

## Clarification of the facts around locus dropping in the FST software developed by the Office of Chief Medical Examiner, New York City

John Buckleton<sup>1,2</sup> and James Curran<sup>1</sup>

### Background

FST is a computer program which assigns an  $LR$  for DNA profiles created by multiplex STR typing. FST was developed for use by the New York City Office of Chief Medical Examiner (OCME), specifically to interpret DNA profiles processed in their laboratory [1]. The FST source code was publicly released in 2017.

FST implements a semi-continuous method of profile interpretation [2], similar to LRMix [3] and Lab Retriever [4, 5]. It was validated to be used for the interpretation of mixed DNA profiles only from two or three contributors.

The FST software has a function that ignores a locus, termed locus dropping, if the sum of the probabilities of the alleles observed in the crime scene at the locus exceeds 0.97 ([6] @ pg 172 ln 10). This rule is applied if this sum is reached in any of the four subpopulations used by FST (Asian, African-American, Caucasian, and Hispanic). This function has been described as a “mistake” and its discovery attributed to code examination<sup>3</sup>. The facts are that this function was deliberately programmed to avoid a problem that was diagnosed late in the development. It was intended to be disclosed. This disclosure was simply inadvertently dropped in editing of the refereed paper<sup>4</sup>. This function does not clearly favour either the prosecution or the defence. It was rediscovered by testing by defence analysts and only subsequently confirmed by code review.

FST applies NRC II [7] recommendation 4.1 to adjust for population structure. This adjusts the homozygote genotype probability upwards by adding a correction for inbreeding. In addition, the probabilities of any rare alleles in the FST database were inflated to  $5/2N$  (where  $N$  is the number of individuals in the database), in an attempt to down-weight the effect of rare alleles on the  $LR$  ([6] @ pg 172 ln 10) It is these actions that can lead to the sum of the allele probabilities exceeding 1.

In testimony, Adele Mitchell [6] outlined that the function was used to exclude the possibility of getting a negative  $LR$  due to the inbreeding adjustment (termed theta below) allowing the genotype probabilities to be greater than one.

*“One of the first few samples that was run generated a negative likelihood ratio, which does not make sense mathematically...if there was a locus that had alleles that added up almost to a 100 percent, when those conservative adjustments (inflating allele frequencies and the theta adjustment) are made, it could bump the frequency above 100 percent, which would be nonsensical in the calculation.”* [6] @ pg 171 ln 13

*“So what had happened in that one sample was there were several common alleles at a locus and a couple of rare alleles that bumped the entire frequency up to 1, and then when we did the second -- the homozygous frequency adjustment, it gave the total genotype frequencies*

---

<sup>1</sup> University of Auckland, Department of Statistics, Auckland, New Zealand

<sup>2</sup> ESR, Auckland, New Zealand

<sup>3</sup> For example at Government Accountability Office Meeting on Forensic Algorithms, January 15-16 at the National Academy of Sciences Building but also in many other situations

<sup>4</sup> Adele Mitchell per comm

(that) go higher than 1. So we needed to avoid getting into that situation in the future.” [6] @ pg 173 ln 7

Mitchell explains that any locus fitting the criterion for removal would be unlikely to have much ability to discriminate between true and false donors [6].

Adams<sup>5</sup> noted that: “During this review, I encountered no notice, either intended or actual, provided to the user of FST that any loci were removed from the likelihood ratio calculation. I found no indication that this behavior is intended during my examination of FST-related publications and the FST Validation materials.” At the time that he wrote this Adams had performed a code review and would have noted a comment in the code announcing this function and giving a very clear signal that it was intended.

## Rediscovery

This function was rediscovered during testing by defense analysts when running some of the validation samples. For one sample described as “pen B” they obtained a different *LR* from the one given in the validation document ([8] @ pg 741). Code review subsequently traced this behavior to the function that drops loci from the calculation ([8] @ pg 742).

