



# A Focused Study on One Major Benefit of Expert System Implementation: The Time Savings

Laura Young<sup>1\*</sup>; Valerie K. Bostwick<sup>1</sup>, MS; Rachel Lovelace<sup>1</sup>; Chris Carney<sup>2</sup>; Thomas Leach<sup>3</sup>, MS; Cleveland Miles<sup>4</sup>, BS; Rhonda K. Roby<sup>5</sup>, PhD, MPH  
<sup>1</sup>Marshall University Forensic Science Center, Huntington, WV 25701; <sup>2</sup>Florida Department of Law Enforcement, Tallahassee, FL;  
<sup>3</sup>New York State Police Forensic Investigation Center, Albany, NY; <sup>4</sup>Georgia Bureau of Investigation, Atlanta, GA; <sup>5</sup>NIJ Technical Director, NEST Project, Fort Worth, TX 76107



## ABSTRACT

Forensic laboratories worldwide have made great advances in forensic DNA testing: technology, chemistry, techniques, instrumentation, robotics, and software. One of the latest advances for high throughput DNA processing is the use of expert systems for the analysis of single source samples. With the adoption of multi-capillary instruments, single amplification kits, and robotics along with the expansion of laws for the collection of convicted offender samples, the bottleneck in sample processing has shifted from laboratory work to the time-consuming technical review of offender sample data. Current laboratory practice without a validated expert system requires two forensic analysts to visually review the data before a DNA profile is entered into the national database. During the review process, analysts must apply a written set of standard operating procedures, which may have many rules and can be difficult to apply consistently.

Expert systems promise to significantly speed the data review process. They apply complex algorithms, automate data analysis, and free analysts to focus on other tasks. The NIJ Expert Systems Testbed (NEST) Project evaluated the ability of three commercially available expert systems—designed specifically for forensic DNA laboratories—to rapidly and accurately review convicted offender single-source DNA samples for the eventual upload into the national DNA database. We evaluated many steps involved from buying the programs to running them. One step not previously reported on is the processing speed and time savings associated with each computer program.

This study focuses on the time savings using those expert systems previously evaluated by the NEST Project Team. It includes timed comparisons between analysts and software to process defined datasets. The procedure for using the expert systems was broken down into multiple steps. We timed several analysts who are very familiar with the software systems in performing each of these steps. Some of the steps evaluated were: 1) the time required to import data; 2) the time to setup the data for analysis; 3) the time to process the dataset by the software; 4) the time to analyze the controls; 5) the time to generate the files for eventual upload into the national DNA database; and, 6) the time to review all flags and rule firings. These times were compared to the overall time it takes for two analysts to manually review the data as the standards require. In summary, using an expert system greatly reduces the analysts' time spent reviewing data. The authors of this poster found that the three programs will help reduce the DNA convicted offender backlog and ensure timely and accurate submission of the data into the national database. Additionally, three laboratories present the time savings and overall changes in their workflows by having a validated expert system in active use.

## INTRODUCTION

One major premise of expert systems used for DNA analysis is that they will save the analyst time in review. But how much time is saved? This important question has not previously been addressed by the NEST Project Team. Each vendor postulates that its software saves time in review as compared to the current method of human review by two analysts. The NEST Project Team calculated the time savings provided by each of the expert systems, and compared those times to human review by two analysts. The human reviews were conducted using GeneMapper® ID v.3.2 (Applied Biosystems, Foster City, CA), and included the time spent for two analysts to manually review all of the data and check the concordance of their results. Implementing an expert system into a laboratory workflow allows exemplary profiles to be uploaded directly into the national DNA database without any human review. The remaining profiles that do not meet the laboratory-defined thresholds contain artifacts or exhibit noisy baseline and are flagged for human review. According to NDIS, any sample that an analyst views must have a second human review to verify the call. Even after incorporating this step, the overall time savings is well-worth the time needed for developmental validation and implementation of an expert system.

Because every software system is unique, direct comparison of each step could not be made. The study was designed to divide the analysis into various steps, and then separate each of those steps into either a human analysis or software analysis category. Due to the varying levels of complexity and design of the software systems, it was decided that a more accurate representation of the time spent in each system was an evaluation of the human analysis time rather than the full processing time. Each expert system is different and offers other features and benefits that may increase the time to process the data. A laboratory should not choose a particular expert system based on the comparison of analysis time between expert systems. Instead, the expert system that best meets the needs of the laboratory should be chosen, since all expert systems in this study significantly reduce the time needed to review data while maintaining accurate allele calls.

