# IDENTIFICATION OF NON-SYNONYMOUS SNPs IN ARCHAEOLOGICAL HAIR PROTEIN: CALCULATION OF MEASURES OF IDENTITY AND BIOGEOGRAPHIC BACKGROUND

Jonathan Hilmer, Katie Giddons, Tami Leppert, Brian Bothner, Mark Leppert, Andrew Wilson, Glendon Parker, Utah Valley University

Genetic variation provides the basis for developing measures of individual identity. The short-tandem repeats in DNA have proven to be particularly effective as means of developing measures of identity. Other genetic features however can also be used to develop measures of identity. Based on the NHLBI Exome Sequencing Project (ESP), 28,200 and 46,200 non-synonymous Single Nucleotide Polymorphisms (nsSNPs; MAFs> 1%), occur in the European-American and the African-American population respectively. These loci of genetic variation have the advantage of being preserved as changes in protein primary structure, which is highly stable and persists in the environment after the more labile DNA has degraded. Shot-gun proteomics can identify peptides containing nsSNPs which have the potential to provide a measure of individual identity when DNA-typing is either not possible or too difficult to interpret. Hair samples from three 18$^{th}$ century and three 19$^{th}$ century individuals were collected from archaeological contexts in London, England. Each sample was washed using 20% Methanol, and milled in 50 mM ammonium bicarbonate and then extracted with acetone and treated with 10 mM DTT and 0.1% Protease-Max (Promega), and 4µg TPCK-treated trypsin (Worthington). The insoluble sample was split into two 1mg aliquots. The first was treated in 20µL of 20 mM ammonium bicarbonate containing 6M Urea, 100 mM DTT and 0.01% Protease-Max. After 16 h 5µL of 0.8M iodoacetamide was added, followed 60 minutes later by 50µL of 50mM ammonium bicarbonate containing 0.017% Protease-Max containing 5µg TPCK-treated trypsin (overnight at RT). The second aliquot of hair was treated with 100 µl of 100mM ammonium bicarbonate containing 100 mM DTT and 0.05% sodium deoxycholate. After 2 h iodoacetamide was added to 2-fold molar excess, followed 2 h later by trypsin in a final volume of 220µL at a concentration of 20µg trypsin per ml. The sample was then acidified with formic acid to precipitate the deoxycholate detergent. The supernatants of each treatment were applied to a Bruker maXis Impact quadrupole Time-of-Flight mass spectrometry instrument. Mass Spectrometry of the 6 samples resulted in identification of 326 to 445 and 686 to 908 unique peptides from the above two methods respectively, when using the default settings for the “Global Proteome Machine” peptide spectra matching algorithm and database. There were considerable differences in the composition of each peptide population and chromatographic profiles as a result of the two different methods. Peptides corresponding to ten nsSNPs were identified in the analysis. Using the genotypic frequencies based on the European cohort in the 1000 Genomes project (n = 379) and using the product rule on nsSNPs in different gene products, measures of identity were calculated (ranging from 1 in 4.9 to 91.6). When the same calculation was conducted using genotypic frequencies from the African cohort in the 1000 Genomes Project (n = 240) the calculated measures of identity ranged from 1 in 26 to 1 in 1400. The quotient of “probabilities” (EUR/AFR) ranged from 4.9 to 17.5, demonstrating that the hair samples were more likely to originate from a European biogeographic context.