Incorrect comments relating to STRmix in State of New Jersey v Corey Pickett

February 16, 2021

Summary

In the case of State of New Jersey versus Corey Pickett a ruling was made preventing the admissibility of some of the forensic DNA evidence until release of the source code underlying the probabilistic genotyping software TrueAllele for review by the defense. Supporting this opinion were briefs from six amicus curiae; Drs Heimdahl and Matthews, the Innocence Project, American Civil Liberties Union, Upturn Inc, The Legal Aid Society, and The Association of Criminal Defense Lawyers of New Jersey. A number of these briefs refer to STRmix. Some of this information is incorrect and we take this opportunity to correct the record. In view of the risk of a repeat of these misstatements, we offer the correct information here.

We make no comment on detail in this ruling outside of the direct comments relating to the software STRmix™. The STRmix team supports Corey Pickett's 6th Amendment right to confront the witnesses against him.

STRmix makes its code available through its Defence Access Policy (https://strmix.com/assets/STRmix/STRmix-PDFs/Access-to-STRmix-Software-by-Defence-Legal-teams-April-2020.pdf) under an NDA. This code has been inspected three times. In no case was a miscode detected by independent code review that affected performance although it is now known that some miscodes were present during those occasions causing intermittent and minor changes to the assigned *LR*.

1. No miscodes in STRmix have been identified by independent code review

Incorrect or misleading statement: Heimdahl and Matthews (H&M) state "*Flaws have been discovered in other PG programs including STRmix and Forensic Statistical Tool ("FST") and in much simpler technologies such as breathalyzers. Those flaws which called into question thousands of convictions frequently went undiscovered until the source code was reviewed as part of the judicial process*". …

Response: None of the miscodes in STRmix have been found by code review and none have affected a conviction.

Of the 14 post production miscodes found in STRmix (https://strmix.com/news/summary-of-miscodes/) three were noticed in use (two by users and one by the STRmix team) where unusual results were detected and investigated by the STRmix team. Eleven (three by one user and eight by the STRmix team) were detected by parallel calculation of intermediate results. The three found by a user were during internal validation. The eight by the STRmix team were often associated with developmental validation of a successor version but existed in earlier versions. None of the miscodes have been found by code review. None have been found as part of a judicial process.

This is further discussed in Buckleton, Curran, Taylor, Bright, What can forensic probabilistic genotyping software developers learn from significant non-forensic software failures? WIREs Forensic Science, 2020, https://doi.org/10.1002/wfs2.1398 .

2. No miscodes in STRmix have affected a criminal trial

Incorrect or misleading statements:

H&M state "*In total, at least thirteen "coding faults" have been found in STRmix, TrueAllele's chief competitor. In one notable example, the miscode impacted 60 criminal cases, requiring new likelihood ratios to be issued in 24 cases. David Murray, "Queensland Authorities Confirm 'Miscode Affects DNA Evidence in Criminal Cases"* , *Courier Mail Mar. 20, 2015 9 https://bit.ly/34DBlZy."*

Innocence project states "*Indeed, doubts about TrueAllele's accuracy are plausible—a competing program, STRmix, that performs the same type of analysis as TrueAllele, was found to have coding errors that impacted the data presented in court. David Murray, Queensland Authorities Confirm "Miscode" Affects DNA Evidence in Criminal Cases, Courier Mail (Mar. 20, 2015).*"

Innocence project: "*Genotyping software in particular has also been found error-prone—source code errors impacting the reliability of STRmix, a competitor to TrueAllele, materially altered match statistics in DNA mixture analysis in over sixty cases. David Murray, Queensland Authorities Confirm "Miscode" Affects DNA Evidence in Criminal Cases, Courier Mail (Mar. 20, 2015)."*

Innocence project: "*For example, STRmix was found to have errors that tainted numerous criminal proceedings*."

Upturn states: *"And in 2015, investigators in Australia encountered an error in their use of STRmix, a probabilistic genotyping software program intended to resolve mixed DNA profiles similar to TrueAllele. David Murray, Queensland Authorities Confirm 'Miscode' Affects DNA Evidence in Criminal Cases, The Courier Mail (Mar. 20, 2015). The error produced incorrect results in at least sixty criminal cases, including a high-profile murder case. Ibid."*

Upturn: "*For probabilistic genotyping in particular, STRmix and FST have both been revealed to have outcome-determinative errors*."

