H95 Developing Biological Models for the Probabilistic Genotyping of Next Generation Sequencing (NGS) Data

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, 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 explain the need for models that describe the behaviors of NGS DNA profiles and the methods for developing said models for probabilistic genotyping software.

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

Conventional Polymerase Chain Reaction (PCR) -based amplification of Short Tandem Repeats (STRs) at different discriminatory loci, and subsequent fragment characterization using Capillary Electrophoresis (CE), is currently the predominant DNA profiling approach used in forensic laboratories. Early methods for the interpretation of the resulting eletropherogram (epg) profiles were very limited, utilizing a binary approach. This binary interpretation was largely threshold-based and did not account for the probability of drop-in and drop-out of alleles in a DNA profile.[1]

As technology developed, DNA profiles increased in complexity with more sensitive DNA profiling methods, resulting in more mixed DNA profiles. This resulted in the need for probabilistic genotyping methods. This is where additional information, such as peak heights and stutter peaks, present in an epg is also considered in the interpretation of a profile.

The analysis of forensic DNA methods using STR and CE methods have been used within forensic DNA laboratories internationally for more than 20 years.[1] More recently, laboratories have started investigating NGS methods for analysis of forensic samples. There are known limitations with a CE-based approach that NGS methods may be able to resolve.[2]

An example of the benefits of an NGS approach is its capability of resolving iso-alleles. These are alleles with the same number of STRs, but with different DNA sequences.[3] Using the sequence information within these iso-alleles, different repeat patterns of the same repeat motif(s) can be observed, allowing for increased discrimination of profiles and possibly improved resolution of contributors.[4]

Analogous to how interpretation methods were limited during the early implementation of CE-based technology, the current methods for the interpretation of DNA profiles obtained through an NGS approach are limited. There have been examples in which existing probabilistic genotyping software, which is developed for the interpretation of CE DNA profiles, is modified to interpret NGS DNA profiles.[5] The assumption is that the models that exist within these probabilistic genotyping software are suited for the interpretation of NGS. There is an increasing need for the development of more sophisticated biological models for use within probabilistic genotyping software for NGS DNA profiles.

In order to develop biological models for the interpretation of NGS DNA profiles, it is first necessary to understand how these profiles behave. This presentation explores the behavior of NGS DNA profiles, including aspects such as stutter, peak height variability, and locus-specific PCR efficiency. The goal is to identify the similarities and differences between CE and NGS DNA profiles. Experiments designed to determine these models and findings are presented.

Reference(s):

1. Coble, M.D. and J.A. Bright. Probabilistic genotyping software: An overview. *Forensic Science International: Genetics*, 2019. 38: p. 219-224.
2. de Knijff, P. From next generation sequencing to now generation sequencing in forensics. *Forensic Science International: Genetics*, 2019. 38: p. 175-180.
3. Warshauer, D.H., J.L. King, and B. Budowle. STRait Razor v2.0: The improved STR Allele Identification Tool—Razor. *Forensic Science International: Genetics*, 2015. 14: p. 182-186.
4. Young, B.A. et al. Estimating number of contributors in massively parallel sequencing data of STR loci. *Forensic Science International: Genetics*, 2019. 38: p. 15-22.
5. Hwa, H.L. et al. Massively parallel sequencing analysis of nondegraded and degraded DNA mixtures using the ForenSeq™ system in combination with Euro *ForMix software. International Journal of Legal Medicine*, 2019. 133(1): p. 25-37.

NGS, Probabilistic Genotyping, DNA

*Presenting Author