Furthermore, time savings results from the Georgia Bureau of Investigation, Florida Department of Law Enforcement, and the New York State Police Forensic Investigation Center are presented. These three labs have developmentally validated GeneMapper® ID, FSS-i<sup>3</sup>™ Expert Systems Software (Promega Corporation, Madison, WI) and TrueAllele® Databank (Cybergenetics, Pittsburgh, PA), respectively, as expert systems in their laboratories. The implementation of these expert systems into their facilities has created changes in their workflows, which will also be addressed.

## METHODS & MATERIALS

- CODIS data were chosen which included samples amplified with the following kits:
  - AmpFLSTR® Identifier®, COfiler®, and Profiler Plus® PCR Amplification Kits (Applied Biosystems)
  - PowerPlex® 16 System (Promega Corporation)
- Analysis was conducted on three expert system software packages:
  - GeneMapper ID® Software v.3.2 (GMID)
  - FSS-i<sup>3</sup>™ Systems Software v.4.1.3 (FSSI<sup>3</sup>) using GeneMapper® ID v.3.2 for height, sizing, and peak detection
  - TrueAllele® Databank v.2.9 (TA)
- Analysis was divided into multiple steps relative to each software program (Table 1)
- Each step was categorized by time spent in human analysis vs. software analysis
- Data were collected by three experienced analysts and times were averaged

## RESULTS

Table 1: The results below show the separate divisions of analysis timed for this study, split into three tables by expert system. Below are legends for each table. In each legend, a brief explanation of each division is given, along with whether it was considered a human analysis step or a software analysis step.

Table 1A: GMID Time for each step (hours:minutes:seconds)

GMID	Import data	Analysis Setup	Software Analysis Time	Analysis of Size Standards	Ladder Check	Human Analysis	CMF Creation
COFILER							
CO Plates 1-3	0:00:14	0:01:05	0:01:05	0:00:37	0:02:27	0:03:41	0:00:17
PROFILER PLUS							
PP Plate 1-3	0:00:14	0:00:57	0:01:13	0:00:18	0:04:27	0:03:20	0:00:15
POWERPLEX 16							
PP16 Plates 1-3	0:00:12	0:00:32	0:01:14	0:00:15	0:04:35	0:05:58	0:00:21
IDENTIFILER							
ID Plates 1-3	0:00:17	0:00:55	0:01:34	0:00:19	0:07:42	0:15:45	0:00:24

NOTE: All times calculated for each analyst were consistent.

Table 1B: FSS-i<sup>3</sup> Time for each step (hours:minutes:seconds)

FSS-i <sup>3</sup>	GMID				FSS-i <sup>3</sup>				
	Import data	Analysis Setup	Software Analysis Time	Analysis of Size Standards	RAW File Creation	Import RAW file	Rule Analysis	Human Analysis	CMF Creation
COFILER									
CO Plates 1-3	0:00:17	0:00:55	0:00:57	0:00:29	0:00:31	0:00:13	0:00:08	0:02:18	0:01:37
PROFILER PLUS									
PP Plate 1-3	0:00:16	0:00:39	0:01:00	0:00:15	0:00:20	0:00:11	0:00:15	0:03:13	0:01:08
POWERPLEX 16									
PP16 Plates 1-3	0:00:14	0:00:27	0:01:07	0:00:13	0:00:19	0:00:15	0:00:07	0:04:30	0:00:54
IDENTIFILER									
ID Plates 1-3	0:00:14	0:00:41	0:01:13	0:00:07	0:00:21	0:00:14	0:00:07	0:07:12	0:00:57

Table 1C: TrueAllele Databank Time for each step (hours:minutes:seconds)

