

**The author(s) shown below used Federal funds provided by the U.S. Department of Justice and prepared the following final report:**

**Document Title: Automated STR Analysis for DNA Databases**

**Author(s): Mark W. Perlin Ph.D.**

**Document No.: 199069**

**Date Received: January 2003**

**Award Number: 2000-IJ-CX-K005**

**This report has not been published by the U.S. Department of Justice. To provide better customer service, NCJRS has made this Federally-funded grant final report available electronically in addition to traditional paper copies.**

**Opinions or points of view expressed are those of the author(s) and do not necessarily reflect the official position or policies of the U.S. Department of Justice.**

199069

# NIJ Grant Final Report

Mark W. Perlin  
Cybergenetics, Pittsburgh, PA

December 20, 2002

*Grant Number:* 2000-IJ-CX-K005  
*Title:* Automated STR Analysis for DNA Databases  
*Grantee:* Cybergenetics, Pittsburgh, PA  
*PI:* Dr. Mark W. Perlin

Cybergenetics © 2002



**Contact information:**

Mark W. Perlin, PhD, MD, PhD  
Chief Executive Officer  
Cybergenetics  
160 North Craig Street  
Suite 210  
Pittsburgh, PA 15213  
USA  
(412) 683-3004  
(423) 683-3005 FAX  
perlin@cybgen.com

PAL 12/20/02

## Abstract

STR technology has enabled the rapid generation of highly informative DNA data for use in human identification. However, these data must be carefully analyzed. With database samples, there is now an acute shortage of skilled data reviewers. With casework samples (including mixtures), much information is not extracted from the data, despite considerable examiner effort. We are rapidly developing novel computational, mathematical and statistical methods that help overcome these limitations. This report focuses on the collaborative validation of these methods as applied to DNA databases.

Convicted offender DNA databases must be accurate. To minimize error, the original STR data are carefully reviewed by two or more people. Moreover, in a troubleshooting capacity, this review helps to continuously maintain high quality lab data. But there are not enough skilled personnel for this arduous, repetitive task. To alleviate this critical labor shortage, we developed the TrueAllele™ expert system. The computer program automates virtually every human review function, and provides consistent quality assessment and allele designation.

The TrueAllele validation began with the original data from 50,000 CODIS genotypes. System parameters were adapted to the instruments (ABI/310, ABI/3700, Hitachi/FMbio) and panels (ProfilerPlus, Cofiler, PowerPlex 1.2) used to generate the data. Computer processing was then done, with automated scoring of the high quality data, followed by limited human review. The computed expert system results were compared against manually scored results. We report here on the relative accuracy and efficiency of the automated approach.

Our report describes novel computer-based methods for assuring data quality and automating DNA database review. We present here the objective results of our validation study, and demonstrate the feasibility of practical automated analysis. Our primary objective is the rapid introduction of validated intelligent data analysis systems for eliminating tedious human STR analysis. This contribution may help free up valuable DNA examiner time for serving justice through forensic science.

## Table of Contents

Abstract .....	2
Table of Contents .....	3
Introduction .....	4
Automated STR Analysis .....	4
Reviewing STR data .....	4
TrueAllele™ technology .....	5
Automated processing .....	6
Quality assurance .....	6
Rule system .....	7
Multi-platform engine .....	7
Databasing Validation .....	8
Methodology .....	8
Rule settings .....	9
Hitachi/FMBio (PowerPlex) .....	10
ABI/310 (ProfilerPlus & Cofiler) .....	10
ABI/3700 (ProfilerPlus & Cofiler) .....	11
Specific Aims Results .....	12
Conclusion .....	13
Acknowledgements .....	13
References .....	14
Figures .....	15
Tables .....	19

## Introduction

STR technology has enabled the rapid generation of highly informative DNA data for use in human identification (1). However, these data must be carefully analyzed for errors to assure forensic usability (2). With database samples, there is a growing shortage of skilled data reviewers for quality assurance (3). With casework samples (including mixtures), much information is not extracted from the data (4, 5), despite considerable examiner effort. We have developed novel computational, mathematical and statistical methods that help overcome these limitations (6-8). This report focuses on the collaborative validation of these new scientific analysis methods as specifically applied to DNA databases.

We present this report in three sections:

- Automated STR Analysis. Current STR analysis is a labor-intensive effort performed by expert forensic examiners. It would be highly useful to deploy expert computer systems that could automate much of their work. This section describes the TrueAllele™ expert system for automated STR analysis.
- Databasing Validation. We have performed a validation study of the TrueAllele™ automated databasing solution on a variety of DNA sequencing instruments and STR panels. In this section, we outline our methods and present the main validation results.
- Specific Aims. The grant application described three broad specific aims. In this section, we discuss the validation results in the context of these aims.

## Automated STR Analysis

We describe here a computer-based solution to automated STR analysis that provides quality assurance, rule-based diagnostics, and runs on most forensic platforms. We provide validation results for this TrueAllele™ expert system in subsequent sections.

### *Reviewing STR data*

Human review (in duplicate) of forensic STR data is currently required to assure that DNA profiles are correctly scored. However, this human analysis step has become a labor-intensive bottleneck impeding the rapid construction of high-quality DNA databases. Any useful scoring approach must address the following issues:

- Computer automation. It would be desirable to have a computer-based analysis technology that can automate most or all of the laborious human data review process.
- Quality assurance. What is the quality of the underlying STR data, including PCR controls and size calibrations? This data quality determines the reliability of the scored genotypes.

