that including non-synonymous SNPs as targets would pose an unacceptably high probability of having to deal with issues of medical genetic counseling. As described in Coble et al. (2005), and illustrated below,

the redundant nature of many discriminatory SNPs makes it possible to rely exclusively on synonymous or non-coding variation without losing much discrimination potential. However, we list below additional non-synonymous discriminatory SNPs that could be employed for additional resolution.

All selected SNPs for each population were organized into multiplex panels (Tables 6-8). Each panel targets specific HV-types, and when different HV-types were resolved by the same SNP targets, these were organized into the same panels when possible. SNP panels were kept relatively small (six to eleven sites per panel) due to the complexity of optimizing multiplex PCR, and as such some HV-types are targeted by more than one SNP panel. For example, 14 target SNPs were identified for Caucasian HV-type H:1. Caucasian multiplex panel A is the primary panel for type H:1 and contains the eleven sites with the greatest discriminatory power. Caucasian multiplex panel F contains the three remaining H:1 SNP targets, and can be used as a secondary panel for H:1 samples in the event that panel A did not provide the necessary discrimination.

Panel A	Panel B	Panel C	Panel D	Panel E	Panel F	Panel G	Panel H
477	477	72	482	4808	64	3826	4688
3010	3010	513	5198	5147	4745	3834	11377
4580	3915	4580	6260	9380	10211	4688	12795
4793	5004	5250	9548	9899	10394	6293	13293
5004	6776	11719	9635	11914	10685	7891	14305
7028	8592	12438	11485	15067	11377	11533	16519
7202	10394	12810	11914	16519	14470	12007	
10211	10754	14770	15355		14560	12795	
12858	11864	15833	15884		16390	15043	
14470	15340	15884	16368		14869	16390	
16519	16519	16519				16519	
11 PLEX	11 PLEX	11 PLEX	10 PLEX	7 PLEX	10 PLEX	11 PLEX	6 PLEX
Targets:	Targets:	Targets	Targets:	Targets:	Targets:	Targets:	Targets:
H:1 (primary)	H:2 (primary)	V:1 (primary)	J:1 (primary)	J:4	V:1 (secondary)	J:1 (secondary)	K:1
	H3 (primary)	H:5	J:2	T:2	H:1 (secondary)	J:3	
	H:6		K:2	T:3	H:2 (secondary)	T:1	
			K:3	H:4	H:3 (secondary)		

Table 6. Caucasian Multiplex Panels

59 non-redundant, discriminatory SNPs were organized into eight multiplex panels, ranging in size from six to eleven sites per panel. HV-types that shared several discriminatory SNPs were grouped into the same panel when feasible. Due to the large number of discriminatory sites identified for some HV-types, these HV-types are targeted by both a primary and a secondary multiplex panel. Additionally, some discriminatory SNPs are shared by many HV-types and thus appear in several multiplex panels (i.e. 16519).

Panel I	Panel J	Panel K	Panel L	Panel M	Panel N
3495	5063	6221	3852	4080	4655
5252	6932	13617	7711	5580	6378
6152	7424	13650	8155	7394	8751
7424	7702	13980	11800	7915	10286
12354	8206	15244	14182	11020	13197
12603	8856	16399	15043	11254	14560
13644	12732		16399	13884	15211
13938	14566			16519	16519
16519	16519				
9 PLEX	9 PLEX	6 PLEX	7 PLEX	8 PLEX	8 PLEX
Targets:	Targets:	Targets:	Targets:	Targets:	Targets:
L2a:1	L1a:2	L2a:4 (secondary)	L3:3	L1b:1	L1b:2
L2a:4 (primary)	L2a:5 (primary)	L2a:5 (secondary)	L3b:2	L2a:2	L2a:6
	L2b:1		L3:2	L2a:3	L3e:2
	L3e:1		L1a:1	L3e:4	

Table 7. African-American Multiplex Panels
42 non-redundant, discriminatory SNPs were organized into six multiplex panels, ranging in size from six to nine sites per panel. HV-types that shared several discriminatory SNPs were grouped into the same panel when feasible. Due to the large number of discriminatory sites identified for some HV-types, these HV-types are targeted by both a primary and a secondary multiplex panel. Additionally, some discriminatory SNPs are shared by many HV-types and thus appear in several multiplex panels (i.e. 16519).

Panel O	Panel P	Panel R
64	4790	493
5081	5081	6719
10595	6308	7241
11314	13920	7262
11653	14364	11440
12696	15629	11989
13350	15805	12696
15229	16519	16519
16391		
16519		
10 PLEX	8 PLEX	8 PLEX
Targets:	Targets:	Targets:
A2:1	A2:3	B2:1
A2:2	D1:1	C1:2
A2:6		J:5
B2:2		

Table 8. Hispanic Multiplex Panels
22 non-redundant, discriminatory SNPs were organized into three multiplex panels, ranging in size from eight to ten sites per panel. HV-types that shared several discriminatory SNPs were grouped into the same panel when feasible. Some discriminatory SNPs are shared by many HV-types and thus appear in several multiplex panels (i.e. 16519).

The number of SNPs meeting our criteria varied between the populations as well as with the number of samples sequenced for each HV-type (Tables 9-11). For example, 241 Caucasian samples sequenced from 18 HV-types resulted in a total of 59 target SNPs. A similar relative number of sites for assay development were identified for the African-American population (140 samples sequenced from 20 HV-types resulted in 40 unique discriminatory sites). However, relatively fewer sites for assay development were identified for US Hispanics (125 samples sequenced from 18 HV-types resulted in only 22 unique discriminatory sites).

HV-types	% of Population Matching HV-type	# of Samples Sequenced	# of discriminatory, non-redundant SNPs per HV-type	SNP Information				
				# 1st Codon Position	# 3rd Codon Position	# CR	# Non-coding	# coding**
H:1	7.92%	31	14	1	10	2	0	1
H:2	2.05%	25	16	2	10	3	0	1
H:3	1.03%	11	7	0	5	1	0	1
H:4	0.54%	8	1	0	1	0	0	0
H:5	1.03%	12	5	1	1	2	1	0
H:6	0.91%	11	5	1	1	2	0	1
H:7	0.66%	7	0	0	0	0	0	0
J:1	1.21%	15	8	0	4	4	0	0
J:2	0.66%	8	5	0	4	1	0	0
J:3	0.79%	13	4	0	4	0	0	0
J:4	0.36%	8	2	0	1	1	0	0
T:1	1.33%	21	7	1	6	0	0	0
T:2	0.60%	10	1	0	1	0	0	0
T:3	0.66%	8	2	0	2	0	0	0
V:1	1.21%	25	11	1	6	4	0	0
K:1	1.09%	14	7	0	5	2	0	0
K:2	0.60%	7	2	0	2	0	0	0
K:3	0.60%	7	3	0	2	0	1	0
	23.25%	241	59*	5*	44*	8*	1*	1*

Table 9. Caucasian Whole Genome Sequencing Results

A listing of the HV-types sequenced, the prevalence of each type in the population, the number of samples sequenced for each HV-type, the number of discriminatory, non-redundant SNPs for each HV-type, and the genome location for each of the sites.

* Totals listed for these columns may not equal the total of all values in the column. This is due to the fact that some HV-types share the same discriminatory, non-synonymous SNPs (for example, 16519). The total values reflect the total number of discriminatory, non-synonymous sites for the population.

** One site, 3010, is located in the 16S rRNA gene.

HV-types	% of Population Matching HV-type	# of Samples Sequenced	# of discriminatory, non-redundant SNPs per HV-type	SNP Location			
				# 1st Codon Position	# 3rd Codon Position	# CR	# Non-coding
L0a:1	0.29%	4	1	0	1	0	0
L0a:2	0.64%	4	2	0	1	1	0
L1b:1	2.22%	14	4	0	4	0	0
L1b:2	1.07%	5	1	1	0	0	0
L2a:1	2.29%	10	5	0	5	0	0
L2a:2	0.72%	4	3	0	2	1	0
L2a:3	0.79%	4	3	0	1	1	1
L2a:4	0.93%	10	10	0	8	2	0
L2a:5	1.15%	16	8	0	6	2	0
L2a:6	1.15%	10	4	0	3	1	0
L2b:1	0.93%	8	4	0	3	1	0
L2c:1	0.50%	4	2	0	2	0	0
L3b:1	0.43%	3	0	0	0	0	0
L3b:2	1.00%	5	4	0	3	1	0
L3e:1	1.22%	12	4	0	4	0	0
L3e:2	1.22%	8	3	0	3	0	0
L3e:3	0.72%	3	0	0	0	0	0
L3e:4	0.43%	3	1	0	0	0	1
L3e:6	1.00%	8	3	0	2	1	0
L3f:1	0.72%	5	0	0	0	0	0
20	**19.42%**	**140**	**42***	**1***	**38***	**2***	**1***

Table 10. African-American Whole Genome Sequencing Results

A listing of the HV-types sequenced, the prevalence of each type in the population, the number of samples sequenced for each HV-type, the number of discriminatory, non-redundant SNPs for each HV-type, and the genome location for each of the sites.

* Totals listed for these columns may not equal the total of all values in the column. This is due to the fact that some HV-types share the same discriminatory, non-synonymous SNPs (for example, 16519). The total values reflect the total number of discriminatory, non-synonymous sites for the population.

HV-types	% of Population Matching HV-type	# of Samples Sequenced	# of discriminatory, non-redundant SNPs per HV-type	SNP Location			
				# 1st Codon Position	# 3rd Codon Position	# CR	# Non-coding
A2:1	3.57%	15	7	0	4	3	0
A2:2	1.35%	7	4	0	3	1	0
A2:3	1.35%	15	5	2	2	1	0
A2:4	1.11%	4	0	0	0	0	0
A2:5	0.56%	1	***				
A2:6	0.56%	5	2	0	1	1	0
B2:1	1.11%	7	2	0	2	0	0
B2:2	0.63%	6	2	0	1	1	0
C1:1	2.62%	12	0	0	0	0	0
C1:2	1.74%	13	5	0	3	2	0
C1:3	0.63%	3	0	0	0	0	0
C1:4	0.56%	2	0	0	0	0	0
D1:1	1.98%	11	5	1	3	1	0
U5:1	0.56%	4	0	0	0	0	0
K:4	2.06%	12	0	0	0	0	0
J:5	0.79%	5	1	0	1	0	0
L1b:3	0.63%	2	0	0	0	0	0
L3e:5	0.56%	1	***				
18	22.37%	125**	22*	2*	16*	4*	0*

H:1	1.03%	Cauc. common type, panels already designed.
19	23.40%	156

Table 11. Hispanic Whole Genome Sequencing Results

A listing of the HV-types sequenced, the prevalence of each type in the population, the number of samples sequenced for each HV-type, the number of discriminatory, non-redundant SNPs for each HV-type, and the genome location for each of the sites. The HV-type H:1 is common among Hispanics, however 31 H:1 samples had already been sequenced to identify discriminatory, non-redundant SNPs for the Caucasian common types. As such, no additional H:1 samples were sequenced.

* Totals listed for these columns may not equal the total of all values in the column. This is due to the fact that some HV-types share the same discriminatory, non-synonymous SNPs (for example, 16519). The total values reflect the total number of discriminatory, non-synonymous sites for the population.

** Number represents the number of sequenced samples used to establish the total number of sites identified, not the total number of samples sequenced.

*** Only one sample located for sequencing, no discriminatory, non-redundant sites were identified

It is of interest to compare the amount of discrimination potential contained within the entire mtGenome for common HV types within the three groups. As noted above, fewer useful SNPs (according to our criteria) were discovered for Hispanics than in the other groups. This is mirrored in the overall amount of variation present within the common HV types (Tables 12-14). In Hispanics, there are two sets of 9 individuals that match exactly over the entire mtGenome, compared to African Americans with two sets of three individuals, and Caucasians with four sets of four individuals. Comparing the African American and Hispanic datasets, with roughly equal frequency distributions of common HV types, and similar sample sizes for each common HV type, we see 114/140 (81.4%) of the African American mtGenomes as unique, but only 72/125 (57.6%) of the Hispanics mtGenomes resolved to unique types. The factors that affect frequency mtDNA distributions within populations, especially in the case of recent continental immigration, are complex. Nonetheless it may be reasonable to suppose that the extreme founder effect associated with Native American population expansion in the relatively recent past (~12,000-14,000 years ago) may account for the lack of variation in some of the common Hispanic HV types (that are predominantly Native American in origin).

As discussed, among our basic selection criteria for target SNPs is that they do not have potential for phenotypic effect, so we have excluded non-synonymous variants. It is useful, however, to evaluate this criterion in relation to the cost it carries with regard to lost discrimination potential. Tables 12-14 show the amount of discrimination obtained by the use of the synonymous SNP panels only, compared to the discrimination obtained when non-synonymous substitutions are included as well. In both cases, we have excluded private polymorphisms that vary in only a single individual (which make no sense as targets for specific SNP assays), and the effect of that restriction can be seen by comparing to the discrimination

observed over the entire mtGenome (third panel of Tables 12-14). In Caucasians, the synonymous SNP panels generate ~85% of the total discrimination that could be obtained from the synonymous and non-synonymous variation combined (again, excluding private polymorphisms). Similarly, the African-American and Hispanic multiplex panels generate, respectively, ~84% and ~95% of the discrimination that could be obtained if non-synonymous sites were included. Tables 15-17 list the small number of non-private, non-synonymous sites that provide additional discrimination with respect to our selected multiplex panels, for each of the population groups.

8 Multiplexes		8 Multiplexes + Non-Synonymous SNPs (excluding private polymorphisms)		Across Entire Genome	
# of types	# individuals/type	# of types	# individuals/type	# of types	# individuals/type
2	14	1	13	4	4
1	9	1	12	5	3
3	8	1	8	10	2
2	7	2	7	190	1
1	6	3	6		
3	5	6	4		
3	4	9	3		
8	3	23	2		
27	2	79	1		
55	1				
TOTAL: 106		TOTAL: 125		TOTAL: 209	

Table 12. Caucasian Discrimination
Haplotype breakdown for the Caucasian dataset: 1) by the application of the eight multiplex panels, 2) by the application of the eight multiplex panels plus non-private, non-synonymous SNPs, and 3) across the entire genome. Includes discrimination provided by HV sequences.

6 Multiplexes		8 Multiplexes + Non-Synonymous SNPs (excluding private polymorphisms)		Across Entire Genome	
# of types	# individuals/type	# of types	# individuals/type	# of types	# individuals/type
2	7	3	5	2	3
1	6	1	4	10	2
1	5	6	3	114	1
4	4	23	2		
7	3	56	1		
18	2				
42	1				
TOTAL: 75		TOTAL: 89		TOTAL: 126	

Table 13. African-American Discrimination
Haplotype breakdown for the African American dataset: 1) by the application of the six multiplex panels, 2) by the application of the six multiplex panels plus non-private, non-synonymous SNPs, and 3) across the entire genome. Includes discrimination provided by HV sequences.

3 Multiplexes		8 Multiplexes + Non-Synonymous SNPs (excluding private polymorphisms)		Across Entire Genome	
# of types	# individuals/type	# of types	# individuals/type	# of types	# individuals/type
2	12	1	12	2	9
1	7	1	11	1	5
1	6	1	7	1	4
6	4	1	6	2	3
5	3	7	4	10	2
11	2	3	3	72	1
27	1	10	2		
		32	1		
TOTAL: 53		TOTAL: 56		TOTAL: 88	

Table 14. Hispanic Discrimination
Haplotype breakdown for the Hispanic dataset: 1) by the application of the three multiplex panels, 2) by the application of the three multiplex panels plus non-private, non-synonymous SNPs, and 3) across the entire genome. Includes discrimination provided by HV sequences.

Site:	Variable in Types:
1719	H:1; T:2
1811	J:1
2772	J:1; H:2
4025	j:1
4639	H:7; V:1
5460	H:1
8433	T:1; H:6
8803	H:5
8839	J:1
8869	V:1
15323	H:1; H:3
15773	V:1, T:1
15924	T:2; H:1

Table 15. Caucasian Discriminatory Non-Synonymous Sites
A listing of non-private, non-synonymous sites in US Caucasians that provide additional discrimination when combined with the selected multiplex panels and HV sequencing.

Site:	Variable in Types:
731	L2a:5
1719	L1b:1, L3e:6
3866	L3b:2
5563	L2c:1
6261	L3e:2
9804	L1b:2
10454	L2a:6
12172	L0a:2
12906	L1b:1
13790	L3b:1
14766	L2a:5
15431	L3e:1
15617	L2a:6, L3e:2

Table 16. African-American Discriminatory Non-Synonymous Sites
A listing of non-private, non-synonymous sites in African Americans that provide additional discrimination when combined with the selected multiplex panels and HV sequencing.

Site:	Variable in Types:
3202	A2:1
3316	B2:1; C1:2
7444	A2:2
7628	C1:1; C1:2
6261	L3e:2
9098	B2:1
12223	B2:1; D1:1
14178	B2:1
15617	A2:1

Table 17. Hispanic Discriminatory Non-Synonymous Sites
A listing of non-private, non-synonymous sites in Hispanics that provide additional discrimination when combined with the selected multiplex panels and HV sequencing.

Optimization and Testing of Caucasian Multiplex SNP Panels using the SNaPshot Platform

The choice of a SNP platform must take into consideration the difficulties encountered when working with mtDNA. mtDNA analysis is typically performed on highly degraded samples, and testing is often complicated by limited extract quantities. These factors necessitate an assay that uses short amplicons, has extremely high sensitivity, and can be easily multiplexed. An additional important assay feature is the ability to detect low-level mixtures and heteroplasmy. A detection platform common to laboratories performing fluorescent fragment analysis for short tandem repeats (STRs) or mtDNA analysis is also desirable.

We chose to focus on an allele-specific primer extension assay (ASPE), SNaPshot™ (Applied Biosystems), as the platform for the multiplex SNP assays. The SNaPshot protocol uses fluorescently-labeled dideoxy terminator incorporation at the site of interest followed by detection on a capillary electrophoresis instrument (such as the Applied Biosystems 3100). Multiplexing is achieved by the varying the length of the extension primers (designed by the

scientist) which provides spatial separation in detection. The SNaPshot assay uses small amplicons, can be multiplexed for ten or more SNPs, has high sensitivity, and good detection of mixture and heteroplasmy. Initial work on the ASPE assay was done in collaboration with Peter Vallone (NIST), who performed the amplification and extension primer design, pristine DNA optimization, and sensitivity testing for multiplex panel A (Vallone et al. [2004]).

Figure 9 provides an overview of the standard ASPE assay protocol from amplification through detection, and a detailed description of our protocol is described in Vallone et al. (2004). A 31-cycle, reverse-touchdown multiplex amplification is performed in a total volume of 15µl. PCR reaction cleanup to remove unincorporated dNTPs and PCR primers is accomplished by the addition of Exonuclease I and Shrimp Alkaline Phosphatase (SAP). The 25-cycle multiplex SNaPshot primer extension reaction combines the commercially available SNaPshot Ready Reaction Mix™ (Applied Biosystems) with PCR products and extension primers in a 10µl reaction. A final clean-up is accomplished by a SAP incubation. Fluorescent detection is performed on an Applied Biosystems 3100 capillary electrophoresis sequence detection system and the data analyzed in GeneScan v3.7 and Genotyper v3.7 or GeneMapperID v3.1(all Applied Biosystems). Appendix III describes all current thermal cycling conditions and 3100 injection parameters.

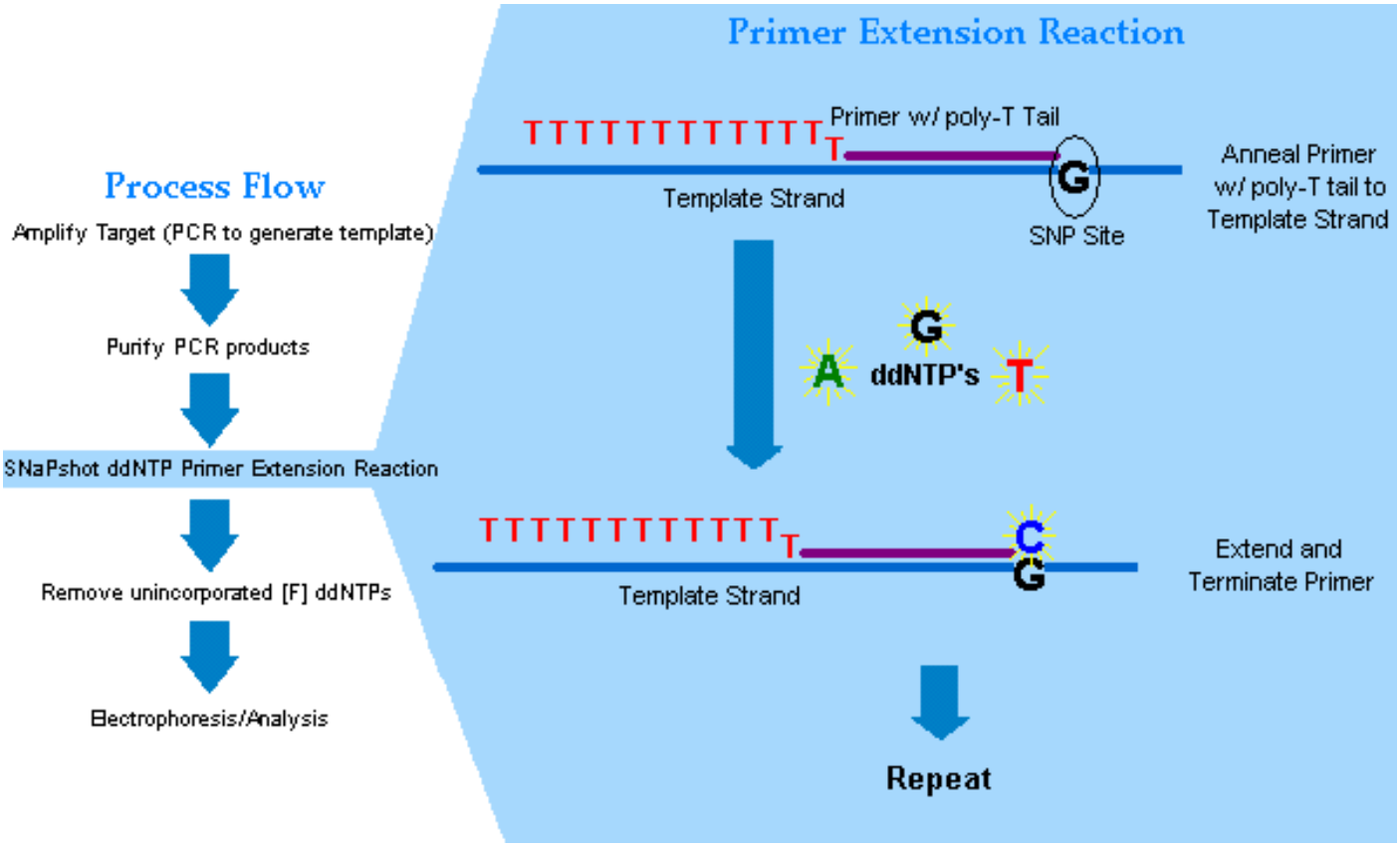


Figure 9. Overview of allele-specific primer extension assay using the SNaPshot platform. The five steps of the assay, including amplification, PCR reaction clean-up, extension, extension reaction cleanup, and detection/analysis. The extension reaction is shown here in more detail. An extension primer anneals to the target strand immediately 3' to the site of interest. A single, fluorescently-labeled dideoxy terminator is incorporated at the SNP site and the extension is terminated. The result is an oligo of a specific size (based on the length of the poly-T tail), the fluorescent label for which identifies the nucleotide at the site of interest.

The development and optimization of working assays for the multiplex panels was accomplished through a step-wise approach. The first step consisted of amplification and extension primer design, and was followed first by singleplex and then multiplex testing of the amplification and extension primers. Once an optimized assay had been developed using pristine positive controls, sensitivity testing and then tests on mixtures were conducted. Finally, the

assays were tested on degraded samples, and the assay conditions optimized to create a new, degraded samples protocol.

Step 1: Amplification and Extension Primer Design and Purchase

Amplification primer design was a critical component of the SNP assay design and optimization. The requirement that the SNP panels be functional for highly degraded samples dictated a small amplicon size, preferably 150 base pairs or smaller. Eliminating approximately 30bp upstream and downstream of the SNP site for later extension primer selection, about 50bp of template sequence further upstream and downstream were entered into the web-based program Primer3 (Rozen and Skaletsky) for forward and reverse amplification primer selection. This primer selection strategy is depicted in Figure 10. The default Primer3 settings (Table 18) were used to select the best candidate amplification primers. These primers were then compared to our whole genome datasets to select primers for which the binding sites would avoid highly polymorphic sites in the target populations. If, based on desired amplicon size, no suitable primers could be identified that would avoid highly polymorphic positions, degenerate primers were designed.

Extension primers were also selected using the Primer3 program with the default settings (Table 18). The nature of the SNaPshot™ single-base extension reaction restricts the extension primer to one of two positions: the sequence immediately 5' to the SNP on either the forward or the reverse strand. For each SNP, approximately 30 bases upstream of the site on both the forward and reverse strands were entered into Primer3 for consideration (Figure 10). While many of the extension primers selected were excellent candidate primers, the position restriction on occasion forced a choice between two undesirable primers.

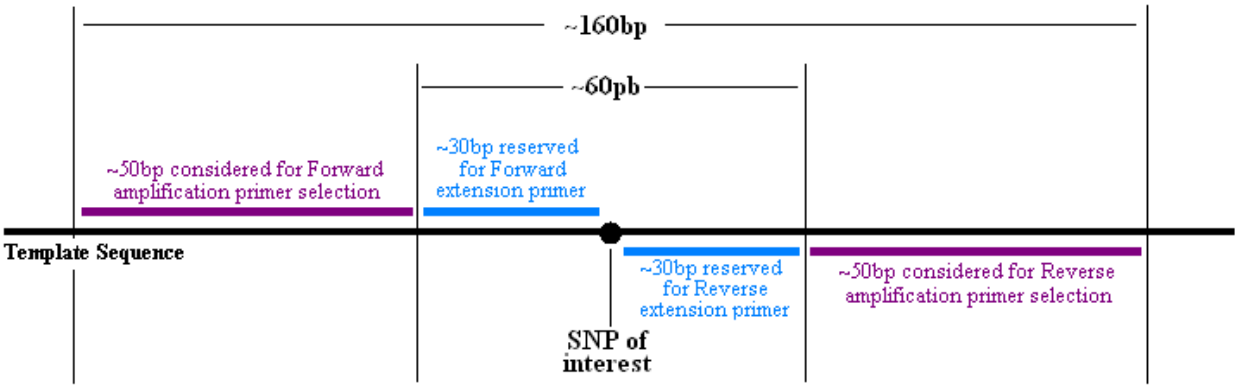


Figure 10. Amplification and Extension Primer Selection Strategy
Extension primers were selected from the first ~30bp upstream or downstream from each SNP site, and amplification primers were selected from the next ~50bp upstream and downstream. Amplicons range in size but most are 150bp or smaller.

Max 3' Stability	9		Max Primer TM	63
Max Mispriming	12		Min Primer GC%	20
Min Primer Size	18		Max Primer GC %	80
Opt Primer Size	20		Max Self-Complimentarity	8
Max Primer Size	27		Max 3' Self-Complimentarity	3
Min Primer TM	57		Max Poly-X	5
Opt Primer TM	60			

Table 18. Primer3 Default Settings
The default Primer3 settings were used for primer selection.

One of the complexities to multiplexing is the potential for significant primer-primer interactions. Once all amplification and extension primers were designed, the sequences for each multiplex panel were run through NIST's AutoDimer program (Vallone and Butler, 2004) to check for such potential interactions. All sequences were compared using the default program settings. On occasion, primer pairs were identified that would be likely to form primer-dimers, and these primers were subsequently redesigned. Appendix IV lists the final amplification and primer sequences for all 8 Caucasian SNP multiplexes.

Issues with primer quality control were problematic throughout the project and resulted in the purchase of oligos from several different sources. Over the course of the 5-year project, primers were purchased from three different companies (Qiagen Operon, MWG Biotech, and Proligo) and with different purification options (none, HPLC, and PAGE) with varying degrees of success in terms of the primer quality control. Primers are currently purchased from Proligo in either unpurified 10 O.D. batches (amplification primers) or PAGE-purified 1 O.D. batches (extension primers).

