National Institute of Standards and Technology
U.S. Department of Commerce

A performance-based approach for examining analytical threshold methods used in interpreting CE-based STR profiles

Gordon Research Conference (GRC)
Forensic Analysis of Human DNA
West Dover, VT (June 19 - 24, 2022)

Email: sarah.riman@nist.gov
Phone: (301) 975-4162

Sarah Riman[1], Hari Iyer[1], Sicen Liu[2], Peter M. Vallone[1]

[1]National Institute of Standards and Technology, 100 Bureau Drive, Gaithersburg, MD 20899, USA
[2]Johns Hopkins University Whiting School of Engineering, Department of Computer Science, Baltimore, MD 21218, USA

Funding: NIST Special Programs Office: Forensic Genetics

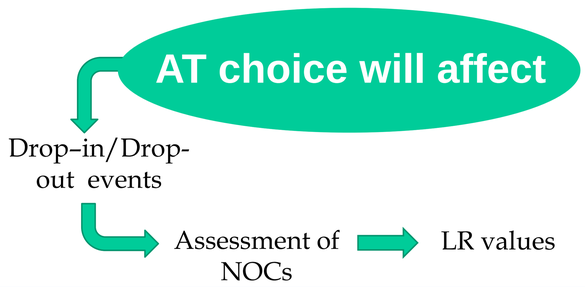
Disclaimer: Points of view in this document are those of the authors and do not necessarily represent the official position or policies of the U.S. Department of Commerce. Certain commercial equipment, instruments, and materials are identified in order to specify experimental procedures as completely as possible. In no case does such identification imply a recommendation or endorsement by NIST, nor does it imply that any of the materials, instruments, or equipment identified are necessarily the best available for the purpose. All work presented has been reviewed and approved by the NIST Human Subjects Protections Office.

Abstract

Continuous probabilistic genotyping software (PGS) model true allelic peaks with their associated PCR byproducts. PGS also model drop-in and drop-out events. Most PGS require: (1) STR profile information with an analytical threshold (AT) pre-applied to remove baseline noise and (2) setting these AT values within the software to model drop-in and drop-out peaks during Likelihood Ratio (LR) assignment. Therefore, setting an AT to help distinguish an allelic signal from electronic noise is necessary until noise models are incorporated into PGS. There are different methods and equations that provide guidance to establish an AT, however these are not normative. Higher AT values will prevent erroneously labeling non-alleles as alleles decreasing drop-in events (false positives). However, a high AT value can lead to drop-out events (false negatives). Therefore, an established AT value should provide a satisfactory balance point between drop-ins and drop-outs. Here, we discuss a performance-based framework to assess different AT methods (AT_{Ms}) and their impact on downstream analysis (i.e. the assignments of the number of contributors (NOCs) and LR values). For illustration, we examine the performance of $AT_{M1}=\mu_{noise} + 10*\sigma_{noise}$ and AT_{M2}=a percentile level based on global drop-out probabilities. Each AT_M was set either globally, or per dye channel, or per locus and was derived using pristine single-source profiles obtained from the GlobalFiler 29 cycles 25s PROVEDIt dataset. The effect of each AT_M on the NOC assessment and LR assignment was determined by applying each AT_M (six in total) to an 'unknown' set of PROVEDIt mixture profiles of varying quality, quantity, experimental NOCs, and mixture ratios. Several quantitative metrics based on LR values were then used to evaluate each AT_M performance.

1 Aim of the study

- ATs are still to be determined until noise models are incorporated into PGS.
- Factors that affect the peak heights of the noise and alleles: amounts of template DNA, fluorescent dye channels, CE injection times, PCR cycle numbers, STR kits, and different models of CE instruments.
- There are different methods and equations that provide guidance to establish AT.
- Here, we conduct a pilot study to illustrate how one might examine the performance of different ATs used in interpreting CE-based STR profiles.



2 Materials and methods

1. Detection of baseline instrument noise

- Raw **single-source** PROVEDIt HID files* (GlobalFiler 29 cycles|3500|25 sec injection time) of varying DNA template amounts (0 – 730 pg) were used for noise detection.
- FaSTR™ software was used to analyze the raw DNA profiles and automatically flag and filter artifacts.
- Known alleles and their respective stutters were removed based on ground truth.
- The remaining peaks were considered as noise (signal not attributed to alleles, stutter types, or artifacts).

2. Setting ATs globally, per-dye channel, and per-locus

- Two methods were used to determine these ATs

$$AT_{M1} = \mu_{noise} + 10 * \sigma_{noise}$$

AT_{M2} = a 99.99th percentile level

The *conservative 99.99th percentile* was selected for this pilot **with the intention of being able to detect and illustrate downstream performance differences** (not that NIST thinks it might be 'best')

- Six different AT_{Ms} were used in this study: Global AT_{M1}, Dye AT_{M1}, Locus AT_{M1}, Global AT_{M2}, Dye AT_{M2}, and Locus AT_{M2}.

