CRITICAL COMPARISON OF HUMAN MITOCHONDRIAL DNA CONTROL REGION WITH HYPERVARIABLE REGION 1 AND HYPERVARIABLE REGION 2

John Crews and Keith McKenney
George Mason University, Manassas, VA

The forensic and molecular anthropological communities routinely analyze sequence information from hypervariable region 1 (HV1) and hypervariable region 2 (HV2) of human mitochondrial DNA (mtDNA) for human identification and divergence estimates. Our laboratory has expanded the HV1 and HV2 regions to include the entire human mtDNA control region (CR, including HV1 and HV2) to create a new CR sequence database of 3,000 individuals from 14 ethnic populations and subpopulations. This large data set clearly demonstrates that non-HV1/HV2 regions of the CR contain numerous informative sites that can be utilized for increased human identification and sample specificity. Single nucleotide polymorphisms (SNPs) are detected by comparing each CR sequence with the "Anderson" reference sequence (Anderson et al., 1981). It is this collection of SNPs that is used to produce a sequence specific haplotype for each individual.

Data presented here was generated from a sample of 401 complete human CR sequences. Though HV1 and HV2 together comprises approximately half of the CR (624 of the 1155 bases in the CR), they contain 70% (236 of 335) of the polymorphic sites in the CR. The remaining 30% of polymorphic sites lay outside of HV1/HV2 and are beneficial in differentiating between an additional 41% (29 out of 70) of sequences that were identical with respect to HV1/HV2 analysis. Of the 99 polymorphic sites found outside of HV1/HV2 in the 401 CR sequences most were observed infrequently, but a few polymorphic sites were observed in a large number of samples.

To further support the use of the CR over HV1/HV2 it is reported that FST values (which are a measure of population heterogeneity) increase overall when analysis was performed on the entire CR as compared to HV1/HV2 only. Likewise, coancestry coefficients (which are a measure of population inbreeding) decreased when analysis was performed on the entire CR as compared to HV1/HV2 only. Both of these parameters indicate that use of the entire CR in place of HV1/HV2 provides a better estimate of population based genetic diversity.