B129 The Implementation of Biological Models for the Probabilistic Interpretation of Next Generation Sequencing (NGS) Autosomal Short Tandem Repeat (aSTR) Mixtures

Kevin Cheng, Institute of Environmental Science and Research, Sandringham, Auckland 1025, NEW ZEALAND; Meng-Han Lin, PhD, Institute of Environmental Science and Research, Auckland 1025, NEW ZEALAND; Jo-Anne Bright, PhD, Institute of Environmental Science and Research, Auckland 1025, NEW ZEALAND; James M. Curran, PhD, University of Auckland, Auckland 1142, NEW ZEALAND; John S. Buckleton, PhD, Auckland, NEW ZEALAND*

Learning Overview: The goal of this presentation is to provide an overview of the models used in the probabilistic genotyping of NGS DNA profiles and show the results of mixture deconvolution using these models.

Impact on the Forensic Science Community: This presentation will impact the forensic science community by describing continuing efforts in modeling of NGS profiles.

Forensic casework examination of biological samples often produces low-level, mixed, and complex DNA profiles. The process in which forensic scientists analyze the DNA evidence typically involves the amplification of Short Tandemly Repeated (STRs) lengths of DNA using a polymerase Chain Reaction (PCR) technique. Most commonly, the amplified products are separated by allele length using Capillary Electrophoresis (CE) techniques, and the resulting signals are visualized as an electropherogram (epg), showing peak heights for observed alleles.

Recently laboratories are adopting NGS, also known as Massive Parallel Sequencing (MPS), for DNA profiling. One of the advantages of applying NGS technology to forensic DNA typing is that the amplified STR products can be separated by the allele sequence rather than allele size. Sequenced-based information can increase discriminatory power of STRs. This discriminatory power has been demonstrated by showing that the Random Match Probability (RMP) was on average 700 times lower when considering sequence-based information, compared with using length-based allele designation,[1] and hence the single-source Likelihood Ratio (LR) was, on average, 700 times higher.[1] Furthermore, many more informative markers may be included in the PCR reactions including Y-chromosomal Short Tandem Repeats (Y-STRs), X-chromosomal Short Tandem Repeats (X-STRs), Single Nucleotide Polymorphisms (SNPs), and phenotypic markers to aid investigation of no-suspect or human-identification cases.[2]

Despite the appeal of sequencing technologies, there are few published models for the interpretation of autosomal STRs in NGS-DNA mixtures using probabilistic genotyping methods. At the time of writing, there is only one probabilistic genotyping software that has been modified to assist with the interpretation of NGS DNA profiles, Whereas there are a number of probabilistic genotyping software implementations used to assist with the interpretation of CE-DNA mixtures.[3-11]

While both CE and NGS methods use PCR techniques to amplify STRs, there are differences in the methodology that leads to differences in the characteristics of the observed profile such as stutter rates, DNA degradation, and locus specific amplification effects. There are a number of studies that have investigated these characteristics.[12-19] The studies help the community better understand how continuous models used in the interpretation of CE-DNA profiles may also be used for the interpretation of NGS DNA profiles.

This presentation describes the adaptation of the allele peak height model introduced by Bright et al. to develop an allele read count model for the continuous interpretation of autosomal STRs in NGS DNA profiles.[20] Bright et al.'s model consists of three key parameters: template, locus-specific amplification efficiencies, and degradation. In this adapted model for allele read counts, it considers how each of these parameters applies to NGS DNA profiles and shows the improved modeling of allele read counts using informed locus-specific amplification efficiency priors in a Markov chain Monte Carlo method.

This allele read count model is implemented in conjunction with the stutter modeling presented in Vilsen et al. and Cheng et al. into a probabilistic genotyping approach, with the goal of showing some results of the continuous interpretation of NGS DNA mixtures.[13,18]

Reference(s):