- **Database integrity.** STR databases must be accurate in order to avoid database matching errors. With a false positive match, the wrong person is identified. With a false negative nonmatch (the more likely error), the correct criminal may never be found – even though their (incorrect) profile resides on the database.
- **Casework and mixtures.** DNA databases of convicted offenders are created primarily for the purpose of identifying suspects later on in casework applications. With mixed or degraded DNA samples, the database match can become inherently ambiguous. Computational methods are needed that can increase the specificity of the match.

In this report, we explore all of these issues.

Any useful solution to the STR data review problem must achieve these key goals:

- **No error.** Allele calling errors should be minimized.
- **High throughput.** The process should rapidly review data, eliminating backlogs.
- **Small staff.** Personnel numbers should remain constant as throughput increases.

The TrueAllele™ technology, described next, achieves all these goals.

### ***TrueAllele™ technology***

The TrueAllele™ process eliminates the STR human review bottleneck (7). Since Cybergenetics has innovated this automation technology for the past decade, its multiple patents protect all aspects of automated STR analysis, including forensic usage, quality scores, PCR stutter handling, and component processing steps (9). Note that the validation study described herein applies solely to the Cybergenetics TrueAllele implementation of these patented expert system processes.

The *input* to the process is the original data generated from a gel, automated or capillary DNA sequencer. The fully automated TrueAllele software then processes the STR data, running on most common computers (Macintosh, Windows, UNIX).

Automated processing steps include:

**Color separation.** TrueAllele can dynamically separate the original channels into their component dye colors.

**Image processing.** With gel-based sequencers, the software analyzes each dye plane image in two dimensions.

**Lane tracking.** TrueAllele tracks the size standards to determine the in-lane size calibration. On gels, the program also automatically tracks the lanes.

**Signal analysis.** The software transforms the signals, and identifies peaks.

**Ladder building.** TrueAllele automatically analyzes the allelic ladders, and uses these ladders for DNA length determination.

**Peak quantification.** The technology models every peak, determining the best fit between the set of modeled peaks and the observed signal data. The result is exquisitely accurate estimation of DNA concentration, enabling a highly quantitative allele calling, quality checking and mixture analysis approach.

**Allele designation.** Using its internal database of DNA sizes and concentrations, TrueAllele designates alleles in DNA length coordinates, using multiple allele calling mechanisms.

**Quality checking.** TrueAllele computes dozens of quality measures for each genotype experiment, compiling statistics on the different phases of laboratory data generation. The program detects outliers (based on user-determined thresholds), and can focus the user review on just the problematic data.

**CODIS reporting.** The program automatically generates reports, including CMF files for automated CODIS DNA database submission.

The software *outputs* quality assured STR profiles to DNA databases.

### ***Automated processing***

The TrueAllele process begins with automated *input* of the data. A site uses an "auto setup" template that knows the source computer directory of original sequencer data runs. Combining these data together with annotating information (sample lane layout, marker information, DNA sequencer used, etc.), TrueAllele transforms the sequencing data of multiple runs into an instrument-independent "DataDisk" format suitable for downstream processing.

The second step is automated gel *image* or capillary *signal analysis*. On all the sequencer runs, the computer performs data filtering, baseline and primer removal, color separation, lane and size tracking, and extraction of 1D lane (or capillary) signals from pixel into size coordinates. Once the computer has completed its work, users do quality assurance (review, accept, edit, reject, etc.) in interfaces such as "ImageView" (Figure 1A). With good data, a run's Q/A typically takes under two minutes.

In the *allelic analysis* (step three), the computer begins by analyzing the allelic ladders, transforming the data into DNA length coordinates, and then (expending considerable effort) accurately quantitating every DNA event (peak or band) in the data. On every genotype, allele calling uses multiple algorithms, assigns quality scores, and applies dozens of rule checks when looking for possible data artifacts in each phase of laboratory processing. After the computer has allele called and quality checked all the genotypes across all sequencer runs, it navigates the user's quality assurance review (accept, edit, reject) through the potentially problematic genotypes (Figure 1B), showing its associated rule analysis (Figure 3).

In the final *output* step, TrueAllele automatically exports the data in format suitable for automated database (e.g., CODIS) import.

### ***Quality assurance***

TrueAllele provides a number of data visualizations and reports that are used primarily for quality assurance. For example, showing the lane data for each marker in size

coordinates provides much information about the relative sizing accuracy and the signal quality (Figure 2A). Also, it is useful to review the size standards and allelic ladders in data views (Figure 2B), overlay plots, with the tracking results shown; large size deviations or low signals can pinpoint potential problems.

### ***Rule system***

TrueAllele's automated rule analysis provides detailed quality assurance information at the individual genotype level (Figure 3). Organized by laboratory DNA processing step (extract, amplify, separate), the rules automatically perform a consistent and detailed technical review. Exportable Excel spreadsheets summarize the rule firings, and highlight potential data problems; such information can be very helpful in troubleshooting and for ongoing quality assurance. Applying a lab's standard operating procedures, a site administrator can customize the values of rule thresholds, and determine whether or not to use a rule. TrueAllele's technical documentation gives detailed descriptions and illustrative examples of each rule.

