

Expert Systems for High Throughput Analysis of Single Source Samples:
A Comparison of GeneMarker® HID v1.71 and GeneMapper® ID v3.2 and
Validation of GeneMapper® ID v3.2

Nicole R. Phillips, BS

ABSTRACT

The NIJ Convicted Offender DNA Backlog Reduction Program has helped laboratories in the U.S. to enhance throughput of single source sample processing technologies and methods with the adoption of robotics, single amplification kits and multicapillary instruments. With these advances, the bottleneck in forensic laboratories today is data analysis. Commercially available expert systems offer automated sizing and analysis, expediting the process of DNA data analysis. The University of North Texas Center for Human Identification Research & Development Laboratory has been contracted for several population database construction projects, requiring a high throughput approach for genetic processing of the single source samples. Summarized in this report are the comparison of two expert systems and the internal validation of a GeneMapper® ID v3.2 (Applied Biosystems, Foster City, CA) as an expert system. GeneMarker® HID v1.71 (SoftGenetics, State College, PA) and GeneMapper® ID v3.2 are the two expert systems chosen for the initial evaluation. The features, user interface, and performance of each software program were compared. GeneMarker HID and GeneMapper ID performed similarly, making accurate allele calls and appropriately directing the analyst's attention

to data that do not meet defined thresholds. The decision was made to internally validate GeneMapper *ID* as an expert system for STR data analysis of single source samples. The expert system was successfully optimized for the analysis of samples amplified with Identifiler® and Yfiler® PCR Amplification Kits (*Applied Biosystems*) and analyzed on the 3130xl Genetic Analyzer (*Applied Biosystems*).

EXPERT SYSTEMS FOR HIGH THROUGHPUT ANALYSIS OF SINGLE
SOURCE SAMPLES: A COMPARISON OF GENEMARKER[®] HID V1.71 AND
GENEMAPPER[®] *ID* V3.2 AND VALIDATION OF GENEMAPPER[®] *ID* V3.2

PRACTICUM REPORT

Presented to the Graduate Council of the Graduate School of Biomedical
Sciences at the University of North Texas Health Science Center in Partial

Fulfillment of the requirements for the degree of

MASTER of SCIENCE

By

Nicole R. Phillips, BS

Fort Worth, TX

May 2009

ACKNOWLEDGMENTS

I would like to acknowledge the staff at SoftGenetics and Applied Biosystems for Technical Support, with specific thanks to Teresa Snyder-Leiby, David Hulce, and John Fosnacht at SoftGenetics. Thanks to Valerie Bostwick from the NIJ's National Expert System Testbed Project who provided technical support and answered many questions.

I would like to greatly acknowledge my team of mentors at the University of North Texas Center for Human Identification. Dr. Rhonda Roby was (and continues to be) so generous with her time, patience and knowledge. Her support and encouragement during this process have greatly facilitated my growth as a professional; for that I am truly grateful. Additionally, Dr. Suzanne Gonzalez, Dr. John Planz, Dr. Arthur Eisenberg, and Dr. Joseph Warren were teachers, mentors and advocates of this research.

I would like to acknowledge my husband Reuben, friends and family for their endless support and optimism. This truly would not have been possible without their tolerance and encouragement.

TABLE OF CONTENTS	Page
CHAPTER 1. INTRODUCTION/PURPOSE.....	1
CHAPTER 2. MATERIALS AND METHODS.....	5
2.1 Preliminary Comparison of Two Expert Systems	5
2.2 Internal Validation of GeneMapper <i>ID</i>	7
CHAPTER 3. RESULTS.....	10
3.1 Preliminary Evaluation/Comparison of GeneMarker HID and GeneMapper <i>ID</i>	11
3.2 Internal Validation of GeneMapper <i>ID</i> for Analysis of Identifier Data: Calibration Phase	13
3.3 Internal Validation of GeneMapper <i>ID</i> for Analysis of Yfiler Data: Calibration Phase	14
3.4 Internal Validation of GeneMapper <i>ID</i> : Concordance Phase	15
CHAPTER 4. DISCUSSION.....	16
4.1 Preliminary Evaluation/Comparison of GeneMarker HID and GeneMapper <i>ID</i>	16
4.2 Internal Validation of GeneMapper <i>ID</i> as an Expert System for High Throughput Analysis of Single Source Samples	37
CHAPTER 5. CONCLUSIONS.....	50
REFERENCES.....	52
APPENDIX A. APPROACH TO CALIBRATING THE EXPERT SYSTEM FOR INTERNAL VALIDATION	54
APPENDIX B. OVERVIEW OF GENEMAPPER <i>ID</i> AND GENEMARKER HID WORKFLOW	55
APPENDIX C. GENEMAPPER <i>ID</i> EXPERT SYSTEM PROTOCOL	56

INDEX OF FIGURES AND TABLES	Page
FIGURE 1. Adding Samples to the Run Wizard in GeneMarker HID	17
FIGURE 2. The Run Wizard Analysis Setup Windows in GeneMarker HID	20
FIGURE 3. The Size Calibration Window in GeneMarker HID	22
FIGURE 4. GeneMarker HID's Main Analysis Window	24
FIGURE 5. Allele Report Filter: <i>Show Only Uncertain Alleles</i>	26
FIGURE 6. Rule Firings for Off Scale Data in GeneMapper <i>ID</i> and GeneMarker HID	33
FIGURE 7. Non-concordant Call between GeneMarker HID and GeneMapper <i>ID</i>	34
FIGURE 8. Inconsistent Rule Firing; Peak Height Ratio rule fired in GeneMapper <i>ID</i>	35
FIGURE 9. Inconsistent Rule Firing; Imbalance rule fired in GeneMarker HID	36
FIGURE 10. Peak Height Ratio (PHR) Rule Firing: Settings #1 vs. Settings #2	38
FIGURE 11. GeneMapper <i>ID</i> 's Minus-A Filtering Capacity: Settings #2 vs. Settings #3	39
FIGURE 12. Sample Electropherogram (59F), Discordant at Amelogenin	41
FIGURE 13. Sample Electropherogram (60F), Discordant at D13S317	42
FIGURE 14. The DYS385 Passing Table	45
FIGURE 15. The DYS19 -2bp Artifact: Settings #1 vs. Settings #2	46
FIGURE 16. The +3bp Stutter Peak Detected at DYS392: Settings #2 vs. Settings #3	47
TABLE 1. Unique features of GeneMarker HID and GeneMapper <i>ID</i>	11
TABLE 2. Concordance Study Results with GeneMarker HID and GeneMapper <i>ID</i>	12
TABLE 3. Identifiler Calibration Phase Results	13
TABLE 4. Yfiler Calibration Phase Results	14
TABLE 5. Results of the Concordance Phase for Internal Validation of GeneMapper <i>ID</i>	15
TABLE 6. Abbreviations and Definitions of Rules Employed by GeneMapper <i>ID</i> and GeneMarker HID	23

CHAPTER 1

INTRODUCTION AND PURPOSE

An expert system, as defined by Feigenbaum in Encyclopedia of Computer Science, is “a computer program that reasons, using knowledge, to solve complex problems” (1). Research in and development of expert systems began in the 1960s. It was in the 1980s that they were first used. In the past 30 years, the number of expert systems being developed has sustained rapid growth. Expert systems are software programs that automate expert problem solving for specific types of scenarios and are a subset of artificial intelligence. They make decisions and reason in particular situations based on prioritized rules and “if-then” statements, which create the foundation of the program’s stored knowledge base. These rules and statements are applied to data to mimic the human decision-making process accurately and consistently. When the system does not have the required information to make a decision, a signal or flag should be prompted; hence, an expert system is limited by the thoroughness of its knowledge base. Such programs, or expert systems, are extensively used in other disciplines such as medical diagnostics, engineering, electronic scheduling/planning, and risk assessments (2). By combining computer power with expert system technology, productivity and quality can be improved.

In forensic sciences, DNA databasing of reference samples is critical for solving crimes and identifying human remains. For example, cases of property crime, sexual assaults, homicides, and mass disasters all require the processing of references to compare to DNA profiles obtained from evidence (3). The number of reference samples in need of DNA testing is

vast. Combining changes in legislation with the time required for processing the increasing number of samples, a substantial backlog has developed. Recent advances in robotics, single amplification multiplex STR kits, and multicapillary instruments have enhanced forensic laboratory efficiency; thus, the bottleneck in DNA processing time has shifted to data analysis (4). Reports indicate that over 50% of the resources expended to designate a genetic profile are used for data analysis (5). Two analysts are required to visually inspect the electropherogram of each sample in order to report the genetic profile (6), which consumes valuable time and resources. An expert system increases laboratory efficiency by automatically interpreting nuclear DNA profile data. As opposed to manually reviewing every allele in the genetic profile, analyst intervention is only necessary for flagged data. Results obtaining high quality scores and meeting all defined criteria are labeled “pass” and are accepted as the actual allele calls without analyst review, thereby, drastically reducing the average analysis time per sample. Although the currently available expert systems do not eliminate the need for analyst interpretation, they greatly increase time and cost efficiency when incorporated into the analysis workflow (7). The National DNA Index System’s (NDIS) DNA Acceptance Standards states that expert system technology is acceptable for review of convicted offender reference samples amplified with an NDIS approved kit and provides guidelines for developmental and internal expert system validation are also provided in Appendix B (8). With the extensive backlog of convicted offender samples serving as motivation, improvements and changes in the field of high throughput forensic DNA databasing are needed and inevitable. Definitively, expert system technologies can play an integral role in reducing the backlog of data analysis, facilitating the upload of critical convicted offender profiles.

Internal validation studies have been conducted to evaluate expert system analyses of single source reference samples. Such studies have found that expert systems technology is accurate and consistent at calling alleles (9). For a laboratory to perform internal or developmental validation of an expert system, it must follow Section 8 of the Quality Assurance Standards for Convicted Offender DNA Databasing Laboratories (10) and other CODIS requirements for acceptability of the data (11). Validating an expert system does not simply refer to the software program; the chemistry kit and the analysis platform are also variables in the system. The NIJ's National Expert System Testbed Project has termed the combination of chemistry kit, analysis platform and expert system as the "complete set". If any component of the "complete set" is altered, additional validation is required. For example, a laboratory that has internally validated an expert system for analysis of samples amplified with Identifiler PCR Amplification Kit (*Applied Biosystems, Foster City, CA*) and analyzed on the 3100 Genetic Analyzer (*Applied Biosystems*) cannot use the expert system for analysis of Identifiler data analyzed on a 3130xl. Such analyses would require further validation.

The University of North Texas Center for Human Identification Research and Development Laboratory (UNTCHI R&D) has recently acquired several requests for population database construction projects. These projects require high throughput autosomal and Y-STR profiling of single source population samples. In preparation for the projects, the UNTCHI R&D team was very interested in implementing expert system technology to expedite data analysis. The NDIS DNA Acceptance Standards are written specifically for the approval of expert systems for convicted offender samples and NDIS approved kits uploaded into CODIS. Population samples and/or samples amplified with Y-STR kits (since there are no NDIS approved Y-STR kits) are not mandated under the NDIS guidelines. The UNTCHI R&D team proceeded with

selecting and validating an expert system for population databasing purposes, using the standards set forth in Appendix B of the NDIS DNA Data Acceptance Standards (8) for the validation design. An internal validation consists of two phases: a calibration phase and a concordance phase. In the calibration phase, the expert system's efficiency is optimized by adjusting the settings and thresholds for rule firings. A set of 200 unique samples is required to test the proposed settings. Once the optimized settings are established, an additional set of unique single source samples are used to validate the settings in the concordance phase of the study. The size of the concordance phase sample set should be 1000 unique samples or 10% of the laboratory's annual single source sample throughput (whichever is fewer).

Illustrated here is the process of selecting and internally validating an expert system for the UNTCHI R&D Laboratory. First, two expert systems were evaluated in a preliminary comparison study: Genemarker[®] HID v1.71 (*SoftGenetics, State College, PA*) and GeneMapper[®] ID v3.2 (*Applied Biosystems*). These software packages were evaluated with respect to their accurate allele designations and appropriate rule firings. Additionally, the workflow, unique features, and user interface were also assessed. GeneMapper ID was then selected since every workstation already has a license for this software, and the expert system was internally validated according to the NDIS Appendix B guidelines (8).

CHAPTER 2

MATERIALS AND METHODS

2.1 Preliminary Evaluation and Comparison of GeneMarker HID and GeneMapper ID

Two hundred buccal swabs were extracted with the Freedom EVO[®] 100 (*Tecan Group Ltd., Männedorf, Switzerland*) using DNA IQ[™] system chemistry (*Promega Corporation, Madison, WI*). One hundred reference specimens were amplified with AmpFLSTR[®] Identifiler[®] PCR Amplification Kit and 100 reference specimens were amplified with PowerPlex 16[®] System (*Promega*). The Identifiler amplified products were electrophoresed on the 3130xl Genetic Analyzer, and the PowerPlex 16 amplified products were electrophoresed on the 3100 Genetic Analyzer (*Applied Biosystems*). The Identifiler data set and the PowerPlex 16 data set were extracted, amplified, and analyzed by the standard operating procedures of the UNTCHI Forensic Laboratory and paternity Laboratory, respectively. The sample files (.fsa files) generated from these 200 reference specimens were used in the concordance study. These electronic data were analyzed using both GeneMarker HID and GeneMapper[®] ID v3.2. When possible, uniform parameter settings were used in all data analyses. A second analysis was performed on the sample sets using a 20% filter (often referred to as a *global filter*), which disregards all peaks within a locus which are less than or equal to 20% of the largest peak in the allele size range. For each analysis (with and without the 20% global filter), resulting allele calls and rule firings (data flagged for violation of an established rule or threshold) were evaluated

side-by-side. All rule firings were independently investigated in both systems. The following information was gathered from the cumulative results for each expert system:

- Total number of peaks detected
- Number of peaks accurately called and no rules fired
- Number of peaks accurately called and a rule fired
- Number of peaks called incorrectly and a rule fired
- Number of peaks called incorrectly and no rules fired

Additionally, the overall operations and features of GeneMarker HID were evaluated and compared to GeneMapper *ID*. A workflow for GeneMarker HID was established using the Help Tree and GeneMarker HID User Manual. The National Institute of Justice's Expert Systems Testbed Project previously established the workflow for GeneMapper *ID* (2). Each step in the process was evaluated from the standpoint of a DNA analyst who is processing samples for DNA databasing. User-interface, functionality and features of each program were reviewed, focusing on the following integral processes:

- Establishing analysis settings and data processing
- Verification of sizing quality
- Analyst navigation: investigation of rules fired and verification of controls
- Allele editing and audit trail
- Exporting results
- Printing results

2.2 *Internal Validation of GeneMapper ID as an Expert System for High Throughput Analysis of Single Source Samples*

2.2.1 Internal Validation of GeneMapper ID for Analysis of Identifiler Data

The expert system calibration phase included 226 samples. All samples were amplified with the Identifiler and analyzed using the 3130xl Genetic Analyzer. This sample set contained 164 family reference samples, amplified and electrophoresed in the Forensic Laboratory, and 62 databasing samples, amplified and electrophoresed in the UNTCHI R&D Laboratory. For the concordance phase of this study, a minimum of 120 single source reference samples (or 10% of annual throughput) is recommended according to Appendix B of the NDIS standards; in this study, 340 samples were used. All 340 samples were amplified and electrophoresed in the Forensic Laboratory. A second calibration sample set containing 257 samples amplified and electrophoresed in the UNTCHI R&D Laboratory was also tested, totaling 597 samples in the Identifiler concordance phase set.