TA	Data Disk Setup/Creation	Image Call	Cap View	Marker Call	Control Check	Allele Call	Allele View	AutoValidate	CMF Creation
CO Plates 1-3	0:00:58	0:00:44	0:01:01	0:00:17	0:00:36	0:04:43	0:02:57	0:01:04	0:01:25
PROFILER PLUS									
PP Plate 1-3	0:00:39	0:00:45	0:01:16	0:00:24	0:01:09	0:07:50	0:03:07		0:00:50
POWERPLEX 16									
PP16 Plates 1-3	0:00:33	0:00:53	0:01:04	0:00:23	0:00:59	0:15:26	0:06:12	-	0:01:18
IDENTIFILER									
ID Plates 1-3	0:00:45	0:00:46	0:01:03	0:00:43	0:03:00	0:15:03	0:07:15	-	0:00:48

Table 2: Percentage of the total number of loci and samples viewed in each expert system

GeneMapper ID		
	Loci Viewed (%)	Samples Viewed (%)
COfiler	3.8	14.9
Profiler Plus	1.9	6.6
PowerPlex 16	2.0	11.8
Identifiler	3.4	33.0
FSS-i <sup>3</sup>		
	Loci Viewed (%)	Samples Viewed (%)
COfiler	2.4	9.7
Profiler Plus	2.4	11.8
PowerPlex 16	1.8	14.2
Identifiler	4.2	28.8
TrueAllele Databank		
	Loci Viewed (%)	Samples Viewed (%)
COfiler	1.7	9.0
Profiler Plus	2.2	14.6
PowerPlex 16	5.0	27.4
Identifiler	5.3	38.9

In this table, the total percentage of loci and samples viewed were calculated. For example, with the COfiler kit in GeneMapper ID, 76 of the 2016 loci were viewed, or approximately 3.8%. If this software was not used as an expert system, all 2016 loci would have to be reviewed by both the first and second analysts. It is apparent that use of any of these three expert systems reduces the number of samples and loci an analyst will review. The quality of the data can greatly influence the percentage of samples that have to be reviewed. According to NDIS, any sample that an analyst views must have a second human review to verify the call. The time required for this step was not included in this study; future studies will be conducted to incorporate the timing of these steps. However, even with the second review of only the flagged samples, there would still be an overall time savings.

Table 3: Expert System analysis time as compared to the current method of analysis: two full reviews in GeneMapper ID.

	Software System	Total Human Analysis Time (hours:minutes:seconds)
Expert System	GeneMapper ID	0:11:55
Expert System	FSS-i <sup>3</sup>	0:06:52
Expert System	TrueAllele Databank	0:10:06
Manual Review	GeneMapper ID	1:08:31

This table presents the time to process a plate amplified with PowerPlex 16 for each software system. Each of the steps that required human interaction (listed in blue font in Table 1) were summed. When manual review is performed, the time for human analysis is considerably more than when using an expert system.

## ACKNOWLEDGEMENTS

The authors of this poster would like to acknowledge the NIJ Expert System Testbed Project Team, especially Amy Christen, Angela Dolph, and Alyse Hale for all of their assistance with this project.

This project was supported by Award No. 2007-MU-BX-K008 awarded by the National Institute of Justice, Office of Justice Programs, U.S. Department of Justice.

The opinions, findings, and conclusions or recommendations expressed in this presentation are those of the authors and do not necessarily reflect those of the Department of Justice.

1. **Import Data:** upload FSA files into GMID
2. **Analysis Setup:** choose sample type, sample export, analysis method, panel, and size standard
3. **Software Analysis Time:** analyzes samples and applies settings
4. **Analysis of Size Standards:** examining all size standards that had a red Sizing Quality (SQ) flag
5. **Ladder Check:** examine ladders at every locus

6. **Human Analysis:** analyze loci with red Genotype Quality (GQ) flag
7. **CMF Creation:** create CMF file for CODIS

1. **Import Data:** upload FSA files into GMID
2. **Analysis Setup:** choose sample type, sample export, analysis method, panel, and size standard
3. **Software Analysis Time:** GMID analyzes samples and applies settings
4. **Analysis of Size Standards:** examine size standards with red Sizing Quality (SQ) flag
5. **RAW File Creation:** sort data by Sample File, Dye, Marker (all ascending), then export table
6. **Import RAW File:** upload exported GMID table into FSS-i<sup>3</sup>

7. **Rule Analysis:** verify ladders, designate alleles and apply rule set
8. **Human Analysis:** analyze loci with rule firings
9. **CMF Creation:** generate output file, create CMF file with export tool

