IDENTIFYING INDIVIDUALS THROUGH PROTEOMIC ANALYSIS

Tatiana Perez[1], Heyi Yang[1], Amy Mundorff[2], Audrey L. Scott[2], David Fenyö[3], Jennifer Teubl[3], Donald Siegel[1]

1) Department of Forensic Biology, NYC Office of Chief Medical Examiner, NY, 212.323.1923; tperez@ocme.nyc.gov
2) Department of Anthropology, University of Tennessee, Knoxville, TN
3) Center for Health Informatics and Bioinformatics, New York University, NY

Introduction: The preeminent method for individual human identification is use of genomic DNA short tandem repeats (**STRs**). However, while this method has very high discriminatory power, it can fail to give full STR profiles when DNA is degraded by taphonomic processes, burning or intentional use of caustic substances. Although mitochondrial DNA is often employed as a backup method it lacks the discriminatory power of genomic DNA and it too is subject to degradation. Additionally, because both methods rely on DNA amplification for signal detection, both techniques i) require intact DNA sequences from primer to primer, ii) are subject to DNA polymerase errors, iii) can fail to amplify in the presence of polymerase inhibitors, iv) are expensive and v) can take a significant amount of time. [1,2] The three latter obstacles can be problematic following a mass disaster that can result in large numbers of fragmentary remains, where rapid identification of victims is necessary for investigative leads and to allow burials within a reasonable timeframe. Consequently, an alternative method to accurately and efficiently identify large numbers of fragmentary human remains is needed. Proteins may offer an alternative as they are i) less liable than nucleic acids, ii) can be informative even when partially degraded, iii) do not require amplification for detection and iv) the assay is inexpensive and relatively fast. Proteins are ideal molecules for testing as they are detectable in very low amounts[3] and long-lasting.[4,5] Besides clinical use in identifying disorders[6], protein- mass spectrometry is currently being used to identify species [7,8,9].

Here we present data for a proteomic mass spectrometry (**MS**) assay to identify informative amino acid polymorphisms in human muscle proteins (confirmed by gene sequencing analysis) that can be used for individual identification. The goal is to develop an accurate, sensitive and rapid method for the identification of comingled remains or large numbers of fragmentary human remains following a mass disaster.

Methods: Muscle and bone samples were collected from cadavers placed in the open during four seasonal trials (winter, spring, summer, autumn) at University of Tennessee's Anthropological Research Facility. Muscle samples were collected from multiple positions over variable time periods for taphonomic analysis. Each trial consisted of three to four cadavers, ranging from ages 42-90. Here we report results on day 0 samples taken from each of the four cadavers during the first (winter) trial and processed for high performance liquid chromatography (**HPLC**) and Matrix-Assisted Laser Desorption/Ionization and tandem time-of-flight mass spectrometry (**MALDI TOF/TOF**). Mass spectra were searched against the NCBI non-redundant mammalian protein database (downloaded September 2015) using ProteinPilot 4.5 (SCIEX), and X!Tandem (The Global Proteome Machine Organization) search algorithms. Peptides with a >95% confidence interval were used for polymorphism identification. Candidate polymorphism selection excluded peptides with isobaric amino acids leucine/isoleucine and lysine/glutamine, as well as those peptides with single amino acid changes that represent less than 30% of that peptide's total coverage. For genomic DNA confirmation, sample extraction, primer design, amplification and Sanger sequencing was performed by Genewiz. Primers were placed in intronic regions flanking the peptide region where polymorphisms were suspected.

Results: HPLC MALDI TOF/TOF from the four individuals of this trial detected an average of 2825 peptides (range = 2207 to 3275) which represented approximately 66.5% of total spectra identified (range = 55.1 to 77.9) and a total of 219 proteins (range = 123 to 280) at 95% confidence interval. Over three thousand peptides were evaluated for polymorphisms using the selection criteria described above and 19 candidates identified. Five polymorphisms were sent for DNA confirmation. Two polymorphisms were confirmed in four individuals for titan (TTN) and enolase (ENO3). Individuals were correctly identified who were homozygous polymorphic, homozygous wild type and heterozygous for both genes. DNA analysis has led us to reevaluate and improve our MS data screening procedures to more accurately identify amino acid polymorphisms. These criteria will be used in ongoing muscle and bone experiments. We are currently working on determining polymorphism population frequencies.

Supported by National Institute of Justice Grant 2014-DN-BX-K014

References:

1. Mundorff, A.Z., *Recovery, Analysis, and Identification of Commingled Human Remains*, ed. J.E.B. Bradley J. Adams. Vol. Chapter 7. 2008.

2. Sledzik, E.J.K.a.P.S., *Recovery, Analysis, and Identification of Commingled Human Remains*, ed. J.E.B. Bradley J. Adams. Vol. Chapter 17. 2008. 317.

3. Masters, P.M., *Preferential preservation of noncollagenous protein during bone diagenesis: Implications for chronometric and stable isotopic measurements.*Geochimica et Cosmochimica, Acta, 1987. Vol. 51, pp.3209-3214.

4. Schweitzer, M.H. et al. *Molecular analyses of dinosaur osteocytes support the presence of endogenous molecules*. Bone, 2013. Vol. 52, pp.414-423.

5. Weiner, S. et al., *Characterization of 80-million year old mollusk shell proteins.* Geology, 1976. Vol. 73, No.8, pp.2541-2545.

6. Moat, S.J., et al., *Newborn blood spot screening for sickle cell disease by using tandem mass spectrometry: implementation of a protocol to identify only the disease states of sickle cell disease.* Clin Chem, 2014. 60(2): p. 373-80.

7. Buckley, M., et al., *Species identification by analysis of bone collagen using matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry.* Rapid Commun Mass Spectrom, 2009. 23(23): p. 3843-54.

8. Yang HY, Z.B., Prinz M, Siegel D *Primate and Other Species Identification Based on Proteomics Data*. in *61st ASMS Conference*. 2013. Minneapolis, MN.

9. Linacre, A., *Capillary electrophoresis of mtDNA cytochrome b gene sequences for animal species identification.* Methods Mol Biol, 2012. 830: p. 321-9.