The expert system analysis parameters were calibrated using 226 unique samples (See Appendix A for the general workflow). The initial settings (Settings #1) were based on the established standard operating procedure, *Interpretation Guidelines for STR Analysis*, validated by the Forensic Laboratory. The passing genotypes were exported and checked for concordance with the manual calls using formulas designed in Microsoft® Office Excel (*Microsoft Corporation, Redmond, WA*). Any incorrect genotypes called by the expert system as “pass” using Settings #1 were investigated and documented. The low quality loci were analyzed based on the rules that fired. Subsequent rule adjustments were made in a step-wise fashion in effort to optimize the expert system’s application. For each setting change, the “pass” genotypes were exported and checked for concordance with the manual calls. The final set of parameters was

selected based on the optimization of the software. Once the parameters were established using the 226 calibration samples, two sets of reference samples (340 from the UNTCHI Forensic Laboratory and 257 from UNTCHI R&D) were used to check for concordance. Alleles called with passing genotype quality were compared to the alleles called by manual review using the current interpretation guidelines of the UNTCHI R&D Laboratory. Any incorrect allele calls designated as “pass” by the expert system were evaluated and documented.

2.2.2 Internal Validation of GeneMapper *ID* for Analysis of Yfiler Data

The expert system calibration phase included 200 samples. All samples were amplified with the Yfiler[®] PCR Amplification Kit (*Applied Biosystems*) and analyzed using the 3130xl Genetic Analyzer in the UNTCHI R&D Laboratory. Since UNTCHI does not regularly process family reference samples with Yfiler, the concordance phase sample set was selected to approximately match the size of the Identifiler concordance phase sample set. A total of 337 samples were amplified and electrophoresed for Yfiler in the UNTCHI R&D Laboratory and used in this study.

The expert system analysis parameters were calibrated using 200 unique samples (see Appendix A for the general workflow). The initial settings (Settings #1) were based on the optimization of GeneMapper *ID* using a subset of samples amplified and electrophoresed in the UNTCHI R&D Laboratory (N=133). The passing genotypes were exported and checked for concordance with the manual calls using concordance formulas designed in Microsoft Office Excel. The rules that fired for the low quality loci were evaluated; adjustments were made in a stepwise fashion in an effort to optimize the expert system’s application. For each setting

change, the “pass” genotypes were exported and checked for concordance with the manual calls. The final set of parameters was selected based on the optimization of the software. Once the parameters were established using the 200 calibration samples, a set of population samples (337 from the UNTCHI R&D Laboratory) were used to check for concordance. Loci called with passing genotype quality were compared to the alleles called through manual review using the current interpretation guidelines of the UNTCHI R&D Laboratory. Any incorrect allele calls deemed “pass” by the expert system were evaluated and documented.

CHAPTER 3

RESULTS

In this chapter, the results of each phase of the study are summarized. Section 3.1 contains the results of the comparison study between GeneMapper *ID* and GeneMarker HID. The key features and limitations of each program are described, as are the results of the performance evaluation. Sections 3.2- 3.4 provide results for the internal validation of GeneMapper *ID* as an expert system for the analysis of Identifiler and Yfiler data electrophoresed on the 3130xl Genetic Analyzer. Section 3.2 details the parameter settings used in the calibration phase of the Identifiler sample set. Similarly, Section 3.3 illustrates the parameter settings used in the Yfiler calibration phase. Section 3.4 provides the concordance phase results for both the Identifiler and Yfiler sample sets.

3.1 Preliminary Evaluation and Comparison of GeneMarker HID and GeneMapper ID

Table 1. Unique features of GeneMarker HID and GeneMapper ID

The following table summarizes *unique* features of each program for each step in the analysis process. This is not a comprehensive list of each program's features.

Process	GeneMarker HID v1.71	GeneMapper ID v3.2
Establishing analysis settings and parameters	<p>The <i>Run Wizard</i>, which guides the setup process, is very intuitive and user friendly.</p> <p>Several parameters can be set specifically for each locus rather than applied uniformly over all loci (e.g., minimum heterozygote and homozygote intensities; heterozygote imbalance ratios).</p>	<p>Marker specific stutter ratios are pre-loaded.</p> <p>-A filtering capabilities.</p> <p>Each sample can be analyzed with different settings, panel, and/or size standard.</p>
Analysis		Automatically saves analysis progress every ten samples.
Verification of sizing quality	The 250bp peak sizes and statistics are provided in the same window with the sizing standard trace and sizing curves.	
Verification of control concordance		The Control Concordance (CC) rule fires if positive control samples are called differently from the known profile.
Review of fired rules and analyst navigation	<p>Main Analysis Window contains the electropherogram, the allele call table, the rules fired and the peak information all on one screen. Less navigation is required.</p> <p>Allele calls of questionable quality are placed in a yellow or red box on the electropherogram to clearly identify which peaks require analyst intervention.</p> <p>The repeated sample names can be hidden in the Allele Report; samples are easily distinguished.</p>	<p>Pink vertical bars indicate where a peak has saturated the detection system, making it easy to identify pull-up peaks in the other spectra.</p> <p>The Allele Number (AN) rule fires if no alleles are detected within a locus.</p> <p>Sorting features (Bring Low Quality Samples to Top, table setting options) allow the user to focus only on the low quality loci.</p>
Allele editing/audit trail	<p>Once an allele has been edited, it is flagged in the Peak Table to clearly identify modified calls.</p> <p>Project Comparison feature verifies that two independently reviewed projects were assigned concordant profiles.</p>	
Exporting results	<p>Capable of exporting electropherograms as JPEG files.</p> <p>Capable of exporting each page of the Print Report as a JPEG file.</p> <p>Capable of exporting tables directly as Excel files (.xls).</p>	Customizable table options using the Table Settings tab of the GeneMapper Manager.
Printing results	Print Report dialogue box and Print Preview feature simplify the printing process, allowing for enhanced user control.	

Table 2. Concordance Study Results with GeneMarker HID and GeneMapper ID

The results from a concordance study with 100 Identifiler and 100 PowerPlex amplified samples. GeneMarker HID (GMKR) and GeneMapper ID (GMID) were the expert systems used for analysis. The first column defines the amplification chemistry used for the 100 samples. The third column reports the total number of true alleles present in the sample sets. Two analyses were conducted; the second applied a 20% global filter as is shown in the fourth column. The fifth column reports the total number of peaks each expert system detected under the specified analysis settings. The sixth column is the number of alleles correctly designated without requiring the analyst's attention (no rules fired). The seventh column reports the number of true allele peaks called but a rule was fired. These are correctly called alleles; however, the analyst was still flagged to review the loci. The eighth column is the number of non-allele or extraneous peaks that each system detected and called; all of these peaks were flagged for analyst intervention. The final column displays that neither system incorrectly called a peak as an allele without flagging the analyst. The sum of columns 5-8 yields the total number of peaks detected by each expert system.

Samples amplified with	# of samples	# of actual alleles	20% filtering applied	Total # of peaks detected by each system		# of correct peaks called and no rule firings		# of correct peaks called and rules fired		# of extraneous peaks called and rules fired		# of extraneous peaks called and no rule firings	
				GMKR	GMID	GMKR	GMID	GMKR	GMID	GMKR	GMID	GMKR	GMID
Identifiler	100	2806	No	2864	2863	2680	2672	126	134	58	57	0	0
			Yes	2806	2806	2750	2733	56	73	0	0	0	0
PowerPlex 16	100	2931	No	3233	3253	2280	2220	651	711	302	322	0	0
			Yes	2943	2953	2667	2648	264	283	12	22	0	0

3.2 Internal Validation of GeneMapper ID for Analysis of Identifier Data: Calibration Phase

Table 3. Identifier Calibration Phase Results

226 unique samples were analyzed under seven different settings/parameters. The expert system was optimized under Settings #7, and these settings were used in the subsequent concordance phase.

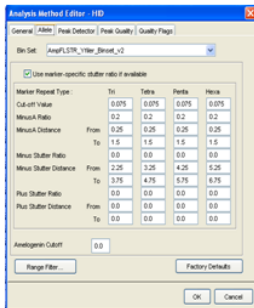

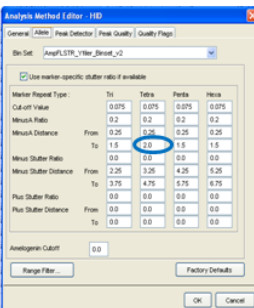

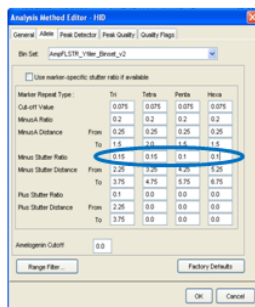

Parameter	Settings #1	Settings #2	Settings #3	Settings #4
Stutter Filter	Marker specific	Marker specific	Marker specific	Marker specific
Cut Off Value	0	0	0	0
Peak detection (all dyes)	100	100	100	100
Min homozygote	200	200	200	250
Min heterozygote	100	100	100	100
Minus A filter	0	0	0.2 (-1.5 to -0.25)	0.2 (-1.5 to -0.25)
PHR	70%	60%	60%	60%
Pull up Ratio	0.05	0.05	0.05	0.05
Max. # alleles	2	2	2	2
Quality Flag settings	Default	Default	Default	Default
Sizing Quality	Pass 0.75-1.00 Fail 0.00-0.25	Pass 0.75-1.00 Fail 0.00-0.25	Pass 0.75-1.00 Fail 0.00-0.25	Pass 0.75-1.00 Fail 0.00-0.25
Genotype Quality	Pass 0.75-1.00 Fail 0.00-0.75	Pass 0.75-1.00 Fail 0.00-0.75	Pass 0.75-1.00 Fail 0.00-0.75	Pass 0.75-1.00 Fail 0.00-0.75
Results				
# passing & concordant loci	2959	2987	3076	3066
# discordant loci	2	2	2	1
# LQ loci	657	629	540	550
Total # loci	3616	3616	3616	3616
Percent of loci passing	81.83%	82.61%	85.07%	84.79%

Parameter	Settings #5	Settings #6	Settings #7
Stutter Filter	Marker specific	Marker specific	Marker specific
Cut Off Value	0	0	0.075
Peak detection (all dyes)	75	100	100
Min homozygote	200	275	200
Min heterozygote	100	100	100
Minus A filter	0.2 (-1.5 to -0.25)	0.2 (-1.5 to -0.25)	0.2 (-1.5 to -0.25)
PHR	60%	60%	60%
Pull up Ratio	0.05	0.05	0.05
Max. # alleles	2	2	2
Quality Flag settings	Default	Default	Default
Sizing Quality	Pass 0.75-1.00 Fail 0.00-0.25	Pass 0.75-1.00 Fail 0.00-0.25	Pass 0.75-1.00 Fail 0.00-0.25
Genotype Quality	Pass 0.75-1.00 Fail 0.00-0.75	Pass 0.75-1.00 Fail 0.00-0.75	Pass 0.75-1.00 Fail 0.00-0.75
Results			
# passing loci	2880	3056	3198
# discordant loci	0	0	2
# LQ loci	736	560	418
Total # loci	3616	3616	3616
Percent of loci passing	79.65%	84.51%	88.44%

3.3 Internal validation of GeneMapper ID for analysis of Yfiler Data: Calibration Phase

Table 4. Yfiler Calibration Phase Results

226 unique samples were analyzed under seven different sets of settings/parameters. The expert system was optimized under Settings #5, and these settings were used in the subsequent concordance phase.

Parameter	Settings A	Settings #1	Settings #2
Allele tab Settings			
Min homozygote	100	100	100
Min heterozygote	50	50	50
PHR	60%	40%	40%
Pull up Ratio	0.05	0.05	0.05
Max. # alleles	2	1	1
Quality Flag settings	Default	Default	Default
Sizing Quality	Pass 0.75-1.00 Fail 0.00-0.25	Pass 0.75-1.00 Fail 0.00-0.25	Pass 0.75-1.00 Fail 0.00-0.25
Genotype Quality	Pass 0.75-1.00 Fail 0.00-0.75	Pass 0.75-1.00 Fail 0.00-0.75	Pass 0.75-1.00 Fail 0.00-0.75
Results			
# passing & concordant loci	2922	2927	3037
# discordant loci	0	0	0
# LQ loci	449	449	163
Total # loci	3200	3200	3200
Percent of loci passing	91.31%	91.47%	94.91%
Parameter	Settings #3	Settings #4	Settings #5
Allele tab Settings			
Min homozygote	100	50	50
Min heterozygote	50	150	150
PHR	40%	40%	40%
Pull up Ratio	0.05	0.05	0.05
Max. # alleles	1	1	1
Quality Flag settings	Default	Default	Default
Sizing Quality	Pass 0.75-1.00 Fail 0.00-0.25	Pass 0.75-1.00 Fail 0.00-0.25	Pass 0.75-1.00 Fail 0.00-0.25
Genotype Quality	Pass 0.75-1.00 Fail 0.00-0.75	Pass 0.75-1.00 Fail 0.00-0.75	Pass 0.75-1.00 Fail 0.00-0.75
Results			
# passing & concordant loci	3039	3071	3136
# discordant loci	0	0	0
# LQ loci	161	129	64
Total # loci	3200	3200	3200
Percent of loci passing	94.97%	95.97%	98.00%

3.4 Internal validation of GeneMapper ID: Concordance Phase

The number of loci genotyped correctly by GeneMapper *ID* without firing any rules (*i.e.*, not requiring human intervention) was totaled. In the concordance phase for the Identifiler sample sets, 95.46% of the 5440 loci in the sample set were correctly genotyped by the expert system without firing any rules, and 85.59% of the 340 samples were correctly genotyped by the expert system at all loci without firing any rules (Table 5). A concordance rate of 100% was obtained for all green or “pass” genotypes called by the expert system. The second concordance sample set (257 samples from UNTCHI R&D) also obtained a 100% concordance rate for all passing genotypes.

In the concordance phase for the Yfiler sample set, 98.39% of the 5392 loci in the sample set were correctly genotyped by the expert system without firing any rules, and 80.42% of the 337 samples were correctly genotyped by the expert system at all loci without firing any rules (Table 5). A concordance rate of 100% was obtained for all green or “pass” genotypes called by the expert system.