The Hitachi genotype experiment shown (Figure 3) raises interesting quality assurance issues. If data bands have low optical density, but otherwise show every sign of high data quality (no rules fired in this genotype's Rule Analysis window), do these data actually require any human review? In one world view, the user-set rule criteria might say "no," and the data would be automatically accepted. From another perspective, the user's rule criteria might say "yes," with the computer enforcing the human review of all such data. Fortunately, TrueAllele is neutral here: the laboratory has the sole authority (via settable rule parameters) to determine their own data review criteria.

### ***Multi-platform engine***

TrueAllele is designed to work with any data, from any DNA sequencing instrument. Results from the Hitachi/FMBio, ABI/310, ABI/3700, and ABI/377 are described in this report. Shown is a TrueAllele run from 96-well plate ABI/3100 16-capillary data (Figure 4A), and the superb signals from Amersham's 96-capillary MegaBACE sequencer (Figure 4B). We work closely with instrument manufacturers to ensure support for all reasonable platforms.

TrueAllele is independent of any instrument or chemistry manufacturer. It works with and displays the actual data. Manufacturer-provided software can always process data extensively for consistently "beautiful" results. TrueAllele, however, shows original STR data, as they have been recorded. We believe that our "never hide the truth" approach is more useful (to prosecution, defense, and society) in a criminal justice setting.



## Databasing Validation

The key results of this report center on our completed scientific study that validates TrueAllele™ for DNA databasing applications. We describe in this section our validation methodology, the rule parameter settings, and accuracy and timing results for the Hitachi FMBio, ABI 310 and ABI 3700 platforms, using both Promega and ABI STR chemistries.

### *Methodology*

Our validation was a concordance study between automated computer results, and a subsequent detailed human review that scrutinized every scored genotype. In actual production use, of course, one would not review genotypes that TrueAllele had “accepted” without rule firings, since such extra human effort would be redundant and inefficient. However, for validation purposes, all genotypes were manually reviewed in this study. Our extensive developmental validation should permit forensic databasing labs to use the validated TrueAllele process efficiently – without redundant human review of TrueAllele’s “accepted” genotypes.

Obtain the original data. The original data was sent to Cybergenetics on CD from participating laboratories. These STR data included:

- about 2,000 database samples of ABI/310 single capillary data amplified in both ABI ProfilerPlus and Cofiler panels (Coffman, FDLE);
- about 2,000 database samples of ABI/3700 96 capillary data amplified in both ABI ProfilerPlus and Cofiler panels (Coffman, FDLE);
- about 1,000 samples of Hitachi FMBio gel data amplified in the Promega PowerPlex 1.2 panel (Crouse, PBSO); and
- developmental Hitachi FMBio data amplified in Promega STR panels from the Virginia lab and their database vendor (Ban, DFS).

The high quality PBSO Hitachi STR data had been generated for casework applications, rather than for high-throughput database processing. This lane arrangement led to considerable manual set up time to determine the location of the single profile data, preparing the gel data for automated analysis.

Process data in the TrueAllele expert system. In broad overview, automated processing entailed the following steps:

- *auto-setup* The original data for each gel or capillary were put into a DataDisk folder. With ABI capillary instruments, these data are just the individual capillary run files. On the Hitachi platform, these data comprise the raw channel images, along with a proto-layout file which indicates what has been run in each lane. In the AutoSetup process, the computer transforms these raw data formats into a platform-independent representation suitable for downstream TrueAllele processing.
- *process run* Automated TrueAllele processing was done. ImageCall was run on the data (ABI capillary, or Hitachi gel), with the computer automatically

processing the signals, removing the primer peaks, separating colors, tracking the sizes and lanes, and extracting the signals in size (bp) coordinates.

- *quality assurance* For Q/A, a human operator reviewed each run (ImageView for gels; CapView for capillaries), performing any necessary editing, with an option to reject the run. On this data set, minimal review (typically several minutes) was needed for each run.
- *call alleles* AlleleCall was run on the 1D electropherogram signals. The module automatically performed peak quantitation, allele calling and quality checking on the data of every genotype. TrueAllele's accurate peak quantitation enabled the use of quantitative allele calling methods and quality checks.
- *apply rules* Several dozen rules were applied to each genotype experiment. The rule thresholds used are described in the next section.
- *check results* The operator then completed a checklist to ensure that all processing steps were properly performed.

The result of this automated processing was the computer's assignment to each genotype experiment a label of "accept," "edit," or "reject."

3. Review all the data. A key concept in the optimized TrueAllele process is "one person, many computers." Using the Timbuktu remote control software (Netopia, Alameda, CA), one computer can control many others over a local area network. Therefore, each TrueAllele operator had available one computer for display, and multiple monitor-less computers for background processing. This way, a person never had to wait for a computer – they could work continuously on their review of the automated processing.

In the TrueAllele process, the computer accepts, rejects or suggests review (with possible editing) of each genotype. Although, the user usually does not examine any computer accepted results, in this validation study all genotypes were reviewed for concordance. The result of this manual review phase was the user's assignment to each genotype experiment a label of "accept," "edit," or "reject."

4. Generate results and statistics. Comparison of the computer and the user labels of "accept," "edit," or "reject" was the core result of the study. Other data were recorded, such as the total human operator time expended on a subset of the runs. The computer generated useful statistics, such as rule firing summaries for every gel or marker; these results helped in assessing the quality of the data.

### ***Rule settings***

TrueAllele's rule thresholds are set by the user, preferably using Cybergenetics optimization protocols. We tuned the parameters on data samples different from those used in the study. Settings for the Hitachi gel data and the ABI capillary are shown (Table 1).

