Combined Extraction Method for Mitochondrial DNA and Proteins from Hair for Human Identification

Award Number: 2017-DN-BX-0159

Contract Period: July 1, 2018 to June 30, 2019

Final Summary Overview

Submitted by:

Kelly Landolfi
Contracting Officer
Battelle Memorial Institute
505 King Avenue
Columbus, OH 43201-2693

Date

Submitted to: Department of Justice
Office of Justice Programs
810 7th Street, NW
Washington, DC 20531

Submission Date: January 24, 2020

Sydney Corkran
Project Manager
CBRNE Defense, Battelle
505 King Ave
Columbus, OH 43201
Office: 614.424.7163
Email: corkran@battelle.org

Craig Bartling
Principal Investigator
Applied Genomics, Battelle
505 King Ave
Columbus, OH 43201
Office: 614.424.5377
Mobile: 614.702.1189
Email: bartlingc@battelle.org

Table of Contents

Page

Acronyms and Abbreviations

Acronym or Abbreviation	Definition
DNA	deoxyribonucleic acid
FBI	Federal Bureau of Investigation
GVP	Genetically Variant Peptide
IRB	Institutional review board
MPS	Massively parallel sequencing
mtDNA	mitochondrial DNA
NGS	Next Generation Sequencing
NIJ	National Institute of Justice
PCR	Polymerase Chain Reaction
PI	Principal Investigator
PM	Program Manager
POC	Point of Contact
SNP	single nucleotide polymorphism
SOPs	Standard Operating Procedures
UC	University of California

1.0 Purpose

The purpose of this study was to develop and demonstrate a method to obtain whole mitochondrial (mt)-genome haplotype and proteomic profiles from a single limited hair sample. The study was led by Battelle Memorial Institute and included participation from the University of California (UC) Davis. The result of research conducted under this grant is an optimized method for the extraction of protein and mtDNA from the same hair sample. The following report summarizes the project, methods, results, and impact at a high level. Further details are provided in our pending manuscript entitled "Demonstration of a Mitochondrial DNA-compatible Workflow for Genetically Variant Peptide Identification from Human Hair Samples" by Catlin et al.

2.0 Project Design, Methods, and Data Analysis

The study conducted under this grant consists of four main technical tasks:

- **Task 1: Sample collection**. Hair (head, pubic, and arm) and saliva samples (and associated metadata) were collected from a diverse cohort of individuals (60 individuals).
- **Task 2: Workflow ideation and planning**. This task included an ideation session between Battelle and UC Davis and concluded with documented plans for workflow testing.
- **Task 3: Baseline analysis**. This task involved establishing a working "detection limit" for current state-of-the art methods [SD (Sodium Dodecanoate) method, Modified SD method, Triple Prep kit, and FASP method] using mitochondrial DNA quantification, mass spectrometry, and protein identification. This task included work at Battelle (bulk of the

work) and UC Davis (four samples at least in duplicate with at least three different hair lengths). The FASP method was selected for further optimization.

Task 4: Workflow down-selection. Using a working set of hair samples, this task involved testing the workflows against established metrics. This task included work at Battelle (bulk of the work) and UC Davis (eight hair samples analyzed at UC Davis). The optimized FASP method was used, followed by reverse-phase microflow liquid chromatography coupled with high resolution quadrupole time-of-flight mass spectrometry (μHPLC-Q-TOF-MS) with protein identification using ProteinPilot and whole genome mtDNA sequencing on an Applied Biosystems Ion S5. Sequencing results were processed using Applied Biosystems Converge software.