One of the significant primer quality control issues encountered was variation, sometimes by as much as 40%, in the concentration of the primer stocks. Similar variation was observed whether oligos were received lyophilized or in solution. To correct for this variation, oligos were ordered lyophilized and subsequently resuspended in-house to ~200µM based on the concentration indicated by the manufacturer. Actual concentrations were determined via UV absorbance read at 260nm on a Tecan™ Genios™ plate reader, using the molar extinction coefficient for each oligo. The concentrations were then adjusted to100µM for storage at -20˚C.

Step 2: Primer Testing

The second step in the multiplex assay development was the testing of the newly designed amplification and extension primers. Amplification primers were tested in singleplex at 1µM final reaction concentrations, and visualized on 1% agarose gels. Extension primers were tested in singleplex at 0.5µM final reaction concentrations. Singleplex extension products were detected on an Applied Biosystems 3100 Genetic Analyzer, and the resulting electropherograms analyzed in Genescan™ v3.7 and GenoTyper™ v.3.7, or GeneMapperID™ v3.1 (all Applied Biosystems). The electropherograms were reviewed to detect any potential problematic results, to include: 1) the presence of any peaks other than the expected peak, 2) the presence of

significant N-X or N+X peaks, indicative of poor extension primer quality, 3) unexpected migration of the expected peak that might cause overlap and incorrect allele assignment with a peak for another SNP in the multiplex, 4) uneven signal intensities between variants for each SNP, and 5) low signal intensity. Primer issues identified in singleplex testing resulted in some primer re-order and/or redesign.

Following successful testing of all primers in singleplex, multiplex amplification with singleplex extension reactions was attempted. For each multiplex panel, the amplification primers were initially tested at equimolar concentrations. For example, Multiplex A has 22 amplification primers (11 each forward and reverse primers), and each primer was at a concentration of 4.55 μM in the multiplex amplification primer mix (0.91 μM final reaction concentration). The resulting electropherograms were examined for, in particular, strong differences in extension product signal intensity between the multiplex amplification results compared with the previous singleplex amplification results. While some decrease in signal intensity might expected due to the competitive nature of multiplex PCR, any unusual changes in signal intensity were investigated. Occasionally, problems with signal intensity resulted in primer re-order and/or re-design.

The final step in primer testing was multiplex amplification combined with multiplex extension. For the initial testing, all extension primers were combined into a mix with a 0.5μM final concentration for each primer. The resulting electropherograms were examined for two critical potential problems: 1) relative signal strength, and 2) extraneous extension peaks. 1) Relative signal strength: Though amplification and extension primers were carefully designed, there was often variation in relative signal strength between sites as a result of varied primer binding efficiency in both the amplification and extension reactions. Attempts were made to

balance the signal strength for each site in a multiplex panel by adjusting the concentration of the extension primer in the multiplex primer mix. When sufficient balancing could not be obtained in this manner, forward, reverse, or both amplification primer concentrations were adjusted in the multiplex primer mix. When signal strength still failed to achieve the minimum balance desired, amplification and/or extension primers were re-ordered or redesigned. 2) Extraneous extension peaks: When peaks were observed on the electropherogram that did not appear to correspond to a particular SNP, experiments were performed to determine the cause. Frequently, through duplex testing of the amplification and/or extension primers, the extraneous peak was identified as resulting from a primer-primer interaction that had not been identified during the primer design phase. In these cases, primers were re-designed and re-tested to attempt to eliminate the extraneous peaks. In some cases however, the extraneous peaks appeared to be primer lot dependant. In such situations, the primers were either re-ordered, or if the extraneous peak would cause no error in allele calling, the extraneous peak was noted and recorded as a known artifact specific to a particular lot of primer.

Figures 11-17 show the balanced multiplex results for Caucasian panels A-C and E-H. Tables 19-25 list the sites in each multiplex, and the expected position and color for each peak. Appendix IV lists the final primer sequences, and the final primer concentrations, for multiplex panels A-C and E-H. Caucasian panel D is still undergoing some optimization and is not yet ready for further testing.

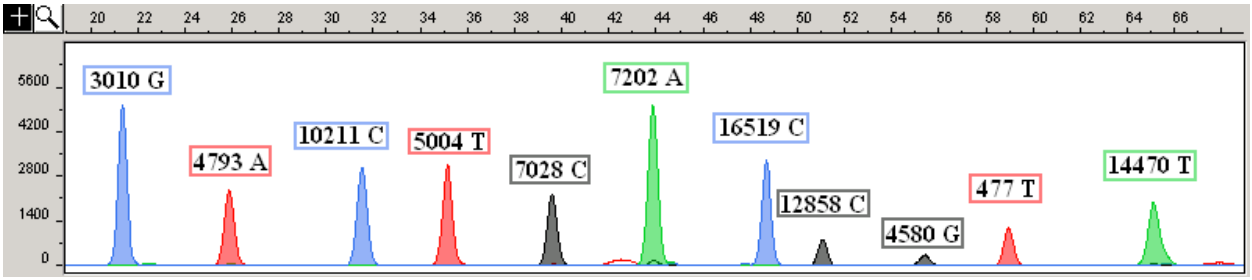


Figure 11. Multiplex Panel A
Plots of fragment size (x-axis) relative to –120LIZ size standard (Applied Systems) and relative fluorescent units (RFUs, y-axis). Results are typical for Multiplex A following extension primer concentration balancing (Appendix II lists the final primer concentrations).

Multiplex A				
Site	**Anderson**	**Variant**	**Variant 2**	**Size Range**
3010 F	Blue / G	Green / A		20.5-23
4793 R	Red / A	Black / G		24-26.5
10211 R	Blue / C	Green / T		31-33
5004 F	Red / T	Black / C		33.5-35.5
7028 F	Black / C	Red / T		39-41
7202 F	Green / A	Blue / G		42-44.5
16519 R	Green / T	Blue / C	Red / A	48-50
12858 F	Black / C	Red / T		50.5-52.5
4580 R	Black / G	Red / A		55-56.5
477 F	Red / T	Black / C		57.5-59.5
14470 R	Green / T	Red / A	Blue / C	63.5-66

Table 19. Multiplex Panel A
Lists the sites, the possible variants, their dye colors on the resulting electropherogram, and the approximate nucleotide size as it will appear following GeneScan™ analysis.

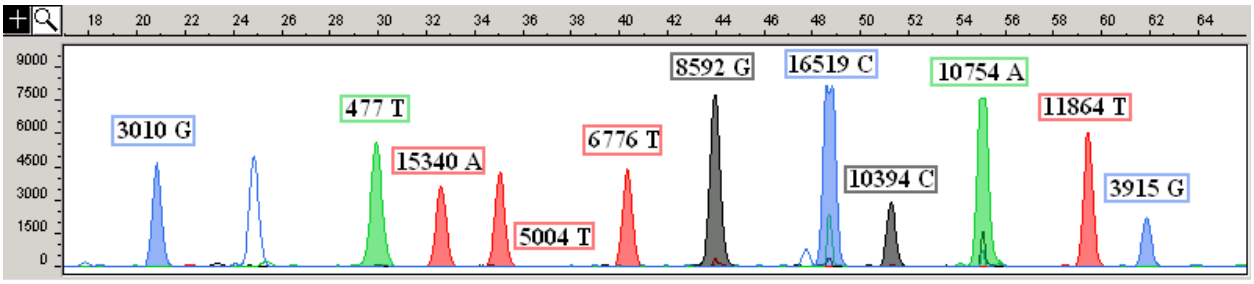


Figure 12. Multiplex Panel B
Plots of fragment size (x-axis) relative to –120LIZ size standard (Applied Systems) and relative fluorescent units (RFUs, y-axis). Results are typical for Multiplex B following extension primer concentration balancing (Appendix II lists the final primer concentrations).

Multiplex B				
Site	**Anderson**	**Variant**	**Variant 2**	**Size Range**
3010 F	Blue / G	Green / A		20-23
477 R	Green / T	Blue / C		28-31
15340 R	Red / A	Black / G		31-33
5004 F	Red / T	Black / C		33.5-35.5
6776 F	Red / T	Black / C		39-41
8592 R	Black / G	Red / A		43.5-45.5
16519 R	Green / T	Blue / C	Red / A	48.5-50
10394 F	Black / C	Red / T		51-52.5
10754 F	Green / A	Black / C		54-55.5
11864 F	Red / T	Black / C		58-60
3915 F	Blue / G	Green / A		61.5-64

Table 20. Multiplex Panel B
Lists the sites, the possible variants, their dye colors on the resulting electropherogram, and the approximate nucleotide size as it will appear following GeneScan™ analysis.

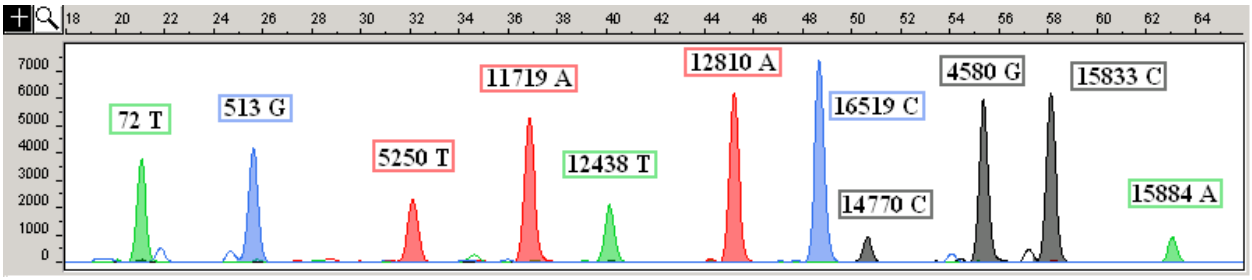


Figure 13. Multiplex Panel C
Plots of fragment size (x-axis) relative to –120LIZ size standard (Applied Systems) and relative fluorescent units (RFUs, y-axis). Results are typical for Multiplex C following extension primer concentration balancing (Appendix II lists the final primer concentrations).

Multiplex C			
Site	**Anderson**	**Variant**	**Position**
72 R	Green / T	Blue / C	20 / 22
513 F	Blue / G	Green / A	25 / 27
5250 F	Red / T	Black / C	31 / 32
11719 R	Black / G	Red / A	36 / 37
12438 R	Green / T	Blue / C	39 / 40
12810 R	Red / A	Black / G	44 / 45
16519 R	Green / T	Blue / C	48 / 49
14770 F	Black / C	Red / T	50 / 51
4580 R	Black / G	Red / A	55 / 56
15833 F	Black / C	Red / T	58 / 59
15884 F	Blue / G	Green / A	62 / 63

Table 21. Multiplex Panel C
Lists the sites, the possible variants, their dye colors on the resulting electropherogram, and the approximate nucleotide size as it will appear following GeneScan™ analysis.

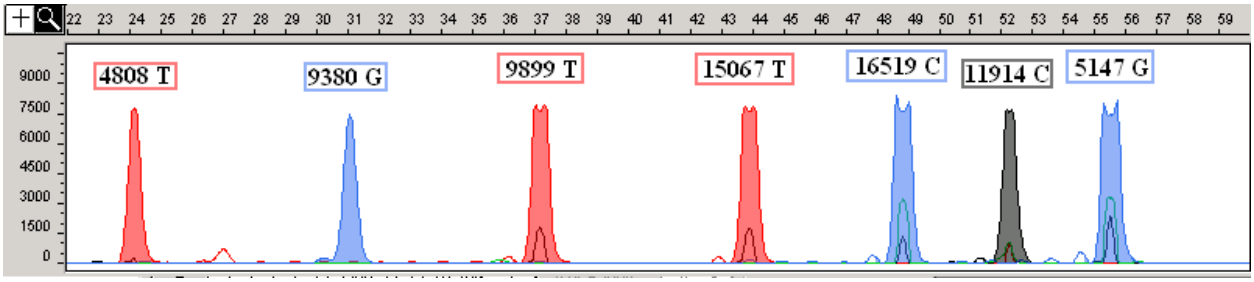


Figure 14. Multiplex Panel E
Plots of fragment size (x-axis) relative to –120LIZ size standard (Applied Systems) and relative fluorescent units (RFUs, y-axis). Results are typical for Multiplex E following extension primer concentration balancing (Appendix II lists the final primer concentrations).

Multiplex E			
Site	**Anderson**	**Variant**	**Position**
4808 F	Black / C	Red / T	23 / 24
5147 F	Blue / G	Green / A	55 / 56
9380 F	Blue / G	Green / A	31 / 32
9899 F	Red / T	Black / C	36 / 37
11914 R	Black / C	Red / T	52 / 53
15067 F	Red / T	Black / C	43 / 44
16519 R	Green / T	Blue / C	49 / 50

Table 22. Multiplex Panel E
Lists the sites, the possible variants, their dye colors on the resulting electropherogram, and the approximate nucleotide size as it will appear following GeneScan™ analysis.

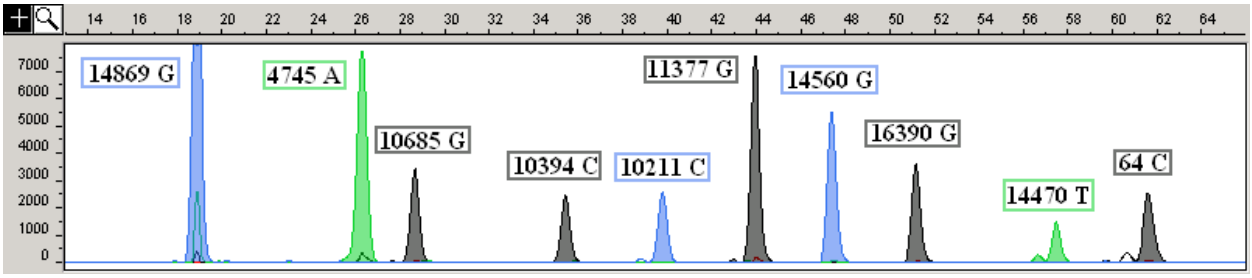


Figure 15. Multiplex Panel F
Plots of fragment size (x-axis) relative to –120LIZ size standard (Applied Systems) and relative fluorescent units (RFUs, y-axis). Results are typical for Multiplex F following extension primer concentration balancing (Appendix II lists the final primer concentrations).

Multiplex F				
Site	**Anderson**	**Variant 1**	**Variant 2**	**Position**
14869 F	Blue / G	Green / A	Black / C	19 / 20 / 20
4745 F	Green / A	Blue / G		25 / 26
10685 R	Black / G	Red / A		29 / 30
10394 F	Black / C	Red / T		35 / 36
10211 R	Blue / C	Green / T		40 / 41
11377 R	Black / G	Red / A		44 / 45
14560 F	Blue / G	Green / A		47 / 48
16390 R	Black / G	Red / A		51 / 52
14470 R	Green / T	Blue / C	Red / A	56 / 57 / 57
64 F	Black / C	Red / T		61 / 62

Table 23. Multiplex Panel F
Lists the sites, the possible variants, their dye colors on the resulting electropherogram, and the approximate nucleotide size as it will appear following GeneScan™ analysis.

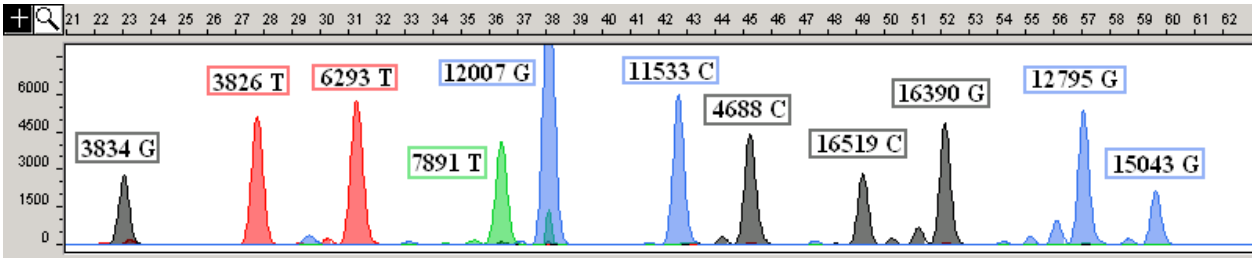


Figure 16. Multiplex Panel G
Plots of fragment size (x-axis) relative to –120LIZ size standard (Applied Systems) and relative fluorescent units (RFUs, y-axis). Results are typical for Multiplex G following extension primer concentration balancing (Appendix II lists the final primer concentrations).

Multiplex G			
Site	**Anderson**	**Variant**	**Position**
3834 R	Black / G	Red / A	23 / 25
3826 F	Red / T	Black / C	27 / 28
6293 F	Red / T	Black / C	30 / 31
7891 R	Blue / C	Green / T	34 / 35
12007 F	Blue / G	Green / A	38 / 39
11533 R	Blue / C	Green / T	43 / 44
4688 F	Red / T	Black / C	46 / 47
16519 F	Red / T	Black / C	49 / 50
16390 R	Black / G	Red / A	52 / 51
12795 F	Blue / G	Green / A	57 / 58
15043 F	Blue / G	Green / A	60 / 61

Table 24. Multiplex Panel G
Lists the sites, the possible variants, their dye colors on the resulting electropherogram, and the approximate nucleotide size as it will appear following GeneScan™ analysis.

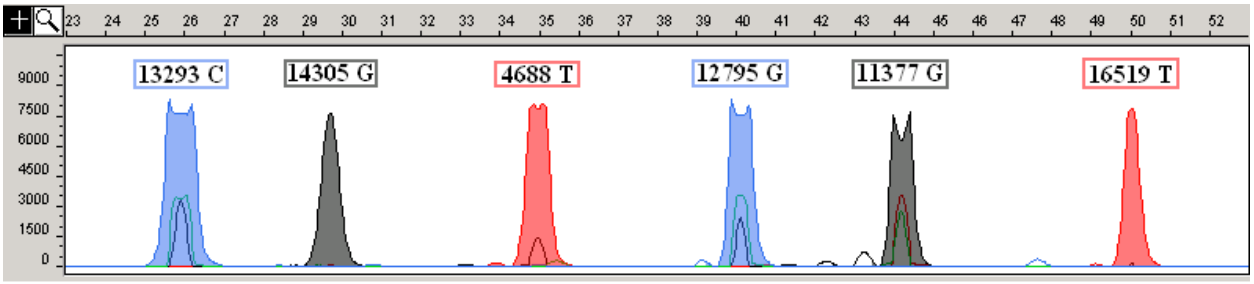


Figure 17. Multiplex Panel H
Plots of fragment size (x-axis) relative to –120LIZ size standard (Applied Systems) and relative fluorescent units (RFUs, y-axis). Results are typical for Multiplex H following extension primer concentration balancing (Appendix II lists the final primer concentrations).

Multiplex H			
Site	**Anderson**	**Variant**	**Position**
13293	Blue / C	Green / T	25.5 / 27
14305	Black / G	Red / A	29 / 30
4688	Red / T	Black / C	34.5 / 33
12795	Blue / G	Green / A	40 / 41
11377	Black / G	Red / A	44 / 45
16519	Red / T	Black / C	50 / 49

Table 25. Multiplex Panel H
Lists the sites, the possible variants, their dye colors on the resulting electropherogram, and the approximate nucleotide size as it will appear following GeneScan™ analysis.

Step 3: Sensitivity Testing

Once each multiplex was optimized, each panel was tested in regards to its sensitivity. Because these SNP assays are designed specifically for use on highly degraded samples, it is especially important that they have high sensitivity, preferably achieving full results with less

than 10pg of genomic DNA. A dilution series from at least two different positive controls was run for each Caucasian multiplex panel. The dilution series concentrations ranged from 500pg to 0.1pg. Figures 18-24 show the sensitivity results for Caucasian multiplex panels A-C and E-H. The figure for each multiplex shows typing results typical of 50pg genomic DNA, 10pg genomic DNA, and then the lowest concentration at which full typing was reproducibly obtained for that multiplex. The threshold for calling a peak was set at 100 relative fluorescent units (rfu). Full results were consistently obtained to 3pg genomic DNA or less, with some multiplexes sensitive to less than 1pg.

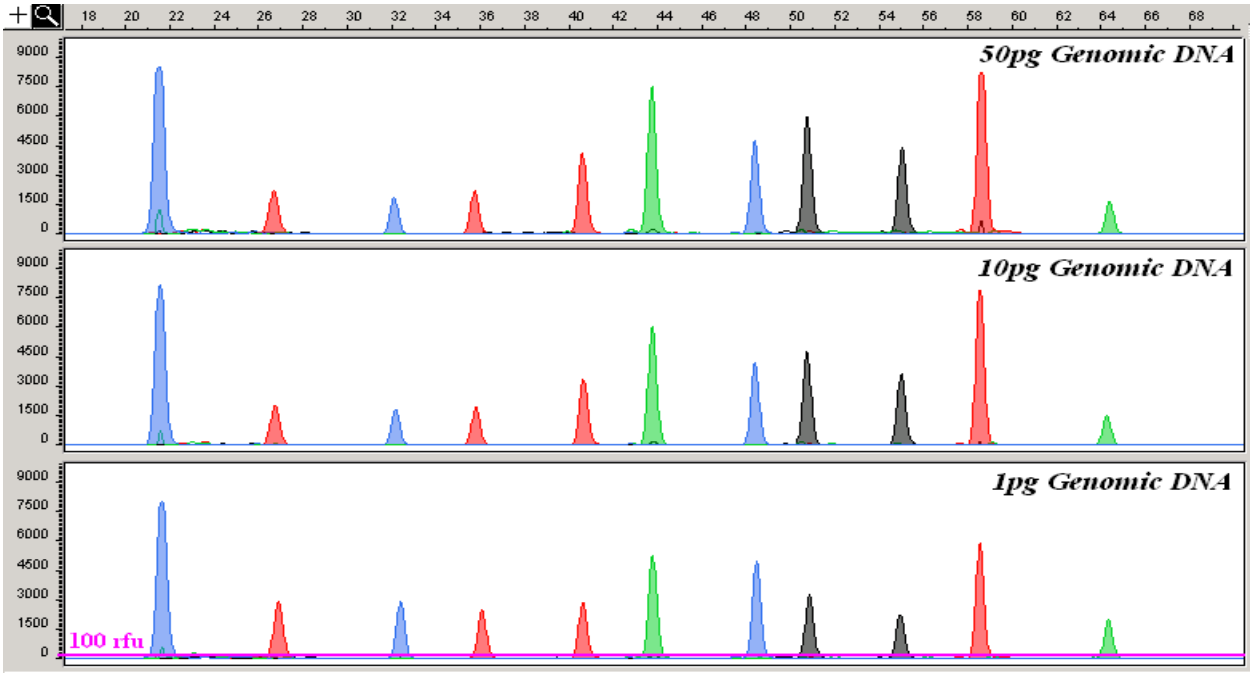


Figure 18. Sensitivity for Multiplex A.
Plots nucleotide size (x-axis) relative to –120LIZ size standard (Applied Systems) and relative fluorescent units (RFUs, y-axis). All 11 multiplexed sites were typed to 1pg.

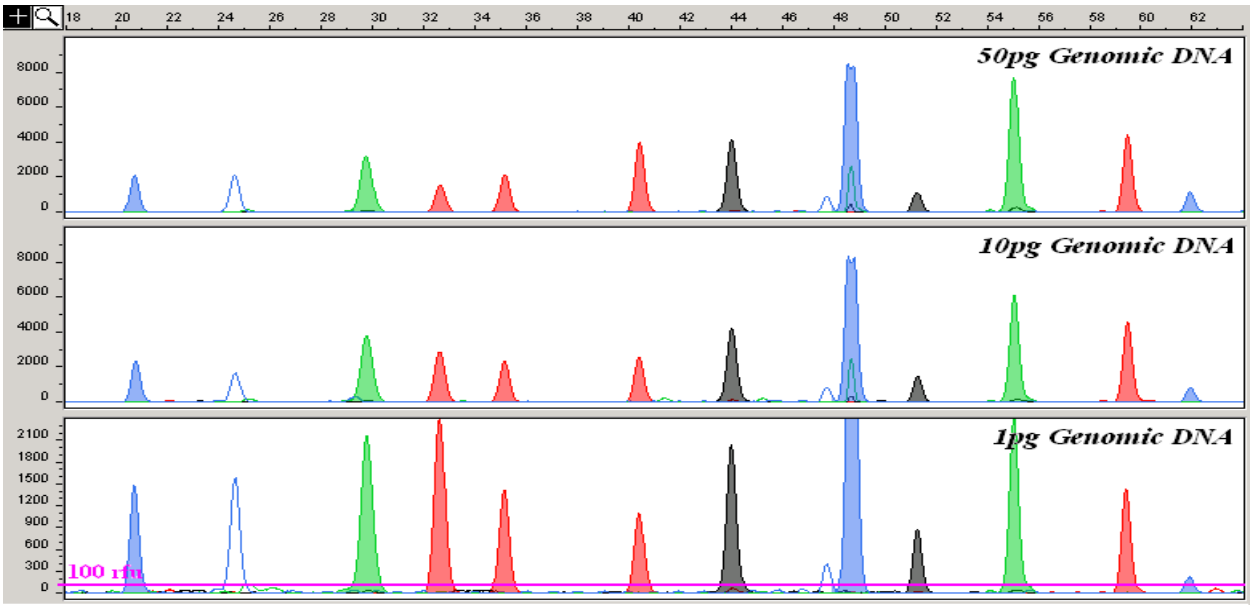


Figure 19. Sensitivity for Multiplex B.
Plots nucleotide size (x-axis) relative to –120LIZ size standard (Applied Systems) and relative fluorescent units (RFUs, y-axis). All 11 multiplexed sites were typed to 1pg.

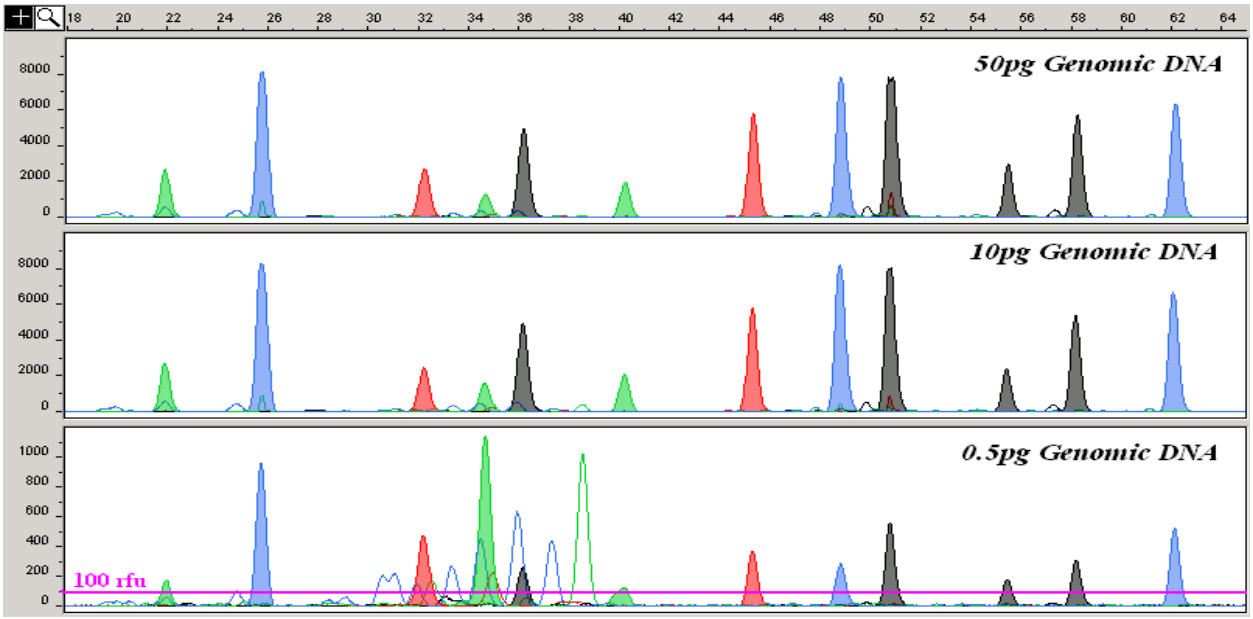


Figure 20. Sensitivity for Multiplex C.
Plots nucleotide size (x-axis) relative to –120LIZ size standard (Applied Systems) and relative fluorescent units (RFUs, y-axis). All 11 multiplexed sites were typed to 0.5pg.