After gaining more experience with the Hitachi platform, we determined that 15,000 (rather than 9,000) was a better value for the High Signal rule. Also, the Peak Height

Cutoff baseline should have been set to 200 (not 50). We did not change these threshold values during the course of the study. However, we note throughout the paper how using these proper values would have affected important results.

### ***Hitachi/FMBio (PowerPlex)***

On the Hitachi FMBio platform, using Promega's PowerPlex 1.2 STR panel, we had TrueAllele process 7,973 PBSO genotypes. The computer performed all gel and allele automated processing. The computer decided that no review was needed for about 80% of these genotypes. Subsequent human review disclosed that all these designations were correct. These results suggest that TrueAllele can eliminate human review of most STR DNA databasing gel data.

The breakdown of the computer vs. human review comparison is shown (Table 2). The first row is for the data that the computer accepts as not requiring any human review; this row provides the most important measures of effectiveness.

- The row's first column gives the *efficiency factor* (72.5%) of the proportion of STR data that is safely scored automatically without requiring human review. That is, both computer "accept" and user "accept" decisions. In fact, the true efficiency was appreciably greater (82.0%) after reprocessing the data with properly tuned Hitachi threshold rule and baseline settings, as indicated above.
- The next two columns measure *accuracy*, specifically, the false negative error rate (0.0%). This critical measure describes how many genotypes might be incorrectly called (computer "accept"), but would actually require human review (user "edit" or "reject"). We want this number to equal (or be very near) zero.

The clear conclusion is that TrueAllele provides both efficient and accurate automated review of STR data for DNA databases on an Hitachi/Promega platform. (Because we used Hitachi data for nonautomation casework, with most human time spent sorting through data rather than actually processing it, there are no useful timings to report.)

### ***ABI/310 (ProfilerPlus & Cofiler)***

On the ABI 310 platform, using ABI's ProfilerPlus and Cofiler STR panels, we had TrueAllele process 23,723 FDLE genotypes. The computer performed all capillary and allele automated processing. The computer determined that no review was needed for about 85% of these genotypes. Subsequent human review disclosed that all the properly set up designations were correct. These results suggest that TrueAllele can eliminate human review of most STR DNA databasing capillary data.

The breakdown of the computer vs. human review comparison is shown (Table 3). We look closely at the first row for computer accepted data, i.e., results that should not entail any human review.

- An *efficiency factor* (computer "accept," user "accept") shows that a high proportion (86.4%) of the STR data was safely scored automatically without requiring

human review. Another useful efficiency measure is the sum of the second row entries (computer "edit") that the computer asks the user to edit (13.3%).

- The *accuracy* shown in the last two columns of the first row (computer "accept," user "edit" or "reject") gave a zero false negative error rate (0.0%). In fact, because of incorrect STR sizing windows (based on manufacturer guidelines), there were several missized alleles. However, after properly adjusting the sizing windows to actual data, all such computer errors were eliminated.

The human processing time for AlleleView editing (95% of the total effort) was recorded for 43% of the data (10,208 genotypes) as 11.25 hours, or 675 minutes. Reducing this time by the efficiency gained from reviewing only computer "edit" data (13.3%, from the second row), the effective review time was only 90 minutes. Dividing the number (10,208 genotypes) by the time spent (90 minutes) yields an average human review time of 113 genotypes per minute, or *almost 2 reviewed genotypes per second*. This rate far exceeds the data generation capability of current sequencer technology. For example, one would need to run a 16-plex panel over 4 sequencers each having 96 lanes or capillaries every hour to generate 100 genotypes per minute. This data generation rate would be on par with the scoring rate of one TrueAllele-enhanced full-time equivalent (FTE) data reviewer.

When people never wait for computers (i.e., "one person, many computers"), the human operator review time was about two genotypes per second. Properly configured, automatically applying 22 data artifact detection rules to every genotype, no false negative scoring errors were observed in ~24,000 genotypes. That is, TrueAllele provided a thorough, diligent and objective automated process that was accurate and entailed relatively little human effort. These results suggest that the expert system may be useful for automated review of STR data for DNA databases on ABI capillary platforms.

### **ABI/3700 (ProfilerPlus & Cofiler)**

On the ABI 3700 platform, using ABI's ProfilerPlus and Cofiler STR panels, we applied TrueAllele to 17,014 FDLE genotypes. The computer performed all capillary and allele automated processing. The computer decided that no review was needed for about 85% of these genotypes. Subsequent human review disclosed that all the designations were correct. These results suggest that TrueAllele can eliminate human review of most STR DNA databasing capillary data.

The breakdown of the computer vs. human review comparison is shown (Table 4).

- An *efficiency factor* (computer "accept," user "accept") shows that a high proportion (84.9%) of the STR data was scored automatically without requiring human review. The sum of the second row entries (computer "edit") that the computer asks the user to edit was low (14.9%).
- The *accuracy* shown in the last two columns of the first row (computer "accept," user "edit" or "reject") gave a zero false negative error rate (0.0%).

The human processing time was recorded for 35% of the data (5,880 genotypes) as 8.5 hours, or 510 minutes. Reducing this time by the efficiency gained from reviewing only computer "edit" data (14.9%, from the second row), the effective review time was 76 minutes. Dividing the number (5,880 genotypes) by the time spent (76 minutes) yields an average human review time of 77 genotypes per minute, or *over 1 reviewed genotype per second*. These results provide additional support for the use of the TrueAllele as an automated review expert system for DNA databases.