Table 5. Results of the Concordance Phase for Internal Validation of GeneMapper *ID* (GMID)

<i>The Complete Set</i>	<i>Total # of samples</i>	<i>% of <u>samples</u> genotyped correctly without any rule firings</i>	<i>Total # of loci</i>	<i>% of <u>loci</u> genotyped correctly without rule firings</i>	<i># of discordant loci/samples</i>
Identifiler/3130xl/GMID (Forensic Family Reference Samples)	340	85.59	5440	95.46	0
Identifiler/3130xl/GMID (UNTCHI R&D Pop. Samples)	257	67.87	4112	93.26	0
Yfiler/3130xl/GMID	337	80.42	5392	98.39	0

CHAPTER 4

DISCUSSION

4.1 Preliminary Evaluation/Comparison of GeneMarker HID and GeneMapper ID

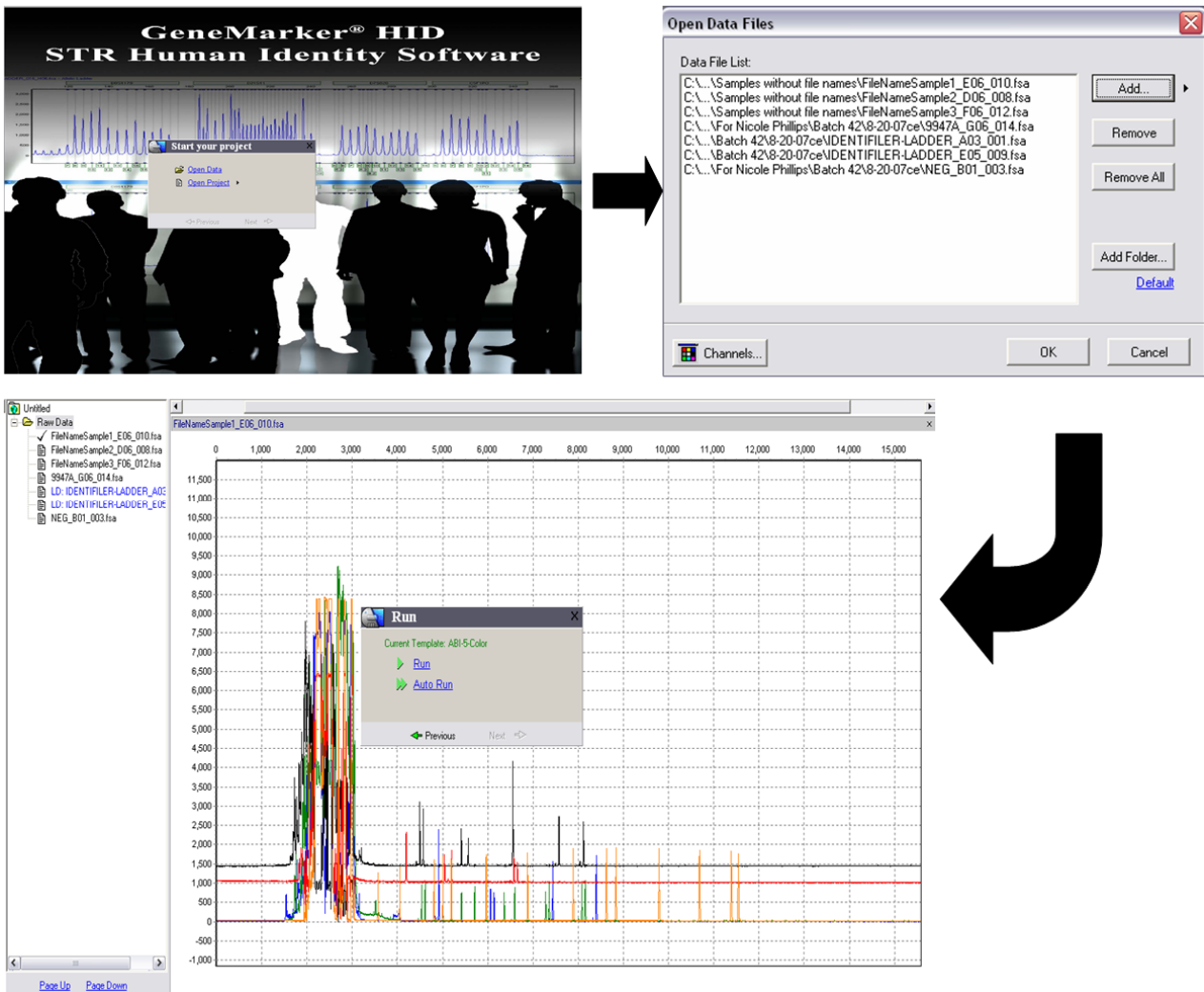
Differences were observed between GeneMarker HID and GeneMapper *ID* when conducting an initial evaluation of the two expert systems. Multiple function, feature and user-interface differences were documented that are unique to each of the software programs used in this study (see Chapter 3, Table 1).

4.1.1 Analysis Setup

GeneMarker HID features a *Run Wizard* which takes the analyst step-by-step through the analysis options. Prompts guide the entire process, from adding raw sample files directly from the instrument through data processing. The software will recognize the user-defined conventions for naming samples to determine which samples are positive controls, negative controls, and ladders. Samples are added to a project using the *Run Wizard* as a guide (Figure 1). After samples are added, the raw data are displayed. The Sample Tree, a hierarchical list of the samples in the project, is displayed in the left panel. The raw data view for individual samples can be launched by double-clicking the file name. The *Run Wizard* signals the user that no analysis has been performed (seen in Figure 1, the third pane). Clicking **Run** initiates a wizard which guides the user in analysis setup.

Figure 1. Adding Samples to the Run Wizard in GeneMarker HID

The Run Wizard setup windows in GeneMarker HID for adding samples to a project. The first frame is the screen seen when opening the software. The *Run Wizard* is launched automatically and prompts the user to either open new data, **Open Data**, or open a previously saved project, **Open Project**, which opens the main analysis window (Figure 3) with all of the sample files. Clicking **Open Data** opens the second window. Here, individual samples can be added by clicking **Add**, or a folder of samples can be added by clicking **Add Folder**. By clicking **OK**, the raw data of selected samples are displayed. The third pane illustrates the raw data window. The Run Wizard is still open as no analysis has yet been performed. Clicking **Run** launches a series of dialog boxes in which the user defines the analysis parameters.

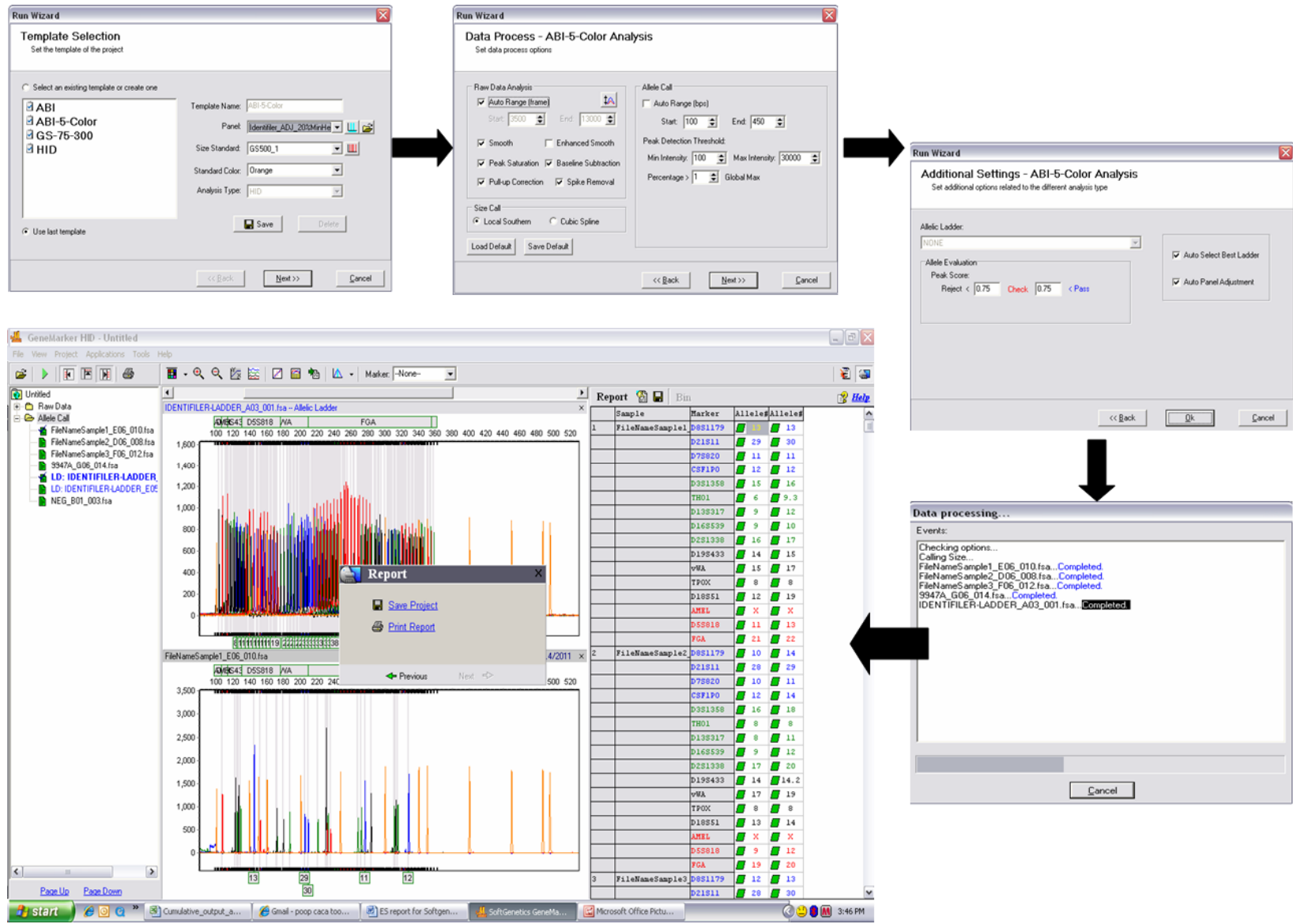


After clicking **Run**, the windows open in series, beginning with the **Template Selection** pane (Figure 2). The user can select a previously saved template name or can

define a new template (kit panel and size standard information). All samples within a project must be analyzed using the same panel and size standard. GeneMarker HID includes panels for all commercially available multiplex STR fluorescently labeled kits from Applied Biosystems and Promega Corporation. Using the Size Template Editor, new size standards can be created. Analysis setup continues with the **Data Process** pane, where the user designates the analysis settings for data processing (*e.g.*, analysis range, smoothing and sizing algorithms, peak detection threshold). The software features *Auto Range* detection for the raw data analysis and allele calls; however, the user may choose to manually assign these ranges. In the **Additional Settings** pane, the user can either manually select the ladder to be used for allele calls or can utilize the *Auto Select Best Ladder* feature. GeneMarker HID calls unknown alleles using one ladder. If *Auto Select Best Ladder* is chosen, the software will select and use the ladder which demonstrates the best panel alignment. This setting is recommended since it saves the analyst from having to manually inspect each ladder's alignments and re-analyze with the chosen ladder. Allelic ladders with poor sizing quality, bin alignment, or peak intensity are flagged accordingly. The user can inspect ladders by selecting the ladder used for analysis from the Sample Tree. The software verifies the proper allele calls in the electropherogram and/or Peak Table. Also displayed in this pane, is the user-defined peak score range in which the software will pass the peak (and resulting genotype) or flag a peak for analyst intervention. Once the settings are established in this window, data processing is automatically launched. The **Data Processing** pane displays analysis progression and reports the time taken to complete the process. GeneMarker HID reports the number of samples analyzed and the time taken for completion after each run. A batch of 47

Identifiler samples was analyzed with each program. GeneMarker HID completed the data processing of these 47 samples in 22.34 seconds. GeneMapper *ID* completed the data processing of the same data set in approximately 32.30 seconds, conducting an automatic save for every ten samples analyzed. These times are not the same with each repetition and will certainly vary with the processing capacity of the computer being used. The reported times were attained on a PC running XP SP3, 1.73 Pentium M, 1 GB RAM. The times reported here indicate that there is little variation in data processing times between the two systems. If choosing between these two expert systems, it is recommended that no weight be placed on the processing times.

Figure 2. The Run Wizard Analysis Setup Windows



Optimization of extraction methods, amplification protocols, instrument parameters, and the expert system are all critical to the results obtained. Likewise, for the optimization of GeneMarker HID, the manufacturer has recommended establishing a modified panel for each instrument. Modified panels are created by analyzing the first run with no panel selected. The software simply sizes the fragments; no allele calls are made. In the Panel Editor, all markers are verified for precise bin alignment (bins can be moved manually). The adjusted panel is named and used for all future analyses with that particular instrument. This process is only necessary when analyzing data from a particular instrument for the first time. Also in the Panel Editor, Minimum Homozygote Intensity, Minimum Heterozygote Intensity and Heterozygote Imbalance are assigned per locus. These specifications allow the user to customize the settings for the characteristic peaks in that locus. For example, the Minimum Homozygote Intensity for smaller STR loci can be set higher than larger STR loci since peak intensity often decreases with allele size in degraded or inhibited samples.

4.1.2 Size Quality Verification

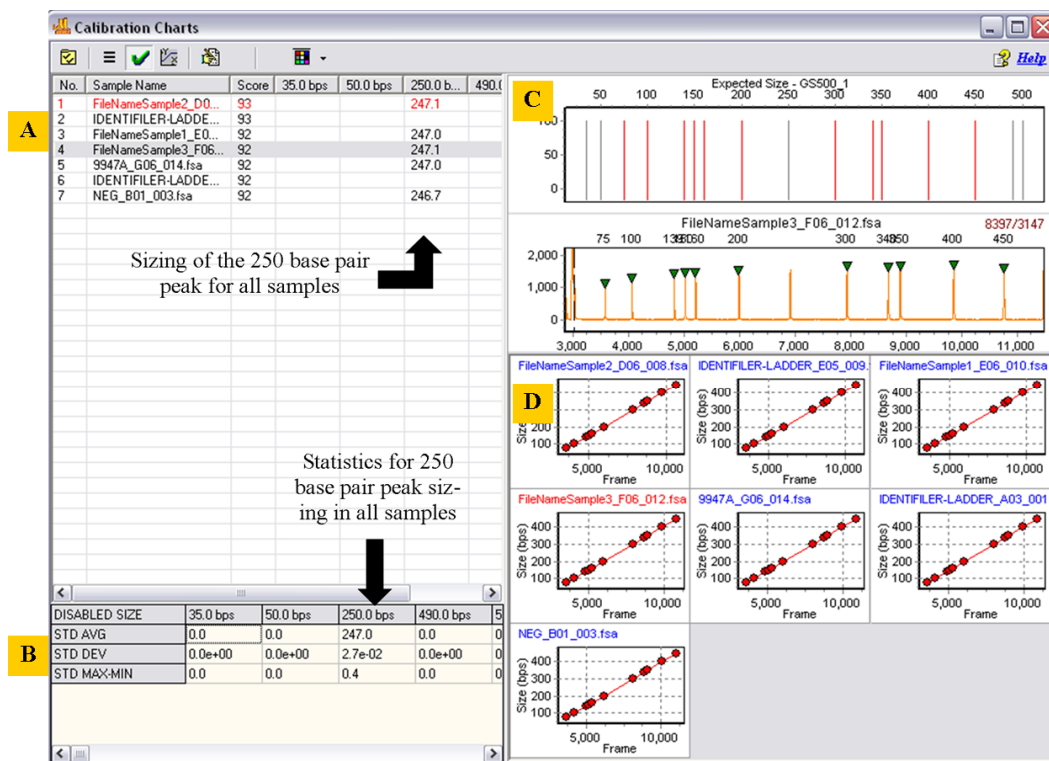
GeneMarker HID and GeneMapper *ID* both automatically assesses the sizing quality of samples. In GeneMarker HID, the Main Analysis Window is launched after data processing, and the sizing quality for each sample is shown in the Sample Tree. Samples with good sizing quality display a green sheet icon whereas those with questionable or poor sizing quality display yellow and red sheet icons, respectively. To view and verify sizing quality, the user clicks the Size Calibration Icon. The Size Calibration window displays the trace and the size calling curve for each sample.

GeneMapper *ID* also displays the sizing quality of each sample using a color indicator.

Unique to GeneMarker HID, the sizing and statistical data for the 250 base pair peak (and/or other omitted peaks) is provided (Figure 3).

Figure 3. The Size Calibration Window in GeneMarker HID

The Calibration Charts display the sample list (Pane A), the 250 bp peak statistics (Pane B), the size standard trace (Pane C), and the sizing curves (Pane D) in one window.



