Final Technical Report for 2003-IJCX-K104

REPORT TITLE: **Mitochondrial DNA Analysis by Denaturing Liquid Chromatography for the Separation of Mixtures in Forensic Samples**

AWARD NUMBER: **2003-IJCX-K104**

AUTHOR: **Phillip B. Danielson**

ABSTRACT:

Overview **-** Mitochondrial DNA (mtDNA) sequencing can provide crucial information to investigators when the quantity and quality of DNA is otherwise limiting. As electrophoretic peak heights of mixed bases are context dependent, however, it is impossible to determine the haplotypes of individual contributors to a DNA mixture by direct inspection of sequencing traces. This can impede the forensic use of mtDNA. Alternate methods (*e.g.*, subcloning) that could separate contributors to a mixture, are time consuming and prone to artifacts from PCR and increased sample manipulation.

Denaturing High-Performance Liquid Chromatography (DHPLC) is a chromatographic means of fractionating natural (heteroplasmic) or situational (multi-contributor) DNA mixtures prior to sequencing. This approach does not require secondary amplification or excessive manipulation. Forensic implementation of this approach can be achieved with minimal training and consumable costs.

Project Objectives **-** The goal of this research program was the developmental validation of DHPLC for the rapid and accurate resolution of mtDNA mixtures. Our core objectives were to:

1) **Rigorously Demonstrate** the reliability of DHPLC to detect and resolve mixtures involving all classes of mtDNA HV1 and HV2 polymorphisms.

2) **Rigorously Demonstrate** the reproducibility of DNA sequencing electropherograms with an emphasis on the statistical correlation between peak height ratios and DNA quantity ratios at mixed base positions.

3) **Rigorously Demonstrate** the reliability of determining individual haplotypes in a DNA mixture through linkage phase analysis of sequence data from DHPLC fractionated samples.

4) **Develop** standard operating procedures for DHPLC mixture resolution and work toward interpretation guidelines.

Results and Conclusions **-** All core objectives have been achieved. DHPLC has been shown to be an accurate method for rapid sequence-specific fractionation of heteroplasmic or multi-contributor DNA mixtures prior to sequencing. Fractionation allows the mtDNA sequence of individual contributors to be unambiguously determined without secondary amplification or excessive manipulation. Although complete physical separation of both contributors of a mixture can be achieved in many cases, this is not essential. The

demonstrated reproducibility of electrophoretic sequence profiles and correlation with changes in DNA quantity ratios makes it possible to determine the linkage phase of individual amplicons from even partially resolved mtDNA mixtures. Moreover, DHPLC can be used to streamline the entire mtDNA analysis process flow by enabling analysts to assess sample complexity, accurately quantify PCR product yield without gels, purify target amplicons to facilitate improved dye terminator labeling and confirm negative controls without sequencing. This approach has been also been validated using casework-type samples.

TABLE OF CONTENTS:

EXECUTIVE SUMMARY:

[Introduction] Sequencing of DNA from Hypervariable Regions 1 and 2 (HV1/HV2) of human mtDNA is the method of choice for the analysis of human mtDNA. This approach has been forensically validated and has withstood several court challenges (*e.g.*, U.S. v. Chase, Superior Court of the District of Columbia, Criminal Division, No.F-7330-99, (2005) and People v. Klinger, 713 N.Y.S.2d 823, 2000 N.Y. Slip Op. 20450 (N.Y.Co.Ct., 2000) (NO. 0849/00)) [1-3].

A mixture of different mtDNA molecules in a single sample, however, presents a significant obstacle to successful mtDNA analyses by standard methods. For example, an individual human can naturally possess more than one mtDNA haplotype – a phenomenon, termed "heteroplasmy" [4]. Alternately, forensic analysts often encounter mixtures of DNA-containing body fluids from separate humans – *i.e.*, a "situational" mixture. The ability to accurately determine the specific haplotypes of individual contributors to a DNA mixture can provide valuable investigative information.

While the size variation of short tandem repeats can facilitate the interpretation of nuclear DNA mixtures, this is not the case with mtDNA haplotypes that differ primarily by base substitution or by very small differences in length - most often arising from the insertion or deletion of a single basepair. Such subtle variations in sequence produce ambiguous base calls due to overlapping electrophoretic peaks. Since electrophoretic peak heights are strongly context dependent, direct inspection of the height of a peak cannot be used to determine the relative abundance of the underlying amplicon. This is in marked contrast to the case with STR traces. Similarly, variations in the length of mtDNA molecules, such as those commonly associated with the HV1 cytosine stretch, result in long stretches of unreadable electropherograms. This reflects the fact that the dideoxy terminated chains of labeled DNA fall out of register with each other starting at the position of the insertion or deletion responsible for the length variant. In these cases, additional sequencing primers targeted to flank the heteroplasmic cytosine stretch must be used to obtain readable sequence. Unfortunately, this reduces the total amount of sequence data that can be used to characterize a given mitochondrial amplicon [2]. The ability to accurately resolve and interpret these types of samples in a timely and cost efficient manner would substantially increase the power of mtDNA analysis by allowing its use in cases where the current approach yields results that are of limited or no utility.

There are a number of molecular strategies that have been proposed to separate DNA mixtures. These include denaturing gradient gel electrophoresis (DGGE), single-strand conformational polymorphism (SSCP) analysis and subcloning into bacterial vectors [5-7]. Both DGGE and SSCP, however, require manual recovery of fractionated DNA from polyacrylamide gels and PCR reamplification to generate enough template for DNA sequencing. Subcloning represents an even more time and labor-intensive approach. Furthermore, it would require forensic scientists to screen and sequence DNA from multiple transformed bacterial colonies to ensure that observed sequence differences reflect genuine contributors to the starting template rather than artificial variants that were introduced as a result of DNA nucleotide misincorporation by *Taq* polymerase. Thus, the tedious and time consuming aspects of these approaches, which are not readily adaptable

to automation, have been obstacles to the implementation of these technologies by forensic laboratories.

By comparison, DHPLC is a rapid and cost-effective method for physically fractionating mtDNA mixtures and then recovering amplicons for characterization by direct DNA sequencing [8, 9]. While doing so, DHPLC addresses the limitations of alternative strategies. The underlying principle of DHPLC is Temperature Modulated Heteroduplex Analysis [10]. Briefly, a mixture of two (or more) DNA amplicons is heat denatured and then allowed to slowly reanneal forming a combination of homo- and heteroduplices. The homoduplices represent the original components of the mixture while the heteroduplices are formed by cross-hybridization of the different contributors to the mixture. Under partially denaturing conditions and an increasing gradient of a nonpolar solvent (*e.g.*, acetonitrile), it is possible to separate both the heteroduplices and the homoduplices from each other by HPLC [11]. Heteroduplices, are inherently less stable as the result of one or more base pair mismatches and thus they appear as early eluting peaks in the chromatographic trace of a DHPLC assay (Figure 1) [12, 13]. The sensitivity of DHPLC to the thermodynamic stability of a DNA amplicon ensures that even very subtle sequence differences (*i.e.*, single nucleotide substitutions, insertions, deletions) between the components of a mixture can be readily detected and fractionated in time [14, 15].

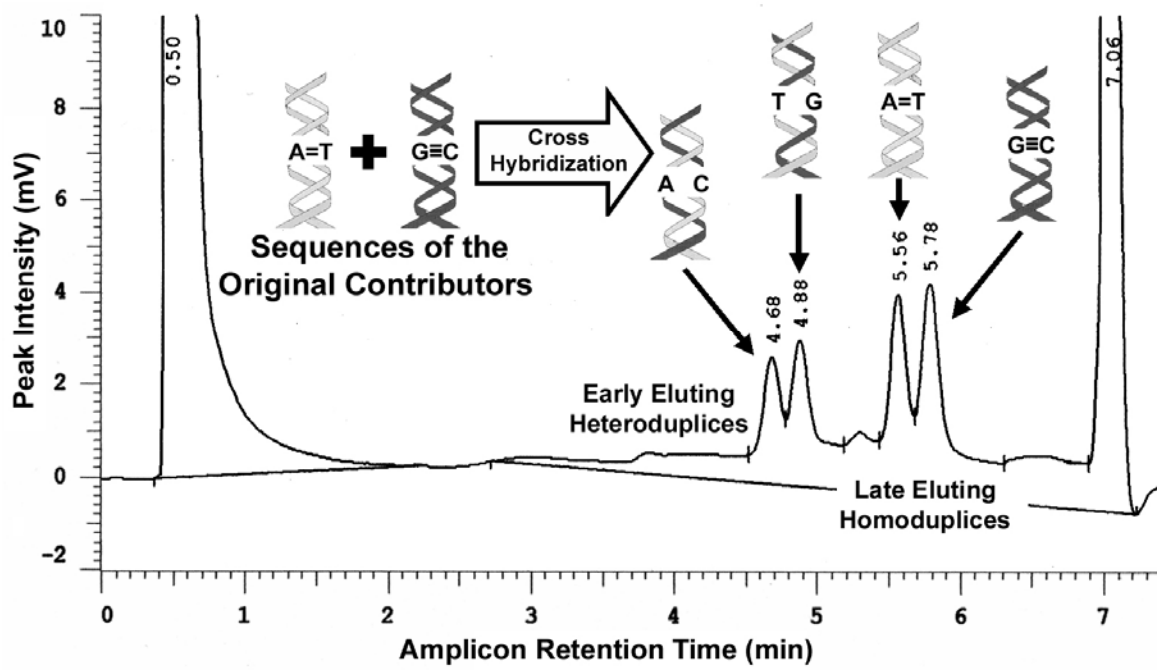


Figure 1: DHPLC chromatograph of an mtDNA mixture illustrating hetero- and homoduplices produced by cross-hybridization of two amplification products. One has an A/T and the other has a G/C basepair at the same site. The earliest eluting peak contains the most destabilizing basepair mismatch (*i.e.*, A/C) while the latest eluting peak has the most stabilizing basepair (*i.e.*, C/G) at the same position.

In the current study, the efficacy of DHPLC for fractionating mtDNA mixtures to recover sequence ready DNA without reamplification is investigated. Also examined is the correlation between a change in the molar ratio of two contributors to a DNA mixture and

the change in the ratio of electrophoretic peak heights at mixed base positions. Such information is crucial to the development of statistically supported interpretation guidelines for accurately determining the linkage phase of mtDNA amplicons representing the individual contributors to a mixture. Finally, DHPLC was evaluated and validated as a "comprehensive approach" to streamlining the standard workflow associated with mtDNA analysis. Specifically, the accuracy and sensitivity with which DHPLC can be used to screen for sample complexity; quantify and purify PCR products; and confirm negative control samples was assessed.

[Methods and Materials]

This research was conducted in compliance with U.S. Federal Policy for the Protection of Human Subjects (Basic DHHS Policy for Protection of Human Research Subjects; 56 FR 28003). Individual buccal swabs were collected from 96 unrelated subjects of diverse ethnicity who had previously provided informed consent to participate in the study.

Mitochondrial DNA Extraction and Sequencing **-** Mitochondrial DNA was extracted from buccal swabs using the EZ1 DNA tissue kit with the Qiagen BioRobot EZ1 (Qiagen Inc., Valencia, CA) in accordance with the manufacturer's protocol. Mitochondrial DNA hypervariable region amplicons (HV1A, HV1B, HV2A and HV2B) were prepared using forensically validated PCR primers and using previously published conditions [3, 16]. Sequencing was performed according to the manufacturer's protocols using the Big Dye® Terminator v1.1 Cycle Sequencing kit (Applied Biosystems, Foster City, CA). Labeled products were purified using Performa DTR V3 96-well Short Plate Kit (Edge BioSystems Gaithersburg, MD) according to the manufacturer's protocol and analyzed on an Applied Biosystems 310 Genetic Analyzer. The resulting sequence data were analyzed using the Sequencher™ DNA analysis software (Gene Codes Corp, Ann Arbor, MI). Appropriate positive and negative controls were carried through the entire sample handling process detect cross contamination and to monitor amplification efficiency.

DHPLC Analysis **-** Mixture amplification reactions were performed using the same PCR primers, and conditions mentioned previously. Amplification efficiency was determined by injecting 5µl of PCR amplification product into a WAVE® 3500HT DNA Fragment Analysis System (Transgenomic Inc., Omaha, NE) containing a DNAsep analytical column packed with alkylated poly(styrenedivinylbenzene) resin [17]. PCR product yield was then determined from the peak area of the resulting DHPLC trace. There is a near perfect correlation (R^2 = 0.9982) between peak area and DNA quantity. This has been internally validated for fragments ranging in size from 200-1000bp and for a quantity range of 1.25-140ng (Data not shown).

Cross hybridization of mixed amplicons was performed by denaturation of the mtDNA at 95˚C for 4 minutes and renaturation by gradual cooling (1.5˚C/min) over a 45 minute period to reach a final temperature of 25˚C [16].

Cross hybridized mixtures were analyzed by Temperature-Modulated Heteroduplex analysis by [14] at empirically predetermined optimal temperatures for each of the four forensic mtDNA amplicons (HV1A, 58˚C; HV1B, 59.2˚C; HV2A, 56.5˚C and HV2B, 57˚C.) The optimal acetonitrile linear gradient generated from differential mixing of buffer A (0.1M TEAA) and buffer B (0.1M TEAA, 25% ACN) were 56% to 65% buffer B

increase in 3.5 minutes for HV1A, HV2A and HV2B and a 55% to 64% buffer B increase in 3.5 minutes for HV1B. All samples were eluted at 0.9ml/min flow rate and detected by UV absorbance at 260nm. Standard DHPLC controls included: zero-volume injections to screen for residual DNA carryover from the column matrix, no-template PCR controls to check for amplification contamination and manufacturer provided mutation control standards for buffer and column quality control.

[Results]

DHPLC Fractionation and recovery of DNA from a two-component mtDNA mixture - The applicability of DHPLC for the fractionation of DNA mixtures requires that the technology be able to chromatographically resolve mtDNA amplicons which differ in sequence but not length. Once resolved in time, it must be possible to physically recover sufficient quantities of DNA for sequencing. At maximum resolution, DHPLC fractionation of a two-component mtDNA mixture will yield a chromatographic trace consisting of two homoduplices and two heteroduplices. The heteroduplices being formed by cross-hybridization of the original amplicons in the mixture. Following fractionation and recovery of DNA from individual peaks, aliquots can be reanalyzed by DHPLC to confirm isolation and recovery of a given target peak. The DHPLC chromatograms from reinjected DNA fractions are characterized by a single major peak having a retention time consistent with that of the target peak in the initial chromatogram.

To demonstrate the applicability of DHPLC to chromatographically fractionate a broad variety of identical or similar length mtDNA amplicons which differ in primary sequence, 576 two-component, equimolar mixtures of previously sequenced amplicons were assayed by DHPLC for the occurrence of early eluting peaks consistent with a mixture of non-identical amplicons. Collectively, these mixtures represented a total of 85 of 246 known variants sites in HV1 and 59 of 160 variant positions across HV2. The variant sites included single and multi-base insertions, deletions and base substitutions which occurred in regions of GC- as well as AT-rich sequence. The detection of an mtDNA mixture by DHPLC was fully concordant with the direct sequencing data for the amplicons which were combined in each pairwise test.

Reproducibility of Relative Peak Height Patterns in Sequencing Electropherograms - Peak heights on sequencing electropherograms reflect the quantity of DNA molecules terminated at a given position. This is a function of the sequence context of each base which influences the efficiency of dye terminator incorporation. As a result, DNA sequence data are characterized by electrophoretic peaks for individual nucleotide position which may differ significantly relative to each other. If DNA sequence data are to be quantitatively compared as would be the case among sequential DHPLC generated fractions of an mtDNA mixture, the reproducibility of the relative differences in peak heights across an electropherogram is essential.

To evaluate the consistency of relative differences among electrophoretic peak heights for each of four forensically relevant mtDNA fragments, HV1A, HV1B, HV2A and HV2B amplicons were sequenced. Each amplicon was sequenced at five different DNA template input concentrations, which encompassed a range frequently encountered by forensic practitioners (*i.e.,* 0.1ng – 1.6ng). All dye terminator labeling reactions were

independently replicated five times in both the forward and reverse direction for statistical accuracy. Figure 2 presents data for positions 16,148 to 16,184 of HV1A. Shown are mean peak heights±SD in relative fluorescence units (RFU) as determined from sequence data representing five independent labeling reactions at each of three different quantities of input DNA (*i.e.*, 1.6ng, 0.8ng and 0.4ng). While the mean height of any one electrophoretic peak may be significantly greater or lesser than that of neighboring peaks, the means were consistent at all three quantities of input DNA shown as well as at 0.2 ng and 0.1 ng. This is exemplified by the relatively low peak heights of 16,155A and 16,179C versus the consistently greater peaks heights at 16,153G and 16167C. Depending on the amplicon and sequencing primer used the variability of any specific peak was generally observed to be lowest for 0.4 to 0.8ng of input DNA and most pronounced for labeling reactions containing 1.6ng of input DNA. Although the data presented in figure 2 represent a small portion of the HV1A fragment, these data are consistent with the results that have been obtained for the entire HVI and HV2 region.

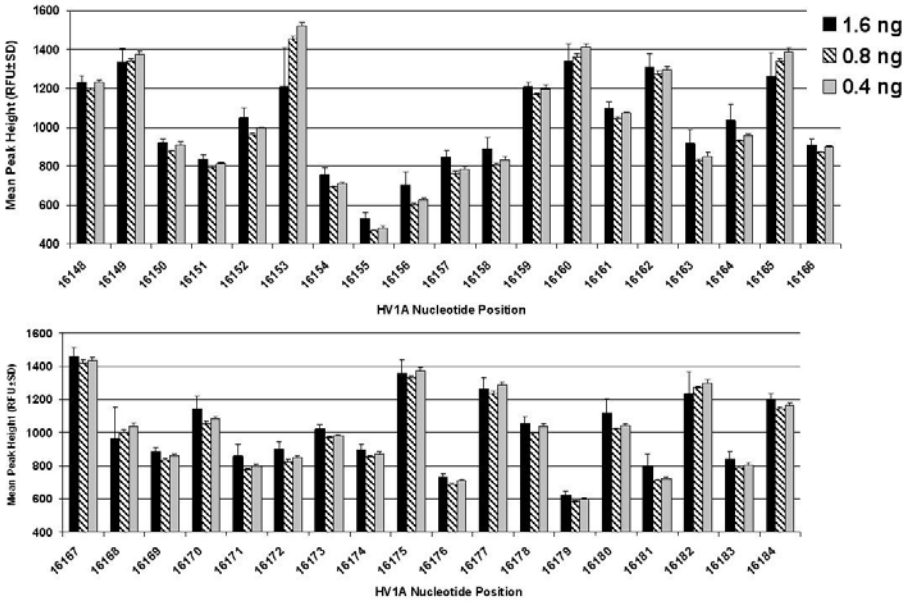


Figure 2: Reproducibility of sequencing electrophoretic peak heights for positions 16,148 to 16,184 of the mtDNA HV1A amplicon. The data represent the mean peak heights and standard deviations from five independent labeling reactions for each of the three DNA quantities. Although there are significant differences in the mean peak heights between among different nucleotide positions, at any individual nucleotide position, the mean peak height remains consistent across the range of DNA input quantities shown.

Correlation between DNA Quantity Ratios and Electrophoretic Peak Height Ratios - When a mixture of two or more DNA amplicons are labeled for sequencing, overlapping electrophoretic peaks are typically observed at those sites where the amplicons differ in primary sequence. Because peak height is sequence context dependent, peak height alone cannot be used to determine the absolute or even relative quantities of DNA representing the individual components of the mixture. Given the reproducibility of individual peak heights, however, it is reasonable to postulate that a change in the ratio of one component to another in a mixture will be reflected by a corresponding change in the ratio of the overlapping fluorescence signals at any mixed base position. Figure 3 illustrates the relationship between differences in the molar ratio of two contributors to a mixture and the electrophoretic peak height ratio at four mixed base positions (16,163G/A; 16,179 C/T; 16,186 C/T; 16189C/T). As the molar ration between two contributors in an HV1A

mtDNA mixture is varied in a stepwise manner from a 50:50 to a 90:10 ratio of contibutor 1 to contributor 2, there is a coordinated shift in the ratio of the overlapping electrophoretic peaks at nucleotide positions where the two contributors differ from each other. Thus, it is this change in the peak height ratio and not an estimate of the actual DNA quantities that is used to determine the linkage phase of an amplicon. In this example, four peaks representing nucleotides associated with contributor 1 (16,163G; 16,179C; 16,186T; 16189C) display a coordinated increase in fluorescence relative to the peaks associated with the contributor 2 (16,163A; 16,179T; 16,186C; 16189T).

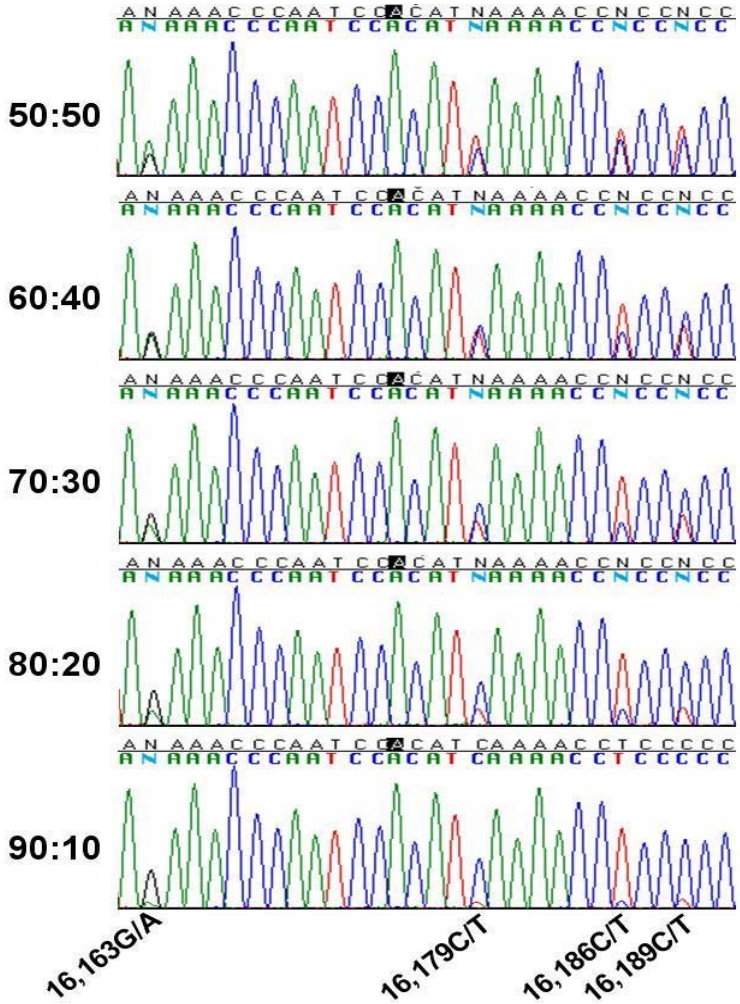


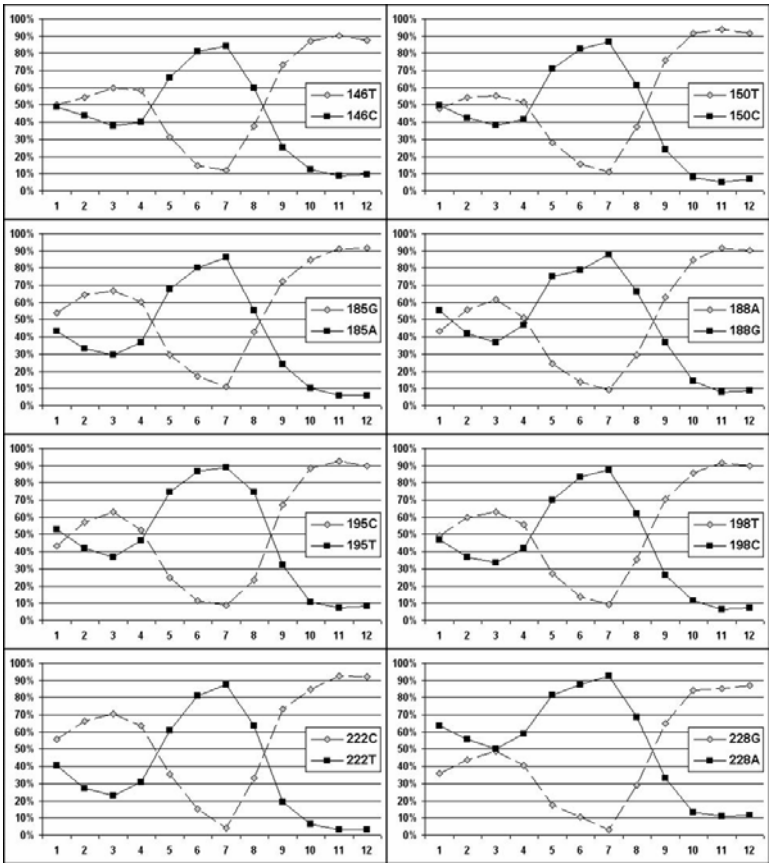
Figure 3: Fluorescence peak height ratios at four mixed base positions as a function of a series of stepwise shifts in DNA quantity from a 50:50 mixture to a 90:10 mixture of two DNA amplicons. The coordinated shift in the overlapping eletrophoretic peaks identifies the linkage phase of contributor 1 (16,163G; 16,179C; 16,186T; 16189C) as distinct from the linkage phase of contributor 2 (16,163A; 16,179T; 16,186C; 16189T).

To quantitatively assess the strength of this correlation, peak height ratios at 202 mixed base positions encompassing both strands of the HV1A, HV1B, HV2A and HV2B amplicons were assayed at nine stepwise mixture ratios ranging from 10:90 to 90:10. Consistent with the high degree of reproducibility that was observed previously for

relative differences in electrophoretic peak heights, a change in the molar ratio of one contributor to another was found to correlate nearly perfectly (R^2 = 0.99948) with a change in the relative fluorescence of overlapping peaks at mixed base positions. This is not to suggest, that an equimolar mixture of two amplicons will necessarily result in perfectly overlapping electrophoretic peaks. On the contrary, there are numerous nucleotide positions where mixtures containing equimolar quantities of two amplicons display skewed electrophoretic peak height ratios as high as 26:74. Nevertheless, even positions that display significant discordance between molar ratios and fluorescence ratios still display a strong correlation between a change in the ratio of input DNA and the change in the ratio of overlapping peak heights at mixed base positions.

Linkage Phase Determination by Sequence Analysis of Sequential DHPLC Fractions - Because of the strong correlation between a change in the ratio of DNA quantities within a mixture and a change in the ratio of overlapping electrophoretic peaks, the linkage phase of the individual amplicons present in a two component mixture can be readily determined. This holds true even in cases where the complete separation of the individual components of a mixture cannot be achieved. In such cases, the linkage phase of each components of the mixture is determined by tracking the ratio of overlapping fluorescent peaks at all mixed base positions between two or more DHPLC fractions. The observation of coordinated shifts in the relative fluorescence ratios for a given set of nucleotides is consistent with them being in the same linkage phase and thus representing the same amplicon. To illustrate this, a two-component mixture of HV2A amplicons was fractionated by DHPLC. A total of twelve fractions were recovered and sequenced. The peak height ratio at each of the eight mixed base positions in the resulting series of electropherograms is presented in figure 4. The linkage phase of the two amplicons can be determined from the coordinated shifts in peak height ratios across the fractions that were sequenced. In this example, the height of the fluorescent peaks associated with nucleotides 146T, 150T, 185G, 188A, 195C, 198T, 222C and 228G increase and decrease in a coordinated manner across the twelve DHPLC fractions. Similarly, the height of the fluorescent peaks associated with nucleotides the 146C, 150C, 185A, 188GA, 195T, 198C, 222T and 228A also shift in a coordinated manner but in the opposite direction. Thus, these two sets of bases define the linkage phase of each of the two contributors to the mixture. These results are in concordance with the known sequence of the individual amplicons used to construct the mixture. Careful inspection of the results for individual mixed base positions also demonstrates that discordance between molar and fluorescence ratios does not compromise the accurate linkage phase determination. This is exemplified by nucleotide positions 222 and 228. Sequence electropherograms generated from DHPLC fractions 1-4 consistently show equal or greater fluorescence for the 222C and 228A pairing than for 222T and 228G pairing. Neither of these nucleotide pairings, however, demonstrate coordinated shifts in relative fluorescence across these DHPLC fractions. Rather, it is the 222C and 228G nucleotide pairing that shows coordinated increases in fluorescence between fractions 1 and 3 and then a coordinated decrease between fractions 3 and 4. Conversely, the 222T and 228A nucleotide pairing shows the complementary shift, *i.e.,* a decrease in relative fluorescence between fractions 1 and 3 and an increase between 3 and 4. It is the nucleotide pairings based on coordinated shifts in relative fluorescence that are concordant with the actual sequence of the amplicons present in the mixture. This underscores the fact that it is the change in relative fluorescence and not the

absolute peak height that allows for linkage phase of the individual amplicons in an mtDNA mixture to be determined. Using this approach to analyze a set of 36 two-component mixtures it was possible to accurately determine the linkage phase in all cases.



12 sequentially collected DHPLC fractions.

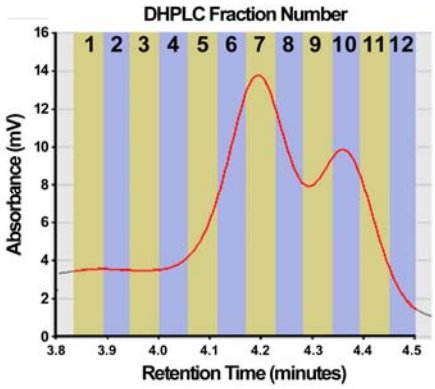


Figure 4:
(A) DHPLC chromatogram from a 70:30 mixture of two HV2A amplicons showing the retention times of the DNA fractions that were recovered. (B) Changes in peak height ratios at eight mixed base positions as determined from the resulting sequence electropherograms for each of the

[Discussion]

The current study provides an initial assessment of the potential utility of DHPLC to resolve mixtures involving mtDNA amplicons commonly used by forensic analysts. The essential first step in DHPLC-based mixture fractionation is the determination of whether or not a sample contains two or more non-identical DNA sequences. Based on differential retention of individual DNA amplicons, the detection of mtDNA mixtures by DHPLC was found to be concordant with direct sequencing.

Although fragment resolution, is a function of amplicon length and/or differences in primary sequence, it is difficult to predict *apriori* the characteristics of the chromatographic trace that will be result from a specific mixture. The number of sequence differences plays a secondary role. Resolution of a 2-component mixture into three or four discernible peaks (*i.e.*, separated homo- and/or heteroduplices) is optimal for the physical separation and recovery of DNA from individual contributors. This minimizes coelution

of non-target DNA and subsequent electropherograms are generally of good quality and are free of underlying electrophoretic signal from a second contributor.

In practice, however, less than optimal chromatographic resolution is common and the resulting coelution of target and non-target DNA produces a series of DHPLC fractions that are enriched for one contributor versus the other. In the absence of complete separation, sequence electropherograms show signal from both contributors with overlapping peaks at those sites where the two amplicons differ in primary sequence. Although the current study demonstrated that peak heights in an electropherogram are highly reproducible, there are significant differences in relative peak heights that are sequence context dependent. As a result, comparative measurements of peak height alone cannot generally be used to accurately determine the sequence of the individual amplicons based on absolute abundance. To circumvent this limitation, the current study sought to determine the linkage phase of individual amplicons by tracking coordinated shifts in fluorescence ratios among mixed base positions across a series of DHPLC fractions. The quantitative comparison of DNA overlapping peak height ratios among DHPLC fractions, requires that any relative differences in peak heights in an electropherogram be highly reproducible. Excessive variability would make the use of peak height ratios at mixed base positions unreliable. The current study demonstrated that peak heights in an electropherogram are highly reproducible for all four mtDNA amplicons; for different sequencing primers and; a range of DNA input quantities.

An analysis of over 200 mixed base positions across a broad range of mixture ratios revealed a statistically strong correlation between changes in the relative quantity of two amplicons and changes in the ratio of overlapping fluorescent peaks at positions where the amplicons differed in sequence. Thus, a change in the relative quantity of one amplicon to the other (*e.g.,* from a 50:50 to a 70:30 mixture of amplicons “1” and “2”), results in an increase in the heights of all bases associated with amplicon “1” and a corresponding decrease in the heights of all bases associated with amplicon “2”. By analyzing such relative peak shifts across a series of DHPLC fractions each of which are differentially enriched for a given contributor, is possible to clearly and reproducibly determine the linkage phase of individual amplicons. It is important to note however, that while there is every indication that DHPLC in combination with linkage phase analysis is a highly reliable means of analyzing mixtures, it would be premature to assume that DHPLC will resolve all mixtures with 100% accuracy as we have not tested every conceivable mtDNA mixture.

The extensive body of information that the current study has collected on peak height reproducibility as a function of input DNA quantity, sequencing primer used and sequence context has facilitated the development of statistically supported interpretation guidelines and statistical tests of linkage phase confidence values. Moreover, the results of this research have improved out understanding of the pattern of relative differences in the heights of adjoining peaks in a DNA sequence electropherogram. This information can be used to assess how closely the electropherogram for a given test sample conforms to that of a pristine reference sample. A thorough understanding of specific factors that may impact the reproducibility of a given electrophoretic trace may also make it possible to rapidly identify both software anomalies and genuine sequence irregularities thereby

facilitating more efficient and accurate interpretation of mtDNA sequence data by lab analysts.

It should be stressed that a number of questions remain to be addressed. For example, one concern is that even if the linkage phase of individual amplicons can be determined, it will not be possible to associate the fractionated HV1 amplicons with the fractionated HV2 amplicons in order to report a complete haplotype. Here the linkage phase approach alone may not be able to resolve this problem. It is postulated, however, that phylogenetic-based methods could be used to accurately associate these fragments. Another area of concern is how readily linkage phase analysis can be applied to mixtures of three or more contributors, whether they are the result of three different individuals or a combination of hetero- and homoplasmic individuals. Preliminary results from the analysis of these more complex mixtures suggest that interpretation will be appreciably more difficult. This is a task that will likely necessitate a bioinformatics solution involving likelihood-based statistics.

[Conclusion]

The difficulty and amount of analyst time required for analyzing mtDNA mixtures has been a significant challenge for forensic practitioners and limits the potential utility of mtDNA in some forensic investigations. DHPLC is an accurate method for rapid, low-cost sequence-specific fractionation of heteroplasmic or multi-contributor DNA mixtures prior to sequencing. The technology also offers analysts a significant opportunity to streamline the standard workflow at nearly every step in the process. On the front side, DHPLC makes it possible to determine with near 100% accuracy whether an amplified mtDNA sample consists of a single amplicon or represents a mixed product (*i.e.*, either a heteroplasmic or situational mixture). If the product is mixed, it is often possible ti determine by DHPLC the approximate level of complexity (e.g., 2 contributors vs >2 contributors). Also on the front end of the mtDNA amplification process, DHPLC makes it possible to very accurately determine the amount of template amplified in the PCR process and to simultaneously purify the target template for optimal dye terminator labeling. This not only eliminates the need for laborious yield gels and expensive PCR product clean up columns but also ensures that the analyst is using an optimal quantity and quality of mtDNA for the dye terminator labeling reaction. In the case of mixed samples, fractionation allows the mtDNA sequence of individual contributors to be unambiguously determined without secondary amplification or excessive manipulation. Although complete physical separation of both contributors of a mixture can be achieved in many cases by DHPLC, this level of chromatographic resolution is not essential. The excellent reproducibility of electrophoretic sequence profiles and correlation with changes in DNA quantity ratios, makes it possible to determine the linkage phase of individual amplicons in an mtDNA mixture by correlating the enrichment of individual amplicons across a series of DHPLC fractions with corresponding shifts in relative peak fluorescence at mixed base positions. Finally, the sensitivity of DHPLC with a fluorescence detector to detect amplified DNA surpasses that of DNA sequencing. Accordingly it is possible to use DHPLC to confirm the absence of amplified DNA in negative controls. This represents a significant potential for cost savings over the traditional approach of always sequencing negative and reagent blank controls. A set of initial standard operating procedures for implementation and integration of DHPLC into the workflow for standard forensic

mtDNA casework have been written (Appendix I) Additionally, software to aid analysts by automating mixture resolution from sequence data files using statistically grounded interpretation guidelines has been developed.

[Executive Summary References]

[1] Fourney, R., *Mitochondrial DNA and Forensic Analysis: A Primer for Law Enforcement.* Journal of the Canadian Society of Forensic Science, 1998. **31**(1): p. 45-53.

[2] Holland, M.M. and T.J. Parsons, *Mitochondrial DNA Sequence Analysis-Validation and use for forensic casework.* Forensic Science Review, 1999. **11**(1): p. 22-49.

[3] Wilson, M.R., et al., *Validation of mitochondrial DNA sequencing for forensic casework analysis.* Int J Legal Med, 1995. **108**(2): p. 68-74.

[4] Lightowlers, R.N., et al., *Mammalian mitochondrial genetics: heredity, heteroplasmy and disease.* Trends Genet, 1997. **13**(11): p. 450-5.

[5] Hanekamp, J.S., W.G. Thilly, and M.A. Chaudhry, *Screening for human mitochondrial DNA polymorphisms with denaturing gradient gel electrophoresis.* Hum Genet, 1996. **98**(2): p. 243-5.

[6] Steighner, R.J., et al., *Comparative identity and homogeneity testing of the mtDNA HV1 region using denaturing gradient gel electrophoresis.* J Forensic Sci, 1999. **44**(6): p. 1186-98.

[7] Barros, F., et al., *Rapid and enhanced detection of mitochondrial DNA variation using single-strand conformation analysis of superposed restriction enzyme fragments from polymerase chain reaction-amplified products.* Electrophoresis, 1997. **18**(1): p. 52-4.

[8] Emmerson, P., et al., *Characterizing mutations in samples with low-level mosaicism by collection and analysis of DHPLC fractionated heteroduplexes.* Hum Mutat, 2003. **21**(2): p. 112-5.

[9] Etokebe, G.E., et al., *Physical separation of HLA-A alleles by denaturing high-performance liquid chromatography.* Tissue antigens, 2003. **61(6)**: p. 443-50.

[10] Hou, Y. and S. Zhang, *Temperature-modulated high-performance liquid chromatography for detecting variation in human genome.* Chung-Hua i Hsueh i Chuan Hsueh Tsa Chih, 2000. **17**(3): p. 145-8.

[11] Huber, C.G., P.J. Oefner, and G.K. Bonn, *High-resolution liquid chromatography of oligonucleotides on nonporous alkylated styrene-divinylbenzene copolymers.* Anal Biochem, 1993. **212**(2): p. 351-8.

[12] Huber, C.G., P.J. Oefner, and G.K. Bonn, *Rapid and accurate sizing of DNA fragments by ion-pair chromatography on alkylated nonporus poly(styrene-divinylbenzene).* Analytical Chemistry, 1995. **67**: p. 578-585.

[13] O'Donovan, M.C., et al., *Blind analysis of denaturing high-performance liquid chromatography as a tool for mutation detection.* Genomics, 1998. **52**(1): p. 44-9.