## Specific Aims Results

This section describes the results of the TrueAllele validation project in the context of the specific aims of the original proposal.

- (1) *Reduce by at least 75% the human review of forensic STR data.*
  - (1a) *Classify data into good (accept), bad (reject), and ambiguous (review).*
  - (1b) *Compute quantitative measures for such classification.*
  - (1c) *Apply decision criteria to the quantitative measures.*

With the rule parameters optimized for each lab, TrueAllele produced no incorrect designations on these data. TrueAllele was more effective on ABI than Hitachi data, due to the higher quality of the ABI instrument. TrueAllele reduced by over 75% the human review of forensic STR databank data. Therefore, a significant reduction in the human review of forensic STR data may be achieved.

- (2) *Perform quality assessments of STR data, and provide lab feedback.*
  - (2a) *Compute quality scores that reflect data confidence.*
  - (2b) *Apply rules to identify data artifacts.*
  - (2c) *Report quality scores and identified artifacts to the laboratory.*

TrueAllele's quality scores accurately reflected the data confidence, and the 20+ designation rules properly identified data artifacts. These assessments appear to be useful in improving laboratory data quality.

- (3) *Fully automate the forensic STR data scoring process.*
  - (3a) *Set up the data and their annotations for processing.*
  - (3b) *Process the data: extract, designate, and assess quality.*
  - (3c) *Operate independently of sequencer, chemistry, or computer.*

The set up of data in the TrueAllele process was entirely automated. Moreover, TrueAllele's processing of forensic STR data was fully automatable. TrueAllele operated independently of both hardware platform and STR chemistry.

## **Conclusion**

We have conducted a concordance validation study of the TrueAllele expert system for DNA database applications. We processed 48,710 genotypes using diverse STR panels (Promega PowerPlex, ABI ProfilerPlus and Cofiler) and automated DNA sequencers (ABI/310, ABI/3700 and Hitachi FMBio). We demonstrated that our automated analysis and quality assurance system is accurate, and that it can reduce the human time, error, effort and costs associated with conventional labor-intensive review of CODIS data. Moreover, our automated methods are objective (done by computer) and comprehensive (enabling analysis of all DNA samples), providing for more uniform application of DNA forensic technology.

## **Acknowledgements**

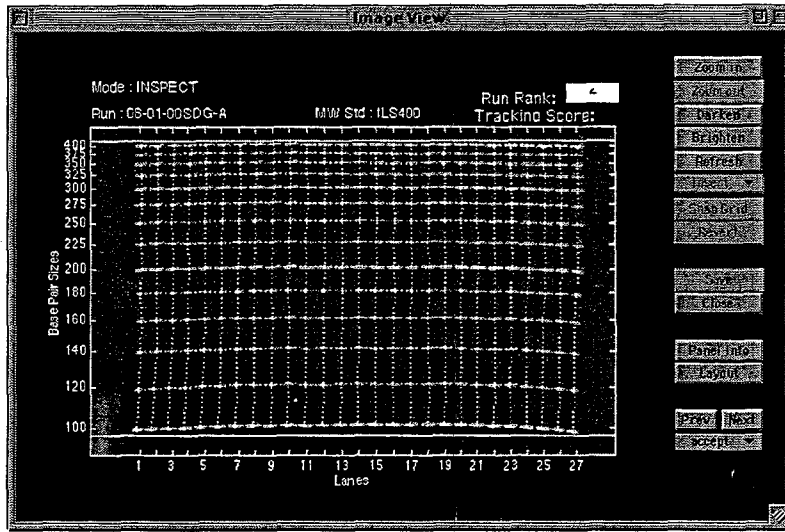
The investigators would like to thank our collaborating laboratories, specifically (1) David Coffman at the Florida Department of Law Enforcement (FDLE) in Tallahassee, FL, (2) Cecelia A. Crouse and Felipe Konotop of the Palm Beach County Sheriff's Office (PBSO) in West Palm Beach, FL, and (3) Jeffrey D. Ban in the Division of Forensic Science (DFS) in Richmond, VA. Drs. Lisa Forman (NIJ) and Barry Brown (FBI) helped in the design of our database validation study. At Cybergenetics, the project was coordinated by Bill Allan, with extensive data review by James Torchia and Alan Enriquez, and expert assistance from Meredith Clarke, Michael Breen and Dr. Mariusz Lubomirski. Orna Taran coordinated the casework validation study. All the validation work described herein was performed in Pittsburgh by Cybergenetics personnel. This research was supported under Award number 2000-IJ-CX-K005 from the Office of Justice Programs, National Institute of Justice, Department of Justice. Points of view in this document are those of the authors, and do not necessarily represent the official positions of Cybergenetics, FDLE, PBSO, DFS, or the U.S. Department of Justice.