3.0 Project Findings

3.1 Combined Protein and mtDNA Extraction Methods

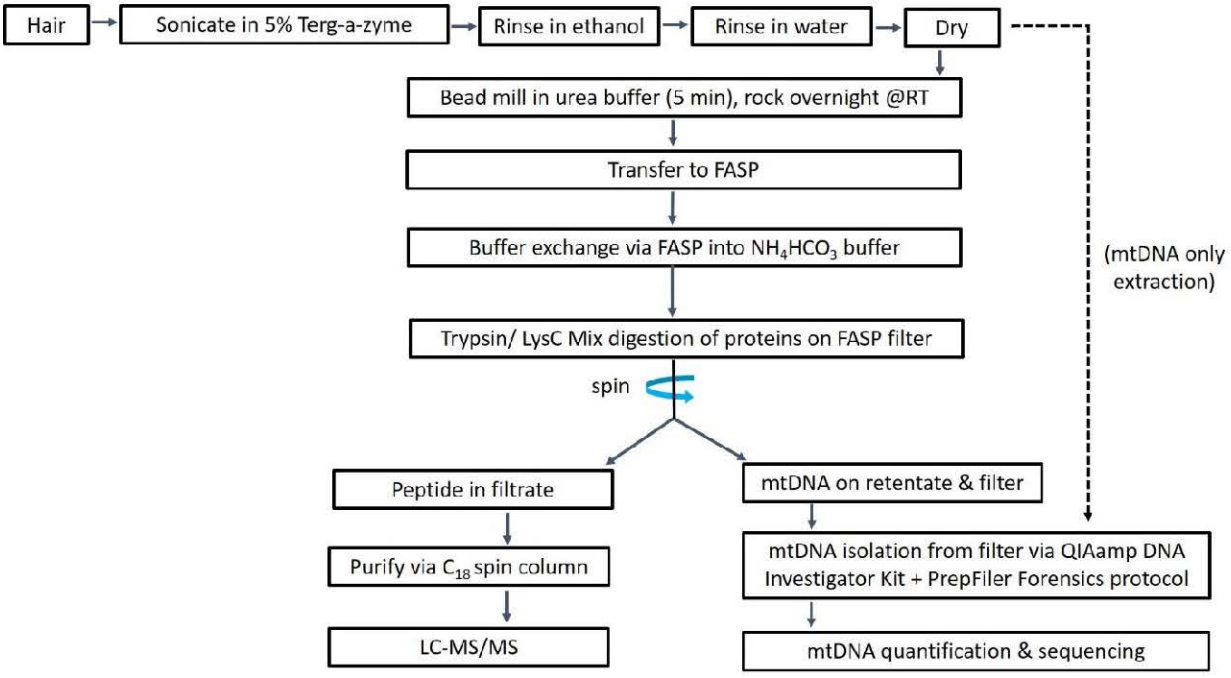
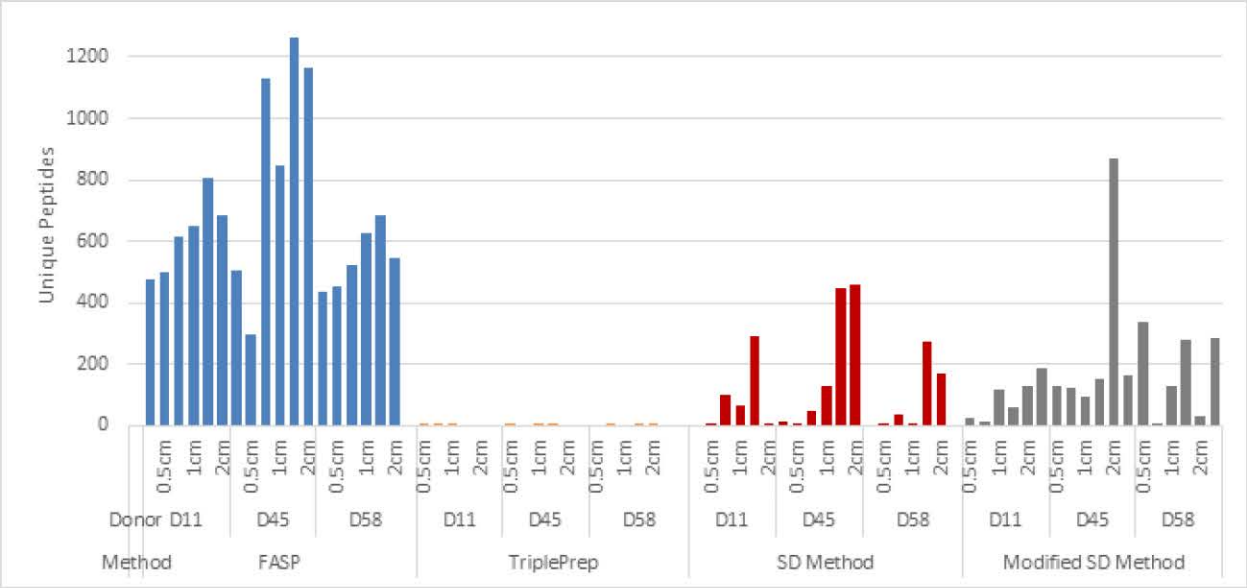