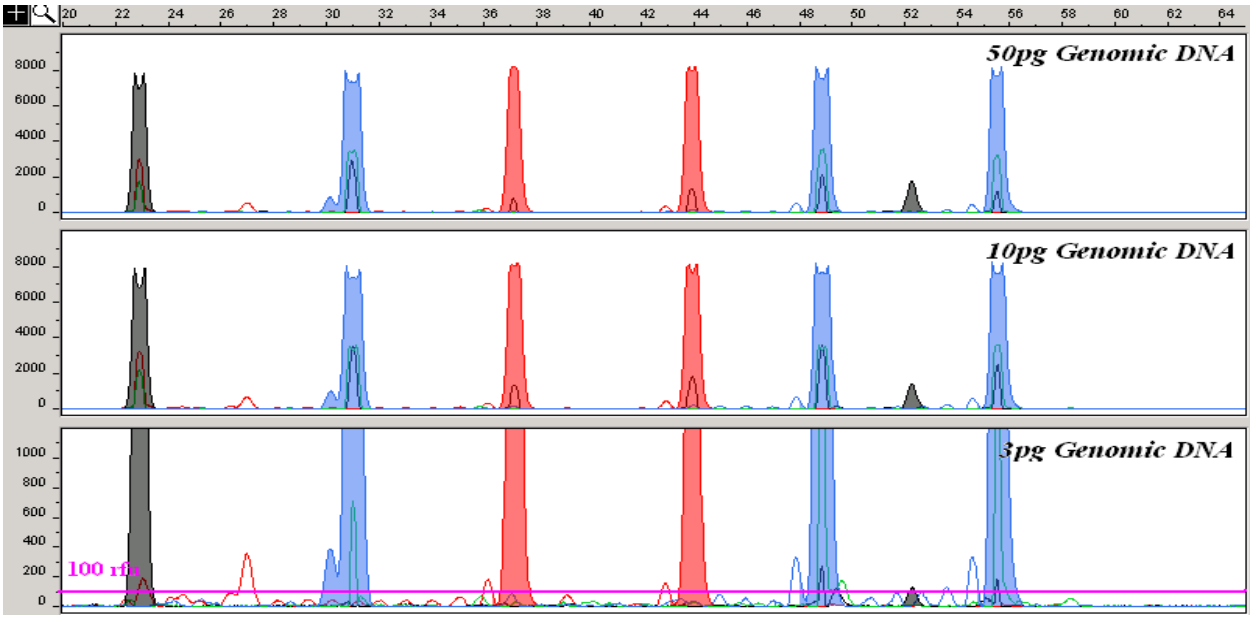


Figure 21. Sensitivity for Multiplex E.
Plots nucleotide size (x-axis) relative to –120LIZ size standard (Applied Systems) and relative fluorescent units (RFUs, y-axis). All 7 multiplexed sites were typed to 3pg.

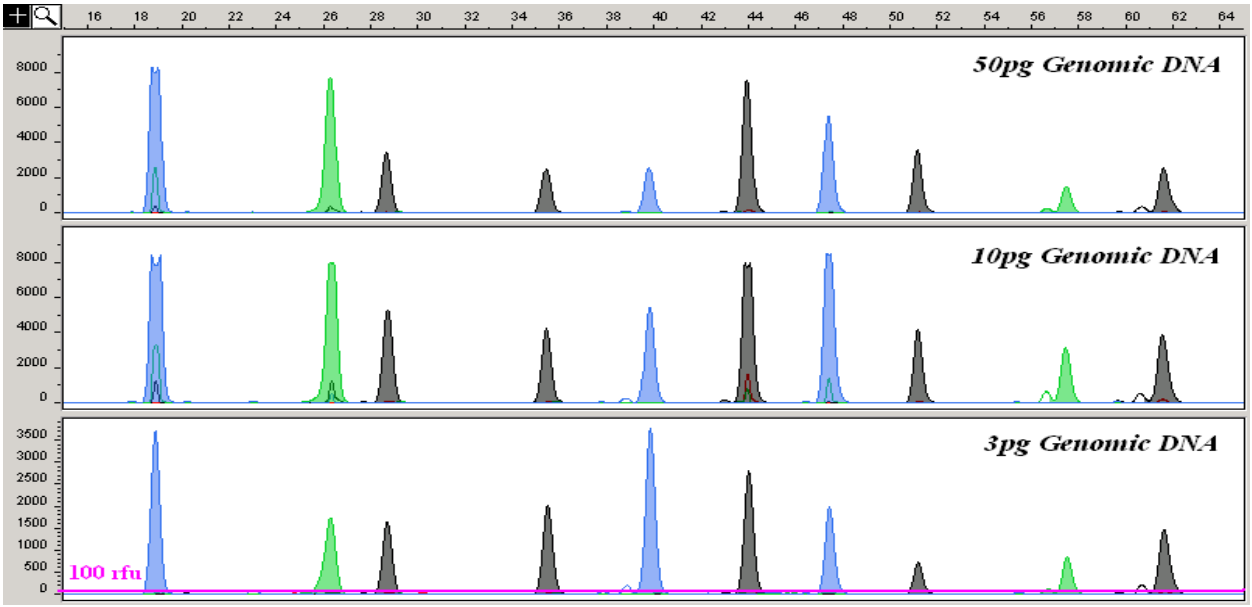


Figure 22. Sensitivity for Multiplex F.
Plots nucleotide size (x-axis) relative to –120LIZ size standard (Applied Systems) and relative fluorescent units (RFUs, y-axis). All 10 multiplexed sites were typed to 3pg.

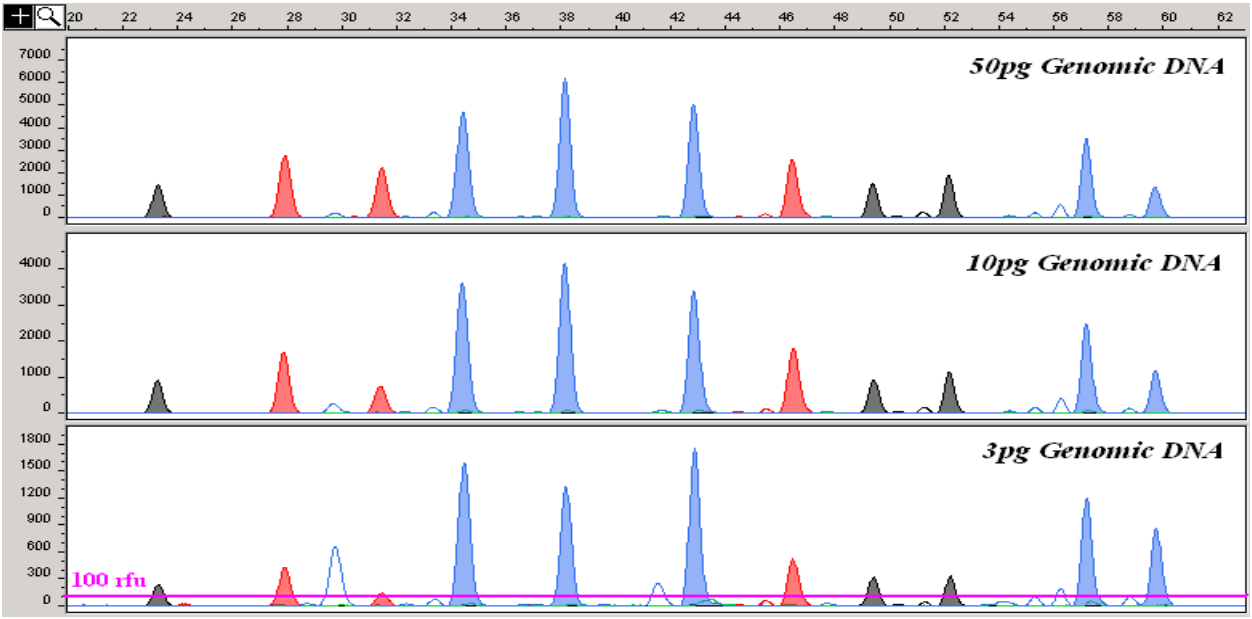


Figure 23. Sensitivity for Multiplex G.
Plots nucleotide size (x-axis) relative to –120LIZ size standard (Applied Systems) and relative fluorescent units (RFUs, y-axis). All 11 multiplexed sites were typed to 3pg.

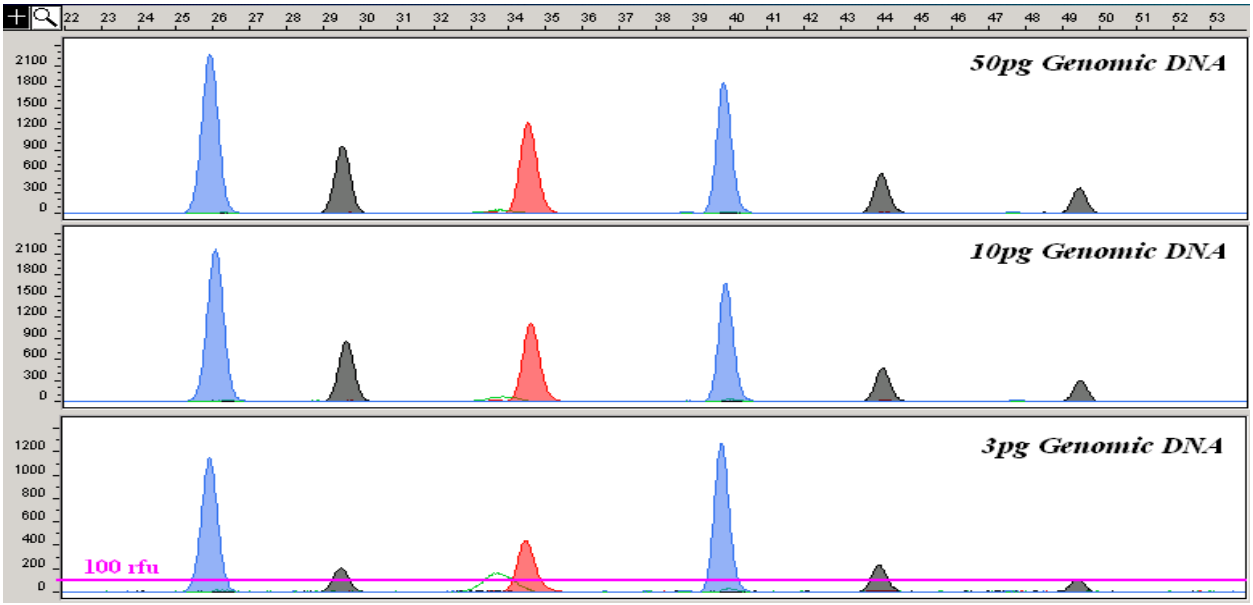


Figure 24. Sensitivity for Multiplex H.
Plots nucleotide size (x-axis) relative to –120LIZ size standard (Applied Systems) and relative fluorescent units (RFUs, y-axis). All 6 multiplexed sites were typed to 3pg.

Step 4: Mixture Testing

Following sensitivity testing, each multiplex was analyzed in regards to mixture/heteroplasmy detection. For each multiplex, samples were identified which would produce mixtures at two or more sites. The samples were quantified by an in-house Alu-repeat Taqman assay, and then the concentrations for each sample were normalized so that the concentrations for each extract were approximately equal. The samples were mixed in the following ratios for mixture testing: 100:0, 90:10, 80:20 (or 70:30), 50:50, 20:80 (or 30:70), 10:90, and 0:100. All mixtures to the 10% level were successfully resolved for each multiplex under standard reaction conditions, with the exception of one site in Multiplex E. Figures 25-31 show mixtures results for each multiplex at 90:10 and 50:50 ratios. In general, the electropherogram peaks resulting from ddG incorporation (blue) exhibited a stronger signal than ddA incorporation (green). As such, 50:50 mixtures do not show even signal intensity for ddG/ddA mixtures.

As discussed in one of our publications (Vallone *et al*., 2004), the peak height ratios for mixed samples have not been characterized well enough to permit a determination of the exact mixture ratio. An additional complication with the SNaPshot assay is the uneven signal intensity for the ddG/ddA mixtures. This system is more than sufficient, however, for identifying a mixed sample or a heteroplasmic SNP. Further, the inability to determine an exact mixture ratio is not limited to this assay, it is a known problem in mtDNA sequencing as well.

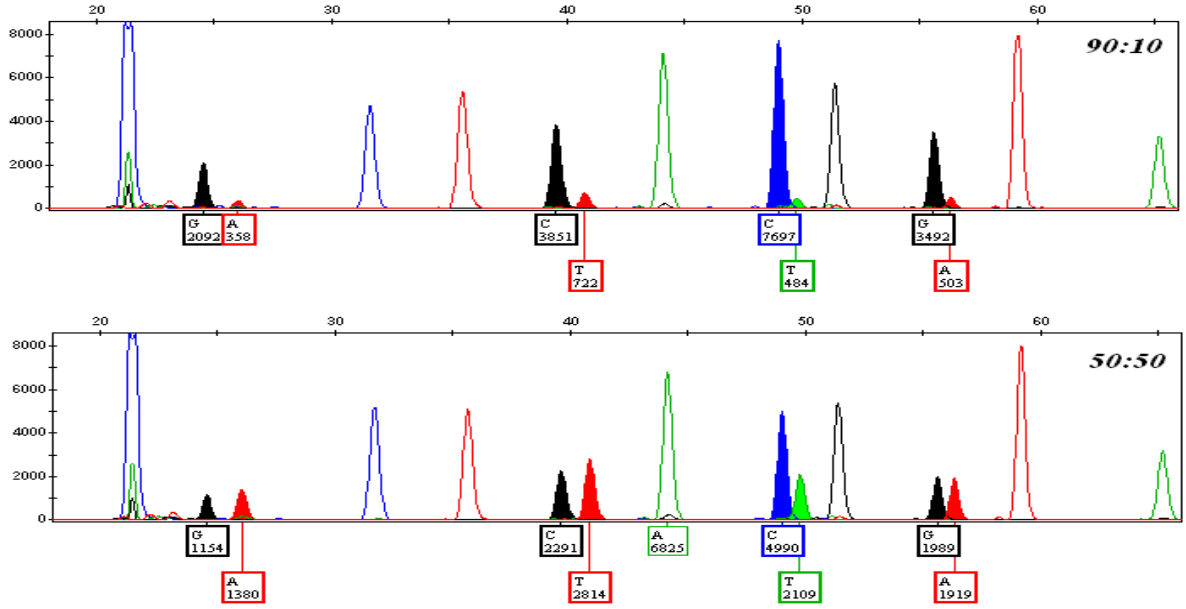


Figure 25. Mixtures at 4 SNPs for Multiplex A.
Plots nucleotide size (x-axis) relative to –120LIZ size standard (Applied Systems) and relative fluorescent units (RFUs, y-axis). Mixture ratios shown are 90:10 and 50:50.

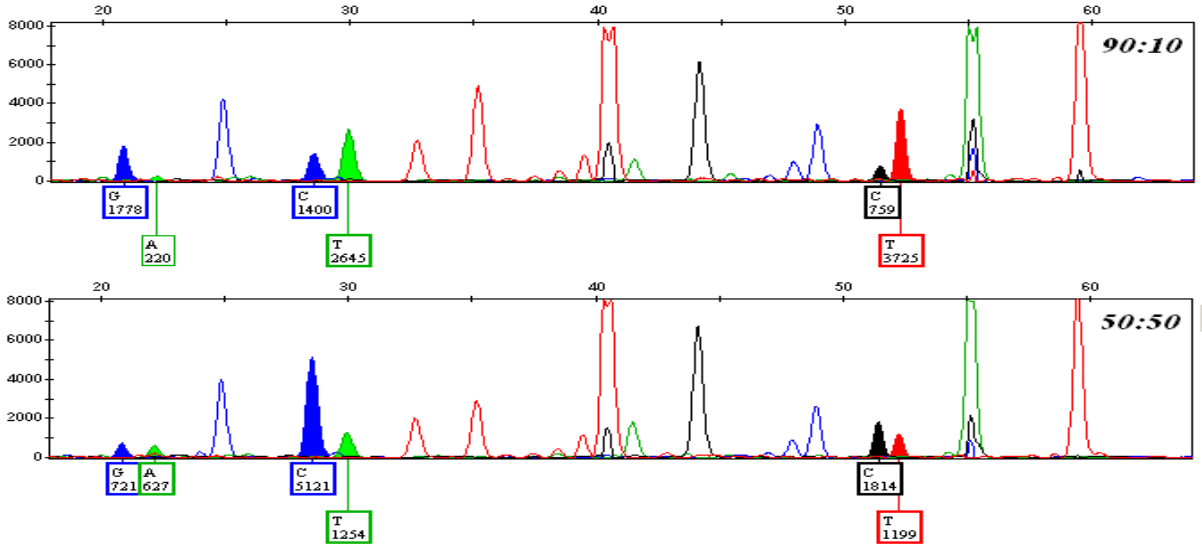


Figure 26. Mixtures at 4 SNPs for Multiplex B.
Plots nucleotide size (x-axis) relative to –120LIZ size standard (Applied Systems) and relative fluorescent units (RFUs, y-axis). Mixture ratios shown are 90:10 and 50:50.

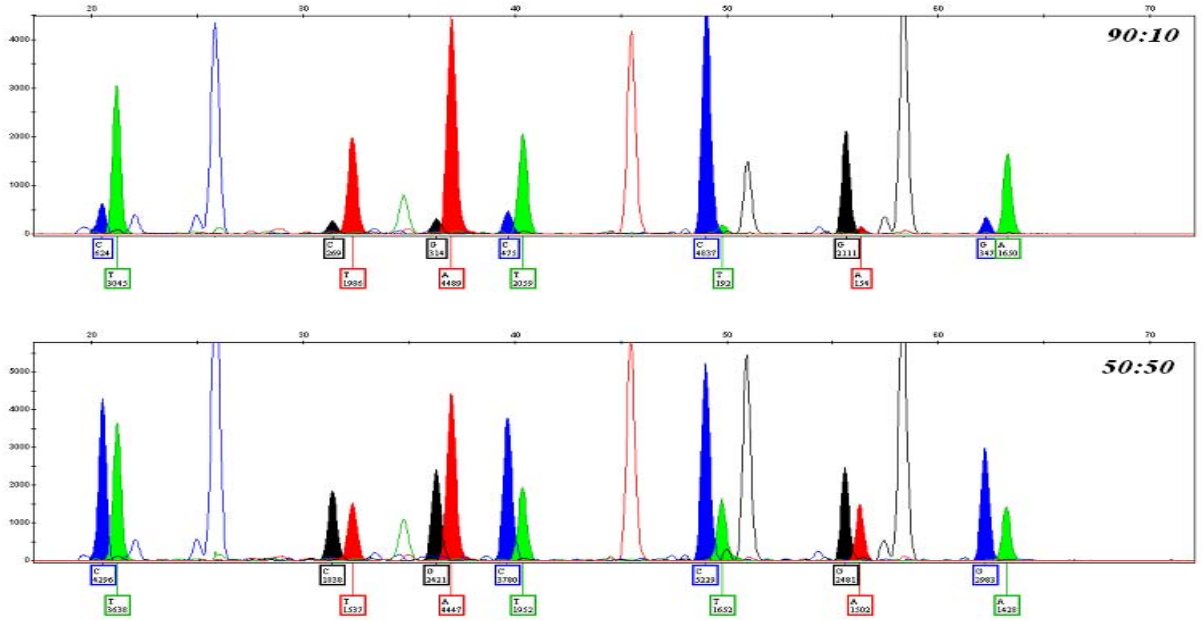


Figure 27. Mixtures at 7 SNPs for Multiplex C.
Plots nucleotide size (x-axis) relative to –120LIZ size standard (Applied Systems) and relative fluorescent units (RFUs, y-axis). Mixture ratios shown are 90:10 and 50:50.

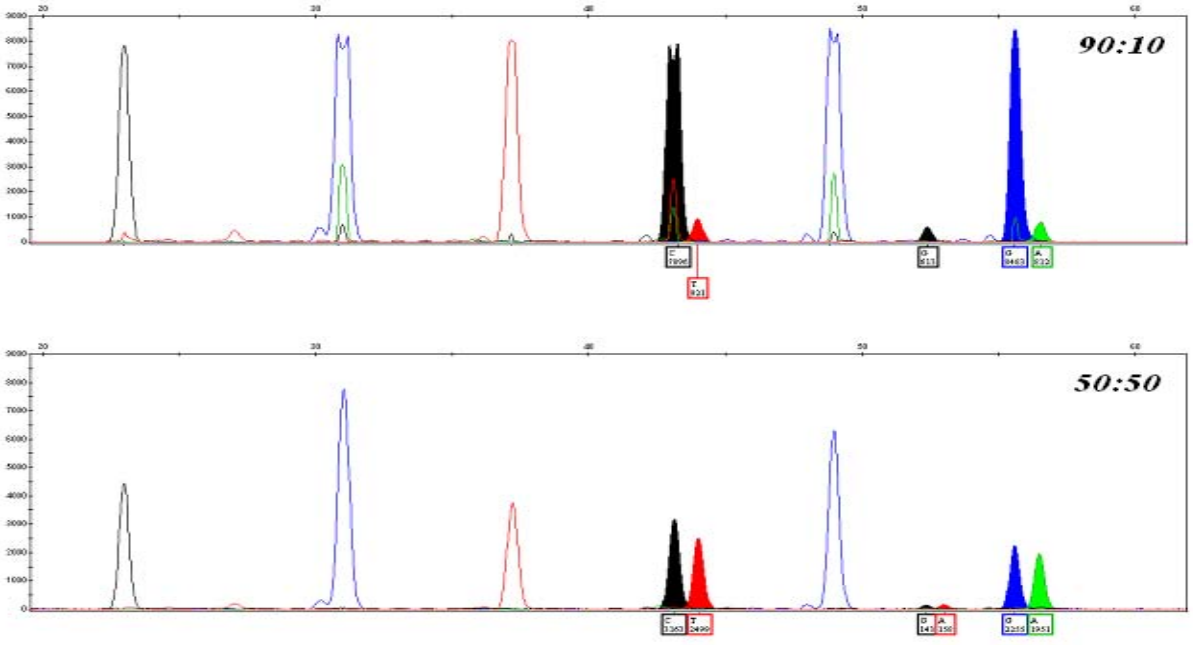


Figure 28. Mixtures at 3 SNPs for Multiplex E.
Plots nucleotide size (x-axis) relative to –120LIZ size standard (Applied Systems) and relative fluorescent units (RFUs, y-axis). Mixture ratios shown are 90:10 and 50:50.

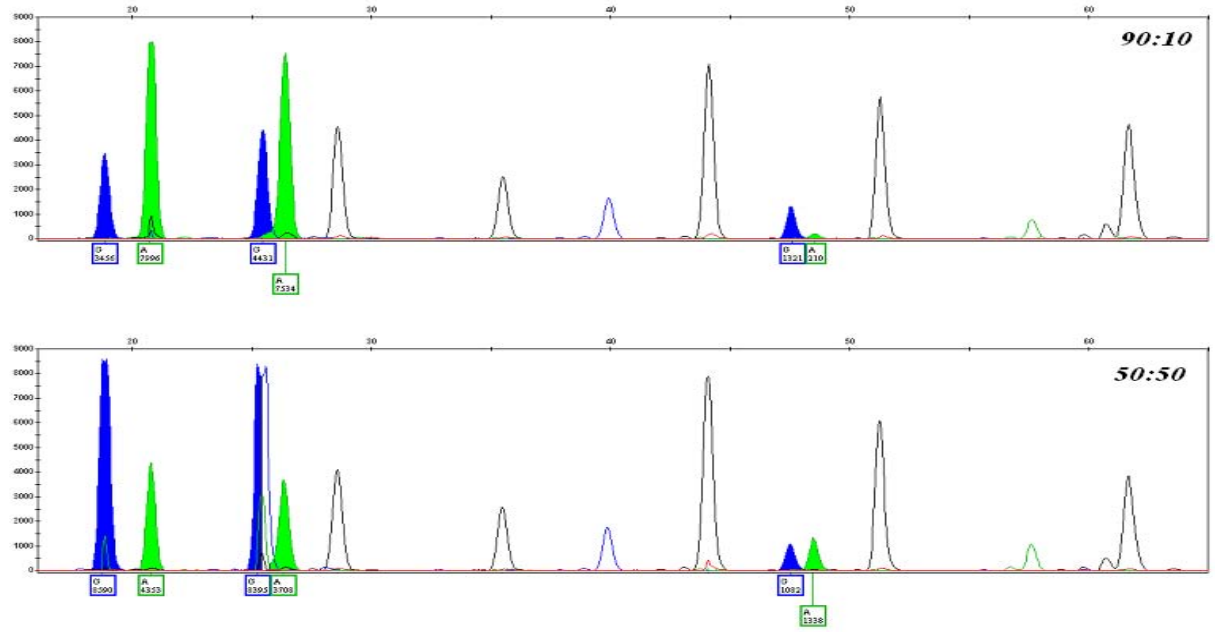


Figure 29. Mixtures at 3 SNPs for Multiplex F.
Plots nucleotide size (x-axis) relative to –120LIZ size standard (Applied Systems) and relative fluorescent units (RFUs, y-axis). Mixture ratios shown are 90:10 and 50:50.

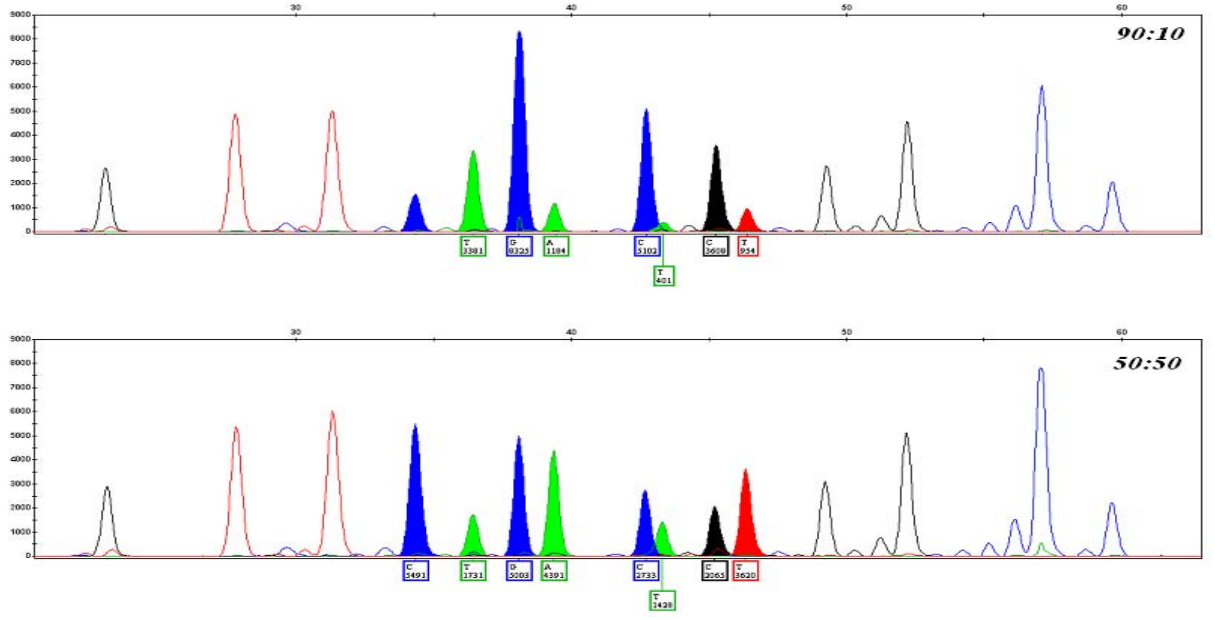


Figure 30. Mixtures at 4 SNPs for Multiplex G.
Plots nucleotide size (x-axis) relative to –120LIZ size standard (Applied Systems) and relative fluorescent units (RFUs, y-axis). Mixture ratios shown are 90:10 and 50:50.