## References

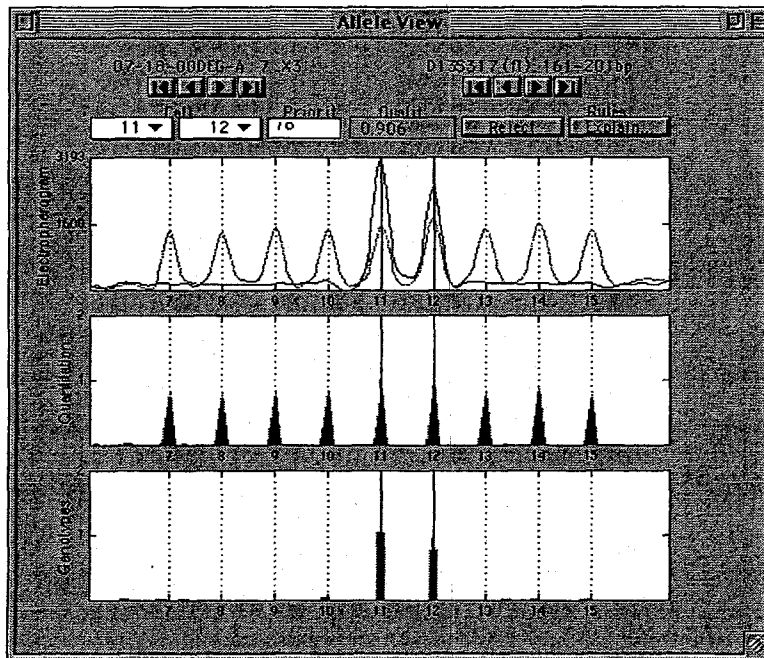
1. Butler JM. Forensic DNA Typing: Biology and Technology Behind STR Markers. New York: Academic Press, 2000
2. Gill P, Sparkes R, Kimpton CP. Development of guidelines to designate alleles using an STR multiplex system. Forensic Sci. Int. 1997;89:185-197.
3. DNA Advisory Board. Quality assurance standards for forensic DNA testing laboratories and for convicted offender DNA databasing laboratories. Forensic Sci Commun (FBI) 2000;2(3).
4. Evett IW, Weir BS. Interpreting DNA Evidence: Statistical Genetics for Forensic Scientists. Sunderland, MA: Sinauer Assoc, 1998
5. Ladd C, Lee HC, Yang N, Bieber FR. Interpretation of complex forensic DNA mixtures. Croat Med J 2001;42(3):244-246.
6. Perlin MW. Computer automation of STR scoring for forensic databases. In: First International Conference on Forensic Human Identification in The Millennium; 1999 Oct 25-27; London, UK: The Forensic Science Service; 1999.
7. Perlin MW. An expert system for scoring DNA database profiles. In: Promega's Eleventh International Symposium on Human Identification; 2000; Biloxi, MS; 2000.
8. Perlin MW, Szabady B. Linear mixture analysis: a mathematical approach to resolving mixed DNA samples. Journal of Forensic Sciences 2001;46(6):1372-1377.
9. Perlin MW. Method and system for genotyping. US Patents; 1996-2000. Report No.: #5,541,067, #5,580,728, #5,876,933 and #6,054,268.

## Figures

**Figure 1. Automated processing.** Quality assurance on Hitachi data after automated (A) gel analysis (step 2) and (B) allelic analysis (step 3).



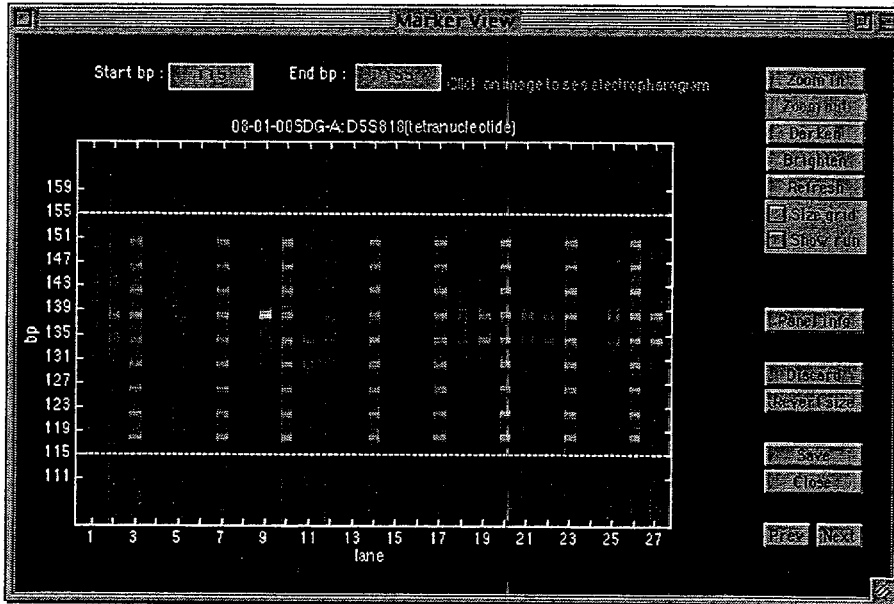
(A)