1. **Data Disk Setup/Creation:** select folder for import, select template, designate analysis parameters, designate controls
2. **Image Call:** TA tracks size standard to determine exact pixel location
3. **Cap View:** review size standards
4. **Marker Call:** determines data peak size by interpolation
5. **Control Check:** verify positive controls, negative controls, and ladders
6. **Allele Call:** designates each sample and applies rules and thresholds

Blue Steps = Human Analysis  
Red Steps = Software Analysis

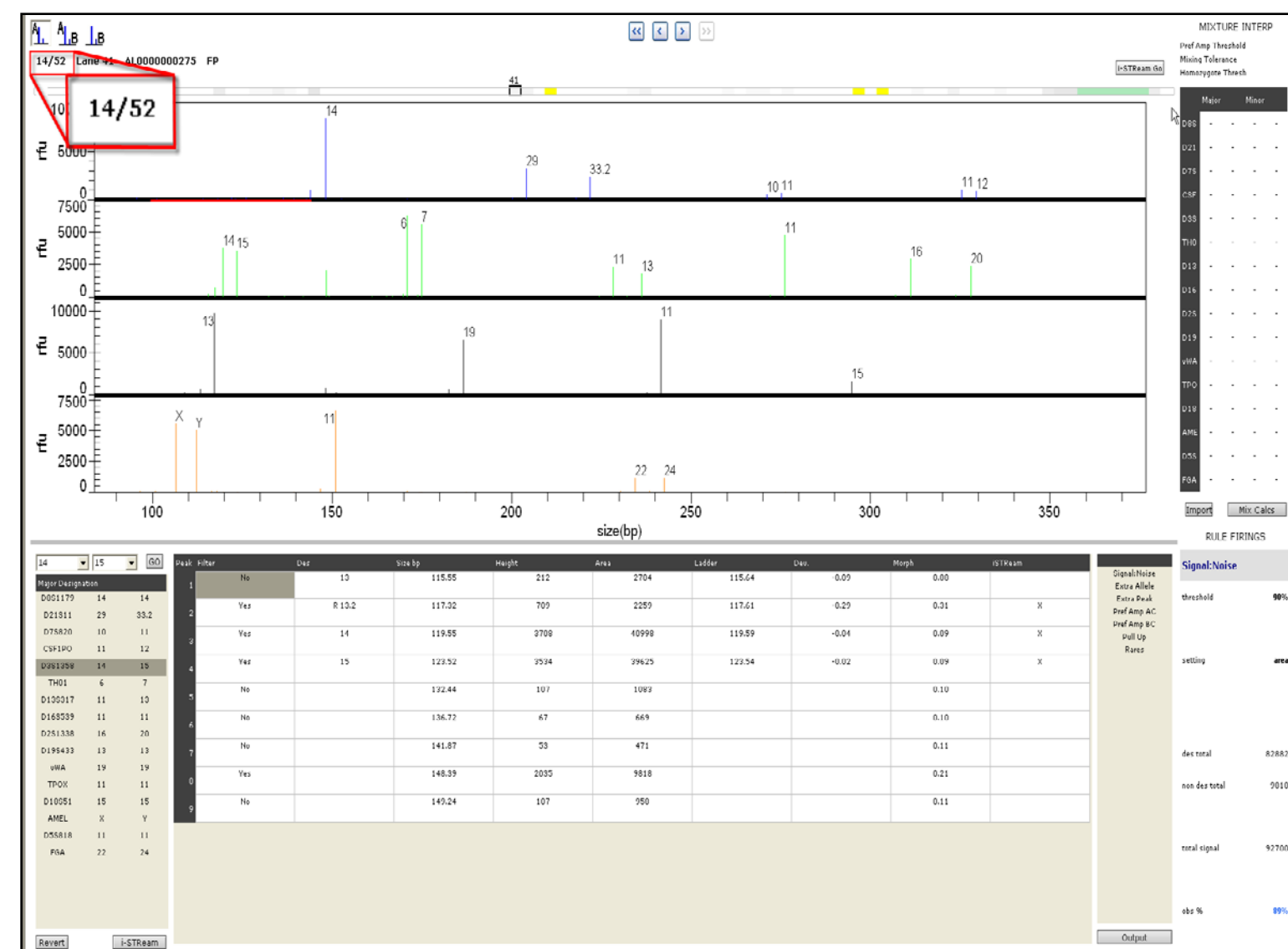


Figure 1: A profile viewed in FSS-i<sup>3</sup> from a plate of Identifier samples. In the red box, the number of loci with rule firings in the entire batch is identified as the number in the denominator. The numerator designates the current locus being evaluated. This number does not indicate the total number of rules fired per locus. In this example, the analyst is viewing the 14<sup>th</sup> locus out of 52 total loci with rule firings. GMID and TA do not offer a similar accounting of flagged loci.