Omitting a locus effectively assigns an *LR* of 1 for that locus. Adams reports that the *LR* for the three loci affected should be D3: 0.53, D13: 3.13, D16: 1.33 [9] @ pg 4. The omission is therefore conservative (a lower *LR*) for D13 and D16 and non-conservative for D3.

## Criticism

This behavior has been criticized, in part, because it was neither mentioned in the publication, the validation materials, nor does the software alert the user to this rule being invoked. In addition, no formal exploration of the potential effects of this function had been reported in the scientific literature.

The FST source code was released in 2017 after a motion by ProPublica<sup>6</sup>. ProPublica commented “*This is not an academic debate: errors in forensic science lead to the imprisonment of innocent people*” [10].

Goldthwaite et al. ([11] @ pg 15 right hand column) argued that the function will always favor the prosecution: “*The defense expert discovered an unexpected function that has never been reported, and which favors the prosecutor’s hypothesis.*” We are informed (Richard Torres per comm) that the Adams affidavit was heavily redacted and this is a plausible explanation why they could advance this argument despite loci D13 and D16, but not D3, giving *LR*s which, if these loci were dropped, would result in a conservative (lower) *LR* overall. However Adams does state, in the redacted statement: “I am not aware of any studies conducted on this feature’s impact on casework samples and the calculation of their statistical weights reported by OCME.” This statement of Adams would seem to rebut the Goldthwaite statement.

Torres [12] gives an account that correctly identifies the respective roles of testing and code review “... *Clinton Hughes, (built)... an open source program using the same exact calculations as FST called reQBT. He recruited college interns with math and computer*

---

<sup>5</sup> <https://johnbuckleton.files.wordpress.com/2019/03/adams-exhibit-c.pdf>

<sup>6</sup> <https://www.propublica.org/article/federal-judge-unseals-new-york-crime-labs-software-for-analyzing-dna-evidence>

*science backgrounds ... in many cases, reQBT got the same answers.. reQBT was getting different answers in some cases. ...a code review ... (by) Adams found that FST performed its calculations differently than what OCME ...”*

Matthews et al. state [13] “*...OCME "fixed" the defect by dropping data that triggered the flaw even if that data might have been important to the defense or prosecution. They notified no one when data was dropped in a particular case and also aggressively resisted expert witness review ...”*

## **Effect**

Matthews et al. [13] note: “*By dropping data, the results skew towards inaccuracy – they skew incorrectly towards more exclusionary (below the line) for real contributors and towards inclusionary (above the line) for non-contributors.*” By eye, as judged from their figures, the effect is quite small. This is the normal behaviour of the LR when the data are removed.

Gasston et al. [14] note “*On average, the dropping of a locus is conservative for six-peak loci and non-conservative for five-peak loci. For persons of interest (POI) with rare alleles, the dropping is usually conservative. For POI with common alleles, the dropping of the locus is often non-conservative.*”

These two empirical studies, the latter one [13] peer reviewed, stand in contrast with a comment: “*A secret function ... was present in the software, tending to overestimate the likelihood of guilt.*”<sup>7</sup>

At the time of creation of the FST software there was a widespread misunderstanding that omitting loci was conservative. This view is incorrect but belief in it was maintained by many publications such as the SWGDAM 2010 guidelines. These dropped loci from the CPI calculation gave the practice of locus dropping a verisimilitude.

## **Open source or inspection under NDA**

Comments in code are text intended for human readers and not part of the operative code. They are intended to assist understanding of the code. (Multi-line) Comments in C#, the programming language in which FST is written, appear between two instances of the symbol `///`. A comment, which appears in the code at the head of the locus dropping section, states:

```
/// This function checks for the total frequencies according to races and removes the alleles (sic) from calculation /// if the sum of frequencies are greater than 0.97.///
```

This is an obvious announcement of the function. It would be found trivially by reading the code either under an NDA or as open source. The failure to include this in the publicly available documentation (validation studies and publications) is entirely consistent as an inadvertent omission. It clearly would have been desirable that the software report to the operator on the screen, or in the output, that loci were dropped.

