## STRSEQ: A RESOURCE FOR SEQUENCE-BASED STR ANALYSIS

L.A. Borsuk, K.B. Gettings, P.M. Vallone
National Institute of Standards and Technology, Applied Genetics Group

The STR Sequencing Project (STRSeq) was initiated to facilitate the description of sequence-based alleles at the Short Tandem Repeat (STR) loci targeted in human identification assays. STRSeq data are maintained as GenBank records at the U.S. National Center for Biotechnology Information (NCBI). Each GenBank record contains: observed sequence of an STR region, annotation of the repeat region ("bracketing" consistent with the guidance of the International Society for Forensic Genetics) and flanking region polymorphisms, information regarding the sequencing assay and data quality, and backward compatible length-based allelic designation. STRSeq GenBank records are organized within a BioProject at NCBI (www.ncbi.nlm.nih.gov/bioproject/380127), which is sub-divided by Commonly used autosomal STR Loci, Alternate autosomal STR Loci, Y-chromosomal STR loci, and X-chromosomal STR loci. Each of these categories is further divided into locus-specific projects. The BioProject will initially contain aggregate alleles across 4,612 samples submitted by four laboratories: National Institute of Standards and Technology (NIST, the project organizer), University of North Texas Health Sciences Center, Kings College London, and University of Santiago de Compostela. In addition to providing a framework for communication among laboratories, the ability to search the BioProject can be leveraged as QC for rare sequences encountered in forensic casework. Future plans for this NIJ-funded effort include a pathway for researchers to submit additional alleles and customized interface tools, which would allow users to easily search the STRSeq data set. As of the time of this abstract submission 1145 records have been uploaded into STRSeq. These records represent the evaluation of approximately 25% of the current data set for 28 autosomal STR loci.