AN INTEGRATED MIXTURE SCREENING ASSAY USING HIGH RESOLUTION MELT CURVE ANALYSIS AND SUPPORT VECTOR MACHINE MODELING

Andrea L. Williams, Hannah E. Wines, M.S., Darianne C. Cloudy, M.S., Edward L. Boone, Ph.D., Tracey Dawson Cruz, Ph.D.
Department of Forensic Science, Virginia Commonwealth University

A primary challenge faced by forensic analysts is the demand for timely analysis of evidence. DNA analysis techniques have drastically increased in sensitivity, allowing for low template DNA samples to be detected and used for identification. Since low template samples are even more problematic when a DNA mixture is present, it would be advantageous to add an assay earlier in the DNA workflow that could detect a mixture and, potentially, determine the number of contributors. Real time PCR instruments have both quantification and high-resolution melt curve analysis (HRM) capabilities allowing for an opportunity to integrate an HRM screening assay into a DNA quantification kit. The melting behavior of DNA varies with nucleotide length/sequence, allowing melt curves to differentiate between single source samples (and their genotypes) and mixtures. In this work, STR loci D5S818 and D18S51 were chosen as amplification targets for HRM analysis along with linear discriminate analysis (LDA) as well as support vector machine analysis (SVM) with linear or radial basis functions for sample classification. After both STR targets were amplified and optimized separately, the loci were integrated into the Qiagen Investigator Quantiplex kit, a commercially available, human-specific qPCR kit. When LDA was used to predict STR genotypes from key HRM curve characteristics of single source melt curves, genotypes were accurately classified at rates of 48.21% and 35.71% for D5 and D18, respectively. When standard groups were combined for closely related genotypes (“geno-groups”), classification rates increased to 68.18% and 62.12%. Most importantly, 92.86% of single source samples and 100% of mixtures were accurately identified as such, providing an overall screening accuracy of 93.94% to distinguish between them. When the entire HRM curve data was used for classification, genotype prediction accuracies increased to 94.64% for D5 and 92.86% for D18. Using the whole curve, again, 100% of mixtures and 87.5% of single source samples were classified as such. The results of this work demonstrate the potential use of HRM analysis at the DNA quantification stage. Our data suggest this mixture-screening assay could be used as a time efficient, cost effective tool with a potential to give insight into the nature of a sample earlier in the forensic DNA workflow. In the future, a simple, user-friendly online analysis tool could be developed that will allow for the easy import of HRM data for automated genotype prediction and identification of single-source and mixture samples.