USE OF HAIR SHAFT PROTEIN TO OBTAIN MEASURES OF IDENTITY

Tami Leppert, Krishna Parsawar, Lisa Baird, Brandon Hanberg, Chad Nelson, Mark Leppert, and Glendon Parker
University of Utah and Utah Valley University, Utah, USA

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 can also be used to develop measures of identity. Over 650,000 non-synonymous Single Nucleotide Polymorphisms (nsSNPs) have been identified and the population frequencies determined. nsSNPs 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. We collected hair shafts and genomic DNA from 19 European American individuals. We cleaned each sample in with water, 20% methanol and then digested with TPCK-treated trypsin (Trypsin Gold, Promega) in the presence of 0.1% Protease-Max (Promega). The resulting peptides were applied to a hybrid LC/LTQ/FT mass spectrometer (Thermo Finnigan) and analyzed using the MASCOT Peptide Spectra Matching algorithm. Up to 300 proteins and over 80 peptides with nsSNPs were identified. Seven detected peptides have allelic frequencies below 50% in the European American Population. If independence is assumed and the product rule employed, measures of identity of up to 1 in 1750 can be calculated using polymorphic peptides from 17 nsSNP alleles (average = 1 in 160). The allelic status at 11 loci was subsequently directly confirmed using PCR sequencing on the genomic DNA of 13 individuals. The nsSNPs predicted by proteomic analysis of hair peptides were consistent with the genomic DNA in every instance. The MV1 and MV2 regions of mitochondrial DNA (mtDNA) were also sequenced in 13 individuals and mtDNA-based measures of identity calculated using the Utah population database. Two individuals had better measures of identity from protein-based measures of identity compared to mtDNA-based measures of identity, resulting in a combined measure of identity of up to 1 in 78500 (average increase 160-fold, average combined measure: 1 in 8100). The frequency of many nsSNPs is a function of the biogeographic background of the individual, with some nsSNP peptides up to 5.1–fold more common in the European American relative to the African-American population. The cumulative likelihood ratio ranged from 1 to 13-fold increased likelihood of European-American ancestry. This is consistent with the study design. Mass spectrometry is particularly sensitive. The data generated was obtained from the equivalent of less than 10µm of hair. In order to be readily applicable protein-based measures of identity need to be robust, consistent, genetically independent and powerful, sensitive and economical. We are currently developing algorithms to process and extract polymorphic peptides from complex proteomic mixtures. This will be a major step in the process of applying this new technology to the forensic context. ⌘