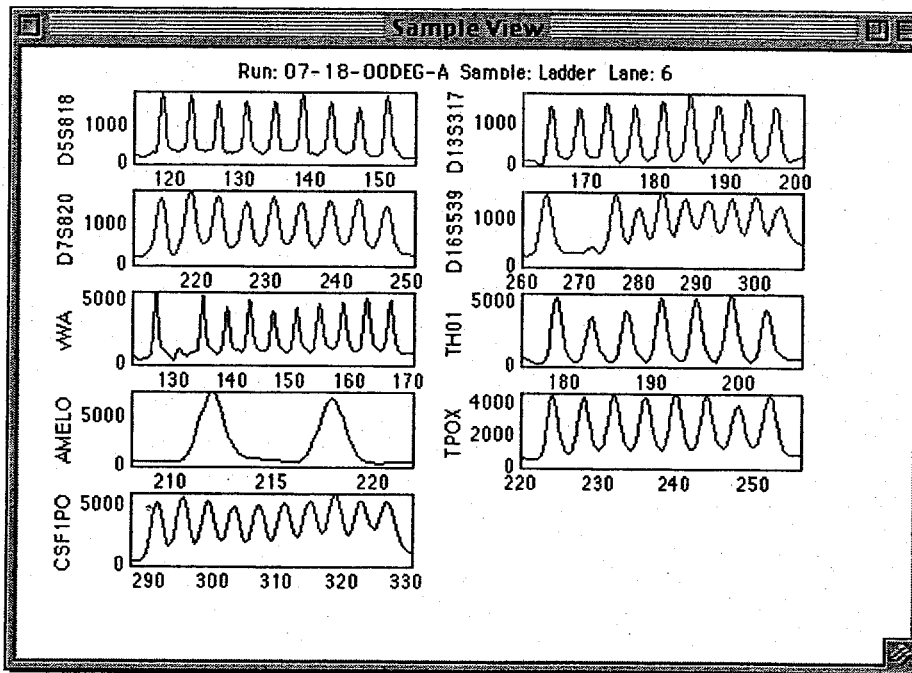
(B)



**Figure 2. Quality assurance.** Some of TrueAllele's many visualizations for viewing the quality of (A) STR data and (B) ladder controls.

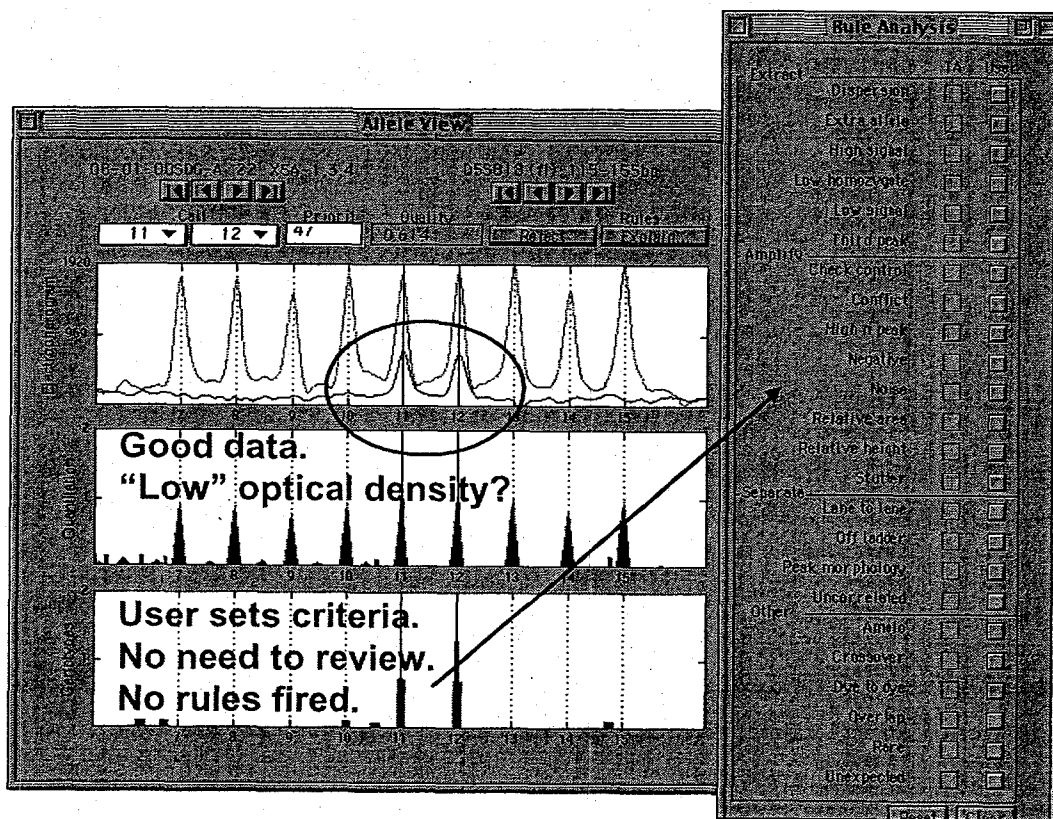


(A)