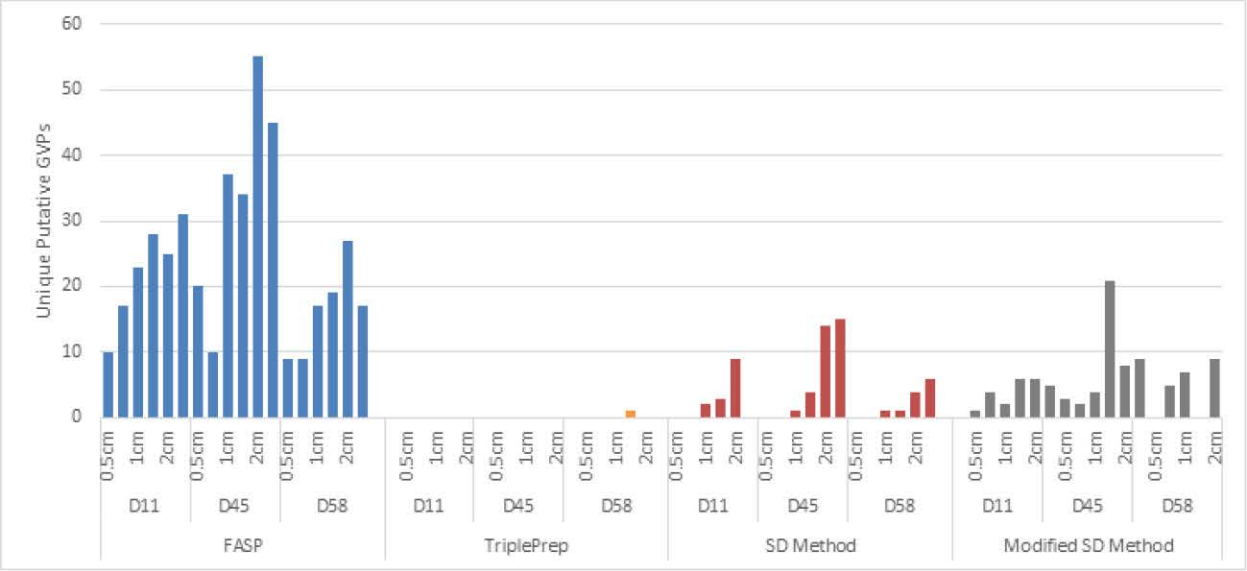
Figure 1. Optimized FASP Workflow.

Different hair lengths from 3 different donors (Donor 11, 58, and 45) were tested using the four different protein extraction methods. All hairs from donors 11 and 58 were head hairs, while the hairs tested for donor 45 were pubic. The TriplePrep kit demonstrated low counts of peptides, putative GVPs, and mtDNA copies in all cases. The vast majority (15/18) of samples did not result in any detectable mtDNA copies. When the SD method was tested, 16 of 18 samples presented mtDNA copies, an average peptide yield of 114 unique tryptic peptides, and an average of 3 unique putative GVPs. All FASP samples (18/18) generated mtDNA results, an average of 678 unique peptides and an average of 24 unique GVPs. The samples processed using the Modified SD method performed poorly for mtDNA; 13 of 18 samples did not generate any mtDNA results. However, the Modified SD method performed similarly to the SD method in terms of unique peptides (avg: 174) and putative GVPs (avg: 5). Overall, a much larger number of GVPs resulted from samples processed using the FASP method when compared to the other three methods tested [Figure 2A-2B]. Additionally, the numbers of unique peptides (and GVPs) identified decreased with decreasing amounts of hair for all methods tested. The SD Method was also performed for these sample at UCD to confirm the trends of decreasing peptide identifications with decreasing hair lengths.

A



B



The FASP and SD methods resulted in measurable mtDNA yields in all cases. For 1cm hair lengths, the FASP method (Average: 1,681) outperformed the SD method (Average: 109). In fact, these tests revealed that the FASP method provided higher quantities of mtDNA than all other combined methods tested. Even down to 0.5 cm of hair, the FASP method resulted in measurable mtDNA yields for all three donors and both replicates tested, where other methods did not. Like the proteomic results, pubic hair samples consistently generated higher mtDNA

copies than the other hair types. As expected, there was an increase of mtDNA copies, number of unique peptides and putative GVPs with the increase of hair length.