4.1.3 Analyst Navigation

GeneMarker HID displays genotype quality rank icons, which can be seen in the main analysis window. GeneMarker HID assigns a quality rank to each allele rather than to the locus. Any allele with a yellow or red flag has fired one or more previously defined rules, resulting in a reduced quality score. Each flagged allele should be investigated, and the rule(s) fired can be found in the Peak Table (Table 6 lists the rule firings in GeneMarker HID and the analogous rules found in GeneMapper *ID*).

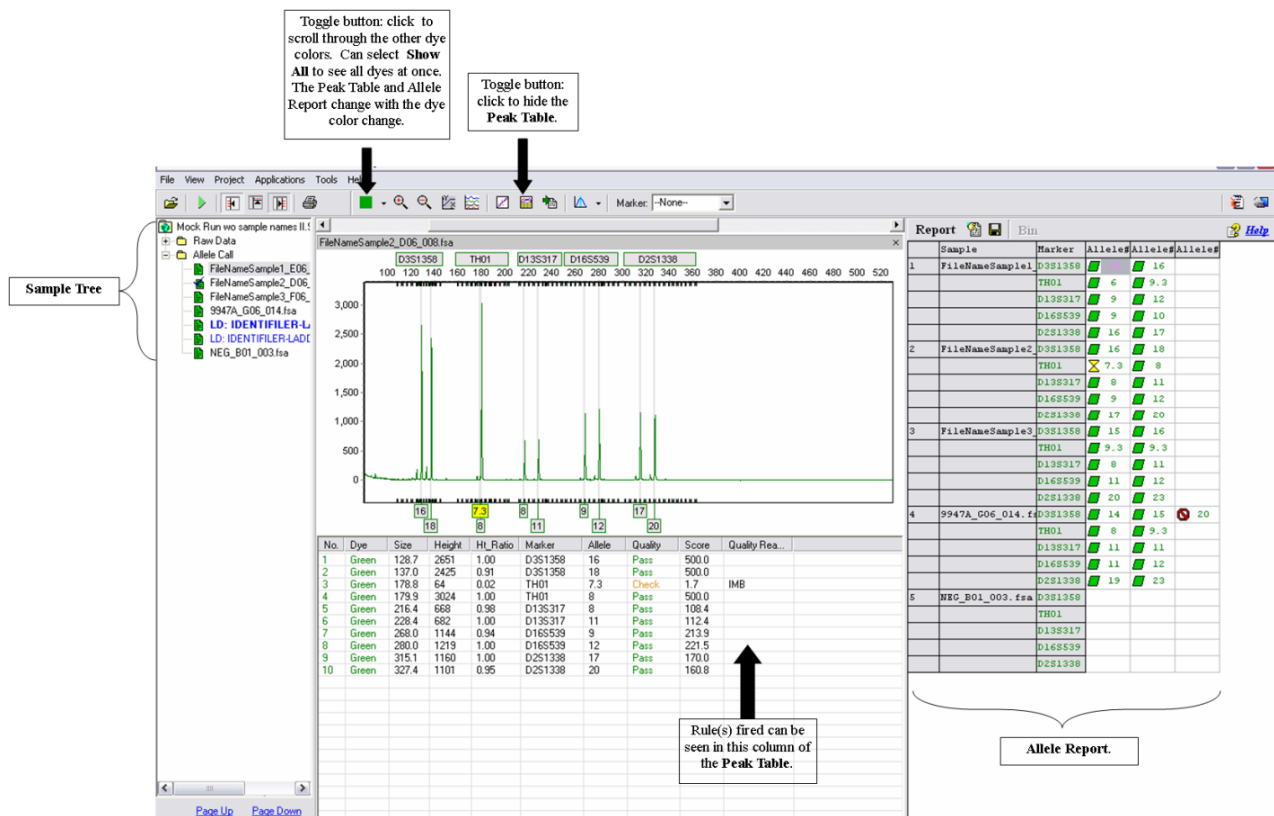
Table 6. Abbreviations and Definitions of Rules Employed by GeneMapper *ID* (GMID) and GeneMarker HID (GMKR)

Rules in GeneMapper <i>ID</i>	Rules in GeneMarker HID
BIN Out of Bin- displayed when the called allele's peak apex is out of the bin boundary.	OB Out of Bin- Peak is within the marker range but outside of a bin.
PHR Peak Height Ratio- two alleles are present and the ratio between the lower allele height and the higher allele height is below a defined threshold.	IMB Peak Imbalance- Peak intensity does not exceed the minimum percentage of the major peak within the marker as set in the <i>Panel Editor Edit Marker</i> box.
AN Allele Number- the number of alleles exceeds the maximum defined number for the organism, or no alleles are found . That number is specified in the analysis method.	PL Beyond Ploidy- When the number of peaks identified within a marker exceeds the maximum number of peaks expected as set in the <i>Panel Editor Edit Panel</i> box. Note: The PL rule does not fire if no alleles are detected.
OS Off Scale- indicates an off scale result in the marker size range. <i>When an off scale peak is detected, GMID places a pink bar at the location in all dye colors. This feature is helpful when identifying pull up peaks.</i>	HI High Intensity- indicated that a peak's intensity approaches and/or exceeds the maximum peak intensity filter as set in the <i>Run Wizard Data Process</i> box. SR/SP Saturated (repaired)- indicates the presence of intense peaks with characteristic morphology are identified and "repaired" for allele calling. Saturated (pull up)- Intense peaks may cause "pull-up" or additional peaks to appear in other dye colors.
LPH Low Peak Height- indicates that the alleles are lower than the specified values and do not result in the proper intensity.	LO Low Intensity- fired when a single peak called below the Minimum Homozygote Intensity threshold because a second peak was detected above the specified percentage value, as set in the <i>Edit Marker</i> box
OVL Overlap- indicates that a peak in the overlap region is called twice.	BC Bin Conflict- indicates that a peak is within the range overlapped by two or more bins.
SPU Spectral Pull Up- fired when the marker signal contains bleed through peaks (pull-up peaks).	
BD Broad Peak- fired if the called allele's peak width is wider than a certain value. The default of this value is 1.5 base pair.	
CC Control Concordance- indicates that the designated control sample does not exactly match the defined alleles for this marker in the panel being scored.	
Note: OL in GMID denotes any allele which does not correspond with a peak in the allelic ladder. It is an allele call, not a rule firing.	OL Off Ladder- fired if a peak is outside of the marker range. Note: If a peak is within the marker range but does not correspond to a peak in the allelic ladder, GMKR reports the size of the fragment (in basepairs) and flags the analyst with the OB rule firing.
	LS Low Score- indicates a low Quality Score (based on Signal-to-Noise Ratio and Peak Morphology and the Pass, Fail, Reject ranges are set in the <i>Run Wizard Additional Settings</i> box).

In GeneMarker HID, the analyst can visualize a sample's electropherogram, peak details, allele calls, and rules fired all in one window (Figure 4). The Peak Table contains a column titled **Ht_Ratio**, containing the peak height ratios for heterozygote loci. This information is an assessment of the heterozygote imbalance at a locus. When viewing the electropherograms in GeneMarker HID, the bins are shown as brackets on the top and bottom of the plot, and the gray vertical bars are indicative of the detected peak's center. These indicators help the analyst visualize the center of the peak falling within the bin.

Figure 4. GeneMarker HID's Main Analysis Window

The right pane is the Allele Report, and the left pane is the Sample Tree. The middle pane displays the electropherogram with the Peak Table is beneath the electropherogram. One dye is displayed in the electropherogram below. The green box indicates passing allele calls. The yellow icon in the Allele Report, seen in Sample 2, indicates the presence of a low quality peak at TH01 (-A shoulder). Also seen in the Allele Report is a red, circular icon by the 20 allele detected in the D3S1358 marker range of the positive control; this allele is a pull up peak from a true allele at D8S1179 (not shown here).



The Allele Report can be customized to show only those alleles with low quality scores allowing the analyst to focus on the alleles requiring review, a feature which compares to the sorting abilities of GeneMapper *ID*. GeneMarker HID, however, does not assign a quality value to loci which do not contain any alleles. The allele column is simply blank which can create confusing results when utilizing the table setting to view low quality loci (Figure 5). It is not mandatory to use the table filter to accurately analyze the low quality alleles, but such filtering features enhance data analysis using an expert system.

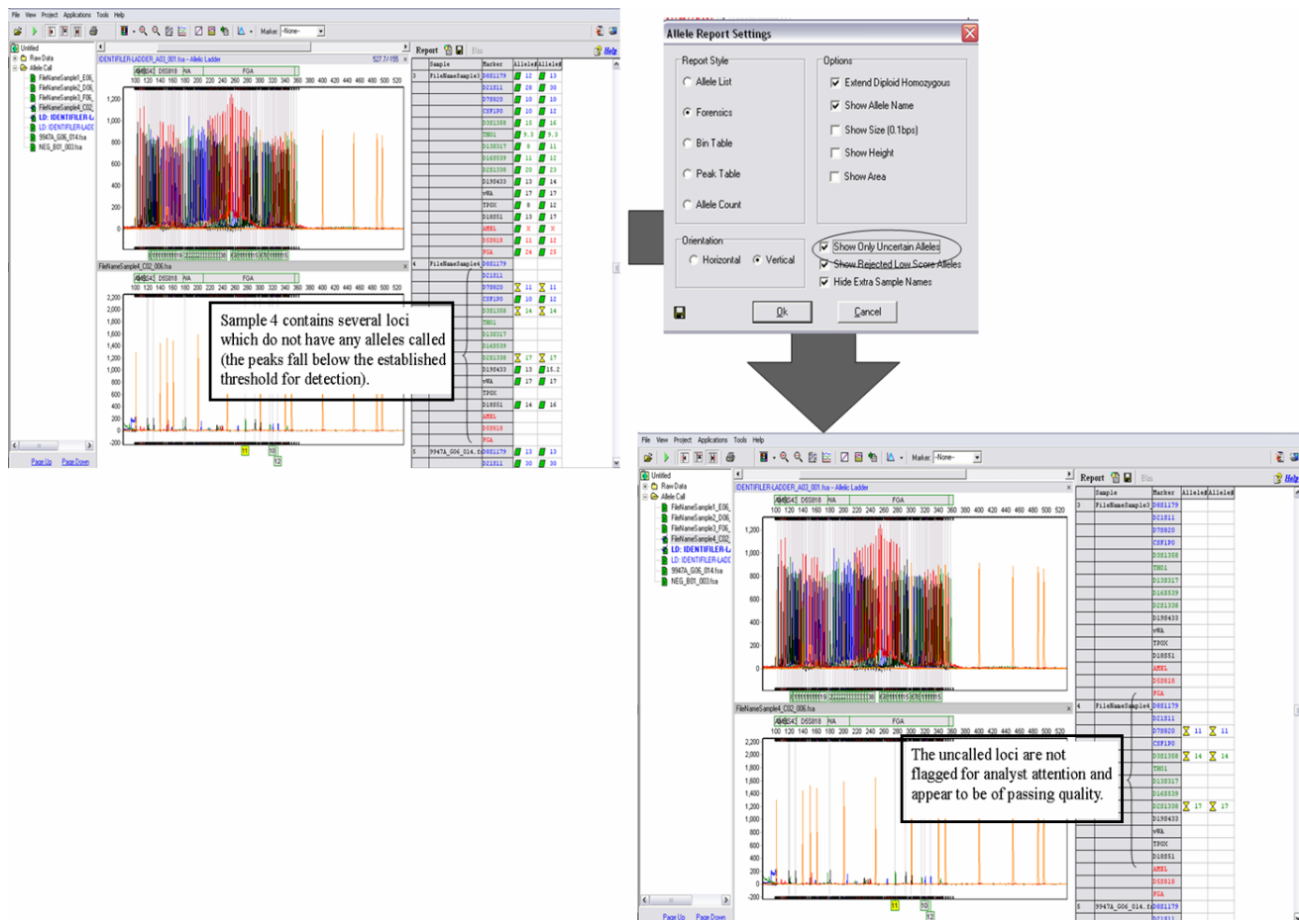
GeneMarker HID does not have an automated control concordance check like other expert systems. The user must manually confirm concordance of the controls with the Profile Comparison feature. The analyst is still responsible for verifying that the known profile is correctly genotyped using the expert system analysis.

GeneMapper *ID* displays the allele calls and genotype qualities in the Genotypes tab. The genotype table lists every analyzed locus in the sample set with the rule firings. This table is highly customizable. Using the Table Settings Editor, the user can create multiple table formats/templates for viewing the genotype results. For example, the table can be filtered to display only the DYS319 loci that have rules fired for Allele Number and Peak Height Ratio. Such filtering capabilities allow the user to sort and troubleshoot data efficiently. GeneMapper *ID* does not provide peak height ratios for heterozygote loci. The Peak Height Ratio rule fires when the set threshold is not met, but the specific ratios are not provided. GeneMapper *ID* automates positive control concordance check. Under the Panel Editor, the user designates the profile of the positive control for each panel. During analysis, the software automatically checks the allele calls for the positive

controls against the expected defined control alleles. If discordance is seen, the Concordance Check rule (CC) is fired. This feature expedites this mandatory quality control procedure and ensures that the positive controls are checked consistently and accurately for every analysis.

Figure 5. Allele Report Filter: *Show Only Uncertain Alleles*

GeneMarker HID allows the user to customize the allele report. The table can be filtered to Show Only Uncertain Alleles, allowing the analyst to focus on the low quality peaks. GeneMarker HID does not assign a quality value to loci without any allele calls. This example illustrates how this can be problematic if using the Show Only Uncertain Alleles filter. Sample 4 contains nine loci which do not have any peaks detected, four loci with passing genotypes and three loci with low quality genotypes. When the filter is applied, the green icons for all passing alleles are hidden and only the yellow and red icons are still seen. Those loci with no alleles detected, and thus no quality icon, appear the same as the passing loci. For example, TH01 failed to genotype yet it appears to be of the same quality as CSF1PO, a passing locus.



4.1.4 Allele Editing and Audit Trail

GeneMarker HID allows the user to edit allele calls. In the Peak Table of the Main Analysis screen, an allele call may be changed by opening the popup menu on any row of the allele calls.. This menu can also be accessed by right-clicking the allele in the electropherogram. Alleles can then be added, deleted or edited. Options of **Confirm** and **Confirm All** are additional in this program. The **Confirm** option labels the selected allele as *Pass*. **Confirm All** labels all alleles within the dye color as *Pass*; however, the initial Quality Reason can still be seen in the Peak Table. Once an allele has been edited, a pink indicator appears in the Peak Table. If the action is removed/negated, the pink indicator will still be seen, documenting that a change has been made. All allele edits can be viewed by selecting View History from the allele change popup menu. In order to View History, **Record Data Edit History** must be selected by the Administrator (under the User Management window of the Help menu) to activate this feature. GeneMarker HID features the Project Comparison application. The software will compare the allele edits and profile designations for a project from the reviews of two independent analysts. Inconsistent genotypes are flagged for further investigation. This feature expedites the review process.

GeneMapper *ID* also allows for allele edits. All edits are tracked in the View History log. The Electropherograms can be customized to either show or hide alleles that have been deleted. When hidden, the analyst can delete an allele call, leaving no evidence of the call on the electropherogram. It is not ideal that such changes can be hidden, even when there is documentation in the Allele Edit History. Additionally, there

is no administrative control of such settings, so analysts have access to turn this feature on or off.