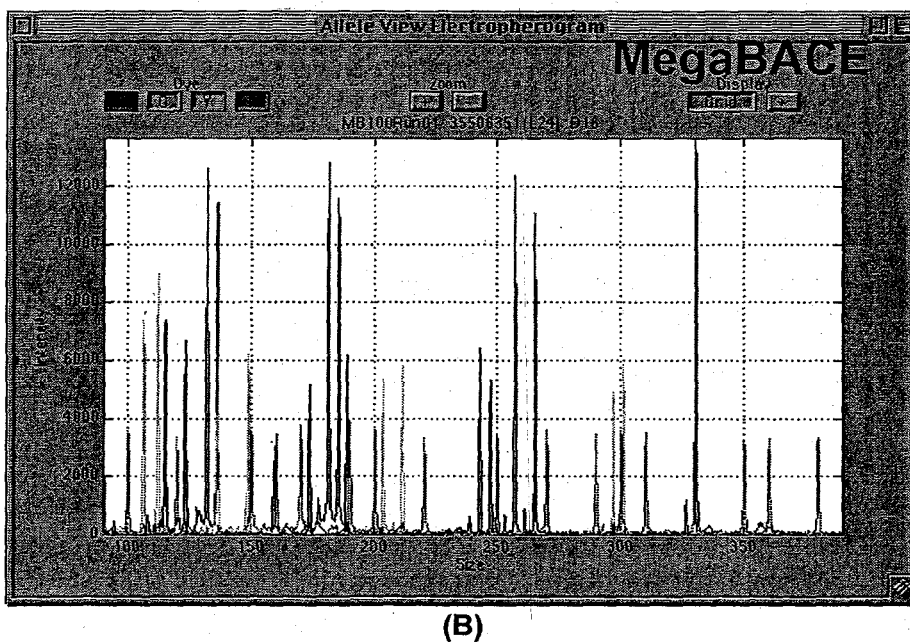
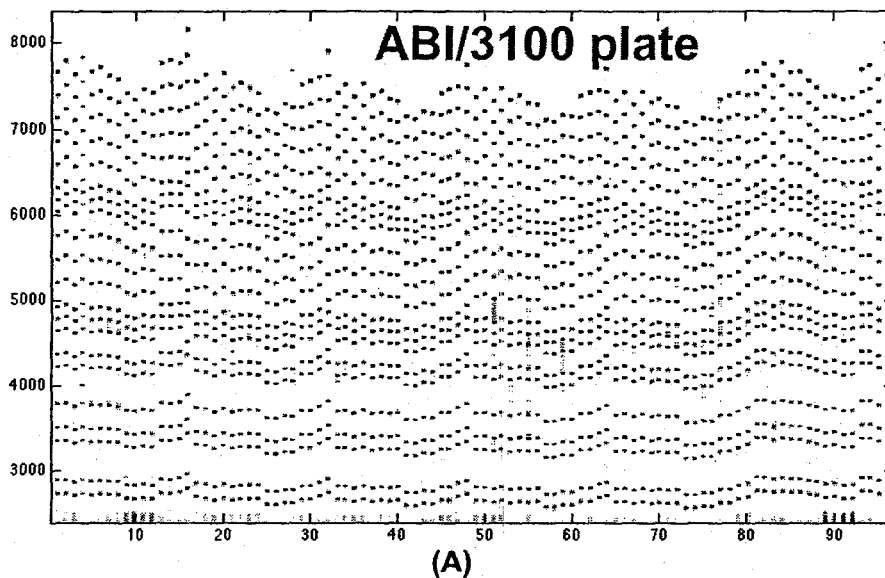


(B)

**Figure 3. Rule system.** The AlleleView navigator, shown together with the Rule Analysis. Although the optical density of the Hitachi bands might be considered "low," the user-set criteria here lead the computer to decide that the data are good (no rules fired), and therefore no human review is necessary. Different user criteria might lead to a different computer decision (e.g., firing a rule). The rule criteria are customized to the needs of the individual laboratory.



**Figure 4. Multi-platform engine.** TrueAllele runs on most DNA sequencers, including **(A)** the 16-capillary ABI/3100 (one 96-well plate in six runs) and **(B)** the 96-capillary Amersham MegaBACE.



## Tables

**Table 1. Rule parameter settings.** Rule settings used for the Hitachi gel and ABI capillary platforms.

PHASE	RULE	GEL		CAPILLARY	
		Status	Value	Status	Value
EXTRACT	Dispersion	On	0.6	On	0.8
	Extra Allele	On	0.25	On	0.15
	High Signal	On	9000	On	10000
	Low Signal	On	150	On	300
	Low Homozygote	On	250	On	500
	Third Peak	On	0.23	On	0.15
AMPLIFY	Check Control	On		On	
	Conflict	On		On	
	High n Peak	On	0.3	On	0.3
	Negative	On		On	
	Noise	On		On	
	Relative Area	On	0.5	On	0.5
	Relative Height	On	0.5	On	0.5
	Stutter	On	0.2	On	0.15
SEPARATE	Lane to Lane	Off		Off	
	Off Ladder	On	0.4	On	0.4
	Peak Morphology	On	0.5	On	0.2
	Uncorrelated	On	0.4	On	0.4
OTHER	Amelo	On		On	
	Crossover	On		On	
	Dye to Dye	Off		Off	
	Overlap	On		On	
	Rare	On		On	
	Unexpected	On	0.25	On	0.15

**Table 2. Hitachi/FMBio (PowerPlex) results.** A table comparing the computer decisions (accept, edit, reject) with the human decisions (accept, edit, reject). \*With more appropriate threshold settings, the accept-accept efficiency rate was actually 82.0%.

		HUMAN REVIEW		
		Accept	Edit	Reject
COMPUTER PROCESS	Accept	72.5% *	0.0%	0.0%
	Edit	23.0%	4.2%	0.3%
	Reject	0.0%	0.0%	0.0%

**Table 3. ABI/310 (ProfilerPlus & Cofiler) results.** A table comparing the computer decisions (accept, edit, reject) with the human decisions (accept, edit, reject).

		HUMAN REVIEW		
		Accept	Edit	Reject
COMPUTER PROCESS	Accept	86.4%	0.0%	0.0%
	Edit	10.9%	2.1%	0.3%
	Reject	0.3%	0.0%	0.0%

**Table 4. ABI/3700 (ProfilerPlus & Cofiler) results.** A table comparing the computer decisions (accept, edit, reject) with the human decisions (accept, edit, reject).

		HUMAN REVIEW		
		Accept	Edit	Reject
COMPUTER PROCESS	Accept	84.9%	0.0%	0.0%
	Edit	10.8%	3.3%	0.8%
	Reject	0.2%	0.0%	0.0%