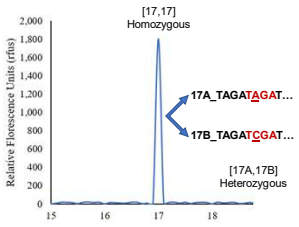
Developmental Validation of a Probabilistic Genotyping Software for NGS-Generated aSTR DNA Profiles

Kevin Cheng[1], Jo-Anne Bright[1], Hannah Kelly[1], Yao-Yuan Liu[1], Meng-Han Lin[1], Maarten Kruijver[1], Duncan Taylor[2] and John Buckleton[1]

[1] Institute of Environmental Science and Research
[2] Forensic Science South Australia

Forensic DNA Mixture Interpretation

- STR and CE methodologies have been used in forensic laboratories for over 30 years.
- Sequencing technologies provide potential advantages.
  - Such as the sequence-level information & its application in other areas of forensic DNA analysis.
- One of the advantages is the resolution of iso-alleles.



Objectives

Current interpretation methods for sequenced DNA profiles are limited. Probabilistic genotyping interpretation methods must be developed to facilitate the uptake of sequencing technologies.

Aims – **Implement published models** into a PG software
– **Developmental validation** of the probabilistic genotyping solution following recommendations from SWGDAM, ISFG, and UK FSR

Sensitivity and Specificity Data

Dataset	Assay	Analysis Software
A	ForenSeq DNA Signature Prep Kit, DNA Primer Mix A & B	UAS v1.3
B*	ForenSeq DNA Signature Prep Kit, DPMB	STRait Razor v3 (ForenSeq v1.25 config)
C	MainstAY	UAS v1.3

*Gettings, K. (2020), Forensic DNA Open Dataset https://doi.org/10.18434/M32157

E/S/R Science for Communities

Sensitivity and Specificity Data

A total of 472 profiles were interpreted and compared to a database of known donors and non-donors (*n*>250). Non-donor profiles were simulated using NIST1036 Caucasian allele frequencies.

	Number of Contributors					
Dataset	1	2	3	4	5	Total
A	36	34	13	19	0	102
B*	17	0	47	21	18	103
C	58	198	7	4	0	267

*Gettings, K. (2020), Forensic DNA Open Dataset https://doi.org/10.18434/M32157

E/S/R Science for Communities

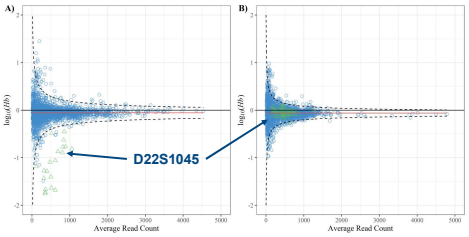
Disclaimer

This developmental validation study uses profiles developed from the DNA Signature Prep Kit and MainstAY.

This does not constitute endorsement, approval, or certification of this particular sequencing technology or kit chemistry. **The probabilistic genotyping solution is designed with the intention to be compatible with other sequencing workflows.**

E/S/R Science for Communities

D22S1045 Heterozygous Balance



DNA Signature Prep Kit MainstAY

Sensitivity and Specificity

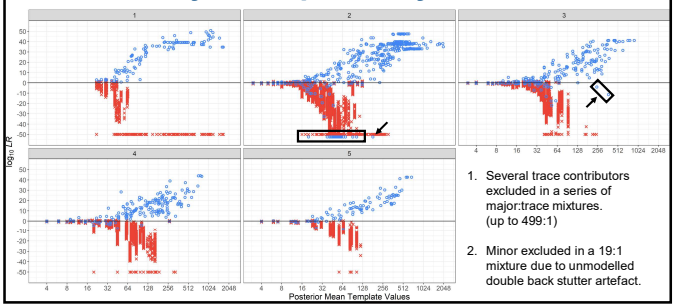
*LR*s were assigned using the **sequence-level information**, the NIST1036 Caucasian allele frequencies, and F_{ST} = 0.01. The propositions considered were:

> ***H_p*:** The DNA originated from the database individual and *N*-1 unknown individuals
>
> ***H_d*:** The DNA originated from *N* unknown individuals

Where *N* is the experimental design number of contributors.

E/S/R
Science for Communities

Sensitivity and Specificity Results



1. Several trace contributors excluded in a series of major:trace mixtures. (up to 499:1)
2. Minor excluded in a 19:1 mixture due to unmodelled double back stutter artefact.

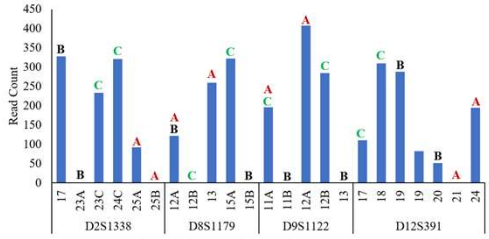
Two-Person Mixture Exclusions

Allele	Sequence	SRv3 Reads	UAS Reads
9.3_A	…[AATG]6 ATG AATG [AATG]2…	9920	10,365
9.3_B	…[AATG]6 ATG AATA [AATG]2…	516	537
9	…[AATG]9…	206	217
8.3	…[AATG]5 ATG [AATG]3…	340	361
7	…[AATG]7…	398	429

- The sequence for the 9.3_B allele has not been previously observed within the published NIST1036 allele frequencies.
- Single base substitution G→A.

E/S/R Science for Communities

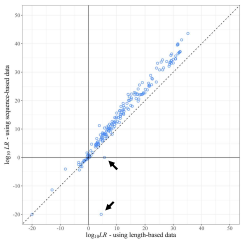
Variable Three-Person Mixture



E/S/R Science for Communities

Allele Length Designation

- The PG solution can also assign *LR*s using the **length-based information**.
- Important for legacy databases and reference samples where sequencing is not an option.
- Demonstrates that additional information in sequencing can increase the *LR*.
- Low-level isoalleles that align with the minor contributor can diffuse the weights, resulting in a lower sequence-based *LR*.



E/S/R Science for Communities

Conclusions

The developmental validation demonstrates that this PG solution is suitable for its intended use for the interpretation of autosomal STRs from forensic profiles generated using NGS technology.

- Developmental validation included testing the accuracy, precision, assigning the length-based *LR*, and alternate propositions.
- NGS profiles have been observed to be generally more variable than CE profiles. This can cause diagnosable false exclusions.

E/S/R Science for Communities

Future Directions

- More research and support for potential casework use
- Improvements in the models and our prototype probabilistic genotyping software
- Improvements in noise modelling
- Expand the work to other assays and other allelotyping software
- Expand mixture interpretation to SNPs

E/S/R Science for Communities

Acknowledgements

THE UNIVERSITY OF AUCKLAND NEW ZEALAND
STATISTICS

NIJ NATIONAL INSTITUTE of JUSTICE
2017-DN-BX-K541
2020-DQ-BX-0022

SBIR
W911-NF-22-C-0002

STRmix. RESOLVE MORE DNA MIXTURES.

Brian Young
Bruce Budowle
Bruce Weir
Daniela Cuenca
James Curran
Jessica Skillman
Jodi Irwin
Judi Morawitz
Kathryn Stephens
Katie Reising
Laurant Baron
Lily Moreno
Melissa Kotkin
Mike Coble
Rachel Oefelein
Rebecca Just
Sanne Aalbers
Swathi Kumar

E/S/R Science for Communities