4.1.5 Exporting Results

GeneMarker HID allows the user to customize the reports to be exported. In the Peak Table, columns can be filtered/added for the exported report. The Allele Table also has several formatting options available, including the option to remove the repeated file names. This feature makes sample distinction much easier in the report. All tabular reports can be exported as tab-delimited text (.txt) or directly exported as Microsoft Office Excel spreadsheet files (.xls). Since many users report in Excel, the ability to directly export as an Excel workbook is beneficial. GeneMarker HID enables the user to export the electropherograms and Print Reports as JPEG files. These images can then be further incorporated into other applications (such as Microsoft Word documents), or they can simply be saved as a paperless reference to the reports an analyst generates.

GeneMarker HID also contains a customizable report format for export to CODIS (CMF1.0 or CMF3.0 reports) which can be exported for upload into the CODIS database.

GeneMapper *ID* allows the user to export any of the tables created in the software. These tables are highly customizable using the **Table Settings** function found in the **GeneMapper Manager**. Tables are exported as tab-delimited text (.txt) files or comma-separated values (.csv) files. These files can then be opened and saved as spreadsheets. GeneMapper *ID* also includes a CODIS compatible export format. Allele reports can be exported as CMF1.0 or CMF3.0 for upload into the CODIS database. Electropherograms are not exportable from GeneMapper *ID*.

4.1.6 Printing Results

GeneMarker HID features a Print Report dialog box which allows the user to select the components of the final report. The Advance button allows for further customization of the printed report. The Preview button opens a Print Preview window, showing the user the exact format of the report. In this view, the user has options to change page settings, further alter report settings and export the report as a JPEG file. The printing options in GeneMarker HID are very user friendly, and they simplify the printing process by giving the user more control over the final product prior to printing.

4.1.7 Concordance of Results and Rules Fired

Table 2 is a detailed accounting and comparison of the number of allele calls and rule firings when using the different software programs and chemistry kits on the same instrument platform. The total number of actual alleles for each sample population was counted. The Identifiler samples contained 2,806 actual alleles, and the PowerPlex 16 samples contained 2,931 actual alleles. The total number of peaks detected (with and without a 20% global filter), including actual allele peaks and extraneous peaks, was calculated. Also, the number of actual alleles called without firing a rule was tabulated. When using an expert system, these calls are accepted as correct without analyst verification. The number of actual (true) alleles flagged for analyst intervention was totaled. These flagged alleles are correctly identified by the expert system but violate a rule (e.g., heterozygote imbalance, allele number). The number of extraneous alleles identified and not flagged for analyst attention was totaled. This result is a critical

measure of an expert system's effectiveness. Summing these four values (columns 5-8 in Table 2) for each system yields the total number of peaks detected.

Two hundred samples were analyzed in duplicate to evaluate concordance of allele calls. Initially, the analysis settings were established without applying a global filter to all loci. Table 2 reports the results for 100 Identifiler and 100 PowerPlex 16 samples without utilizing a 20% peak filter. Neither GeneMarker HID nor GeneMapper *ID* called an extraneous allele without flagging the analyst. The systems performed similarly when analyzing the Identifiler samples. GeneMarker HID called and flagged 174 peaks of the 2,864 peaks identified. Of the flagged peaks, 58 were extraneous alleles and 126 were actual alleles requiring analyst intervention. GeneMapper *ID* called and flagged 191 peaks of the 2,863 peaks identified. Of the flagged peaks, 57 were extraneous alleles and 134 were actual alleles requiring analyst intervention. For the PowerPlex 16 samples, GeneMarker HID called and flagged 953 peaks out of 3233 peaks identified. Of the flagged peaks, 302 were extraneous alleles and 651 were actual alleles requiring analyst intervention. GeneMapper *ID* called and flagged 1033 peaks out of 3253 peaks identified. Of the flagged peaks, 322 were extraneous alleles and 711 were actual alleles requiring analyst intervention. Based on the analytical results evaluated for the 100 PowerPlex 16 samples, GeneMarker HID required less analyst intervention and time than GeneMapper *ID*. Both expert systems can potentially increase the analyst's efficiency by correctly calling the majority of the alleles without intervention and correctly identifying problem samples quickly.

The PowerPlex 16 data analyzed in this study had high baseline due to poor color separation. Many samples, PowerPlex 16 and Identifiler, had -A shoulders because too

much template DNA was added to the amplification, leading to incomplete adenylation of the PCR product. The same data sets were analyzed a second time using a global filter of 20% with all other settings remaining constant. When processing high quality, high yield, single source samples, this filter is efficient at reducing the number of baseline and artifact peaks detected. Many laboratories responsible for high-throughput databasing use filtering of this type. Table 3 displays the analysis results using the global filter. Neither system incorrectly called a peak without flagging the analyst's attention when using the 20% filter. Remarkably, neither GeneMarker HID nor GeneMapper *ID* called any extraneous peaks in the Identifiler samples. The majority of rule firings in the 100 samples were due to PHR/IMB (Peak Height Ratio/Imbalance) results, where a locus displays preferential amplification. The remaining rule firings were mostly due to off-scale intensities, indicating that the filter is very efficient at removing the extraneous baseline and -A peaks. GeneMarker HID and GeneMapper *ID* performed similarly for the PowerPlex 16 data. The 20% filter greatly reduced the number of baseline peaks and artifacts (-A shoulders) detected and called in the PowerPlex 16 samples. Many laboratories may not be comfortable applying a non-specific filter to all loci. Note that GeneMapper *ID* has the option to filter -A shoulder peaks specifically. This feature was not utilized in this study but would have improved the GeneMapper *ID* results without applying the 20% global filter.

Off-scale peaks cause pull-up peaks in other dye colors due to spectral overlap. GeneMapper *ID* recognizes detector (CCD) saturation and flags the high intensity peaks as off-scale (OS). The system then highlights the areas where pull-up peaks may be seen with a pink bar; these peak(s) are flagged for analyst attention since off-scale data may

present problems for interpretation, especially due to spectral pull up. In contrast, GeneMarker HID fires the High Intensity rule (HI) when a peak exceeds the assigned maximum detection threshold. The program also recognizes the truncated morphology of a saturated peak and will “repair” the peak for allele calling. When a peak has been repaired, the SR (Spectral Repair) and SP (Spectral Pull Up) rules fire to alert the analyst of off-scale data and the resulting possibility of pull-up peaks. Pull-up peaks are automatically corrected when processing the raw data and are removed completely when the Pull Up Correction feature is selected. For this study, the Pull Up Correction feature was selected. Several loci were flagged by GeneMapper *ID* for OS peaks and Spectral Pull Up (SPU), but the loci were not flagged in GeneMarker HID because either the pull-up peaks were removed during raw data processing or the high intensity peak did not exceed the maximum detection threshold (set at 8000RFU) (Figure 6). It is recommended that the saturation limit for detection be validated through independent laboratory studies when using an expert system that applies a threshold to ensure that high intensity samples are adequately examined prior to assigning the genotype.

Due to differences in smoothing algorithms, which function to reduce the noise level of the baseline, different software systems do not report the same peak heights. These differences must be taken into consideration when adopting an expert system. In this case, GeneMapper *ID* consistently reports lower peak heights than GeneMarker HID. In this study, all samples were analyzed using an established peak detection and allele calling threshold of 100 RFU. On occasion, GeneMapper *ID* did not detect a baseline “noise” peak because it falls beneath the established threshold of 100 RFU for peak detection and allele calling. In contrast, GeneMarker HID does detect and call this peak

since the peak height assigned by this software is higher. The “noise” peak was measured as 130 RFU in GeneMarker HID as opposed to 95 RFU as in GeneMapper *ID* (Figure 7). Thresholds should be established as they apply to the software being utilized as opposed to adopting previously used thresholds. Such settings should be determined through a laboratory’s independent validation study for adopting an expert system.

Figure 6. Rule Firings for Off Scale Data in GeneMapper *ID* and GeneMarker HID
GeneMapper *ID* detects CCD saturation and alerts the analyst of such data using magenta bars and the Off Scale (OS) rule firing. GeneMarker HID does not have this capability. The High Intensity (HI) rule fires when data exceed a user-defined threshold, set at 8000RFU in this example. Additionally, GeneMarker HID has a pull up correction feature. The pull up peak seen in the GeneMapper *ID* window is not present in the GeneMarker HID window.

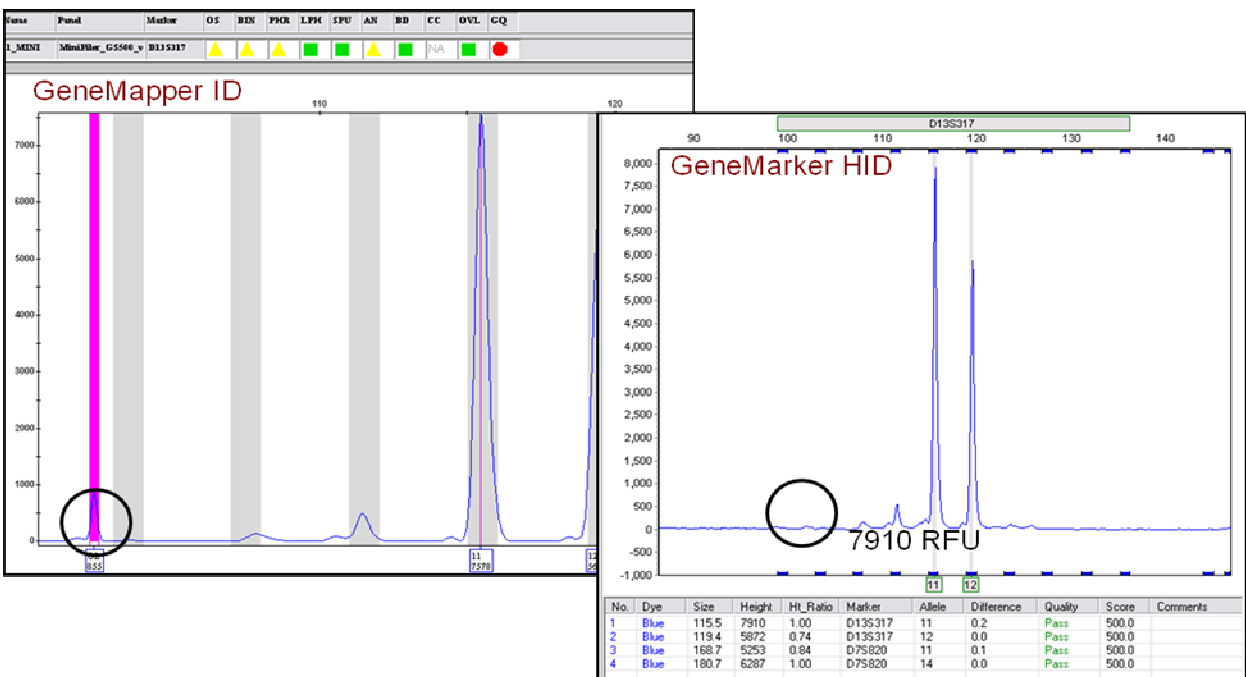
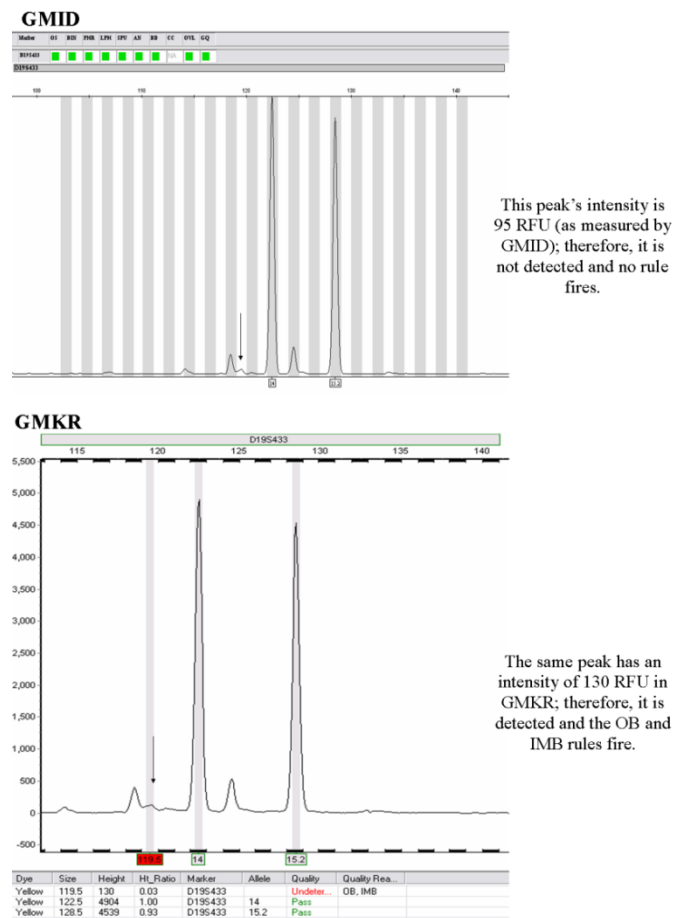


Figure 7. Non-concordant Call between GeneMarker HID and GeneMapper ID

The following figure is an example of GeneMapper ID (GMID) and GeneMarker HID (GMKR) having non-concordant peak calls. Since the reported peak heights are not consistent between the two systems, the baseline peak is detected with GeneMarker HID and is not detected with GeneMapper ID. Note that the stutter peaks are filtered appropriately in both expert systems.



Several loci violate the heterozygote peak balance rule in one program but not the other. These inconsistencies are also a result of differences in smoothing algorithms. On several occasions, only one expert system fired a rule for heterozygote peak imbalance (Figures 8 and 9). In one example, GeneMapper ID fired a rule for PHR violation, but the same locus did not fire the IMB rule in GeneMarker HID, although the filtering parameter was set at 70% for both expert systems (Figure 8). The respective rule should fire when a heterozygote peak is less than 70% of the major peak. The two peaks were reported as 280 RFU and 401 RFU in GeneMapper ID, which is a ratio of 69.82%. In GeneMarker HID, the two peaks were measured as 293 RFU and 417 RFU, and the system calculated the ratio as 70% (calculated as 70.26%). The PHR rule was fired in

GeneMapper *ID* but not in GeneMarker HID because of the differences in peak intensities. A similar inconsistency in PHR/IMB rule firings was observed in another case, except GeneMarker HID was the system which fired the IMB rule and GeneMapper *ID* did not. The smoothing algorithms reduced baseline noise but altered major peak intensities (Figure 9).

Figure 8. Inconsistent Rule Firing; Peak Height Ratio rule fired in GeneMapper *ID*

In this example GeneMapper *ID* flags the locus for PHR (Peak Height Ratio) violation but GeneMarker HID does not fire the analogous rule, IMB (Imbalance). The ratios of the peak intensities as reported by GeneMapper *ID* and GeneMarker HID are 69.82% and 70.26% respectively. Both systems fired the rule appropriately according to the established settings.