3. NOC assessment in FaSTR™ at different AT_{Ms}

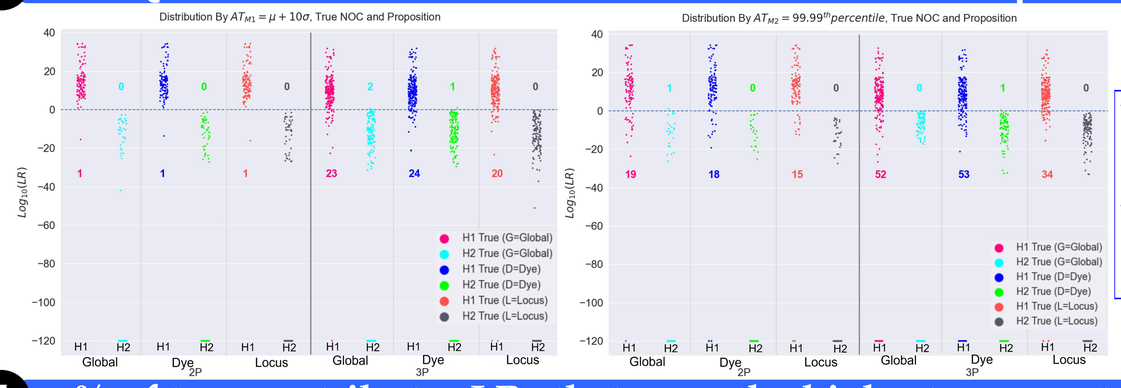
- Raw **mixture** PROVEDIt HID files* of ground-truth NOCs = 2 and 3 and of varying DNA quality, quantity, and mixture ratios were used for NOC assessment.

4. LR calculations and data analysis:

- Sub-source LRs for H1-true and H2-true tests were assigned in STRmix™ v2.9 at each AT setting using the NOC proposed by FaSTR™.

* https://lftdi.camden.rutgers.edu/provedit/files/

3 Qualitative assessment of the discrimination performance



- Numbers below the line at $\log_{10}(LR) = 0$ are known contributors with $\log_{10}(LRs) < 0$.
- Numbers above the line at $\log_{10}(LR) = 0$ are known non-contributors with $\log_{10}(LRs) > 0$.

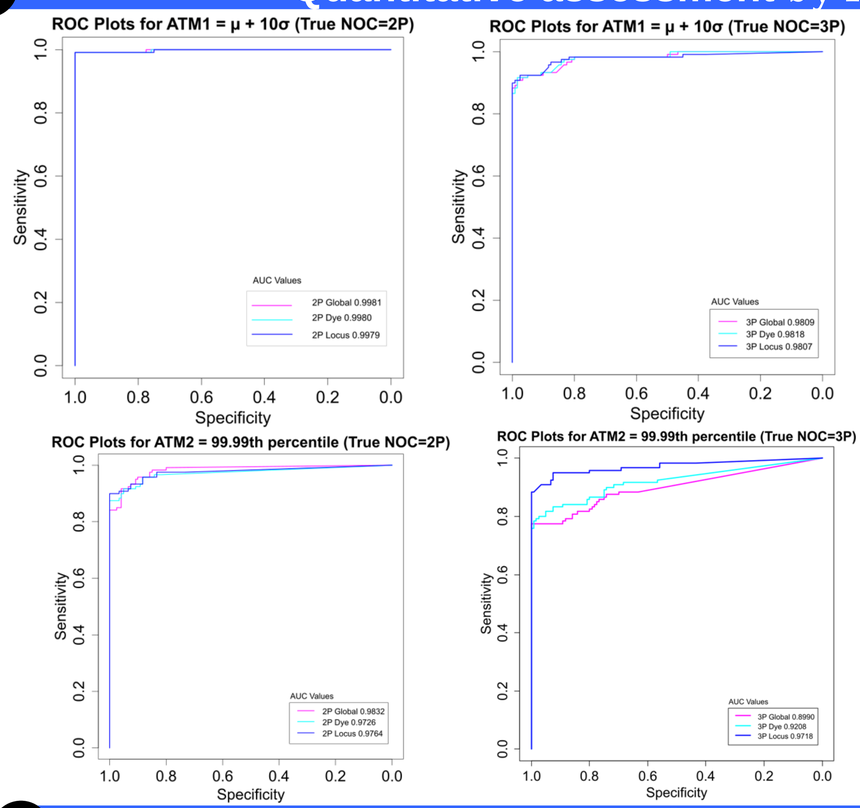
4 % of true contributor LRs that are > the highest non-contributor LRs

	$AT_{M1} = \mu + 10\sigma$			AT_{M2} = 99.99th percentile		
Empirical NOCs	**Global AT_{M1}**	**Dye AT_{M1}**	**Locus AT_{M1}**	**Global AT_{M2}**	**Dye AT_{M2}**	**Locus AT_{M2}**
2P	99.2%	99.2%	99.2%	84.2%	87.5%	90%
3P	89.3%	87.1%	91.1%	76.9%	75.6%	89.3%

Using this metric, initial results indicate that:

- Locus AT_{M1} had higher rates than Global and Dye AT_{M1s} for the considered 3P data.
- Locus AT_{M2} appears to be performing better than Global and Dye AT_{M2s} for the sample set considered.

5 Quantitative assessment by ROC plots



Empirical NOCs	$AT_{M1} = \mu + 10\sigma$		
	Global AUC	**Dye AUC**	**Locus AUC**
2P	*0.9981*	*0.9980*	*0.9979*
3P	*0.9809*	*0.9818*	*0.9807*

- Empirical ROCs with their respective AUCs indicate that performance did not greatly vary between Global AT_{M1}, Dye AT_{M1}, and Locus AT_{M1}.

Empirical NOCs	AT_{M2} = 99.99th percentile		
	Global AUC	**Dye AUC**	**Locus AUC**
2P	*0.9832*	*0.9726*	*0.9764*
3P	*0.8990*	*0.9208*	*0.9718*

- Initial results indicate that Locus AT_{M2} appears to have a better discrimination performance than Global and Dye AT_{M2} with 3P data.

6 Conclusions

- This work focuses on a strategy to assess the performance of an AT method relative to the types of samples encountered by labs.
- We use an *LR-based approach* as a metric to evaluate the performance of different AT methods.
- Initial results indicate that AT_{M1} (μ+10σ) shows a higher degree of discrimination than AT_{M2} (99.99th percentile).