VALIDATION OF KINSHIP SOFTWARE FOR APPLICATION OF FORENSIC GENETIC GENEALOGY FOR MISSING PERSONS IDENTIFICATIONS

Bruce Budowle[1,2], V.P. Nagraj[3], Matthew Scholz[3], Shakeel Jessa[3], Carlos Acevedo[3], Jianye Ge[1,2], Jonathan L. King[1], Amy Smuts[1], August E. Woerner[1,2], Stephen D. Turner[3]
[1]Center for Human Identification, University of North Texas Health Science Center
[2]Department of Microbiology, Immunology, and Genetics, University of North Texas Health Science Center
[3]Signature Science

The majority of forensic genetic casework is based on direct comparisons of STR profiles derived from an evidence item(s) and a person of interest(s). Indirect comparisons (i.e. kinship) are conducted when relatives are the reference source to effect identifications (or associations) of, for example, human remains. Indirect comparisons allow for a greater range of reference sample sources than do direct comparisons to potentially identify an individual. However, indirect comparisons have not been used extensively to identify the donor of the DNA on an evidence item (such as a bloodstain or semen stain), primarily because of the limited power of association with the current core CODIS STR loci. Generally, only first degree relatives are compared (extended beyond to some degree with lineage marker systems). However, massively parallel sequencing (MPS) and genotyping array technology make it feasible to assay hundreds of thousands to millions of SNPs simultaneously, thereby increasing the kinship distance between reference and target samples that can be inferred. The overall approach of dense SNP analysis and subsequent public record searching to identify the source of an evidence item(s) has been termed forensic (or investigative) genetic genealogy (FGG). While there is substantial support that the molecular biology and kinship estimation and relationship inference algorithms are relatively robust on high-quality reference samples, the genetic analyses and interpretation processes of FGG have not been validated according to forensic requirements, especially with lower-quality samples typically encountered in forensic analyses. The Department of Justice issued an interim policy for use of FGG which requires, among other things, that "STR DNA typing must be performed, and the suspect's STR DNA profile must be directly compared to the forensic profile previously uploaded to CODIS." However, there are samples where the quantity of DNA is too low or too degraded (such as hairs) to yield a CODIS eligible STR profile. Yet, MPS may be able to obtain useful SNP data to enable direct or indirect comparisons. Therefore, for the genetic analyses portion of FGG to stand on its own, the methodology should be fully validated. In this regard, we assessed a selection of genome-wide relatedness measures (e.g. KING) and IBD segment-based approaches (e.g. IBIS and Hap-IBD) for their ability to detect relationships up to third degree relatives and for their sensitivity to low call rates and genotyping error with forensic-type samples. Genome-wide relatedness measures have much lower data and resource requirements and may be more suitable for close (1-3 degree) relationships, while IBD segment approaches are required for inferring more distant relationships. The former class of methods performs well with a smaller set of SNPs (as might be encountered with degraded DNA samples) in linkage equilibrium, and the approaches do not require any upstream preprocessing such as phasing and imputation. In contrast, IBD segment methods use SNPs to detect shared IBD segments between a pair of individuals and may require phasing and imputation to fill in missing genotypes. Using simulated genome-wide SNP data in individuals where the true underlying relationships were known, we assessed the performance of these different classes of methods for their robustness to challenges typically associated with forensic samples. For closer relationships, the genome- wide relatedness method, KING, performs as well as IBD segment methods, while IBD segment methods are much more sensitive to genotyping error, where accuracy degrades quickly with increasing error. These validation studies provide support for using KING for close (first to third degree) relationship inference for missing and unidentified person applications.