A SIMULATION MODEL OF DNA TEMPLATE QUALITY, USED IN VALIDATING A GENETIC ANCESTRY ESTIMATION SYSTEM FOR FORENSIC APPLICATIONS

Jason Bryan, BS, Marc Bauchet, Ph.D., Victoria Vance, MS, Dan Hellwig, MFS, Lars Mouritsen, BS
Sorenson Genomics

Sorenson Forensics recently released a genetic ancestry estimation test known as Investigative LEAD^{SM} (Law Enforcement Ancestry DNA). This new test provides a means for law enforcement agencies to identify the genetic ancestry of suspects and/or victims. Software systems used to estimate genetic ancestry may vary based on the type and number of genetic markers, generally Single Nucleotide Polymorphisms (SNPs), observed and statistical algorithms used. The common premise of these systems, however, is to measure the genetic affinity of an individual in relation to representative parental populations. The result is generally reported in terms of relative percentages for each population. Laboratory test systems used to generate SNP data may vary by chemistry and analytical method, which can play a role in the number, quality, and accuracy of the genotypes used to derive ancestry estimations for a given DNA sample. Forensic-type samples often have low DNA copy numbers due to minimal sample amount or degradation and frequently contain inhibitors. These factors can lead to lower recovery rates of targeted loci and allele dropout or stochastic effect. When DNA quality or quantity is compromised, it is important to determine that the accurate genetic ancestry estimations can still be made. Sorenson Forensics' I-LEAD test makes use of the Applied Biosystems TaqMan OpenArray technology to genotype 192 autosomal SNPs. A proprietary algorithm developed at Sorenson is used to create the estimations of ancestry. An ancillary software program (Genotype Degrader) was created to assist in the validation of the algorithm used for the I-LEAD test system.

Initially, full genotype profiles for 190 SNPs were generated from DNA samples with known genetic ancestry. Settings were selected within the Genotype Degrader simulation program to randomly introduce specific amounts of locus dropout (up to 50%) and stochastic effect into each given genotype set (up to 30%). The simulation for each parent genotype profile can be run as many times as desired creating innumerable, unique combinations of the genotype set at a desired 'quality-level'. The ancestry estimation of the "degraded" genotypes can then be compared to that of the parent genotype to assess the impact random degradation would have on ancestry estimations. The Genotype Degrader software tool allowed for well-controlled experimentation to demonstrate the accuracy of the I-LEAD test when locus dropout and stochastic effects are observed in genotype data. These simulations greatly reduced the time and expenses for this type of study over actual sample testing in the forensic laboratory.