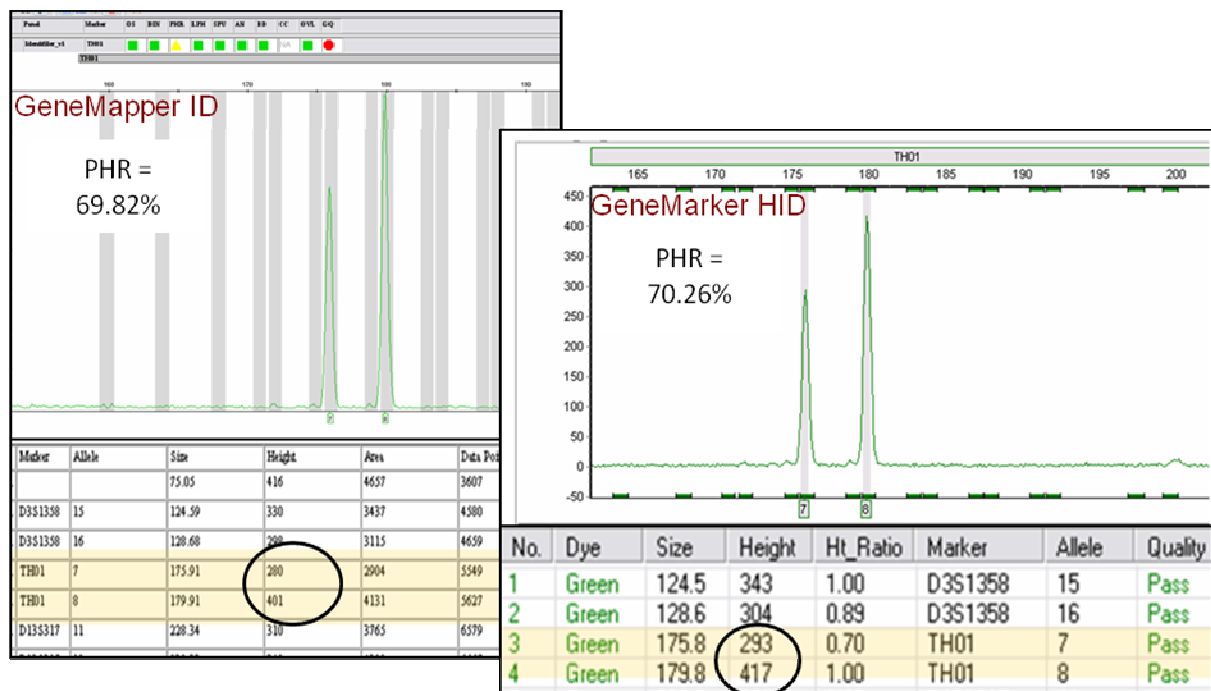
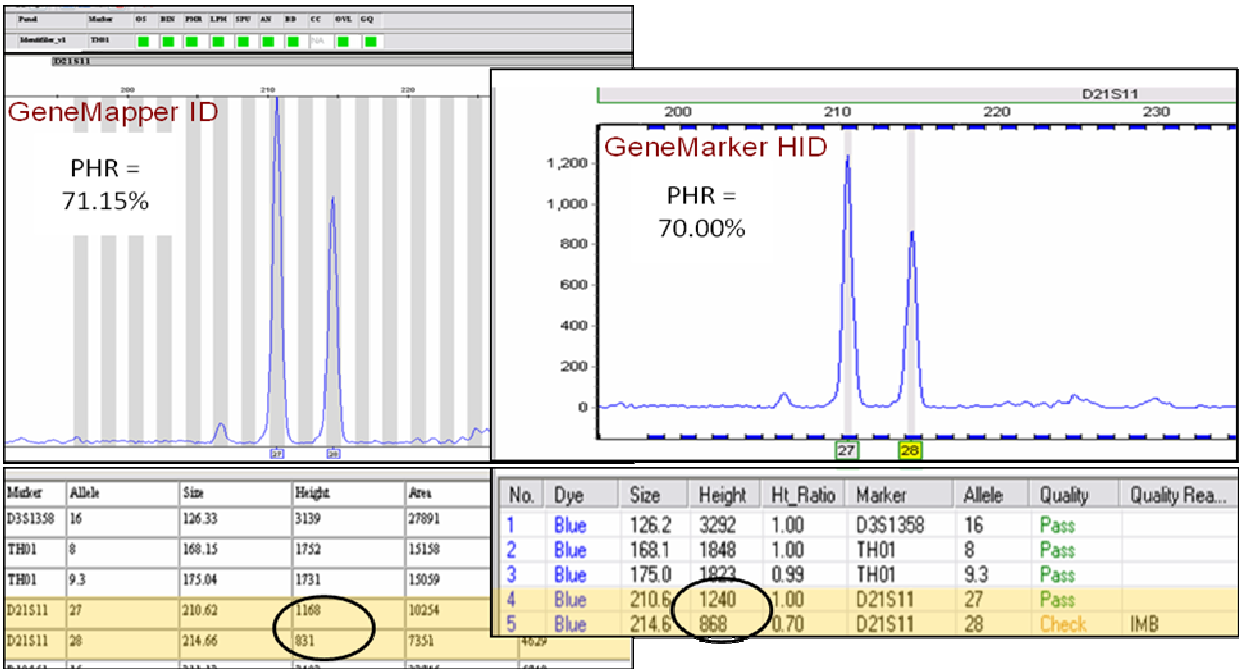


Figure 9. Inconsistent Rule Firing; Imbalance rule fired in GeneMarker HID

In contrast to Figure 6, this example shows a locus in which GeneMapper *ID* does not fire the PHR (Peak Height Ratio) rule but, GeneMarker HID fires the analogous IMB (Imbalance) rule. The ratios of the peak intensities as reported by GeneMapper *ID* and GeneMarker HID are 71.15% and 70.00% respectively. Both systems fired the rule appropriately according to the established settings.



4.2 *Internal Validation of GeneMapper ID as an Expert System for High Throughput Analysis of Single Source Samples*

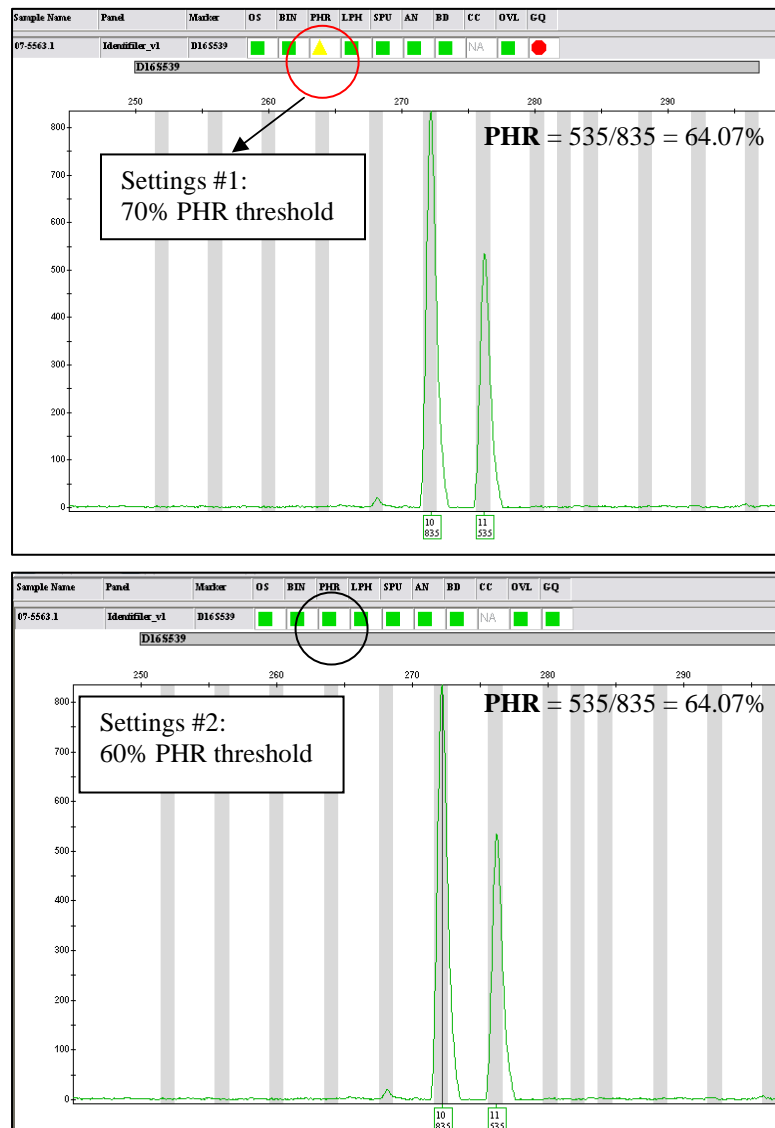
The preliminary comparison/evaluation of GeneMarker HID and GeneMapper *ID* confirmed that both software programs are viable expert systems. Although GeneMarker HID offers many unique features and a user-friendly interface, the UNTCHI R&D team opted for GeneMapper *ID* as their expert system for the immediate future. This decision was primarily based in the fact that GeneMapper *ID* is already available and familiar to the UNTCHI R&D analysts. Since the analysts already have a strong working knowledge of the software, the learning curve for implementing this expert system is greatly decreased. In addition, the complete set of Identifiler/3130xl/GMID has already been developmentally validated, requiring only an internal validation for incorporation. Adopting a novel expert system such as GeneMarker HID, requires a much more extensive developmental validation, which the UNTCHI R&D Laboratory did not have time to conduct prior to the impending population databasing project.

4.2.1 Internal Validation of GeneMapper *ID* for Identifiler Data: Calibration

Table 3 in Chapter 3 describes the expert system parameters as they were applied sequentially to the Calibration Sample Set. Parameters highlighted in purple are new or modified from the previous run. If a setting is accepted, it appears in blue for the subsequent runs. The starting point, Settings #1, is representative of the current parameter settings utilized by the Forensic Laboratory for manual interpretation. 81.83% of the loci in this sample set qualified as “pass” by GeneMapper *ID*. Of the 2961 passing loci, 99.93% (2959) of the genotypes were called correctly by the expert system without rule firings (discussion of the 2 discordant loci found on pages 40-43). For Settings #2,

the peak height ratio threshold was decreased to 60%. Many loci were flagged for analyst's attention due to this violation alone (Figure 10). Decreasing the threshold by 10% allowed for 28 additional loci to be called as "pass", and they were correctly genotyped by the expert system. This adjusted parameter increased the efficiency of the expert system to 82.61%.

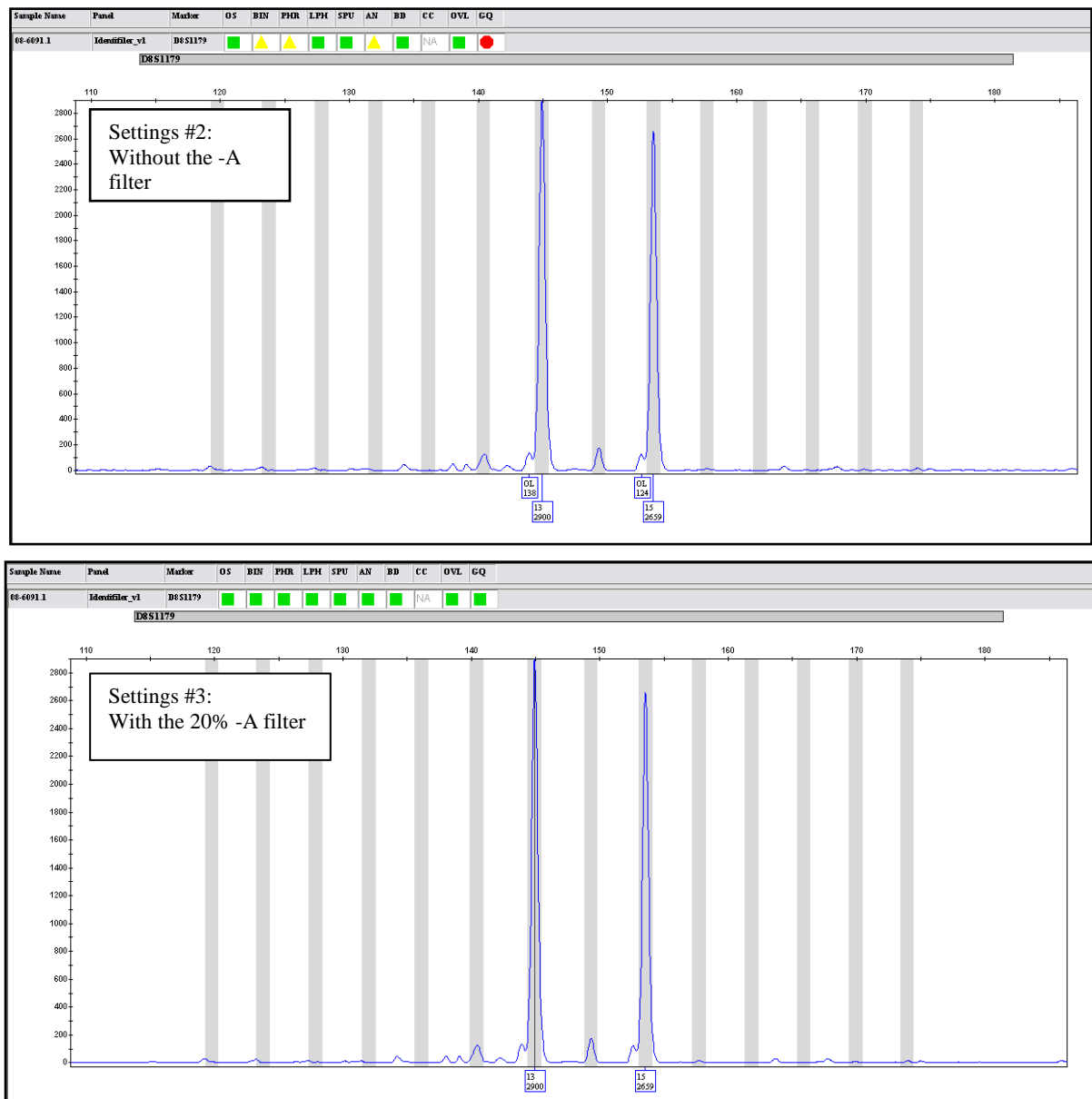
Figure 10. Peak Height Ratio (PHR) Rule Firing: Settings #1 vs. Settings #2
The top panel illustrates a rule firing for Peak Height Ratio (PHR) under Settings #1 which employs a 70% PHR threshold. These peaks represent a true heterozygous result at D16S539, indicating that the PHR threshold is possible set to high. By decreasing the threshold to 60%, the result seen in the bottom panel was obtained.



-A shouldering was repeatedly detected in the Calibration Sample Set. Settings #3 added a 20% -A peak filter (Figure 11). The expert system correctly genotyped an additional 89 loci when this filter was employed, increasing the efficiency of the expert system to 85.07%.

Figure 11. GeneMapper ID's -A Filtering Capacity: Settings #2 vs. Settings #3

The top panel illustrates a result obtained under Settings #2. The -A shoulders are detected, causing rule firings for a peak out of bin (BIN), peak height ratio (PHR), and allele number (AN). These rules are resolved by applying a -A filter, as seen in the lower panel.



Settings #7 were selected for the concordance phase of the validation, which include the decreased peak height ratio threshold and a -A filter. These settings also included a 7.5% global filter to decrease the number of extraneous baseline peaks called as alleles. The global filter allowed for an additional 142 loci to be correctly genotyped by the expert system. Under these settings, the expert system was most efficient at calling alleles correctly (88.44% of total loci qualified as “pass” and were concordant) while filtering extraneous and/or artifact peaks. It should be noted that this percentage is lower than one would expect for an optimized expert system. As described in Chapter 2.1, the calibration phase sample set (N=226) included 62 databasing samples from a visiting scientists work in the UNTCHI R&D Laboratory. The concentration of these samples was not determined prior to Identifiler amplification, and the resulting sample set contained a high degree of off-scale data. This data caused a decrease in the optimization during calibration phase. This data set is not representative of the data produced by the UNTCHI R&D Laboratory

In Table 3, discordant loci are documented under multiple settings. These loci (2) are from two different samples, and each sample was independently reviewed. An increase in peak height imbalance due to sub-optimal input DNA for amplification caused stochastic effects, leading to a higher degree of heterozygote peak imbalance. The overall signal of these two samples was uncharacteristically low (Figures 12 and 13). In both cases of discordance, the more intense peak was greater than the minimum homozygote intensity threshold (200 RFU), and the lesser peak was below the detection threshold (100 RFU). GeneMapper *ID* called each locus as homozygote and no rule fired because the lesser peak was not detected.