[14] Kuklin, A., et al., *Detection of single-nucleotide polymorphisms with the WAVE™ DNA Fragment Analysis System.* Genetic Testing, 1997. **1**: p. 201-206.

[15] Underhill, P.A., et al., *Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography.* Genome Res, 1997. **7**(10): p. 996-1005.

[16] LaBerge, G.S., R.J. Shelton, and P.B. Danielson, *Forensic utility of mitochondrial DNA analysis based on denaturing high-performance liquid chromatography.* Croat Med J, 2003. **44**(3): p. 281-8.

[17] Huber, C.G., P.J. Oefner, and G.K. Bonn, *Rapid analysis of biopolymers on modified non-porous polystyrene-divenylbenzene particles.* Chromatographia, 1993. **37**: p. 653-658.

FINAL TECHNICAL REPORT (MAIN BODY):

[I. Introduction: Statement of the problem]

The Challenge of DNA Mixtures: Analysis of mtDNA is currently accomplished by sequencing the DNA from hypervariable regions 1 and 2 (HV1/HV2) of the displacement loop (nt 16024-16400; 19-410). This approach has been extensively used to identify human remains and has withstood several court challenges (see www.denverda.org for specific case law examples).

Although the protocol for sequencing mtDNA is well established, the presence of a mixture of different mtDNA molecules in a single sample is a significant obstacle to successful mtDNA analyses by standard methods. Individual humans can possess more than one mtDNA sequence (*i.e.*, mitotype), which differ by length or sequence. This phenomenon, termed "heteroplasmy", is well known by biomedical researchers (1).

Heteroplasmy can be considered a "natural" DNA mixture while a mixture of body fluids containing DNA often encountered in forensic science can be considered a "situational" mixture. The mixture of bodily fluids can provide forensic scientists with valuable investigative information providing that the individual contributors to a mixture can be resolved. For example, sexual assault cases often start as DNA mixtures between the victim's DNA and the perpetrator's DNA with the most common situation involving the mixture of semen with vaginal fluids.

In forensics, natural or situational mixtures of mtDNA molecules that differ by sequence or length can be problematic in determining the true level of discrimination between sequences that are determined to "match" (2). Variations in sequence result in ambiguous base calls while variations in the length of the HV1 cytosine-stretch, for example, result in large stretches of unreadable electropherograms since the dideoxy terminated chains of labeled DNA fall out of register starting at the position of the base responsible for the length variant (Figure 4). In these cases, additional sequencing primers that flank the cytosine stretch can be tried in an effort to obtain readable sequence. Unfortunately, this addresses only one of many possible sites of length heteroplasmy and it reduces the total amount of sequence data that can be used to characterize a given mitotype (3).

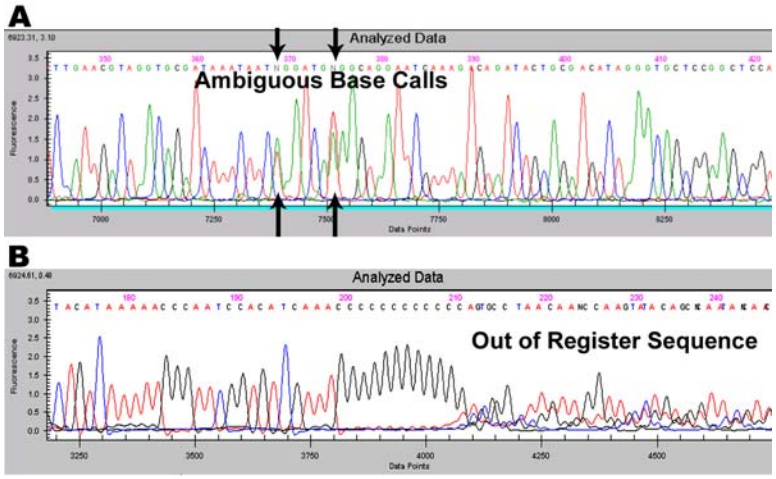


Figure 4: DNA Sequencing electropherogram showing ambiguous base calls as indicated by overlapping green and red electrophor-etic peaks (arrows in **Figure 4A**) and unreadable sequence data resulting from a mixture of mtDNA length variants (**Figure 4B**). Such variants cause the sequence to fall out of register beginning at the site of the insertion/deletion.

Of specific concern to analysts conducting mtDNA testing is the observation from extensive casework records that a significant proportion of hairs were heteroplasmic (11.4%) or displayed a mixed profile (8.7%). Moreover, the occurrence of mixed mtDNA profiles appears to increase with the age of a sample and is usually not ameliorated even following the use of extensive validated cleaning methods (**4**). This likely represents only the "tip of the iceberg" since samples which are suspected to yield mixtures are often not even submitted for analysis. A reliable means of fractionating the individual sequences within a mixture could greatly aid investigators by increasing the range of casework samples suitable for mtDNA testing.

Current Mixture Separation Strategies: The current PCR and DNA sequencing-based procedures employed for mtDNA analysis are very sensitive to minute amounts of DNA and mitotypes are characterized on the basis of a relatively small number of differences in base pair sequence. While this may be advantageous in cases with limited amounts of evidentiary material, heteroplasmic samples and evidentiary samples that present as a DNA mixture represent a significant challenge for analysis by direct sequencing. The ability to accurately resolve and interpret these types of samples in a timely and cost efficient manner would substantially increase the power of mtDNA analysis by allowing its use on more types of forensic cases where the current approach is limited or fails.

There are a number of molecular strategies that could be employed to separate DNA mixtures into their individual components. These include separation by denaturing gradient gel electrophoresis (DGGE) (5) or single-strand conformational polymorphism (SSCP) analysis (6-8) and subcloning into bacterial vectors. These approaches are generally time consuming, necessitate multiple handling steps, require laborious product purification and are not readily adaptable to automation. These factors have all been obstacles to the implementation of these technologies by forensic laboratories. Both DGGE and SSCP require manual recovery of fractionated DNA from polyacrylamide gels and a second round of PCR amplification to generate enough template for DNA sequencing. Subcloning is an even more time and labor-intensive approach. It would require forensic scientists to screen and sequence DNA from multiple transformed bacterial colonies to ensure that observed sequence differences reflect genuine contributors to the starting template rather than artificial variants that were introduced as a result of DNA nucleotide misincorporation by the *Taq* polymerase used for PCR.

Mixture Separation by Denaturing High-Performance Liquid Chromatography (DHPLC): DHPLC is an extremely accurate, rapid and cost-effective method for separating mtDNA mixtures into their individual components in preparation for subsequent characterization by direct DNA sequencing. While doing so, DHPLC addresses the limitations of the alternative strategies outlined above.

DHPLC enables the chromatographic fractionation of a mixture of nucleic acids based on nucleotide sequence and/or sequence length. Thus it is possible to clearly identify and separate two components of a DNA mixture even in cases where the two DNA molecules are identical in length but differ in sequence by as little as one base pair. The fundamental principle that makes this possible is Temperature Modulated Heteroduplex Analysis. Upon heat denaturation and renaturation of a mixture containing two (or more) DNA sequences, a combination of homoduplices and heteroduplices are produced. The homoduplices represent the original components of the mixture while the heteroduplices are formed as

the result of cross-hybridization between the different components of the mixture. Under partially denaturing conditions and an increasing gradient of a nonpolar solvent (*e.g.*, acetonitrile), it is possible to separate both the heteroduplices and the homoduplices from each other by HPLC (9) (See Figure 1 from Executive Summary). Heteroduplices, are inherently less stable as the result of one or more base pair mismatches and thus they appear as early eluting peaks in the chromatographic profile of a DHPLC assay (10, 11). The sensitivity of DHPLC ensures that even very subtle sequence differences (*i.e.*, single nucleotide substitutions, insertions, deletions) between the components of a mixture can be readily detected (12, 13) and thus isolated from each other for subsequent characterization by standard DNA sequencing.

The accurate detection of sequence differences between two or more components of a mixture is an essential first step in the separation and recovery of the individual contributors to a mixture. In comparative studies of DHPLC-based detection of sequence differences versus alternate sequence detection systems, DHPLC has been shown to provide better sensitivity and accuracy. Relative to DHPLC, the use of SSCP analysis has been shown to detect 50-97.5% of mutations depending upon the length and base sequence of the fragment being analyzed (14, 15). Similar results were found for other methods including DDGE (11, 16, 17). The overall concordance between sequence detection by DHPLC and direct DNA sequencing ranges from 95-100%. In our hands, simple adjustments to the DHPLC run conditions (temperature or acetonitrile gradient) have made it possible to correctly detect the presence of sequence polymorphisms in each of >1200 mixtures of amplicons from the HV1/HV2 regions of the mtDNA that have forensic relevance (18).

Time/Cost Savings of DHPLC vs. Alternative Strategies: Relative to gel-based electrophoretic methodologies (*e.g.*, DGGE and SCCP) or subcloning into bacterial vectors, DHPLC is a much more rapid and economical method of fractionating mixtures and recovering the individual components for subsequent analysis by DNA sequencing.

Following forensically validated standard operating procedures, a DHPLC assay to detect the presence of a mixture consumes approximately 25 ng of amplified DNA. In our hands, this is equivalent to 5µl of the standard 50 µl reaction volume. If desired, the detection sensitivity can be increased by up to100 fold if fluorescence detection is used in place of UV detection. For the complete fractionation of a 2-component mixture, 100 ng of PCR product yields excellent results though less can be used in cases where amplified DNA is limiting. It is important to emphasize that in the case of especially precious samples, DHPLC makes it possible to recover virtually all the DNA used in any assay. For example, the DNA used for the PCR yield analysis is recovered and used as the sequencing template to obtain the reference electropherogram for the original mixture. Similarly, any DNA used for a series of temperature scan assays can be recovered for use in subsequent assays. This is because although the eluent is normally directed to waste, the system can direct all DNA containing portions of the eluent to a chilled 96-well recovery plate.

The cost of separating and recovering the individual components of a DNA mixture by DHPLC is significantly less than it would be for the far more labor-intensive electrophoretic techniques or subcloning. Part of this is due to the low reagent cost of each

DHPLC assay (≈$0.50/injection including column replacement). Thus, the separation and recovery of a two-contributor mixture of DNA (including initial temperature optimization for maximum chromatographic peak separation) could be performed for less than $10 in reagent cost. All DHPLC reagents are now commercially available from Transgenomic, Inc. (Omaha, NE) at ready-to-use concentrations.

Understandably, there are a number of practical concerns for any forensic laboratory that would want to make use of DHPLC. The QA/QC of the WAVE instrument is initially certified and calibrated by the manufacturer using a commercial set of DNA mutation and molecular size standards. The instrument can be recalibrated by the user as needed or as part of a scheduled semiannual preventive maintenance service call which is included in all service contracts. As with any major instrument system, there are several levels of service contract that are available and the decision of whether or not to purchase coverage should reflect the instrument's level of use. As part of Phase 2 of the current project, we sought to develop an SOP for the internal validation of the instrument for forensic practitioners. This internal validation followed the revised SWGDAM validation guidelines. With respect to column use, we have not seen any difference in the quality of chromatographic fractionation between a column that is kept continuously on the instrument with a continuous flow of solvent and a column that is used sporadically and which is removed from the instrument during periods of non-use and stored at room temperature. This issue, is one that was addressed as part of the developmental validation of DHPLC technology in Phase 2 of this project. Specifically we tracked the reproducibility of a defined set of mtDNA mixtures representing each of the four forensically relevant HV1 and HV2 amplicons for the duration of the project. With regard to the life span of the DNAsep columns, there is no manufacturer specified expiration date. Rather each column is warranted for 4000 injections if users prepare their own DHPLC buffers and 6000 injections if users purchase commercially prepared buffers at 1X concentration from the manufacturer. There is not a significant cost advantage to users who prepare their own buffers.

With almost any approach to mixture separation, however, the cost of reagents is generally small compared to the cost of skilled labor. The labor cost savings associated with DHPLC is one of its most attractive features. This is because virtually all steps in a DHPLC assay from sample injection to the recovery of DNA from individual contributors to a mixture can be automated without risk of cross contamination (19, 20). Even under stringent conditions, a DHPLC assay takes less than 7 minutes from start to finish. Thus the separation and recovery of a two-contributor mixture of DNA (including optimization for maximum chromatographic resolution) can be achieved in as little as 2 hours of instrument run time with minimal hands-on manipulation of the sample. Seven minutes is the time required to fractionate a mixture of one of the four mtDNA amplicons used in forensic assays. Two hours is the total amount of DHPLC time required to quantify the DNA yield from the initial amplification, screen for the presence of a mixture, fractionate each of the four fragments and wash the column with 100% acetonitrile between fractionations to ensure that there is no carry over of residual DNA between fractionations for all four mtDNA fragments.

By contrast, electrophoretic methodologies (*e.g.*, DGGE and SCCP) require tedious manual recovery and reamplification of small amounts of fractionated DNA from

acrylamide gels. This requires significant hands-on manipulation by a skilled analyst and thus significantly higher labor costs. This is particularly true in the case of subcloning DNA mixtures into bacterial vectors where it is necessary to purify amplified PCR products; ligate them into a plasmid cloning vector; transform competent cells; screen for transformants, streak the transformants out to single colonies and; sequence multiple colonies to ensure that sequence variants reflect real sequence differences rather than artificial variants generated through nucleotide misincorporation by DNA polymerases. This not only translates into greater reagent and labor costs, but the increased manipulation of the sample greatly increases the opportunity for inadvertent contamination of the sample. With DHPLC-based mixture separation, the fractionated products are sequenced directly without extensive manipulation or reamplification.

The equipment start up cost for DHPLC-based mixture separation is also quite reasonable compared to other analytical instruments used in forensic DNA laboratories. For example, the list price of a fully loaded WAVE® Nucleic Acid Fragment Analysis system with a computer, all necessary software, an autoinjector, a UV detector, an integrated fraction collector, an optional fluorescence detector/HSX unit and on-site training is $135K. This is less expensive than the ABI 3130 Genetic Analyzer commonly used with the AmpFLSTR® Cofiler, Profiler Plus, and Identifiler kits. This puts a commercial DHPLC instrument within the reach of most crime laboratories especially those specializing in mtDNA.

Training for DHPLC-Based Mixture Resolution: DHPLC technology is also extremely user friendly. Using a commercially available DHPLC system from Transgenomic Inc. (Omaha, NE), it is possible for a forensic scientist with a week of training to competently manage all basic operational aspects of the instrument. The instrument's operating software, integrates HPLC control and data acquisition with instrument functions.

Based on our extensive experience with DHPLC and training in forensic sciences we anticipate that training for DHPLC-based mixture separation will be no more extensive than the mtDNA sequencing course offered by the FBI. As with any technology, however, appropriate internal validation studies will need to be conducted in each crime laboratory. In fact, one of the objectives of phase one was to begin to develop standard operating procedures (SOP) for DHPLC mixture fractionation and DNA recovery to facilitate the forensic validation of DHPLC in accordance with DAB standards for developmental and internal validation of the technology.

Research Program Objectives - In the current research program we sought to evaluate the potential forensic utility of DHPLC for the analysis and resolution of mtDNA mixtures. To achieve this, our research activities encompassed the following specific objectives:

1) **Rigorously Demonstrate** the reliability of DHPLC to detect and chromatographically resolve mixtures involving all classes of human mtDNA HV1 and HV2 polymorphisms.

2) **Rigorously Demonstrate** the reproducibility of DNA sequencing electropherograms with an emphasis on the statistical significance of any correlation between peak height ratios and DNA quantity ratios at mixed base positions.

3) **Rigorously Demonstrate** the reliability of determining the individual haplotypes in a DNA mixture through linkage phase analysis of sequence data from DHPLC fractionated mixtures.

4) **Develop** standard operating procedures for DHPLC mixture resolution and work toward interpretation guidelines.

[Analytical Methods]

Human Subjects **-** This research was conducted in compliance with U.S. Federal Policy for the Protection of Human Subjects (Basic DHHS Policy for Protection of Human Research Subjects; 56 FR 28003). Individual buccal swabs were collected from 96 unrelated subjects of diverse ethnicity who had previously provided informed consent to participate in the study. Additional samples of human tissues for casework-type samples used for developmental validation studies (*e.g.*, blood, saliva, semen and hair) were collected from unrelated subjects who had previously provided informed consent to participate in the study. Buccal swabs, peripheral blood, saliva, semen and hair (0.5cm with root) were collected from individuals who had provided informed consent to participate in the study. To investigate the influence of PCR inhibition on DHPLC-based mixture fractionation and analysis, simulated casework samples were prepared by applying 10µL aliquots of whole blood to Pur-Wraps® sterile cotton tipped applicators (Hardwood Products, Guilford, ME) that had previously been dipped in a slurry of top soil in nanopure water (50:50 w/v) and then allowed to air dry for 2 hours at room temperature. To test the potential impact of environmental contaminants, 10µL aliquots of undiluted blood or semen were applied to sterile cotton tipped applicators that had previously been dipped in laundry detergent, used motor oil, gasoline, sodium hydroxide or glacial acetic acid and allowed to dry for 2 hours. All samples were prepared in triplicate and then stored at -20°C until DNA quantification. Additional simulated casework included cigarette butts from a study participant and 100µL aliquots of blood saliva or semen that had been applied to nylon carpeting, cotton cloth or blue denim fabric.

Mitochondrial DNA Extraction, Purification and Amplification **-** All DNA samples were extracted using the EZ1 DNA Tissue kit run on the BioRobot® EZ1 with DNA Forensic Card (Qiagen Inc., Valencia, CA) in accordance with the manufacturer's recommend protocol for DNA extraction from trace samples. To maximize the recovery of genomic DNA, whole swabs, or cuttings from stained blue denim, cotton or nylon carpeting) were immersed in 190µL of the Qiagen's proprietary "G2 Buffer" and 10µL of proteinase K solution (600mAU/mL), vortexed for 10 sec. and incubated at 56˚C for 15 min. The digested supernatant was transferred to a sterile 2mL sample tube and loaded onto the BioRobot® EZ1. Following automated DNA extraction and elution of the DNA into 200µL of TE (10mM Tris, 1mM EDTA), each tube was visually inspected for the presence residual paramagnetic beads which have the potential to interfere with qRT-PCR quantification. When detected, the beads were removed by brief centrifugation and transfer to of the cleared supernatant to a new sterile 2mL sample tube.

Validated primer pairs (Table 1) and PCR amplification conditions were used to amplify four forensically relevant regions of the human mitochondrial control region (*i.e.*, HV1A,

HV1B, HV2A and HV2B amplified by primer sets, A1/B2, A2/B1, C1/D2 and C2/D1, respectively). Two minor modifications to these procedures were made, the exclusion of bovine serum albumin (BSA) from the amplification reaction and the supplementation of AmpliTaq GOLD® with higher fidelity *Pfu* DNA polymerase.

The presence of >10% w/v BSA in a sample reduces the functional lifespan of the DNASep® column. BSA, however, is often used to ameliorate the effects of endogenous and environmental PCR inhibitors in some DNA extracts. These include IgG and heme in blood [23], melanin in hair [24], collagen in bone [25] and such environmental inhibitors as humic acid in soil [26] which are not efficiently removed by inorganic extraction methodologies (*e.g.*, Chelex™). In our laboratory, both endogenous and environmental inhibitors have been effectively removed from hair, bone and environmentally contaminated samples by the use of organic extraction, silica-gel membranes or by systems that employ paramagnetic particles such as the Qiagen® EZ1 DNA tissue kit. If the use of BSA is unavoidable, purification of the PCR amplification products by organic extraction or other protein removal method is recommended prior to DHPLC analysis.

The fidelity of the DNA polymerase used for amplification may also impact the accurate interpretation of DHPLC chromatograms. The relatively high nucleotide misincorporation rate of Taq DNA polymerase (8.0×10-6 mutations/bp/duplication) [27] typically produce DHPLC chromatograms from single sequence samples which erroneously appear to indicate a DNA mixture. A mixture of *Taq* and *Pfu* DNA polymerases at a 9:1 ratio effectively eliminates these anomalies. The 3′→5′ exonuclease activity and lower nucleotide misincorporation rate (1.3×10-6 mutations/bp/duplication) of Pfu yields greater uniformity in amplification products [27].

Table 1. Validated PCR Primers Targeted to mtDNA HV1 and HV2

Hypervariable Region 1A (HV1A)		
Primer A1	(L15997)	**5′ CAC CAT TAG CAC CCA AAG CT 3′**
Primer B2	(H16236)	**5′ CTT TGG AGT TGC AGT TGA TG 3′**
Hypervariable Region 1B (HV1B)		
Primer A2	(L16159)	**5′ TAT TTG ACC ACC TGT AGT AC 3′**
Primer B1	(H16391)	**5′ GAG GAT GGT GGT CAA GGG AC 3′**
Hypervariable Region 2A (HV2A)		
Primer C1	(L48)	**5′ CTC ACG GGA GCT CTC CAT GC 3′**
Primer D2	(H285)	**5′ GGG GTT TGG TGG AAA TTT TTT G 3′**
Hypervariable Region 2B (HV2B)		
Primer C2	(L172)	**5′ ATT ATT TAT CGC ACC TAC GT 3′**
Primer D1	(H408)	**5′ CTG TTA AAA GTG CAT ACC GCC A 3′**

Amplification reactions were purified by denaturing high-performance liquid chromatography on a WAVE® DNA Fragment Analysis System (Transgenomic Inc., Omaha, NE). Briefly, unpurified PCR products were bound to a DNAsep® analytical column at 50ºC in the presence of 0.1M triethylamine acetate (TEAA) pH 7.0 as an ion pairing reagent. Target PCR amplicons were selectively eluted using a linear gradient of increasing acetonitrile produced by differential mixing of buffer A (0.1M TEAA) and

buffer B (0.1M TEAA: 25% ACN). The optimal gradients employed a 56% to 65% buffer B increase in 3.5 minutes for HV1A, HV2A and HV2B and a 55% to 64% buffer B increase in 3.5 minutes for HV1B. All samples were eluted at 0.9ml/min flow rate and detected by UV at 260nm. Eluted peaks were captured using an automated fraction collector; dried by vacuum centrifugation (30 minutes at 50ºC, followed by a 20-minute cool down to ambient); resuspended in ddH_2O; quantified by 260 / 280nm UV spectrophotometry and stored frozen until dideoxy terminator labeling.

DHPLC-Based Temperature-Modulated Heteroduplex Analysis (TMHA) and Mixture Fractionation **-** DNA sequence data for each of four forensically-relevant amplicons (HV1A, HV1B, HV2A and HV2B) were compared to identify individuals with different mtDNA haplotypes. These were combined at a 1:1 molar ratio to generate a series of pair-wise mixtures which were cross-hybridized by denaturation at 95˚C for 4 minutes and gradual cooling (1.5˚C/min) to a final temperature of 25˚C.

The cross-hybridized mixtures were analyzed by TMHA on a WAVE® 3500HT DNA Fragment Analysis System (Transgenomic Inc., Omaha, NE) using a DNASep® analytical column. Triethylammonium acetate (TEAA) pH 7.0 at a final concentration of 0.1M served as an ion-pairing reagent. Cross-hybridized mixtures were analyzed under partially denaturing conditions at the empirically determined optimal temperatures for each of the four mtDNA amplicons (*i.e.*, HV1A, 58˚C; HV1B, 59.2˚C; HV2A, 56.5˚C and HV2B, 57˚C). The optimal acetonitrile linear gradient generated from differential mixing of buffer A (0.1M TEAA) and buffer B (0.1M TEAA, 25% ACN) were a 56% to 65% buffer B increase in 3.5 minutes for HV1A, HV2A and HV2B and a 55% to 64% buffer B increase in 3.5 minutes for HV1B. All samples were eluted at a 0.9ml/min flow rate and detected by UV absorbance at 260nm. Fractions were recoverd into 96-well round-bottom plates using Transgenomic Navigator software in combination with an automated fraction collector. Standard DHPLC controls included: zero-volume injections to screen for DNA carryover between assays; no-template PCR controls to check for reagent contamination; and manufacturer mutation control standards for buffer and column quality control. Where necessary for enhanced sensitivity, temperate titrations were conducted in ±0.5˚C increments.

Mitochondrial DNA Sequencing **-** DNA sequencing reactions were prepared for each of four control region amplicons. Use of the eight aforementioned PCR primers for labeling allowed each amplicon to be sequenced in both the forward and reverse directions. Reactions prepared using the Applied Biosystems BigDye® Terminator v1.1 cycle sequencing kit used five different DNA input quantities (0.1ng, 0.2ng, 0.4ng, 0.8ng and 1.6ng) while those prepared using the BigDye® Terminator v3.1 cycle sequencing kit used three different quantities (0.4ng, 0.8ng and 1.6ng). For all samples and conditions tested, five independently-labeled replicate 10µl reactions were prepared in accordance with the manufacturer's protocol. Appropriate positive and negative controls were carried through the entire sample handling process to monitor for contamination and reagent integrity.

Dideoxy-terminated products were purified by EdgeBio Performa DTR V3 96-well short plate purification columns (Edge Biosystems, Gaithersburg, MD) per the manufacturer's protocol and resolved on an Applied Biosystems Prism 310 Genetic Analyzer using POP 6 polymer and 47cm x 50µm capillaries. The raw electrophoretic traces were analyzed using

the KB basecaller together with the dye set mobility files indicated for the labeling kit used. The resulting sequence data were analyzed using the Sequencher™ DNA analysis software (Gene Codes Corp, Ann Arbor, MI).

Quantitative data on peak heights for statistical analyses of electrophoretic patterns were obtained by first analyzing raw electrophoretic data using the KB basecaller which is part of the Prism 310 Sequencing Analysis Software version 5.1 (Applied Biosystems). The resulting .scf output file was then ported to the CEQ™ 8000 Genetic Analysis System version 8.0 software package (Beckman-Coulter, Fullerton, CA) to allow output of the electrophoretic peak height data for each channel and time point as a .txt file. Text files were then opened in Microsoft Excel for subsequent statistical analyses. The height of each peak in relative fluorescence units (RFU) was compared across the five replicates for each amplicon, input DNA quantity, and sequencing chemistry tested.

[**Results and Conclusions]**

The reliability of DHPLC to detect and chromatographically resolve mixtures involving all classes of human mtDNA HV1 and HV2 polymorphisms

The critical first step in DHPLC-based mixture fractionation is the determination that a given sample contains a mixture of two or more DNA sequences. When there is a mixture of at least two DNA amplicons that differ in primary base sequence, a combination of homo- and heteroduplices will be produced upon renaturation of a heat-denatured sample. Subsequent DHPLC analysis under a partial denaturing temperature results in the differential elution of homo- and heteroduplices from the DHPLC column. Namely, heteroduplex molecules denature and elute at lower concentrations of acetonitrile than do corresponding homoduplices. Accordingly, the presence of sequence differences (*e.g.*, single nucleotide substitutions, insertions, deletions) between DNA molecules is readily indicated by the appearance of one or more early-eluting peaks in the chromatograms. The exact retention time at which individual heteroduplex and homoduplex peaks resolve is determined by the base composition as well as by the number and type of nucleotide difference(s) between the molecules being assayed. As a result, DNA mixtures are indicated by the appearance of one or more early-eluting heteroduplex peaks (See Figure 1 from Executive Summary).

Optimization of DNASep® Column Assay Temperatures **-** Comparative DHPLC-based DNA sequence analysis relies on temperature-dependent chromatographic separation of the nucleic acids under partially denaturing conditions. In accordance with manufacturer recommendations and empirical observations by the authors and other researchers, optimal resolution is generally obtained using amplicons of less than 400bp in length and a column temperature where the average of nucleic acid helicity across an amplicon is approximately 75% to 85% [28, 29]. Under these conditions, heteroduplices with a single base mismatch elute from the DNASep® column before identical-length homoduplices.

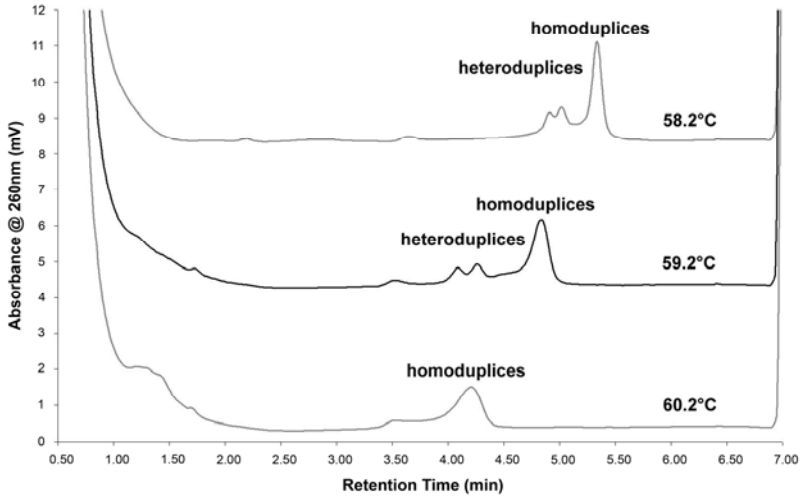


Figure 5: Empirical determination of the optimal column temperature for comparative sequence analysis of HV1B. While the hetero- and homoduplices can be resolved at 58.2ºC, optimal resolution of clearly discernible chromato-graphic peaks is obtained at 59.2ºC. At 60.2ºC, the heteroduplices elute too readily to form distinct peaks.

The mtDNA primer pairs used in the current study yield amplicons within a size range that are well suited for DHPLC analysis (*i.e.*, HV1A = 278bp, HV1B = 271bp, HV2A = 278bp, HV2B = 277bp). Although the appropriate partially denaturing tempera-ture for a given amplicon is approximated by the system software based on nearest neighbor interactions [30], the optimal temperature must be empirically confirmed. This was done by monitoring the separation of homo- and heteroduplices in multiple mixed samples at temperatures flanking the predicted value. The optimal temperature is that at which the greatest chromatographic resolution is achieved. This is illustrated in Figure 5 where the greatest resolution for the mixture of HV1B amplicons being assayed is at 59.2˚C. This is not to suggest that all mixtures of HV1B amplicons will be optimally resolved at this temperature. Rather, this represents the optimal initial assay temperature at which mixtures of HV1B amplicons of unknown haplotype can be screened. Similarly, the optimal initial assay temperatures for the HV1A, HV2A, and HV2B amplicons were determined to be 58.0˚C, 56.5˚C and 57.0˚C, respectively.

Reproducibility of DHPLC Chromatograms – The reproducibility of DHPLC generated chromatographic peaks is also essential to the reliability of the assay as a tool for the presumptive detection of sequence identity/non-identity. To assess the reproducibility of peak height retention time and peak area across multiple injections, ten replicate injections of a two-component HV2A mixture were sequentially assayed by DHPLC under the optimized conditions described above. The resulting data which encompass two early eluting heteroduplex peaks and two late eluting homoduplex peaks indicate that DHPLC analyses are highly reproducible in terms of both the maximum height and retention time of each individual peak (Table 2). Assay-to-assay variability was negligible between independent injections as evidenced by the small standard deviations in the parameters measured for each peak across the replicates tested. It should be noted that the excellent reproducibility of DHPLC assays does not require the use of added retention time standards but rather reflects the intrinsic reproducibility of the assay system. Commercial calibration standards are used, however, to regularly confirm and, if needed, to calibrate the DNASep® column oven.

Table 2 – Ten sequential injections of a two-contributor HV2A mixture showing the reproducibility of DHPLC assays on a DNASep® column.

HV2A		Injection											
		1	2	3	4	5	6	7	8	9	10	Average	Std Dev
Peak 1	Peak height	2.82	2.82	2.78	2.75	2.76	2.79	2.75	2.78	2.74	2.76	2.82	0.004
	Ret. Time	3.49	3.47	3.46	3.48	3.47	3.47	3.48	3.48	3.47	3.48	3.48	0.01
Peak 2	Peak height	8.34	8.53	8.37	8.31	8.44	8.28	8.41	8.37	8.34	8.47	8.39	0.08
	Ret. Time	3.83	3.81	3.81	3.83	3.82	3.81	3.82	3.83	3.83	3.81	3.82	0.01
Peak 3	Peak height	4.50	4.51	4.47	4.51	4.48	4.47	4.49	4.45	4.45	4.51	4.48	0.02
	Ret. Time	4.04	4.02	4.02	4.04	4.03	4.03	4.03	3.97	4.04	4.02	4.02	0.02
Peak 4	Peak height	11.13	11.13	10.91	11.02	11.22	11.25	11.33	11.11	11.31	11.35	11.17	0.14
	Ret. Time	4.38	4.37	4.36	4.38	4.37	4.37	4.37	4.39	4.38	4.36	4.37	0.01

DHPLC Detection of mtDNA Mixtures **-** A major objective of the this research program was to evaluate the potential utility of DHPLC as a means of assaying amplified human mtDNA samples for sequence concordance/non-concordance in the context of a forensic laboratory. Sequence non-concordance within a sample would be consistent with either naturally occurring heteroplasmy or a situational mixture of two or more contributors. Initial experiments using subcloned mtDNA haplotypes (data not shown) demonstrated that DHPLC accurately detects even single base differences between two mtDNA amplicons. Subclones, however, are artificially pure templates which do not reflect the molecular complexity of a human tissue-derived DNA extract. To more thoroughly assess the reliability of DHPLC as a tool for comparative sequence analysis, the approach was tested on blood samples from 95 research volunteers.

Exclusive of length polymorphisms associated with the homopolymeric cytosine-stretch in HV2, the 95 subjects in this phase of the study represented 83 distinct haplotypes. Seventy-seven of these were unique within the study population. Of the 6 haplotypes which occurred more than once, the most common (263G, 315.1C) was observed 8 times while the remaining 5 haplotypes each occurred twice. Relative to the revised Cambridge Reference Sequence (rCRS) [31, 32], the haplotypes represented in the current study encompass 84 polymorphisms in HV1 and 46 polymorphisms in HV2, including cytosine-stretch length polymorphisms.

Based on pair-wise comparisons, individual haplotypes differed from each other at 0-22 positions (0-11 in HV1A; 0-13 in HV1B; 0-13 in HV2A; and 0-12 in HV2B). On average, there were 8.71 positional differences between haplotypes. *In toto*, the study population encompassed a broad diversity of haplotypes and thus is well suited for evaluating the utility of DHPLC for the accurate detection of sequence polymorphisms throughout the mtDNA control region. This is essential for the validation of DHPLC as a tool for comparative sequence analysis.

A total of 920 pair-wise combinations of amplicons from the 95 individuals in this study were prepared, denatured and allowed to gradually reanneal. Of these, 72 (22 in HV1A, 8 in HV1B, 17 in HV2A, and 25 in HV2B) represented combinations of amplicons that were from different individuals but which had identical DNA sequences. DHPLC analyses of these samples all produced clear chromatograms consisting of a single symmetrical homoduplex peak. This pattern is consistent with sequence concordance and is in 100% agreement with direct sequencing data for these amplicons.

To assess the reliability of DHPLC to detect sequence non-concordance, 849 combinations of amplicons (209 in HV1A, 222 in HV1B, 213 in HV2A, and 205 in HV2B) which differed in sequence were assayed. These differences encompassed a broad diversity of polymorphisms distributed throughout the HV1 and HV2 regions including transitions, transversions, insertions and deletions (Table 3). Positional differences were located internally as well as near the termini of amplicons and encompassed regions of varying GC richness. In all, the mixtures assayed in this study included sequence variants at 39, 62, 38 and 30 different positions in HV1A, HV1B, HV2A and HV2B, respectively. Figure 6 illustrates the distribution of the variant positions and the frequency with which they were assayed.

Table 3 - *Sequence polymorphisms assayed by DHPLC in the HV1 and HV2 regions.*

	HV1	HV2
Transitions	16041G 16051G 16069T 16093C 16104T 16111T 16114T 16124C 16126C 16129A 16136C 16144C 16145A 16147T 16148T 16153A 16154C 16163G 16172C 16173T 16176T 16179T 16186T 16187T 16189C 16192T 16193T 16209C 16215G 16217C 16221T 16222T 16223T 16224C 16239T 16241G 16243C 16244A 16249C 16256T 16258G 16261T 16266T 16270T 16274A 16278T 16286T 16287T 16290T 16291T 16293G 16294T 16295T 16296T 16298C 16299G 16301T 16304C 16309G 16311C 16316G 16320T 16325C 16327T 16343G 16352C 16353T 16354T 16355T 16357C 16359C 16362C 16390A	55C 64T 72C 73G 89C 92A 93G 114T 131C 143A 146C 150T 152C 153G 185A 185A 188G 189G 192C 194T 195C 196C 198T 199C 200G 204C 207A 217C 222T 225A 226C 228A 234G 239C 242T 250C 285T 293C 295T 319C 340T
Transversions	16111A 16129C 16176A 16182C 16183C 16184A 16265C 16290G	57G 207C 280G
Insertions	16192.1T 16193.1C	309.1C 309.2C
Deletions		249: 299: 309:

Using the aforementioned initial assay temperature for each mtDNA amplicon, DHPLC analyses correctly indicated the presence of a mixture of non-concordant amplicons in 836 (209 in HV1A, 222 in HV1B, 203 in HV2A, and 202 in HV2B) of the combinations tested. The remaining 13 mixtures (10 in HV2A and 3 in HV2B) of non-identical amplicons, yielded chromatographic traces with a single eluent peak, a result erroneously suggesting sequence concordance. Careful examination of the amplicon mixtures which were not detected by DHPLC reveals that these aberrant results are limited to a very small number of challenging positions. Mirroring results obtained with subcloned mtDNA, it was not possible to detect as non-concordant, combinations of HV2A amplicons that differ only at positions 72 or 73. Taken together, these two positions account for all of the undetected non-concordant mixtures in HV2A.

In HV2B, DHPLC was not able to detect as non-concordant, combinations of amplicons that differed only at position 295. This position lies in a narrow stretch of sequence immediately adjacent to a large GC-rich region. Given the thermodynamic stability of this region, a single base mismatch may not sufficiently destabilize the surrounding helix such that an early eluting peak can be discerned. This is a postulate supported by the observation that mixtures of amplicons that possess an additional mismatch in this same region are readily detected.

An inverse relationship was generally observed between the number of positional differences associated with a given pair of non-identical amplicons and the relative heights of the hetero- versus homoduplex peak(s). The height and retention time of a heteroduplex peak is a function of the stability and base sequence of the helix. The more stable a heteroduplex, the more readily it should form relative to the competing homoduplices and

thus the greater its peak height (indicative of quantity) on the DHPLC chromatograms. The corollary of this is that the formation of less stable helices is less favored and should be associated with smaller and earlier-eluting peaks on the DHPLC chromatograms. In theory, this could compromise the ability of DHPLC to detect as non-concordant some combinations of amplicons that differ at a large numbers of positions. In the current study, however, no examples of such "heteroduplex dropout" were observed.