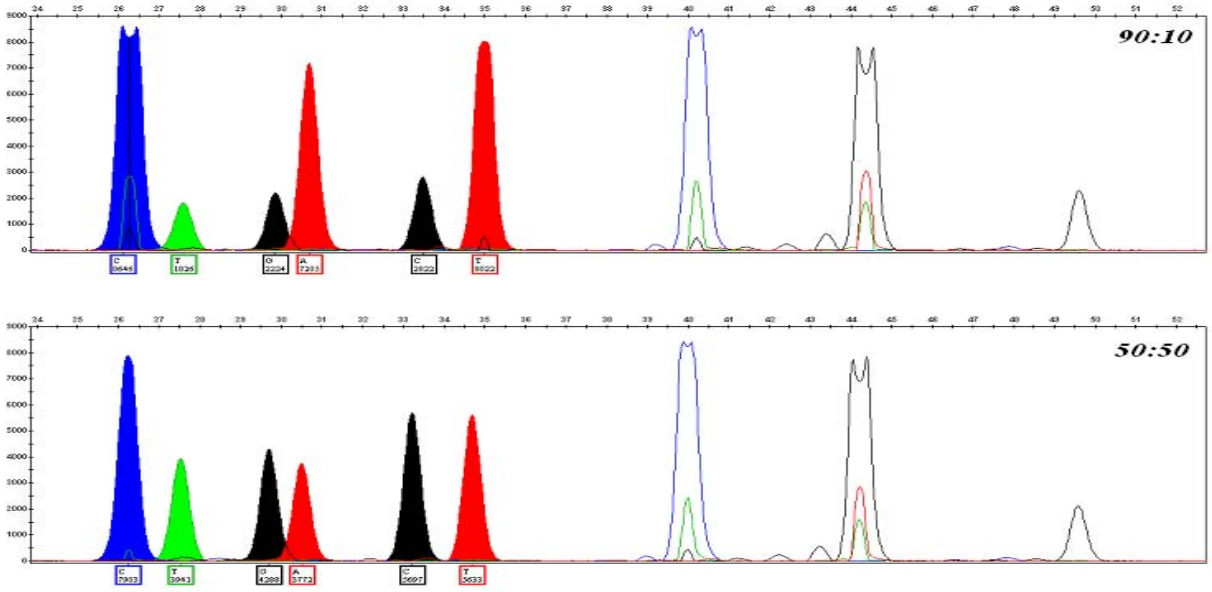


Figure 31. Mixtures at 3 SNPs for Multiplex H.
Plots nucleotide size (x-axis) relative to –120LIZ size standard (Applied Systems) and relative fluorescent units (RFUs, y-axis). Mixture ratios shown are 90:10 and 50:50.

Step 5: Degraded Samples Testing

We tested the multiplex panels on several non-probative extracts typical of AFDIL degraded skeletal mtDNA cases (standard STR typing failed to give results) to test the multiplex panels on degraded samples. Caucasian Multiplex A, being the first multiplex panel completed, was the first to be run on any degraded samples (Figure 32). Initial results at standard amplification conditions were poor, and most samples had only a few sites the exceeded the 100rfu threshold. Additional experimentation to increase success included increased reaction volume, decreased extract concentration, and increased Taq polymerase concentration. The final, optimized reaction conditions for degraded samples incorporates all of these modified conditions into a 25µl total volume, increased Taq concentration, degraded samples protocol (as opposed to the typical 15µl total volume standard reaction protocol, Table 26), and has resulted

in full typing on all degraded extracts thus far for all multiplexes tested (Figures 32-38) (Vallone *et al*., 2004). A single exception to this was site 3915 for Multiplex B, for which primer quality control issues prevented typing at either reaction volume prior to the completion of this report.

Components	25ul Rxn 1X	15ul Rxn 1X
Taq Gold PCR Buffer	*2.5ul*	*1.5ul*
dNTP's	*2.5ul*	*1.5ul*
BSA	*1.25ul*	*0.75ul*
MgCl2	*2.0ul*	*1.25ul*
Taq Gold	*1.25ul*	*0.5ul*
Primer Mix	*5.0ul*	*3.0ul*
dH20	*9.5ul*	*4.5ul*
Total volume per reaction	*24.0ul*	*13.0ul*
Extract Volume to be Added	*1.0ul*	*2.0ul*

Table 26. Standard and Degraded Samples Amplification Reaction Volumes.
The 15µl total volume reaction with standard Taq polymerase concentration is the standard protocol for pristine DNA extracts. The modified 25µl total volume reaction with increased Taq polymerase concentration is the protocol optimized for degraded samples.

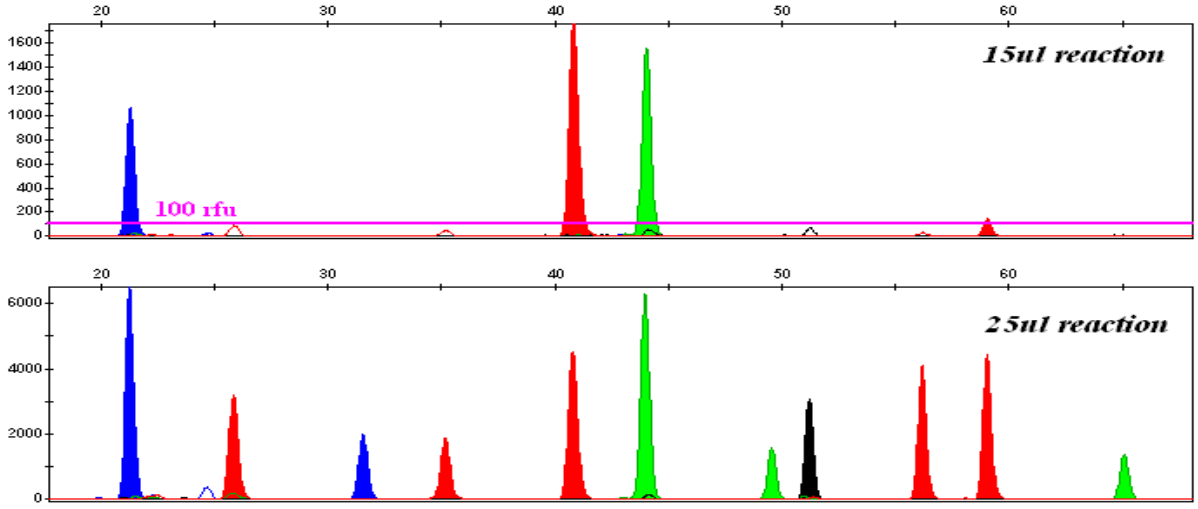


Figure 32. A degraded sample typed for Multiplex A using the 15 and 25µl protocols.
Plots of fragment size (x-axis) relative to –120LIZ size standard (Applied Systems) and relative fluorescent units (RFUs, y-axis). Only 4 sites exceeded the 100rfu threshold using the 15µl total reaction volume standard protocol, however all 11 sites in the multiplex could be typed using the modified 25µl total reaction volume, increased Taq polymerase concentration protocol.

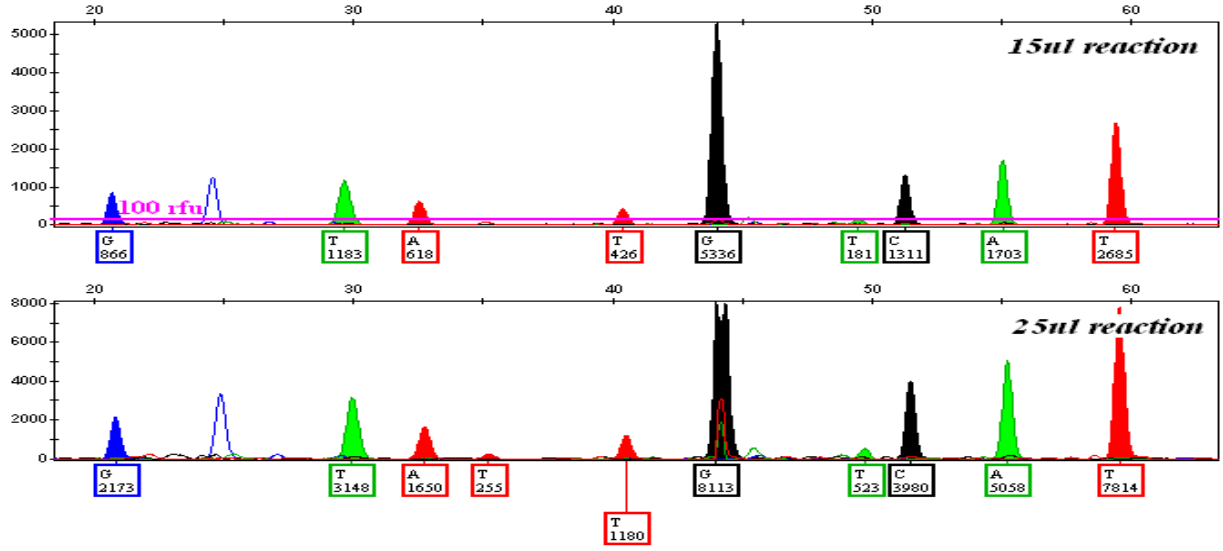


Figure 33. A degraded sample typed for Multiplex B using the 15 and 25µl protocols.
Plots of fragment size (x-axis) relative to –120LIZ size standard (Applied Systems) and relative fluorescent units (RFUs, y-axis). 9 of 11 sites exceeded the 100rfu threshold using the 15µl total reaction volume standard protocol, and 10 of 11 sites in the multiplex could be typed using the modified 25µl total reaction volume, increased Taq polymerase concentration protocol. The single remaining site that could not be typed at either reaction volume was the result of a poor primer lot, and additional runs will confirm typing of this site at the 25 µl reaction volume.

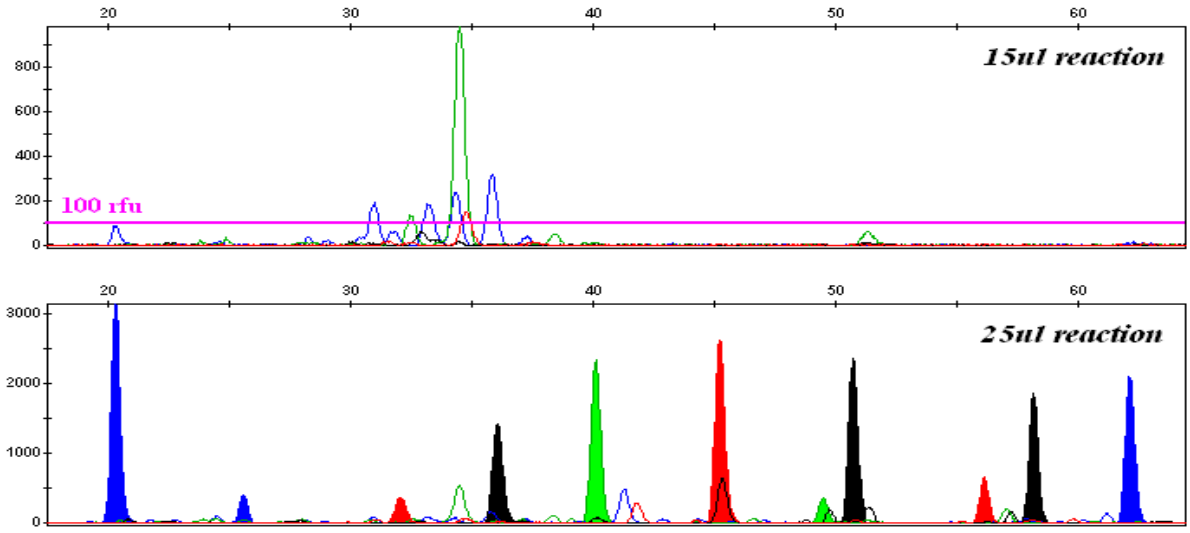


Figure 34. A degraded sample typed for Multiplex C using the 15 and 25µl protocols.
Plots of fragment size (x-axis) relative to –120LIZ size standard (Applied Systems) and relative fluorescent units (RFUs, y-axis). No sites exceeded the 100rfu threshold using the 15µl total reaction volume standard protocol, however all 11 sites in the multiplex could be typed using the modified 25µl total reaction volume, increased Taq polymerase concentration protocol.

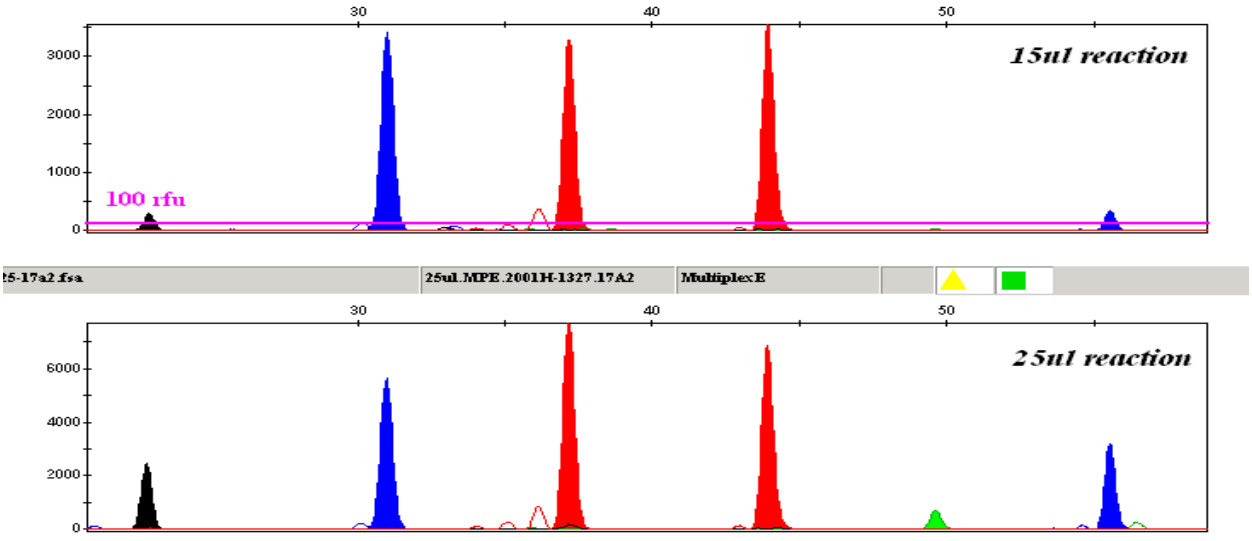


Figure 35. A degraded sample typed for Multiplex E using the 15 and 25µl protocols.
Plots of fragment size (x-axis) relative to –120LIZ size standard (Applied Systems) and relative fluorescent units (RFUs, y-axis). Five of six multiplexed sites exceeded the 100rfu threshold using the 15µl total reaction volume standard protocol, however all six sites in the multiplex could be typed using the modified 25µl total reaction volume, increased Taq polymerase concentration protocol.

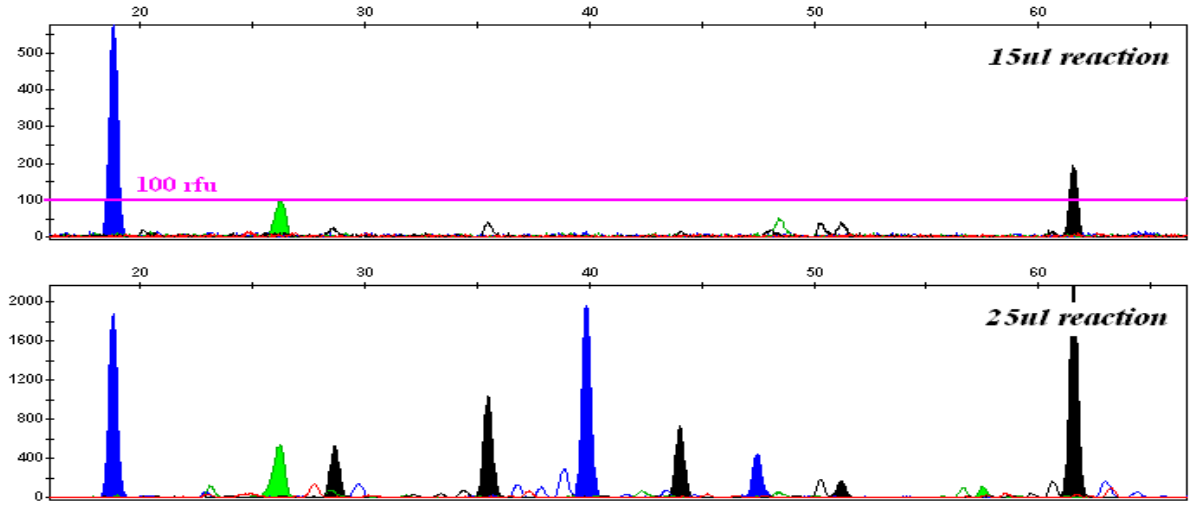


Figure 36. A degraded sample typed for Multiplex F using the 15 and 25µl protocols.
Plots of fragment size (x-axis) relative to –120LIZ size standard (Applied Systems) and relative fluorescent units (RFUs, y-axis). Only three of the 10 multiplexed sites exceeded the 100rfu threshold using the 15µl total reaction volume standard protocol, however all 10 sites in the multiplex could be typed using the modified 25µl total reaction volume, increased Taq polymerase concentration protocol.

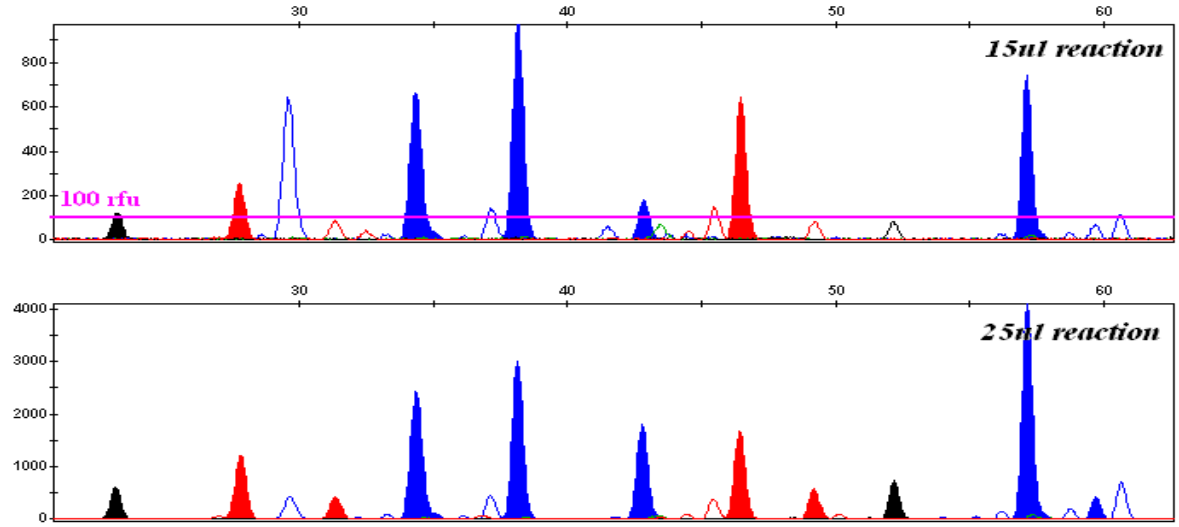


Figure 37. A degraded sample typed for Multiplex G using the 15 and 25µl protocols.
Plots of fragment size (x-axis) relative to –120LIZ size standard (Applied Systems) and relative fluorescent units (RFUs, y-axis). Seven of the 11 multiplexed sites exceeded the 100rfu threshold using the 15µl total reaction volume standard protocol, however all 11 sites in the multiplex could be typed using the modified 25µl total reaction volume, increased Taq polymerase concentration protocol.

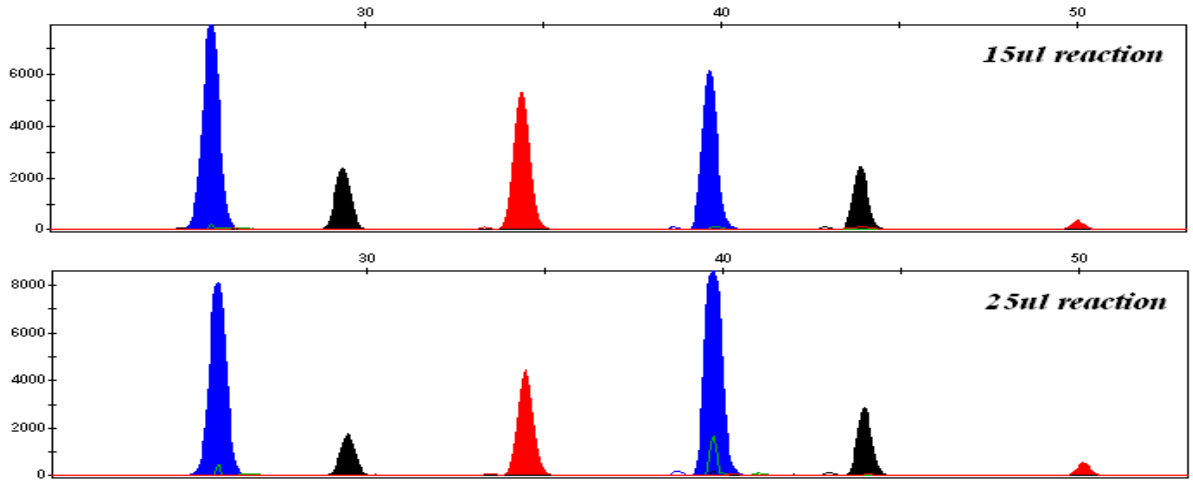


Figure 38. A degraded sample typed for Multiplex H using the 15 and 25µl protocols. Plots of fragment size (x-axis) relative to –120LIZ size standard (Applied Systems) and relative fluorescent units (RFUs, y-axis). Full typing was obtained using both the 15µl total reaction volume standard protocol and the modified 25µl total reaction volume, increased Taq polymerase concentration protocol.

Casework Application:

Multiplex panel A has now been used to assist in the resolution of multiple AFDIL cases in which CR sequencing alone did not give sufficient discrimination. The most common situation where the SNP panel has been applied are cases where CR sequencing showed only a single difference between samples. AFDIL reporting guidelines, typical for forensic labs, require at least two mtDNA sequence differences in order to report an exclusion. In other instances, the SNP panel testing has permitted both reassociation of unknown samples, and exclusion of references. Figure 39 depicts a typical example where Multiplex A typing was able to resolve a case by revealing an additional sequence difference, permitting a definitive exclusion of one reference family, and strengthening the evidentiary significance of a match with another reference family.

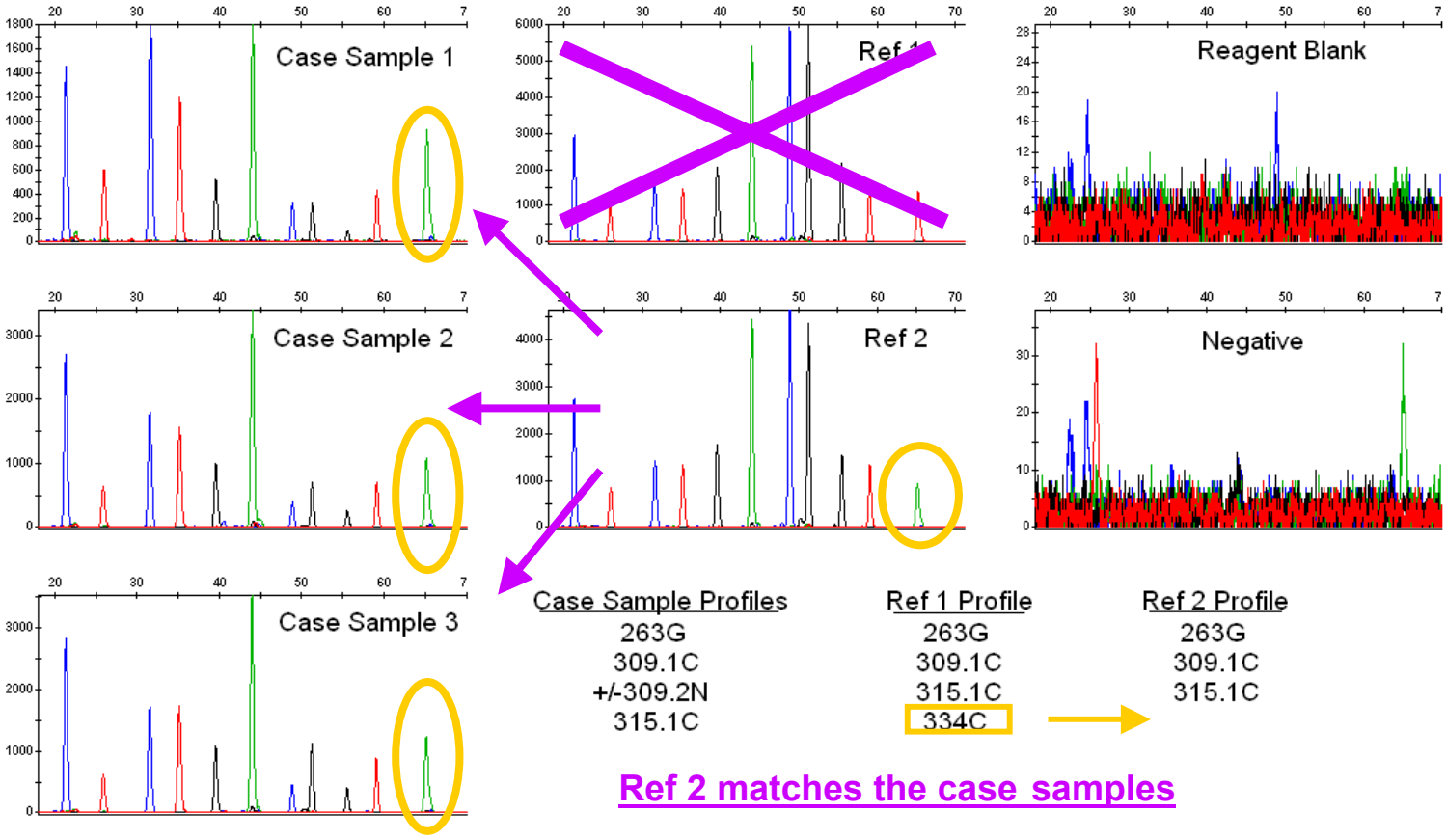


Figure 39. Multiplex A applied to an AFDIL missing person's case.
HVI/HVII sequencing on 3 case samples matched one reference family, but differed from a second reference family by only a single polymorphism (+/-334C). AFDIL policy dictates that an exclusion cannot be made on the basis of a single difference, and thus the case would have been reported as "inconclusive". SNP typing using Multiplex A provided resolution by identifying an additional position at which the mismatching reference differed from the case samples (+/- 14470A).

SNP Databasing

284 Caucasian population samples (source: NIST) were typed for seven Caucasian multiplex panels (A-C and E-H). Amplification set-up was performed in a 96-well format on a Corbett CAS single-probe instrument, and most post-PCR set-ups were performed in 96-well format on an 8-probe Tecan Genesis workstation. Detection was performed as previously

described on an Applied Biosystems 3100. For standard databasing, amplification thermal cycling parameters were decreased to a 26-cycle protocol. Some samples, however, were amplified with a 31-cycle amplification protocol due to low sample concentrations. In addition, some samples required longer or shorter injection times than the standard 13-second injection.

Sample analysis was performed in GeneMapper ID v3.2 (Applied Biosystems). Custom panels and bin sets were designed for each multiplex from samples previously tested and a set of analysis parameters (Table 27) was determined for analysis of all database samples. Following a structure similar to that described for sequencing, data analysis involved confirmation by at least two scientists , and a total of three levels of review. Primary and Secondary analyses were performed in GeneMapper, then imported into a custom Microsoft Excel macro for comparison. Once any discrepancies were resolved, the data was then exported to a secure Microsoft Access database for analysis of the results and long-term storage. Custom queries within Access were used to compile the results of the entire database.

In the course of databasing, null alleles were detected for a small number of samples for various multiplexes. For the NIST database samples that have not been sequenced for the mtGenome, we don't know the cause for the null alleles. However, past experience indicates that null alleles can readily result from mutation in the extension primer binding site. We have also observed that mutations in amplification primer binding sites can have large effects on signal strength, and this might also be a cause of the null alleles. Care was taken in selection of SNPs and primers to avoid binding site polymorphisms within the target common HV types, however random HV types, for which the multiplexes were not designed, could well harbor polymorphisms. For unknown reasons, 8 database samples did not produce full profiles for Multiplex C. In the presentation of results that follows, we omit samples that had either null

alleles or incomplete typing; this accounts for slight variation in sample sizes reported for the different multiplexes.