Figure 12. Sample Electropherogram (59F), Discordant at Amelogenin

At the Amelogenin locus for this sample, the intensity of the X allele was measured at 88 RFU. The Y allele intensity is reported at 266 RFU. The expert system incorrectly genotyped this locus without firing a rule because the Y peak is greater than 200 RFU (the minimum homozygote intensity threshold) and the X peak is less than 100 RFU (the peak detection threshold).

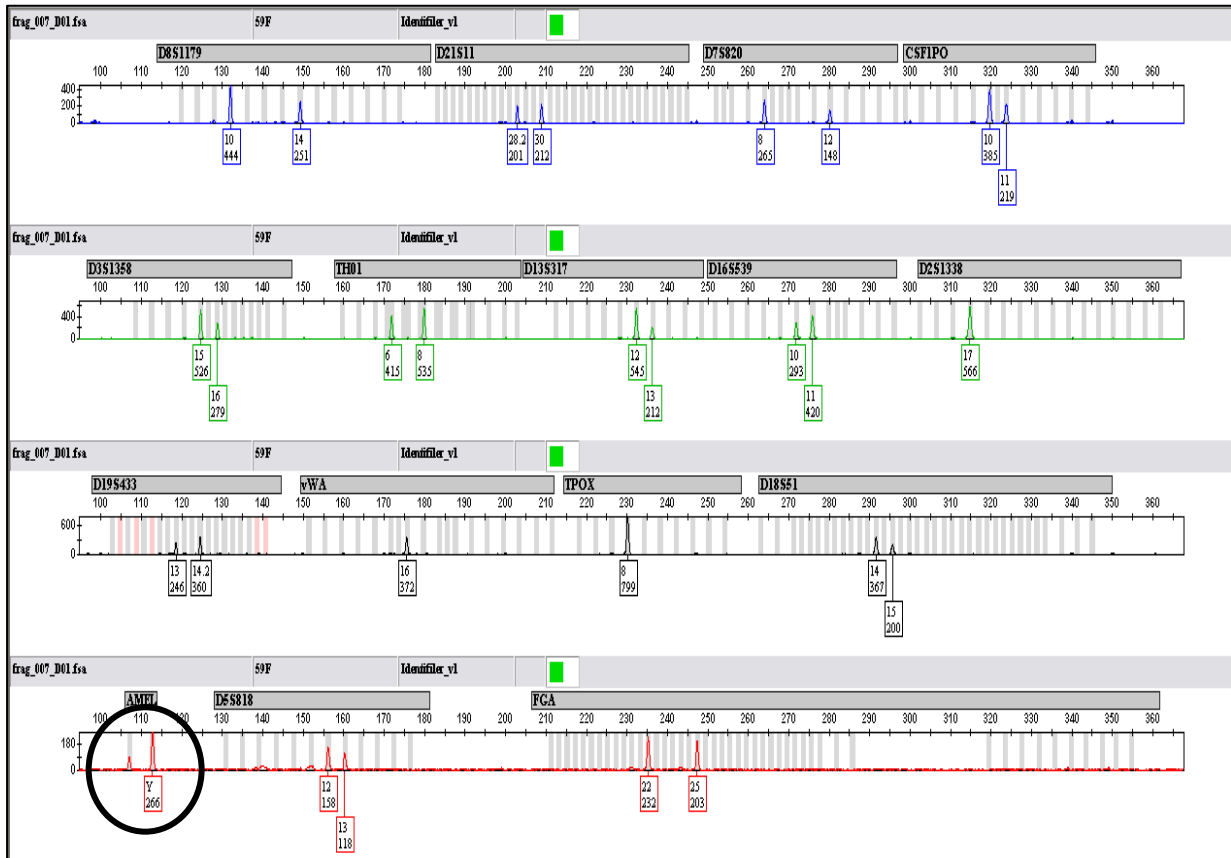
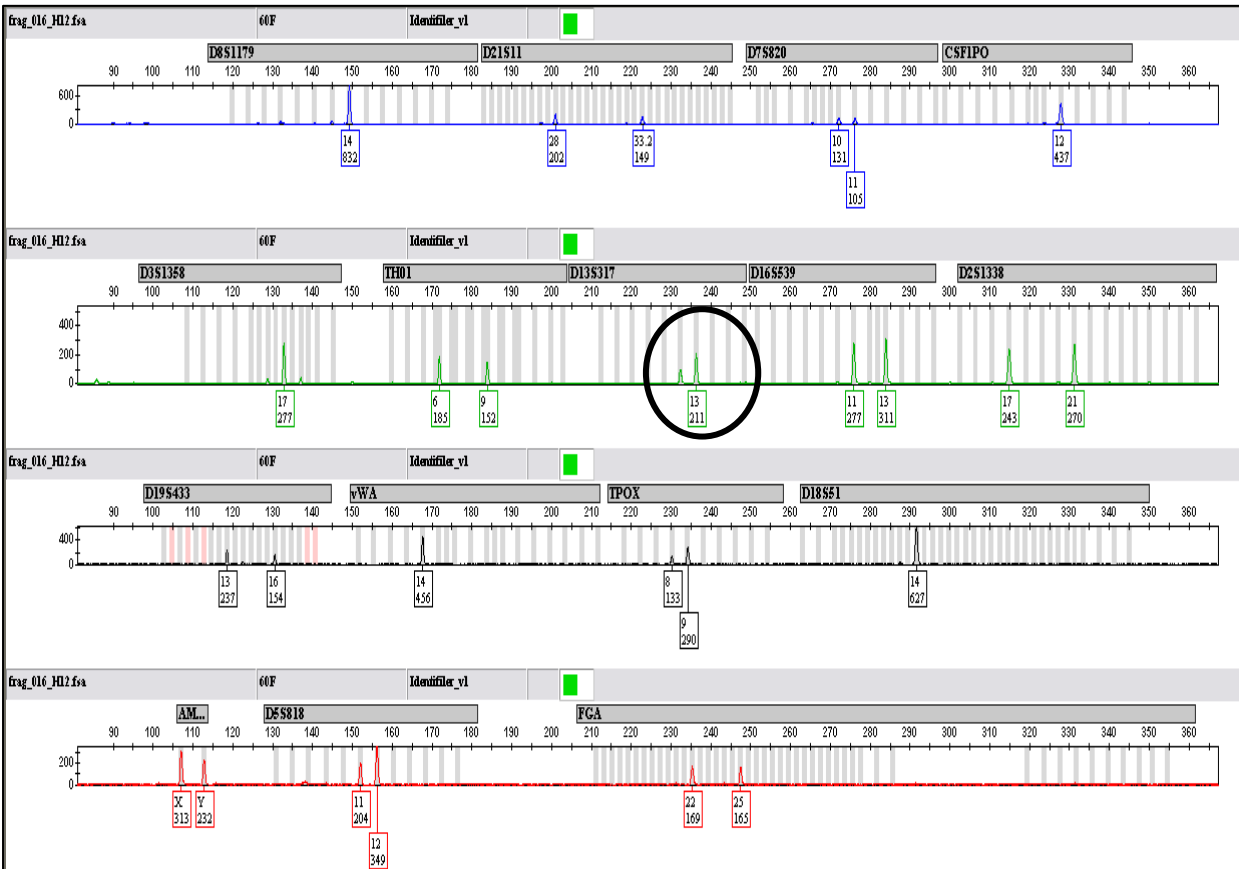


Figure 13. Sample Electropherogram (60F), Discordant at D13S317

At the D13S317 locus for this sample, the intensity of the 11 allele was measured at 99 RFU. The 13 allele intensity is reported at 211 RFU. Because the 13 allele is greater than 200 RFU (the minimum homozygote intensity threshold) and the 11 allele is less than 100 RFU (the peak detection threshold), the expert system incorrectly genotyped this locus without firing a rule.



Successful attempts were made to resolve the discordant genotypes (Settings #4-#6), but the efficiency of the expert system was compromised. Settings #5 decreased the peak detection threshold in order for the expert system to detect the lesser peak. Changing the peak detection threshold to 75 RFU resulted in GMID firing a rule at both loci for peak height ratio (PHR), resolving the discordant “pass” call. However, the efficiency of the expert system is significantly decreased due to increased baseline

detection. Settings #4 and #6 were attempts to resolve the discordant calls by increasing the minimum homozygote intensity threshold. This approach resolved the discordant “pass” calls as well because GMID fired a rule for low peak height (LPH) and the expert system efficiency was not decreased as significantly as with Settings #5. However, since these samples are not characteristic of the data produced in the UNTCHI Laboratories, they are not considered of great consequence for the optimization of GeneMapper *ID* as an expert system. Considering also the significant amount of flexibility in the stringency of searches within reference databases, it can be expected that a successful hit/match can be obtained when allele dropout is undetected at only one locus. For these reasons, the concordance phase commenced using Settings #7.

Note: Two additional loci were flagged as discordant in the concordance check spreadsheet (not provided here) due to differences in allele notation between the manual call and the expert system call. When the allele size is greater than the standard range for that locus, the Forensic laboratory uses a greater-than (>) notation, whereas the expert system calls the allele based on the number of STR repeats detected. Since the expert system correctly called the allele, these are not true inconsistencies.

4.2.2 Internal Validation of GeneMapper *ID* for Yfiler Data: Calibration

Table 4 of Chapter III describes the expert system parameters as they were applied sequentially to the Yfiler Calibration Sample Set. Parameters circled in blue are new or modified from the previous run. If a setting is accepted, it is retained in the settings for subsequent runs. The starting point, Settings A, is the starting point for calibration of GeneMapper *ID* for Yfiler data. These settings were established in a preliminary study (data not included here) using 133 population samples amplified and analyzed in the UNTCHI R&D Laboratory. 91.31% of the loci in this sample set qualified as “pass” by

GeneMapper *ID* and all were concordant. For Settings #2, the peak height ratio threshold was decreased to 40% because the expected number of alleles was changed to 1. The peak height ratio threshold was made less stringent since there is only one locus with two peaks. This change increased the expert systems efficiency slightly; these settings were implemented to add more confidence in the allele assignments for all loci other than DYS385. The result is a rule firing for DYS385 every time there are two alleles detected. In order to counter these rule firings, a table setting was established in the genotypes tab which selects the DYS385 loci that are flagged solely because there are two peaks (Figure 14).

Figure 14. The DYS385 Passing Table

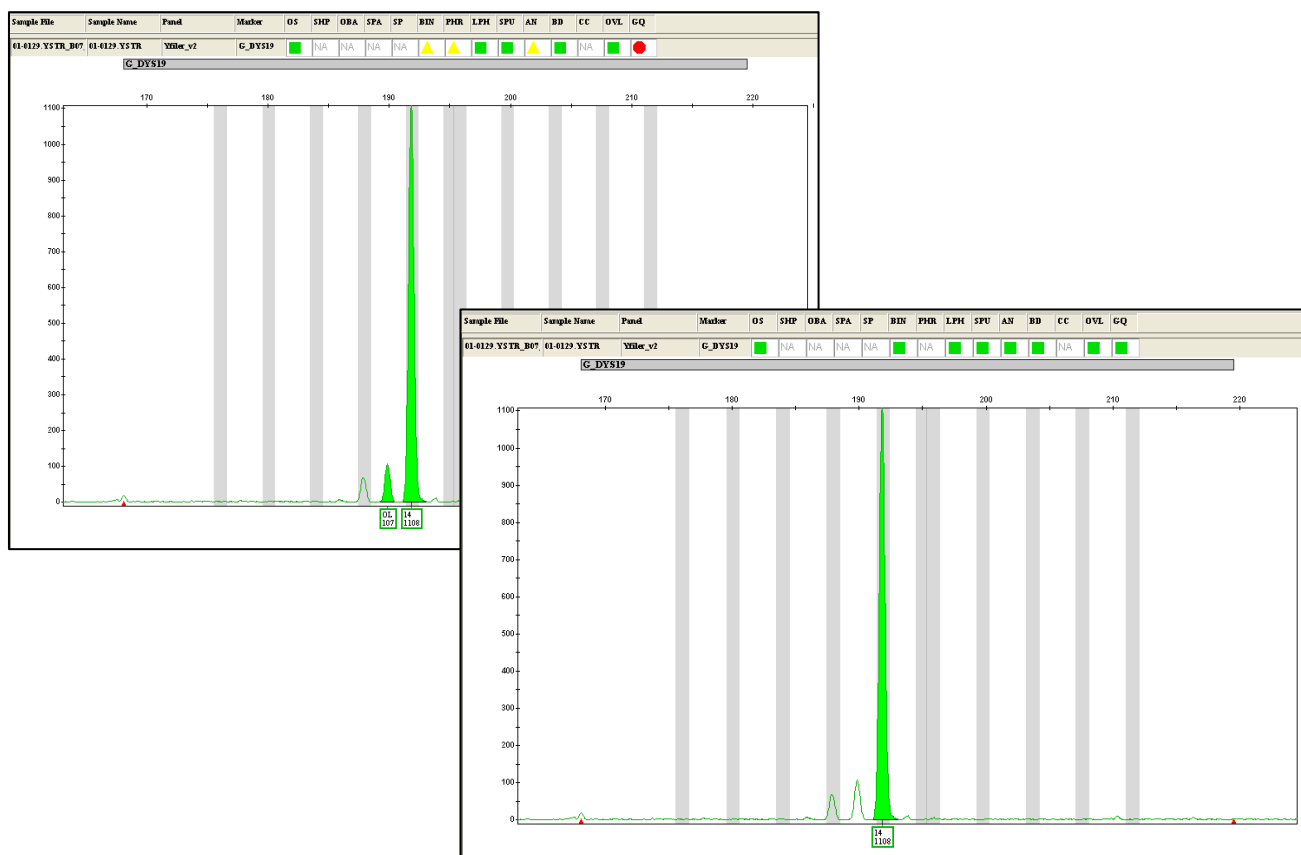
Using the filtering capabilities of GeneMapper *ID*, the Genotypes table can be set to display only the DYS385 loci that only have the Allele Number (AN) rule firing. The analyst can export these allele calls as true/correct, even though the Genotype Quality (not shown here) is still red.

Table Setting: DYS385 passing																			
Samples	Genotypes																		
	Sample Name	Run Name	Panel	Marker	Dye	Allele 1	Allele 2	AE Comrr	AE Comrr	ADO	AE	OS	BIN	PHR	LPH	SPU	AN	BD	
39	01-0121.YSTR	0358_Batch2.YS	Yfiler_v2	G_DYS385	G	11	14					■	■	■	■	■	▲	■	
71	01-0123.YSTR	0358_Batch2.YS	Yfiler_v2	G_DYS385	G	11	14					■	■	■	■	■	▲	■	
87	01-0124.YSTR	0358_Batch2.YS	Yfiler_v2	G_DYS385	G	16	17					■	■	■	■	■	▲	■	
100	01-0126.YSTR	0358_Batch2.YS	Yfiler_v2	G_DYS385	G	13	14					■	■	■	■	■	▲	■	
119	01-0127.YSTR	0358_Batch2.YS	Yfiler_v2	G_DYS385	G	13	17					■	■	■	■	■	▲	■	
151	01-0128.YSTR	0359_Batch2.YS	Yfiler_v2	G_DYS385	G	11	14					■	■	■	■	■	▲	■	
167	01-0129.YSTR	0359_Batch2.YS	Yfiler_v2	G_DYS385	G	11	15					■	■	■	■	■	▲	■	
199	01-0131.YSTR	0359_Batch2.YS	Yfiler_v2	G_DYS385	G	14	16					■	■	■	■	■	▲	■	
215	01-0132.YSTR	0359_Batch2.YS	Yfiler_v2	G_DYS385	G	11	14					■	■	■	■	■	▲	■	

Settings #2 increased the filtering distance for - A shouldering to -2 basepair. The filter threshold was maintained at 20%. DYS19 exhibits a documented -2 stutter artifact which was detected recurrently (Figure 15). By increasing the 20% filter distance to -2 basepair, the majority of these rule firings were eliminated which increased the efficiency of GeneMapper *ID* to 95.97%.