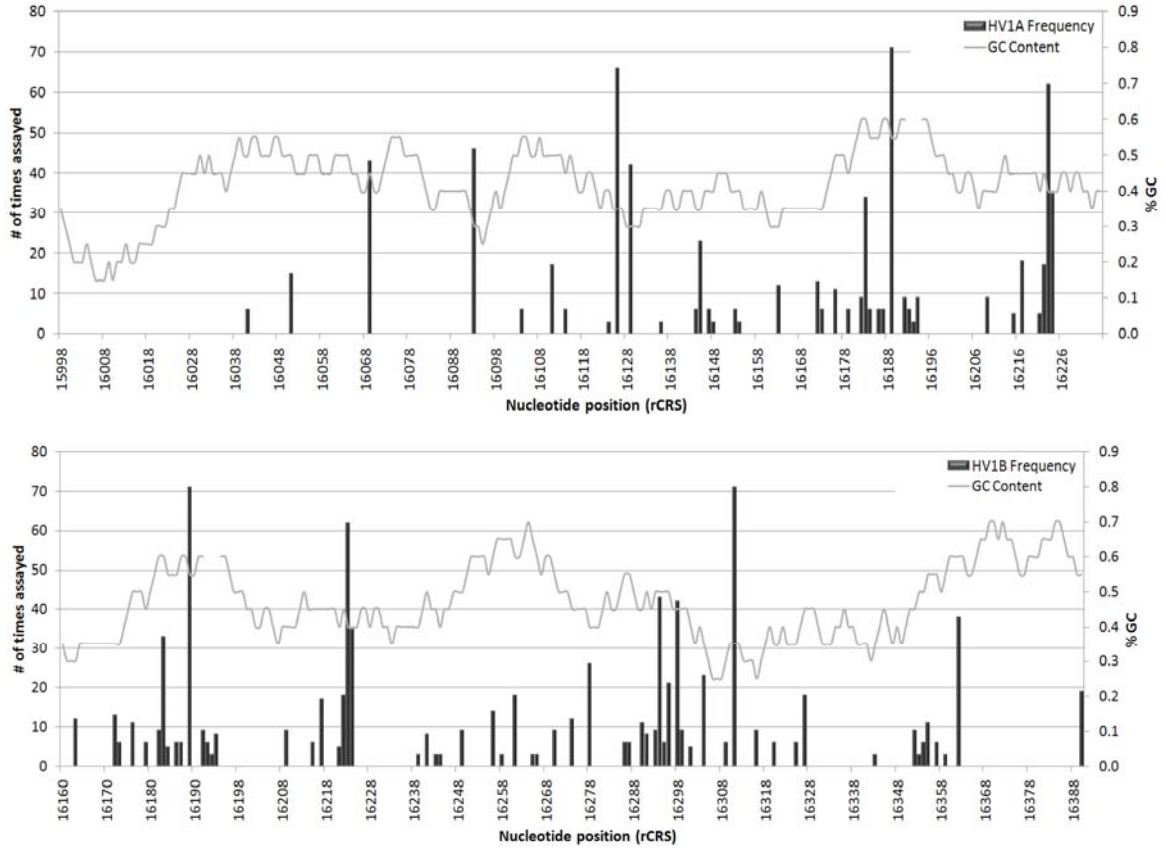


Figure 6: A histogram showing the distribution of nucleotide positions within the HV1A, HV1B, HV2A and HV2B amplicons and the number of times that sequence polymorphisms at each position were assayed by DHPLC. The polymorphisms assayed are broadly distributed throughout both GC and AT rich regions as illustrated by the light gray line which indicates %GC content across each fragment.

Heteroplasmy Detection **-** As a naturally occurring combination of two haplotypes, heteroplasmy represents an important but potentially confounding factor for DHPLC-based analyses. Chromatograms obtained from assays of known heteroplasmic samples typically show the presence of early-eluting heteroduplices. In these cases, DHPLC can often provide an indication of the nature of the heteroplasmy. Using either non-denaturing (column temperature = 50ºC) or partially denaturing conditions, length heteroplasmy results in a chromatogram characterized by a markedly broadened peak often characterized by multiple shoulders (Figure. 7). This represents the combination of the various length amplification products and the characteristic shape of such peaks makes it possible to readily screen for heteroplasmy associated with the cytosine-stretches of HV1 and HV2B. Using partially denaturing conditions, point (*i.e.*, base substitution) heteroplasmy can also be readily detected. It is typically indicated by the presence of a

defined heteroduplex peak or a significant shoulder in association with the main homoduplex peak.

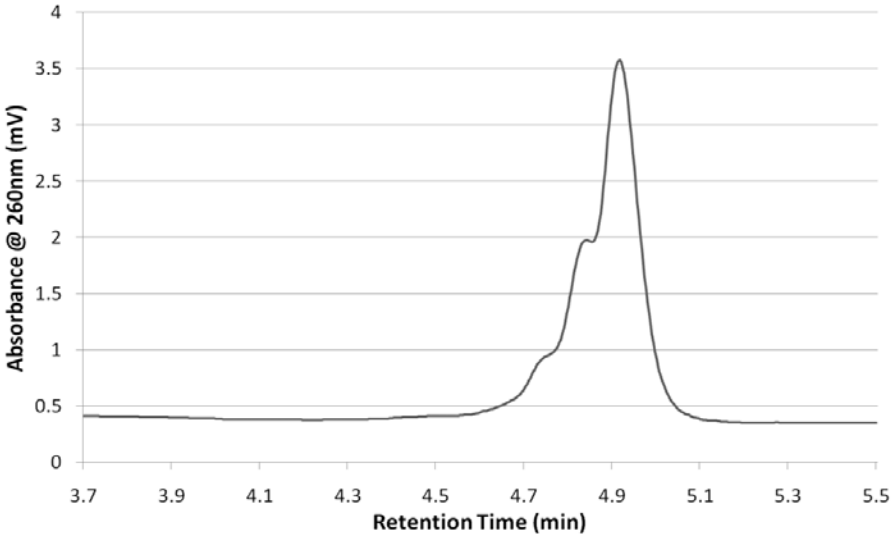


Figure 7: An example of length heteroplasmy as detected by DHPLC at a non-denaturing DNASep® column temperature of 50°C. Secondary and tertiary length variants are detected as shoulders prior to the primary length

When analyzing questioned samples, however, it is not possible to reliably distinguish between heteroplasmy and a situational mixture where mtDNA has been contributed by two or more individuals. Heteroplasmy, however, can reduce the discriminatory power of DHPLC in some cases. In the current study, this was most often observed for samples with significant length heteroplasmy. The broad chromatographic peak associated with these samples can obscure the presence of additional heteroduplices formed as a result of cross hybridization between the amplicons present.

While DHPLC circumvents many of the limitations of alternate approaches to detecting mtDNA mixtures, it is important to consider very carefully the types of samples for which such an approach might be indicated. Within an mtDNA sequencing laboratory, screening by DHPLC makes it possible to detect samples that contain mixtures of non-concordant amplicons immediately after PCR amplification and without having to sequence them. For both heteroplasmic and situational mixtures this approach would enable the analyst to identify potentially challenging samples and mark them for “special handling” – whether that be use of alternate sequencing primers to avoid C-stretch polymorphisms or the application of emerging technologies for resolving mixed samples [33-35].

The reproducibility of DNA sequencing electropherograms and the statistical correlation between peak height ratios and DNA quantity ratios at mixed base positions

Electrophoretic Relative Peak Height Consistency **-** Peak heights on sequencing electropherograms reflect the quantity of DNA molecules terminated at a given position. This is a function of the sequence context of each base which influences the efficiency of dye terminator incorporation. As a result, DNA sequence data are characterized by electrophoretic peaks for individual nucleotide position which may differ significantly relative to each other. If DNA sequence data are to be quantitatively compared as would be the case among sequential DHPLC generated fractions of an mtDNA mixture, the reproducibility of the relative differences in peak heights across an electropherogram is essential.

Therefore, the extent to which relative differences in peak heights throughout the entire HV1 and HV2 regions was conserved was one of the first issues that we sought to resolve. We also sought to determine whether any specific sites appeared to display significantly greater variability. Our preliminary data showed that the height of any given electrophoretic peak relative to the height of surrounding peaks was highly reproducible and that even the degree of variability across a range of input DNA concentrations was minimal.

The reproducibility of electrophoretic traces across the mtDNA HV1 and HV2 region was evaluated. A total of five different DNA template input quantities (*i.e.*, 0.1ng, 0.2ng, 0.4ng, 0.8ng, 1.6ng); encompassing a range frequently encountered by forensic practitioners were used. For these analyses, each of four forensically-relevant mtDNA amplicons was sequenced in both the forward and reverse directions. The resulting eight dye terminator labeling reactions were independently replicated five times each for statistical purposes.

Figure 8 presents an example of peak height and variability data (RFU ± SD) for nucleotide positions 140 - 180 of HV2A at DNA input quantities of 0.4ng, 0.8ng and 1.6ng. Across these DNA input quantities, the heights of individual electrophoretic peaks and, therefore, the overall “pattern of relative peak height differences” was highly reproducible. This is exemplified by the consistently low peaks for 147C and 164C versus the consistently high peaks for 143G and 160A. Although the data presented in Figure 8 represent a small portion of the HV2A amplicon, they are consistent with the results that have been obtained for the entire HV1 and HV2 region. The sequencing traces obtained using lower DNA input quantities (0.1ng and 0.2ng) were similarly consistent (data not shown).

Impact of DNA input quantity on peak height reproducibility **-** Shown in Table 4 are peak height variability data for each of the eight labeling reactions used to sequence the HV1 and HV2 regions. These values reflect the variability in peak height averaged across all the peaks in the amplicon sequenced. Also shown is the combined peak height variability average of all sequencing reactions at each of five DNA input quantities. In order to compensate for the often significant differences in peak heights between any two positions, variability data were expressed as a percentage of peak height.

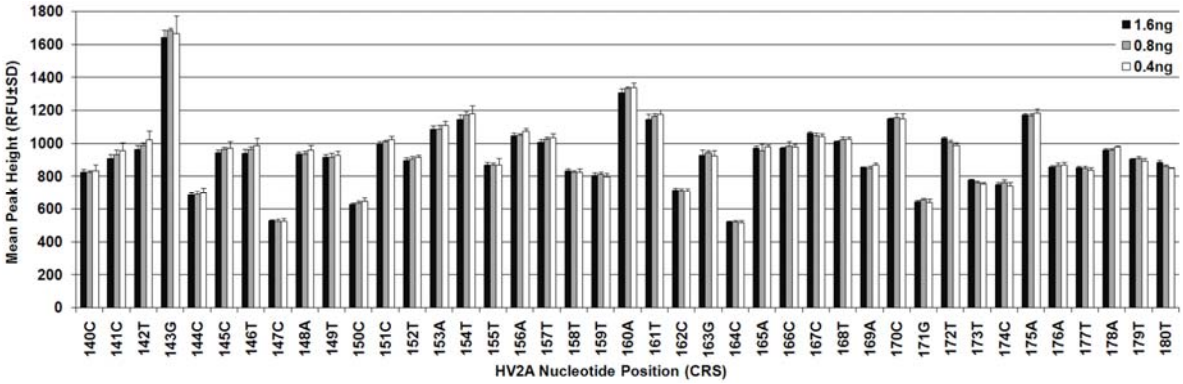


Figure 8: Reproducibility of electrophoretic peak heights for positions 140 to 180 of the mtDNA HV2A amplicon. The data represent the mean peak heights ± standard deviations from five independent labeling reactions for each of the three DNA input quantities. Although there are significant differences in mean peak heights between different nucleotide positions, the mean peak heights for each individual nucleotide position are highly consistent across the range of DNA input quantities shown.

Table 4 - Total variability (standard deviation normalized to individual peak heights) associated with direct sequence results for each of the four forensically relevant mtDNA amplicons which encompass HV1 and HV2.

DNA Quantity	A1	B2	A2	B1	C1	D2	C2	D1	Average
0.1ng	6.0%[a]	8.6%[a]	8.5%	3.6%[a]	5.3%[a]	3.8%	5.4%	10.0%[a]	**6.4%**
0.2ng	2.7%[a]	6.5%[a]	4.4%	2.6%[a]	4.3%[a]	5.7%	5.1%	7.5%[a]	**4.9%**
0.4ng	2.7%	5.7%	3.5%	2.2%	5.6%	1.4%	8.6%	8.3%	**4.8%**
0.8ng	2.2%[b]	8.6%	4.6%	3.2%	2.3%	1.4%	3.4%	6.0%	**4.0%**
1.6ng	7.2%[b,c]	7.5%[b]	2.6%[b]	1.5%[b]	2.2%[b,c]	3.5%[b]	4.3%	8.0%[b]	**4.6%**

[a] elevated noise at the beginning of the electropherogram

[b] signal deterioration at the end of the electropherogram

[c] signal saturation across the electropherogram

Using the BigDye® Terminator v1.1 cycle sequencing kit, 0.8ng of input DNA yielded the lowest overall variability (±4.0%) across all amplicons. The lowest input DNA quantity tested (0.1ng) had the highest average peak height variability at ±6.4%. At lower input DNA quantities (0.1ng and 0.2ng), there was a marked decrease in the signal-to-noise ratio particularly at the beginning of an electropherogram. Conversely, electropherograms generated using 1.6ng of input DNA were frequently characterized by fluorescence signal saturation, especially of G and T peaks and/or a marked degradation in peak resolution near the end of the trace.

DNA template quantities have a significant influence on peak height reproducibility and sequence quality. Across all four forensically-validated mtDNA amplicons, 0.4 - 0.8ng of input DNA yielded the lowest average peak height variability. Electropherograms generated using these optimal input quantities showed minimal noise associated with unincorporated dyes at the beginning of the sequence and minimal signal deterioration at the end of the sequence. Elevated peak height variability and lower signal-to-noise ratios

were frequently seen with low DNA input (0.1 - 0.2ng) while signal saturation at high input (1.6ng) produced false peaks due to spectral pull up.

Deconvolution of saturated electrophoretic peaks, which is required for basecalling, results in both an underestimation of a peak's true height and greater peak height variability. Signal saturation also produces pull-up of the overlapping dye spectra. This artifact remains following application of the dye matrix and typically results in the appearance of false peaks immediately before or after the saturated peak. Figure 9 illustrates this in a comparison of a 43bp stretch of HV1A sequence data for two DNA input quantities (1.6ng and 0.8ng). Signal saturation seen in the raw data for the 1.6ng reaction (Figure 9A) resulted in the appearance of five clearly defined false peaks in the

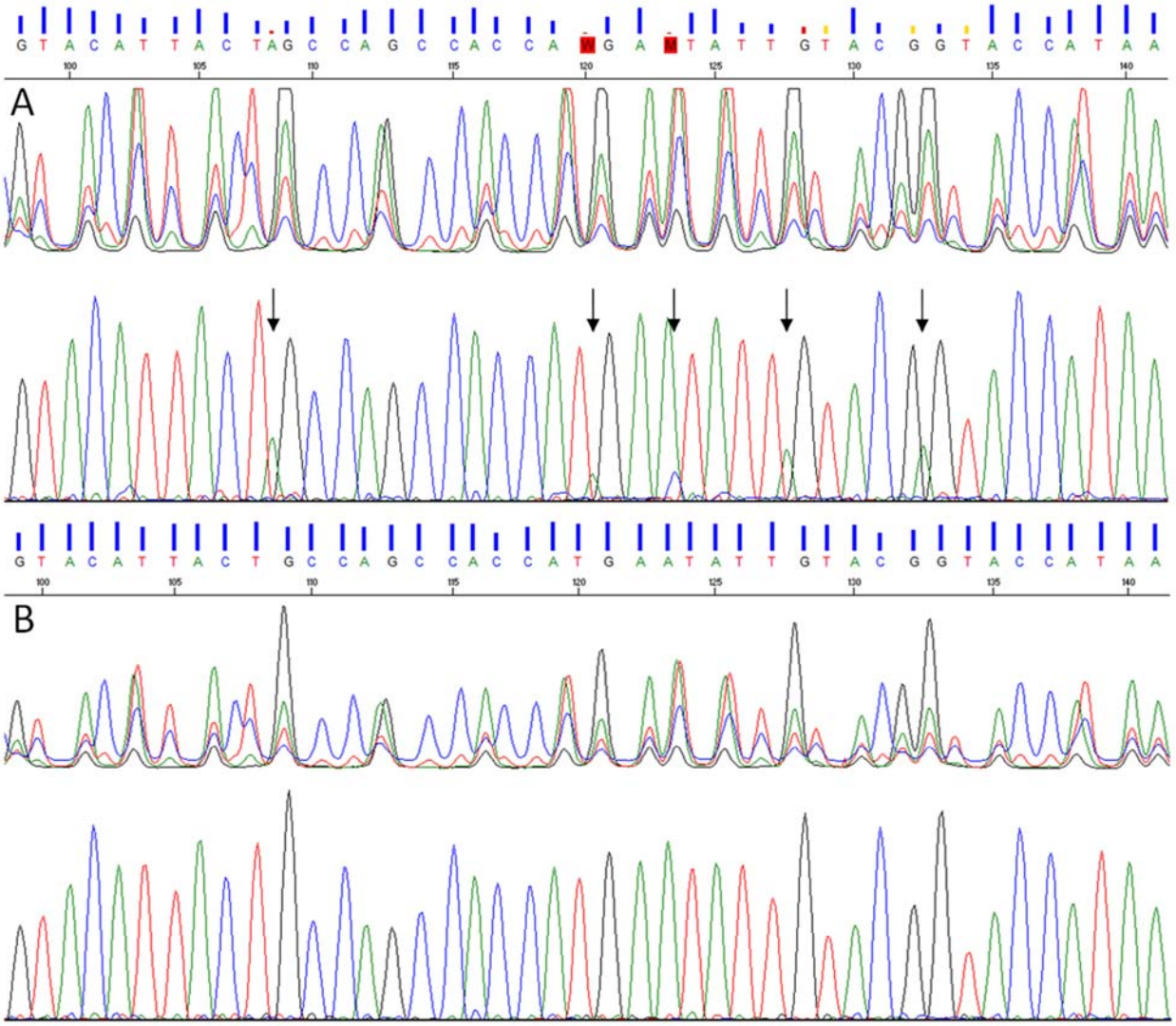


Figure 9: Comparison of raw and analyzed electrophoretic sequence data for HV1A sequenced using the A1 primer at two DNA input quantities. (9A) Use of 1.6ng of template DNA in a labeling reaction was associated with signal saturation of the raw data; false peaks produced by spectral pull-up in the analyzed data and frequent low basecall quality scores. (9B) Use of 0.8ng of template DNA in a labeling reaction eliminated the signal saturation seen at higher DNA input quantities and the resulting false peaks in the analyzed data. Basecall quality scores were also consistently high.

analyzed data – three of which were erroneously called by the sequencing software. In the absence of signal saturation, these were eliminated as illustrated by the analyzed data for the 0.8ng reaction (Figure 9B). This underscores the importance of accurate template quantification prior to dye terminator labeling.

Impact of positional location and sequence context on peak height reproducibility - To determine whether a base's positional location within an amplicon has an influence on electropherogram reproducibility, peak height variability was compared across the length of each sequenced amplicon at five DNA input quantities. Although the average variability across an amplicon (Table 4) may range from ±1.4% (HV2A sequenced using the D2 primer at 0.4 - 0.8ng input DNA) to ±10.0% (HV2B sequenced using the D1 primer at 0.1ng input DNA), variability was found to be very position dependent. As illustrated by an analysis of HV1A amplicon sequenced using the B2 primer (Figure 10), peak height variability is typically greatest immediately adjacent to the primer binding site. Variability decreases steadily towards the center of the amplicon; normally reaching a minimum 85 - 114bp downstream of the primer binding site. Variability then rises steadily as one approaches the end of the labeled product. This phenomenon is generally more pronounced at either low (0.1 - 0.2ng) or high (1.6ng) input DNA quantities.

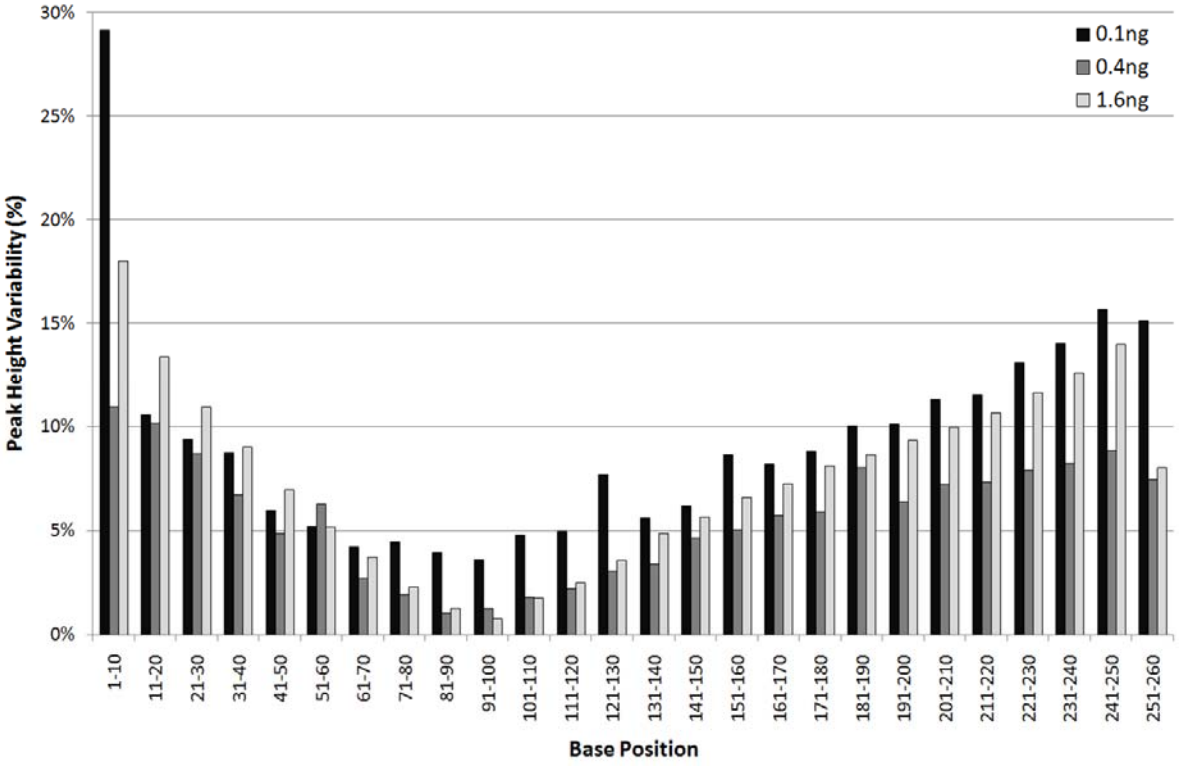


Figure. 10: Variability of electrophoretic peak heights across the HV1A amplicon sequenced using the B2 primer. To compensate for relative differences in peak heights, variability data (averaged for groups of ten bases) are expressed as a percentage of peak height. Values are based on five independent labeling reactions for each of the three DNA input quantities. Variability which is greatest adjacent to the primer binding site decreases towards the center of the amplicon and then rises toward the end of the labeled product. This pattern is more pronounced at input DNA quantities of 0.1ng and 1.6ng.

Nearest neighbor interactions are important to the thermodynamic stability of base pairing in nucleic acids and thus may impact the efficiency of nucleotide incorporation during dye terminator labeling. To determine whether this is a significant concern, peak height variability was examined as a function of the preceding base for all HV1 and HV2 amplicons at all five input DNA quantities. Analysis of the resulting data by two-way ANOVA showed no significant influence on peak height variability as a result of the preceding base at any DNA input quantity (F_s= 1.99 [0.1ng], 1.77 [0.2ng], 0 .20 [0.4ng], 0.19[0.8ng], 0.25[1.6ng]; df = 3,112; $p > 0.05$).

Given the sequence dependent folding characteristics of a DNA molecule, the context specificity of a base might be expected to influence the incorporation of dideoxy dye terminators. Although small differences in peak height reproducibility were seen as a function of the preceding base, these differences did not approach statistical significance. The degree of electropherogram reproducibility was, however, found to vary as a function of location within an amplicon. Peak height variability was elevated both in the region immediately downstream of the labeling primer and at the end of the amplicon leaving the most reproducible peak heights clustered toward the center of the amplicon. Although this was seen at all DNA input quantities, use of optimal input DNA quantities (0.4 - 0.8ng) minimized the phenomenon.

Impact of base substitution on proximal peak heights - To assess the impact of base substitution on the overall "pattern" of relative peak heights, was sequenced in both the forward and reverse directions at five different DNA input quantities (0.1ng - 1.6ng).

A base substitution may influence the peak heights of surrounding bases. Based on analyses of 150 base substitution sites, the height of peaks within a range of -2 to +5 bases of the substitution site are most typically affected. This is not to suggest, however, that every peak within this window will necessarily be affected nor that such proximal effects cannot occur beyond it. Figure 11 illustrates this for a T→C base substitution at position

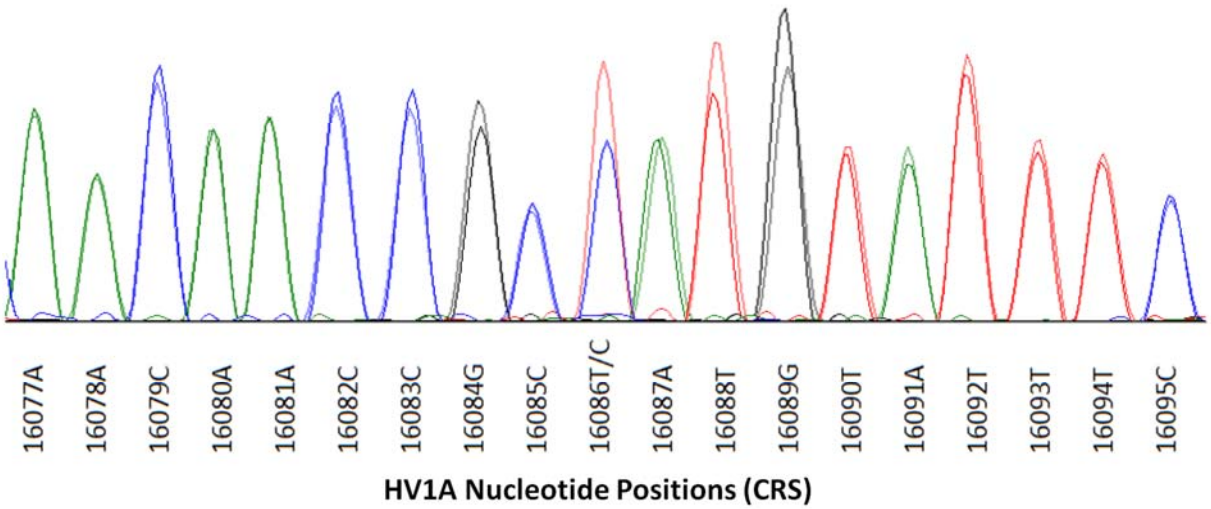


Figure 11: Superimposed electropherograms for a region of HV1A sequenced using the A1 primer illustrated the influence of a 16086T→C base substitution on the heights of proximal peaks. The largest changes involved decreases in peak heights at 16084G and 16088T and an increase at 16089G. This proximal effect did not impact the heights of the intervening peaks at 16085C and 16087A.

16086. In addition to the expected change in channel and height of the electrophoretic peak at the substitution point, the heights of the nearby peaks also showed a reproducible change in height. The largest changes were seen at 16084G, 16088T and 16089G. The heights of 16084G and 16088T decreased while that of 16089G increased following the base substitution at 16086. The intervening peak heights for 16085C and 16087A, by contrast were unchanged.

Other anomalous impacts on peak heights **-** Although the majority of cases in which a perturbation was seen in the expected peak height were attributable to base substitutions, some peak height deviations appeared to have a different etiology. Among peak height deviations detected in the course of the current study, 75.0% could be traced to poor signal quality typically at the beginning or end of a sequencing electropherogram. The remaining 25.0% of peak height deviations, however, were present in regions of high quality signal. In an effort to elucidate the underlying cause of the remaining atypical peak heights, raw and analyzed electropherograms were compared for these anomalous positions which were present in all four mtDNA amplicons.

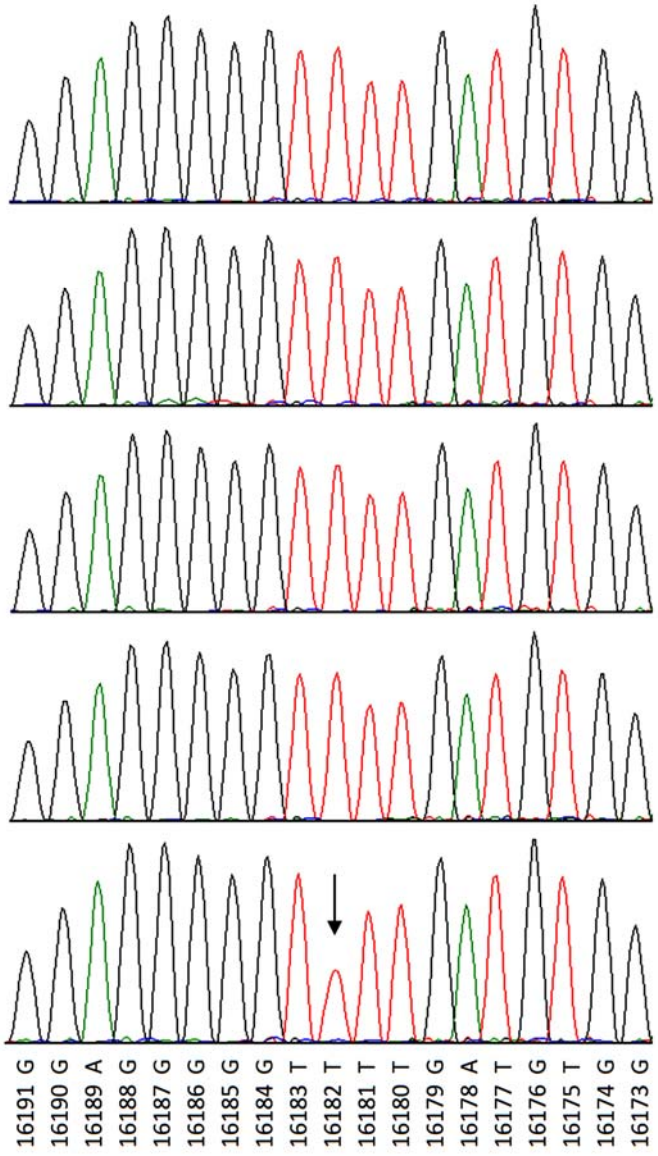


Figure 12: Replicates of five sequencing reactions from HV1A sequenced with the B2 primer and the overall peak height consistency for each peak across multiple independent reactions. Note that the 16182T peak (arrow) of the last replicate is significantly lower than the corresponding peak in the preceding four replicates.

Figure 12 shows the analyzed electrophoretic traces for five replicate sequencing reactions of part of the HV1A amplicon sequenced with the B2 primer. While there is almost complete consistency in the height for each peak across these independent reactions, the peak corresponding to 16182T has a significantly reduced height in the fifth replicate sample. This results in an atypical peak height when compared to a reference sequence. Analyses of such atypical peaks in regions of good signal quality revealed a similar pattern (*e.g.*, positions 16182 in HV1A; 16267 and 16284 in HV1B; and 158 in HV2A). In each case, the anomalous peak was nearly always the second base of a 3 - 5 base homopolymeric A- or C-

stretch; reduced in height relative to the reference pattern; and did not appear to have any influence on the heights of neighboring bases.

The detection of such unexpected peak height deviations may facilitate the detection of subtle underlying sequence anomalies that are not evident from base quality scores. In the current study, for example, the occurrence of some "atypical peak height deviations" revealed the presence of anomalies that were not linked to base substitutions, mixed base positions or regions of elevated peak height variability. Figure 13 highlights two bases (16267C and 16284A) which were found to be independently associated with atypical

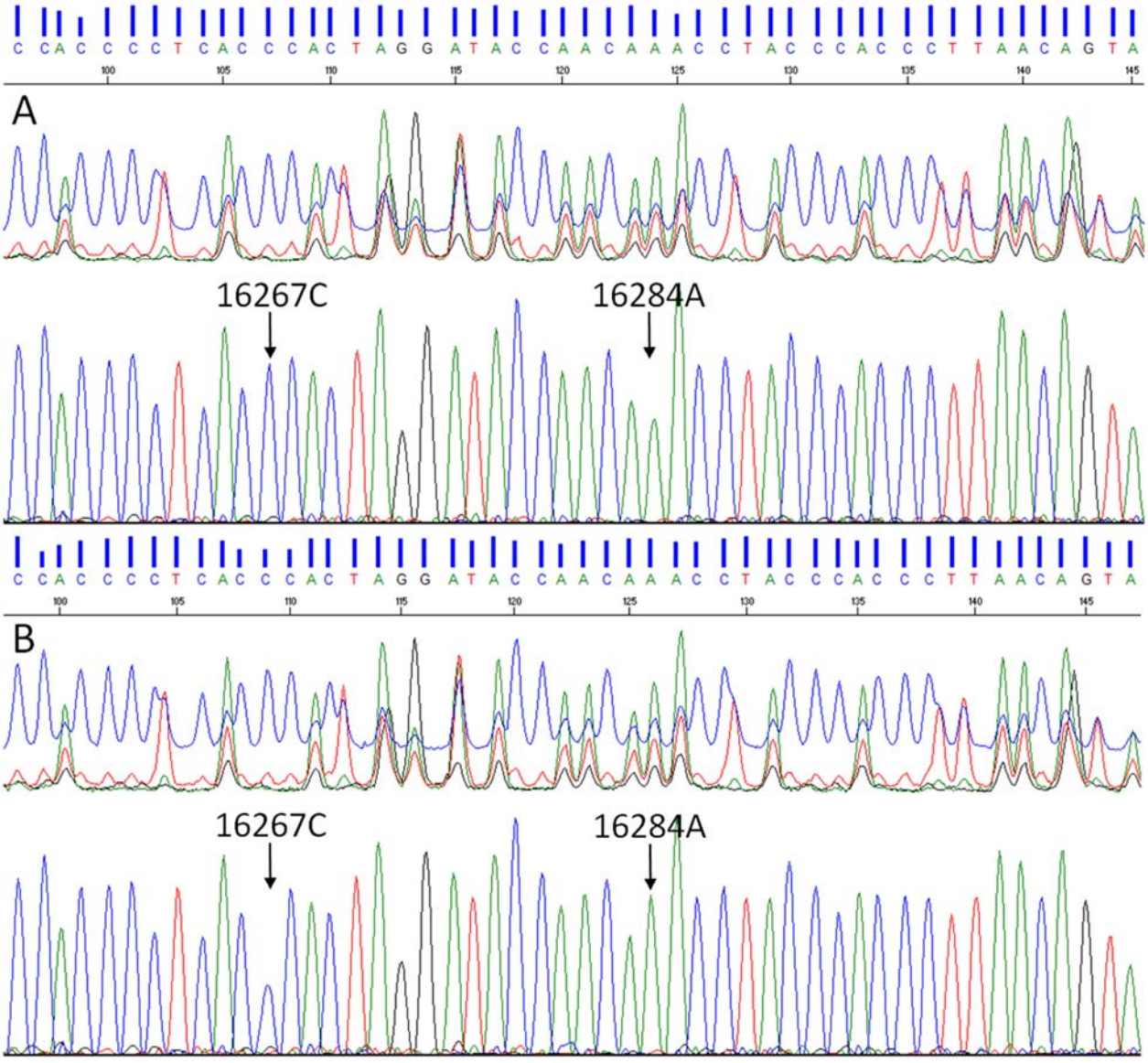


Figure 13: Illustration of the independent occurrence of two peak height anomalies within the same region of two of five replicates sequencing reactions. (13A) A low peak height anomaly is present at position 16267C while the peak height for 16284A is within expectations. (13B) A low peak height anomaly is present at position 16284A while the peak height for 16267C is normal. Although the anomalies are clearly evident in the analyzed data, comparison of the corresponding peaks in the raw data traces show no significant difference in peak height suggesting that the anomaly is due to a factor other than a reduction in fluorescence signal.

peak heights. In Figure 13A, the height of the 16267C peak is consistent with expectations based on the reference electrophoretic pattern while 16284A is abnormally low. Conversely, in Figure 13B, this relationship is reversed such that 16267C is abnormally low while the 16284A peak is consistent with the reference electrophoretic pattern. These anomalously low peaks, however, showed no indication of an anomalously low fluorescence signal in the raw data. Rather, the raw data appears indistinguishable between "normal" traces and those in which an atypical peak was seen. Furthermore, there does not appear to be an increase in noise or an obvious change in peak morphology that could account for this anomaly. In consultation with software engineers at Applied Biosystems, the consensus opinion is that the anomaly is likely the result of an unidentified bug within the KB basecaller (Applied Biosystems, personal communication). The potential for this to adversely impact basecalling accuracy has not been determined.

DNA quantity vs. peak height ratio determination **-** We also sought to determine the extent to which a change in the relative quantities of input DNA for two amplicons correlated with a change in the ratio between the resulting electrophoretic peak heights at mixed base positions. Preliminary data provided a strong indication that a statistically significant correlation existed. This was based on an initial analysis of nine mixed base positions and nine DNA quantity ratios. To much more rigorously establish the statistical reliability of this correlation, representatives of each of the four forensically validated mtDNA amplicons (*i.e.*, HV1A, HV1B, HV2A and HV2B) that collectively encompassed 100 variant sites across these regions were originally examined. In the course of theses analyses it was noted that the relationship between relative quantities of input DNA and the ratio of electrophoretic peak heights at mixed base positions was not identical when sequenced in the forward and reverse directions. Therefore, a more rigorous approach was taken to demonstrate the relationship between relative quantities of input DNA and the ratio of electrophoretic peak heights at mixed base positions by including an additional 102 mixed base positions in the reverse sequencing direction (*i.e.*, a total of 202 mixed base positions, (Table 5).

Table 5 - Sequence polymorphism mixtures assayed at 13 different DNA quantity ratios in the HV1 and HV2 regions.

