PROBABILISTIC GENOTYPING AND WHOLE GENOME KINSHIP ESTIMATION

August E. Woerner[1,2], Sammed Mandape-[1,2], Benjamin Crysup [1,2], Bruce Budowle[1,2]
[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

Investigative genetic genealogy (IGG) is an emerging field in forensic genomics that holds substantial promise. IGG often is concerned with identifying close relationships (≤ 3rd degree), potentially in low quality or quantity samples. A number of the methods to estimate relatedness operate on sample genotypes and have a limited tolerance for error. Accurate genotyping, however, may be an unreasonable assumption in some IGG contexts. Further, inaccurate genotyping tends to inflate estimates of kinship for most relationship-types. There is a class of bioinformatics tools, however, that were originally designed to work on low template (coverage) samples that may be applicable to IGG. One such tool is NGSrelate. NGSrelate considers the most likely genotype at a site (as most tools otherwise would) as well as the likelihood of the other genotypes given the data. For example, if only a few reads are apparent at a site, while one may not be able to say with certainty what the true genotype is, with NGSrelate relationships can be estimated by considering the likelihoods of being heterozygous or homozygous. In this study, NGSrelate was used to estimate the kinship coefficients in relatives up to 3rd degree using both empirical and simulated whole genome sequence data from a well-described pedigree. The distribution of kinship coefficients was inferred in a variety of relationship contexts and a corresponding bioinformatics pipeline was developed to take and process low-template sequence data for relationship estimation.