B180 Sequence-Based Analysis of Stutter at Short Tandem Repeat (STR) Loci: Implementation and Utilization

Rachel Aponte, 9815 Bristol Square Lane, Apt 301, Bethesda, MD 20814; Katherine B. Gettings, PhD, NIST, 100 Bureau Drive, MS 8314, Gaithersburg, MD 20899; David L. Duewer, PhD, 100 Bureau Drive, Gaithersburg, MD 20899; Becky Hill, MS, 100 Bureau Drive, MS 8311, Gaithersburg, MD 20899; Michael D. Coble, PhD, 100 Bureau Drive, MS 8312, Gaithersburg, MD 20899-8312; and Peter M. Vallone, PhD, 100 Bureau Drive, Gaithersburg, MD 20899-8311*

After attending this presentation, attendees will understand the properties of Polymerase Chain Reaction (PCR) stutter artifacts observed in data obtained using Next Generation Sequencing (NGS) technologies and Capillary Electrophoresis (CE) methods.

This presentation will impact the forensic science community by laying a foundation for artifact interpretation guidelines, which will be required prior to implementation of NGS technologies. Observing how stutter artifacts characterized by NGS systems compare to those in traditional CE systems will help establish these guidelines. The characterization of stutter with NGS will aid in single-source data interpretation, mixture interpretation, and allow for sequence-based stutter thresholds to be set.

NGS technologies provide the potential for deeper analysis of forensic DNA samples compared to the currently employed length-based CE methods because NGS yields the full DNA sequence of each allele opposed to only the number of repeat units obtained by CE. Access to the full sequence can further individualize a DNA fingerprint by allowing specific base changes or changes within the repeat patterns of more complex loci to be observed. This is particularly useful when multiple alleles at a locus overlap with length-based genotyping but can be distinguished with NGS based on differences within the sequence motif.

NGS systems hold promise for advances in forensic DNA typing, but they will require a new understanding of artifacts and the corresponding interpretation guidelines in order to be beneficial for the forensic community. Stutter artifacts may behave differently with NGS data than what has been observed in CE data due to the differences in the system workflows. Characterization of stutter events between NGS and CE was obtained using the Promega® AutoSeq™ and PowerPlex® Fusion STR multiplex amplification kits, respectively. This study will be expanded to evaluate if stutter events are consistently observed across assays between the instrument platforms.

For the common tetra-nucleotide STR loci, stutter is commonly seen at the n-4 position in both CE and NGS data, but may be observed more frequently at the n+4 and n-8 positions in NGS data. It is possible that the higher frequency of stutter is due to a higher sensitivity of the instrumentation, but it is also possible that these go undetected due to the thresholds set for CE instruments. Lowering the threshold for CE data down to ten Relative Fluorescent Units (RFUs) allows for a better comparison with stutter from NGS systems, which currently have no established thresholds. Comparing stutter between the two systems will provide insight into how the differences of NGS workflows may affect artifacts, and thus, data interpretation. Analyzing each locus and the various types of stutter associated with each allele by sequence will allow for the establishment of allele- and sequence-specific stutter thresholds, which may further benefit current models used for mixture interpretation.

Access to the sequence of each allele also allows for general characterization of where stutter occurs within the repeat motif and whether it is primarily related to the Longest Uninterrupted Stretch (LUS) or to the total number of repeats within an allele.[1] Compound and complex loci — including D2S1338, D3S1358, D8S1179, D19S433, D21S11, FGA, and vWA — have an abundance of alleles with different motifs and would therefore benefit from sequence-based stutter thresholds. Observing the LUS of an allele opposed to observing alleles solely based on the total repeat number may provide a more accurate representation of stutter artifacts for each sequence. For example, results from data analysis of a 14 allele at locus D8S1179 demonstrates differences in stutter ratios when comparing simple and compound repeat motifs. The compound 14 allele with a LUS of 12 exhibited less stutter than the simple motif 14 allele. Therefore, targeting the LUS will allow for even more accurate stutter thresholds to be implemented.

Reference(s):

1. Bright J.A., Stevenson K.E., Coble M.D., Hill C.R., Curran J.M., Buckelton J.S. Characterising the STR locus D6S1043 and examination of its effect on stutter rates. *Forensic Sci Int Genet.* 2014:8(1):20-3

Next Generation Sequencing, Stutter Artifacts, STR Loci

*Presenting Author