ANALYSIS PARAMETERS					
General	Analysis Type	SNaPshot			
Allele	Allele Cut-off value	0.01			
Peak Detector	Algorithm	Basic			
	Minimum Peak Heights (RFUs)	Blue	100	Red	100
		Green	100	Orange	50
		Yellow	100		
Peak Quality	Signal Level	HM min peak height			100
		HT min peak height			100
	Heterozygote balance	Min peak height ratio			0.3
	Peak morphology	Max peak width (bps)			1.5
	Pull-up peak	Pull-up ratio			0.1
	Allele number	Max expected alleles			1
	SNP	Double peak			0.5
Quality Flags	Quality Flag Settings	Spectral pull-up	0.5	Control Concordance	0.5
		Broad peak	0.5	Low peak height	0.5
		Narrow bin	0.5	Off-scale	0
		Double peak	0	Peak Height Ratio	0.5
	PQV Thresholds	**Range:**	**Pass**	**Low Quality**	
		Sizing Quality	0.75-1.0	0.0-0.25	
		Genotype Quality	0.75-1.0	0.0-0.25	

Table 27. GeneMapperID ID v3.2 Analysis Parameters.

Figures 40-46 depict the breakdown of SNP-types when each multiplex panel is applied to a set of random Caucasian population samples. A SNP-type is defined here as the profile generated using only a single multiplex panel, and excluding HV sequence information. Multiplex A (Figure 40), targeting HV-type H:1, provides the greatest discrimination among a set of random samples. Its application results in 17 SNP-types, and no single type has a frequency greater than 30%. This result is not unexpected given the high frequency of the H:1 type in the Caucasian population (~7%) and in our dataset (~6%, or 17 samples out of the 281

samples data based for Multiplex A). Multiplex F (Figure 44), the secondary multiplex for HV-types H:1, H:2, H:3, and V:1, provides the least stand-alone discrimination among a set of random samples. Though application of this multiplex results in 13 SNP-types, a single SNP-type occurs with a frequency of nearly 90% and more than half of the SNP-types occur with a frequency of less than 1%. This result is also not unexpected given that Multiplex F is comprised of the second tier of discriminatory SNPs identified for the four HV-types the panel targets.

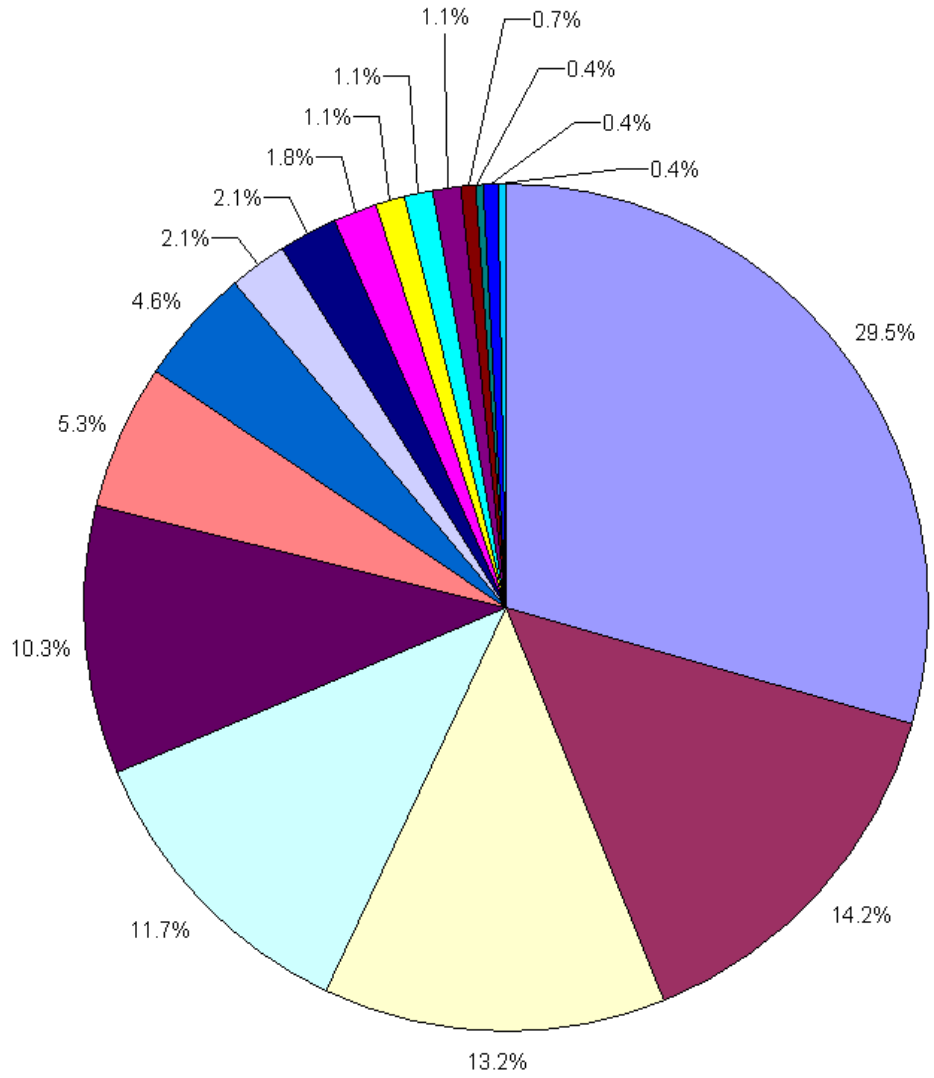


Figure 40. Multiplex A applied to 281* Caucasian samples.
The 11 multiplexed SNP sites of Multiplex A produced 17 SNP-types, 3 of which were unique.
*Samples with null alleles (3) were not used in the analysis of this panel.

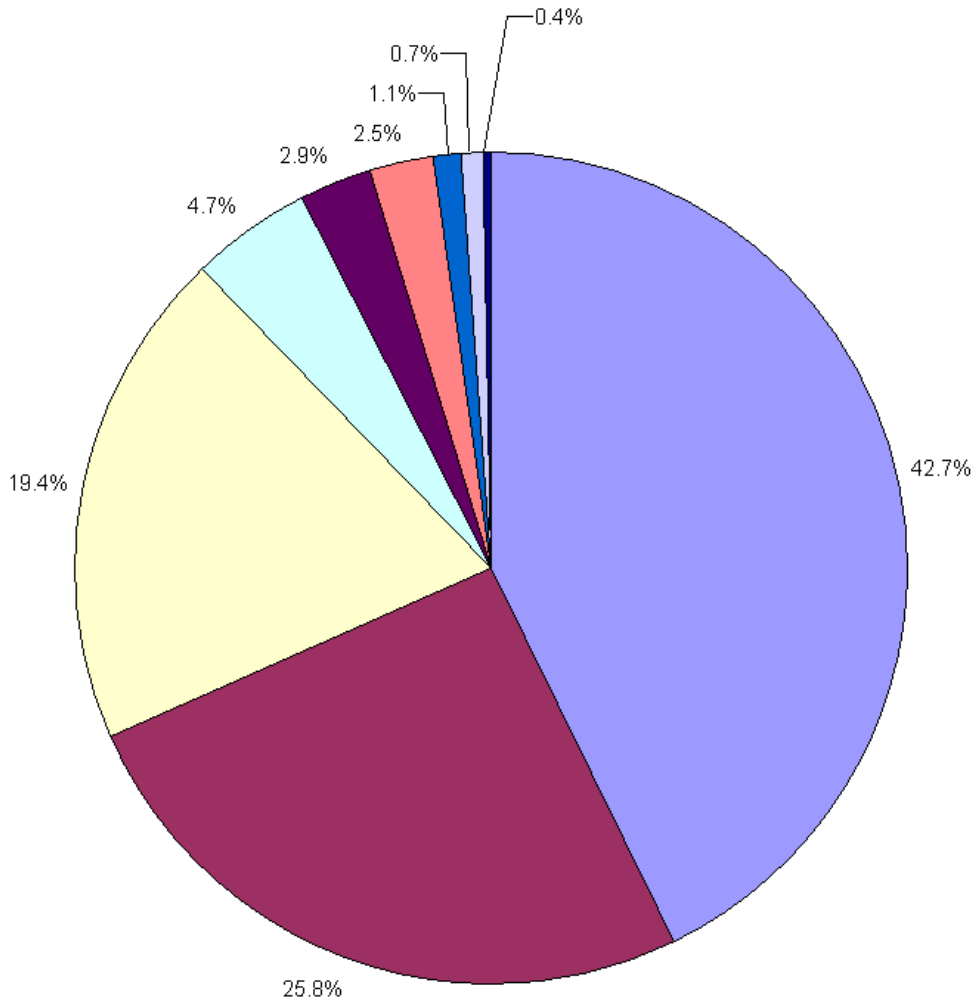


Figure 41. Multiplex B applied to 279* Caucasian samples.
The 10^{+} multiplexed SNP sites of Multiplex B produced 9 SNP-types, 1 of which was unique.
*Samples with null alleles (5) were not used in the analysis of this panel.
+Site 3915 was not used for data analysis, as this site gave unusually low signal strength during the particular databasing run, and we haven't yet resolved this presumably sporadic problem.

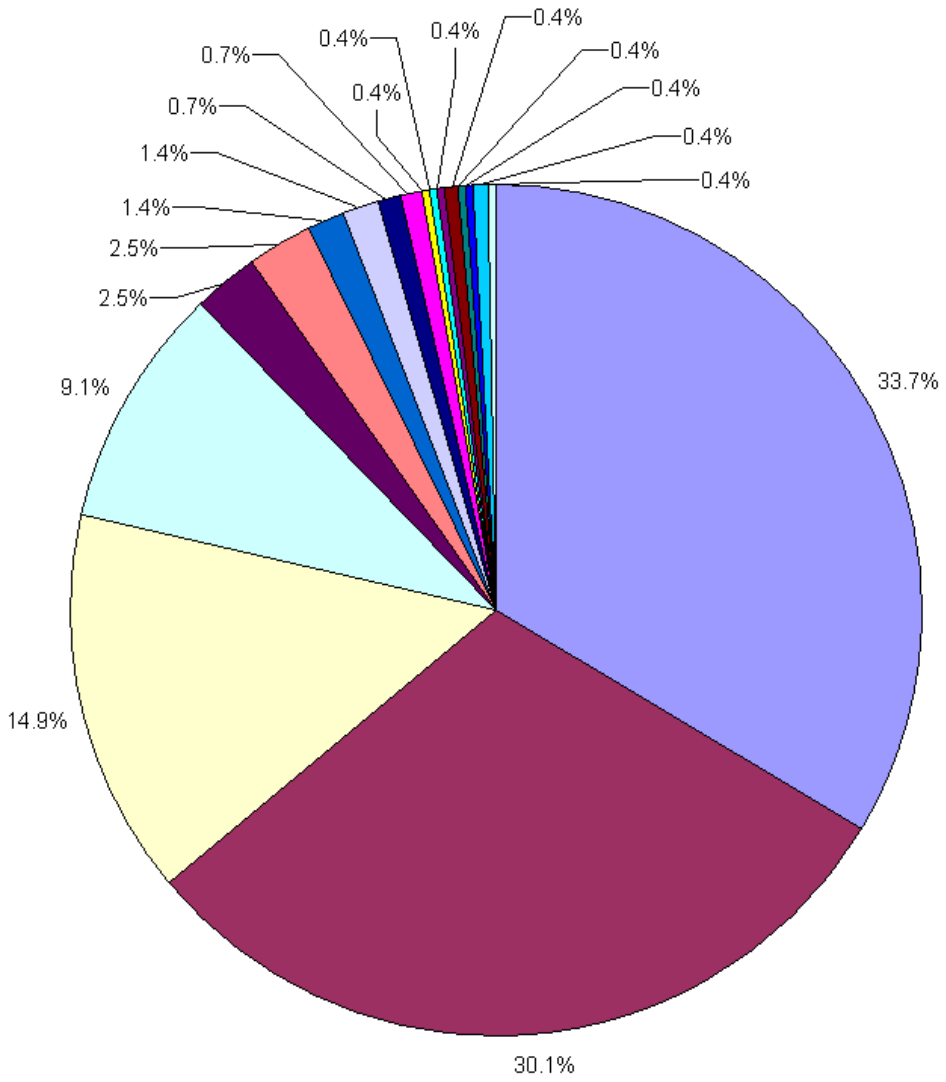


Figure 42. Multiplex C applied to 276* Caucasian samples.
The 11 multiplexed SNP sites of Multiplex C produced 18 SNP-types, 8 of which were unique.
*Samples with partial profiles (8) were not used in the analysis of this panel.

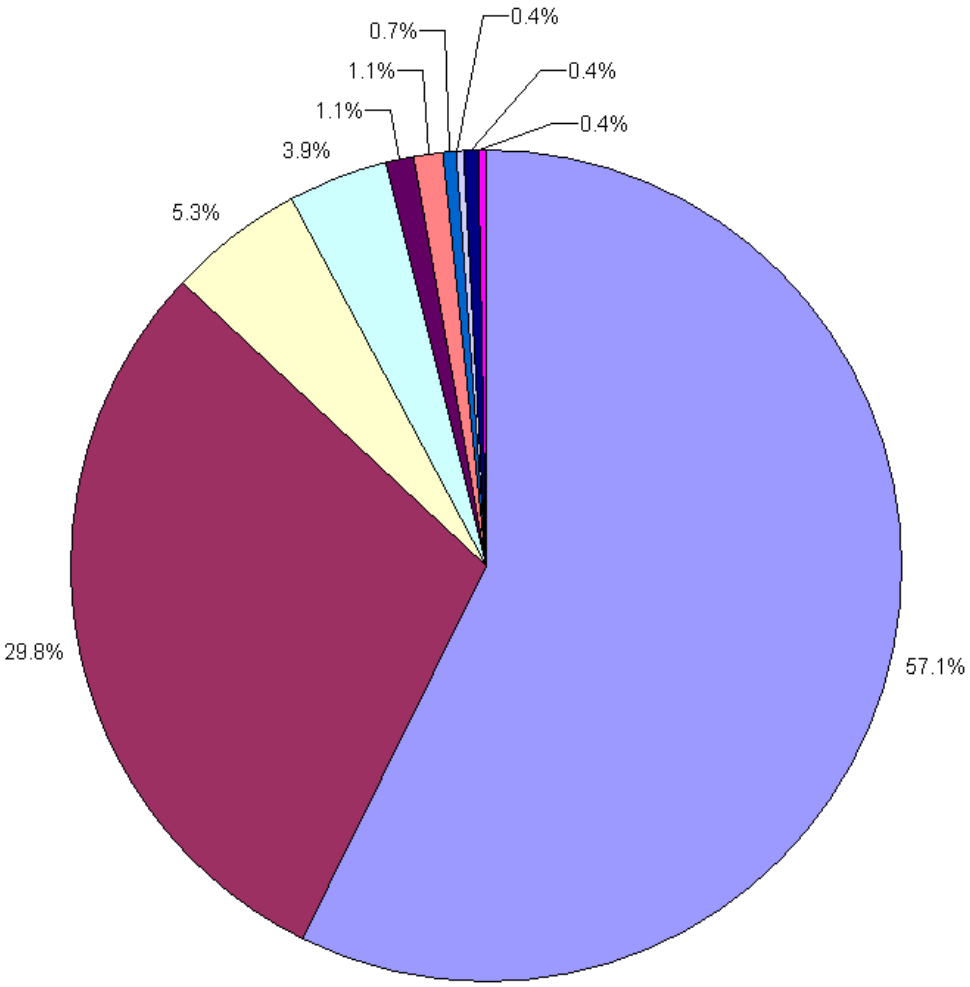


Figure 43. Multiplex E applied to 282* Caucasian samples.
The 7 multiplexed SNP sites of Multiplex E produced 10 SNP-types, 3 of which were unique.
*Samples with null alleles (2) were not used in the analysis of this panel.

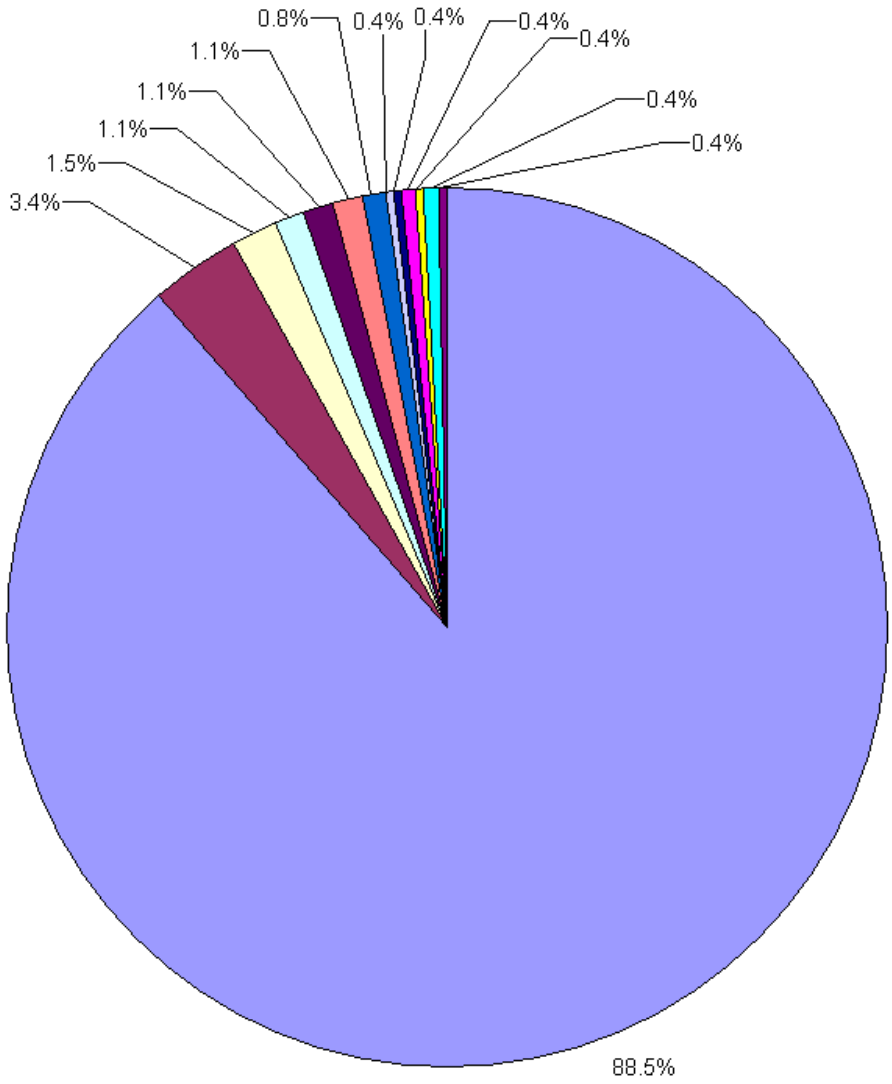


Figure 44. Multiplex F applied to 261* Caucasian samples.
The 10 multiplexed SNP sites of Multiplex F produced 13 SNP-types, 6 of which were unique.
*Samples with null alleles (23) were not used in the analysis of this panel.

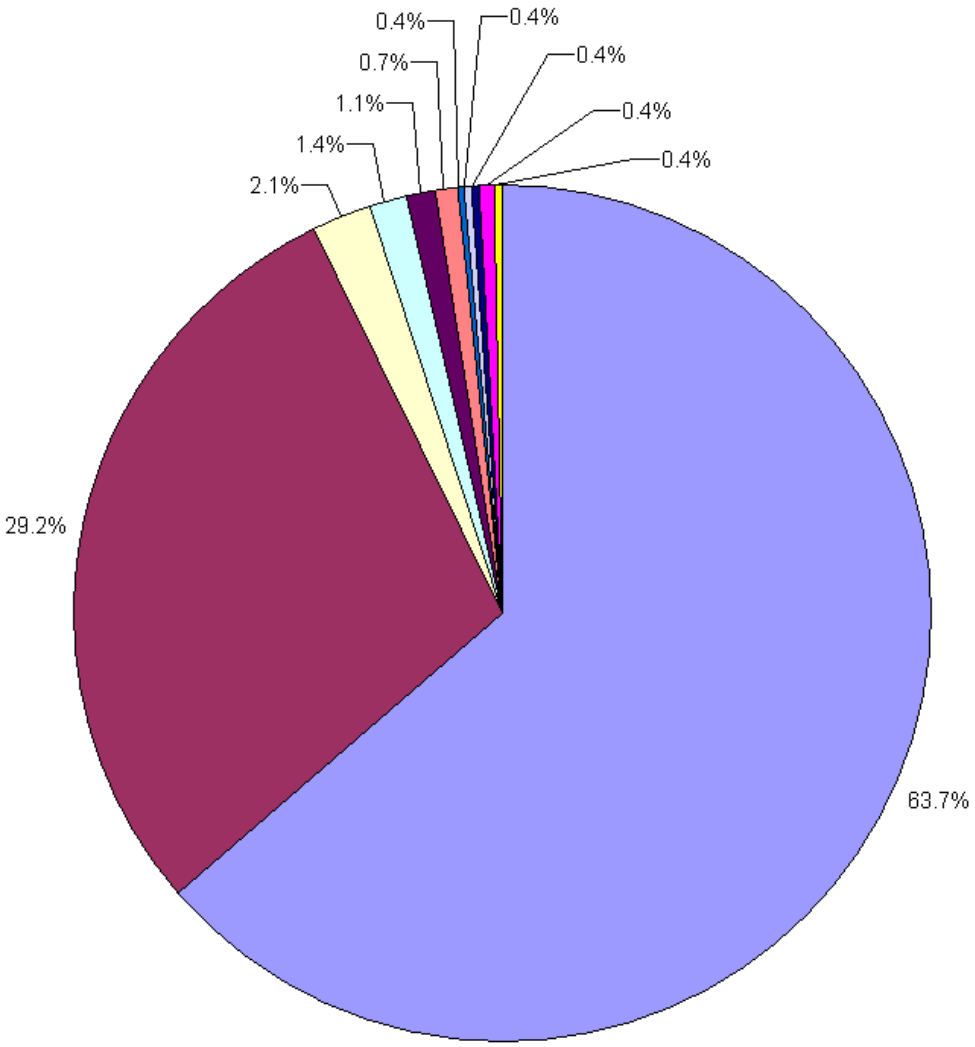


Figure 45. Multiplex G applied to 281* Caucasian samples.
The 11 multiplexed SNP sites of Multiplex G produced 11 SNP-types, 5 of which were unique.
*Samples with null alleles (3) were not used in the analysis of this panel.

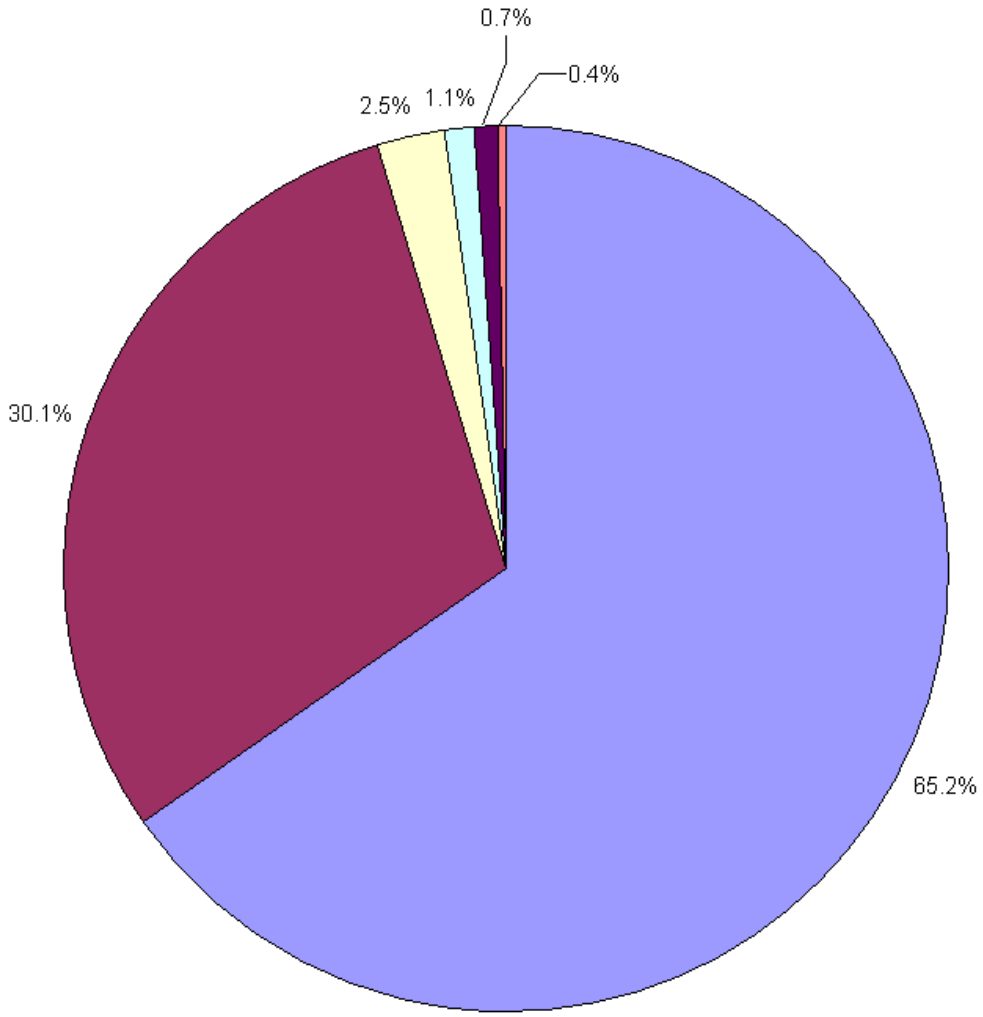


Figure 46. Multiplex H applied to 282* Caucasian samples.
The 6 multiplexed SNP sites of Multiplex H produced 6 SNP-types, 1 of which was unique.
*Samples with null alleles (2) were not used in the analysis of this panel.

Tables 28-37 depict the breakdown of HV/SNP-types among HV-types that occurred more than once in the random Caucasian dataset. An HV/SNP-type is defined here as the profile generated when HV sequence data and SNPs from a multiplex panel are taken together. Multiplex A discriminates well among the 17 samples matching HV-type H:1, resulting in six HV/SNP-types, four of which are unique (Table 28). The results depicted in Table 28 also

indicate that Multiplex A discriminates well among samples that a match a near-H:1 type (here defined as an HV-type that differs from H:1 by only a single polymorphism), providing additional discrimination for three out of the four near-H:1 types duplicated in our dataset. Multiplex B, which targets additional haplogroup H HV-types, also performed well on the targeted and closely-related HV-types (Table 29).

HV Type	HV Profile	Total	Breakdown
H:1	263G 315.1C	17	(8) (5) (1) (1) (1) (1)
H:1 + 16189	16189C 263G 315.1C	5	(3) (1) (1)
H:1 + 16311	16311C 263G 315.1C	3	(1) (1) (1)
H:1 + 16355	16355T 263G 315.1C	2	(2)
H:1 + 93	93G 263G 315.1C	3	(1) (1) (1)
H:2	152C 263G 315.1C	7	(4) (2) (1)
H:3	16129A 263G 315.1C	3	(2) (1)
H:5	16304C 263G 315.1C	6	(6)
H:7	16162G 16209C 73G 263G 315.1C	5	(5)
J:3	16069T 16126C 73G 185A 188G 228A 263G 295T 315.1C	2	(2)
J:3 + 185	16069T 16126C 73G 146C 185A 188G 228A 263G 295T 315.1C	2	(2)
K:1	16224C 16311C 73G 146C 152C 263G 315.1C	2	(2)
K:3 + 195	16224C 16311C 73G 195C 263G 315.1C	2	(2)
T:1	16126C 16294T 16296T 16304C 73G 263G 315.1C	2	(2)
T:2	16126C 16163G 16186T 16189C 16294T 73G 152C 195C 263G 315.1C	4	(4)
T:3	16126C 16294T 16296T 73G 263G 315.1C	3	(3)
V:1	16298C 263G 315.1C	3	(2) (1)
-	16069T 16126C 16145A 16231C 16261T 73G 150T 152C 195C 215G 263G 295T 310.1T 315.1C 319C	6	(5) (1)
-	16126C 16172C 16266T 16292T 16294T 73G 152C 263G 315.1C	2	(4)
-	16126C 16187T 16294T 16296T 16304C 73G 151T 263G 315.1C	2	(2)
-	16147A 16154C 16172C 16223T 16248T 16320T 16355T 73G 152C 199C 204C 263G 315.1C	2	(2)
-	16256T 16270T 73G 263G 315.1C	4	(2)
-	16293G 16311C 195C 263G 315.1C	2	(1) (1)
-	16294T 16304C 198T 263G 315.1C	2	(2)
-	16362C 239C 263G 315.1C	2	(2)
		93	**1 null allele***

Table 28. Breakdown of 93* non-unique HV-type samples by Multiplex A.
The application of Multiplex A to 25 non-unique HV-types resulted in 42 different HV/SNP-types including 18 unique types. This panel targets the H:1 HV-type, highlighted in yellow.
*One sample with a null allele was not included for data analysis.