---

<sup>7</sup> [https://www.eff.org/files/2018/07/30/champion\\_article\\_-\\_lacambra\\_forensic\\_software\\_may\\_2018\\_07102018.pdf](https://www.eff.org/files/2018/07/30/champion_article_-_lacambra_forensic_software_may_2018_07102018.pdf)

## Conclusions

The overwhelming evidence is that a small function was included in FST late in development to avoid a downstream problem caused by interventions in allele (minimum allele probability) and genotype probabilities (NRC recommendation 4.1). External alerts to the operator were omitted. These actions were thought to be conservative on average.

This function was rediscovered by testing and confirmed by code review. The commenting in the code is sufficiently explicit that this function could have been found whether reading open source or under NDA.

## Acknowledgements

This work was supported in part by grant 2017-DN-BX-K541 from the US National Institute of Justice. Points of view in this document are those of the authors and do not necessarily represent the official position or policies of their organizations. We gratefully thank Jo-Anne bright for comments that greatly improved this paper.

## References

1. Mitchell, A.A., et al., *Validation of a DNA mixture statistics tool incorporating allelic drop-out and drop-in*. Forensic Science International: Genetics, 2012. **6**(6): p. 749-761.
2. Gill, P., et al., *An investigation of the rigor of interpretation rules for STR's derived from less than 100 pg of DNA*. Forensic Science International, 2000. **112**(1): p. 17-40.
3. Gill, P. and H. Haned, *A new methodological framework to interpret complex DNA profiles using likelihood ratios*. Forensic Science International: Genetics, 2013. **7**(2): p. 251-263.
4. Balding, D.J. and J. Buckleton, *Interpreting low template DNA profiles*. Forensic Science International: Genetics, 2009. **4**(1): p. 1-10.
5. Lohmueller, K. and N. Rudin, *Calculating the weight of evidence in low-template forensic DNA casework*. Journal of Forensic Sciences, 2013. **58**(1): p. 234-259.
6. Mitchell, A., *Testimony Adele Mitchell*, in *HB6HJON3*, McKay, Editor. 2017. p. 170-180.
7. NRC II, *National Research Council Committee on DNA Forensic Science, The Evaluation of Forensic DNA Evidence*. 1996, Washington, D.C.: National Academy Press.
8. Adams, N., *Adams testimony Dean Jones*.
9. Adams, N., *Declaration of Nathaniel Adams*. 2017: New York.
10. Schulz, D.A. and H. Bloch-Wehba, *Memorandum in support of application by Propublica for leave to intervene, lift the protective order and unseal judicial records*. 2017, Media freedom and information access clinic: New York.
11. Goldthwaite, J., C. Hughes, and R. Torres, *Mixing it up: legal challenges to probabilistic genotyping programs for mixture analysis*. The Champion, 2018(May): p. 12-24.
12. Torres, R. *OPEN SOURCE SOFTWARE: APPLICATIONS IN FORENSICS AND THE COURTROOM*. November 29, 2017 Available from: <https://forensicstats.org/blog/2017/11/29/open-source-software-applications-forensics-courtroom/>.

13. Matthews, J.N., et al., *When Trusted Black Boxes Don't Agree: Incentivizing Iterative Improvement and Accountability in Critical Software Systems*. Proceedings of the 2020 AAAI/ACM Conference on AI, Ethics, and Society (AIES '20), February 7–8, 2020, New York, NY, USA. ACM, New York, NY, USA, <https://doi.org/10.1145/3375627.3375807>.
14. Gasston, J., et al., *An examination of aspects of the probabilistic genotyping tool: Forensic Statistical Tool*. WIREs Forensic Science, 2019. **e1362**(10.1002/wfs2.1362).