To benchmark these combined methods against mtDNA extraction protocols, the "mito only" extraction protocol was also performed. The FASP protocol did not yield as much mtDNA as the "mito only" method. However, subsequent optimization and repetition of our FASP method improved mtDNA yields closer to the "mito only" protocol.

3.2 FASP Method Optimization

As demonstrated in the previous section, the FASP protocol resulted in high numbers of unique peptides and putative GVPs as well as measurable mtDNA yields, but optimization of various steps of the method improved results. Initial investigation focused on the transfer step from the bead mill tube to the FASP filter. Subsequent repetitions of the method involved washing the beads following initial transfer, and specific care was taken to transfer all residual material. This modification to the protocol resulted in greatly improved mtDNA recovery and more consistent peptide results. While this modification improved recovery, we also investigated the impact of other steps in the process, including hair washing and reagent preparation procedures, the reduction step, and the filtration steps as well. Additional rinsing steps with Terg-a-zyme produced marginally more unique peptides as well as fewer unique peptides in the negative controls, suggesting a possible gain in sensitivity and contamination control with the use of Terg-a-zyme. We also identified that the passivation of filters with TWEEN®-20 slightly increased the number of unique peptides compared to no passivation.

Additional studies into the selection of reducing agent (DTT or TCEP) and concentration (50nM or 500nM) were also investigated. The total unique peptide counts for each concentration of the reducing agents were compared to the original method (50mM DTT) using the t-test. The use of 500mM TCEP proved to have significantly lower values of mtDNA (t-test, p=0.01) and unique

number of peptides (t-test, p=0.04) compared to the 50mM DTT method. No other comparison resulted in a significant statistical value.

The FASP method recovers approximately 10% of the mtDNA compared to the mito only protocol when starting with hair samples. The control experiments with 2800M described suggest that mtDNA is not significantly lost in the FASP process. However, the 2800M spike experiments differ from hair-derived mtDNA in that it was spiked at higher levels and the mtDNA source differs (2800M is isolated from blood and likely has different levels of degradation than mtDNA from hair). Thus, some hair-derived mtDNA may still be lost in the FASP process during the several spin filtrations in the protocol (Figure 1).

We removed urea from the extraction buffer, as it may denature DNA, causing it to more easily pass through the filter membrane; however, removal of urea from the buffer did not improve recovery. Further, quantification of filters and flow throughs over time from the various filtration steps indicated that the mtDNA was not degraded or significantly lost during the washes or incubation steps. Results showed that an average of 8% of the total mtDNA was lost in the final peptide elution. This average of 8% had a standard deviation of 10%, with 12/22 resulting in <2% of total mtDNA in this fraction, suggesting that this elution step may cause some, albeit small and variable, loss.

Tests were performed to determine whether disrupting the filter would improve recovery. Millipore filters were spiked with 2800M then pre-treated with either, Phenol:Chloroform:Isoamyl Alcohol (PCIA), Punch Solution, or dried (no chemical treatment) then underwent the "mito only" extraction procedure. These results demonstrate that the "mito only" procedure sufficiently extracts mtDNA from the filter. The highest recovery was garnered when the "mito only" protocol was performed using filters that were allowed to dry overnight at 2-8°C rather than proceeding immediately after the DNA was spiked in.

Based on the experiments described above, we determined that the initial transfer from the bead tube to the FASP filter is critical to mtDNA yields. No other single step causes significant peptide or mtDNA loss, but repetitive filtrations, washes, and sample transfers may lead to some small losses across the entire process.

3.3. Verification of the Optimized Method

To verify the reproducibility of our mtDNA-compatible proteomic workflow (Figure 1), hair samples of 1 cm in length from the same donors were processed using the optimized FASP method at Battelle and at the University of California-Davis (Parker Laboratory). Similar results were generated by both laboratories. While results are not directly comparable due to differences in instrumentation and peptide identification workflows, we did demonstrate that at a high level, the

To establish our workflow for a range of hair types, the optimized FASP method was tested using arm, head, and pubic hair samples from a diverse cohort of 22 individuals. The results verify that GVPs can be identified using our method (GVPs identified in all samples) and putative GVP profiles can be obtained. The highest numbers of unique peptides and GVPs were found in the pubic hair samples (GVPs: 49.6±10, Peptides: 12,688±2010). Fewer unique peptides and GVPs were identified in head hair (GVPs: 24.4±10.6, Peptides: 9,639±3344) and in arm hair (GVPs: 16.5±8.6, Peptides: 5,106±2,493).