HV Type	HV Profile	Total	Breakdown
H:1	263G 315.1C	17	(6) (5) (2) (1) (1) (1) (1)
H:1 + 16189	16189C 263G 315.1C	6	(4) (1) (1)
H:1 + 16311	16311C 263G 315.1C	3	(2) (1)
H:1 + 16355	16355T 263G 315.1C	2	(2)
H:1 + 93	93G 263G 315.1C	3	(1) (1) (1)
H:2	152C 263G 315.1C	7	(2) (2) (1) (1) (1)
H:3	16129A 263G 315.1C	3	(1) (1) (1)
H:5	16304C 263G 315.1C	6	(6)
H:7	16162G 16209C 73G 263G 315.1C	5	(5)
J:3	16069T 16126C 73G 185A 188G 228A 263G 295T 315.1C	2	(2)
J:3 + 185	16069T 16126C 73G 146C 185A 188G 228A 263G 295T 315.1C	2	(2)
K:1	16224C 16311C 73G 146C 152C 263G 315.1C	2	(2)
K:3 + 195	16224C 16311C 73G 195C 263G 315.1C	2	(2)
T:1	16126C 16294T 16296T 16304C 73G 263G 315.1C	2	(2)
T:2	16126C 16163G 16186T 16189C 16294T 73G 152C 195C 263G 315.1C	4	(4)
T:3	16126C 16294T 16296T 73G 263G 315.1C	3	(3)
V:1	16298C 263G 315.1C	3	(2) (1)
-	16069T 16126C 16145A 16231C 16261T 73G 150T 152C 195C 215G 263G 295T 310.1T 315.1C 319C	6	(5) (1)
-	16126C 16172C 16266T 16292T 16294T 73G 152C 263G 315.1C	2	(2)
-	16126C 16187T 16294T 16296T 16304C 73G 151T 263G 315.1C	2	(2)
-	16147A 16154C 16172C 16223T 16248T 16320T 16355T 73G 152C 199C 204C 263G 315.1C	2	(2)
-	16256T 16270T 73G 263G 315.1C	4	(4)
-	16293G 16311C 195C 263G 315.1C	2	(1) (1)
-	16294T 16304C 198T 263G 315.1C	2	(2)
-	16362C 239C 263G 315.1C	2	(2)
Also targets H:6 HV type		**94**	

Table 29. Breakdown of 94 non-unique HV-type samples by Multiplex B.
The application of Multiplex B to 25 non-unique HV-types resulted in 45 different HV/SNP-types including 20 unique types. This panel targets H:2 and H:3 HV-types, highlighted in yellow.

HV Type	V Profile	Total	Breakdown
H:1	263G 315.1C	16	(13) (2) (1)
H:1 + 16189	16189C 263G 315.1C	6	(5) (1)
H:1 + 16311	16311C 263G 315.1C	3	(2) (1)
H:1 + 16355	16355T 263G 315.1C	2	(2)
H:1 + 93	93G 263G 315.1C	3	(2) (1)
H:2	152C 263G 315.1C	7	(6) (1)
H:3	16129A 263G 315.1C	3	(2) (1)
H:5	16304C 263G 315.1C	5	(3) (2)
H:7	16162G 16209C 73G 263G 315.1C	5	(5)
J:3	16069T 16126C 73G 185A 188G 228A 263G 295T 315.1C	2	(2)
J:3 + 185	16069T 16126C 73G 146C 185A 188G 228A 263G 295T 315.1C	2	(2)
K:1	16224C 16311C 73G 146C 152C 263G 315.1C	2	(2)
K:3 + 195	16224C 16311C 73G 195C 263G 315.1C	2	(1) (1)
T:1	16126C 16294T 16296T 16304C 73G 263G 315.1C	2	(2)
T:2	16126C 16163G 16186T 16189C 16294T 73G 152C 195C 263G 315.1C	4	(4)
T:3	16126C 16294T 16296T 73G 263G 315.1C	3	(3)
V:1	16298C 263G 315.1C	3	(2) (1)
-	16069T 16126C 16145A 16231C 16261T 73G 150T 152C 195C 215G 263G 295T 310.1T 315.1C 319C	6	(5) (1)
-	16126C 16172C 16266T 16292T 16294T 73G 152C 263G 315.1C	2	(2)
-	16126C 16187T 16294T 16296T 16304C 73G 151T 263G 315.1C	2	(2)
-	16147A 16154C 16172C 16223T 16248T 16320T 16355T 73G 152C 199C 204C 263G 315.1C	2	(2)
-	16256T 16270T 73G 263G 315.1C	4	(4)
-	16293G 16311C 195C 263G 315.1C	2	(1) (1)
-	16294T 16304C 198T 263G 315.1C	2	(2)
-	16362C 239C 263G 315.1C	2	(2)
		92	**2 partial profiles***

Table 30. Breakdown of 92* non-unique HV-type samples by Multiplex C.
The application of Multiplex C to 25 non-unique HV-types resulted in 37 different HV/SNP-types including 12 unique types. This panel targets H:5 and V:1 HV-types, highlighted in yellow.
*Samples with partial profiles (2) were not included for data analysis.

HV Type	HV Profile	Total	Breakdown
H:1	263G 315.1C	17	(15) (2)
H:1 + 16189	16189C 263G 315.1C	6	(5) (1)
H:1 + 16311	16311C 263G 315.1C	3	(2) (1)
H:1 + 16355	16355T 263G 315.1C	2	(2)
H:1 + 93	93G 263G 315.1C	3	(2) (1)
H:2	152C 263G 315.1C	7	(6) (1)
H:3	16129A 263G 315.1C	3	(2) (1)
H:5	16304C 263G 315.1C	6	(6)
H:7	16162G 16209C 73G 263G 315.1C	5	(4) (1)
J:3	16069T 16126C 73G 185A 188G 228A 263G 295T 315.1C	2	(2)
J:3 + 185	16069T 16126C 73G 146C 185A 188G 228A 263G 295T 315.1C	2	(2)
K:1	16224C 16311C 73G 146C 152C 263G 315.1C	2	(2)
K:3 + 195	16224C 16311C 73G 195C 263G 315.1C	2	(2)
T:1	16126C 16294T 16296T 16304C 73G 263G 315.1C	2	(2)
T:2	16126C 16163G 16186T 16189C 16294T 73G 152C 195C 263G 315.1C	4	(4)
T:3	16126C 16294T 16296T 73G 263G 315.1C	3	(3)
V:1	16298C 263G 315.1C	3	(2) (1)
-	16069T 16126C 16145A 16231C 16261T 73G 150T 152C 195C 215G 263G 295T 310.1T 315.1C 319C	6	(5) (1)
-	16126C 16172C 16266T 16292T 16294T 73G 152C 263G 315.1C	2	(2)
-	16126C 16187T 16294T 16296T 16304C 73G 151T 263G 315.1C	2	(2)
-	16147A 16154C 16172C 16223T 16248T 16320T 16355T 73G 152C 199C 204C 263G 315.1C	2	(2)
-	16256T 16270T 73G 263G 315.1C	4	(4)
-	16293G 16311C 195C 263G 315.1C	2	(1) (1)
-	16294T 16304C 198T 263G 315.1C	2	(2)
-	16362C 239C 263G 315.1C	2	(2)
Also targets H:4 and J:4 HV types		**94**	

Table 31. Breakdown of 94 non-unique HV-type samples by Multiplex E.
The application of Multiplex E to 25 non-unique HV-types resulted in 35 different HV/SNP-types including 10 unique types. This panel targets H:4, J:4, T:2, and T:3 HV-types, highlighted in yellow.

HV Type	HV Profile	Total	Breakdown
H:1	263G 315.1C	17	(15) (1) (1)
H:1 + 16189	16189C 263G 315.1C	6	(6)
H:1 + 16311	16311C 263G 315.1C	3	(3)
H:1 + 16355	16355T 263G 315.1C	2	(2)
H:1 + 93	93G 263G 315.1C	3	(3)
H:2	152C 263G 315.1C	7	(6) (1)
H:3	16129A 263G 315.1C	3	(1) (1) (1)
H:5	16304C 263G 315.1C	6	(6)
H:7	16162G 16209C 73G 263G 315.1C	5	(5)
J:3	16069T 16126C 73G 185A 188G 228A 263G 295T 315.1C	2	(2)
J:3 + 185	16069T 16126C 73G 146C 185A 188G 228A 263G 295T 315.1C	2	(2)
K:1	16224C 16311C 73G 146C 152C 263G 315.1C	2	(2)
K:3 + 195	16224C 16311C 73G 195C 263G 315.1C	2	(2)
T:1	16126C 16294T 16296T 16304C 73G 263G 315.1C	1	(1)
T:2	16126C 16163G 16186T 16189C 16294T 73G 152C 195C 263G 315.1C	2	(2)
T:3	16126C 16294T 16296T 73G 263G 315.1C	1	(1)
V:1	16298C 263G 315.1C	3	(3)
-	16069T 16126C 16145A 16231C 16261T 73G 150T 152C 195C 215G 263G 295T 310.1T 315.1C 319C	6	(6)
-	16126C 16172C 16266T 16292T 16294T 73G 152C 263G 315.1C	1	(1)
-	16126C 16187T 16294T 16296T 16304C 73G 151T 263G 315.1C	1	(1)
-	16147A 16154C 16172C 16223T 16248T 16320T 16355T 73G 152C 199C 204C 263G 315.1C	2	(2)
-	16256T 16270T 73G 263G 315.1C	4	(4)
-	16293G 16311C 195C 263G 315.1C	2	(2)
-	16294T 16304C 198T 263G 315.1C	2	(2)
-	16362C 239C 263G 315.1C	2	(2)
		87	**7 null alleles***

Table 32. Breakdown of 87* non-unique HV-type samples by Multiplex F.
The application of Multiplex F to 21 non-unique HV-types resulted in 26 different HV/SNP-types including 6 unique types. This panel provides additional discrimination for H:1, H:2, H:3, and V:1 HV-types, highlighted in light yellow.
*Samples with null alleles (7) were not included for data analysis.

HV Type	HV Profile	Total	Breakdown
H:1	263G 315.1C	17	(14) (2) (1)
H:1 + 16189	16189C 263G 315.1C	6	(6)
H:1 + 16311	16311C 263G 315.1C	3	(2) (1)
H:1 + 16355	16355T 263G 315.1C	2	(2)
H:1 + 93	93G 263G 315.1C	3	(2) (1)
H:2	152C 263G 315.1C	7	(6) (1)
H:3	16129A 263G 315.1C	3	(2) (1)
H:5	16304C 263G 315.1C	6	(6)
H:7	16162G 16209C 73G 263G 315.1C	5	(5)
J:3	16069T 16126C 73G 185A 188G 228A 263G 295T 315.1C	2	(2)
J:3 + 185	16069T 16126C 73G 146C 185A 188G 228A 263G 295T 315.1C	2	(2)
K:1	16224C 16311C 73G 146C 152C 263G 315.1C	2	(2)
K:3 + 195	16224C 16311C 73G 195C 263G 315.1C	2	(1) (1)
T:1	16126C 16294T 16296T 16304C 73G 263G 315.1C	2	(2)
T:2	16126C 16163G 16186T 16189C 16294T 73G 152C 195C 263G 315.1C	4	(4)
T:3	16126C 16294T 16296T 73G 263G 315.1C	3	(3)
V:1	16298C 263G 315.1C	3	(2) (1)
-	16069T 16126C 16145A 16231C 16261T 73G 150T 152C 195C 215G 263G 295T 310.1T 315.1C 319C	6	(5) (1)
-	16126C 16172C 16266T 16292T 16294T 73G 152C 263G 315.1C	2	(2)
-	16126C 16187T 16294T 16296T 16304C 73G 151T 263G 315.1C	2	(2)
-	16147A 16154C 16172C 16223T 16248T 16320T 16355T 73G 152C 199C 204C 263G 315.1C	2	(2)
-	16256T 16270T 73G 263G 315.1C	4	(4)
-	16293G 16311C 195C 263G 315.1C	2	(1) (1)
-	16294T 16304C 198T 263G 315.1C	2	(2)
-	16362C 239C 263G 315.1C	2	(2)
Also targets J:1 HV type		**94**	

Table 33. Breakdown of 94 non-unique HV-type samples by Multiplex G.
The application of Multiplex G to 25 non-unique HV-types resulted in 35 different HV/SNP-types including 11 unique types. This panel targets J:1, J:3, and T:1 HV-types, highlighted in yellow.

HV Type	HV Profile	Total	Breakdown
H:1	263G 315.1C	17	(15) (2)
H:1 + 16189	16189C 263G 315.1C	6	(6)
H:1 + 16311	16311C 263G 315.1C	3	(2) (1)
H:1 + 16355	16355T 263G 315.1C	2	(2)
H:1 + 93	93G 263G 315.1C	3	(2) (1)
H:2	152C 263G 315.1C	7	(6) (1)
H:3	16129A 263G 315.1C	3	(2) (1)
H:5	16304C 263G 315.1C	6	(6)
H:7	16162G 16209C 73G 263G 315.1C	5	(5)
J:3	16069T 16126C 73G 185A 188G 228A 263G 295T 315.1C	2	(2)
J:3 + 185	16069T 16126C 73G 146C 185A 188G 228A 263G 295T 315.1C	2	(2)
K:1	16224C 16311C 73G 146C 152C 263G 315.1C	2	(1) (1)
K:3 + 195	16224C 16311C 73G 195C 263G 315.1C	2	(2)
T:1	16126C 16294T 16296T 16304C 73G 263G 315.1C	2	(2)
T:2	16126C 16163G 16186T 16189C 16294T 73G 152C 195C 263G 315.1C	4	(4)
T:3	16126C 16294T 16296T 73G 263G 315.1C	3	(3)
V:1	16298C 263G 315.1C	3	(2) (1)
-	16069T 16126C 16145A 16231C 16261T 73G 150T 152C 195C 215G 263G 295T 310.1T 315.1C 319C	6	(5) (1)
-	16126C 16172C 16266T 16292T 16294T 73G 152C 263G 315.1C	2	(2)
-	16126C 16187T 16294T 16296T 16304C 73G 151T 263G 315.1C	2	(2)
-	16147A 16154C 16172C 16223T 16248T 16320T 16355T 73G 152C 199C 204C 263G 315.1C	2	(2)
-	16256T 16270T 73G 263G 315.1C	4	(4)
-	16293G 16311C 195C 263G 315.1C	2	(1) (1)
-	16294T 16304C 198T 263G 315.1C	2	(2)
-	16362C 239C 263G 315.1C	2	(2)
		94	

Table 34. Breakdown of 94 non-unique HV-type samples by Multiplex H.
The application of Multiplex H to 25 non-unique HV-types resulted in 34 different HV/SNP-types including 10 unique types. This panel targets the K:1 HV-type, highlighted in yellow.

HV Type	HV Profile	Total	Breakdown
H:1	263G 315.1C	17	(7) (4) (1) (1) (1) (1) (1) (1)
H:1 + 16189	16189C 263G 315.1C	5	(3) (1) (1)
H:1 + 16311	16311C 263G 315.1C	3	(1) (1) (1)
H:1 + 16355	16355T 263G 315.1C	2	(2)
H:1 + 93	93G 263G 315.1C	3	(1) (1) (1)
H:2	152C 263G 315.1C	7	(3) (2) (1) (1)
H:3	16129A 263G 315.1C	3	(1) (1) (1)
H:5	16304C 263G 315.1C	6	(6)
H:7	16162G 16209C 73G 263G 315.1C	5	(5)
J:3	16069T 16126C 73G 185A 188G 228A 263G 295T 315.1C	2	(2)
J:3 + 185	16069T 16126C 73G 146C 185A 188G 228A 263G 295T 315.1C	2	(2)
K:1	16224C 16311C 73G 146C 152C 263G 315.1C	2	(2)
K:3 + 195	16224C 16311C 73G 195C 263G 315.1C	2	(2)
T:1	16126C 16294T 16296T 16304C 73G 263G 315.1C	1	(1)
T:2	16126C 16163G 16186T 16189C 16294T 73G 152C 195C 263G 315.1C	2	(2)
T:3	16126C 16294T 16296T 73G 263G 315.1C	1	(1)
V:1	16298C 263G 315.1C	3	(2) (1)
-	16069T 16126C 16145A 16231C 16261T 73G 150T 152C 195C 215G 263G 295T 310.1T 315.1C 319C	6	(5) (1)
-	16126C 16172C 16266T 16292T 16294T 73G 152C 263G 315.1C	1	(1)
-	16126C 16187T 16294T 16296T 16304C 73G 151T 263G 315.1C	1	(1)
-	16147A 16154C 16172C 16223T 16248T 16320T 16355T 73G 152C 199C 204C 263G 315.1C	2	(2)
-	16256T 16270T 73G 263G 315.1C	4	(4)
-	16293G 16311C 195C 263G 315.1C	2	(1) (1)
-	16294T 16304C 198T 263G 315.1C	2	(2)
-	16362C 239C 263G 315.1C	2	(2)
		86	**8 null alleles***

Table 35. Breakdown of 86 non-unique HV-type samples by Multiplexes A and F.
The application of Multiplexes A and F to 21 non-unique HV-types resulted in 42 different HV/SNP-types including 23 unique types. Multiplexes A and F are the primary and secondary multiplexes for HV-type H:1, highlighted in orange.
*Samples with null alleles (8) were not included for data analysis.

HV Type	HV Profile	Total	Breakdown
H:1	263G 315.1C	17	(6) (4) (2) (1) (1) (1) (1) (1)
H:1 + 16189	16189C 263G 315.1C	6	(4) (1) (1)
H:1 + 16311	16311C 263G 315.1C	3	(2) (1)
H:1 + 16355	16355T 263G 315.1C	2	(2)
H:1 + 93	93G 263G 315.1C	3	(1) (1) (1)
H:2	152C 263G 315.1C	7	(2) (2) (1) (1) (1)
H:3	16129A 263G 315.1C	3	(1) (1) (1)
H:5	16304C 263G 315.1C	6	(6)
H:7	16162G 16209C 73G 263G 315.1C	5	(5)
J:3	16069T 16126C 73G 185A 188G 228A 263G 295T 315.1C	2	(2)
J:3 + 185	16069T 16126C 73G 146C 185A 188G 228A 263G 295T 315.1C	2	(2)
K:1	16224C 16311C 73G 146C 152C 263G 315.1C	2	(2)
K:3 + 195	16224C 16311C 73G 195C 263G 315.1C	2	(2)
T:1	16126C 16294T 16296T 16304C 73G 263G 315.1C	1	(1)
T:2	16126C 16163G 16186T 16189C 16294T 73G 152C 195C 263G 315.1C	2	(2)
T:3	16126C 16294T 16296T 73G 263G 315.1C	1	(1)
V:1	16298C 263G 315.1C	3	(2) (1)
-	16069T 16126C 16145A 16231C 16261T 73G 150T 152C 195C 215G 263G 295T 310.1T 315.1C 319C	6	(5) (1)
-	16126C 16172C 16266T 16292T 16294T 73G 152C 263G 315.1C	1	(1)
-	16126C 16187T 16294T 16296T 16304C 73G 151T 263G 315.1C	1	(1)
-	16147A 16154C 16172C 16223T 16248T 16320T 16355T 73G 152C 199C 204C 263G 315.1C	2	(2)
-	16256T 16270T 73G 263G 315.1C	3	(3)
-	16293G 16311C 195C 263G 315.1C	2	(1) (1)
-	16294T 16304C 198T 263G 315.1C	2	(2)
-	16362C 239C 263G 315.1C	2	(2)
		86	**8 null alleles***

Table 36. Breakdown of 86 non-unique HV-type samples by Multiplexes B and F.
The application of Multiplexes B and F to 21 non-unique HV-types resulted in 42 different HV/SNP-types including 21 unique types. Multiplexes B and F are the primary and secondary multiplexes for HV-types H:2 and H:3, highlighted in orange.
*Samples with null alleles (8) were not included for data analysis.

HV Type	HV Profile	Total	Breakdown
H:1	263G 315.1C	16	(11) (2) (1) (1) (1)
H:1 + 16189	16189C 263G 315.1C	6	(5) (1)
H:1 + 16311	16311C 263G 315.1C	3	(2) (1)
H:1 + 16355	16355T 263G 315.1C	2	(2)
H:1 + 93	93G 263G 315.1C	3	(2) (1)
H:2	152C 263G 315.1C	7	(5) (1) (1)
H:3	16129A 263G 315.1C	3	(1) (1) (1)
H:5	16304C 263G 315.1C	5	(3) (2)
H:7	16162G 16209C 73G 263G 315.1C	5	(5)
J:3	16069T 16126C 73G 185A 188G 228A 263G 295T 315.1C	2	(2)
J:3 + 185	16069T 16126C 73G 146C 185A 188G 228A 263G 295T 315.1C	2	(2)
K:1	16224C 16311C 73G 146C 152C 263G 315.1C	2	(2)
K:3 + 195	16224C 16311C 73G 195C 263G 315.1C	2	(1) (1)
T:1	16126C 16294T 16296T 16304C 73G 263G 315.1C	1	(1)
T:2	16126C 16163G 16186T 16189C 16294T 73G 152C 195C 263G 315.1C	2	(2)
T:3	16126C 16294T 16296T 73G 263G 315.1C	1	(1)
V:1	16298C 263G 315.1C	3	(2) (1)
-	16069T 16126C 16145A 16231C 16261T 73G 150T 152C 195C 215G 263G 295T 310.1T 315.1C 319C	6	(5) (1)
-	16126C 16172C 16266T 16292T 16294T 73G 152C 263G 315.1C	1	(1)
-	16126C 16187T 16294T 16296T 16304C 73G 151T 263G 315.1C	1	(1)
-	16147A 16154C 16172C 16223T 16248T 16320T 16355T 73G 152C 199C 204C 263G 315.1C	2	(2)
-	16256T 16270T 73G 263G 315.1C	4	(4)
-	16293G 16311C 195C 263G 315.1C	2	(1) (1)
-	16294T 16304C 198T 263G 315.1C	2	(2)
-	16362C 239C 263G 315.1C	2	(2)
		85	**9 not included***

Table 37: Breakdown of 86 non-unique HV-type samples by Multiplexes C and F.
The application of Multiplexes C and F to 21 non-unique HV-types resulted in 37 different HV/SNP-types including 17 unique types. Multiplexes C and F are the primary and secondary multiplexes for HV-type V:1, highlighted in orange.
*Samples with null alleles (7) or partial profiles (2) were not included for data analysis.

Conclusions and Future Directions

Accomplishing the goals of this project has resulted in the development of a refined automated system, relying on robotics and bioinformatics, for high throughput mtDNA sequencing [we note that infrastructure, instrumentation, and software development costs were independent of NIJ funding]. In order to identify samples matching common HV types for entire mtGenome sequencing, new databases of entire control region sequences were established for an African American sample, and four regional Hispanic databases. Significant differences in haplogroup and haplotype representation among the regional Hispanic databases suggests that additional attention should be paid to the issue of population subdivision, especially involving inherently heterogeneous population categories such as "Hispanics."

This work represents the first comprehensive demonstration that variation in the mtDNA coding region can provide substantial levels forensic discrimination, over and above the variation present in the hypervariable regions of the control region. Admittedly, knowledge of relative rates of sequence evolution in the control region versus the coding region, and their relative size, made that a bit of a foregone conclusion. However, what was very much an open question at the outset of this project was whether the additional variation was distributed in such a manner that it could be practically accessed within the context of real-world forensic mtDNA testing. It is clearly—for the immediately foreseeable future—unrealistic to sequence the entire mtDNA genome for mtDNA casework samples. Normally, retrieving sequence from the control region leaves only a small amount of evidentiary extract left, so that only small portions of the coding region could be expected to be recovered. As a result of this project, we moreover know that the distribution of variation in the coding region is such that sequencing any practically attainable number of small fragments, using the same general approach for any case in question, will rarely

provide additional forensic discrimination (this is presented in detail in Coble *et al*., 2005). Using high-throughput mtGenome sequencing of common HV types, we have identified SNP sites that are highly useful for providing additional discrimination for those HV-types (and their near relatives) that will most often be involved in cases where additional discrimination is needed. This approach was particularly useful in the case of common HV types in the US Caucasian population. Smaller available sample sizes for whole mtGenome sequencing somewhat limited the ability to identify a large number of discriminatory sites in African American and Hispanic common HV types; and in some Hispanic types we simply observed low levels of variation (a limitation for any approach). Nonetheless, for all population groups studied, useful SNP panels were identified. For Caucasians, multiplex ASPE assays were designed, optimized, and demonstrated to work effectively on degraded samples.

Based on the results reported from this project, the primary use of the SNP panels is to distinguish among multiple matching samples, refining inclusion and exclusion from within a closed set of compared sequences. An alternative desirable application would be to simply increase the evidentiary significance of a single match involving one of the common types. For example, instead of an HV type that matches 7% of the population, combining SNP data might result in a SNP/HV type present in 0.5% of the population, increasing the significance of the match. However, for this purpose, large population databases that combine SNPs and HV sequences are necessary. AFDIL is presently planning to add SNP panel typing to current family reference databasing, which will quickly increase to large numbers, and will make that data available as the project is implemented. In this case, it would be only be necessary to apply the SNP panels to those population samples for which they are relevant. However, in addition to having the databases available, wide use of this approach will also require labs to have access to

population search engines that accommodate the SNP data (AFDIL's LIMS system already incorporates this, but there are no current plans for the search function to be openly available). As the field progresses toward use of discriminatory SNPs, modifications to the publicly available searchable databases such as CODISmt and EMPOP would be highly desirable.

This project is undoubtedly only the beginning of identifying methods for obtaining data ranging over the entire mtDNA genome, to powerfully increase forensic discrimination. The ASPE assays we designed give forensic laboratories this ability, now, for many of the circumstances where need will arise. However, the implementation of the ASPE assay, while simple in practice, carries a high burden of effort regarding primer quality control, and we suspect this will be a limitation to very widespread adoption of this technique. The field of forensic DNA analysis is aggressively developing new SNP typing methods, and maximum benefit to the field would be in commercial availability of a (fast, cheap, and easy) SNP typing kit, where primary QC issues are removed from individual laboratory. Fortunately, interest has been expressed by a number of companies in this regard. We note that, with DNA sequencing becoming ever less expensive and time consuming, that it may be feasible to sequences the entire mtGenome from reference samples, in the not-distant future. This could then direct the investigation of degraded evidentiary samples to particular portions of the mtGenome that could be targeted for a high probability of excluding unrelated individuals. In any case, the primary mtGenome data generated for this study will be a highly useful reference for guiding the development of new approaches.

References:

Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, Howell N (1999) Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nature Genetics* **23(2):** 147.

Bandelt, H.-J., Quintana-Murci, L., Salas, A., and Macaulay, V. (2002) "The Fingerprint of Phantom Mutations in Mitochondrial DNA Data." *Am. J. Hum. Genet*. 71:1150-1160.

Bandelt, H.-J., Salas, A., and Lutz-Bonengel, S. (2004a) "Artificial recombination in forensic mtDNA population databases." *Int. J. Legal Med*. 118:267-273.

Bandelt, H.-J., Salas, A., and Bravi, C. (2004b) "Problems in FBI mtDNA Database." *Science* 305:1402-1404.