American Civil Liberties Union (ACLU) state *"Likewise, when STRMix (another probabilistic DNA tool similar to TrueAllele) was analyzed by independent researchers, they found programming errors that created false results in 60 cases in Queensland, Australia."*

Response: The reference is to a tabloid newspaper. It is disappointing to see the reliance on such an unreliable source in a serious judicial matter. The factual errors in the Courier Mail article were addressed at the time by the STRmix team. A statement is available at https://www.strmix.com/assets/STRmix/STRmix-PDFs/Statement-relating-to-STRmix-miscodes-180316.pdf. A review of over 20,000 cases in South Australia and New Zealand only identified 22 instances of this miscode firing.

In Queensland, the 24 statements were amended with a minor change to the *LR* in some cases. In all instances this change was before the court case was heard. The changes were minor not "material" or "outcome-determinative" and affected only a subset of the 60 cases. For example, Australia round all *LR*s down to 10^{11} (one hundred thousand million or one hundred billion) and hence in some of the 24 reissued statements the *LR* was exactly the same.

The miscode was discovered in use by the laboratory analysts and not investigators or independent researchers, it did not affect the result in all 60 cases, and the reissued *LR* in what we assume was the "high profile murder case" (the subject of the Courier Mail article) was identical. This is even stated in the Courier Mail article itself:

"*Joan Ryther, who was eight weeks pregnant, was raped and murdered while walking to work in May 2013. Mr Bosscher was advised the case had been affected by the STRmix coding error and was sent a replacement witness statement. The DNA likelihood ratios in both the new and original statements appear to be the same*."

3. Comparing non-equivalent scientific methods as equal is misleading.

Incorrect or misleading statement: ACLU brief *"Furthermore, the sample analysis is dictated by the assumptions programmed into the software. This creates the worrisome reality that softwares like STRMix and TrueAllele provide divergent probability calculations from one another—a discrepancy that can mean the difference between exculpation and inculpation. See Commonwealth v. Foley, 38 A.3d 882, 887, 890 (Pa. Super. Ct. 2012) (noting that TrueAllele calculated a match statistic of 189 billion, compared to a competitor's estimate of 13,000—a more than 14-million-fold difference)."*

Response: This is inappropriately deceptive. The comparison was between TrueAllele and CPI (termed the inclusion method in the reference to follow) and not between TrueAllele and STRmix. CPI is a non-computer based technique that precedes probabilistic genotyping and routinely returns lower values for a comparison to a true donor. We quote from the Cybergenetics webpage[1] "*For the 7% minor unknown contributor, the inclusion method produced a genotype having a 13 thousand match score, obligate allele analysis one of 23 million, and TrueAllele computer interpretation a score of 189 billion. The difference was entirely attributable to how well each method preserved the identification information present in the mixture data. TrueAllele's quantitative interpretation used more of the data, while the other methods ignored peak heights and inclusion further ignored the victim's genotype*." CPI is known to be wasteful of information.

4. Access to the STRmix source code is available as per our Defence Access Policy

Incorrect or misleading statements: Defense brief "*An Illinois court has compelled the disclosure of the source code that runs STRMix, another probabilistic genotyping program, pursuant to a similar protective order."*

Defense brief: "*Mr. Adams reviewed STRMix's computer code in 2014 and found errors that impacted the reliability of the results. This review is discussed in State v. Fair, No. 10-1-09274-5 SEA (King County, WA, Superior Court). The review cited in Fair was pursuant to a strongly worded protective order that prevented Adams from disclosing exactly what is included in STRMix's code. Id. However, the Fair court observed Adams "was able to identify potential issues in STRMix's source code that negatively affected the functioning of the software and could not have been learned from any other source." Id.*

---

[1] https://www.cybgen.com/information/presentations/2010/AAFS/Perlin_Three_match_statistics_one_verdict/page.shtml

Response: No compulsion is required. As per our Defence Access Policy, STRmix offers to disclose its code to representatives of the defense under an NDA. In this case, the code was released under a negotiated protective order and not an NDA.

Adams actual testimony in Fair[2] was "*Let me ask you this question and see whether you can answer it without running afoul of your agreement with STRmix. Did you learn things from your review of STRmix's source code that you could not learn from any other source? A. Yes. Q. Those things that you learned from STRmix's source code that you could not learn from any other source, were they important in understanding STRmix's functioning? A. Yes*.."

The defense do correctly quote the ruling in Fair. However, in this instance the ruling is not based on the actual testimony which does not include the statement "*negatively affected the functioning of the software"* The transcript is in the public domain and should have been consulted by defense before offering incorrect opinions to a court.

---

[2] https://johnbuckleton.files.wordpress.com/2018/08/09-21-16-nathaniel-adams-00400075xbe3c8.pdf