	HV1									HV2						
Transitions	16051 G/A	16051 T/C	16069 C/T	16069 G/A	16086 A/G	16086 C/T	16093 G/A	16093 T/C	16104 C/T	64 A/G	64 T/C	72 C/T	72 G/A	73 A/G	73 C/T	89 A/G
	16104 G/A	16111 C/T	16111 G/A	16114 A/G	16114 T/C	16126 A/G	16126 C/T	16129 A/G	16129 C/T	89 T/C	92 C/T	92 G/A	93 A/G	93 C/T	143 A/G	143 T/C
	16136 T/C	16144 C/T	16144 G/A	16145 C/T	16148 A/G	16148 T/C	16163 C/T	16163 G/A	16172 A/G	146 A/G	146 T/C	150 C/T	150 G/A	151 G/A	151 T/C	152 A/G
	16172 C/T	16173 A/G	16173 T/C	16179 C/T	16179 G/A	16184 G/A	16184 T/C	16186 A/G	16186 T/C	152 C/T	153 A/G	153 T/C	182 G/A	182 T/C	185 A/G	185 C/T
	16187 C/T	16187 G/A	16189 A/G	16189 C/T	16192 A/G	16192 T/C	16193 A/G	16193 T/C	16209 A/G	188 A/G	188 C/T	189 C/T	189 G/A	192 C/T	192 G/A	194 A/G
	16209 C/T	16217 C/T	16217 G/A	16222 G/A	16222 T/C	16223 A/G	16223 C/T	16224 C/T	16224 G/A	194 T/C	195 A/G	195 C/T	196 C/T	196 G/A	198 C/T	198 G/A
	16239 A/G	16239 T/C	16241 C/T	16241 G/A	16243 A/G	16244 A/G	16244 T/C	16249 A/G	16249 T/C	199 C/T	199 G/A	200 A/G	200 T/C	204 C/T	204 G/A	207 A/G
	16256 A/G	16256 T/C	16261 G/A	16261 T/C	16266 G/A	16270 A/G	16270 T/C	16274 C/T	16274 G/A	207 T/C	217 C/T	217 G/A	222 A/G	222 C/T	225 C/T	225 G/A
	16278 A/G	16278 C/T	16286 C/T	16286 G/A	16291 A/G	16291 T/C	16293 C/T	16293 G/A	16294 A/G	226 A/G	226 T/C	228 A/G	228 C/T	239 A/G	247 A/G	247 C/T
	16294 C/T	16295 A/G	16295 T/C	16296 C/T	16296 G/A	16298 G/A	16298 T/C	16299 C/T	16299 G/A	248 T/C	250 C/T	250 G/A	285 T/C	295 C/T	297 A/G	297 T/C
	16304 A/G	16304 C/T	16309 A/G	16309 C/T	16311 A/G	16311 C/T	16320 C/T	16320 G/A	16325 A/G	309 T/C	310 C/T	316 C/T	316 G/A	340 A/G	340 T/C	
	16325 T/C	16327 C/T	16327 G/A	16343 C/T	16343 G/A	16352 C/T	16352 G/A	16353 A/G	16353 T/C							
	16355 C/T	16355 G/A	16360 C/T	16360 G/A	16362 A/G	16362 T/C	16390 G/A									
Transversions	16126 G/C	16129 C/G	16145 T/A	16182 A/C	16183 A/C	16265 A/C	16265 G/T	16286 C/G	16290 C/G	186 A/C	186 G/T	189 C/A	189 T/G	249 A/T	250 A/T	280 C/G

Given the reproducibility of individual peak heights in replicate sequencing reactions, it is reasonable to suggest that a change in relative molar ratio of one component to another in a mixture will be reflected by a corresponding change in the ratio of overlapping fluorescence peaks at any mixed base position. The relationship between differences in

molar ratio of two contributors to a mixture and the electrophoretic peak height ratio at six mixed base positions (182 T/C; 186 A/C; 189 C/A; 195 C/T; 198 T/C; 200 G/A) is illustrated in figure 14. As the molar ratio between two contributors in a HV2A mtDNA mixture was varied in a stepwise manner, from a 90:10 to a 50:50 ratio of contributor 1 to contributor 2, there was an observed coordinated shift in the ratio of overlapping electrophoretic peaks at nucleotide positions where the two contributors differ from each other. In this example, six peaks representing nucleotides associated with contributor 1 (182 T; 186 A; 189 C; 195 C; 198 T; 200 G) display a coordinated decrease in fluorescence intensity relative to peaks associated with contributor 2 (182 C; 186 C; 189 A; 195 T; 198 C; 200 A). Figure 15 illustrates the consistency in peak height for 182 T/C across nine of the DNA quantity ratios tested. As expected, peak heights, for this and the other 201 mixed

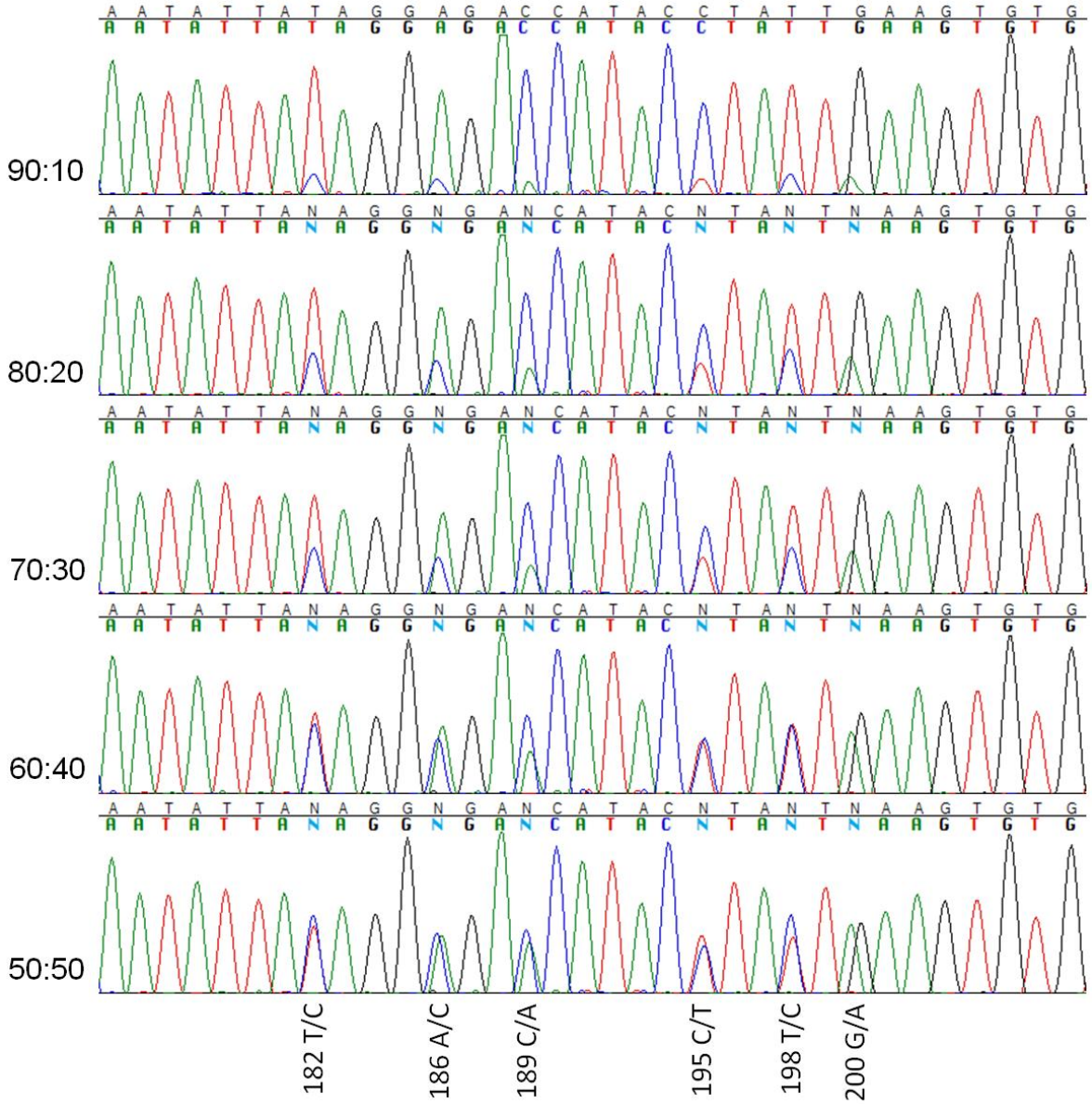


Figure 14: Fluorescence peak height ratios at six mixed base positions as a function of stepwise shifts in DNA quantity from a 90:10 mixture to a 50:50 mixture of two DNA amplicons.

base positions included in this study, were highly consistent across three replicate sequencing reactions. This goes to confirm previously published data on peak height consistency and shows that mixtures of two components can be analyzed by enrichment of one contributor over the other.

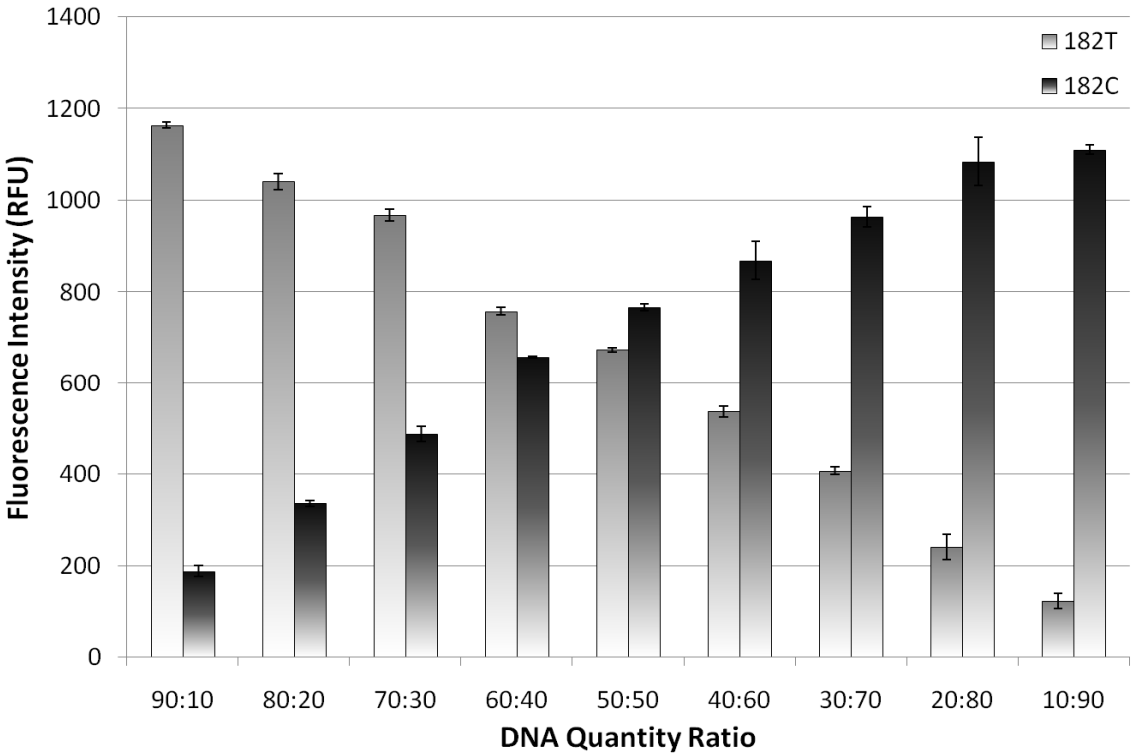


Figure 15: Peak height consistency for nine DNA quantity ratios for three replicate reactions at nt182. Both components of the mixture show a significant degree of peak height consistency across all tested DNA quantities.

Correlation between DNA quantity and fluorescence intensity ratios - The analyzed peak height data for 347 nucleotide mixtures revealed a highly significant correlation between DNA input quantity ratio and the corresponding change in peak height ratios in the sequencing electropherogram. Consistent with the high degree of reproducibility that was observed for relative differences in electrophoretic peak heights, a change in molar ratio of one contributor to another was found to correlate nearly perfectly (R^2 = 0.9996) with the change in relative fluorescence of overlapping peaks at mixed base positions (Figure 16). This is not to suggest that all mixed base positions behave identically with regard to relative peak height ratios at a given position. Instead, the relative peak height ratios observed at different nucleotide positions displayed highly reproducible context-specific peak height characteristics. This is best demonstrated by the highly reproducible differences that were seen at various nucleotide positions with respect to relative electrophoretic peak heights at equivalent input DNA quantities. Figures 17 and 18 illustrates the fluorescent shift in apparent 50:50 ratio for all mixed bases tested and their respective frequencies. Although the majority of fluorescent ratios (69.5%) fall within ±10% of the apparent 50:50 ratio, some mixtures display a significant discordance between molar and fluorescence ratios, even as far as 21:79. Nevertheless, these positions still display a strong correlation between a change in ratio of input DNA and change in

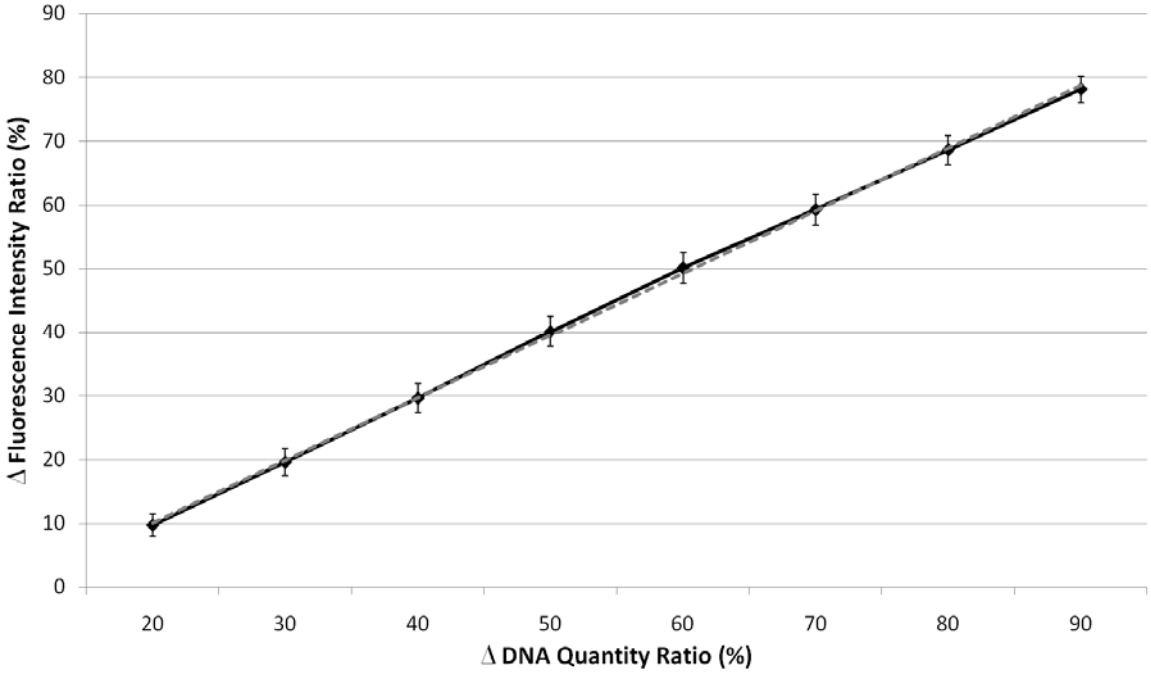


Figure 16: Linear relationship between a change in relative DNA quantity and the corresponding change in peak height ratio. Data shown are based on averages obtained from assay of 347 mixtures covering 202 nucleotide positions at 9 DNA quantity ratios ranging from 10:90 to 90:10. The dashed line indicates line of best fit (R^2 = 0.9996).

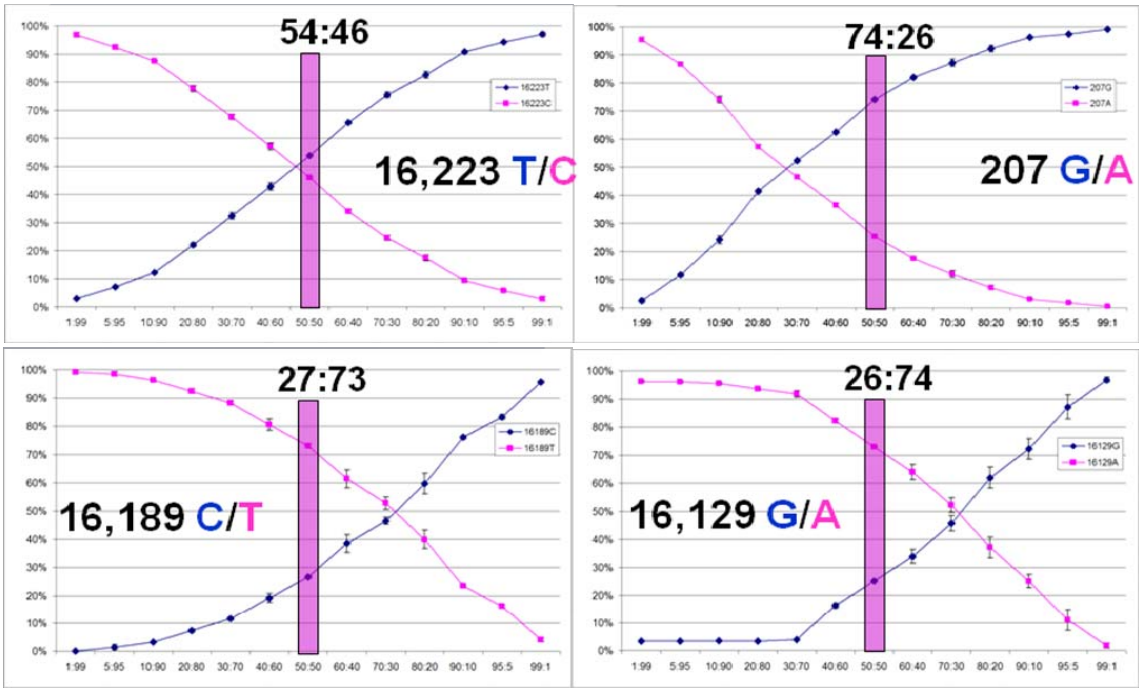


Figure 17: Relationship between DNA quantity input and resulting electrophoretic peak heights. Shown are average fluorescence for each base and error bars at 4 mixed base positions (207G/A; 16129G/A; 16189C/T; 16223C/T) at thirteen stepped DNA quantity ratios. Color bars indicate fluorescence ratios at equimolar DNA input ratios. Although nucleotide context may result in skewed fluorescence ratios, this phenomenon is highly reproducible. Mixed samples were sequenced using ABI BigDye v1.1 dye terminator chemistry.

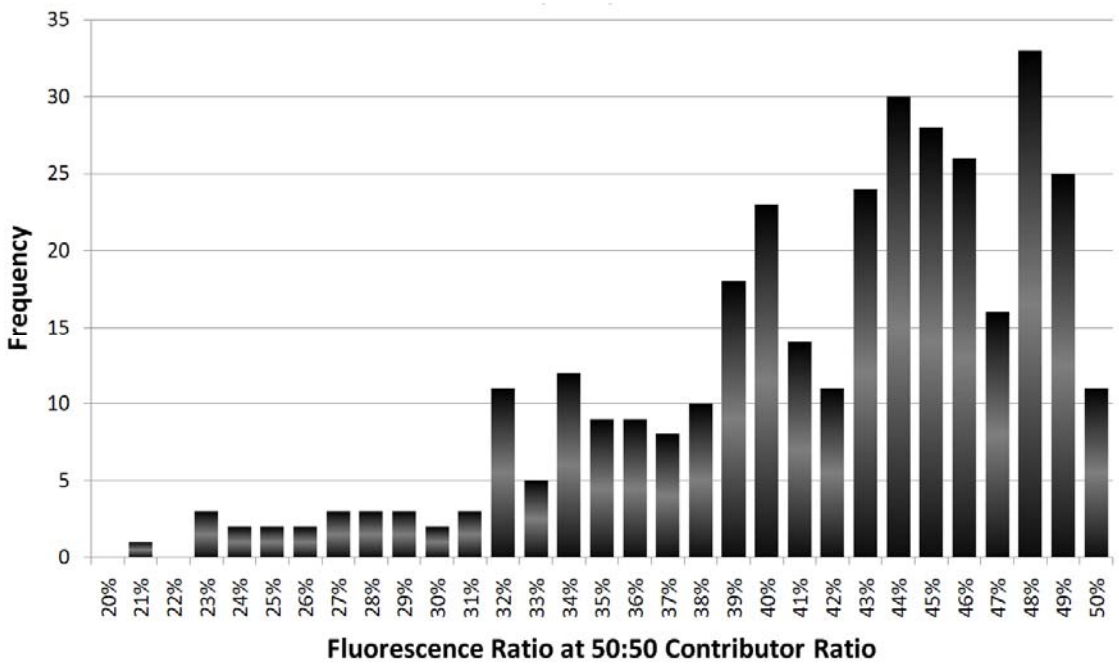


Figure 18: Frequency histogram showing the relationship between peak height ratios and DNA quantity ratios. The x-axis represents the relative fluorescence contribution of the lower peak at a mixed base positions for 50:50 DNA quantity input ratios. Although 69.5% of 50:50 contributor ratios result in a fluorescence ratio that is with ±10%, some mixed base postions display significantly skewed ratios such that a 50:50 ratio of DNA input quantities results in a 21:79 peak height fluorescence ratio.

ratio of overlapping peak heights. Therefore, when the minor component peak is less than 20% of the major component peak, the major component basecalls obtained from direct sequencing of mixtures can be declared as being associated with the major contributor. Interpretation of mtDNA mixtures can, therefore, differ from interpretation of suspected single component data. As current forensic guidelines on mtDNA sequencing analysis state that any secondary peak, with greater than 10% intensity of the primary peak, should be identified and considered a mixed site.

Impact of peak overlap on fluorescence intensity ratio - Direct sequencing of mixed mtDNA templates yields overlapping peaks in the analyzed electropherogram. In most cases, these peaks overlap completely (*i.e.*, peak maxima occur at the same sampling point), resulting in maximum detection of the secondary component. However, on occasion the secondary peak is shifted significantly up or downstream of the major peak resulting in underestimation of the minor component peak height. This phenomena is highly consistent across all ranges of DNA input quantities (Figure 14, 200G/A) and has been noted to a varying degree at a number of locations. Figure 19 illustrates an extreme example of minor component shifting for a 60:40 mixture of 239G/A, sequenced with the D2 primer. Figure 19A shows the 239A nucleotide of the minor component as an insertion due to peak shifting of the mixed nucleotide. Figure 19B, illustrates that the same peak spacing exists, at all bases, between the mixed and pure template containing 239G. As a result of this gap in peak spacing demonstrated for this section of the HV2A amplicon sequenced with the D2 primer, however, the analysis algorithm attempts to add the 239A

in to the string of basecalls as if it were a true insertion (239.1A). In a suspected single source sample this could be considered an insertion and stretches the importance of implementing technologies, such as denaturing high performance liquid chromatography (DHPLC), capable of mixture detection.

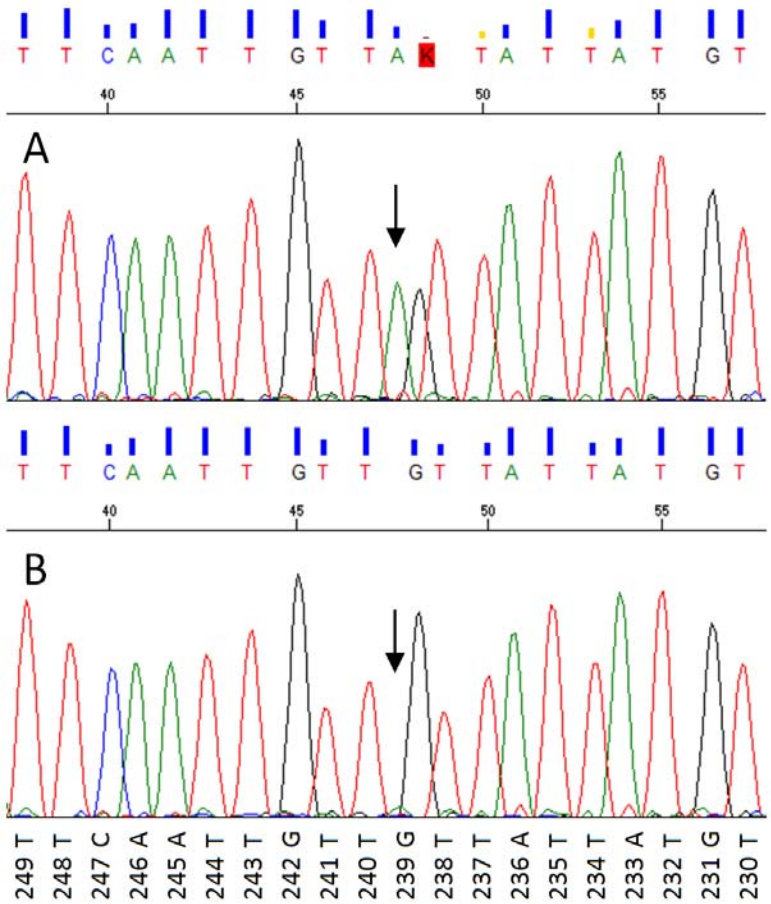


Figure 19: Secondary peak height shifting as a result of inefficient peak spacing during sequencing analysis. 19A illustrates a phantom insertion prior to 239G as a result of shifted peak spacing in the questioned region of the HV2A amplicon sequenced with the D2 primer. 19B illustrates the same peak spacing in a single source sample, demonstrating that the mixed site at 239G/A does not affect peak spacing during analysis of sequencing data.

Minor Contributor ID in Low Level Mixtures - Direct sequence-based mixture determination has generally been classified as any secondary peak having a fluorescence intensity >10% of that of the major component. Although generally true, a number of factors can cause a mixed base position to either under or over represent the relative quantity of DNA contributed by the secondary component. Analysis of 202 unique low level mixed base positions by direct DNA sequencing yielded a shift in fluorescence intensities for several tested positions. A total of 694 comparisons were performed at each contributor quantity ratio. At a quantity ratio of 20:80, 3.6% of minor component peaks yielded a fluorescence intensity ratio lower than the cutoff for mixed nucleotide detection (>10% major component peak height). At the 30:70 quantity ratio, however, no minor component peaks were below the 10% mixture calling threshold. As a result, some minor contributor mixtures representing less than 30% of total DNA quantity will result in peak dropout when analyzed by automated analysis of dye-terminator sequencing, even if they represent >10% of the total input DNA quantity. An example of this is that a low level mixture may yield a minor component haplotype where one or more nucleotides substitutions are not detected, as a result of low level nucleotide incorporation at those specific sites. This can have significant implications for haplotype determination of the minor component. In these cases, fractionation of cross-hybridized PCR products by DHPLC can be employed to enrich for nucleotides belonging to the minor contributor. In doing so, nucleotides

yielding significant decrease in apparent 50:50 fluorescence intensity ratio can be identified.

These results raise questions of how often contributor quantity ratio of 5:95 yield a fluorescence intensity ratio greater than the cutoff for mixed nucleotide detection. Out of 694 comparisons performed, 15.6% of minor components yielded a fluorescence intensity ratio greater than the automated 10% cutoff. Part of these can undoubtedly be associated with sequencing noise, however, due to knowledge of 50:50 fluorescence intensity ratio shift the majority show significant peak height fluorescence above baseline. Implications of these are that a sample can appear to be heteroplasmic at a single site, while in reality it contains a low level of secondary component. Again, employing DHPLC fractionation can aid in these types of analysis by enriching for the minor component and allowing for complete haplotype determination. Alternatively, DHPLC analysis can purify out the minor component contaminant, yielding a clean sequence for the major component. In all cases, knowledge of nucleotide incorporation rates associated with mixture determination is imperative.

The reliability of determining individual haplotypes in a DNA mixture through linkage phase analysis of sequence data from DHPLC fractionated mixtures

Statistical Basis for Mixture Resolution by Linkage Phase Analysis **-** A near perfect correlation (R^2 = 0.9996) exists when mixtures are analyzed for changes in relative DNA quantities versus changes in peak height ratios This demonstrates that a change in peak height ratio of two components in a mixture will be reflected by a corresponding change in relative DNA quantity at any mixed base position. Based on these data, exhaustive statistical analyses were performed using Tukey's honestly significant differences (HSD) test. In these analyses, the relative shift in peak height fluorescence was calculated by comparing fluorescence intensity ratios across thirteen DNA quantity ratios (99:1 – 1:99) for each of 202 mixed nucleotide positions. The resulting 27,066 comparison enrichments were then analyzed via Tukey's HSD analysis for statistical significance determination. The results of these analyses indicate an even tighter statistical correlation and have made it possible to precisely determine the level of statistical confidence associated with any detectible shift in electrophoretic peak height ratios. The calculated significance values for 27,066 comparisons of peak height fluorescence ratios illustrate the reliability of this relationship (Table 6 and Figure 20). Based on these statistical calculations, as little as 6% shift in fluorescence intensity of one contributor over the other, obtained by sequencing a series of DHPLC fractions, makes it possible to accurately determine the linkage phase of the questioned mixture with a >99% confidence. Enrichment by 17% increases the statistical confidence of linkage phase to >99.9%.

Linkage Phase Analysis of a Two-Component mtDNA Mixture **-** Because of the strong correlation between a change in the ratio of DNA quantities within a mixture and a change in the ratio of overlapping electrophoretic peaks, the linkage phase of individual amplicons present in a two component mixture can be readily determined. This holds true even in cases where complete separation of individual components of a mixture cannot be achieved. In such cases, the linkage phase of each component of the mixture is determined by tracking the ratio of overlapping fluorescent peaks at all mixed base positions between two or more DHPLC fractions. The observation of coordinated shifts in the relative

Table 6 - Peak height enrichment table showing corresponding average and minimum confidence of separation as well as the number of comparisons performed for each enrichment percentage point.

	Confidence					Confidence			
Above Δ %	Average	Stdev	Minimum	# of Samples	Above Δ %	Average	Stdev	Minimum	# of Samples
0%	55.998%	31.945%	1.776%	121	50%	99.995%	0.023%	99.730%	245
1%	88.635%	16.568%	15.103%	200	51%	99.995%	0.016%	99.887%	231
2%	94.838%	10.169%	40.710%	269	52%	99.996%	0.015%	99.817%	291
3%	97.594%	6.351%	49.178%	320	53%	99.996%	0.014%	99.863%	283
4%	98.663%	3.182%	75.688%	352	54%	99.995%	0.024%	99.618%	285
5%	98.760%	2.974%	80.905%	375	55%	99.995%	0.016%	99.822%	241
6%	99.308%	1.781%	82.861%	407	56%	99.996%	0.016%	99.795%	220
7%	99.470%	1.232%	89.437%	451	57%	99.997%	0.008%	99.921%	232
8%	99.678%	0.996%	87.545%	605	58%	99.997%	0.008%	99.920%	236
9%	99.731%	0.673%	92.094%	586	59%	99.996%	0.014%	99.876%	265
10%	99.717%	0.832%	92.259%	546	60%	99.998%	0.004%	99.971%	211
11%	99.777%	0.613%	92.674%	464	61%	99.997%	0.009%	99.913%	246
12%	99.835%	0.482%	96.265%	449	62%	99.998%	0.005%	99.942%	263
13%	99.873%	0.426%	95.614%	455	63%	99.996%	0.019%	99.749%	230
14%	99.868%	0.366%	95.620%	303	64%	99.996%	0.014%	99.878%	251
15%	99.883%	0.298%	97.805%	347	65%	99.997%	0.011%	99.866%	222
16%	99.894%	0.340%	96.077%	323	66%	99.998%	0.004%	99.966%	217
17%	99.943%	0.151%	98.870%	381	67%	99.997%	0.009%	99.918%	221
18%	99.945%	0.175%	97.357%	360	68%	99.998%	0.006%	99.943%	223
19%	99.956%	0.149%	97.757%	422	69%	99.998%	0.007%	99.939%	229
20%	99.959%	0.163%	97.445%	414	70%	99.998%	0.005%	99.954%	208
21%	99.971%	0.086%	99.078%	436	71%	99.998%	0.007%	99.927%	247
22%	99.966%	0.111%	98.720%	411	72%	99.998%	0.007%	99.928%	208
23%	99.967%	0.084%	99.381%	368	73%	99.998%	0.004%	99.976%	175
24%	99.972%	0.077%	99.318%	344	74%	99.997%	0.008%	99.935%	213
25%	99.973%	0.091%	99.174%	318	75%	99.997%	0.007%	99.944%	187
26%	99.975%	0.091%	98.832%	306	76%	99.997%	0.007%	99.951%	208
27%	99.976%	0.065%	99.319%	315	77%	99.997%	0.012%	99.852%	172
28%	99.986%	0.036%	99.544%	299	78%	99.998%	0.007%	99.945%	192
29%	99.976%	0.106%	98.631%	344	79%	99.998%	0.007%	99.918%	213
30%	99.981%	0.054%	99.523%	339	80%	99.998%	0.003%	99.979%	184
31%	99.981%	0.066%	99.290%	336	81%	99.998%	0.006%	99.962%	204
32%	99.991%	0.027%	99.782%	345	82%	99.998%	0.006%	99.950%	190
33%	99.987%	0.037%	99.648%	319	83%	99.998%	0.004%	99.967%	214
34%	99.988%	0.047%	99.389%	361	84%	99.998%	0.006%	99.943%	203
35%	99.990%	0.028%	99.719%	302	85%	99.997%	0.011%	99.869%	205
36%	99.988%	0.057%	99.136%	286	86%	99.998%	0.004%	99.973%	224
37%	99.990%	0.028%	99.728%	269	87%	99.998%	0.006%	99.920%	198
38%	99.987%	0.042%	99.554%	290	88%	99.998%	0.007%	99.930%	178
39%	99.993%	0.022%	99.744%	270	89%	99.998%	0.009%	99.879%	221
40%	99.988%	0.057%	99.242%	241	90%	99.999%	0.003%	99.975%	181
41%	99.993%	0.024%	99.762%	291	91%	99.998%	0.007%	99.920%	171
42%	99.992%	0.033%	99.594%	296	92%	99.999%	0.003%	99.983%	144
43%	99.994%	0.017%	99.870%	332	93%	99.999%	0.002%	99.987%	136
44%	99.993%	0.020%	99.852%	294	94%	99.999%	0.001%	99.993%	125
45%	99.994%	0.022%	99.710%	294	95%	99.999%	0.003%	99.979%	108
46%	99.993%	0.021%	99.849%	271	96%	99.999%	0.003%	99.983%	74
47%	99.994%	0.018%	99.754%	263	97%	99.999%	0.002%	99.987%	41
48%	99.994%	0.020%	99.822%	239	98%	99.999%	0.002%	99.994%	12
49%	99.995%	0.013%	99.859%	255	99%	99.9999%	0.0001%	99.9998%	4

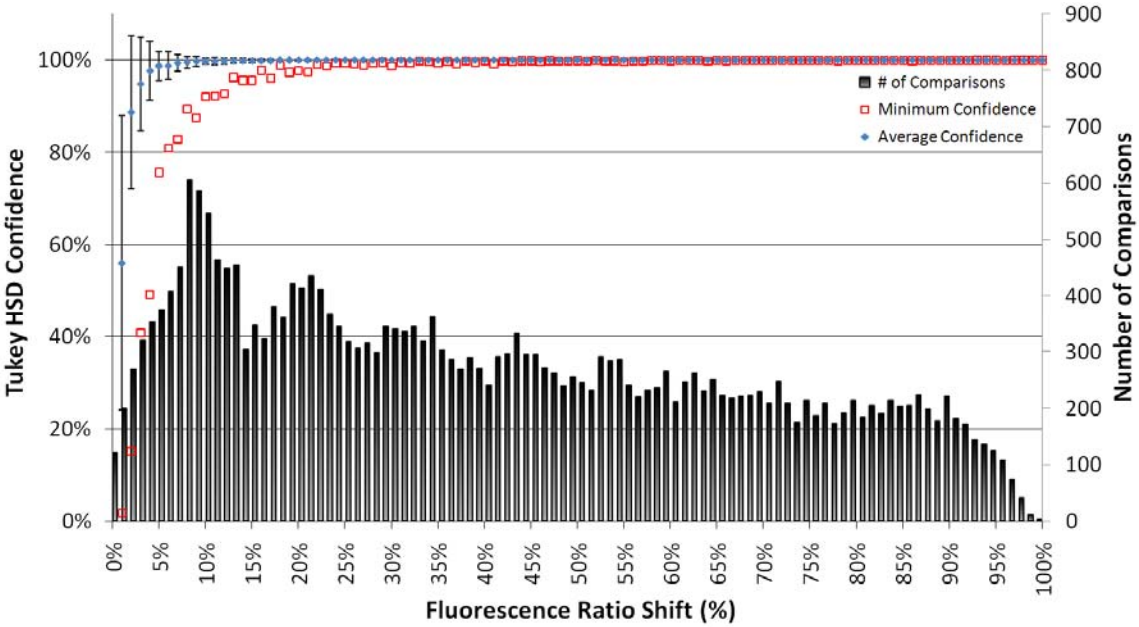
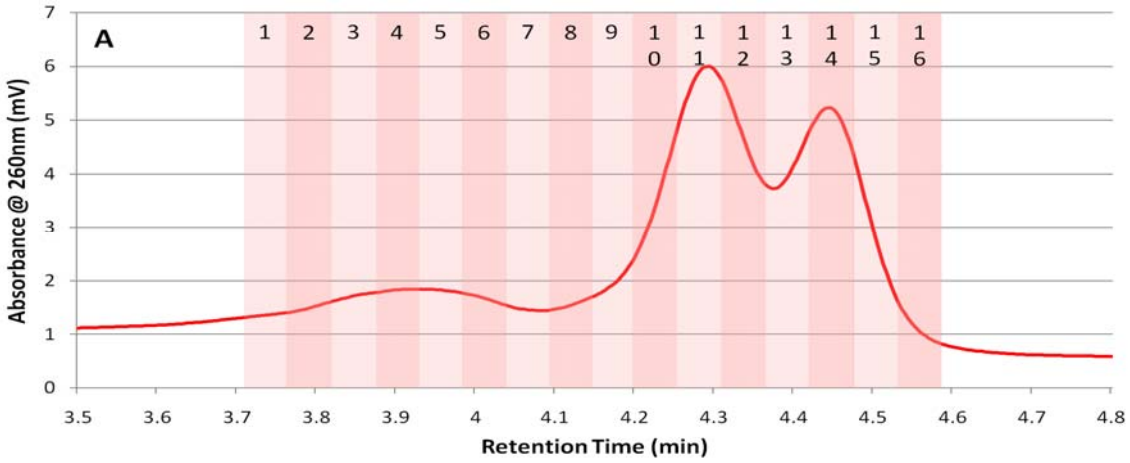


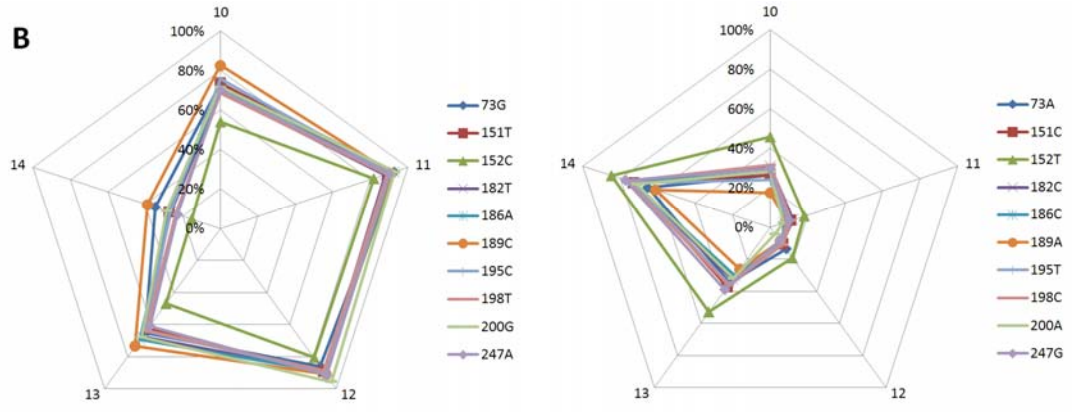
Figure 20: Tukey HSD confidence associated with changes in electrophoretic peak height ratios at mixed base positions. Shown are average and minimum confidence values as a function of the percentage change in peak height fluorescence along with the number of data points behind each calculation. Greater than 99% confidence is obtained for a >7% shift in fluorescence.

fluorescence ratios for a given set of nucleotides is consistent with them being in the same linkage phase and thus representing the same amplicon. Figure 21 represents the complete process of DHPLC fractionation followed by linkage phase analysis (LPA) of mixed nucleotide positions detected as a result of dye-terminator sequencing. A two-component mixture of HV2A amplicons containing ten mixed nucleotide positions (73G/A, 151T/C, 152T/C, 182T/C, 186A/C, 189C/A, 195C/T, 198T/C, 200G/A and 247A/G) was purified and fractionated into 16 individual fractions by DHPLC (Figure 21A). Five of the captured fractions (10-14) were sequenced and the mixed nucleotide positions were analyzed for relative peak height shifts (Figure 21B). Subsequently these fractions were analyzed by LPA for determination of nucleotides corresponding to each contributor (Figure 21C). Separation of mixed nucleotides based on LPA identified the two contributor mixture with greater than 99.99% average basecalling confidence into contributor 1 (73A, 151C, 152T, 182C, 186C, 189A, 195T, 198C, 200A and 247G) and contributor 2 (73G, 151T, 152C, 182T, 186A, 189C, 195C, 198T, 200G and 247A). Although figures 21B and 21C illustrate only results obtained in the forward sequencing direction, concurrent results were obtained when samples were sequenced in the reverse direction.