Brandstätter, A., Peterson, C.T., Irwin, J.A., Mpoke, S., Koech, D.K., Parson W., and Parsons, T.J. (2004) "Mitochondria DNA Control Region Sequences from Nairobi (Kenya): Inferring Phylogenetic Parameters for the Establishment of a Forensic Database." *Int. J. Legal Med.* **118:**294-306.

Brandstätter, A., Parsons, T.J., Niederstätter, H., and Parson, W. (2003). Rapid Screening of mtDNA Coding Region SNPs for the Identification of west European Caucasian Haplogroups. ***Int. J. Legal Med.*** **117**:291-298.

Brandstätter, A., Parsons, T.J., and Parson, W. (2004). MtDNA Coding Region SNPs for Rapid Screening and Haplogroup Identification of Forensic Samples. International Congress Series 1261, *Elsevier Press*, pp 422-424.

Budowle, B., Gyllensten, U., Chakraborty, R., and Allen, M. (2005). Forensic analysis of the mitochondrial coding region and association to disease. *Int. J. Legal Med*. Electronic pre-print (on-line first) 10.1007/s00414-005-0543-y.

Coble, M.D. (2004) The identification of single nucleotide polymorphisms in the entire mitochondrial genome to increase the forensic discrimination of common HV1/HV2 types in the Caucasian population, Ph.D. Dissertation, George Washington University, 206 pp. (http://www.cstl.nist.gov/div831/strbase/pub_pres/Coble2004dis.pdf).

Coble, M. D., Just, R. S., O'Callaghan, J. E., Letmanyi, I. H., Peterson, C. T., and Parsons, T. J. (2004) Single nucleotide polymorphisms over the entire mtDNA genome that increase the forensic power of mtDNA testing in Caucasians. Int. J. Leg. Med. 118: 137-146.

Coble, M.D., Vallone, P.M., Just, R.S., Diegoli, T.M., Smith, B.C., and Parsons, T.J. (2005). Effective strategies for forensic analysis in the mtDNA coding region. *Int. J. Legal Med*. Electronic pre-print (on-line first) DOI: 10.1007/s00414-005-0044-z.

Holland MM and Parsons TJ (1999) Mitochondrial DNA Sequence Analysis - Validation and Use for Forensic Casework. *Forensic Science Review* **11(1):** 21-50.

Just, R.S., Irwin, J.A., O'Callaghan, J.E., Saunier, J.L., Coble, M.D., Vallone, P.M., Butler, J.M., and Parsons, T.J. (2004). Research toward increasing the utility of mtDNA in forensic identifications. *For. Sci. Int.* 146S s147-s149.

Grignani, P., Peloso, G., Achilli, A., Turchi, C., Tagliabracci, A., Alu, M., Beduschi, G., Ricci, U., Giunti, L., Robino, C., Gino, S., Previdere, C. (2005). Subtyping Haplogroup H by SNaPShot minisequencing and its application in forensic individual identification. *Int. J. Legal Med*. Electronic pre-print (on-line first) DOI: 10.1007/s00414-005-0059-5.

Niederstätter, H., Coble, M.D., Grubwieser, P., Parsons, T.J., and Parson, W. (2005) Characterization of mtDNA SNP typing and mixture ratio assessment with simultaneous real-

time PCR quantification of both allelic states. *Int. J. Legal Med*. Electronic pre-print (on-line first) DOI: 10.1007/s00414-005-0024-3

Rozen, S. and Skaletsky, H.J. (2000). Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (eds) *Bioinformatics Methods and Protocols: Methods in Molecular Biology*. Humana Press, Totowa, NJ, pp 365-386
Source code available at http://fokker.wi.mit.edu/primer3/.

Vallone, P.M. and Butler, J.M. (2004) AutoDimer: a screening tool for primer-dimer and hairpin structures. Biotechniques 37: 226-231.

Vallone, P. M., Just, R. S., Coble, M. D., Butler, J. M., and Parsons, T. J. (2004) A multiplex allele-specific primer extension assay for forensically informative SNPs distributed throughout the mitochondrial genome. Int. J. Leg. Med. 118**:** 147-157

Appendix I: Control Region Amplification and Sequencing Primer Sequences

Primer	Type	Sequence	Notes
F15971	Amp	TTAACTCCACCATTAGCACC	also used for sequencing
R599	Amp	TTGAGGAGGTAAGCTACATA	also used for sequencing
F16190	Sequencing	CCCCATGCTTACAAGCAAGT	
F16450	Sequencing	GCTCCGGGCCCATAACACTTG	
F34	Sequencing	GGGAGCTCTCCATGCATTTGGTA	
F314	Sequencing	CCGCTTCTGGCCACAGCACT	
R285	Sequencing	GTTATGATGTCTGTGTGGAA	
R16400	Sequencing	GTCAAGGGACCCCTATCTGA	
F15971	Sequencing	TTAACTCCACCATTAGCACC	
R599	Sequencing	TTGAGGAGGTAAGCTACATA	

Appendix II: Entire mtDNA Genome Amplification and Sequencing Primer Sequences

Amp	Primer	Type	Sequence	Notes
1	F361	Amp	ACAAAGAACCCTAACACCAGC	also used for sequencing
1	R2216	Amp	TGTTGAGCTTGAACGCTTTC	also used for sequencing
2	F1993	Amp	AAACCTACCGAGCCTGGTG	also used for sequencing
2	R3557	Amp	AGAAGAGCGATGGTGAGAGC	also used for sequencing
3	F3441	Amp	ACTACAACCCTTCGCTGACG	also used for sequencing
3	R4982	Amp	GTTTAATCCACCTCAACTGCC	also used for sequencing
4	F4797	Amp	CCCTTTCACTTCTGAGTCCCAG	also used for sequencing
4	R6526	Amp	ATAGTGATGCCAGCAGCTAGG	also used for sequencing
5	F6426	Amp	GCCATAACCCAATACCAAACG	also used for sequencing
5	R8311	Amp	AAGTTAGCTTTACAGTGGGCTCTAG	also used for sequencing
6	F8164	Amp	CGGTCAATGCTCTGAAATCTGTG	also used for sequencing
6	R9848	Amp	GAAAGTTGAGCCAATAATGACG	also used for sequencing
7	F9754	Amp	AGTCTCCCTTCACCATTTCCG	also used for sequencing
7	R11600	Amp	CTGTTTGTCGTAGGCAGATGG	also used for sequencing
8	F11403	Amp	GACTCCCTAAAGCCCATGTCG	also used for sequencing
8	R13123	Amp	AGCGGATGAGTAAGAAGATTCC	also used for sequencing
9	F12793	Amp	TTGCTCATCAGTTGATGATACG	also used for sequencing
9	R14388	Amp	TTAGCGATGGAGGTAGGATTGG	also used for sequencing
10	F14189	Amp	ACAAACAATGGTCAACCAGTAAC	also used for sequencing
10	R15396	Amp	TTATCGGAATGGGAGGTGATTC	also used for sequencing
11	F15260	Amp	AGTCCCACCCTCACACGATTC	also used for sequencing
11	R16084	Amp	CGGTTGTTGATGGGTGAGTC	also used for sequencing
12	F15878	Amp	AAATGGGCCTGTCCTTGTAG	also used for sequencing
12	R649	Amp	TTTGTTTATGGGGTGATGTGA	also used for sequencing
1	F361	Sequencing	ACAAAGAACCCTAACACCAGC	also used for amplification

1	F1234	Sequencing	CTCACCACCTCTTGCTCAGC	
1	F1657	Sequencing	CTTGACCGCTCTGAGCTAAAC	
1	F756	Sequencing	CATCAAGCACGCAGCAATG	
1	F873	Sequencing	GGTTGGTCAATTTCGTGCCAG	
1	R1425	Sequencing	AATCCACCTTCGACCCTTAAG	
1	R1769	Sequencing	GCCAGGTTTCAATTTCTATCG	
1	R2216	Sequencing	TGTTGAGCTTGAACGCTTTC	also used for amplification
1	R921	Sequencing	ACTTGGGTTAATCGTGTGACC	
2	F1993	Sequencing	AAACCTACCGAGCCTGGTG	also used for amplification
2	F2105	Sequencing	GAGGAACAGCTCTTTGGACAC	
2	F2417	Sequencing	CACTGTCAACCCAACACAGG	
2	F2834	Sequencing	CCCAACCTCCGAGCAGTACATG	
2	F3234	Sequencing	AGATGGCAGAGCCCGGTAATC	
2	R2660	Sequencing	AGAGACAGCTGAACCCTCGTG	
2	R3006	Sequencing	ATGTCCTGATCCAACATCGAG	
2	R3557	Sequencing	AGAAGAGCGATGGTGAGAGC	also used for amplification
3	F3441	Sequencing	ACTACAACCCTTCGCTGACG	also used for amplification
3	F3931	Sequencing	TCAGGCTTCAACATCGAATACG	
3	F4392	Sequencing	CCCATCCTAAAGTAAGGTCAGC	
3	R3940	Sequencing	TGAAGCCTGAGACTAGTTCGG	
3	R4162	Sequencing	TGAGTTGGTCGTAGCGGAATC	
3	R4982	Sequencing	GTTTAATCCACCTCAACTGCC	also used for amplification
4	F4797	Sequencing	CCCTTTCACTTCTGAGTCCCAG	also used for amplification
4	F5318	Sequencing	CACCATCACCCTCCTTAACC	
4	F5700	Sequencing	TAAGCACCCTAATCAACTGGC	
4	F6242	Sequencing	CGCATCTGCTATAGTGGAGG	
4	R5882	Sequencing	GCTGAGTGAAGCATTGGACTG	
4	R6526	Sequencing	ATAGTGATGCCAGCAGCTAGG	also used for amplification
5	F6426	Sequencing	GCCATAACCCAATACCAAACG	also used for amplification
5	F7075	Sequencing	GAGGCTTCATTCACTGATTTCC	
5	F7645	Sequencing	TATCACCTTTCATGATCACGC	
5	R7255	Sequencing	TTTCATGTGGTGTATGCATCG	
5	R7792	Sequencing	GGGCAGGATAGTTCAGACGG	
5	R8311	Sequencing	AAGTTAGCTTTACAGTGGGCTCTAG	also used for amplification
6	F8164	Sequencing	CGGTCAATGCTCTGAAATCTGTG	also used for amplification
6	F8539	Sequencing	CTGTTCGCTTCATTCATTGCC	
6	F8903	Sequencing	CCCACTTCTTACCACAAGGC	
6	F9309	Sequencing	TTTCACTTCCACTCCATAACGC	
6	R9059	Sequencing	GTGGCGCTTCCAATTAGGTG	
6	R9403	Sequencing	GTGCTTTCTCGTGTTACATCG	
6	R9848	Sequencing	GAAAGTTGAGCCAATAATGACG	also used for amplification
7	F10127	Sequencing	ACTACCACAACTCAACGGCTAC	
7	F10386	Sequencing	GGATTAGACTGAACCGAATTGG	
7	F11001	Sequencing	AACGCCACTTATCCAGTGAACC	
7	F9754	Sequencing	AGTCTCCCTTCACCATTTCCG	also used for amplification
7	R10275	Sequencing	AAAGGAGGGCAATTTCTAGATC	
7	R10556	Sequencing	GGAGGATATGAGGTGTGAGGG	
7	R11166	Sequencing	CATCGGGTGATGATAGCCAAG	

7	R11267	Sequencing	TGTTGTGAGTGTAAATTAGTGCG	
7	R11600	Sequencing	CTGTTTGTCGTAGGCAGATGG	also used for amplification
8	F11403	Sequencing	GACTCCCTAAAGCCCATGTCG	also used for amplification
8	F11908	Sequencing	AACCACGTTCTCCTGATCAAA	
8	F12357	Sequencing	AACCACCCTAACCCTGACTTCC	
8	F12601	Sequencing	TTCATCCCTGTAGCATTGTTCG	
8	R12876	Sequencing	GATATCGCCGATACGGTTG	
8	R13123	Sequencing	AGCGGATGAGTAAGAAGATTCC	also used for amplification
9	F12793	Sequencing	TTGCTCATCAGTTGATGATACG	also used for amplification
9	F13188	Sequencing	CACTCTGTTCGCAGCAGTATG	
9	F13518	Sequencing	CATCATCGAAACCGCAAAC	
9	F13899	Sequencing	TTTCTCCAACATACTCGGATTC	
9	R13343	Sequencing	TTGAAGAAGGCGTGGGTACAG	
9	R13611		TCGAGTGCTATAGGCGCTTGTC	
9	R13935		TGTGATGCTAGGGTAGAATCCG	
9	R14388	Sequencing	TTAGCGATGGAGGTAGGATTGG	also used for amplification
	F14189	Sequencing	ACAAACAATGGTCAACCAGTAAC	also used for amplification
10	F14470		TCCAAAGACAACCATCATTCC	
10	F14909	Sequencing	TACTCACCAGACGCCTCAACCG	
10	R14996	Sequencing	CGTGAAGGTAGCGGATGATTC	
10	R15396	Sequencing	TTATCGGAATGGGAGGTGATTC	also used for amplification
11	F15260	Sequencing	AGTCCCACCCTCACACGATTC	also used for amplification
	F15574	Sequencing	CGCCTACACAATTCTCCGATC	
11	R15774		ACTGGTTGTCCTCCGATTCAGG	
11	R16084	Sequencing	CGGTTGTTGATGGGTGAGTC	also used for amplification
	F15878	Sequencing	AAATGGGCCTGTCCTTGTAG	also used for amplification
12	F16190		CCCCATGCTTACAAGCAAGT	
12	F16450	Sequencing	GCTCCGGGCCCATAACACTTG	
	F314	Sequencing	CCGCTTCTGGCCACAGCACT	
12	R16175		TGGATTGGGTTTTTATGTA	
12	R16400	Sequencing	GTCAAGGGACCCCTATCTGA	
12	R285	Sequencing	GTTATGATGTCTGTGTG AA	
12	R649	Sequencing	TTTGTTTATGGGGTGATGTGA	also used for amplification
1	F1587	Sequencing	TGCACTTGGACGAACCAGAG	alternate primer
1	F370	Sequencing	CCTAACACCAGCCTAACCAG	alternate primer
1	F629	Sequencing	TCACATCACCCCATAAACAA	alternate primer
1	F895	Sequencing	CACCGCGGTCACACGAT	alternate primer
1	F1095	Sequencing	TAGCCCTAAACCTCAACAGT	alternate primer
1	R1309	Sequencing	TACTTGCGCTTACTTTGT	alternate primer
1	F1372	Sequencing	CTACCCCAGAAAACTACGAT	alternate primer
1	F1594	Sequencing	GGACGAACCAGAGTGTAG	alternate primer
1	F1729		AATGGTTTGGCTAAGGTTGT	alternate primer
1	F1892	Sequencing	AAGCTAAGACCCCCGAAACC	alternate primer
	R1973	Sequencing	CTTCCCACTATTTTGCTACA	alternate primer
1	R2194		AATTGGTGGCTGCTTTTAGG	alternate primer
1	R965	Sequencing	GGGGAGGGGGTGATCTAAAACA	alternate primer
2	F2972	Sequencing	ATAGGGTTTACGACCTCGATG	
3	F3635	Sequencing	GCCTAGCCGTTTACTCAATCC	

3	F4183	Sequencing	TTTCTACCACTCACCCTAGCATTAC	alternate primer
	F4447	Sequencing	TTGGTTATACCCTTCCCGTAC	
3	R4728	Sequencing	TTATGGTTCATTGTCCGGAGAG	alternate primer
4	F4976	Sequencing	ATTAAACCAGACCCAGCTACG	alternate primer
4	F5999	Sequencing	TCTAAGCCTCCTTATTCGAGC	alternate primer
4	R4983	Sequencing	GGTTTAATCCACCTCAACTGCC	alternate primer
4	R5553		AGGGCTTTGAAGGCTCTTG	alternate primer
4	R6262	Sequencing		alternate primer
4	F5529	Sequencing	ACAGACCAAGAGCCTTCAAA	
4	R5994	Sequencing	TGCCTAGGACTCCAGCTCAT	alternate primer
5	F6744	Sequencing	GGCTTCCTAGGGTTTATCGTG	alternate primer
5	F7215	Sequencing	CGACGTTACTCGGACTACCC	alternate primer
5		Sequencing	TGAACCTACGAGTACACCGACTAC	alternate primer
5	R7030	Sequencing	TGGGCTACAACGTAGTACGTG	alternate primer
5	R8215	Sequencing		alternate primer
6	R8669	Sequencing	CATTGTTGGGTGGTGATTAGTCG	
6	F9449	Sequencing	CGGGATAATCCTATTTATTACCTCAG	alternate primer
	R9995	Sequencing	AGAGTAAGACCCTCATCAATAGATGG	alternate primer
7	F10157	Sequencing	ATCCACCCCTTACGAGTGC	alternate primer
7	F10773	Sequencing	TCGTCCCAACAATTATATTACTACCA	alternate primer
7	F10704	Sequencing	GTCTCAATCTCCAACACATATGG	alternate primer
7	R10524	Sequencing	TTCCTAGAAGTGAGATGGTAAATGC	alternate primer
7	R10715	Sequencing	CGTAGTCTAGGCCATATGTGTTG	alternate primer
8	F11760		ACGAACGCACTCACAGTCG	alternate primer
8		Sequencing	TGCTAGTAACCACGTTCTGGTG	alternate primer
8	R11927	Sequencing	TTGATCAGGAGAACGTGGTTAC	alternate primer
	R12189	Sequencing	AAGCCTCTGTTGTCAGATTCAC	alternate primer
8	F12236	Sequencing	GCCCCCATGTCTAACAACAT	alternate primer
8	R12770	Sequencing	TCTCAGCCGATGAACAGTTG	alternate primer
9		Sequencing	GAAGCCTATTCGCAGGATTTC	alternate primer
9	R14118	Sequencing	TGGGAAGAAGAAAGAGAGGAAG	alternate primer
10	R14926	Sequencing	TGAGGCGTCTGGTGAGTAGTGC	alternate primer
11	F15699	Sequencing		alternate primer
11	R16042	Sequencing	CTGCTTCCCCATGAAAGAAC	alternate primer
12	F15971	Sequencing	TTAACTCCACCATTAGCACC	alternate primer
12	F16097	Sequencing	TACATTACTGCCAGCCACCATG	alternate primer
	R16451	Sequencing	GCGAGGAGAGTAGCACTCTTG	alternate primer
12	R336	Sequencing	TTAAGTGCTGTGGCCAGAAG	alternate primer
12	R419	Sequencing	TGCATACCGCCAAAAGATAA	alternate primer
12		Sequencing	TGAGATTAGTAGTATGGGAG	alternate primer

Appendix III. Thermal Cycling conditions and 3100 Run Parameters for ASPE

Amplification

Hold	95°C for 5 minutes
3 cycles	95°C for 30 seconds 50°C for 55 seconds 72°C for 30 seconds
19 cycles	95°C for 30 seconds 50°C for 55 seconds* 72°C for 30 seconds
9 cycles`	95°C for 30 seconds 55°C for 55 seconds 72°C for 30 seconds
Hold	72°C for 7 minutes
Hold	4°C for infinity

EXO/SAP (post-PCR)

Hold	37°C for 90 minutes
Hold	80°C for 20 minutes
Hold	4°C for infinity

Extension (SNaPshot)

Hold	96°C for 10 seconds
25 cycles	96°C for 10 seconds 50°C for 5 seconds 60°C for 30 seconds
Hold	4°C for infinity

SAP (post-extension)

Hold	37°C for 30 minutes
Hold	80°C for 20 minutes
Hold	4°C for infinity

3100 Genetic Analyzer Run Parameters

#	Parameter Name	Value
1	Run Temp	60C
2	Cap Fill Volume	184st
3	Current Tolerance	100uA
4	Run Current	100uA
5	Voltage Tolerance	0.6kV
6	Pre Run Voltage	15kV
7	Pre Run Time	60s
8	Injection Voltage	1kV
9	Injection Time	13s
10	Run Voltage	15V
11	Number Of Steps	10nk
12	Voultage step Interval	20s
13	Data Delay Time	200s
14	Run Time	1200s

Appendix IV: ASPE Amplification and Extension Primer Sequences

MP	Site	Type	Primer Name	Sequence (5' -> 3')	Concentration (uM)
A	477	AMP	A-477-F(A)	CTTTTGGCGGTATGCACTTT	4.55
A	477	AMP	A-477-R(A2)	GTTTGGTTGGTTCGGGGTAT	4.55
A	3010	AMP	A-3010-F(A)	GCGCAATCCTATTCTAGAGTCC	4.55
A	3010	AMP	A-3010-R(A)	TCACGTAGGACTTTAATCGTTGA	4.55
A	4580	AMP	A-4580-F(A)	TCTTTGCAGGCACACTCATC	4.55
A	4580	AMP	A-4580-R(A)	GCAGCTTCTGTGGAACGAG	4.55
A	4793	AMP	A-4793-F(F2)	TGAACCATAACCAATACTACCAATC	4.55
A	4793	AMP	A-4793-R(A)	ATGTCAGAGGGGTGCCTTG	4.55
A	5004	AMP	A-5004-F(A)	TCCATCATAGCAGGCAGTTG	4.55
A	5004	AMP	A-5004-R(A)	TGGTTATGTTAGGGTTGTACGG	4.55
A	7028	AMP	A-7028-F(A)	GGCCTGACTGGCATTGTATT	4.55
A	7028	AMP	A-7028-R(A)	AAGCCTCCTATGATGGCAAA	4.55
A	7202	AMP	A-7202-F(A)	ACGCCAAAATCCATTTCACT	4.55
A	7202	AMP	A-7202-R(A)	TTCATGTGGTGTATGCATCG	4.55
A	10211	AMP	A-10211-F(A)	ACCACAACTCAACGGCTACA	4.55
A		AMP	A-10211-R(A2)	GGGTAAAAGGAGGGCAATTT	4.55
A	12858	AMP	A-12858-F(A)		4.55
A	12858		A-12858-R(A)	TGTGGGTCTCATGAGTTGGA	4.55
A	14470	AMP	A-14470-F(A)	CAAGACCTCAACCCCTGACC	
A	14470	AMP	A-14470-R(A)	GGGGGAGGTTATATGGGTTT	4.55

A	16519	AMP	A-16519-F(A)	ACCACCATCCTCCGTGAAAT	4.55
A	16519	AMP	A-16519-R(A)	AGACCTGTGATCCATCGTGA	4.55
A	477	SNP	477-F(A)	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTCCCTCCCACTCCCATACTAC	5.00
A	3010	SNP	3010-F(A)	TGTTGGATCAGGACATCCC	1.00
A	4580	SNP	4580-R(A)	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGGTTAGAACTGGAATAAAAGCTAG	7.00
A		SNP	4793-R(A)	TTTTTCAGAAGTGAAAGGGGGC	15.00
A	5004	SNP	5004-F(A)	TTTTTTTTTTTTTTTAAACCCAGCTACGCAAAATC	10.00
A	7028	SNP	7028-F(A)	TTTTTTTTTTTTTTTTTTTGACACGTACTACGTTGTAGC	5.00
A	7202	SNP	7202-F(A)	TTTTTTTTTTTTTTTTTTTTTTTCCACAACACTTTCTCGGCCT	1.00
A	10211	SNP	10211-R(A)	TTTTTTTTTTACTAAGAAGAATTTTATGGA	15.00
A	12858	SNP	12858-F(A)	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGCAGCCATTCAAGCAATCCTATA	10.00
A	14470	SNP	14470-R(A)	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGGGAATGATGGTTGTCTTTGG	15.00
A	16519	SNP	16519-R(A)	TTTTTTTTTTTTTTTTTTTTTTTTTTTGTGGGCTATTTAGGCTTTATG	10.00
B	477	AMP	A-477-F(B)	CTTTTGGCGGTATGCACTTT	4.17
B	477	AMP	A-477-R(B2)	GTTTGGTTGGTTCGGGGTAT	4.17
B	3010	AMP		GCGCAATCCTATTCTAGAGTCC	4.17
B	3010	AMP	A-3010-R(B)	TCACGTAGGACTTTAATCGTTGA	4.17
B	3915	AMP	A-3915-F(B)	TTTATCTCCACACTAGCAGAGACC	8.33
B	3915	AMP	A-3915-R(B2)	CTGCGGCGTATTCGATGT	8.33
B	5004	AMP	A-5004-F(B)	TCCATCATAGCAGGCAGTTG	4.17
B	5004	AMP	A-5004-R(B)	TGGTTATGTTAGGGTTGTACGG	4.17
B	6776		A-6776-F(B)	TGGATACATAGGTATGGTCTGAGC	4.17
	6776		A-6776-R(B)	CGGAGGTGAAATATGCTCGT	4.17
B	8592		A-8592-F(B)	CTGTTCGCTTCATTCATTGC	4.17
B	8592		A-8592-R(B)	TGTTGATGAGATATTTGGAGGTG	4.17
B	10394		A-10394-F(B2)	TCATCCCTCTTATTAATCATCATCC	4.17
B	10394		A-10394-R(B))	TGAGTCGAAATCATTCGTTTTG	4.17
B	10754		A-10754-F(B)	TGGGCCTAGCCCTACTAGTCT	4.17
B	10754		A-10754-R(B2)		4.17
B			A-11864-F(B)	AAGGACTTCAAACTCTACTCCCACT	4.17
B	11864		A-11864-R(B)	CACAGAGAGTTCTCCCAGTAGG	
B	15340		A-15340-F(B)	TCACTTCATCTTGCCCTTCA	4.17
B	15340		A-15340-R(B)	TTTTATCGGAATGGGAGGTG	
B	16519		A-16519-F(B)	ACCACCATCCTCCGTGAAAT	4.17
B	16519		A-16519-R(B2)	GTGATCCATCGTGATGTCTT	4.17
B	477	SNP	477-R (B2)	TTTTCGGGGGTTGTATTGATGAGATT	3.80
B	3010	SNP	3010-F(B)	TGTTGGATCAGGACATCCC	0.38
B	3915		3915-F(B2)	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTCTTCGACCTTGCCGAAGG	18.98
B	5004	SNP	5004-F(B)	TTTTTTTTTTTTTTAAACCCAGCTACGCAAAATC	18.98
B			6776-F(B)	TTTTTTTTTTTTTTTTTTGTTTATCGTGTGAGCACACCA	3.80
B	8592		8592-R(B)	TTTTTTTTTTTTTTTTTTTCAATAGAGGGGGAAATAG	1.90

