

## CHARACTERISTICS OF DNA PROFILE DATABASES IN THE PRESENCE OF RELATIVES

**Hee Lee<sup>1</sup>, Juck-Joon Hwang<sup>2</sup> and Ranajit Chakraborty<sup>1</sup>**

<sup>1</sup>*Center for Genome Information, Department of Environmental Health, University of Cincinnati College of Medicine, Cincinnati, OH*

<sup>2</sup>*Department of Legal Medicine, College of Medicine, Korea University, Seoul, South Korea*



Quantitative assessment of strength of DNA evidence in caseworks of DNA forensics relies on several statistical properties of population databases of DNA profiles. For example, in addition to extracting allele frequencies from the database, assumptions of allelic independence within as well as across loci are needed in evaluating the chance of coincidental matches, and expected frequency of occurrence of a specific profile, conditioned on finding the same profile in other individuals. This presentation provides empirical support of the notion that departures from the assumption of Hardy-Weinberg equilibrium (HWE, i.e., allelic independence within loci), and linkage equilibrium (LE; i.e., independence of alleles across pairs of loci) may be caused in the presence of biological relatives in such databases. More specifically, with data on 155 multigenerational pedigrees consisting of 882 individuals from the South Korean population, typed for 12 STR loci (D3S1358, vWA, FGA, D8S1179, D21S11, D18S51, D2S818, D13S317, D7S820, THO1, TPOX, and CSF1PO), of the CODIS system of forensic loci, we show that allele frequencies at individual loci are not statistically biased with inclusion of relatives in the database. However, when relatives are included in the analyses, 4 of the 12 loci showed significant deviation from HWE, and 58 of the 66 pairwise combinations of loci deviated from the assumption of LE. In contrast, in the dataset consisting of unrelated individuals of these pedigrees, no locus deviated from HWE, and 7 pairs of loci showed deviation from LE mainly contributed by the rare alleles at these loci. Inclusion of relatives shifted the distribution of number of shared alleles between pairs of individuals towards higher values, in comparison with their random expectations. By classifying individuals in increasing order of kinship categories, the observed allele sharing distributions were in accordance with their expectations under the assumptions of HWE and LE within each kinship category. For example, pairs of parent-offspring, full-sibs, and first cousins shared on average 14.8, 15.3, and 10.4 alleles in their 12-locus DNA profiles, while unrelated pairs shared 8.5 alleles at these CODIS loci. All of these are in accordance with their respective expectations under the assumptions of HWE and LE in the population. Analytical results also assert the unbiasedness of allele frequency estimates for these forensic markers in databases containing relatives. In summary, these observations provide an empirical support of the notion that the presence of relatives in a forensic database may cause deviations from HWE and LE, and tendencies of increased allele/genotype sharing, without biasing the allele frequency estimates. Thus, when allele frequencies from such databases are used in conjunction with other conservative approaches of evaluating the strength of DNA evidence, the presence of relatives does not compromise the utility of a forensic database. (Research supported by US Public Health Service Research grants to RC).