B141 An Exploration of a Sequence-Specific Stutter Model for STR Analysis in Next Generation Sequencing

Damani Johnson, Sam Houston State University, Huntsville, TX; Tim Kalafut, Sam Houston State University, Department Forensic Science, Huntsville, TX*

Learning Overview: After attending this presentation, attendees will have learned about a proposed stutter model for Next Generation Sequencing (NGS) based on sequence characteristics rather than locus or allele number. Attendees will hear about the potential for more accurate stutter filters for human interpretation and improved stutter modeling in downstream Probabilistic Genotyping (PG) tools.

Impact Statement: This presentation will impact the forensic science community by introducing a novel stutter model based on the sequence of Short Tandem Repeat (STR) markers, which appears to predict the occurrence of stutter more accurately in complex loci than current models using simple locus thresholds or allele number. This model could be used to improve stutter filtering accuracy in Short Tandem Repeat (STR) analysis and probabilistic genotyping software and ease the adoption and validation of NGS technology in forensic laboratories.

Amplification of STRs using Polymerase Chain Reaction (PCR) is the foundation of current forensic DNA analysis methods throughout the world. While it remains the gold standard, an inevitable drawback of this process is the production of stutter artifacts due to strand slippage during PCR. Stutter appears as smaller peaks typically positioned one or two repeat units away from an allele peak on an electropherogram and can cause issues in interpretation or mixture deconvolution if not accounted for effectively. Current DNA analysis software attempts to recognize the expected occurrence of stutter and allow analysts to remove it from the allelic data. These stutter models have evolved over time from locus-specific thresholds to allele-specific filters based on a linear relationship between stutter peak height ratio and allele number.[1] While this allele-specific stutter modeling offers improvements over the locus-specific thresholds in more accurately predicting the occurrence of stutter in simple loci, loci with more complex repeat units tend to show major deviation from the linear model, suggesting that additional factors may still need to be considered.

The advancement and growing adoption of NGS technology offers exciting new possibilities for stutter modeling. NGS still allows laboratories to assess STR marker length but provides the additional ability to view their sequences, which can reveal differences in STR alleles that would have appeared identical on CE. These differences could potentially be used to develop a more accurate model for stutter filtering based not only on allele, but also on characteristics of the sequence.

This presentation details part of a study that sequenced batches of 96 and 32 single-source samples for autosomal and Y-chromosomal Short Tandem Repeats (Y-STRs) using the QIAGEN® ForenSeq™ MainstAY Kit. These sequence data were used to develop a model for sequence-specific stutter to explore its potential effectiveness in comparison to locus- or allele-based stutter filter model. While preliminary data suggests that additional sequence characteristics may still need to be considered, this presentation will demonstrate stutter ratio plots showing improved R2 measures of linearity at complex loci when considering the Longest Uninterrupted Stretch (LUS) in the repeat unit of the STR sequence rather than allele number.[2] Others have looked at various characteristics of NGS stutter before; however, this stutter model is the first to also incorporate the simplicity of the Sequence Identifier (SID) nomenclature first described by Young et al.[3-5]

These NGS stutter models will ultimately need to be incorporated into analysis and PG software to take full advantage of the potential interpretation and deconvolution improvements. While stutter data from many more samples will be needed, along with validation tools and downstream software, this presentation will serve as an introduction of sequence-specific stutter to the forensic community.

References:

1. Kalafut T, Schuerman C, Sutton J, Faris T, Armogida L, Bright JA, et al. Implementation and validation of an improved allele specific stutter filtering method for electropherogram interpretation. *Forensic Sci Int Genet*. 2018;35:50-6. 10.1016/j.fsigen.2018.03.016.
2. Bright JA, Taylor D, Curran JM, Buckleton JS. Developing allelic and stutter peak height models for a continuous method of DNA interpretation. *Forensic Sci Int Genet*. 2013;7(2):296-304. 10.1016/j.fsigen.2012.11.013.
3. D'Angelo O, Vandepoele ACW, Adelman J, Marciano MA. Assessing non-LUS stutter in DNA sequence data. *Forensic Sci Int Genet*. 2022;59:102706. 10.1016/j.fsigen.2022.102706.
4. Agudo MM, Aanes H, Roseth A, Albert M, Gill P, Bleka O. A comprehensive characterization of MPS-STR stutter artefacts. *Forensic Sci Int Genet*. 2022;60:102728. 10.1016/j.fsigen.2022.102728.
5. Young B, Faris T, Armogida L, Meles V. The sequence identifier (SID): Nomenclature for mixed-DNA casework using MPS. *Forensic Science International: Genetics Supplement Series*. 2019;7. 10.1016/j.fsigss.2019.10.125.

Next Generation Sequencing; Modeling; STRs

Copyright 2024 by the AAFS. Permission to reprint, publish, or otherwise reproduce such material in any form other than photocopying must be obtained by the AAFS.