1. Devesse L., Davenport L., Borsuk L., Gettings K., Mason-Buck G., Vallone P.M., et al. Classification of STR allelic variation using massively parallel sequencing and assessment of flanking region power. *Forensic Science International: Genetics*. 2020; 48.
2. Greytak E.M., Moore C., Armentrout S.L. Genetic genealogy for cold case and active investigations. *Forensic Science International*. 2019; 299 103-13.
3. Guo F., Yu J., Zhang L., Li J. Massively parallel sequencing of forensic STRs and SNPs using the Illumina® ForenSeq™ DNA Signature Prep Kit on the MiSeq FGx™ Forensic Genomics System. *Forensic Science International: Genetics*. 2017; 31 135-48.
4. Vilsen S.B., Tvedebrink T., Eriksen P.S., Bøsting C., Hussing C., Mogensen H.S., et al. Stutter analysis of complex STR MPS data. *Forensic Science International: Genetics*. 2018; 35 107-12.
5. Vilsen S. *Statistical modelling of Massively Parallel Sequencing data in forensic genetics*. Aalborg University Publishers: Aalborg University, 2018.
6. Hussing C., Huber C., Bytyci R., Mogensen H.S., Morling N., Børsting C. Sequencing of 231 forensic genetic markers using the MiSeq FGx™ forensic genomics system—An evaluation of the assay and software. *Forensic Sciences Research*. 2018; 3(2): 111-23.
7. Didier M., Welti S., Skillman J., Hickey S., Smith J., Peck M., et al. Establishing STR and identity SNP analysis thresholds for reliable interpretation and practical implementation of MPS gDNA casework. *Forensic Science International: Genetics Supplement Series*. 2019; 7(1): 363-64.

8. Li R., Wu R., Li H., Zhang Y., Peng D., Wang N., et al. Characterizing stutter variants in forensic STRs with massively parallel sequencing. *Forensic Science International: Genetics*. 2020; 45 102225.
9. Cheng K., Skillman J., Hickey S., Just R., Moreno L., Bright J.A., et al. Variability and additivity of read counts for aSTRs in NGS DNA profiles. *Forensic Sci Int Genet*. 2020; 48 102351.
10. Riman S., Iyer H., Borsuk L.A., Vallone P.M. Understanding the characteristics of sequence-based single-source DNA profiles. *Forensic Science International: Genetics*. 2020; 44 102192.
11. Bleka Ø., Storvik G., Gill P. EuroForMix: An open source software based on a continuous model to evaluate STR DNA profiles from a mixture of contributors with artefacts. *Forensic Science International: Genetics*. 2016; 21 35-44.
12. Manabe S., Morimoto C., Hamano Y., Fujimoto S., Tamaki K. Development and validation of open-source software for DNA mixture interpretation based on a quantitative continuous model. *PLOS ONE*. 2017; 12(11): e0188183.
13. Bright J.A., Cheng K., Kerr Z., McGovern C., Kelly H., Moretti T.R., et al. STRmix collaborative exercise on DNA mixture interpretation. *Forensic Sci Int Genet*. 2019; 40 1-8.
14. Bright J.-A., Taylor D., McGovern C.E., Cooper S., Russell L., Abarno D., et al. Developmental validation of STRmix™, expert software for the interpretation of forensic DNA profiles. *Forensic Science International: Genetics*. 2016; 23 226-39.
15. Perlin M.W., Legler M.M., Spencer C.E., Smith J.L., Allan W.P., Belrose J.L., et al. Validating TrueAllele® DNA mixture interpretation. *Journal of Forensic Sciences*. 2011; 56 1430-47.
16. Hwa H.-L., Wu M.-Y., Chung W.-C., Ko T.-M., Lin C.-P., Yin H.-I., et al. Massively parallel sequencing analysis of nondegraded and degraded DNA mixtures using the ForenSeq™ system in combination with EuroForMix software. *International Journal of Legal Medicine*. 2019; 133(1): 25-37.
17. Just R.S., Le J., Irwin J.A. LUS+: Extension of the LUS designator concept to differentiate most sequence alleles for 27 STR loci. *Forensic Science International: Reports*. 2020; 2 100059.
18. Bleka Ø., Just R., Le J., Gill P. An examination of STR nomenclatures, filters and models for MPS mixture interpretation. *Forensic Science International: Genetics*. 2020; 48 102319.
19. Just R.S., Irwin J.A. Use of the LUS in sequence allele designations to facilitate probabilistic genotyping of NGS-based STR typing results. *Forensic Science International: Genetics*. 2018; 34 197-205.
20. Bright J.-A., Taylor D., Curran J.M., Buckleton J.S. Developing allelic and stutter peak height models for a continuous method of DNA interpretation. *Forensic Science International: Genetics*. 2013; 7(2): 296-304.

NGS, Probabilistic Genotyping, Mixture Interpretation

*Presenting Author