

CANDIDATE REFERENCE FAMILY DATA: A TOOL FOR VALIDATING KINSHIP ANALYSIS SOFTWARE

Kristen Lewis O'Connor, Ph.D.¹, Steven P. Myers, M.S.², Erica L.R. Butts, M.S.¹, Carolyn R. Hill, M.S.¹, John M. Butler, Ph.D.¹, and Peter M. Vallone, Ph.D.¹

¹ National Institute of Standards and Technology, Applied Genetics Group, Gaithersburg, MD 20899-8312

² California Department of Justice, Jan Bashinski DNA Laboratory, Richmond, CA 94804

Using DNA testing and statistical calculations, kinship analysis evaluates the strength of proposed familial relationships between individuals. Kinship analysis has a variety of applications: criminal and civil paternity cases, mass disaster victim identifications, missing persons identifications, military identifications, and immigration cases. Many software tools are commercially or freely available to aid kinship analysis; however, there is no standard dataset of familial genotypes to help validate calculations made by a software program. Currently, a laboratory must generate pedigrees and genotypes for individuals with known familial relationships. These genotypes are either simulated or taken from previous casework in the laboratory.

The goal of our work is to develop standard reference family data (SRFD) as a tool to aid laboratory validation of kinship analysis software. We are developing an artificial four-generation pedigree as candidate SRFD based on data collected from six different family groups analyzed with 46 autosomal STRs and 17 Y-STRs. The genotypes of the pedigree reflect observed Mendelian inheritance patterns, including mutations and rare alleles, within real families. The pedigree structure allows for kinship testing of pairwise comparisons (parent-offspring, full siblings, half siblings, first cousins, etc.), paternity trios, and motherless paternity. Due to the size of the pedigree, more complex tests (e.g., incest) can be constructed in the future. The SRFD can be used to verify the functionality of calculations performed within kinship analysis programs including the handling of mutations and rare alleles. Illustrations of how the pedigree data can be used will be demonstrated with GeneMarker® HID v. 1.90, a commercially available program from SoftGenetics, and KIn CALC v. 4.0, an Excel®-based freeware program developed at the California Department of Justice. To assist validation for the kinship testing community, the SRFD and pedigree, allele frequency data from major U.S. populations (1-3), and published likelihood ratio formulas will be made available on STRbase (<http://www.cstl.nist.gov/strbase/>), an online resource for the forensic genetics community (4).

References:

1. Butler, J.M., Schoske, R., Vallone, P.M., Redman, J.W., Kline, M.C. (2003) Allele frequencies for 15 autosomal STR loci on U.S. Caucasian, African American, and Hispanic populations. *J. Forensic Sci.* 48(4):908-911
2. Hill, C.R., Kline, M.C., Coble, M.D., Butler, J.M. (2008) Characterization of 26 miniSTR loci for improved analysis of degraded DNA samples. *J. Forensic Sci.* 53(1):73-80
3. Hill, C.R., Duewer, D.L., Kline, M.C., Sprecher, C.J., McLaren, R.S., Rabbach, D.R., Krenke, B.E., Ensenberger, M.G., Fulmer, P.M., Stort, D.R., Butler, J.M. (2010) Concordance and population studies along with stutter and peak height ratio analysis for the PowerPlex® ESX 17 and ESI 17 Systems. *Forensic Sci. Int. Genet.* (in press)
4. Ruitberg, C.M., Reeder, D.J., Butler, J.M. (2001) STRBase: a short tandem repeat DNA database for the human identity testing community. *Nucleic Acids Res.* 29: 320-322