Although fragment resolution by DHPLC, is a function of amplicon length and/or differences in primary sequence, it is difficult to predict apriori the characteristics of the chromatographic trace that will result from a specific mixture. Optimal separation and recovery is generally achieved when a two-component mixture is resolved into four discernible peaks (*i.e.*, separated homo- and/or heteroduplices). This minimizes co-elution of mixed DNA profiles and subsequent electropherograms are generally of good quality and are free of underlying electrophoretic signal from a second contributor. In practice,

however, optimal chromatographic resolution is uncommon, therefore resulting in the co-elution of target and non-target DNA. This produces a series of DHPLC fractions that are enriched for one contributor versus the other. In the absence of complete separation, sequence electropherograms show signal from both contributors with overlapping peaks at those sites where the two amplicons differ in primary sequence.





C

Linkage Phase Determination	Position	Comparison 11 and 14	Average Confidence	Confidence Std Dev	Minimum Confidence	Samples Behind Stats	Increasing	Decreasing
	73	57.8%	99.997%	0.008%	99.921%	232	73A	73G
	151	61.9%	99.997%	0.009%	99.913%	246	151C	151T
	152	66.5%	99.998%	0.004%	99.966%	217	152T	152C
	182	61.4%	99.997%	0.009%	99.913%	246	182C	182T
	186	65.5%	99.997%	0.011%	99.866%	222	186C	186A
	189	52.2%	99.996%	0.015%	99.817%	291	189A	189C
	195	63.6%	99.996%	0.019%	99.749%	230	195T	195C
	198	64.9%	99.996%	0.014%	99.878%	251	198C	198T
	200	64.3%	99.996%	0.014%	99.878%	251	200A	200G
	247	67.1%	99.997%	0.009%	99.918%	221	247G	247A

Figure 21: (21A) Denaturing HPLC chromatogram, showing collected fractions, including fractions submitted to dye-terminator sequencing. (21B) Relative quantitation of fluorescence peak height ratios (y-axis) for 10 mixed base positions from sequence electropherograms representing DNA recovered from five DHPLC fractions (x-axis). (21C) Linkage phase analysis determination and identification of individual contributors to the mixture using relative fluorescent intensity at fractions 11 and 14 as determined by DHPLC fractionation.

The results of the statistical analysis described above illustrate that only a 7% shift in relative peak height fluorescence at mixed nucleotide positions is required to yield greater than 99% statistical confidence of separation, while a 17% peak height shift yields greater than 99.9% confidence in base calling. Therefore, only a slight shift in peak height fluorescence is required in order to obtain a highly statistically significant linkage phase resolution of a mixture. It is worth noting that the majority of mixtures analyzed to date show peak height enrichment >17%.. In addition, all mixtures analyzed by LPA, no matter the amount of enrichment, have accurately identified each contributor to a mixture. As a safeguard against incorrect determination of contributors, however, analyses where peak height enrichment is <7% (*i.e.*, <99% confidence) should be viewed with caution.

Linkage Phase Analysis of Minor-Component mtDNA Mixtures - In forensic casework in general, mtDNA samples are not considered mixtures unless two or more nucleotide positions yield a secondary peak height of >10% that of the major component nucleotide. A 10% secondary peak height, however, can represent a mixture ranging anywhere from a 95:5 to a 70:30, as a result of context specific incorporation of dye terminators.

A known 95:5 mixture of HV2A, which appeared to be single source with a single heteroplasmic site, as determined by direct DNA sequencing, was separated into ten fractions by DHPLC and analyzed by LPA. Two of the isolated fractions yielded no mixture at all, while 5 yielded apparent heteroplasmy at one site as determined by the 10% cutoff in secondary peak height. However, two of the fractions yielded 8 mixed positions ranging from 15 – 53% secondary peak height. LPA analysis confirmed the existence of a mixture at all eight sites with an average confidence of 99.985%.

Identification of suspected low level mixtures that might not be visible by direct sequencing without DHPLC fractionation shows the tremendous power of this approach. By enriching for the minor contributor as a function of heteroduplex analysis and DHPLC fractionation, it is possible to determine not only that the original sample was a mixture, but also to determine the sequence of the minor contributor to a high degree of statistical significance. The presented results show that samples thought to be of a single source as determined solely by direct DNA sequencing might contain significant amounts of additional contributors. Until now, however, no method has been available to determine otherwise.

The ability to fractionate and analyze low level mixtures will allow for identification of minor contributors which, to date; have been entirely overlooked by direct dye-terminator sequencing analysis. These samples, which may appear only slightly heteroplasmic at one site, can however, yield significant information in cases where a second contributor is suspected. Until now, only by subcloning mixed templates has it been possible to identify the presence of additional contributors to these types of mixtures.

Additional Examples of Two-Component mtDNA Mixtures Resolved by Linkage Phase Analysis – As indicated above, because of the strong correlation between a change in the ratio of DNA quantities within a mixture and a change in the ratio of overlapping electrophoretic peaks, the linkage phase of the individual amplicons present in a two component mixture can be readily determined. The linkage phase of each components of the mixture is determined by tracking the ratio of overlapping fluorescent peaks at all

mixed base positions between two or more DHPLC fractions. The observation of coordinated shifts in the relative fluorescence ratios for a given set of nucleotides is consistent with them being in the same linkage phase and thus representing the same amplicon. Using this approach to analyze a set of forty-two two-component mixtures it was possible to accurately determine the linkage phase in all cases.

Figure 22 illustrates a two-component mixture of HV2A amplicons was fractionated by DHPLC. A total of twelve fractions were recovered and sequenced. The peak height ratio at each of the eight mixed base positions in the resulting series of electropherograms was determined and plotted. The linkage phase of the two amplicons can be determined from the coordinated shifts in peak height ratios across the fractions that were sequenced. In this example, the height of the fluorescent peaks associated with nucleotides 146T, 150T, 185G, 188A, 195C, 198T, 222C and 228G increase and decrease in a coordinated manner across the twelve DHPLC fractions. Similarly, the height of the fluorescent peaks associated with nucleotides the 146C, 150C, 185A, 188GA, 195T, 198C, 222T and 228A also shift in a coordinated manner but in the opposite direction. Thus, these two sets of bases define the linkage phase of each of the two contributors to the mixture. These results are in concordance with the known sequence of the individual amplicons used to construct the mixture.

Careful inspection of the results for individual mixed base positions also demonstrates that discordance between molar and fluorescence ratios does not compromise the accurate linkage phase determination. This is exemplified by nucleotide positions 222 and 228. Sequence electropherograms generated from DHPLC fractions 1-4 consistently show equal or greater fluorescence for the 222C and 228A pairing than for 222T and 228G pairing. Neither of these nucleotide pairings, however, demonstrate coordinated shifts in relative fluorescence across these DHPLC fractions. Rather, it is the 222C and 228G nucleotide pairing that shows coordinated increases in fluorescence between fractions 1 and 3 and then a coordinated decrease between fractions 3 and 4. Conversely, the 222T and 228A nucleotide pairing shows the complementary shift, *i.e.*, a decrease in relative fluorescence between fractions 1 and 3 and an increase between 3 and 4. It is the nucleotide pairings based on coordinated shifts in relative fluorescence that are concordant with the actual sequence of the amplicons present in the mixture. This underscores the fact that it is the change in relative fluorescence and not the absolute peak height that allows for linkage phase of the individual amplicons in an mtDNA mixture to be determined. Additional examples of these analyses are shown in Figures 23-27.

Three or More Component Mixtures **-** Three or more component mixture detection by DHPLC yields significantly more complicated chromatographic traces than two-component mixtures as a result of significant increase in the number of heteroduplices formed with the addition of each contributor. To date, 32 three to five component mixtures at various contributor ratios have been analyzed by DHPLC, yielding anywhere from 4 to 10 chromatographic peaks. Sequencing results and subsequent LPA of three component mixtures have shown that determination of one or two contributors can be achieved. However, application of statistical analyses for these samples remains difficult since shifting in peak height fluorescence ratio is skewed as a result of multiple overlapping nucleotide contributions at any one mixed site. Analysis of three to five component mixtures by LPA, therefore, has not been determined. Analyses and modifications to the

LPA algorithm to enable this represents a possible area for further research. DHPLC analysis of these types of mixtures, however, can yield significant information as to the complexity of the mixture. From these results, a forensic analyst can then determine whether to proceed with downstream sequencing. This not only gives the analyst prior knowledge of the complexity of the analyses ahead, but also the choice on whether or not to proceed with time and cost intensive sequencing analyses and interpretation.

Identification of Heteroplasmic Components by LPA - Identification of heteroplasmy in mtDNA sequencing analysis has been proven to be of tremendous utility for human identification. Due to the nature of DNA sequencing, low level heteroplasmic identification can often prove difficult to distinguish from underlying sequence chatter and noise. Utilizing DHPLC fractionation and LPA, identification of these can be made significantly easier, allowing for confirmation of suspected heteroplasmy in difficult samples. One sample displaying a broadened DHPLC profile revealed low level heteroplasmy at 16183. This sample was subjected to DHPLC fractionation and LPA, resulting in 25.6% and 24.0% enrichment of the minor component, depending on the direction of sequencing. This confirmed the existence of a heteroplasmic 16183C/A with an average confidence of 99.973±0.091% and 99.973±0.077% depending on sequencing direction.

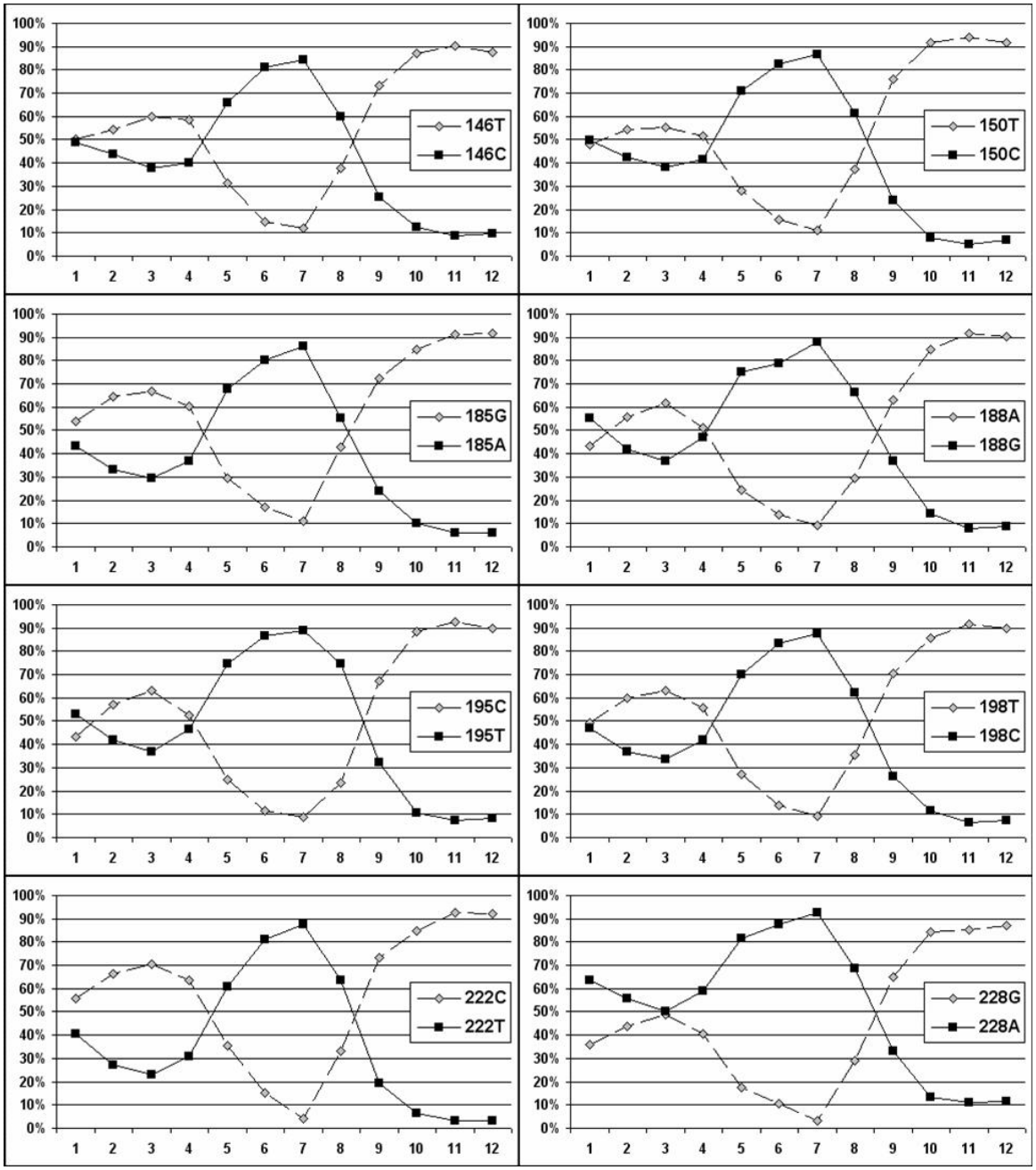


Figure 22: Linkage phase determination by quantitation of fluorescence peak height ratios at eight mixed base positions from sequence electropherograms representing DNA recovered from serial DHPLC fractions (shown on the x-axis). The linkage phase is determined from the coordinated shifts in peak height ratios across the fractions that were sequenced. The coordinated shift in the overlapping eletrophoretic peaks identifies the linkage phase of contributor 1 (146T, 150T, 185G, 188A, 195C, 198T, 222C and 228G) as distinct from the linkage phase of contributor 2 (146C, 150C, 185A, 188GA, 195T, 198C, 222T and 228A).

	rCRS	C10	G9
16311	a	G	A
16294	g	A	G
16286	g	C	G
16278	g	G	A
16265	t	G	T
16223	g	A	A

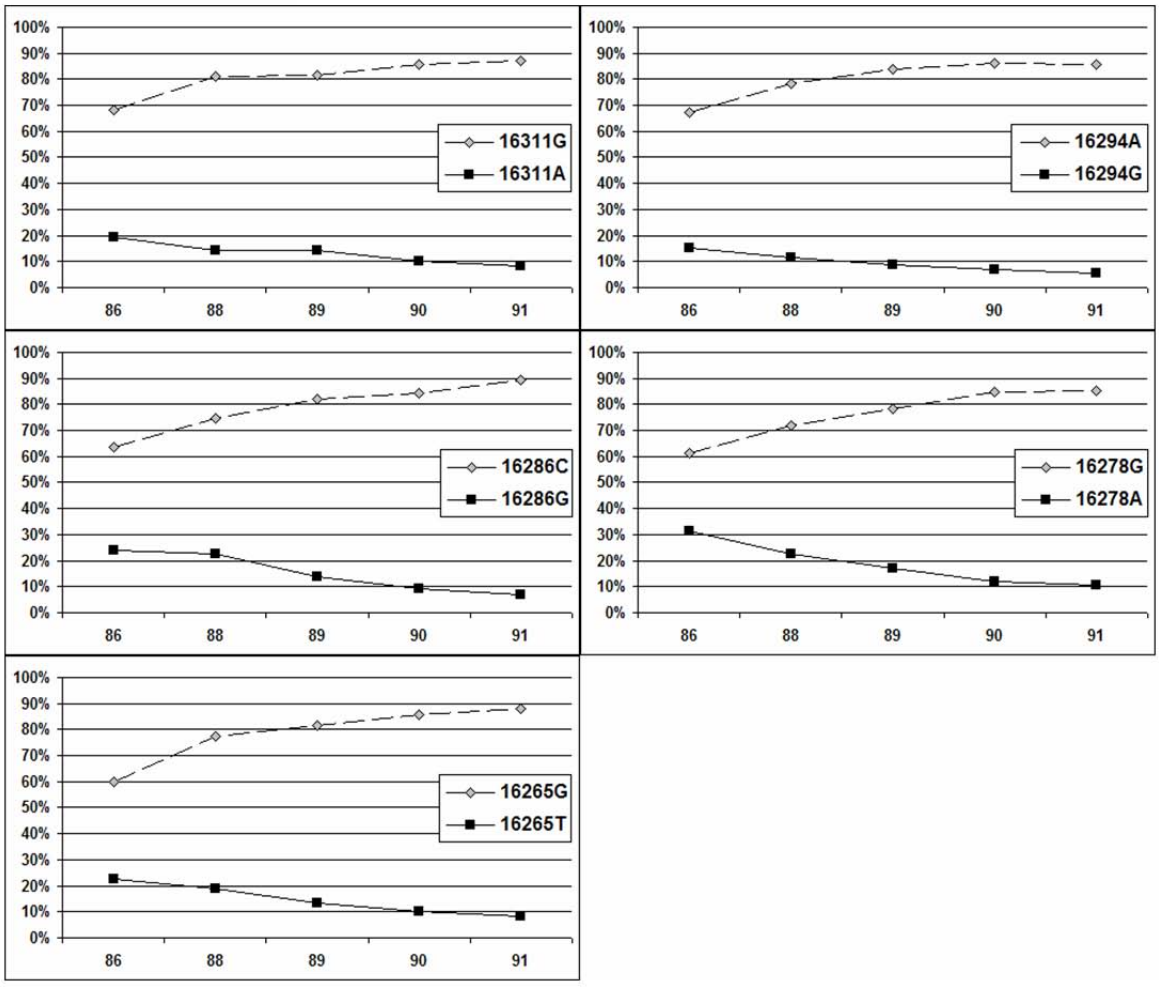


Figure 23: Linkage phase determination by quantitation of fluorescence peak height ratios at five mixed base positions from sequence electropherograms representing DNA recovered from 5 DHPLC fractions from a 20/80 mixture of C10 (16311G, 16294A, 16286C, 16278G, 16265G) and G9 (16311A, 16294G, 16286G, 16278A, 16265T)..

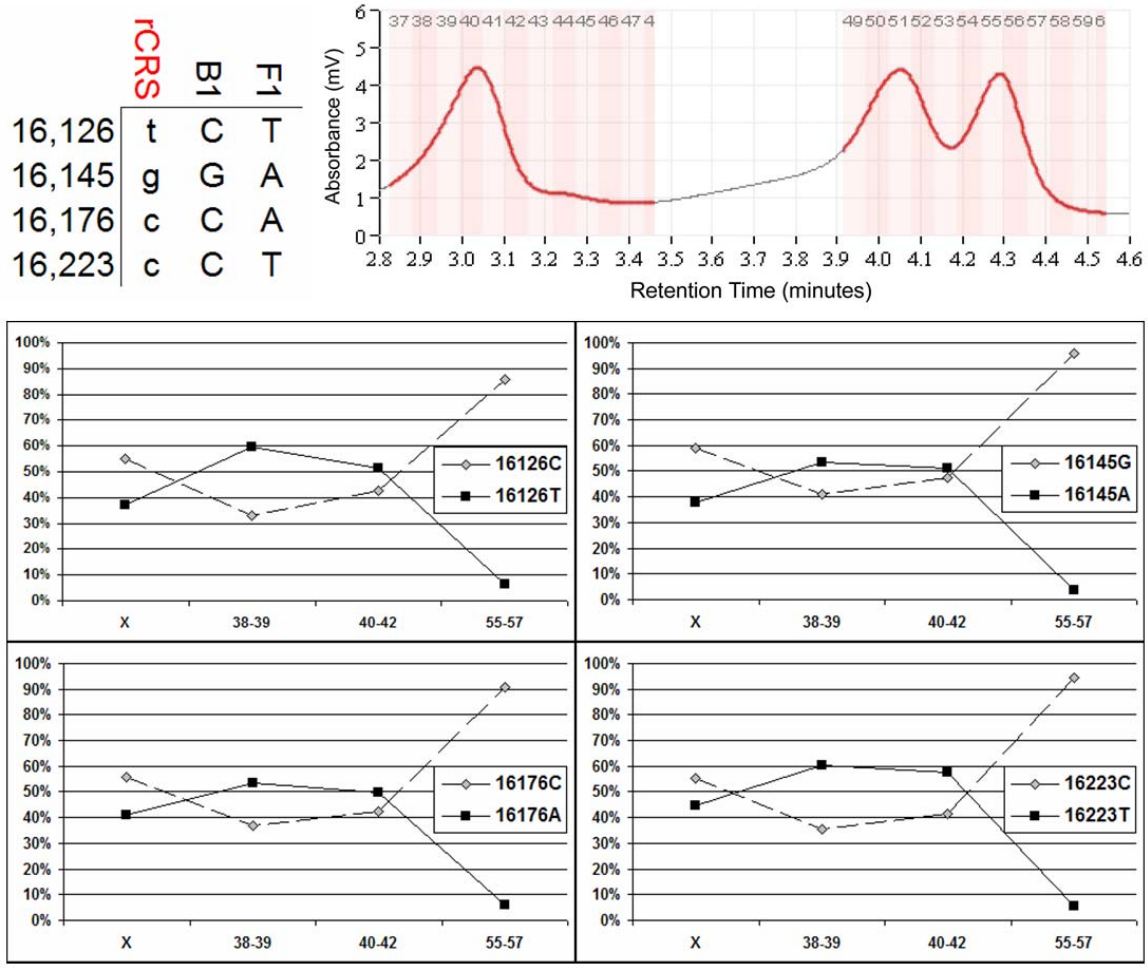


Figure 24: Linkage phase determination by quantitation of fluorescence peak height ratios at four mixed base positions from sequence electropherograms representing DNA recovered from 3 DHPLC fractions. Fraction X represents a non-fractionated 50/50 mixture of B1 (16126C, 16145G, 16176C, 16223C) and F1 (16126T, 16145A, 16176A, 16223T).

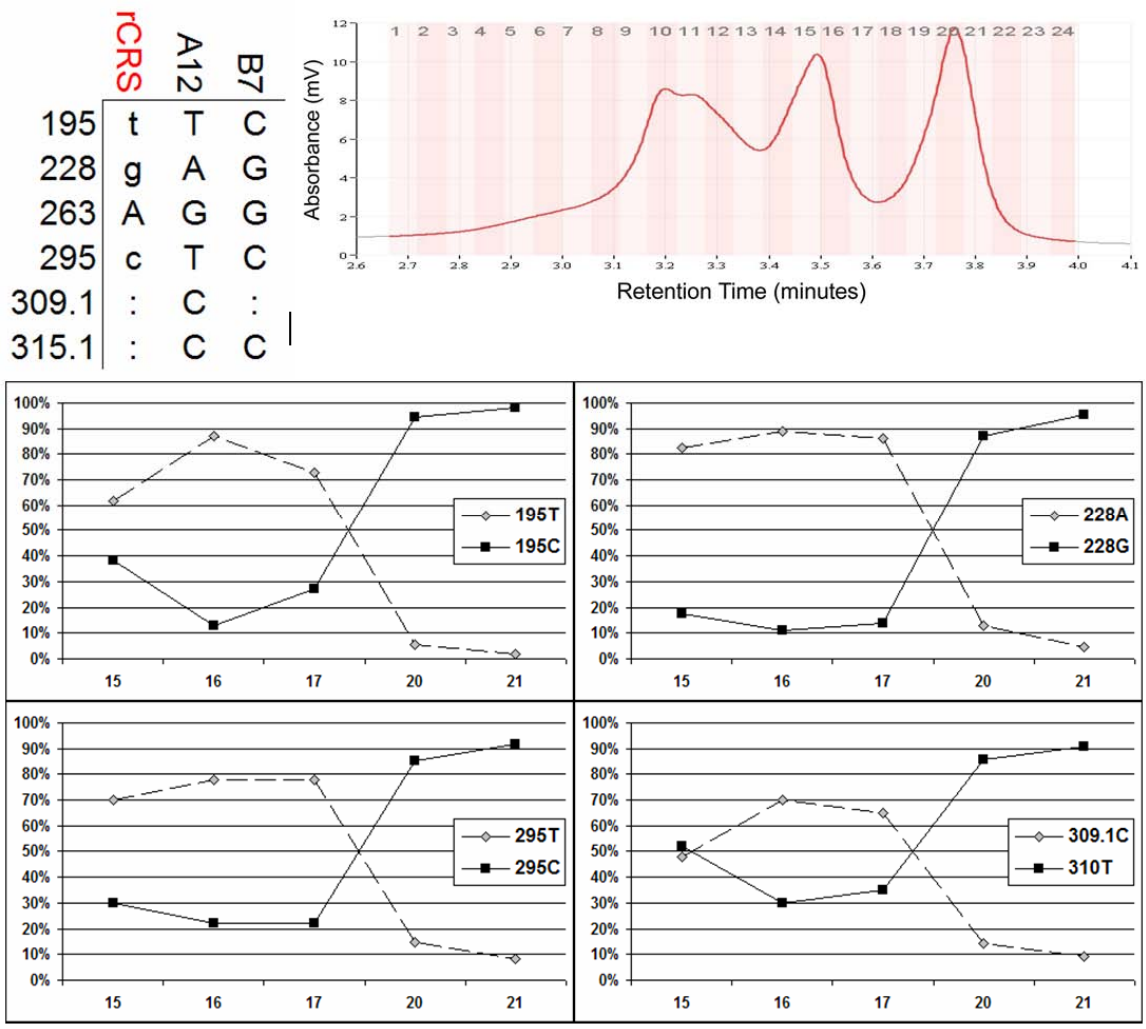


Figure 25: Linkage phase determination by quantitation of fluorescence peak height ratios at four mixed base positions from sequence electropherograms representing DNA recovered from 5 DHPLC fractions of a 40/60 mixture of A12 (195T, 228A, 295T, 309.1C) and B7 (195C, 228G, 295C, 310T).

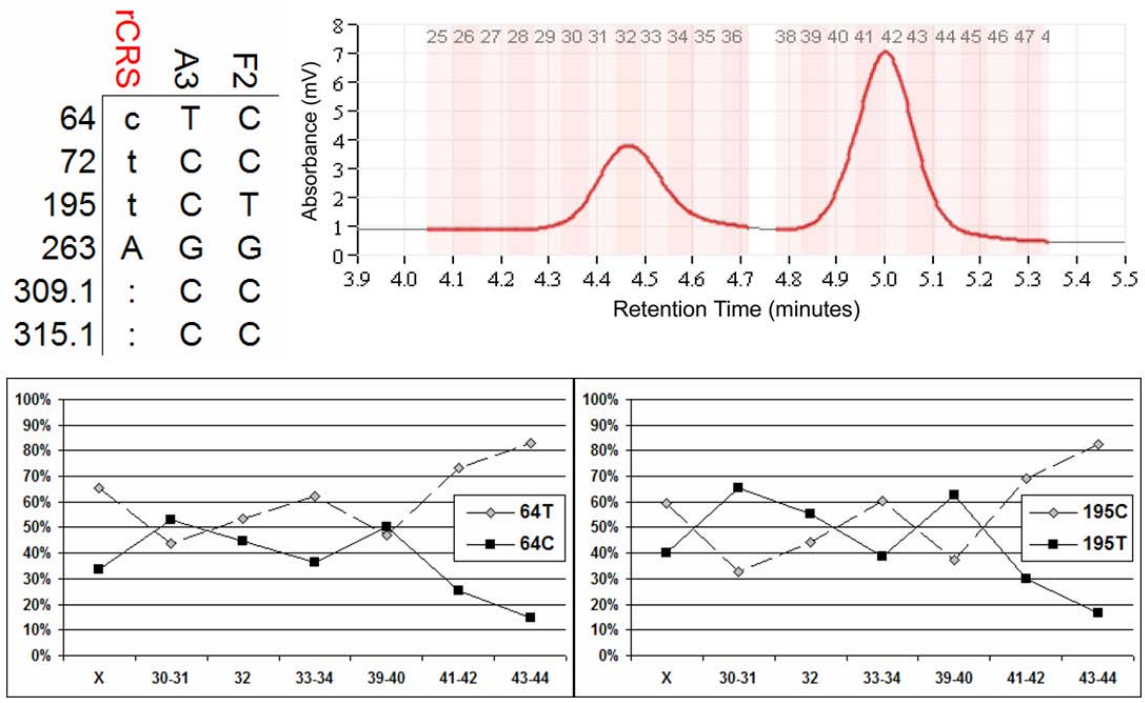


Figure 26: Linkage phase determination by quantitation of fluorescence peak height ratios at two mixed base positions from sequence electropherograms representing DNA recovered from 6 DHPLC fractions. Fraction X represents a non-fractionated 50/50 mixture of B1 A3 (64T, 195C) and F2 (64C, 195T).

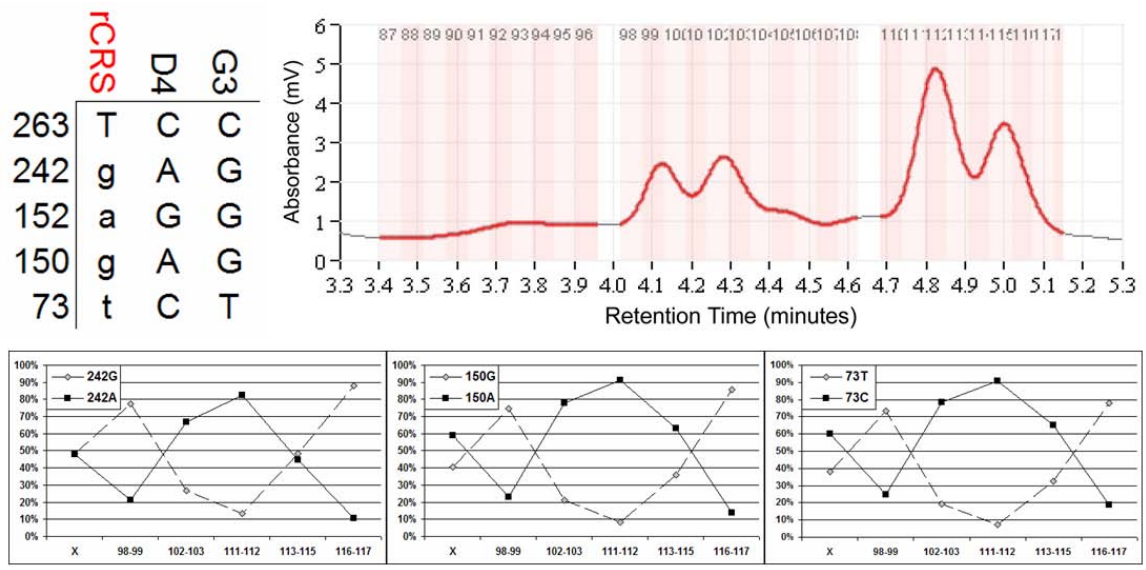
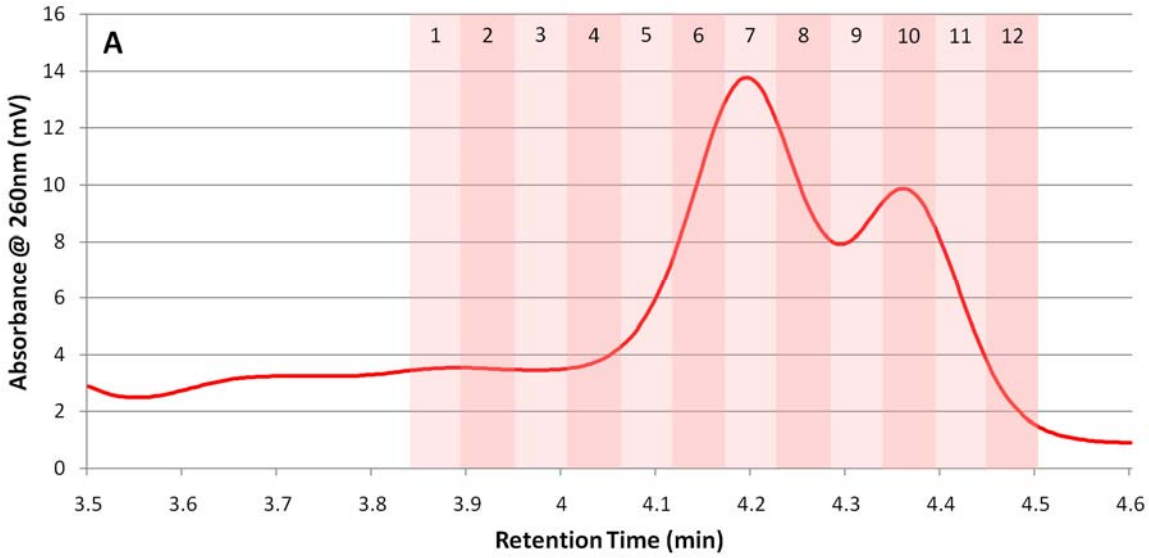


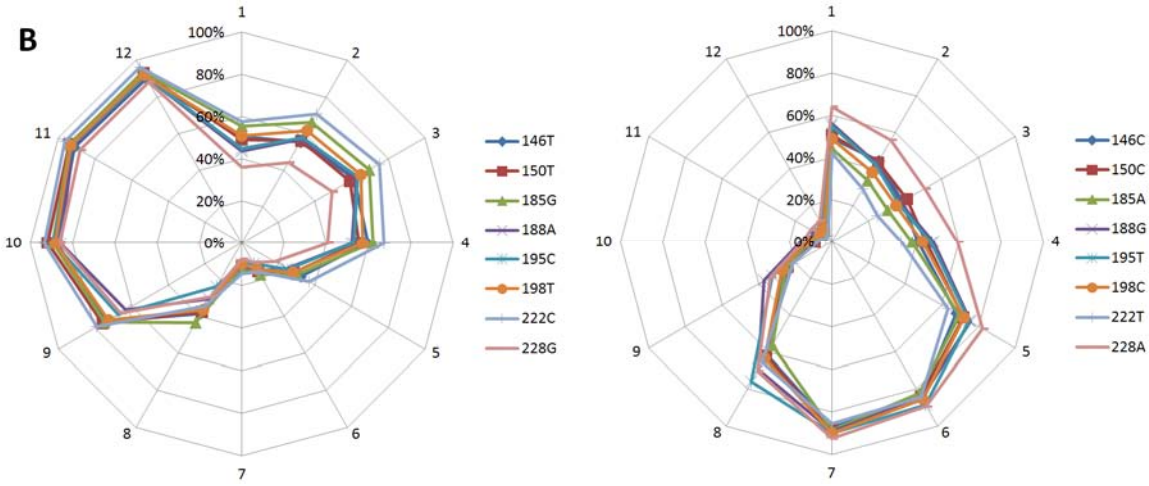
Figure 27: Linkage phase determination by quantitation of fluorescence peak height ratios at three mixed base positions from sequence electropherograms representing DNA recovered from 5 DHPLC fractions. Fraction X represents a non-fractionated 50/50 mixture of D4 (242A, 150A, 73C) and G3 (242G, 150G, 73T).

Efficiency of Component Identification as a Function of Chromatographic Trace Characteristics **-** DHPLC trace characteristics have a weighty effect on the efficiency of mixture separation. Mixture analysis by TMHA, yields chromatographic traces ranging from a major peak complimented by a significant shoulder, to multiple peaks representing each of the homo- and heteroduplices formed during denaturation and subsequent renaturation of the DNA present in the sample. In a two-component mixture the optimal DHPLC trace displays complete separation of each homo- and heteroduplex peaks.

Mixtures which display optimal resolution, i.e, where even the two homoduplices resolve into two distinct peaks of close or equal height, generally yield the most pronounced coordinated shift in peak height fluorescence ratios. Figure 28 illustrates the fractionation and LPA analysis of a two-component mixture with optimal resolution. The two homoduplices were captured by automated fraction collection (Figure 28A) and peak height ratios determined by dye terminator sequencing (Figure 28B). Comparison of all fraction combinations sequenced, yielded an average peak height ratio shift of 34.2 ± 21.1%. Of these, the comparison of fractions 7 and 11 yields the greatest coordinated shift in peak height fluorescence ratios across all mixed bases, with an average of 81.8 ± 1.5% and base calling confidence greater than 99.99% for all mixed base positions (Figure 28C). Confidence determinations of shifts in peak height fluorescence ratios are based on developmental validation studies performed on 202 mixed nucleotide positions. Comparison of fractions 11 and 12 yields the least shift in peak height fluorescence ratios, with a mixed base average of 0.77 ± 0.47%. The particularly low shift in peak height ratio between fractions 11 and 12 is due to comparisons between fractions that, for the most part, represent the same amplicon within the mixture. This comparison, therefore, adds no value to the overall analysis of the mixture. Furthermore, the shift in peak height between fractions 11 and 12 is comparable to the internal variability noted in replicate sequencing reactions of mixtures consisting of the same amplicon ratios as well as the inherent peak height variability of individual peaks in any sequencing reaction. Although fractions 7 and 11 yielded the greatest amount of fluorescence ratio shift across all mixed nucleotides, the majority of fraction comparisons (49 out of 66) yield an average base calling confidence greater than 99.9%. The remaining 17 fraction comparisons, where average base calling confidence is less than 99.9%, are as a result of two phenomena. These are either the direct comparison of fractions subsequent or very close to each other (*i.e.*, fractions 6 and 7) and as a result of fractions flanking the same peak (*i.e.*, fractions 5 and 8).

Table 7 illustrates the results obtained from 51 two-component mixtures at component molar ratios ranging from 50:50 to 95:5. These samples display a range of DHPLC trace characteristics, from one major peak and a corresponding shoulder, to complete separation of all four homo- and heteroduplex peaks. All amplified samples were fractionated accordingly by DHPLC under partially denaturing conditions. Three to twelve fractions for each sample were analyzed by direct sequencing. All fractions sequenced were analyzed for correct base calls and subsequently identified mixed base location were analyzed by LPA in order to determine the coordinated shift in peak height fluorescence ratios. Eighty-two percent of analyzed mixtures yielded average confidence of greater than 99.9%, when comparing fractions resulting in the greatest shift in peak height fluorescence. Of the nine remaining mixtures, only two yielded average base calling confidence of less than 99% (CS-38; 98.922% and CS-49; 98.557%).





