SEPARATING THE WHEAT FROM THE CHAFF: MATHEMATICALLY DISTINGUISHING LOW LEVEL ALLELES FROM ARTIFACTS AND NOISE

George Riley[1], Ashley Flack[2], Robert Goor[1], Douglas Hoffman[1], Tim Kalafut[3], Charity Noreuil[4], Curt Schuerman[3], Taylor Scott[2], Lindsey Smith[3], Joel Sutton[3], Stephen Sherry[1]
[1]National Center for Biotechnology Information, National Library of Medicine, NIH, Bethesda, MD,
[2]Illinois State Police, DNA Indexing Laboratory, Springfield, IL
[3]U.S. Army Criminal Investigation Laboratory, Defense Forensic Science Center, Forest Park, GA,
[4] Illinois State Police, Springfield Forensic Science Laboratory, Springfield, IL

Forensic DNA laboratories are attempting to find analytical thresholds that utilize the greatest amount of information possible in their STR profiles without mistaking small artifacts and noise for alleles. The SWGDAM Interpretation Guidelines for Autosomal STR Typing indicate that laboratories should not use their analytical threshold to avoid artifacts, rather should apply analytical thresholds that take the laboratory's empirical noise levels into account. This allows laboratories to utilize the maximum information in STR profiles, and not "leave data on the table." As laboratories begin to use analytical thresholds below 30 RFU to analyze complex mixed profiles, it becomes critical be able to distinguish actual alleles from small STR artifacts and excursions in the baseline noise. Software can employ various mathematically determined metrics to accomplish this.

OSIRIS, the Open Source Independent Review and Interpretation System, downloadable from the NCBI OSIRIS homepage, was created in response to recommendations arising from the World Trade Center victim identification effort. OSIRIS is in use as an expert system for CODIS samples and in clinical and forensic caseworking labs to analyze low level mixtures. OSRIS implements several unique STR analysis metrics and a variety of artifact signatures that give the software exceptional capabilities when analyzing low level STR profiles.

OSIRIS computes new metrics for peak morphology, peak shifting, sample to ladder fit, and channel specific baseline noise. OSIRIS also employs various mathematical artifact signatures, which embody different peak shapes, and applies these metrics and signatures to discriminate allele peaks from artifacts and noise. Low level artifacts include pull-up, non-specific peaks and random deviations in baseline noise that can appear to be peaks. Low level alleles include those masked as shoulders on larger alleles and can be mistaken for artifacts or noise. OSIRIS uses its calculated metrics and peak signatures to identify and annotate these and other artifacts.

Accurately and robustly identifying these artifacts increases the assurance of quality profiles and reduces both the editing burden and the number of conflicting analyst/reviewer calls that require resolution. In this way, OSIRIS improves the efficiency of analysts that are interpreting profiles. This gives analysts more time to do the important work of case interpretation.