Like the peptide analysis, the different hair types also contained different ranges of mtDNA copies: 2-3,066 copies in head hair; 250-20,104 copies in pubic hair; 11-508 copies in arm hair. While processing the hairs, it was clear that the pubic hairs tended to be thickest, while arm hairs most frequently appeared as very thin and fine. The difference in the hair thickness could contribute to the greater number of peptides and mtDNA copies present.

A subset of the mtDNA extracts were subjected to DNA sequencing. A total of 4 sequencing runs were performed, each containing 32 multiplexed samples. A consistent haplogroup call was obtained for hairs with 0.5-2 cm in length from the same individual, with the exception of donor 11 at 0.5 cm. This sample contained many dropouts, affecting the resulting haplogroup call. Regardless, the results demonstrate that mt-genome data can be generated for the optimized FASP method and the SD method. Sequencing was not performed for the samples from TriplePrep and Modified SD methods due to low mtDNA and peptide yields.

To verify that our FASP method can result in viable mt-genome profiles, another sequencing analysis was done for the large cohort of 22 donors. Overall, the haplogroup calls are consistent within each donor across hair types and buccal samples, with some minor exceptions. The samples that presented low mtDNA copy numbed were the ones with the highest number of variant drop-outs. The different hair types and buccal samples for one donor presented similar number of variants. However, the Converge software designated a different haplogroup to the pubic hair sample. After further examination, it was noticed that Converge assigns the haplogroup before any calls are manually adjusted and that the haplogroup call does not update after manual changes are made. EMPOP (https://empop.online/) was then used to further verify the haplogroup, and all results obtained were concordant for all sample types. Subsequently, the same verification using EMPOP was performed for all samples from the remaining donors, and all haplogroup were concordant between EMPOP and Converge.

There were also some trends with hair types and characteristics, although these observations would need to be confirmed with further testing. Overall, the samples with the lowest mtDNA input were arm hairs, which were visibly finer in texture than the other hair types. Amongst the head hair samples, those with under 3,000 mtDNA copy number input were attributed to donors

with dyed hair. And five out of eight donors whom had dyed hair, six had done so recently (within 6 months of sample collection).

4.0 Data Dissemination

An abstract was submitted to the 2019 National Institute of Justice Forensic Science Research and Development Symposium but was not accepted. A manuscript was completed in June and published by Forensic Science International: Genetics.

> Catlin, L.A., RM Chou, ZC Goecker, LA Mullins, DSBSS Silva, RR Spurbeck, GJ Parker, CM Bartling; "Demonstration of mitochondrial DNA-compatible workflow for genetically variant peptide identification from human hair samples"; Forensic Science International: Genetics; 43:2019; 1-12

In addition, an abstract was accepted to the 2019 International Symposium on Huma Identification.

5.0 Impact

We expect the results from this study to have an immediate impact on current proteomics research for human identification purposes. Researchers will be able to analyze proteins from valuable samples without sacrificing DNA quantity or quality. We consider this program to be the first step in the path to ultimate implementation of our method into routine forensic laboratory analysis. Further, the method can potentially be modified for other sample types, such as touch samples.

The GVP identifications presented here are putative based on proteomic results and may require validation of imputed SNPs using exome, targeted, or SNP sequencing. Our experiments provide foundational data for testing the discrimination power of GVPs panels alone and in combination with mtDNA profiles. Further statistical testing would provide valuable insights into the forensic

utility of our method. Our data set is also a source of information to explore correlations between metadata variables (hair color, race, hair source, etc.) and GVP or mtDNA profiles. Statistical comparisons such as principal component analysis (PCA) and clustering techniques could be explored for identifying relationships between the metadata and GVP profiles, overall proteome profiles, and/or mt-haplotypes. Researchers have used such techniques to determine protein signatures that allow for distinguishing hair source (e.g., head versus pubic) and race (e.g., Caucasian versus African American) as well as other unique properties of the sample.

It can be done