B2 Population Analysis and Forensic Utility of Sequence-Based X-Chromosomal Short Tandem Repeat (X-STR) Loci

Lisa Borsuk, MS, National Institute of Standards and Technology, Gaithersburg, MD 20899; Toni M. Diegoli, PhD, Beltsville, MD; Katherine B. Gettings, PhD, National Institute of Standards and Technology, Gaithersburg, MD 20899; Becky Steffen, MS, National Institute of Standards and Technology, Gaithersburg, MD 20899; Kevin Kiesler, MS, Gaithersburg, MD 20899; Peter M. Vallone, PhD, National Institute of Standards and Technology, Gaithersburg, MD 20899-8314*

Learning Overview: After attending this presentation, attendees will learn about sequenced-based variation at seven X-STR loci, representing the four X chromosome linkage groups, sequenced from over one-thousand United States population samples with a commercially available sequencing assay. Attendees will understand how this sequence-based diversity compares to the length-based diversity of X-STR loci.

Impact on the Forensic Science Community: This presentation will impact the forensic community by providing additional information regarding forensically-relevant X-STR loci. Specific impacts include facilitating forensic practitioners' understanding of the performance and informational gains obtained through sequencing X-STR loci. The sequence-based allele and haplotype frequencies in four U.S. population groups will assist with the adoption of sequence-based X-STRs by the forensic practitioner.

X-STR markers are recognized as useful tools to supplement kinship testing in the forensic setting. Numerous studies of allele and haplotype frequencies based on traditional length-based analyses of these loci have been reported in the literature for various population groups. More recently, new technologies capable of providing sequenced-based information with a higher level of marker multiplexing have been investigated for characterization of forensic loci, including X-STRs. The details of sequencing and analysis of seven X-STRs in U.S. populations will be presented.

The National Institute of Standards and Technology (NIST) U.S. Population Sample Set consists of 1,036 unrelated individuals (1,032 male, 4 female) with four population groups represented: African American (*n*=342), Asian (*n*=97), Caucasian (*n*=361), and Hispanic (*n*=236). These samples have been sequenced using the MiSeq FGx Forensic Genomics System, including the ForenSeq DNA Signature Prep Kit, which targets important STR markers commonly used for human identification and relationship testing.[1] Seven X-STR loci are included in this assay: DXS10135, DXS10074, DXS7132, DXS10103, DXS7423, DXS8378, and HPRTB, with at least one marker representing each of the four linkage groups found on the X-chromosome.[2] The core repeat region as well as flanking region variation was assessed with a customized bioinformatic approach. This approach also detected two additional X-STR loci (DXS10148 and DXS8377) which are sequenced with the assay but not reported in the associated Universal Analysis Software (UAS). These two 'extra' loci are being evaluated for potential inclusion in the population set.

Sequence-based allele and haplotype frequencies along with other relevant population genetic parameters for each population group will be presented. Results from this study will be compared to allele calls and frequencies derived from previous analyses using length-based methods.[3] The magnitude of the increase in the number of unique alleles and haplotypes will be presented in the context of potential gains in discriminatory power between the methods. The information provided in this study will serve to facilitate the application of sequence-based methods to X STR profiling in the forensic setting. The sequence data will be made publicly available at NCBI STRSeq X-Chromosomal STR Loci BioProject accession PRJNA380348.[4]

Reference(s):

1. Gettings, K.B.; Borsuk, L.A.; Steffen, C.R.; Kiesler, K.M.; Vallone, P.M.; U.S. Population Sequence Data for 27 Autosomal STR Loci. *Forensic Science International: Genetics.* 2018; https://doi.org/10.1016/j.fsigen.2018.07.013
2. Verogen, Inc. *ForenSeq DNA Signature Prep Reference Guide*. Sept. 2015. Document# 15049528 v01
3. Diegoli, T.M.; Linacre, A.; Vallone, P.M.; Butler, J.M.; Coble, M.D.; Allele frequency distribution of twelve X-chromosomal short tandem repeat markers in four U.S. population groups, *Forensic Science International Genetics Supplement Series 3* (2011) e481-e483.
4. Gettings, K.B.; Borsuk, L.A.; Ballard, D.; Bodner, M.; Budowle, B.; Devesse, L.; King, J.; Parson, W.; Phillips, C.; Vallone, P.M.; STRSeq: A catalog of sequence diversity at human identification Short Tandem Repeat loci. *Forensic Science International: Genetics.* 2017 Nov.;31:111-117; https://doi.org/10.1016/j.fsigen.2017.08.017

X-STR Loci, Sequencing, United States Population