

## Abstract 7

---

### DATA REVIEW PRODUCTIVITY FOR DATABANK PROFILES

**Mark W. Perlin, PhD, MD, PhD, Alexander Sinelnikov, PhD, Bill Allan, Matt Legler, Meredith A. Clarke**

*Cybergenetics, Corp., 160 North Craig Street, Suite 210, Pittsburgh, PA 15213, U.S.A.*

---

DNA databanks can serve a vital role in preventing and solving crime. Forensic scientists estimate that thorough use of DNA databanks in the United States could prevent 300 stranger rapes a day [1]. However, in order for society to construct useful and reliable DNA databanks, the STR data must first be reviewed (by man or machine) in order to ensure the accuracy of the genetic profiles.

To build these databanks rapidly and cost-effectively, it would be helpful to understand the determinants of data review productivity. High human review productivity can rapidly populate DNA databanks, whereas low productivity diverts forensic scientists from solving crimes. This study examines the factors that determine data review productivity.

We examined convicted offender STR profile data from 16 different laboratory processes. The STR panels included Cofiler, ProfilerPlus, SGMplus, Identifier, PowerPlex 1 and 2, and PowerPlex 16. The DNA sequencers included the ABI/310, ABI/3100, ABI/3700, ABI/377 and Hitachi FM/BIO. In total, more than 200,000 STR experiments (locus amplifications) were studied across the 16 data sets.

Each data set was reviewed using two different review protocols: dual human review (two people looking at all the data) and expert system review (one person examining only potentially problematic profiles following computer interpretation). For objective data quality assessment, we measured the average signal-to-noise ratio and size precision of each data set. Expert system rule firings were also used to assess data quality. All the data review and analysis was conducted in Cybergenetics TrueAllele® System 2, a databank and reference profile expert system [2, 3].

We found that data quality correlates highly with data review productivity, as measured by human review time. In particular:

- greater data quality (signal-to-noise or size precision) yielded higher productivity
- expert system rule firings were a useful measure of data quality
- some STR panels or sequencers produced higher quality data than others
- data analysis time (prior to interpretation) was highly sensitive to data quality
- data interpretation review time is prolonged by using lower quality data
- human TrueAllele review of all the data doubles the review time (on average), compared with an expert system review that focuses people on just the problematic profiles
- higher quality data further reduces the time that people spend reviewing problematic profiles
- with reasonable quality data, human review time is under 5 minutes for 100 samples

For national DNA databank policy, and for workflow in an individual laboratory, it is important to understand human productivity in data review. Current expert system computer technology enables one person working a few hours a week to review all of a laboratory's databank data. However, this level of productivity requires the generation and use of high quality STR data.

#### References

[1] Wickenheiser R. The business case for using forensic DNA technology to solve and prevent crime. *J. Biolaw & Bus.* 2004;7(3).

[2] Perlín MW. An expert system for scoring DNA database profiles. In: *Promega's Eleventh International Symposium on Human Identification*; 2000; Biloxi, MS.

[3] Perlín MW, Coffman D, Crouse CA, Konotop F, Ban JD. Automated STR data analysis: validation studies. In: *Promega's Twelfth International Symposium on Human Identification*; 2001; Biloxi, MS.