C

Linkage Phase Determination	Position	Comparison 7 and 11	Average Confidence	Confidence Std Dev	Minimum Confidence	Samples Behind Stats	Increasing	Decreasing
	146	78.7%	99.998%	0.007%	99.945%	192	146T	146C
	150	82.8%	99.998%	0.006%	99.950%	190	150T	150C
	185	82.4%	99.998%	0.006%	99.950%	190	185G	185A
	188	81.3%	99.998%	0.006%	99.962%	204	188A	188G
	195	82.7%	99.998%	0.006%	99.950%	190	195C	195T
	198	83.6%	99.998%	0.004%	99.967%	214	198T	198C
	222	82.2%	99.998%	0.006%	99.950%	190	222C	222T
	228	80.8%	99.998%	0.003%	99.979%	184	228G	228A

Figure 28: (28A) Denaturing HPLC chromatogram, showing collected fractions, including fractions submitted to dye-terminator sequencing. (28B) Relative quantitation of fluorescence peak height ratios (y-axis) for 8 mixed base positions from sequence electropherograms representing DNA recovered from 12 DHPLC fractions (x-axis). (21C) Linkage phase analysis determination and identification of individual contributors to the mixture using relative fluorescent intensity at fractions 7 and 11 as determined by DHPLC fractionation.

Table 7 - Summary of mtDNA mixtures analyzed by DHPLC illustrating nucleotide enrichment and confidence ranges of mixed nucleotides detected.

						All Sequenced Fractions		Maximum Δ Fluorescense		
Sample	Ratio	DHPLC peaks	Primer	Fractions Sequenced	Mixed sites detected	Ave Δ Fluorescence Range	Ave Δ Fluorescence	Δ Fluorescence Range	Confidence Ave.	Confi Stc
CS-1	50:50	3	A1	4	4	6.2% - 55.2%	29.7%	52.7% - 57.7%	99.996%	0.0
CS-2	50:50	3	A1	4	2	1.1% - 10.4%	5.4%	9.3% - 11.6%	99.754%	0.0
CS-3	50:50	4	A2	7	3	0.7% - 44.4%	19.5%	42.5% - 47.2%	99.993%	0.0
CS-4	50:50	3	B1	7	8	2.1% - 26.0%	13.1%	19.8% - 30.2%	99.974%	0.0
CS-5	50:50	4	B1	7	5	3.0% - 31.6%	15.2%	21.6% - 39.2%	99.982%	0.0
CS-6	50:50	3	B2	3	4	4.2% - 8.3%	5.7%	4.8% - 10.0%	99.451%	0.5
CS-7	50:50	3	B2	5	2	1.8% - 18.8%	8.8%	13.5% - 24.0%	99.923%	0.0
CS-8	50:50	4	C1	5	10	1.8% - 62.5%	30.1%	52.2% - 67.1%	99.997%	0.0
CS-9	50:50	3	C1	6	10	2.1% - 59.1%	31.0%	33.5% - 75.0%	99.996%	0.0
CS-10	50:50	5	C1	4	3	1.0% - 62.2%	31.3%	60.9% - 63.0%	99.997%	0.0
CS-11	50:50	4	C1	7	4	3.6% - 57.1%	21.6%	54.0% - 65.0%	99.995%	0.0
CS-12	50:50	4	C1	6	5	1.7% - 74.6%	29.2%	71.3% - 77.9%	99.997%	0.0
CS-13	50:50	2	C1	7	2	2.2% - 44.9%	19.7%	39.9% - 49.9%	99.994%	0.0
CS-14	50:50	4	C1	6	3	3.3% - 74.9%	33.7%	72.8% - 76.2%	99.998%	0.0
CS-15	50:50	4	D2	6	10	1.3% - 74.4%	36.7%	57.8% - 92.3%	99.998%	0.0
CS-16	50:50	2	D2	7	2	2.5% - 44.5%	19.7%	40.0% - 49.0%	99.993%	0.0
CS-17	50:50	4	D2	6	3	3.9% - 75.4%	36.0%	73.2% - 77.6%	99.998%	0.0
CS-18	60:40	2	B1	6	7	3.1% - 28.5%	15.3%	21.0% - 35.9%	99.977%	0.0
CS-19	60:40	2	B1	3	8	7.8% - 22.5%	15.0%	16.3% - 27.0%	99.959%	0.0
CS-20	60:40	4	C1	5	10	7.2% - 78.2%	44.2%	70.5% - 84.7%	99.998%	0.0
CS-21	60:40	4	C1	6	10	7.1% - 76.2%	39.3%	65.0% - 88.8%	99.997%	0.0
CS-22	60:40	4	C1	3	10	3.9% - 76.3%	51.0%	66.7% - 85.7%	99.998%	0.0
CS-23	60:40	4	C2	3	5	4.0% - 24.4%	16.5%	16.7% - 32.5%	99.960%	0.0
CS-24	60:40	4	C2	10	6	1.8% - 70.3%	25.3%	58.5% - 85.1%	99.997%	0.0
CS-25	60:40	4	D2	6	10	3.8% - 72.9%	38.6%	60.4% - 79.7%	99.998%	0.0
CS-26	60:40	4	D2	6	10	7.1% - 71.1%	36.5%	56.4% - 79.3%	99.997%	0.0
CS-27	60:40	4	D2	3	10	6.0% - 77.7%	52.7%	58.9% - 87.6%	99.998%	0.0
CS-28	70:30	2	A1	13	4	0.3% - 29.9%	13.0%	24.5% - 35.8%	99.980%	0.0
CS-29	70:30	2	A1	13	4	0.4% - 30.2%	13.8%	27.3% - 34.3%	99.979%	0.0
CS-30	70:30	3	A1	7	2	1.4% - 26.3%	11.5%	20.1% - 32.4%	99.975%	0.0
CS-31	70:30	2	B2	7	2	0.9% - 26.2%	12.9%	26.0% - 26.3%	99.974%	0.0
CS-32	70:30	2	B2	9	2	0.4% - 32.0%	13.0%	26.9% - 37.1%	99.982%	0.0
CS-33	70:30	3	B2	10	2	0.7% - 25.1%	11.1%	24.5% - 25.7%	99.973%	0.0
CS-34	70:30	4	C1	12	8	0.8% - 81.8%	34.2%	78.7% - 83.6%	99.998%	0.0
CS-35	70:30	3	C2	7	4	1.1% - 15.3%	7.8%	12.1% - 17.6%	99.897%	0.0
CS-36	70:30	4	D2	12	8	1.1% - 82.0%	34.9%	73.7% - 87.1%	99.998%	0.0
CS-37	80:20	2	B1	7	7	1.9% - 20.8%	10.3%	15.2% - 27.9%	99.943%	0.0
CS-38	80:20	4	C1	4	9	0.7% - 7.3%	4.5%	2.1% - 12.6%	98.922%	1.5
CS-39	80:20	4	C1	5	10	0.9% - 46.0%	24.1%	34.7% - 54.1%	99.993%	0.0
CS-40	80:20	4	C1	6	10	2.1% - 78.1%	32.2%	67.2% - 89.1%	99.998%	0.0
CS-41	80:20	3	C2	6	6	2.6% - 38.5%	21.8%	22.3% - 55.6%	99.984%	0.0
CS-42	80:20	4	D2	5	10	0.6% - 29.5%	13.2%	17.8% - 45.2%	99.976%	0.0
CS-43	80:20	4	D2	5	10	1.7% - 45.4%	24.0%	26.5% - 59.8%	99.992%	0.0
CS-44	80:20	4	D2	5	10	1.9% - 22.8%	10.6%	5.7% - 37.5%	99.832%	0.3
CS-45	90:10	3	A2	7	3	0.7% - 18.3%	7.6%	8.0% - 26.1%	99.871%	0.1
CS-46	95:5	2	A1	7	2	0.5% - 26.8%	12.3%	19.0% - 34.6%	99.967%	0.0
CS-47	95:5	2	A2	7	3	0.4% - 12.9%	5.1%	5.4% - 18.5%	99.525%	0.6
CS-48	95:5	2	B2	7	2	0.3% - 22.1%	8.6%	11.2% - 33.0%	99.884%	0.1
CS-49	95:5	2	C2	7	4	0.2% - 4.5%	1.6%	3.4% - 6.5%	98.557%	0.7
CS-50	95:5	2	D2	10	8	0.4% - 34.5%	11.0%	26.7% - 43.8%	99.985%	0.0
CS-51	95:5	2	C1	12	8	0.3% - 40.8%	17.0%	26.8% - 47.8%	99.990%	0.0

Analysis of sample CS-38 by DHPLC using TMHA (Figure 29) shows separation of individual homoduplices into a major peak and a trailing shoulder (separated into six 50µL fractions), as well as slight formation of minor heteroduplices (not captured). The major reason for lack of apparent separation for CS-38 when analyzed by LPA is that the trailing edge of the shoulder was not sequenced (fractions 5 and 6). When compared to sequencing results obtained from the opposite sequencing primer (CS-42; where fractions 1-5 were sequenced), the shift in peak height fluorescence was significantly greater. These results illustrate that, while blind fraction sequencing yields significant information as to the haplotypes constructing the mixture, considerably improved results are attained when strategic fraction sequencing is performed.

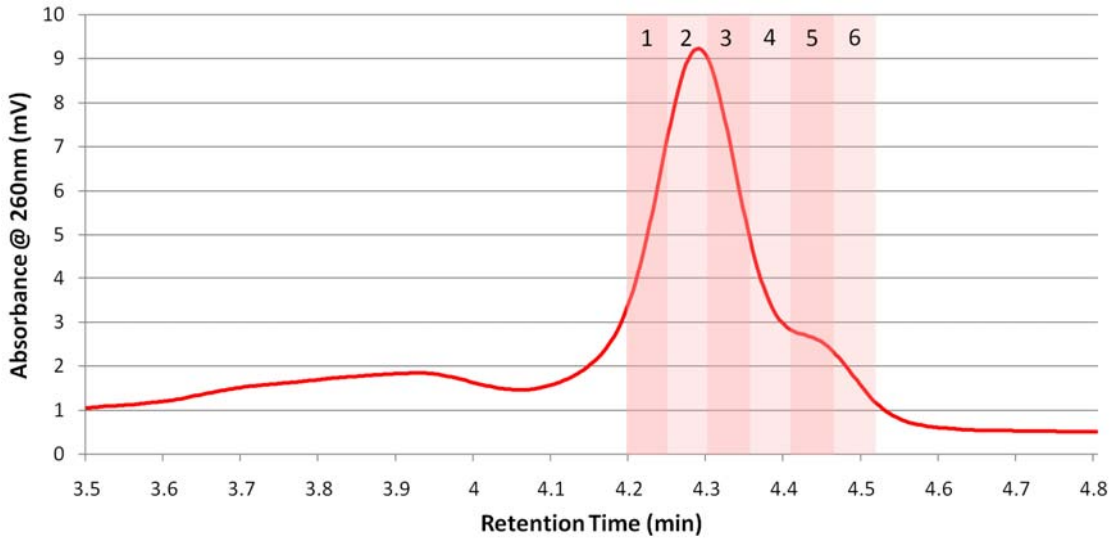


Figure 29: Analysis of sample CS-38 by DHPLC using showing separation of individual homoduplices into a major peak and a trailing shoulder.

Analysis of sample CS-49 by TMHA (Figure 30) shows the separation and capture of a single broad homoduplex peak (fractions 4-7) preceded by a separate single heteroduplex peak (fractions 1-2). Although only slight due to the nature of the mixture (a 95:5 component ratio), LPA shows the greatest coordinated shift in peak height fluorescence ratio between fractions 4 and 7. This illustrates that the broad nature of the homoduplex peak is due to the slight retention time differences between the two homoduplices formed during cross-hybridization. This may not be a visible difference as the minor component homoduplex is hidden by the major component peak; however, the variation in retention time is detected by LPA of sequences obtained from these fractions.

Of the samples displayed in table 7, samples displaying two, three and four peak resolution by DHPLC account for 29.4%, 21.6% and 47.1% of the samples tested. Comparison to a larger study of 849 mixtures subjected to DHPLC profiling across all four forensically relevant regions of the human mtDNA [7] reveals a slight decrease in the amount of samples yielding two and three peak resolution. In that study, samples displaying two, three and four peak resolution comprised 50.3%, 36.6% and 11.6% of all samples tested, respectively.

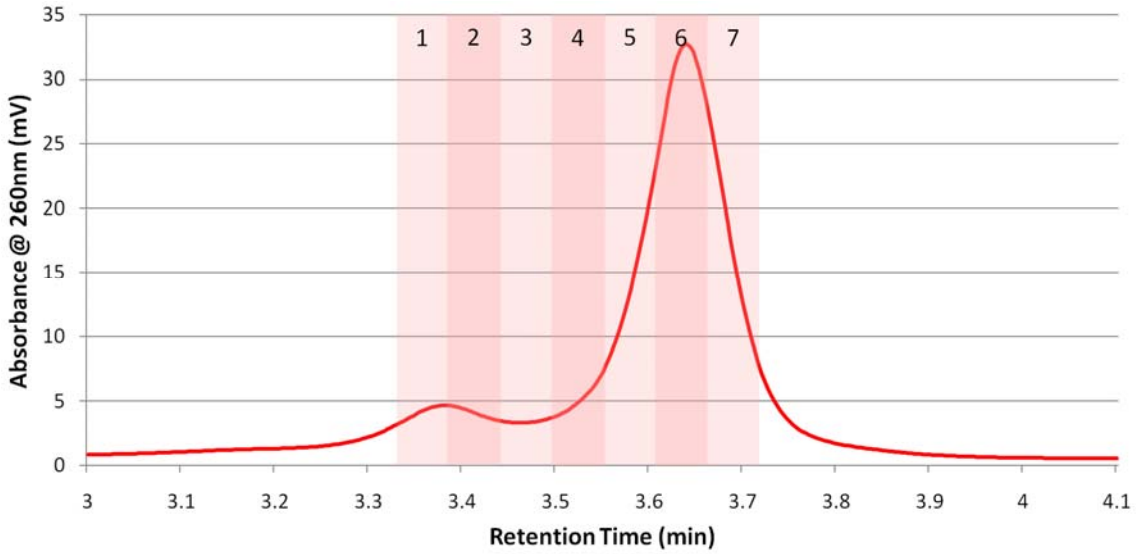


Figure 30: Analysis of sample CS-49 by DHPLC shows the separation of a 95:5 mixture into a single broad homoduplex peak (fractions 4-7) preceded by a separate single heteroduplex peak (fractions 1-2).

Efficiency of Component Identification as a Function of Mixture Ratio - Comparison of multiple mixture ratios for a two-component mixture yielded information regarding which fraction comparisons generate the most useful and consistent coordinated shifts in peak height fluorescence. Analysis of samples CS-8, 20, 21, 38 and 39 sequenced with the C1 primer and samples CS-15, 25, 26, 42 and 43 sequenced with the D2 primer illustrates which fraction comparison yield the greatest and the least coordinated shift. In all cases, the homoduplicies formed during cross hybridization of the amplified PCR products resolve into two distinct peaks, each representing one contributor. If the first peak is the tallest, the greatest coordinated shift in peak height fluorescence is achieved by comparing the lagging edge of the second peak to the top of the first peak (Figure 28A; fractions 7 and 11). If the second peak is the tallest, the same can be achieved by comparing the leading edge of the first peak to the top of the second peak. As described above, comparison of fractions subsequent to each other yields limited shift in peak height fluorescence due to minimal changes in contributor ratios between those fractions.

When homoduplices are not resolved into separate peaks, comparisons between the heteroduplex peaks and the top or lagging edge of the homoduplex peak generally yields the greatest coordinated shift in peak height fluorescence. This can be illustrated when comparing fractions sequenced for CS-29 (Figure 31; fractions 1-3 compared to fractions 11-13). Consistently, a large shift in peak height fluorescence ratio is observed when comparing a number of fractions on the lagging side of the heteroduplex peak to the top or lagging side of the homoduplex peak. When this is done, a consistent 20% - 30% coordinated shift is noted for all fraction comparisons performed between the homo- and heteroduplex peaks. Only fractions comparing the same peaks yield a low shift in peak height fluorescence.

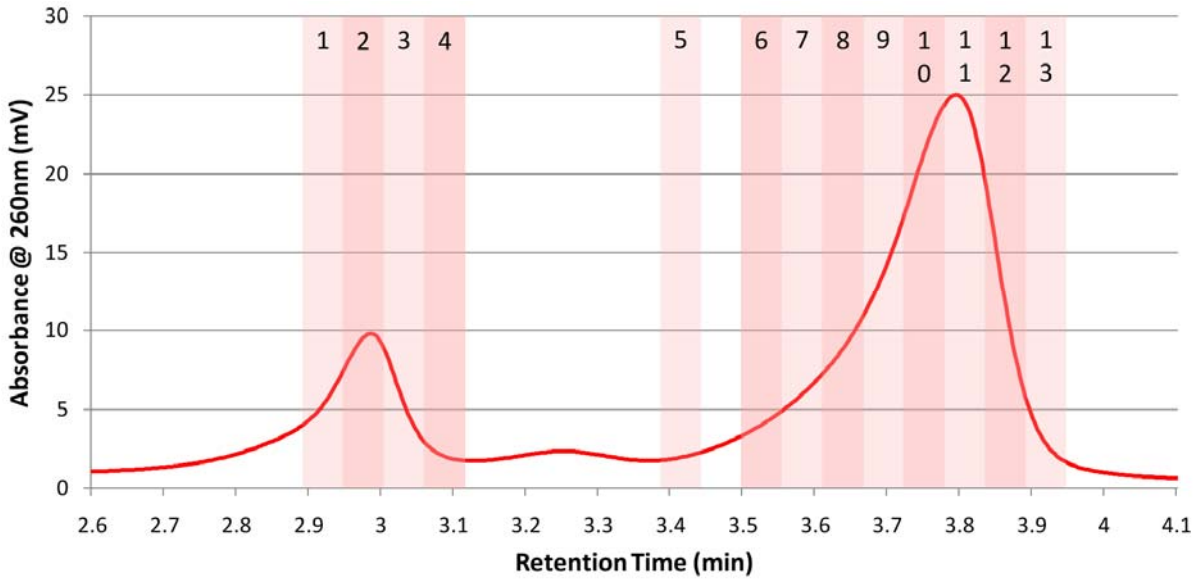


Figure 31: Analysis of sample CS-29 by DHPLC shows an example of separation where neither hetero- nor homoduplices are resolved into clearly separate peaks.

Body Fluid Mixtures – Forty-two two-component body fluid mixtures were stained onto a variety of substrates including denim, leather, wood, carpet, nylon and wallboard and aged at room temperature for four weeks followed by DNA extraction. No two donor samples used in this study had identical mtDNA haplotypes as previously determined by sequencing analysis of known buccal swabs. Amplification of control region fragments was performed on a subset of questioned mixtures using forensically validated primer sets and conditions. All amplified samples were fractionated accordingly by DHPLC under partially denaturing conditions. Five to nine fractions for each sample were analyzed by direct sequencing, followed by LPA of mixed base locations. Mixture sequencing yielded clean base calls for all but mixed nucleotides, which resulted in overlapping fluorescent peaks. Identification by LPA of individual haplotypes to each of the 18 mixtures tested was achieved in all cases for a success rate of 100% and all identified haplotypes were in concordance with sequencing results obtained from individual non-mixture samples. Mixtures were separated with a high degree of average base calling confidence depending on substrate and mutation, ranging from an average of 99.926% to 99.994% (Table 8) when comparing the greatest shift in peak height fluorescence.

Comparison of all fractions sequenced yielded an average change in peak height fluorescence ratio ranging from 7.7% - 24.5%, with fractions neighboring each other yielding the lowest shift in peak height fluorescence (0.2% - 2.7%). Greatest shift in peak height fluorescence ratio ranged from 16.3% - 51.5% and were found to fit the same general trend in optimal DHPLC fraction selection as the control samples presented above.

Environmental Insult Mixtures **-** Thirty six two-component body fluid mixtures were subjected to a variety of environmental insults including gasoline, soil, laundry detergent, used motor oil, sodium hydroxide and acetic acid followed by DNA extraction. Amplification of control region fragments was performed on a subset of mixtures using forensically validated primer sets and conditions. All amplified samples were fractionated

accordingly by DHPLC under partially denaturing conditions. Five to nine fractions for each sample were analyzed by direct sequencing, followed by LPA of mixed base locations. Mixture sequencing yielded good quality base calls for all but mixed nucleotides, which resulted in overlapping fluorescent peaks. Identification by LPA of individual haplotypes to each of the 18 amplified mixtures was achieved, for a success rate of 100% and all identified haplotypes were in concordance with sequencing results obtained from individual non-mixture samples. Mixtures were separated with a high degree of average base calling confidence depending on substrate and mutation, ranging from an average of 99.926% to 99.993% (Table 8) when comparing the greatest shift in peak height fluorescence.

Comparison of all fractions sequenced yielded an average change in peak height fluorescence ratio for all mixed bases ranging from 8.0% - 23.2%, with fractions neighboring each other yielding the lowest shift in peak height fluorescence (0.3% - 1.7%). Greatest shift in peak height fluorescence ratio ranged from 12.2% - 49.9% and were also found to fit the same general trend in optimal DHPLC fraction selection as the control samples.

Hair and Bone Mixtures - In a more forensically important experiment conducted, 9 hair samples were contaminated with various body fluids. Head, axial and pubic hairs collected were immersed in blood, semen and saliva and allowed to air dry, followed by DNA extraction. All amplified samples were fractionated accordingly by DHPLC and 5 to 8 fractions for each sample were analyzed by direct sequencing, followed by LPA of mixed base locations. Identification by LPA of individual haplotypes to each of the 9 amplified mixtures was achieved, for a success rate of 100% and all identified haplotypes were in concordance with sequencing results obtained from individual non-mixture samples. Mixtures were separated with a high degree of average base calling confidence depending on substrate and mutation, ranging from an average of 99.84% to 99.98% (Table 8) when comparing the greatest shift in peak height fluorescence.

Comparison of all fractions sequenced yielded an average change in peak height fluorescence ratio ranging from 5.2% - 18.5%, with fractions neighboring each other yielding the lowest shift in peak height fluorescence (0.2% - 2.5%). Greatest shift in peak height fluorescence ratio ranged from 11.0% - 44.5% and were also found to fit the same general trend in DHPLC fraction selection as the control samples.

Four bone samples obtained from the University of Montana's Department of Anthropology were pulverized followed by DNA extraction. One bone sample yielded no amplifiable product even following a second round of extraction and amplification, while two of the samples yielded a mixed mtDNA profile. The lack of amplifiable product for one of the bones was most likely due to extensive degradation of the samples, as all of the bones had been stored for multiple years at room temperature. Mixture analysis of SCS-119 revealed a two-component mixture at five nucleotides. Mixture fractionation and LPA of six captured fractions yielded 48.7% - 61.6% shift in peak height fluorescence when comparing the greatest difference in peak height ratio. This corresponded to an average base calling confidence of greater than 99.99%.

The shift in peak height fluorescence ratio for six fractions captured from SCS-120 revealed five mixed nucleotide positions. Downstream LPA revealed a secondary component as the contributor to two of the mixed sites (average base calling confidence >99.9), while a third component was identified as the contributor to the additional three mixed sites (average base calling confidence >99.9). Although the third component was not identified by DHPLC, due to its low concentration, LPA of the resulting sequencing analysis clearly revealed a third contributor to the mixture.

Chromatographic characteristics or number of mutations within each amplicon mixture revealed no trend in the efficiency of separation for all of the samples tested in this study, with two peak traces yielding the lowest and greatest average fluorescence shifts.

Table 8 - Summary of simulated mtDNA mixtures analyzed by DHPLC illustrating nucleotide enrichment and confidence ranges of all mixed nucleotides detected.

								All Sequenced Fractions			Maximum Δ Fluorescense			
Sample	Substrate	Contributor 1	Contributor 2	DHPLC peaks	Primer	Fractions Sequenced	Mixed sites	min	max	Ave Δ Fluorescence	min	max	Confidence Ave.	Confidence Stdev.
SCS-2	denim	blood	blood	5	B1	9	4+LH	2.4%	35.8%	17.0%	29.3%	41.2%	99.987%	0.007%
SCS-3	leather	blood	blood	5	B1	9	4+LH	1.5%	38.9%	18.9%	29.1%	51.0%	99.988%	0.008%
SCS-6	wood	blood	blood	5	B1	7	4+LH	2.7%	35.0%	17.4%	28.4%	40.7%	99.987%	0.006%
SCS-7	carpet	semen	semen	5	B1	9	4+LH	0.8%	40.2%	18.8%	28.4%	51.5%	99.989%	0.007%
SCS-10	nylon	semen	semen	5	B1	8	4+LH	0.2%	40.7%	19.7%	31.6%	46.9%	99.989%	0.004%
SCS-13	carpet	saliva	saliva	4	A2	9	4	1.3%	26.5%	11.7%	19.2%	30.4%	99.973%	0.013%
SCS-15	leather	saliva	saliva	4	A2	9	4	0.4%	26.6%	11.7%	17.4%	31.8%	99.971%	0.020%
SCS-17	wallboard	saliva	saliva	4	A2	9	4	0.3%	25.2%	11.6%	21.4%	27.0%	99.970%	0.004%
SCS-18	wood	saliva	saliva	4	A2	9	4	0.7%	23.6%	10.4%	17.2%	27.5%	99.963%	0.014%
SCS-19	gasoline	blood	blood	5	B1	8	4+LH	0.6%	37.5%	19.7%	29.0%	46.7%	99.987%	0.007%
SCS-21	motor oil	blood	blood	5	B1	9	4+LH	1.5%	41.1%	19.2%	33.0%	49.9%	99.991%	0.004%
SCS-22	detergent	blood	blood	5	B1	8	4+LH	1.3%	37.8%	19.7%	28.8%	45.8%	99.988%	0.006%
SCS-24	HAC	blood	blood	5	B1	7	4+LH	0.8%	39.5%	21.8%	27.7%	49.4%	99.987%	0.011%
SCS-29	NaOH	semen	semen	4	B1	6	4+LH	3.2%	38.6%	16.4%	31.0%	48.9%	99.989%	0.006%
SCS-31	gasoline	saliva	saliva	4	A2	9	4	1.5%	27.4%	12.1%	20.2%	33.0%	99.972%	0.015%
SCS-33	motor oil	saliva	saliva	4	A2	9	4	1.6%	28.4%	12.1%	24.7%	33.3%	99.977%	0.007%
SCS-36	HAC	saliva	saliva	4	A2	8	4	1.1%	26.9%	12.6%	19.1%	30.4%	99.972%	0.013%
SCS-38	denim	blood	semen	3	C1	7	2	1.5%	18.4%	7.7%	16.3%	20.4%	99.920%	0.047%
SCS-39	leather	blood	semen	3	C1	7	2	0.6%	21.5%	8.4%	16.8%	26.1%	99.930%	0.061%
SCS-42	wood	blood	semen	3	C1	7	2	0.3%	25.5%	10.0%	19.3%	31.8%	99.969%	0.023%
SCS-43	carpet	saliva	semen	2	C1	5	3	0.7%	28.7%	14.2%	19.9%	33.8%	99.977%	0.020%
SCS-46	nylon	saliva	semen	2	C1	6	3	0.7%	27.8%	12.5%	19.2%	32.6%	99.976%	0.020%
SCS-49	carpet	blood	saliva	2	A2	7	3	0.8%	32.1%	16.3%	31.1%	33.1%	99.986%	0.001%
SCS-51	leather	blood	saliva	2	A2	7	3	0.9%	44.6%	24.5%	43.0%	47.4%	99.994%	0.000%
SCS-53	wallboard	blood	saliva	2	A2	7	3	0.6%	42.8%	23.1%	42.3%	43.5%	99.993%	0.001%
SCS-54	wood	blood	saliva	2	A2	7	3	1.0%	42.1%	23.2%	39.8%	46.1%	99.992%	0.001%
SCS-55	gasoline	blood	semen	3	C1	7	2	0.6%	20.8%	8.3%	17.3%	24.3%	99.957%	0.021%
SCS-56	soil	blood	semen	3	C1	6	2	0.5%	20.8%	8.0%	15.6%	26.1%	99.925%	0.069%
SCS-58	detergent	blood	semen	3	C1	6	2	0.8%	22.1%	9.4%	17.0%	27.1%	99.957%	0.021%
SCS-59	NaOH	blood	semen	3	C1	6	2	0.7%	19.8%	9.0%	17.6%	22.1%	99.955%	0.018%
SCS-61	gasoline	saliva	semen	2	C1	6	3	0.7%	18.4%	8.2%	12.2%	21.7%	99.911%	0.094%
SCS-65	NaOH	saliva	semen	2	C1	5	3	1.7%	26.5%	12.8%	24.8%	28.1%	99.975%	0.004%
SCS-69	motor oil	blood	saliva	2	A2	7	3	0.3%	43.4%	23.2%	41.5%	47.2%	99.992%	0.002%
SCS-70	detergent	blood	saliva	2	A2	7	3	0.6%	39.3%	21.4%	37.1%	41.5%	99.990%	0.002%
SCS-72	HAC	blood	saliva	2	A2	7	3	0.5%	41.9%	23.1%	39.4%	46.1%	99.992%	0.001%
SCS-119	Aged	Bone 19	Bone 21	4	C1	6	4	0.8%	55.4%	20.7%	49.6%	61.3%	99.996%	0.001%
SCS-120	Aged	Bone 20	Bone 21	3	C1	6	5	0.3%	21.6%	13.0%	11.2%	39.1%	99.911%	0.089%
SCS-H2	dyed	dyed hair	blood	5	A1	6	2+LH	1.0%	18.6%	8.7%	15.6%	21.7%	99.920%	0.063%
SCS-H3	None	head hair	blood	5	A1	8	2+LH	1.5%	32.3%	11.9%	31.3%	33.4%	99.987%	0.001%
SCS-H4	None	axial hair	blood	3	A1	7	2	0.2%	19.0%	8.2%	13.7%	24.3%	99.920%	0.073%
SCS-H5	None	pubic hair	semen	5	A1	8	2+LH	0.9%	29.8%	11.4%	28.8%	30.7%	99.980%	0.002%
SCS-H6	permed	permed hair	blood	4	A1	5	3+LH	2.0%	13.2%	6.6%	11.0%	16.1%	99.821%	0.058%

Linkage Phase Analysis software and Standard operating procedures for DHPLC mixture resolution

While individual contributors to an mtDNA mixture are resolved by determining the linkage phase of nucleotides at mixed base positions. this process requires complex sequence alignments and computationally intensive analyses of enormous datasets. This task is too large for a single person to manage "in their head". Accordingly, it was necessary to develop a bioinformatics application to organize and analyze these datasets. This will facilitate adoption of DHPLC by laboratories and the courts.

The bioinformatics solution developed to analyze mixture data and identify contributors has been termed FLiPARS (Fractional Linkage Phase Analysis Resource System). This is a conglomeration of several smaller applications integrated into a Microsoft Windows Visual Basic application. It is designed to streamline the data analysis process of electropherogram data derived from the mtDNA mixture fraction linkage phase analysis method. The smaller portions of the application are Perl modules which may be accessed by the FLiPARS Windows application in order to gather sequence and electropherogram data from Raw format sequence files (Tab delimeted .txt files [Beckman Format]). In addition these smaller portions of the program also perform sequence alignment using the Needleman-Wunsch alignment algorithm. Though very basic, this algorithm provides satisfactory subject to reference sequence alignment for the human hypervariable regions. The software allows rapid linkage phase determination through a three-step process as follows:

Step 1: Data Input

- Analyst inputs case and sample specific information via a graphical user interface.
- Analyst selects one or more DHPLC fractions for inclusion in linkage phase determination. This utilizes a direct connection to the WAVE 3500 DHPLC Oracle database.
- Sequence files from the Prism 310 Genetic Analyzer (electropherograms) are then associated with each DHPLC fraction.

Step 2: Data Extraction

- Perl module parses all relevant DNA sequence data corresponding to each selected DHPLC fraction.
- Sequence data are aligned to a reference sequence using the Needleman-Wunsch global alignment algorithm.
- All ambiguous base positions are identified and stored in a 2D-array.

Step 3: Linkage Phase Analysis

- Ambiguous base positions are selected for analysis via the graphical user interface.
- Subroutines gather all fluorescence data for each selected base and consolidate them into data structures.

- Consolidated data are statistically weighted via linkage phase analysis.
- Calculated data are output as a report to an Excel worksheet.
- Linkage phase of individual contributors is identified and assigned a statistical confidence (Tukey's HSD) for each base position called.

Like many open-source, collaborative software applications, FLiPARS has always been a constantly evolving entity – It saw it's beginning as a very rough Perl script and progressed into a fully functioning Graphical User Interface.

However, FLiPARS is still a work in progress and several areas remain to be improved, these include but are not limited to:

- Enhanced data processing – less lag time on computationally intensive processes via more efficient coding.
- Greater database integration – Expansion of the software capabilities to store and retrieve data from data sources such as Microsoft Access databases or various SQL database servers.
- "Smoother" user interface – Many aesthetic enhancements may be made.

FLiPARS is a fully-functional software application with a great potential for growth, it is ideally suited as an open-source software application which may be improved upon via testing and feedback from selected users or testers.

An initial set of Standard Operating Procedures for the use of DHPLC in mtDNA analysis has been developed and is being used by the Danielson Laboratory with entry level DNA analysts. These are provided in Appendix I.

[Implications for policy and practice]

DHPLC as a Reliable Means of Mixture Resolution and as a Means of Streamlining the mtDNA Analysis Workflow – The successful completion of this research now has the potential to greatly benefit forensic practitioners engaged in the analysis of mtDNA by making it possible to efficiently fractionate and interpret mixtures of mtDNA so that the mitotypes of the individual contributors can be accurately determined. This makes it possible to obtain useful forensic information from samples that are not amenable to analysis by direct sequencing alone. Given the ability of DHPLC to separate and collect fractions of mixtures, one could ascribe a component of the mixture to a missing person while the remaining fraction could be ascribed to a putative suspect. This will allow for the uploading/comparison of additional profile to the coming CODIS plus mtDNA database. In the case of a missing person, a mixture may reflect a sample that is heteroplasmic in nature and it may be possible to confirm this by comparison to maternal relatives. In such cases, the ability to resolve and subsequently sequence the individual mtDNA components of a heteroplasmic mixture will strengthen the resolving power of the assay. Thus, DHPLC offers the opportunity to facilitate and expand the use of mtDNA in criminal investigations.

The federal rules of evidence provide the standard to admitting expert scientific testimony in the federal courts, Daubert v. Merrell Dow Pharmaceuticals, Inc., 509 U.S. 579, 593-94 (1993). Many state courts have also superseded *Frye's* "general acceptance" test, Frye v. United States, 293 F. 1013, 1014 (D.C. Cir. 1923), with the federal rules of evidence, specifically Rules CRE 403 and CRE 702. The developmental validation studies completed in the course of this research program coupled with publication of findings in peer-reviewed journals will help to place the technology on sound legal footing with regard to the reliability standard set forth in CRE 702. The current studies have shown that the method has been tested, that the underlying reasoning is scientifically valid, that the potential error rates and limitations have been evaluated and that appropriate standards have been adopted to control the operation of the system. The successfully completed research program on DHPLC was conducted with both the *Frye* and *Daubert* standards in mind.

In addition to its utility as a means of fractionating DNA mixtures, DHPLC can also be used as an alternative to standard gel electrophoresis. Under non-denaturing conditions the molecular size of a target amplicon and the quantity of DNA present in a peak can be accurately determined. After extraction and amplification using established forensic SOPs, it is possible to accurately quantity the PCR yield from the initial chromatographic trace. This is a function of the fact that chromatographic peaks reflect the quantity of DNA eluted from the HPLC column, the area of a target DNA peak provides an accurate means of assessing PCR yield for a target amplicon (R^2 = 0.995). In the course of the current study, the potential utility for use of a commercial HPLC system to quantify PCR product yield and simultaneously purify target amplicons for downstream applications such as DNA sequencing was evaluated. In order to demonstrate the efficacy of such an approach, the linearity of DNA quantitation by HPLC was assessed across a range of amplicon sizes and concentrations of potential utility to forensic analysts.

The size resolution by DHPLC is superior to that of standard agarose electrophoresis and approaches that of polyacrylamide gel electrophoresis. The time required for a size and quantity analysis is usually less than 7 minutes for fragments under 500 bp. Thus DHPLC streamlines the processing of forensic samples by replacing yield gels and the column purification of PCR products prior to the sequencing labeling reaction. PCR products can usually be loaded directly onto the DHPLC column and the captured fractions vacuum dried and directly sequenced without additional processing. Even in cases where a DNA mixture is not present, DHPLC can be used to purify PCR products in preparation for direct sequencing. As a general clean up method, DHPLC is equivalent to PAGE purification but is much less labor intensive and can be completed in a matter of minutes. Similarly, the suitability of DHPLC as a means of amplicon purification was assessed in comparison to existing PCR product cleanup methods employed by forensic laboratories. Based on the resulting signal-to-noise ratio seen in sequenceing electropherograms for amplicons purified by ExoSAP-IT®, Amicon Centricon-50® spin columns and DHPLC, the purity of samples prepared by DHPLC was equivalent to or slightly better than the alternate cleanup methods tested.

Thus, if a laboratory sought to fully integrate a DHPLC system into their process flow, it becomes possible accurately determine the quantity of amplified product and then to screen all amplified samples for the presence of a mixture, either a genuine DNA mixture or naturally occurring heteroplasmy. Finally, DHPLC could be used to fractionate mixtures (when present) or to simply purify non-mixed samples for sequencing.

The results of the current study have also demonstrated that DHPLC analysis of pair-wise combinations of identical mtDNA amplicons, accurately and reliably produces a single chromatographic peak consistent with sequence concordance. These results were in 100% agreement with DNA sequence data. Conversely, in pair-wise combinations of non-identical amplicons, DHPLC successfully detects a diversity of sequence differences throughout the HV1 and HV2 regions. These differences, which include a wide variety of base substitutions as well as insertions/deletions, are typically indicated by the presence of more than a single peak in the resulting chromatogram. As such, DHPLC may have significant forensic utility in several areas. These include: a presumptive test of mtDNA concordance between known and questioned samples; a screen for mixed samples prior to direct sequencing and; a preparative tool for the physical fractionation of the individual contributors to an mtDNA mixture prior to sequencing.

Although DHPLC is not a replacement for direct sequencing of mtDNA, it does offer advantages as a potential screening tool. First, the assay is relatively simple and fast. It uses raw PCR products thereby avoiding the time and expense associated with amplicon cleanup. Following cross-hybridization, each assay takes only seven minutes to run and interpretation of the results is straightforward. DHPLC provides a comprehensive assessment of sequence concordance across an entire amplicon without the challenging task of trying to obtain quality base sequence information immediately adjacent to primer binding sites. Finally, the presence of heteroplasmy, while a potentially confounding variable, does not necessarily preclude comparative sequence analysis by DHPLC.

Compared to alternate mtDNA screening strategies based on oligonucleotide probes or linear arrays, DHPLC consumes less DNA and is not limited by the need to design probes

for the detection of known mutations at predetermined polymorphic sites. This reduces the potential for false inclusions and eliminates the need to design custom probes for unique or rare sequence variants. Similarly, DHPLC assays are not subject to the "null" or "blank" results that arise when hybridization of the target sequence is impeded by other nearby polymorphisms. On the contrary, additional sequence variants typically make it easier to detect sequence non-concordance between two amplicons.

