DETERMINATION OF FAMILIAL RELATIONSHIPS USING HIGH-THROUGHPUT DNA SEQUENCING OF SINGLE-NUCLEOTIDE POLYMORPHISM (SNP) MARKERS AND CUSTOM SOFTWARE

Martha S. Petrovick, Eric Schwoebel, Darrell O. Ricke, Tara Boettcher, Christina E. Zook, Johanna Bobrow, Anna Shcherbina, James Harper and Edward Wack, MIT Lincoln Laboratory

The familial relationships of 43 members of a large pedigree were assessed using a panel of over 5000 single-nucleotide polymorphism (SNP) markers with low minor-allele frequencies, referred to as the kinship panel. The PCR primers for the kinship panel were designed by Ion Torrent's Ampliseq Designer tool and allowed amplification of all SNP loci in 2 reactions for each sample. The reactions for each sample were combined before preparing them for sequencing, and sequencing was done on Ion Torrent's Proton instrument, 6 samples per run. The kinship panel was refined empirically by removing SNP loci that did not produce an adequate number of sequences, as well as those that consistently produced poor data or ambiguous allele calls (where the ratio of the number of minor allele sequences to total sequences was something other than 0, 0.5 or 1).

Genotypes and relationships were determined using 2 methods. The first is based on the KING algorithm (Manichaikul et al., 2010). This algorithm calculates two metrics, the Kinship Coefficient, KC, and the probability of identity by state equaling zero, P(IBS=0). The KC is a measure of the similarity of two individuals, and the P(IBS=0) is a measure of how many dissimilar alleles are shared between a pair of individuals. A support vector machine was trained on the KING algorithm output using a set of nine features. Feature values were computed for each pair of individuals and fed to a classifier to determine the individuals' relationship. The second analysis method uses basic statistical calculations to estimate the likelihood of each possible relationship between 2 individuals, based on Mendelian inheritance and available allele frequencies (dbSNP and ALFRED). The results from both methods demonstrate good ability to use SNP-panel sequencing to identify related individuals to 3rd degree relationships (great-grandparents/great-grandchildren, great-uncles aunts/great-nephews nieces, first cousins), and a lesser ability to identify 4th degree or higher relationships.

This work is sponsored by the ASD(R&E) under Air Force Contract #FA8721-05-C-0002. Opinions, interpretations, recommendations and conclusions are those of the authors and are not necessarily endorsed by the United States Government.