## IMPLEMENTATION OF EXPERT SYSTEMS

**GeneMapper ID Software**

- Implemented GeneMapper ID v. 3.2 as an expert system with ABI PRISM® 3130xl Genetic Analyzer (Applied Biosystems) and Identifier in February 2008;
- Receives approximately 2500 offender samples per month;
- Reduces review time approximately 50%;
- Uses software's PQVs (Process Component-Based Quality Values); and,
- Expedites review of samples.

**GEORGIA BUREAU OF INVESTIGATION**  
DIV. OF FORENSIC SCIENCES  
Atlanta, GA

**FSS-i<sup>3</sup> Expert System Software**

- Validated GeneMapper ID v.3.2 and FSS-i<sup>3</sup> v.4.1.3 as an expert system with an ABI PRISM 3130xl Genetic Analyzer (Data Collection v.3.0) and Identifier in May 2007;
- Not yet implemented for upload to NDIS; and,
- Validating additional instrument platform and expert system software.

**Florida Department of Law Enforcement**  
Tallahassee, FL

**TrueAllele Databank**

- Implemented TrueAllele v.2.9 as an expert system with an ABI PRISM 3100 Genetic Analyzer (Data Collection v1.1) and Profiler Plus and COfiler in July 2007;
- Validated TrueAllele v2.9 as an expert system with an ABI PRISM 3130xl (Data Collection v3.0) and Profiler Plus and COfiler and approved by NDIS in June 2008;
- Reduced backlog by 50% without an increase in overtime and utilizing only two TrueAllele qualified analysts for primary and technical review (see table below);
- First laboratory in the U.S.A. to process samples through an expert system and directly upload into NDIS; and,
- Utilize TrueAllele for technical review of outsourced data (Bode Technology) and in-house data.

**New York State DNA Databank**  
Forensic Investigation Center  
Albany, NY

Total Number of Samples Processed Through TrueAllele v2.9 and Reviewed by Two Qualified Analysts (JUL 07 – JAN 08)

Month	Technical Review of Out-Sourced Data	In-House Review	Total per Month
JUL 07	1,613	0	1,613
AUG 07	2,976	0	2,976
SEPT 07	4,647	3,063	7,710
OCT 07	3,717	6,110	9,827
NOV 07	4,925	6,014	10,939
DEC 07	4,937	4,631	9,568
JAN 08	5,020	6,720	11,740
			TOTAL = 54,373

## CONCLUSIONS

- An expert system for convicted offender data analysis significantly reduces the time spent in data review;
- Each expert system is similar in the time savings it offers;
- Each expert system varies the time spent between human analysis vs. software analysis; and
- The quality of the data can greatly influence the time savings of the expert systems.

## DISCUSSION

The results of this study support the claim that using an expert system will reduce the time spent in technical review of convicted offender samples. The time savings achieved are so significant that implementation of an expert system into a streamlined laboratory will help reduce the backlog of convicted offender samples, and ensure timely and accurate submission of data into the national database. Another important feature of an expert system is the consistency in allele calling and applying rules to the data. Analysts, however, can vary when applying rules or following standard operating procedures. Use of an expert system allows an analyst to be directed to the loci that do not meet pre-defined criteria. The increased analyst efficiency along with the reduced number of samples requiring review prove that expert system use results in a considerable time savings.

## REFERENCES

[1] Roby, R. and Jones, J.P. "Evaluating Expert Systems for Forensic DNA Laboratories," Forensic News, Foster City: Applied Biosystems, 2005.  
<http://www.promega.com/geneticidproc/ussymp16proc/abstracts/robby.pdf>

[2] Roby, R.K. and Christen, A.D. (2007). Validating Expert Systems: Examples with the FSS-i3 Expert Systems Software. *Profiles in DNA* 10(2), 13-15.