It has been reasonably argued by experienced practitioners in the field that that it is best to avoid the use of mtDNA screening method on limited or irreplaceable evidentiary material [36, 37] A presumptive DHPLC screen for mtDNA sequence concordance, however, could serve as a useful tool for investigators in special situations such as the investigation of property crimes. The limited budgets of many law enforcement agencies make it extremely difficult for investigators to justify the expense of mtDNA testing in the majority of criminal offences that do not involve crimes against persons – particularly when there is no assurance *a priori* that the test results will necessarily advance an investigation. A presumptive screen for sequence concordance between a suspect and an item of evidence, however, could provide sufficient justification to submit the sample for confirmatory analysis by direct sequencing. Such screening might also help to readily eliminate from consideration such non-probative samples as hairs consistent with a victim that were recovered from the victim's home or vehicle in the case of a burglary or auto theft. In short, this could help investigators to focus their efforts on the most probative samples and thereby maximize the efficient use of investigative resources.

Employing a presumptive test of mtDNA sequence concordance in the manner described above shifts the process of DNA extraction from the dedicated mtDNA sequencing laboratory to the local law enforcement laboratory. This necessitates that additional consideration is given to the handling of these samples. The presence of evidentiary material with large quantities of mtDNA (*e.g.*, blood, saliva and seminal fluids, etc.) in local laboratories can pose a significant risk of cross contamination. Accordingly, an mtDNA sequencing laboratory accepting a DNA extract for analysis would almost certainly require the submission of a co-extracted reagent blank control that could be tested to detect the presence of spurious mtDNA contamination. Similarly, the submission of amplified PCR products for direct sequencing would also need to be accompanied by the appropriate positive and negative PCR controls.

[Implications for further research]

In toto, the central objective of this research program was to develop and demonstrate the use of Denaturing High Performance Liquid Chromatography (DHPLC) as a rapid, accurate and cost effective method for fractionating mtDNA mixtures prior to direct DNA sequencing. Through a comparative analysis of the sequence electropherograns representing DNA from two or more chromatographic fractions it has been shown that the linkage phase (and thereby the specific mitotypes) of the individual components of a mixture can be determined. If adopted by forensic practitioners, this will increase the number of forensic samples for which definitive sequence analyses can be conducted. The current research program has been completed in accordance with DAB developmental

validations standards for sensitivity, reproducibility and accuracy. The future of this research, therefore, will depend on achieving three critical objectives. These are:

1) **Rigorous Interlaboratory Validation by Practitioners** of the reliability of DHPLC in combination with linkage phase analysis of direct DNA sequence data resolve mtDNA mixtures in a practioner (caseworking) laboratory. This will likely necessitate participation by respected practitioners with extensive experience in forensic mtDNA analysis. It is anticipated that NIJ resources will be needed to enable interested practitioner laboratories to initially lease the necessary instrumentation and to hire the additional personnel to dedicate to the next phase of evaluating this technology.

2) **Rigorously Evaluate by Practitioners** of the initial Standard Operating Procedures and statistically-supported interpretation guidelines for mtDNA mixture analysis that have been developed in the course of the current study. It is anticipated that this process will not only involve training on the DHPLC system software but also in the handling of samples at all stages of a DHPLC-facilitated work flow. Although some of the researchers involved in the development of DHPLC for forensic applications are no longer employed by the University of Denver, these skilled individuals have agreed to provide training on a continuing basis as needed by the practitioner community.

3) **Further Development of Essential User Friendly Software Tools** to aid practitioners in linkage phase analysis of DHPLC-fractionated DNA. In the course of the current research program a first generation software solution, FLiPARS (Fractional Linkage Phase Analysis Resource System) was developed to automate the process of linkage phase analysis. While FLiPARS is a fully-functional software application, it has with a great potential for growth and will certainly be improved upon through feedback from practitioners. Some features that could be improved include faster data processing; greater database integration to allow for the storage and retrieval of data from Microsoft Access databases or SQL database servers; and a “Smoother” user interface. This will necessitate a continuing conversation between the developer and forensic practitioner community along with resources to obtain the necessary level of programming expertise.

4) **Extend Validation to Additional Primer Sets** in recognition of the fact that forensic practitioners often employ a variety of primer pairs in addition to those with which DHPLC was validated. While the standard HV1A through HV2B primer sets yield amplification products that are approximately 270bp in length, Highly degraded samples frequently contain DNA molecules that are severely restricted in size (*e.g.*, <150bp). To facilitate the analysis of such highly degraded material, a variety “mini-primer” pairs have been developed that span each of the HV regions with an average amplicon size of 140 bp. Although these were not evaluated as part of the current research, it is anticipated that they will perform even more efficiently with DHPLC since heteroduplex formation tend to be more readily resolved the shorter the amplicon length.

[References Cited]

1. Lightowlers RN, Chinnery PF, Turnbull DM, Howell N. Mammalian mitochondrial genetics: heredity, heteroplasmy and disease. Trends in Genetics 1997;13(11):450-5.

2. Fourney R. Mitochondrial DNA and Forensic Analysis: A Primer for Law Enforcement. Journal of the Canadian Society of Forensic Science 1998;31(1):45-53.

3. Holland MM, Parsons TJ. Mitochondrial DNA Sequence Analysis-Validation and use for forensic casework. Forensic Science Review 1999;11(1):22-49.

4. Melton T, Dimick G, Higgans B, Lindstrom L, Nelson K. Forensic Mitochondrial DNA Analysis of 691 Casework Hairs. Journal of Forensic Sciences 2005;50(1):73-80.

5. Hanekamp JS, Thilly WG, Chaudhry MA. Screening for human mitochondrial DNA polymorphisms with denaturing gradient gel electrophoresis. Hum Genet 1996;98(2):243-5.

6. Barros F, Lareu MV, Salas A, Carracedo A. Rapid and enhanced detection of mitochondrial DNA variation using single-strand conformation analysis of superposed restriction enzyme fragments from polymerase chain reaction-amplified products. Electrophoresis 1997;18(1):52-4.

7. Alonso A, Martin P, Albarran C, Garcia O, Sancho M. Rapid detection of sequence polymorphisms in the human mitochondrial DNA control region by polymerase chain reaction and single-strand conformation analysis in mutation detection enhancement gels. Electrophoresis 1996;17(8):1299-301.

8. Steighner RJ, Tully LA, Karjala JD, Coble MD, Holland MM. Comparative identity and homogeneity testing of the mtDNA HV1 region using denaturing gradient gel electrophoresis. Journal of Forensic Sciences 1999;44(6):1186-98.

9. Huber CG, Oefner PJ, Bonn GK. Rapid analysis of biopolymers on modified non-porous polystyrene-divenylbenzene particles. Chromatographia 1993;37:653-658.

10. Hou Y, Zhang S. Temperature-modulated high-performance liquid chromatography for detecting variation in human genome. Chung-Hua i Hsueh i Chuan Hsueh Tsa Chih 2000;17(3):145-8.

11. O'Donovan MC, Oefner PJ, Roberts SC, Austin J, Hoogendoorn B, Guy C, et al. Blind analysis of denaturing high-performance liquid chromatography as a tool for mutation detection. Genomics 1998;52(1):44-9.

12. Kuklin A, Munson K, Haefele R, Gjerde D, Haefele R, Taylor R. Detection of single-nucleotide polymorphisms with the WAVE™ DNA Fragment Analysis System. Genetic Testing 1997;1:201-206.

13. Underhill PA, Jin L, Lin AA, Mehdi SQ, Jenkins T, Vollrath D, et al. Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography. Genome Res 1997;7(10):996-1005.

14. Dobson-Stone C, Cox RD, Lonie L, Southam L, Fraser M, Wise C, et al. Comparison of fluorescent single-strand conformation polymorphism analysis and denaturing high-performance liquid chromatography for detection of EXT1 and EXT2 mutations in hereditary multiple exostoses. European Journal of Human Genetics 2000;8(1):24-32.

15. Gross E, Arnold N, Goette J, Schwarz-Boeger U, Kiechle M. A comparison of BRCA1 mutation analysis by direct sequencing, SSCP and DHPLC. Human Genetics 1999;105(1-2):72-8.

16. Liu W, Smith DI, Rechtzigel KJ, Thibodeau SN, James CD. Denaturing high performance liquid chromatography (DHPLC) used in the detection of germline and somatic mutations. Nucleic Acids Research 1998;26(6):1396-400.

17. Jones AC, Austin J, Hansen N, Hoogendoorn B, Oefner PJ, Cheadle JP, et al. Optimal temperature selection for mutation detection by denaturing HPLC and comparison to single-stranded conformation polymorphism and heteroduplex analysis. Clinical Chemistry 1999;45(8 Pt 1):1133-40.

18. Buyse IM, Fang P, Hoon KT, Amir RE, Zoghbi HY, Roa BB. Diagnostic testing for Rett syndrome by DHPLC and direct sequencing analysis of the MECP2 gene: identification of several novel mutations and polymorphisms. American Journal of Human Genetics 2000;67(6):1428-36.

19. Hecker KH, Turpie B, Kuklin A. Optimization of cloning efficacy by pre-cloning DNA fragment analysis. Biotechniques 1999;26(2):216-8.

20. Oefner PJ, Huber CG, Umlauft F, Berti GN, Stimpfl E, Bonn GK. High-resolution liquid chromatography of fluorescent dye-labeled nucleic acids. Analytical Biochemistry 1994;223(1):39-46.

21. Budowle B, Wilson MR, DiZinno JA, Stauffer C, Fasano MA, Holland MM, et al. Mitochondrial DNA regions HVI and HVII population data. Forensic Science International 1999;103(1):23-35.

22. Ivanov PL, Wadhams MJ, Roby RK, Holland MM, Weedn VW, Parsons TJ. Mitochondrial DNA sequence heteroplasmy in the Grand Duke of Russia Georgij Romanov establishes the authenticity of the remains of Tsar Nicholas II. Nature Genetics 1996;12(4):417-20.

23. Al-Soud WA, Rådström P. Purification and characterization of PCR-inhibitory components in blood cells. J Clin Microbiol 2001;39(2):485-93.

24. Eckhart L, Bach J, Ban J, Tschachler E. Melanin binds reversibly to thermostable DNA polymerase and inhibits its activity. Biochem Biophys Res Commun 2000;271(3):726-30.

25. Scholz M, Giddings I, Pusch CM. A polymerase chain reaction inhibitor of ancient hard and soft tissue DNA extracts is determined as human collagen type I. Anal Biochem 1998;259(2):283-6.

26. Watson RJ, Blackwell B. Purification and characterization of a common soil component which inhibits the polymerase chain reaction. Can J Microbiol 2000;46(7):633-42.

27. Cline J, Braman JC, Hogrefe HH. PCR fidelity of pfu DNA polymerase and other thermostable DNA polymerases. Nucleic Acids Res 1996;24(18):3546-51.

28. Navigator Software Manual. Omaha, NE: Transgenomic, Inc., 2003.

29. Cooksey RC, Morlock GP, Holloway BP, Limor J, Hepburn M. Temperature-mediated heteroduplex analysis performed by using denaturing high-performance liquid chromatography to identify sequence polymorphisms in Mycobacterium tuberculosis complex organisms. J Clin Microbiol 2002;40(5):1610-6.

30. Fixman M, Freire JJ. Theory of DNA melting curves. Biopolymers 1977;16(12):2693-704.

31. Anderson S, Bankier AT, Barrell BG, de Bruijn MH, Coulson AR, Drouin J, et al. Sequence and organization of the human mitochondrial genome. Nature 1981;290(5806):457-65.

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

33. Danielson PB, Kristinsson R, Shelton RJ, Laberge GS. Separating human DNA mixtures using denaturing high-performance liquid chromatography. Expert Rev Mol Diagn 2005;5(1):53-63.

34. Danielson PB, Sun HY, Melton T, Kristinsson R. Resolving mtDNA mixtures by denaturing high-performance liquid chromatography and linkage phase determination. Forensic Sci Int Gen 2007;1(2):148-53.

35. Andreasson H, Nilsson M, Budowle B, Frisk S, Allen M. Quantification of mtDNA mixtures in forensic evidence material using pyrosequencing. Int J Legal Med 2006;120(6):383-90.

36. Melton T, Holland CA, Nelson K. Commentary on: Divne A-M, Nilsson M, Calloway C, Reynolds R, Erlich H, Allen M. Forensic casework analysis using the HVI/HVII mtDNA linear array assay. J Forensic Sci 2006;51(4):935-6; author reply 937-8.

37. Divne AM, Nilsson M, Calloway C, Reynolds R, Erlich H, Allen M. Forensic casework analysis using the HVI/HVII mtDNA linear array assay. J Forensic Sci 2005;50(3):548-54.

[VI. Dissemination of Research Findings]

An important part of the forensic validation and courtroom acceptability of DHPLC and linkage phase analysis will be: the publication of research results in professional forensic and other science journals; formal presentations of the approach at appropriate forensic and other scientific conferences and; an ongoing discourse with leading practitioners in the field of mtDNA analysis to gain their support and guidance for the implementation of the technology. To this end we have done the following:

Manuscripts:

Richard Kristinsson and **Phillip B. Danielson** (2008) Characterization of Peak Height Reproducibility in Human Mitochondrial DNA Sequence Electropherograms. Electrophopresis (manuscript under review)

Danielson, P.B. and Kristinsson, R. (2008) Detection of Sequence Polymorphisms in the HV1 and HV2 Regions of Human Mitochondrial DNA by Denaturing High-Performance Liquid Chromatography. Forensic Science International (revised manuscript under review)

Westring, CG, Kristinsson, R., Gilbert, DM, and **Danielson, P.B.** (2007) Reduced Volume Protocol for Optimized Quantification of Genomic DNA using the Quantifiler™ Human DNA kit. Journal of Forensic Sciences 52(5):1035-43

Danielson, P.B., Sun, H-Y, Melton, T., and Kristinsson, R. (2007) Resolving mtDNA Mixtures by Denaturing High-Performance Liquid Chromatography and Linkage Phase Determination. FSI Genetics 1:148-153. (INVITED MANUSCRIPT)

Danielson, P.B. (2005) Mitochondrial DNA Analysis by Denaturing High-Performance Liquid Chromatography for the Characterization and Separation of Mixtures in Forensic Samples. Peer reviewed final research report for electronic release under NIJ cover.

Danielson, P.B., Kristinsson, R., Shelton, R.J. and LaBerge, G.S. (2005) Separating Human DNA Mixtures using Denaturing High-Performance Liquid Chromatography (DHPLC). Expert Review of Molecular Diagnostics 5(1):53-63. INVITED REVIEW ARTICLE

LaBerge, G.S., Shelton, R.J. and **Danielson, P.B.** (2003) Forensic Utility of Mitochondrial DNA Analysis based on Denaturing High-Performance Liquid Chromatography. Croatian Medical Journal 44(3):281-288.

Danielson, P.B., LaBerge, G.S. and Shelton, R.J. (2003) Clinical Applications of Denaturing High-Performance Liquid Based-based Genotyping. Croatian Medical Journal 44(4):447-454.

Manuscripts (final draft stage):

Richard Kristinsson, Terry Melton and Phillip B. Danielson. Simultaneous Purification and Quantification of PCR Yield by High-Performance Liquid Chromatography on DNASep® Columns. (2008 target submission)

Richard Kristinsson, Hong-Yu Sun, Terry Melton and Phillip B. Danielson. Quantitative Determination of Mixed Nucleotide Populations in Mitochondrial DNA Sequencing. (2008 target submission)

Richard Kristinsson, Hong-Yu Sun, Terry Melton and Phillip B. DanielsonLinkage Phase Analysis of Mitochondrial DNA Mixtures: Determination of Individual Contributors. (2008 target submission)

Professional Research Conference Presentations on mtDNA analysis by DHPLC:

2008 Advancing Forensic DNA Analysis of Challenging Evidentiary Material: Resolving mtDNA Mixtures by DHPLC. Exempla St. Joseph's Hospital, Denver, CO

2007 Advancing Forensic DNA Analysis of Challenging Evidentiary Material: Resolving mtDNA Mixtures by DHPLC. University of Northern Colorado, Greeley, CO

2006 DNA Mixture Fractionation and Analysis by Denaturing High-Performance Liquid Chromatography. III International Symposium on Biochemistry and Molecular Biology; Federación Latinoamericana de Asociaciones Químicas. (Treasury Department License CT-8218) Havana, Cuba

2006 mtDNA Mixture Fractionation and Analysis by Denaturing High-Performance Liquid Chromatography. 85. Jahrestagung der Deutschen Gesellschaft für Rechtsmedizin und DNA in Forensics 2006. Innsbruck, Austria.

2006 National Institute of Justice, "Forensic mtDNA Analysis and Mixture Separation by Denaturing High-Performance Liquid Chromatography", 7th Annual DNA Grantees' Meeting. Arlington, VA.

2005 National Institute of Justice, "mtDNA Mixture Fractionation and Analysis by Denaturing High-Performance Liquid Chromatography", Forensic DNA Research and Development Program Review Conference. Washington, DC.

2005 Louisiana State University Health Sciences Center "Rapid Mitochondrial DNA Analysis for Evidence Screening and Separation of Forensic Mixtures". Keynote Speaker for the Annual HSC Graduate Student Research Conference, Shreveport, LA.

2005 National Institute of Justice, "Forensic mtDNA Analysis and Mixture Separation by Denaturing High-Performance Liquid Chromatography", 6th Annual DNA Grantees' Meeting. Washington, DC.

2004 National Institute of Justice, "mtDNA Analysis by DHPLC for the characterization and separation of mixtures in forensic samples", Forensic DNA Research and Development Program Review Conference. Washington, DC.

2004 Cambridge Healthtech Institute, "Human mtDNA Mixture Separation by Denaturing High-Performance Liquid Chromatography" DNA Forensics: Enabling Investigative Examination, 6th Biannual Meeting. McLean, VA. (Session Chair)

2004 National Institute of Justice, "Forensic mtDNA Analysis and Mixture Separation by Denaturing High-Performance Liquid Chromatography", 5th Annual DNA Grantees' Meeting. Washington, DC.

2004 Poster Presentation. 15th International Symposium on Human Identification. Danielson, P.B., Kristinsson, R and LaBerge G.S. "Forensic mtDNA Analysis and Mixture Separation by Denaturing High-Performance Liquid Chromatography". Phoenix, Arizona.

Meetings with Forensic Practitioners, Researchers and other Law Enforcement Professionals:

2008 DNA 101 and the Future of Forensic DNA. Colorado Association of Criminal Justice Educators, Denver, CO.

2007 Mitotyping Technologies Incorporated, "Forensic Utility of Denaturing High-Performance Liquid Chromatography to Analyses of Human Mitochondrial DNA", State College, PA.

2006 University of Connecticut, Center for Applied Genetics and Technology, Department of Molecular and Cell Biology, "Forensic Analysis of mtDNA: Fractionating Mixtures by Denaturing High-Performance Liquid Chromatography", Storrs, CT.

2006 Forensic Applications of mtDNA Analysis and Denaturing High-Performance Liquid Chromatography. Shenyang Regional Police Agency, Shengyang, People's Republic of China.

2006 Mitotyping Technologies Incorporated, "Forensic Utility of Denaturing High-Performance Liquid Chromatography to Analyses of Human Mitochondrial DNA", State College, PA.

2006 Forensic Applications of Denaturing High-Performance Liquid Chromatography to mtDNA Analysis. National Institute of Scientific Investigation. Seoul, Republic of Korea.

2006 Forensic Applications of mtDNA Analysis and Denaturing High-Performance Liquid Chromatography. China Central Police Agency, Beijing, People's Republic of China.

2006 Forensic Applications of mtDNA Analysis and Denaturing High-Performance Liquid Chromatography. Shanghai Regional Police Agency, Shanghai, People's Republic of China.

2006 Transgenomic Inc. "mtDNA Mixture Fractionation and Analysis by Denaturing High-Performance Liquid Chromatography", Omaha, NE.

2005 Mitotyping Technologies Incorporated, "mtDNA Mixture Fractionation and Analysis by Denaturing High-Performance Liquid Chromatography: Progress Report", State College, PA.

2004 National Law Enforcement and Corrections Technology Center, “Rapid Mitochondrial DNA Analysis: Evidence Screening and Separation of Forensic Mixtures”, Regional Advisory Council Meeting, Whitefish, MT.

2004 National Center for Missing and Exploited Children, “Rapid Mitochondrial DNA Analysis: Evidence Screening and Separation of Forensic Mixtures”, Alexandria, VA.

2004 Royal Canadian Mounted Police and Transgenomic Corporation, “Forensic Utility of Denaturing High-Performance Liquid Chromatography to Analyses of Human Mitochondrial DNA”, Ottawa, Canada.

[Appendix I]

Draft Standard Operating Procedures for DHPLC Analysis of mtDNA

(Excerpted from the SOP Manual for the Danielson Laboratory)

Introduction

Mitochondrial DNA (mtDNA) sequencing can provide crucial information to forensic investigators when the quantity and quality of DNA would otherwise be limiting. Unlike short tandem repeats, mtDNA amplicons are identical or nearly identical in length and thus are characterized on the basis of nucleotide sequence. This makes it impossible to determine the mitotypes of individual contributors in a DNA mixture by direct sequencing. While alternate methods (*e.g.*, subcloning) could separate the contributors to a mixture, such approaches are time consuming; some subclones will contain artifactual sequence variants due to nucleotide misincorporation by the amplification; and contamination becomes a major concern as sample manipulation increases.

The technique enables sequence-specific separation of natural (heteroplasmic) or situational (multi-contributor) DNA mixtures prior to sequencing. This is achieved without secondary amplification or excessive sample manipulation.

Control Region Overview

DHPLC analysis of mtDNA requires amplification of four regions within the control region (D-loop) of human mtDNA (Figure 7.1). This is achieved using forensically validated mtDNA primers and amplification procedures. The D-loop is divided into two main fragments (HV1 and HV2). For general forensic sequencing and DHPLC analysis, each fragment is divided into two smaller fragments (HV1A, HV1B, HV2A and HV2B).

DHPLC analysis follow a validated flowpath for efficient determination of sample integrity and amplification efficiency (PCR check) as well as determination of whether the sample is of a pure source (Denatured control) (Figure 7.2). If a sample is determined to be a mixture, separation via DHPLC fractionation and Linkage phase analysis can be achieved.

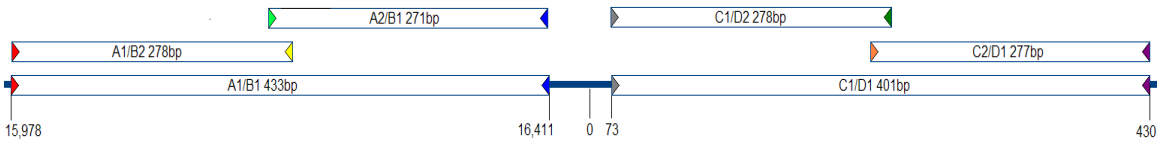


Figure 7.1 Human mtDNA Control region (D-loop) fragments used in forensics

Flowchart

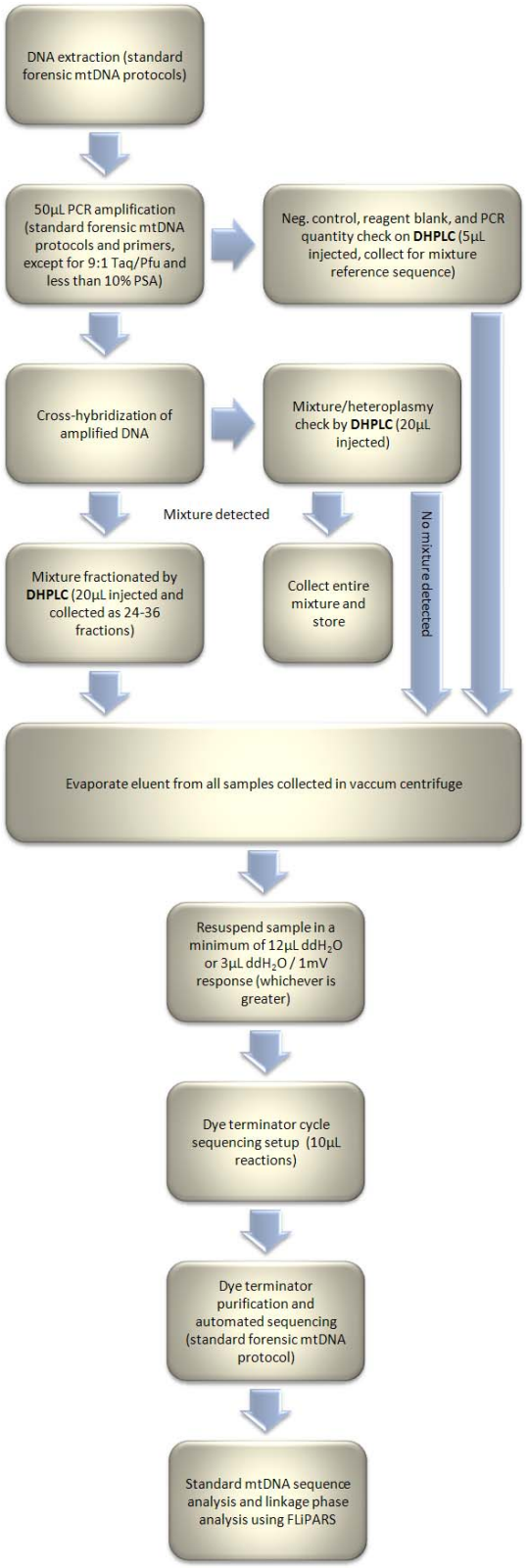


Figure 7.2 Flowchart of DHPLC fragment characterization.

DNA Extraction

DNA for DHPLC analysis may be extracted from fresh or frozen whole blood, peripheral blood lymphocytes, blood stains, sperm cells, hair, tissue, bone, and other types of biological samples. Each type of specimen may require different extraction procedures, which are outlined in Chapter 3 (DNA Extraction).
Regardless of the method used for DNA extraction, all samples must be handled carefully to prevent sample-to-sample contamination or contamination by extraneous DNA. DNA handling procedures outlined earlier in this manual are also recommended for DHPLC analysis.

DNA Quantitation

DNA quantitation is optimal to optimize each PCR reaction performed, however not necessary due to the nature of the analyses. All genomic DNA quantitation should be performed using the Quantifiler DNA quantification kit for accurate measurement of amplifiable genomic DNA, including inhibition and degradation effects. Mitochondrial DNA quantitation should be performed using mitochondrial specific primers and probes. DNA Quantitation is outlined in Chapter 4 (DNA Quantitation).

DHPLC PCR Amplification

1. DHPLC PCR amplification 50µL reaction volume:
   - Prepare the DHPLC PCR master mix by adding the following volumes of reagents to a 1.5mL or 0.6ml microcentrifuge tube:
     - Number of samples to be amplified X 5.0µL of 10X AmpliTaq GOLD Buffer
     - Number of samples to be amplified X 5.0µL of 10X dNTP's
     - Number of samples to be amplified X 4.0µL of 25mM $MgCl_2$
     - Number of samples to be amplified X 5.0µL of 10X Forward Primer (see Appendix II for appropriate forward mitochondrial primers)
     - Number of samples to be amplified X 5.0µL of 10X Reverse Primer (see Appendix II for appropriate reverse mitochondrial primers)
     - Number of samples to be amplified X 0.5µL of 9:1 AmpliTaq Gold DNA polymerase/Pfu Polymerase
     - Number of samples to be amplified X 20.5µL of milli-Q 18.2µOHM water
2. When working with reduced volumes the proportions of the reactants are maintained. Mix thoroughly by vortexing at medium speed for 5 seconds.
3. Spin the tube briefly in a microcentrifuge to remove any liquid from the cap.
4. Dispense 45µL of master mix to each 0.2mL microcentrifuge tube.
5. Add 5µL of DNA sample to be amplified to each tube.
6. Add the positive control DNA and negative amplification control DNA to the appropriately labeled tubes.
7. Briefly vortex and centrifuge the reaction tubes.

8. Place the tubes in the ABI 9600 thermalcycler and start the appropriate cycling program. The cycling parameters are as follows:
    - Initial incubation to activate AmpliTaq GOLD 10 minutes at 94˚C
    - Cycling 32 times:
        i. Denaturation 20 seconds at 94˚C
        ii. Annealing 30 seconds at 60˚C
        iii. Extension 45 seconds at 72˚C
    - Final Extension 10 minutes at 72˚C
    - Hold 4˚C forever
9. Following amplification, remove the tubes from the instrument block and store the amplified DNA products. For short periods store at 2-6˚C. For longer periods store the tubes at –15 to –25˚C.

PCR Amplification Check by DHPLC

NOTE: For specific details concerning the WAVE DHPLC sample setup refer to the Navigator Software manual.
DHPLC Buffers are purchased at 1X concentration from Transgenomic Inc., Omaha, Nebraska.
Buffers are stored at 4˚C after arrival to the laboratory and need to be equilibrated to room temperature prior to loading on the instrument.

Transfer 7µL of amplified sample to a sterile 0.2mL centrifuge tube and locate on the sample plate of the DHPLC autosampler. The PCR check will be obtained using the DS Single Fragment Application type (0.9mL/min, 50˚C) for the appropriate size fragment (A1B2: 278bp; B1A2: 271bp; C1D2: 278bp; D1C2: 277bp) using 5µL of sample. This sample should be collected into 12 wells on a 96-well plate labeled with the sample ID using the DHPLC fragment collector (see figure 7.3 for sample wells).

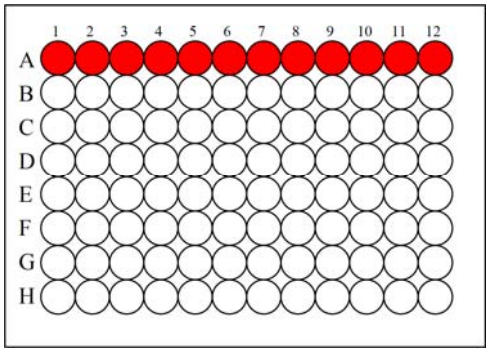


Figure 7.3 PCR Check sample wells after sample collection.

Results Interpretation

In the absence of length heteroplasmy or a mixture of two length products, pure mtDNA from individuals runs as a single peak using the methods designed for the PCR check. In the presence of length heteroplasmy or a mixture of two length products, the chromatographic profile will be represented as a shoulder or a significant broadening of the resulting peak.

PCR Product Concentration Determination

Efficiency of PCR and amount of DNA can be determined from the peak area of the PCR Check peak. Amplified DNA can be estimated by y=1230.9x + 712.99, where x is the DHPLC analysis estimated peak area and y is the DNA concentration. PCR Check DNA concentration can be determined using: PCR Check DNA concentration calc

Proceed to section 7.10 for further handling and downstream analysis.

Cross-Hybridization of Amplified DNA

PCR products are cross-hybridized by heating to 95°C for 4 min and cooling over a period of 45 minutes (1.5°C decrease/min) on the MJ Research, Inc PTC-100™ Thermocycler.

Cross-Hybridized Mixture Check by DHPLC

Transfer 7µL of cross-hybridized sample to a sterile 0.2mL centrifuge tube and locate on the sample plate of the DHPLC autosampler. The Mixture / Heteroplasmy check will be obtained using the Mutation Detection Application type (0.9mL/min) and the appropriate standard fragment analysis Method (HV1A, HV1B, HV2A or HV2B) developed by the laboratory. This sample should be collected using the DHPLC fragment collector and labeled: Mixture Check (see figure 7.4 for sample wells)

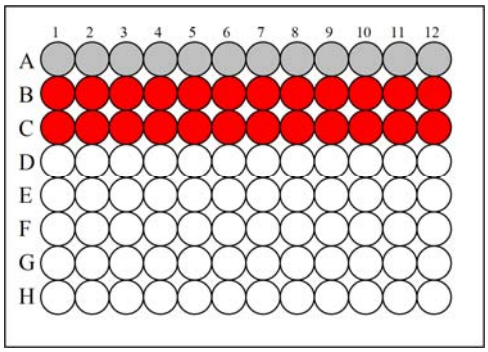


Figure 7.4 Mixture / Heteroplasmy Check sample wells after sample collection.

Results Interpretation

In the absence of heteroplasmy or mixture, pure mtDNA from individuals runs as a single peak using the methods designed for the Mixture / Heteroplasmy check. In the presence of heteroplasmy, the chromatographic profile will give more than one peak, usually represented as a shoulder or a significant broadening of the main peak. Mixed samples of DNA (*i.e.* from two sources) produce characteristic patterns of up to four separate peaks that are easily distinguishable from pure DNA from a single individual. The pattern of peaks (graphically and numerically) seen in the mixed sample is unique to the specific DNA molecules that comprise the mixture.

If no mixture is detected proceed to section 7.10 for further sample handling and downstream analysis.

Mixture fractionation by DHPLC

Load the cross-hybridized sample on the DHPLC autosampler. The Mixture fractionation will be obtained using the Mutation Detection Application type (0.9mL/min) and the appropriate standard fragment analysis Method (HV1A, HV1B, HV2A or HV2B) developed by the laboratory. This sample should be fractionated using the DHPLC fragment collector and labeled: Mixture Separation. Using the Mixture / Heteroplasmy chromatographic result create a fractionation method using the Fragment Collection generation tab of the Navigator software (see figure 7.5 for fraction wells).

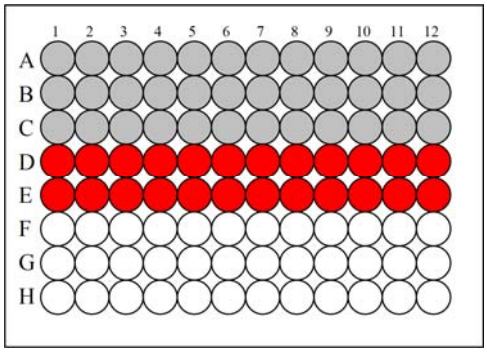


Figure 7.5 Mixture fractionation sample wells after sample fractionation.

Determination of Fractions Selected for Sequencing

- Resolution of homoduplex peaks:
  - When homoduplicies resolve into two distinct peaks, comparison of the top of the tallest peak to the edge of the secondary peak that is the furthest away from the tallest peak yields the greatest coordinated shifts in peak height fluorescence. This can be done by comparing the leading edge of the first peak to the top of the second peak, if the second peak is the tallest. Or compare the lagging edge of the second peak to the top of the first peak if the first peak is the tallest.

- No resolution of homoduplex peaks:
  - When homoduplices are not resolve into separate peaks, comparisons between the top or lagging edge of a heteroduplex peak to the top or lagging edge of the homoduplex peak, generally yields the greatest coordinated shift in peak height fluorescence.

- Fractions that yield limited shift in peak height fluorescence:
  - Comparison of fractions subsequent to each other yields limited peak height fluorescence shift due to minimal change in contributor ratios between those fractions.

Sample Evaporation

Each well (50µL) that contains more than 0.5mV of response above baseline should be transferred from the 96-well plate into a 0.6 sterile microcentrifuge tube. The samples should then be loaded into the Labconco Centri-Vac system for solvent evaporation.

Sample Resuspension

Following solvent evaporation, each sample should be resuspended in at least 12µL of milli-Q 18.2µOHM water or 3µL milli-Q 18.2µOHM water / 5ng of DNA (1mV). Proceed to section 8.8 for downstream sequencing analysis.

Mitochondrial DNA Sequencing

Introduction

Mitochondrial sequencing has been proven to be a useful tool in the field of forensics. It is often employed as an alternative when conventional STR fragment analysis fail. The mitochondrial DNA is extremely small and exists in very high copy numbers within cells, making it an ideal candidate for severely degraded forensic samples.

DNA Extraction

DNA for mitochondrial DNA sequencing analysis may be extracted from fresh or frozen whole blood, peripheral blood lymphocytes, blood stains, sperm cells, hair, tissue, bone, and other types of biological samples. Each type of specimen may require different extraction procedures, which are outlined in Chapter 3 (DNA Extraction).
Regardless of the method used for DNA extraction, all samples must be handled carefully to prevent sample-to-sample contamination or contamination by extraneous DNA. DNA handling procedures outlined earlier in this manual are also recommended for sequencing analysis.

DNA Quantitation

DNA quantitation is necessary to optimize each PCR reaction performed. Mitochondrial DNA quantitation should be performed using mitochondrial specific primers and probes. DNA Quantitation is outlined in Chapter 4 (DNA Quantitation).

PCR Amplification

1. PCR amplification 45µL reaction volume:
   - Prepare the PCR master mix by adding the following volumes of reagents to a 1.5mL or 0.6ml microcentrifuge tube:
     - Number of samples to be amplified X 4.5µL of 10X AmpliTaq GOLD Buffer
     - Number of samples to be amplified X 4.5µL of 10X dNTP's
     - Number of samples to be amplified X 3.6µL of 25mM $MgCl_2$
     - Number of samples to be amplified X 4.5µL of 10X Forward Primer (see Appendix II for appropriate forward mitochondrial primers)
     - Number of samples to be amplified X 4.5µL of 10X Reverse Primer (see Appendix II for appropriate reverse mitochondrial primers)
     - Number of samples to be amplified X 0.45µL of 9:1 AmpliTaq Gold DNA polymerase/Pfu Polymerase
     - Number of samples to be amplified X 18.45µL of milli-Q 18.2µOHM water
2. When working with reduced volumes the proportions of the reactants are maintained. Mix thoroughly by vortexing at medium speed for 5 seconds.
3. Spin the tube briefly in a microcentrifuge to remove any liquid from the cap.
4. Dispense 40.5µL of master mix to each 0.2mL microcentrifuge tube.

5. Add 4.5µL of DNA sample to be amplified to each tube.
6. Add the positive control DNA and negative amplification control DNA to the appropriately labeled tubes.
7. Briefly vortex and centrifuge the reaction tubes.
8. Place the tubes in the ABI 9600 thermalcycler and start the appropriate cycling program. The cycling parameters are as follows:
    - Initial incubation to activate AmpliTaq GOLD 10 minutes at 94˚C
    - Cycling 32 times:
        i. Denaturation 20 seconds at 94˚C
        ii. Annealing 30 seconds at 60˚C
        iii. Extension 45 seconds at 72˚C
    - Final Extension 10 minutes at 72˚C
    - Hold 4˚C forever
9. Following amplification, remove the tubes from the instrument block and store the amplified DNA products. For short periods store at 2-6˚C. For longer periods store the tubes at –15 to –25˚C.

DHPLC PCR Product Purification

NOTE: For specific details concerning the WAVE DHPLC sample setup refer to the Navigator Software manual.
DHPLC Buffers are purchased at 1X concentration from Transgenomic Inc., Omaha, Nebraska.
Buffers are stored at 4˚C after arrival to the laboratory and need to be equilibrated to room temperature prior to loading on the instrument.

Load the sample on the DHPLC autosampler. The PCR product purification reference and the product purification will be obtained using the DS Single Fragment Application type (0.9mL/min, 50˚C) for the appropriate size fragment (A1D1: 1021bp; A1B2: 278bp; B1A2: 271bp; C1D2: 278bp; D1C2: 277bp) using 20µL. Obtain a reference trace of the sample by injecting 20µL of the sample. Then use the Fragment Collection tab of the Navigator software to generate a fraction collection method for the product purification. This sample should be collected into 12 wells on a 96-well plate using the DHPLC fragment collector.

Sample Evaporation

Each well (50µL) that contains more than 1mV of response should be transferred from the 96-well plate into a 0.6 or 1.5mL sterile microcentrifuge tube. The samples should then be loaded into the Labconco Centri-Vac system for solvent evaporation.