				AATGAT	
B	10394		10394-F(B)	TTTTTTTTTTTTTTTTTTTTTTTTTTTTAGTGACTACA AAAAGGATTAGA	18.98
B	10754		10754-F(B)	TTTTTTTTTTTTTTTTTTTTTTTTTTTTCCTAGACTAC GTACATAACCTAAACCT	7.59
B	11864		11864-F(B)	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT TAGCCTCGCTAACCTCGCC	14.23
B	15340		15340-R(B2)	TTTTTTTTTTTTTCCCGTTTCGTGCAAGAA	9.49
B	16519		16519-R(B)	GGCTTTATG	1.90
C	72	AMP	A-72-R(C)	TACTGC GAC ATA GGG TGC TC	4.44
C	72	AMP	A-72-F(C)	TCACCCTATTAACCACTCACG	4.44
C	513	AMP	A-513-F(C)	CTTTTGGCGGTATGCACTTT	4.44
C	513	AMP	A-513-R(C)	GTTTGGTTGGTTCGGGGTAT	4.44
C	4580	AMP	A-4580-F(C)	TCTTTGCAGGCACACTCATC	4.44
C	4580	AMP	A-4580-R(C)	GCAGCTTCTGTGGAACGAG	4.44
C	5250	AMP	A-5250-F(C)	AATTCCATCCACCCTCCTCT	7.78
C	5250	AMP	A-5250-R(C)	TGGGGATGATGAGGCTATTG	7.78
C	11719	AMP	A-11719-R(C)	TGAGTGCGTTCGTAGTTTGAG	4.44
C	11719	AMP	A-11719-F(C)	ATCCAAACCCCCTGAAGC	4.44
C	12438	AMP		CCATCCTTACCACCCTCGTT	4.44
C		AMP	A-12438-R(C)	TGTTGTGGGGAAGAGACTGA	4.44
	12810	AMP	A-12810-R(C)	CGGTTGTATAGGATTGCTTGAA	4.44
C	12810		A-12810-F(C)	GGCTGAGAGGGCGTAGGA	4.44
C	14770	AMP	A-14793-R(C)	GCCGAAGTTTCATCATGC	5.56
C	14770	AMP		CCAATGATATGAAAAACCATCG	5.56
C	15833/ 15884	AMP	A-15833-F(C)	GCTACCCTTTTACCATCATTGG	5.56
C	15833/ 15884	AMP	A-15833-R(C)	CCGGTTTACAAGACTGGTGTATT	5.56
C	16519	AMP	A-16519-R	AGACCTGTGATCCATCGTGA	4.44
C	16519	AMP	A-16519-F	ACCACCATCCTCCGTGAAAT	4.44
C	72	SNP		CAATGCTATCGCGTGCA[Y-Q]	3.33
C	513	SNP	513-F(C)	TTTTTTTTCCCCGCCCATCCTACCCA	0.83
C	4580	SNP	4580-R(C)	CTGGAATAAAAGCTAG	3.33
C	5250	SNP	5250-F(C)	TTTTTTTTTTTTCCCCCGCTAACCGGCTTT	35.00
C	11719	SNP	11719-R(C)	TTTTTTTTAGGCAGAATAGTAATGAGGATGTAAG	8.33
C	12438	SNP	12438-R(C)	TTTTTTTTTTTTTTTGATGCGACAATGGATTTTACA TA	3.33
C	12810	SNP	12810-R(C)	TTTTTTTTTTTTTTTTTTTTTTTTGGCATCTGCTCGG GCGTA	5.00
	14770	SNP	14770-F(C)	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTGACCCCAAT ACGCAAAACTAA	13.33
C		SNP	15833-F(C)	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTCATCCG TACTATACTTCACAACAATC	8.33
C	15884	SNP	15884-F(C)	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTAA TTGAAAACAAAATACTCAAATGG	20.00
C	16519	SNP	16519-R(B)	TTTTTTTTTTTTTTTTTTTTTTTTGTGGGCTATTTA GGCTTTATG	6.67
D	482	AMP	A-482-F(D)	CTTTTGGCGGTATGCACTTT	5.00

D	482	AMP	A-482-R(D)	GTTTGGTTGGTTCGGGGTAT	5.00
D	5198	AMP	A-5198-F(D)	CCACGACCCTACTACTATCTCG	5.00
D	5198	AMP	A-5198-R(D)	GGGCAAAAAGCCGGTTAG	5.00
D	6260	AMP	A-6260-R(D)	GGTGGGAGTAGTTCCCTGCT	5.00
D	6260		A-6260-F(D)	TTTCCCCGCATAAACAACAT	5.00
D	9548	AMP	A-9548-F(D)	TTTCTGAGCCTTTTACCACTCC	5.00
D	9548	AMP	A-9548-R(D)	GAGTGGGACTTCTAGGGGGATTT	5.00
D	9635	AMP	A-9635-R(D)	TTGAATTATTTGGTTTCGGTTG	5.00
D	9635	AMP	A-9635-F(D2)	CCGTATTACTCGCATCAGGA	5.00
D	11485	AMP	A-11485-R(D)	CGTAGGCAGATGGAGCTTGT	5.00
D	11485	AMP	A-11485-F(D)	CGCTGGGTCAATAGTACTTGC	5.00
D	11914	AMP	A-11914-F(D)	CTCGCTAACCTCGCCTTACC	5.00
D	11914	AMP	A-11914-R(D)	TGTGACTAGTATGTTGAGTCCTGTAA	5.00
D	15355	AMP	A-15355-F(D)	TCACTTCATCTTGCCCTTCA	5.00
D	15355	AMP	A-15355-R(D)	TTTTATCGGAATGGGAGGTG	5.00
D	15884	AMP	A-15833-R(C)	CCGGTTTACAAGACTGGTGTATT	5.00
	15884	AMP	A-15833-F(C)	GCTACCCTTTTACCATCATTGG	5.00
D	16368		A-16368-R(D)	ATTTCACGGAGGATGGTGGT	5.00
D		AMP	A-16368-F(D)	CATAGTACATAAAGCCATTTACCG	5.00
D	482	SNP			
D	5198	SNP	mtSNP5198-R(short)	AGAGGAGGGTGGATGGAAT	
D	6260	SNP	6260-R(D)	TTTTTTTTACCTGTTCCTGCTCCGGC	
D	9548	SNP		In Progress	
D	9635	SNP	mtSNP9635-F	TTTTTTTTTTTTTCTCGCATCAGGAGTATC	
D	11485	SNP		In Progress	
D	11914	SNP		In Progress	
D	15355	SNP	15355-R(D)	TTTTTTTTTTTTTTCCTAGGGGGTTGTTTGATCC	
D	15884	SNP	15884-F(CD)	TTTTTTTTTTTTTTAATTGAAAACAAAATACTCAAATGG	
D	16368	SNP	16368-R(D)	TTTTTTTTTTTTTTTTTTTTTTTCTATCTGAGGGGGGTCATCC	
E	4808	AMP	A-4808-R(E)	AAGAAGCAGGCCGGATGT	6.00
E	4808	AMP	A-4808-F(E)	CCAATACTACCAATCAATACTCATCA	6.00
E	5147	AMP	A-5147-R(E)	AATTAAGGGTGTTAGTCATGTTAGC	6.00
E	5147	AMP		CCGTACAACCCTAACATAACCA	6.00
E	9380	AMP	A-9380-F(E)	CATAACGCTCCTCATACTAGGC	6.00
	9380	AMP	A-9380-R(E)	GGACAGGTGGTGTGTGGTG	6.00
E	9899	AMP	A-9899-F(E)	TTTCCTCACTATCTGCTTCATCC	6.00
E	9899	AMP	A-9899-R(E)	TCAAACCACATCTACAAAATGC	6.00
	11914	AMP	A-11914-F(E2)	AGCAAGCCTCGCTAACCTC	14.00
E	11914	AMP	A-11914-R(E2)		14.00
E	15067	AMP		TTCTTTATCTGCCTCTTCCTACA	6.00
E	15067	AMP		TGCTATAGTTGCAAGCAGGAG	6.00
E		AMP	A-16519-F	ACCACCATCCTCCGTGAAAT	6.00
E	16519	AMP	A-16519-R	AGACCTGTGATCCATCGTGA	6.00
E	4808		4808-F(E)	GGAATAGCCCCCTTTCACTT	20.00
E	5147	SNP	5147-F(E2)	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTCAACTTAAACTCCAGCACCAC	6.67

E	9380	SNP		TTTTCAACACACTAACCATATACCAATGATG	1.67
E	9899	SNP	9899-F(E)	TTTTTTTTTACTAATATTTCACTTTACATCCAAACA	20.00
E	11914	SNP	11914-R(E3)	TTGATCAGGAGAA	33.33
E	15067	SNP	15067-F(E)	TTTTTTTTTTTTTTTTTTCGAGGCCTATATTACGGATCATT	6.67
E	16519	SNP	16519-R(B)	TTTTTTTTTTTTTTTTTTTTTTTTTGTGGGCTATTTAGGCTTTATG	10.00
F	64	AMP	A-64-R(F)	TACTGCGACATAGGGTGCTC	5.56
F		AMP	A-64-F(F)	ACATCACGATGGATCACAGG	5.56
F	4745	AMP	A-4745-R(F)		3.33
F	4745		A-4745-F(F)	TCCTTCTAATAGCTATCCTCTTCAA	3.33
F	10211	AMP	A-10211-F(F)	ACCACAACTCAACGGCTACA	6.67
F	10211	AMP		GGGTAAAAGGAGGGCAATTT	6.67
F		AMP	A-10394-F(F)	CCATGAGCCCTACAAACAACT	6.67
	10394	AMP	A-10394-R(F)		6.67
F	10685	AMP	A-10685-R(F2)		5.56
F	10685		A-10685-F(F2)	CCCACTCCCTCTTAGCCAAT	5.56
F	11377	AMP	A-11377-R(F2)		3.33
	11377	AMP		ACTCACTCTCACTGCCCAAGA	3.33
F		AMP	A-14470-R(F)	GGGGGAGGTTATATGGGTTT	6.67
F	14470	AMP	A-14470-F(F)	CAAGACCTCAACCCCTGACC	6.67
F	14560	AMP	A-14560-F(F)	CAACCATCATTCCCCCTAAA	5.56
F	14560	AMP	A-14560-R(F3)	TTGTGGGGTTTTCTTCTAAGC	5.56
F	14869	AMP	A-14869-R(F)	AGGCGTCTGGTGAGTAGTGC	1.11
F	14869	AMP	A-14869-F(F)	CACCCCATCCAACATCTCC	1.11
F	16390	AMP	A-16390-F(F)	GCACATTACAGTCAAATCCCTTC	5.56
F	16390	AMP	A-16390-R(F)	TAGCACTCTTGTGCGGGATA	5.56
F	64	SNP	64-F(F4)	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTCCATGCATTTGGTATTTTCGT	6.00
F	4745	SNP		GAACCATAACCAATACTACCAATCA	2.00
F	10211	SNP	10211-R(F)	TTTTTTTTTTTTTTTTTTACTAAGAAGAATTTTATGGA	22.00
F	10394	SNP	10394-F(F)	TTTTTTTTTTTTAGTGACTACAAAAAGGATTAGA	30.00
F	10685	SNP	10685-R(F2)	TTTTTTTTTTACTAGTAGGGCTAGGCCCAC	6.00
F	11377	SNP	11377-R	TTTTTTTTTTTTTTTTTTTTAGTGGAGTCCGTAAAGAGGTAT	2.00
F	14470	SNP		TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGGGAATGATGGTTGTCTTTGG	22.00
F	14560	SNP	14560-F(F)	TTTTTTTTTTTTTTTTTTTTTTCCAAAATTCAGAATAATAACACACCC	2.00
F	14869	SNP	14869-F(F)	ACTCCTTGGCGCCTGCCT	0.66
F	16390	SNP	16390-R	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGAGGATGGTGGTCAAGGGAC	6.00
G	3826, 3834	AMP	A-3826-F(G)	TCCTTTAACCTCTCCACCCTTA	6.67
G	3826, 3834	AMP	A-3826-R(G)		6.67
G	4688	AMP		CAAGTATTTCCTCACGCAAGC	6.67
G	4688	AMP	A-4688-R (G)	TTGATTGGTAGTATTGGTTATGGTT	6.67
G		AMP	A-6293-F(G)	TCCTACTCCTGCTCGCATCT	3.33

G	6293	AMP		GGTGTAAGGAGAAGATGGTTAGG	3.33
G	7891	AMP	A-7891-F(G)	CATAACAGACGAGGTCAACGA	3.33
G		AMP	A-7891-R(G)	TGGGGGAAGTATGTAGGAGTTG	3.33
G	11533	AMP	A-11533-F(G)	CGCTGGGTCAATAGTACTTGC	3.33
G	11533	AMP	A-11533-R(G)	CGTAGGCAGATGGAGCTTGT	3.33
G	12007	AMP	A-12007-F(G)	CAGGACTCAACATACTAGTCACAGC	3.33
G	12007	AMP	A-12007-R(G)	TTCTCGTGTGAATGAGGGTTT	3.33
G	12795	AMP	A-12795-F(G)	CCAACTGTTCATCGGCTGAG	3.33
G	12795	AMP	A-12795-R(G)	CTTGAATGGCTGCTGTGTTG	3.33
G	15043	AMP	A-15043-F(G)	CGCTACCTTCACGCCAAT	6.67
G	15043	AMP	A-15043-R(G)	ATGCCGATGTTTCAGGTTTC	6.67
G	16390	AMP	A-16390-R(G)	TAGCACTCTTGTGCGGGATA	6.67
G	16390	AMP	A-16390-F(G2)	ACCGTACATAGCACATTACAGTCAA	6.67
G	16519	AMP	A-16519-F(G2)	CCATAACACTTGGGGGTAGC	6.67
G	16519	AMP	A-16519-R(G2)	AGAGCTCCCGTGAGTGGTTA	6.67
G	3826	SNP		TTTTCACAACACAAGAACACCTCTGA	5.00
G	3834	SNP	3834-R(G)	GGCCAAGGGTCATGATGG	5.00
	4688	SNP	4688-F(G2)	TTTTTTTTTTTTTTTTTTTTCGCATCCATAATCCTTCTAATAGC	5.00
G	6293	SNP	6293-F(G2)	TTTTTTGGAACAGGTTGAACAGTCTACCC	10.00
G	7891	SNP	7891-R(G2)	TTTTTTTTTGTACTCGTAGGTTCAGTACCATTG	2.50
G	11533	SNP	11533-R(G2)	TTTTTTTTTTTTTTTTTTGGGATAGTACAAGGAAGGGGTA	2.50
G	12007	SNP	12007-F(G)	TTTTTTTTTTTCCTCTACATATTTACCACAACACAATG	2.50
G	12795	SNP	12795-F(G2)	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGGCGTAGGAATTATATCCTTCTT	2.50
G	15043	SNP	15043-F(G2)	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGCCTCTTCCTACACATCGG	10.00
G	16390	SNP	16390-R(G2)	TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGAGGATGGTGGTCAAGGGAC	10.00
G	165119	SNP	16519-F(G)	TTTTTTTTTTTTTTTTTTTTTTTTACATCTGGTTCCTACTTCAGGG	10.00
H	4688	AMP	A-4688-F (H)	CAAGTATTTCCTCACGCAAGC	2.00
H	4688	AMP	A-4688-R (H)	TTGATTGGTAGTATTGGTTATGGTT	2.00
H	11377	AMP	A-11377-F(H2)	CAAACTCCTGAGCCAACAACT	2.00
H	11377	AMP	A-11377-R(H2)	GGCTTCGACATGGGCTTT	2.00
H	12795	AMP	A-12795-F(H)	CCAACTGTTCATCGGCTGAG	2.00
H	12795	AMP	A-12795-R(H)	CTTGAATGGCTGCTGTGTTG	2.00
H	13293	AMP	A-13293-F(H)	CGTAGCCTTCTCCACTTCAA	2.00
H	13293	AMP	A-13293-R(H)	CTTTGAAGAAGGCGTGGGTA	2.00
H	14305	AMP	A-14305-F(H)	CCAATAGGATCCTCCCGAAT	2.00
H	14305	AMP	A-14305-R(H)	AAAGAGTATGATGGGGTGGTG	2.00
H	16519	AMP	A-16519-F(H)	ACCACCATCCTCCGTGAAAT	2.00
H	16519	AMP	A-16519-R(H)	AGACCTGTGATCCATCGTGA	2.00
H	4688	SNP	4688-F(H)	TTTTTTTTCGCATCCATAATCCTTCTAATAGC	10.00
H	11377	SNP	11377-R(H)	TTTTTTTTTTTTTTTTTTTTAGTGGAGTCCGTAAAGAGGTAT	10.00
	12795	SNP	12795-F(H)	TTTTTTTTTTTTTTTGGCGTAGGAATTATATCCTTCTT	5.00

H	13293	SNP	13293-R(H)		1.00
H	14305	SNP	14305-R(H)	TGTGGTAAACTTTAATAGTGTAGGAAG	5.00
H	16519	SNP	16519-F(H2)	TTTTTTTTTTTTTTTTTTTTTTTTTTTACATCTGGTTCCTACTTCAGGG	10.00

Appendix V: Publications and Presentations

Publications

2005 Niederstätter H, Coble MD, Parsons TJ, Parson W. Characterization of mtDNA SNP Typing using Quantitative Real-Time PCR with Special Emphasis on Heteroplasmy Detection and Mixture Ratio Assessment, Proceedings of the 21st Congress of the International Society for Forensic Genetics (2005) *in press*.

2005 Coble MD, Vallone PM, Just RS, Diegoli TM, Smith BC, Parsons TJ. Effective strategies for forensic analysis in the mtDNA coding region, Int. J. Legal Med. (2005) Electronic pre-print (on-line first) DOI: 10.1007/s00414-005-0044-z.

2005 Niederstätter H, Coble MD, Grubwieser P, Parsons TJ, Parson W. Characterization of mtDNA SNP typing and mixture ratio assessment with simultaneous real-time PCR quantification of both allelic states,Int. J. Legal Med (2005) Electronic pre-print (on-line first) DOI: 10.1007/s00414-005-0024-3

2004 Just RS, Irwin JA, O'Callaghan JE, Saunier JL, Coble MD, Vallone PM, Butler JM, Barritt SM, Parsons TJ. Short Communication: Toward increased utility of mtDNA in forensic identifications, Forensic Sci. Int. 146S (2004) S147-S149.

2004 Coble MD, Just RS, O'Callaghan JE, Letmanyi IH, Peterson CT, Irwin JA, Parsons TJ. Single nucleotide polymorphisms over the entire mtDNA genome that increase the power of forensic testing in Caucasians, Int. J. Legal Med. 118 (2004) 137-146.

2004 Vallone PM, Just RS, Coble MD, Butler JM, Parsons TJ. A multiplex allele-specific primer extension assay for forensically informative SNPs distributed throughout the mitochondrial genome, Int. J. Legal Med. 118 (2004) 147-157.

2004 Coble MD. The Identification of Single Nucleotide Polymorphisms in the Entire Mitochondrial Genome to Increase the Forensic Discrimination of Common HV1/HV2 Types in the Caucasian Population. PhD Dissertation, The George Washington University, 206 pp.

2003 Brandstätter A, Parsons TJ, Niederstätter H, Parson w. Rapid Screening of mtDNA Coding Region SNPs for the Identification of west European Caucasian Haplogroups. Int. J. Legal Med. 117 (2003) 291-298.

2003 Levin, B.C., Holland, K.A., Hancock, D.A., Coble, M., Parsons, T.J., Kienker, L.J., Williams, D.W., Jones, M.P., and Richie, K.L. Comparison of the complete mtDNA genome sequences of human cell lines - HL-60 and GMIO742A - from individuals with pro-myelocytic leukemia and leber hereditary optic neuropathy, respectively, and the inclusion of HL-60 in the NIST human mitochondrial DNA standard reference material - SRM 2392-I. Mitochondrion 2: 387-400.

2001 Parsons TJ, Coble MD. Increasing the forensic discrimination of mitochondrial DNA testing through analysis of the entire mitochondrial DNA genome, Croat. Med. J. 42 (2001) 304-309.

Oral Presentations

2005 Just RS, Parsons TJ. mtDNA Genome SNPs for Increased Forensic Discrimination in U.S. Caucasians, African Americans, and Hispanics. NIJ 6th Annual DNA Grantees Workshop, Washington, DC.

2004 Parsons TJ. Oral Presentation. *Weathering Stormy Seas*: Challenges and Progress in mtDNA Forensic Applications. IV. International Forensic Y-User Workshop – Haploid DNA Markers in Forensic Genetics, Berlin, Germany.

2004 Coble MD, Just RS, Saunier JL, O'Callaghan JE, Letmanyi IH, Peterson CT, Irwin JA, Vallone PM, Butler JM, Parsons TJ. Oral Presentation. Identification of SNPs in the Mitochondrial Genome to Resolve Common HVI/HV2 Types in Caucasian Populations. Taipei Symposium on the Application of mtDNA Technology in Forensic Science, Taipei, Taiwan.

2004 Irwin JA. Oral Presentation. Research Efforts at the Armed Forces DNA Identification Laboratory to Improve Forensic DNA Testing. George Washington University Seminar.

2004 Just RS, Irwin JA, O'Callaghan JE, Saunier JL, Coble MD, Vallone PM, Butler JM, Barritt SM, Parsons TJ. Oral Presentation. Research Toward Increasing the Utility of mtDNA in Forensic Identifications. Mediterranean Academy of Forensic Sciences Workshop, Calabria, Italy.

2004 Irwin JA, Just RS, O'Callaghan JE, Saunier J, Parsons TJ. Oral Presentation. Automated, High-throughput Production of Global mtDNA Population Databases. Challenges and Changes: 17thInternational Symposium on the Forensic Sciences: The Australian and New Zealand Forensic Science Society Biennial Conference, Wellington, New Zealand.

2003 Coble MD. Oral Presentation – Dissertation Seminar. The Identification of Single Nucleotide Polymorphisms in the Entire Mitochondrial Genome to Increase the Forensic Discrimination of Common HVI/HVII Types in the Caucasian Population. Institute of Biomedical Sciences and Program in Genetics, George Washington University, Washington, D.C.

2003 Parsons TJ, Coble MD, Hamm RS, O'Callaghan JE, Barritt SM, Vallone PM, Butler JM, Brandstaetter A, Niederstaetter H, Parson W. Oral Presentation. Forensic SNP Testing in the mtDNA Genome. Third European Academy of Forensic Science Meeting, Istanbul, Turkey.

2003 Hamm RS, Coble MD, Vallone PM, O'Callaghan JE, Saunier JL, Letmanyi IH, Peterson CT, Irwin JA, Butler JM, Parsons TJ. Multiplex SNP Panels for Increased Discrimination in Forensic mtDNA Testing. Oral Presentation. European-American School in Forensic Genetics and Mayo Clinic Course in Advanced Molecular and Cellular Medicine, Zagreb, Croatia.

2003 Hamm RS, Coble MD, O'Callaghan JE, Vallone PV, Parsons TJ. SNPs in Forensic mtDNA Testing. Oral presentation. Hamilton College, Clinton, NY.

2003 Hamm RS, Coble MD, Letmanyi IH, Vallone PV, Parsons TJ. Multiplex Fluorogenic Assays for Increasing Discrimination of mtDNA Testing. Oral presentation. Professional Staff Conference, Armed Forces Institute of Pathology, Washington, DC.

2003 Vallone PM. Development of Multiplexed Assays for Evaluating SNP and STR Forensic Markers. Oral Presentation. George Washington University Department of Biological Sciences, Washington, DC.

2003 Parsons TJ. Increasing forensic discrimination of mtDNA: variation outside of HV1/HV2. Invited presentation. Advanced DNA Technical Workshop, Bode Technologies, Hawkes Caye, FL.

2003 Vallone PM. Development of Multiplexed SNP Assays from Mitochondrial and Y Chromosome DNA for Human Identity Testing. Oral Presentation. NIJ DNA Grantees Workshop, Washington DC.

2003 Vallone PM. Typing Single Nucleotide Polymorphisms (SNPs) Located on the Y Chromosome and in the Mitochondrial Genome. Oral Presentation. NIST Biotechnology Division Seminar, Gaithersburg, MD.

2003 Vallone PM. Multiplex SNP Assays for the Evaluation of Forensic Markers. Oral Presentation. Royal Institute of Technology, Department of Biotechnology, Stockholm, Sweden.

2003 Vallone PM. A Multiplex Primer Extension Assay for Probing 11 SNPs Located in the Mitochondrial Genome. Oral Presentation. Uppsala University, Department of Genetics and Pathology, Uppsala, Sweden.

2002 Parsons TJ, Coble MD, Hamm RS, Eyster JL, Letmanyi IH, Niederstaetter H, Parson W, Vallone PM. Multiplex Mito SNPs: How do we really get there, and what do we do with them? Oral Presentation. 13th Promega Symposium on Human Identification, Phoenix, AZ.

2002 Coble MD, Hamm RS, Eyster JL, Letmanyi IH, Parsons TJ. Increasing mtDNA Discrimination for Common Haplotypes: Targeting of Additional Information in the Entire Mitochondrial Genome. Oral Presentation. 13th Promega Symposium on Human Identification, Phoenix, AZ.

2002 Parson W, Niederstaetter H, Coble M, Parsons TJ. Characterization of mtDNA SNP-Typing Using Real-Time PCR. Oral Presentation, 13th Promega Symposium on Human Identification, Phoenix, AZ.

2002 Parsons TJ, Coble MD, Letmanyi IH, Hamm RS, Niederstaetter H, Parson W. Quick and Easy SNP Assays to Increase the Power of Discrimination of mtDNA Testing. Oral Presentation. National Institute of Justice DNA Grantees Workshop. Washington D.C.

2002 Vallone PM, Coble MD, Letmanyi IH, Butler JM, Parsons TJ. Multiplex Detection of 10 SNPs Located in the Coding Region of the Mitochondrial Genome. Oral Presentation. 52nd Annual American Society for Human Genetics Conference, Baltimore, MD.

2002 Parsons TJ, Coble MD, Letmanyi IH, Niederstaetter H, Parson W. MtDNA Genome SNP Discovery and Development for Increased Forensic Discrimination and Heteroplasmy Detection. Oral Presentation. Mitochondrial DNA Workshop, American Academy of Forensic Sciences 54th Annual Meeting, Atlanta, GA.

2002 Vallone PM. Analyzing Single Nucleotide Polymorphisms. Oral Presentation. Forensic Mitochondrial DNA Analysis: A Community Forum Workshop for American Academy of Forensic Sciences, Atlanta, GA.

2002 Coble MD, Letmanyi IH, Hamm RS, Harvey, C, Irwin JA, Parsons TJ. Increasing mtDNA Discrimination for Common Haplotypes: Targeting of Additional Information in the Entire Mitochondrial Genome. Oral Presentation. American Chemical Society Mid-Atlantic Regional Meeting, Fairfax, VA.

2002 Coble MD, Letmanyi IH, Harvey, C, Irwin JA Parsons TJ. Increasing mtDNA Discrimination for Common Haplotypes: Targeting of Additional Information in the Entire Mitochondrial Genome. Oral Presentation. Professional Staff Conference, The Central Identification Laboratory-Hawaii, Honolulu, Hawaii.

2001 Parsons TJ. Increasing the Power of mtDNA Forensic Testing by SNP Assays over the Entire mtDNA Genome. Oral Presentation. Second European-American Intensive Course in Clinical and Forensic Genetics. Dubrovnik, Croatia.

2001 Parsons TJ, Coble MD. Increasing the Power of mtDNA Forensic Testing by SNP Assays over the Entire mtDNA Genome. Oral Presentation. National Institute of Justice Second Annual DNA Grantee's Workshop, Washington D.C.

2001 Coble MD, Harvey C, Letmanyi IH, Parsons TJ. SNP Analysis and Fluorescent Detection of Sequence Variation in the Entire Human Mitochondrial Genome. Oral Presentation. Joint Meeting of the Mid-Atlantic and Southern Associations of Forensic Sciences, Williamsburg, VA.

2001 Coble MD, Harvey C, Letmanyi IH, Parsons TJ. Analysis of Sequence Variation in the Entire Human Mitochondrial Genome. Oral Presentation**.** 53rd Annual Meeting of the American Academy of Forensic Sciences, Seattle, Washington.

2000 Coble MD, Harvey C, Letmanyi IH, Parsons TJ. Analysis of the Entire mtDNA Genome for Polymorphisms to increase the Discrimination of common HV1/HV2 haplotypes. Oral Presentation. Professional Staff Conference, Armed Forces Institute of Pathology, Washington, DC.

Poster Presentations

2005 Just RS, Los CW, Miller C, Leney M, Barritt-Ross SM, Parsons TJ. Using Mitochondrial SNP Typing to Resolve Common HV1/HV2 Types In Highly Degraded Samples From a Missing Persons Case. 16th Promega Symposium on Human Identification, Dallas, TX.

2004 Just RS, Los CW, Leney M, Barritt SM, Parsons TJ. Multiplex mtDNA SNP Typing on Degraded Skeletal Remains. Poster Presentation. The 7th International Conference on Ancient DNA and Associated Biomolecules, Brisbane, Australia.

2004 Niederstatter H, Coble MD, Parsons TJ, Parson W. Characterization of mtDNA SNP typing using quantitative real-time detection PCR for forensic purposes. Poster Presentation. The 1st International qPCR Symposium and Application Workshop, Freising-Weihenstephan, Germany.

2003 Brandstätter A, Parsons TJ, Parson, W. mtDNA Coding Region SNPs for Rapid Screening and Haplogroup Identification of Forensic Samples. 20th International Congress of International Society of Forensic Genetics, Arcachon, France, Sept. 9-13, 2003.

2002 Hamm RS, Coble MD, Letmanyi IH, Vallone PV, Parsons TJ. Multiplex Real Time Fluorogenic SNP Assays for Increasing Discrimination for mtDNA Testing. Poster presentation. 13th International Symposium on Human Identification, Phoenix, AZ.