Sample Resuspension

Following solvent evaporation, each sample should be resuspended in at least 12µL of milli-Q 18.2µOHM water or 3µL milli-Q 18.2µOHM water / 5ng of DNA (1mV).

Dye-Terminator Cycle Sequencing Labeling Amplification

PCR purified samples are the best candidates for efficient labeling by dye terminators due to the absence of inhibitors and non-specific DNA. Dye terminator sequencing can be performed using either ABI PRISM® BigDye™ Terminator v1.1 or v3.1 Ready Reaction Cycle Sequencing Kits.

1. Labeling 10µL reaction volume:
   - Prepare the labeling master mix by adding the following volumes of reagents to a 1.5mL or 0.6ml microcentrifuge tube:
     - Number of samples +1 to be labeled X 2µL of Terminator Ready Reaction Mix
     - Number of samples +1 to be labeled X 1µL of Sequencing Buffer
     - Number of samples +1 to be labeled X 1µL of 10X Sequencing Primer (see Appendix II for appropriate mitochondrial sequencing primers)
     - Number of samples +1 to be labeled X 4µL of milli-Q 18.2µOHM water (amount of water depends on amount of template being added).
2. Mix well and spin briefly.
3. Aliquot 8µL of labeling master mix to each of the reaction tubes.
4. Add 2µL of DNA (0.4ng/µL) template to each of the reaction tubes.
5. Mix well and spin briefly.
6. Place the tubes in the ABI 9600 thermalcycler and start the appropriate cycling program. The cycling parameters are as follows:
   - Cycling 25 times:
     i. Denaturation 10 seconds at 96˚C
     ii. Annealing 5 seconds at 50˚C
     iii. Extension 4 minutes at 60˚C
   - Hold 4˚C forever

Extension product purification

Ethanol Precipitation

NOTE: Use only a refrigerated centrifuge at 4˚C when performing extension product purifications. This will allow for significantly greater recovery of labeled products.

Prepare a fresh batch of stop solution. Stop solution consists of 1µL 3M Sodium Acetate, pH 5.2; 1µL 0.5M Na_2EDTA, pH 8.0; 1µL 20mg/mL glycogen per reaction.

1. Spin the tube briefly in a microcentrifuge to remove any liquid from the cap.
2. Dispense 3µL of stop solution to each 0.2mL microcentrifuge tube.
3. Vortex briefly to allow sufficient mixing of the stop solution.

4. Transfer the entire contents of the 0.2mL tube to a labeled 0.6mL microcentrifuge tube.
5. Add 30µL of ice-cold 95% ethanol to each of the 0.6mL tubes.
6. Vortex for 5 seconds.
7. Spin at 4˚C for 15min.
8. Carefully remove the 0.6mL tube from the centrifuge, try to not disturb the pellet. Remove supernatant using a 200µL gel tip.
9. Add 100µL of 75% ice-cold ethanol. Spin at 4˚C for 2 min.
10. Remove supernatant.
11. Repeat steps 9 and 10.
12. Evaporate the remaining liquid using the Labconco Centri-Vac system for 15min.
13. Resuspend the purified product in 25µL of Hi-Di Formamide.

Column purification

Column purification of dye-terminator extension products is performed using the Performa® DTR V3 96-Well Short Plates or Performa® DTR Gel Filtration Cartridges (EdgeBio Systems). The columns are pre-packed with a fully hydrated matrix to afford optimal handling and performance characteristics. To minimize the potential for interference with sequencing applications, no preservatives, salts or buffers are used in the preparation of these columns.

Performa DTR V3 96-Well Short Plates

Both ends of the Performa DTR V3 96-Well Short Plates are sealed to prevent drying. The sample can be spun directly into the ABI PRISM® MicroAmp® Optical 96 Well Reaction Plate or equivalent (96-Well Semi-Skirted Capillary Plates) thereby saving a transfer step.

1. Remove the bottom adhesive tapes from a V3 96-well Short Plate.
    - Ensure that the plate remains horizontal to avoid losing any gel.
2. Stack the V3 96-Well Short Plate on top of a 96-well waste plate.
3. Remove the top adhesive tape from the V3 96-well Short Plate.
4. Place assembly on a rubber cushion in the centrifuge carrier.
5. Centrifuge for 3 minutes at 850 *x g*. Discard eluate.
6. Transfer the reaction samples to the center of each well in the V3 96-Well Short Plate. Pipet slowly. Do not touch the sides of the wells. Cover with lid.
7. Stack the V3 96-Well Short Plate on top of a 96-well Semi-Skirted Capillary Plate. Place the assembly on a rubber cushion in the centrifuge carrier designed to hold deep-well 96-well plates.
8. Centrifuge for 5 minutes at 850 *x g*. Retain eluate.

- The eluate contains purified sample ready for loading on sequencers.

Performa® DTR Gel Filtration Cartridges

1. Centrifuge the gel filtration cartridge assembly for 3 minutes at 850 *x g* with the cap closed.
2. Transfer the cartridge to the provided 1.5-ml microcentrifuge tube and add the sample to the packed column. Be sure the fluid runs into the gel. Discard 1.5ml tube containing original eluate.
3. Close the cap and centrifuge for 3 minutes at 850 *x g*. Retain eluate.
   - The eluate contains purified sample ready for loading on sequencers.
4. Transfer eluate to a 96-well Semi-Skirted Capillary Plate or 0.2ml PCR strip tubes for sequencing.

Sequencing Product Resolution on the ABI 310 Genetic Analyzer

NOTE: For specific details concerning the ABI 310 sequencing sample setup refer to the ABI PRISM® BigDye™ Terminator v1.1 Ready Reaction Cycle Sequencing Kit Protocol Manual.

1. Vortex the tubes to mix then spin briefly in a microcentrifuge.
2. Seal the tube strips with a septum.
3. Denature each sample by placing in a thermalcycler or appropriate dry bath at 95˚C for 2 minutes.
4. Chill tubes (snap cool) for at least 3 minutes in a 96 well ice plate.
5. Samples can now be placed into the 310 instrument for electrophoresis. (See Chapter 5 for detailed description of setting up the 310 for capillary electrophoresis). The electrophoresis will be conducted on a 47cm 50µm internal i.d. capillary using POP6 separation polymer at 50˚C. The software module Seq POP6 Rapid (1mL) E under ABI 310-collection software is used to collect raw data from the samples.

NOTE: In order for sequencing analysis to accurately determine sequence composition it is essential that peaks be no higher than 8000 RFU.

Dye Terminator sample analysis

Following sample resolution samples are analyzed for base calling and sequence quality using ABI Sequencing Analysis Software. Sample analyses are performed using the KB basecaller, KB310_POP6_BDTv1_36Rapid.mob Dye/Primer Set and flat profile display using the ABI Sequencing Analysis Software.

Analysis and Interpretation of Mitochondrial DNA Sequencing Results

Criteria for Acceptance of Data

Data should be discarded under the following conditions:

A. Insufficient data or no data – Collection software gives a "Run Log Error" and no data is collected.
B. Visual inspection of electropherogram – Readable sequence indeterminable, *i.e.* peaks are of scale or contains excessive background peaks.
C. Sequence information appears poor and does not align with the revised Cambridge reference sequence.
D. Sequence information appears good, but does not align with the revised Cambridge reference sequence.

Results Interpretation

Single source samples

Results from sequencing should be analyzed using Gene Codes Sequencher™ software. All mitochondrial sequences should be aligned to the Cambridge Reference sequence and a difference map constructed. All sequencing base calls and sequence quality should be reviewed by another DNA sequencing expert in the laboratory. For sequence ambiguities, consult Appendix III for base calling rules and interpretations.

Mitochondrial DNA mixtures

Samples determined to be mixtures based on DHPLC analysis should be sequenced and mutations corresponding to each contributor determined by linkage phase analysis. Following alignment of sequencing results mixed base locations can be determined.

Project Development Using Sequencher Plus

A. Sample sheet "Comments" field must contain the following information:
    a. Region, sequencing, primer, opposite amplification primer/… (*e.g.*: HV2, R381, F15/…; HV1, F15989, R16158/…)-Do not add spaces

Note: Create a separate contig for each sample/control [*e.g.*, CILHI Family Reference Specimen (CILHI FRS)] or batch of samples and controls (*e.g.*, CILHI case).

B. Double-click on the Sequencher Plus icon. A new project should appear into which sequences can be imported. If not, chose "File" then "New Project" for a new project or "Open Project" for an existing project

C. Choose "Windows" then "User Preferences"

    1. (Recommended) Display:

Chromatogram:	Chromatogram Height = Tall
Contig Chrom.:	Columns for Mult. Traces = 1
Format Ruler:	Break every 10 bases/Font Size = 12

2. General:
   Confidence: Low = 0-24 / High = 24-60

Note: Sequences that are assigned an overall score below the threshold will be placed into a refrigerator file, entitled "Low Quality Fragments", within the project. Sequences can be manually placed into the refrigerator by highlighting the sequence and the refrigerator file and choosing "Edit" then "Refrigerate". To remove the sequence from the refrigerator, double-click on the refrigerator then highlight that sequence, then click on the "Move Selected Items to "Project Window" button.

3. Close "User Preferences: window.

D. Choose "File: then "Import" then "Sequences" Or "Folder of Sequences". Double-click on sequences to be imported. Click on the "Done" button. Alternatively, highlight the sequences to be imported and drag them to the open project window. To remove a sequence, highlight the sequence then click on the trash can icon button. Click on the "Throw them away" button.

E. Highlight all sequences and choose "Sequence" then "Trim Ends".

If the "Comments" field on the sample sheet was entered incorrectly, an error message will be displayed and no trimming recommendations will be given. To rectify this problem:

1. Remove the affected sequences from the contig, highlight the lane of data to be edited. Choose "File" window from the menu. Under the file menu, chose "Get info". The comments can then be edited under the info screen window for each lane of data individually.

   Or

2. Remove the affected sequences from the contig, highlight the lane of data to be edited. Chose "Edit" window from the menu. Under the file menu, chose "Edit Comments". The comments can then be edited under the info screen window for each lane of data individually.

F. Accept or reject the recommended trimming by marking the appropriate boxes. The bases flanking the trim point can be observed by pressing "Show Bases" button.

G. Click on the "Trim Checked Items" button. Click on the "Trim" button. Close "AFDIL Trimming" window.

H. Import appropriate reference sequence(s) [filenames must end with ".caf" for the trimming mechanism to function properly].

I. If you are processing the entire control region you can set the circular genome size (16569) by clicking on the "Set Circular Genome Size" under the "Sequence Tab".

J. Click on the "Assembly Parameters" button under the "Contig" tab. Select "Clean Data", "Minimum Match Percentage" = 80, "Minimum Overlap" = 20. Click on the "OK" button. Highlight all sequences that need to be compared to one another (including reference sequence) ad click on the "Assemble Automatically" button. Click on the "OK" button. Sequences from different regions will be placed into separate contig files. To edit the name assigned to a contig file, click on the filename then type or choose "Contig" then "Rename Contig" then type then click on the "OK" button. Contig files may be dissolved by choosing "Contig", then "Dissolve Contig" and click on on the "Dissolve Contig" button.

   Note: Assembly Parameters may need to be adjusted for successful assembly.

   Note: Sequences may need to be added to the contig file one at a time for successful assembly. If assembly is not successful, they can be mindlessly joined. To do this, e sure that the individual sequences to be joined are appropriately aligned (forwards are forwards, reverses are reverses), choose "Contig", then "Assemble Contigs", then "Mindlessly Join". Choose "All Left" from pop-up window and then manually line each sequence up to the reference sequence.

   Note: Sequences may need to be manually trimmed for successful assembly. Double-click on sequence, then press "Show Chromatogram" button, then trim as needed (highlight bases to be trimmed and delete).

K. Double-click on contig file icon. The overview will provide you with an idea of the amount of confirmed sequence data obtained prior to editing. Click on the "Options" button. Select "Scale Diagram to Window Size"

and "Names of Fragments" only. Click on the "OK" button. (These options should stay consistent between analysis projects.)

L. To view bases, click on the "Bases" button. Low quality positions will be dark blue highlighted. High quality positions will be light blue highlighted. If highlighting does not appear, choose "View" then "Display Base Confidences".

Note: The order in which the sequences appear may be changed by clicking and dragging the sample names.

M. Choose "Contig" then "Consensus by Plurality". Highlight the reference sequence in the Contig. Choose "Sequence" then "Reference Sequence" to exclude the reference sequence from the consensus. If the base numbering of the sequences is incorrect, highlight any known positions in the consensus sequence. Choose "Sequence" then "Set Base Number" then "As Base Number". Type in the known base number. Click on the "OK" button.

N. To view all chromatograms simultaneously, highlight the bases of interest in the consensus sequence then press "Show Chromatograms" button. To view a portion of a single chromatogram, highlight the bases of interest in that sequence then press "Show Chromatograms" button. To view an entire chromatogram, double-click on the sequence then highlight the bases o interest in that sequence window then press "Show Chromatogram" button. The remainder of the chromatogram may be viewed by using the scroll bar.

Note: Editing may not be preformed in an individual sequence window or a single chromatogram window. Editing must occur in contig window or contig chromatogram window.

O. Edit each position in each sequence, as necessary. Editing bases should appear magenta, bold, and lower case. If edited bases do not appear this way, choose "View" then "Base Edits As" then "bOLD and cASE cHANGE".

Note: Bases edited by mistake may be returned to the original base call by choosing "Edit" then "Undo"; however, this function only works if performed immediately following the erroneous edit. Bases edited by mistake may also be returned to the original base call by highlighting the original base in the electropherogram window, selecting "Sequence", and then "Revert to Experimental Data"

P. To automatically export consensus sequence: Highlight reference sequence. Choose "Sequence" then "Compare to" then "Consensus". This screen then can be printed or exported to a text file. To export as a text file, click on the "Export As Text" button. Select appropriate directory, type export filename, and click on the "Save" button. Close comparison results window.

Q. Print overview of contig by pressing "Overview" button and choosing "File" then "Print" (landscape format).

R. Print summery of contig by pressing "Summary" button, then pressing "Ruler" button and adjusting left and right margins, then choosing "File" then "Print" (landscape format).

S. Choose "File" then "Close Project". Click on the "Yes" button.

T. If the consensus sequence has been exported, double-click on the exported text file (icon). Edit text. Print text by choosing "File" then "Print". Save file by choosing "File" then "Save".

Criteria for Accepting or Editing a Base Call

It is necessary either to accept or edit the base call at each position made by the analysis software program. The term "authentic" refers to a base that, in the educated opinion of the scientist, is the base present in the sample sequence. The printed electropherogram should be examined prior to finalizing any base calls.

General Rules:

A. A base call must not be edited without proper justification.

B. An "N" position is determined at the discretion of the scientist. Some background is acceptable.

C. Significant differences in relative peak intensity may be used as a criterion for editing a base call. However, substantial weight should be given to the base call made by the analysis software.

D. Discrete peaks are not necessary to justify keeping sequence data. However, data should be excluded where peaks become continuous (*i.e.*, no definition between shoulders).

E. There are certain patterns that can be expected when employing dye-terminator chemistry to generate DNA sequence information that may justify editing a base call.

F. Specific justifications that support both base acceptance and base editing are given below:

a. The authentic base is clearly apparent and is called by the analysis program

b. An additional base is called by the computer software in the authentic sequence (often due o peak broadening near the ends of sequence data). Delete the additional base.

c. The computer software omits an authentic base from the sequence, however, there is a discrete peak in the electropherogram to which it can be assigned. Insert the authentic base.

d. Polycytosine stretches [C stretch(es)] are often difficult to interpret. A possible cause may be the presence of a mixture of length variants in the mtDNA of an individual. A predominant length species is often apparent; however, the frequency of a particular length species cannot be determined accurately and may vary between maternal relatives. When there is mutation at nt 16189 in the C stretch of Hypervariable Region 1, the sequence reported represents the first 10 cytosines observed, beginning at nt 16184. If the sequence cannot be read past the C stretch, all data following should be deleted. The sequence reported for the C stretch in Hypervariable Region 2 represents the number of cytosines present in the ***predominant*** base sequence. When a predominant base sequence cannot be determined, the insertions that could not be confirmed are designated by an "N".

e. When a T to C polymorphism at nt 310 is observed in Hypervariable Region 2, there is an irresolvable number of cytosines in the polycytosine stretch from nt 303-315. Therefore, the sequence reported for Hypervariable Region 2 should represent the first 13 cytosines observed, beginning at nt 303. Irresolvable bases within the first 13 cytosines between nt 303-315 should remain an "N".

f. In the HV2 region, nt 515-524, there is a quintuple AC tandem repeat. Deletion of the fifth AC tandem repeat is not uncommon. The deletion of this repeat is to be searched and reported as 523 and 524 deletions. Tandem repeat insertions will be designated at position 524 (*i.e.*524.1A, 524.2C, 524.3A, 524.4C…).

g. Unincorporated dye with underlying readable bases: the authentic base is apparent in some cases, but obscured by excess unincorporated dye. Bases that cannot be confirmed should be designated by an "N" or "n".

h. Compression of bases: an authentic base is omitted from the sequence. Insert the authentic base.

Summary of the Sequencher Layout

A. Highlight all polymorphisms relative to the revised Cambridge Reference Sequence (Standard) on the summary layout. Highlight that same position in Anderson.

B. For each sample, document the range of sequence confirmed (*e.g.*,35-464), all polymorphisms (*e.g.*, 263 A-G or 263 G), and all positions that are unconfirmed ("N"). The reported sequence will be the most conservative of the two analyses or only the bases confirmed by both scientists.

C. Each page of the summary layout should contain the date, case number and handwritten initials of the scientist who created the project. The lead analyst must also initial each page.

Documentation of Electropherograms

A. All electropherograms for evidentiary samples should be labeled with the following information:

1. Approximate beginning and end of confirmed sequence used

2. Polymorphisms present in the sample, excluding HV2 C-stretch, where appropriate

3. Unconfirmed positions for that sample

B. In order to aid in review, all electropherograms for negative controls, reagent blanks, and/or substrate controls (where appropriate) should be labeled with the following information:

1. Approximate beginning and end of confirmed sequence used

2. Polymorphisms present

C. Neither polymorphic positions nor start and stop points in the positive amplification control electropherograms need to be labeled.

D. Neither polymorphic positions nor start and stop points in samples sequenced in a 96-well format need to be labeled; however, the lead analyst must review each printed electropherogram for heteroplasmy or mixtures.

E. Handwritten initials of the lead analyst should be present on all electropherograms. The date and case number should be electronically printed or handwritten on them.

Criteria for Reporting Data

For evidentiary specimens, data can be reported when:

A. Two extractions were performed ad at least one amplification reaction was successful from each extract. Therefore, the sequence data obtained is consistent and can be used to include and/or exclude references (*i.e.*, victim, suspect, maternal family member).

B. Two extractions were performed and at least two amplification reactions were successful from one of the extracts. Therefore, the sequence data obtained is consistent and can be used to include and/or exclude references (*i.e.*, victim, suspect, maternal family member).

C. One extraction was performed and at least two amplification reactions were successful from that extract. Therefore the sequence data obtained is consistent and can be use to include and/or exclude references (*i.e.*, victim, suspect, maternal family member).

D. For evidentiary specimens, data cannot be reported when one or two extractions were performed and only one amplification reaction was successful. Therefore, consistent sequence information could not be obtained and the sequence data cannot be reported.

E. For specimens that are known to have a polymorphism present in the binding site of the primer, interpretation will be evaluated on a case-bye-case basis. Several factors need to be taken into account, such as, but not limited to location within the primer binding site, number of extractions, difficulty of amplification, whether an alternate primer was used, etc.

Guidelines for Controls

Negative Controls

A. For manual sequencing of casework samples:

1. If one or both negative controls produce a detectable positive result on an agarose gel, and sufficient sample or control extract is remaining, the amplification should be repeated. The negative control(s) must be sequenced and analyzed to determine the source of contamination for quality control purposes.

2. If a negative control yields a positive result and a limited amount of extract or reagent blank exists for a sample, the sample data may be used with approval from a supervisor providing the negative control produces a sequence different from the sample.

3. If a negative control does not produce a detectable positive result on an agarose gel, but yields a readable sequence when processed, then the sample data may be used if the sequence obtained from the negative is different from the sample.

B. For high throughput sequencing of high quantity/quality DNA extracts in a 96-well format:

1. If a negative control yields readable sequence in one primer, the plate of samples should be reinjected and/or resequenced. If the negative control continues to produce readable sequence data, all samples should be reamplified, if possible.

2. If a negative control yields readable sequence in more than one primer, the amplification should not be used and the plate of samples should be reamplified, if possible.

3. If a negative control yields a positive result and a limited amount of extract or reagent blank exists for a sample, the sample data may be used with approval from a supervisor providing the negative control produces a sequence different from the sample.

Reagent Blanks

A. For manual sequencing of casework samples:

1. If the reagent blank gives a readable positive result:

A. If the sequence is the same for any sample extracted with that reagent blank, then that data cannot be used for that primer set using that extract. If it is suspected that the contaminant is an amplification contaminant and not an extraction contaminant, the amplification may be repeated twice using the same conditions (AmpliTaq Gold® DNA polymerase volume, template volume, and primer pair) with clean and/or different results before reporting. The amplification in which the reagent blank showed the contaminant that was consistent with the sample cannot be used for reporting.

Note: Careful attention must be given to the extract volume if it is to be repeated. The analyst should focus on the primer set that would give the most diagnostic information for the case and may choose not to repeat the amplification or report that region.

B. If the reagent blank gives usable data in one or more amplifications, but is not consistent with the sample data for any of these amplifications, the sample data can be reported.

B. For high throughput sequencing of high quantity/quality DNA extracts in a 96-well format:

    1. If a reagent blank yields readable sequence in one primer (and the negative controls are clean), the plate of samples should be resequenced.

        a. If the reagent blank continues to produce readable sequence data, where multiple amplifications are present for that reagent blank and all but one (1) are clean, the samples should be reamplified.

        b. If all amplifications for that reagent blank continue to produce readable sequence data, all samples should be reextracted, if possible.

    2. If a reagent blank yields readable sequence in more than one primer (and the negative controls are clean):

        a. And multiple amplifications are present where all but one (1) amplification are clean for that reagent blank, the samples should be reamplified, if possible. If the reagent blank continues to produce readable sequence data, the samples should be reextracted, if possible.

        b. And all amplifications for that reagent blank produce readable sequence data, all samples should be reextracted, if possible.

    3. If a reagent blank yields readable sequence in any primer and a negative control also produces readable sequence, the amplification should not be used and all samples should be reamplified.
    4. If a limited amount of extract or reagent blank exists for a sample and a reagent blank yields a positive result, the sample data may be used with approval from a supervisor providing the reagent blank produces a sequence different from the sample.

Substrate Controls

A. A substrate control is optional and at the discretion of the analyst with the exception of paraffin blocks. If a substrate control is provided as an additional sample, as in a proficiency test, the substrate control should be extracted along with the unknown sample of the same type.

B. If the substrate control gives a detectable positive result:

    a. If the sequence is the same for any sample extracted with that substrate control, then that data cannot be used for that primer set using that extract. If possible, the evidentiary sample should be re-extracted along with the substrate control to eliminate the possibility of cross-contamination.

b. If the substrate control gives usable data in one or more amplifications, but is not consistent with the sample data for any of these amplifications, the sample data can be reported.

Positive Controls

A. The positive control (HL-60 cell line or extracted DNA from blood drawn from RK) must produce sequence that is consistent with the known polymorphisms. Since the positive control is a known sample and amplification control, it is considered confirmed with only one amplification. Reinjecting and/or resequencing may be necessary. Two (2) independent analyses of the positive control data must be performed.

B. The known polymorphisms for RK are as follows:

263	A-G
315.1	C

C. The known polymorphisms for HL60 are as follows:

16069	C-T	73	A-G
16193	C-T	1560	C-T
16278	C-T	152	T-C
16362	T-C	263	A-G
		295	C-T
		315.1	C
		489	T-C

Match Criteria

Consistency

When two mtDNA sequences from separate sources are consistent with each other in the overlapping regions, *e.g.*, from two pieces of evidence or from evidence and a maternal family reference source, the two sources cannot be excluded as originating from the same person or from having a maternal relationship, respectively. Polycytosine region differences are ignored because mixtures or length variants are often difficult to interpret and can differ between maternal relatives.

Inconclusive

A. The resulting comparison will be considered inconclusive when:
   1. Two mtDNA sequences from separate sources differ by one polymorphism with no evidence of heteroplasmy, excluding position 16093 and the hypervariable region two (HV2) polycytosine region. Position 16093 and the HV2 polycytosine regions are excluded due to the highly heteroplasmic nature of these regions.

2. Two mtDNA sequences from separate sources differ by just position 16093 with no evidence for heteroplasmy.

B. In these cases, other reference sources and/or further testing in order to obtain additional sequence data may be helpful.

C. Closed population comparisons will be evaluated on a case by case basis.

Exclusion

The resulting comparison will be considered an exclusion when two mtDNA sequences from separate sources differ by two ore more polymorphisms excluding position 16093 and the HV2 polycytosine region. Position 16093 and the HV2 polycytosine region are excluded due to the highly heteroplasmic nature of these regions.

Linkage Phase Analysis

Linkage Phase analysis is the study of peak height comparisons over several fractions of DNA generated and separated by DHPLC. The analysis compares peak heights for mixed bases in a sequencing electropherograms across several sequenced fractions, allowing the analyst to determine the composition of an mtDNA mixture in a streamlined and automated fashion.

Mitochondrial DNA mixture separation can be performed without the application of Linkage Phase analysis. This, however, requires that a clean separation of the two components of the mixture is achieved (*i.e.* that the resulting sequence yields a secondary peak lower than 10% of the overall height of the major peak in the chromatogram). This is rarely achieved and even when achieved linkage phase analysis gives a more comprehensive view of the resulting analysis. The application of linkage phase analysis allows for the determination of the contributors to any two component mixture. The steps outlined in the following section demonstrate how linkage phase analysis is applied to any resulting sequencing data following analysis by the ABI sequencing analysis software.

Sequencing Analysis Data Extraction

Due to the nature of the ABI file system, there is no way currently to extract exact peak heights from the analyzed sequencing electropherograms using only ABI software applications. It is therefore necessary to import the analyzed files into the Beckman CEQ Sequencing analysis software for peak height determination.

ABI sequencing file extraction

All ABI sequencing results are stored locally on the hard drive of the computer running the automated sequencer. These files are generally located on the D drive of the computer. Locate the Run folder and either copy the whole folder or the specific selected files that correspond to the files that are to be used to perform linkage phase analysis. The files can either be copied to a network drive that the computer which has the Beckman CEQ analysis software can access or to a USB flash drive (or any other desired media).

Beckman CEQ Analysis Software import and peak height analysis

Using the Beckman CEQ Sequencing Analysis Software, samples are imported using the import function under the File menu. Following importation, all results need to be closed and saved into the SQL database that controls the software. This is done by using the Close all function under the Windows menu. Ones all of the samples have been saved and closed; the sequencing analysis software can be exited.

Beckman CEQ Result Extraction

Now all of the samples can be accessed in the Database manager of the Beckman CEQ Software system. All of the sample results have been stored in the current default database. Under the current default database, the sequencing results are stored under the Sequence Results tab. Once located and selected, they can be exported by right clicking on the file and clicking export. The Export tab contains multiple sample elements that can be exported. The elements necessary for linkage phase analysis are Results Data and

Quality Parameters. Under Options, make sure that the Remove CEQ Tracking Suffix box is checked. This will remove the Beckman CEQ database tracking information from the file name. The files can then be exported by leaving Default in the File name box and then saved as Text (Tab Delimited) (*.txt). Export the files to an appropriate location, either on the hard drive or the file server.

Excel Data Analysis

All linkage phase analysis have been performed using specially developed linkage phase analysis spreadsheets that are pretty involved and require that the user know some of the intricacies of excel functions. This has been done due to the lack of specific software capable of performing the detailed analysis of sequence peak heights.

There are 2 spreadsheets needed for the complete data analysis of mtDNA mixtures by linkage phase analysis. The first spreadsheet workbook (Alignment Template.xls) aligns the initial analyzed sequences. It will perform 7 alignments at most while also identifying mixed base positions and the percent of each of the two major components. If needed more sample results can be aligned for linkage phase analysis, however, these will have to be performed in a new alignment workbook template. The second spreadsheet workbook (Linkage Phase Template.xls) performs the linkage phase analysis. It is designed to retrieve information from one or two alignment workbooks and display resolved mixed base positions as well as the confidence behind each resolved base.

Alignment Workbook

The first sheet displayed in the workbook (Sequence Information) contains all of the important information for sequence identity. It is also where all of the mixed bases are identified by the analyst as well as the Anderson reference base location for the mixed bases.

The second spreadsheet (%) shows the alignment of the data that will be imported and the percent composition of each of the four fluorescent dyes for each of the locations.

The third sheet (Data Import) is where the analyst imports the sequencing index files for each of the fractions to be analyzed. The workbook then parses through the index files in the subsequent sheets to analyze the mixed bases identified by the analyst on the sequence information sheet.

A stepwise guide to the spreadsheet:

1. Enter sequence file name in the yellow box in the sequence information tab
2. Enter the fraction label from the DHPLC corresponding to the fraction that was captured and sequenced in the yellow box following the file name. (If a sample is an unfractionated sample denote it with an X.)
3. Push F9 to update the spreadsheet.
4. Go to the Data Import tab and paste each sequencing index file obtained in section 9.1.3
5. Push F9 to update the spreadsheet.

6. Go to the % tab and identify the mixed base positions in each of the fractions analyzed. Once identified proceed to the sequence information tab and enter the numbers under each corresponding fraction.
7. Push F9 to update the spreadsheet.
8. Review the data and save the alignment workbook using a unique name.

NOTE: If the intensities of the two major fluorescent data points for a particular base are the same (*i.e.* A (475 RFU) and G (475 RFU)) then the spreadsheet will make an error as it will consider them the same and therefore picks the third channel of absorbencies as the second highest.

Linkage Phase Workbook

This workbook only requires the analyst to update the source links to the alignment files in order to perform linkage phase analysis. It requires no input on the analyst's behalf other than to save the file appropriately prior to printing.
The workbook performs comparisons to all of the fractions entered into the alignment templates as well as the mixed bases identified. The calculations performed will identify the greatest average enrichment for all of the mixed bases and use that greatest average difference to pick fractions for linkage phase analysis. Individual mixed base compositions are then used for linkage phase analysis and confidence determinations. The spreadsheet will then display the average confidence, its standard deviation, minimum confidence and the count of samples behind the statistics.

To update links:

1. On the Excel menu bar click Edit – Links
2. Select Alignment Template - 1.xls and hit Change Source
3. Locate the source for which you want to perform linkage phase analysis
4. If you are only performing linkage phase analysis on one alignment template, hit close. If you are also linking to an additional alignment template for additional fraction information, select Alignment Template – 2.xls and Change Source. Locate the source and then close.
5. Push F9 to update the spreadsheet.
6. Review the data and save the alignment workbook using a unique name.
7. Use the Print Report button to print the linkage phase report.

Interpretation Guidelines

Linkage phase analysis interpretation relies on a significant amount of statistical analysis for which the data were gathered during the validation study. The data are based on the analysis of 202 unique mixed sites and a total of 347 mixed bases at 13 different mixture ratios. These yielded 27066 sample comparisons for confidence determination. The stepped percent enrichment has been summarized and averaged to yield a statistical confidence table (Appendix IV). This table is then used to determine the average confidence, the confidence standard deviation, minimum confidence and the count of the data points behind the statistical values for the percent enrichment achieved by mixture fractionation.

1. Linkage phase analysis should be performed on samples that are identified as mixtures by DHPLC analysis and confirmed by DNA sequencing and interpreted as such under the sequencing interpretation guidelines.
2. Linkage phase analysis can be used to determine the two contributors to a mixed mixture based on the statistics obtained and previously described.
3. Linkage phase analysis can be used to identify sequencing ambiguities not linked to the mixture but only to dye terminator noise.
4. Linkage phase analysis interpretation becomes extremely hard to perform following apparent major length heteroplasmy. Minor length heteroplasmy should be treated with caution, especially in cases where minimal separation is achieved.
5. Linkage phase analysis is not recommended unless the sequencing electropherograms are of high quality.
6. Linkage phase confidence calls are divided into three categories:
    a. Green – enrichment of equal to or greater than 18% yields confidence of base separation greater than 99.9% on average.
    b. Yellow – enrichment of greater than 7%, but lower than 18% yields base separation confidence of greater than 99% on average.
    c. Red – enrichment of less than 7%, yields base separation confidence below 99% on average.

Interpretation guideline – the application of linkage phase analysis fails when two samples appear to be length heteroplasmic. In these cases, all bases including the site of length heteroplasmy and following the length heteroplasmy must be excluded from the analysis of linkage phase (ex. E6 D10 C2 70-30 Pre-amplification Mixture Separation)

It remains to be determined at what position it is appropriate to start performing linkage phase analysis due to common sequence ambiguities at the very beginning of a sequence trace. (Noe: this area lies outside of the sequence of forensic relevance)

Statistical Analysis

The application of statistics in mtDNA analysis will greatly increase the confidence in results obtained by mtDNA sequencing. This should be performed on any sample that is sequenced by BigDye v1.1 Dye terminator sequencing.

Appendix V – Peak height consistency lists the peak heights of individual peaks obtained at 5 different DNA concentrations for each of the eight primers used for forensic mtDNA profiling. Although the optimal DNA input is 0.8ng for DNA sequencing other input concentrations are sometimes unavoidable and are therefore listed as well, if needed.

Appendix IV – Enrichment Confidence lists the average, minimum, standard deviations and sample counts of the statistical confidence for each achieved separation percentage point. It is to be applied when a suspected mixture of two components is being investigated.

Single Source Statistics

Samples considered being of a single source (*i.e.* no clear peaks identified as having a secondary underlying peak of greater than 10%) should be statistically analyzed for peak height consistency. This is done by comparing the individual peak heights of the questioned sequence to the results obtained from the peak height consistency validation study. Samples are compared by using a moving eleven base window to which the central base peak height is evaluated as a fraction of the total fluorescence of all eleven peaks. If the evaluated peak is above or below 2 standard deviations of that of the reference database it is flagged as an anomaly.

Fractionated Linkage Phase Analysis Resource System (FLiPARS)

FLiPARS Home Folder: C:\Program Files\FLiPARS\ - Contains:

- Excel template for LiPhA (modifiable, just be sure to make the new version the same name as original).
- Access database containing reference sequence information.
- Simulated casework folder with 16 demonstration analyses.

FLiPARS Software General Information

FLiPARS (Fractional Linkage Phase Analysis Resource System) is a conglomeration of several smaller applications integrated into a Microsoft Windows Visual Basic application. It is designed to streamline the data analysis process of electropherogram data derived from the mtDNA mixture fraction linkage phase analysis method.

The smaller portions of the application are Perl modules which may be accessed by the FLiPARS Windows application in order to gather sequence and electropherogram data from Raw format sequence files (Tab delimeted .txt files [Beckman Format]). In addition these smaller portions of the program also perform sequence alignment using the Needleman-Wunsch alignment algorithm. Though very basic, this algorithm provides satisfactory subject to reference sequence alignment for the human hypervariable regions.

FLiPARS Software Limitations

Like many open-source, collaborative software applications, FLiPARS has always been a constantly evolving entity – It saw it's beginning as a very rough Perl script and progressed into a fully functioning Graphical User Interface.

However, FLiPARS is a work in progress and several areas remain to be improved, these include but are not limited to:

- Enhanced data processing – less lag time on computationally intensive processes via more efficient coding.
- Greater database integration – Expansion of the software capabilities to store and retrieve data from data sources such as Microsoft Access databases or various SQL database servers.
- "Smoother" user interface – Many aesthetic enhancements may be made.

FLiPARS is a fully-functional software application with a great potential for growth, it is ideally suited as an open-source software application which may be improved upon via testing and feedback from selected users or testers.

FLiPARS Use Instructions – Close all unnecessary processes that are running!

Be patient – It is a very CPU intensive program and may lag 20-40 seconds. Or it may lag for only 5 seconds. It all depends on the data and computer.

Rule 1: If it crashes, simply restart it and all will be well.

Rule 2: It is best to load the application and run one analysis, and then use the File | New Analysis function before demonstration. It will improve speed.

1. Start the application by clicking the icon on the desktop "FLiPARS v0.24b".
   The application may take 15-20s to load on slower machines. You will see the splash screen appear.

2. The main application interface will then appear.

3. From the upper left hand corner (Project Settings), select the "Primer type" drop down box and select the primer (A1, C1, or A2 for this demo). *(Note the cancel button works in case the wrong primer is selected)*

4. Click the "Import Fractions" Button

5. The Open File Dialog will appear (Should be at casework directory, OR:
   Navigate to C:\Program Files\FLiPARS\Casework\)

6. Enter the Folder appropriate for the selected primer (Clearly marked) and double click on the sub-folder containing the desired batch of sequence data

7. You will see several *.txt files ***click and drag from the top (file with "1 –" prefix)*** until all files are selected, release and click "OK"

8. **THE APPLICATION WILL HANG FOR A MOMENT, THIS IS NORMAL AS DATA PROCESSING IS INTENSIVE.**

9. You will then see the fractions you imported in individual cells:
   - You may view the sequence alignments by pressing "View alignment"
     This compares that unique sequence/fraction file to the reference sequence.

10. Upon satisfactory inspection/demonstration, click the "View Fractions Button"

11. **THE APPLICATION WILL HANG FOR A MOMENT, THIS IS NORMAL AS DATA PROCESSING IS INTENSIVE.**

12. You will then seen a massive grid with the rCRS numbering.
    - The rows (vertical) contain the position seen as differing from the rCRS by the alignment.
    - The columns contain the fraction number / file.
    - The cell (row, column) is the basecall and composition data (%A,C,G,T) for that position (row) within that file (column).
      i. *Cells filled in orange/red are those that are pulled from the corresponding file via the alignment algorithm with respect to the other cells in that row*

(position). Generally, the less Orange/Red cells in a row – The more accurate the analysis (because of possibly undiscovered bugs).

13. Select positions to analyze by clicking on the checkboxes for desired positions underneath the "Analyze" column.

14. Once all are selected, click "Analyze" button, Excel should open.

15. ********ONCE EXCEL OPENS AND IS IDLE, PRESS F9 Inside of Excel in order to perform the calculations (Could be set to automatic, but creates lag).**************

16. Compare report to those in binder OR Save in origin folder as an .xls file.

17. Close Excel/View Fractions Window and to start a new analysis you must click in the toolbar menu File | New Analysis…

FLiPARS Known Bugs, Issues, Areas for Improvement

1. The application display will hang while importing fractions or viewing fractions or analyzing. This can be remedied by implementing "BackgroundWorker" objects in Visual Basic (requires fairly extensive re-working of code but would be for the best).
2. For reverse primers, Insertion numbering starts at "Position.99" when it should be "Position.1". This may be remedied by implementing a custom sorting scheme in the MSDataGridView. All other numbering is ok.
3. Deletions may appear as "D" in the Reference Sequence column of the Fraction analysis window. This just has to do with order of sequence importation and just requires the code to refer to the reference sequence upon import in order to pull the reference basecall.