Figure 15. The DYS19 -2bp Artifact: Settings #1 vs. Settings #2

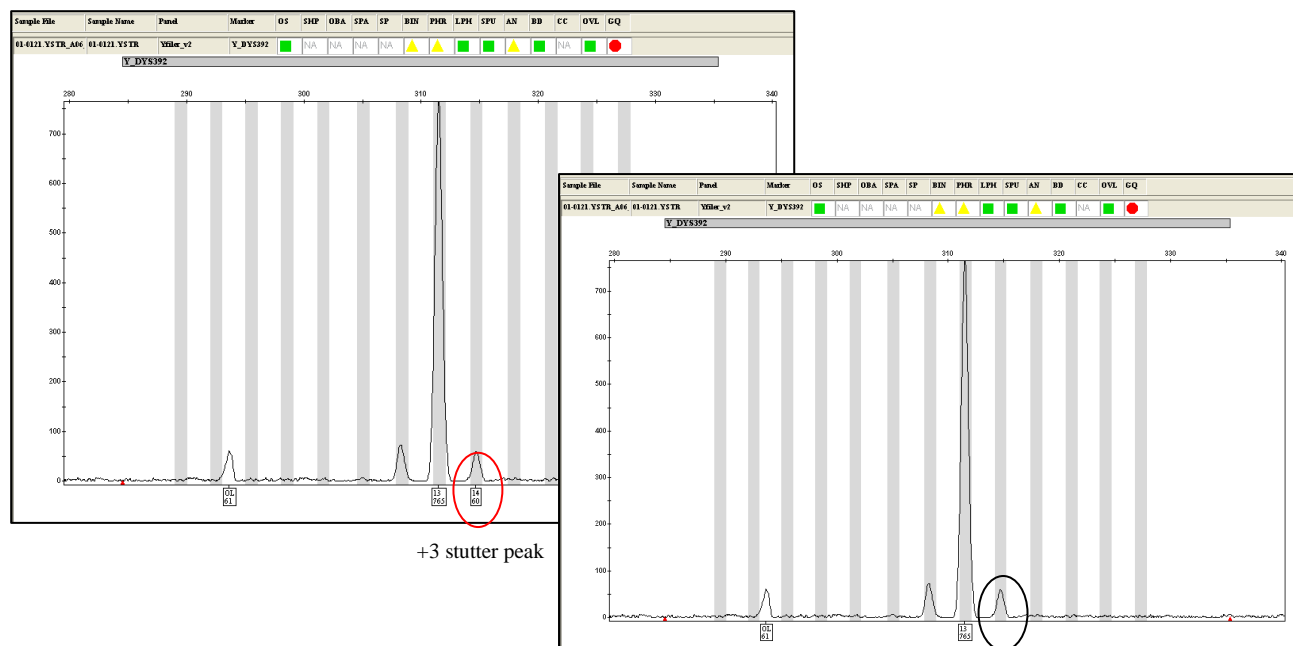
The left panel illustrates a -2bp stutter peak detected under Settings #1. The Out of Bin (BIN), Peak Height Ratio (PHR), and Allele Number (AN) rules fired for this locus. In contrast, the peak was effectively filtered under Settings #2, and the locus was not flagged for analyst intervention.



Settings #3 incorporate a +3 stutter filter for the trinucleotide locus DYS392, since loci with smaller repeat units exhibit a higher degree of plus stutter. Although the artifact was not a large contributor to the rule firings, there were several samples in which the +3 stutter peak was detected, and a rule fired as a result (Figure 16).

Figure 16. The +3bp Stutter Peak Detected at DYS392: Settings #2 vs. Settings #3

The left panel illustrates a +3bp stutter peak detected under Settings #2. The Out of Bin (BIN), Peak Height Ratio (PHR), and Allele Number (AN) rules fired for this locus. In contrast, the peak was effectively filtered under Settings #3. The Genotype Quality (GQ) is still red because of the Off Ladder (OL) pull up peak present.



Many stutter peaks were detected in this sample set, even when the marker specific stutter ratios were applied. This was especially evident at the tetranucleotide loci and DYS392.

Settings #4 increased the stutter percentage beyond the marker specific stutter ratios that are programmed in the software. For the tri- and tetranucleotide loci, the stutter ratio was increased to 15%. For the penta- and hexanucleotide loci, the stutter percentage was set at 10% since loci

containing larger repeat units tend to exhibit a lower degree of stutter. The ratio of loci called correctly by GeneMapper *ID* was then increased to approximately 96%. The final settings, Settings #5 increased the global filter to 12% (from 7.5%). This allowed for more efficient filtering of high baseline peaks and other artifacts such as spectral pull up. Since this global filter is applied across the entire locus, any peak that is less than 12% of the primary peak in the marker range is filtered (stutter and -A peaks included). For that reason, the previously employed 10% stutter filter for the penta- and hexanucleotide loci is overridden. Since it was no longer the parameter filtering the stutter peaks, the stutter value was set at zero. Under these settings, the ratio of loci called correctly by GeneMarker *ID* was finally optimized at 98%.

4.2.3 Internal Validation of GeneMapper *ID*: Concordance Phase

For the internal validation's concordance phase of Identifiler/3130xl/GeneMapper *ID*, two sample sets were analyzed under Settings #7. Table 5 in Chapter 3.3 provides the percent of samples and loci that were correctly genotyped by GeneMapper *ID*. Notably, no discordant loci were seen during this phase for either sample set. The first sample set contained family reference samples from the Forensic Laboratory. Settings #7 proved highly optimized for this sample set, with over 95% of the loci called correctly by GeneMapper *ID*. A second sample set was analyzed for the concordance phase, 257 population samples amplified and analyzed under the UNTCHI R&D's standard operating procedures. This sample set was used to verify the accuracy and optimization of Settings #7 on data typical of this laboratory. These settings also proved effective for the population samples, with over 93% of the loci called correctly by GeneMapper *ID*.

Note: There are two loci which appear to be discordant in the concordance check spreadsheet (not provided here). These were independently reviewed, and it was determined that the cause was a difference in notation when calling alleles which size outside of the designated range for that locus. The expert system correctly called the alleles and these are not true inconsistencies.

The internal validation's concordance phase of Yfiler/3130xl/GeneMapper *ID*, a sample set of 337 population samples were analyzed under Settings #5. Table 5 in Chapter 3 provides the percent of samples and loci that were correctly genotyped by GeneMapper *ID*. Notably, no discordant loci were seen during this phase for either sample set.

CHAPTER 5

CONCLUSIONS

This study evaluated the performance, workflow, user-interface and features of GeneMarker HID and GeneMapper *ID*. The role of an expert system is to alert the analyst of extraneous peaks and rule violations; we closely evaluated the allele calls made using these expert systems. The concordance study demonstrated the ability of GeneMarker HID to accurately designate alleles as compared to routine manual review. GeneMapper *ID*, an expert system familiar to the forensic community, was also evaluated in the concordance study. Both expert systems appropriately fires rules when loci contain peaks that violate the established parameters. Out of the two hundred samples included in the preliminary study, neither system called any incorrect peaks without firing a rule; therefore, they performed as expected for an expert system. The two systems performed similarly in the number of alleles correctly called without requiring analyst intervention. The workflow and user-interface of GeneMarker HID is intuitive. The Main Analysis Window is comprehensive, displaying the electropherogram, peak data, rule firings, and allele table all in one window. GeneMarker HID also offers many unique features: the *Run Wizard* prompts the analyst through analysis setup; the Size Calibration Window includes statistics on the 250 base pair internal sizing standard peak; the Project Comparison feature facilitates quick comparison of second reads; and many printing and exporting options, including electropherogram export as a JPEG file. However, GeneMarker HID does not automatically verify the positive control profile which is performed by the GeneMapper *ID* software. Additionally, GeneMarker HID analyzes all samples within the run

using the same analysis settings, panel, and size standard (whereas, GeneMapper *ID* allows the analyst to custom define the parameters for each sample individually). Since the two software programs performed so similarly, both are presented as a viable expert system option and offers many user-friendly features and an intuitive user interface.

GeneMapper *ID* was selected for the UNTCHI R&D Laboratory and was internally validated according to the established standards. The internal validation of GeneMapper *ID* was successful. Through a series of calibration tests for both Identifiler and Yfiler data, the system was optimized. These optimized settings were then applied to a larger concordance sample set, yielding a 100% concordance rate between the passing genotypes assigned by GeneMapper *ID* and calls made thorough manual review. For the Yfiler data set, the table filtering abilities of GeneMapper *ID* proved very useful in optimizing the expert system. Appendix C provides the protocol for using GeneMapper *ID* as an expert system for Identifiler and Yfiler data. The UNTCHI R&D Laboratory has incorporated the protocol.

Expert systems are a needed addition to laboratories conducting high throughput processing of reference samples and databasing projects. This technology greatly decreases the time required for DNA data analysis by automatically sizing and calling alleles of a genetic profile using a strict set of rules and thresholds that must be met. Expert systems allow analysts to focus solely on challenging samples, thus increasing laboratory efficiency, decreasing the probability of human error and reducing cost. The sizable backlog of reference samples is motivation for laboratories to adopt an expert system.

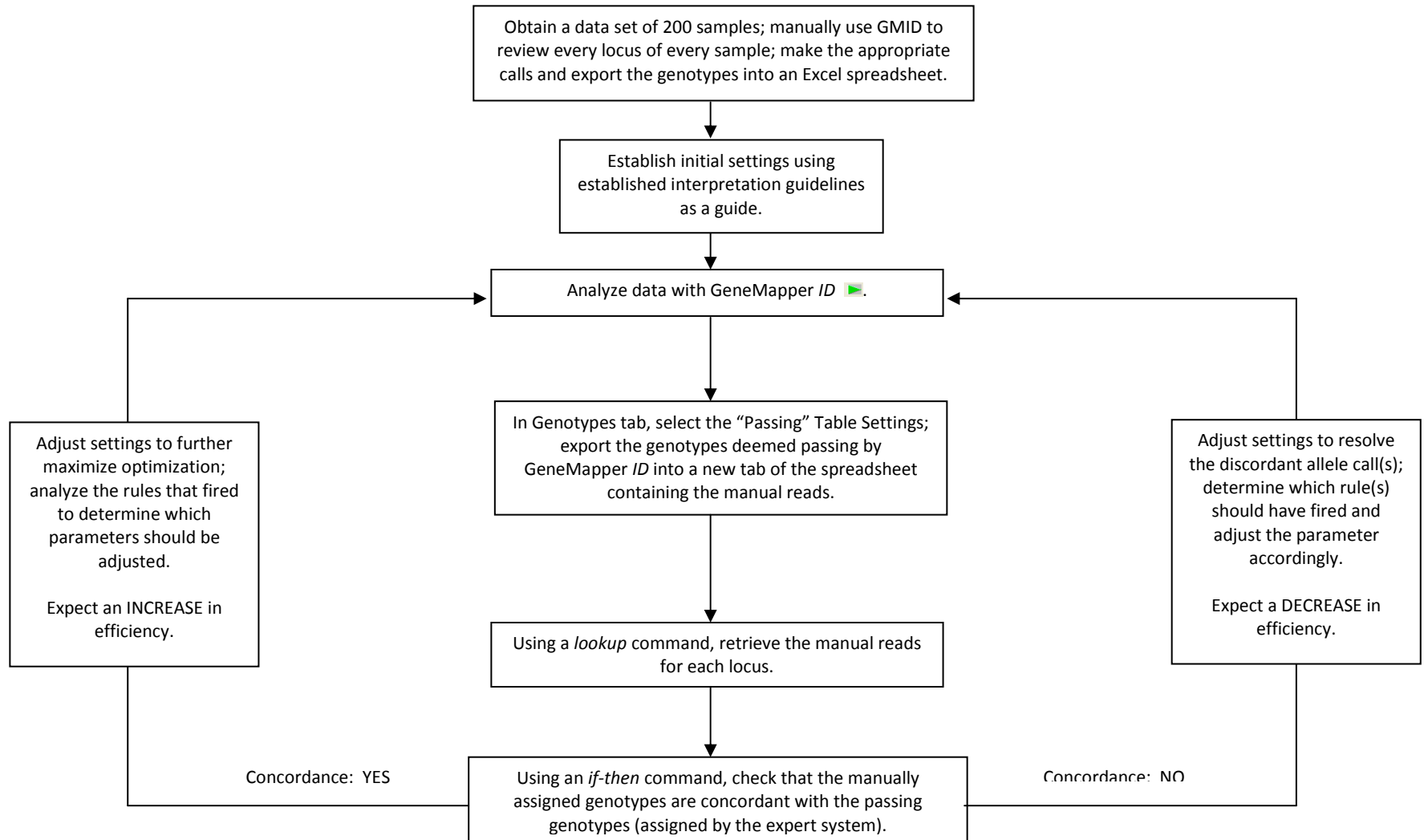
REFERENCES

1. Knowledge-based systems in Japan: Introduction [homepage on the Internet]. Maryland: Loyola College. May 1993. Available from: http://www.wtec.org/loyola/kb/c1_s1.htm.
2. Roby RK, Tincher BM. Expert systems: High-throughput analysis of single source samples for forensic DNA databasing. 218. Huntington, WV: U.S. Department of Justice; May 2006.
3. Perlin MW. In: Real-time DNA investigation. Proceedings of the sixteenth symposium on human identification: 2005; November 4, 2005; Grapevine, TX. Madison, WI: Promega Corporation.
4. Roby RK, Jones JP. Evaluating expert systems for forensic DNA laboratories. Forensic News. 2005: August 8, 2008. Available from: http://www3.appliedbiosystems.com/cms/groups/applied_markets_marketing/documents/general_documents/cms_042230.pdf.
5. Butler JM. Forensic DNA Typing: Biology, Technology, and Genetics of STR Markers. Second Edition ed. Burlington, MA: Elsevier Academy Press (USA); 2005. p. 413-433.
6. Roby RK. Expert systems help labs process DNA samples. NIJ Journal. 2008 July 2008(260).
7. Perlin MW. In: An expert system for scoring DNA database profiles. Proceedings of the eleventh international symposium on human identification: 2000; October 29, 2000; Biloxi, MI. Madison, WI: Promega Corporation.

8. National DNA Index System (NDIS), DNA Data Acceptance Standards Operational Procedures, Appendix B, Guidelines for Submitting Requests for Approval of an Expert System for Review of Offender Samples, revised May 19, 2005.
9. Kadash K, Kozlowski BE, Biega LA, Duceman BW. Validation study of the TrueAllele automated data review system. J Forensic Sci. 2004 Jul;49(4):660-7.
10. DNA Advisory Board. Quality assurance standards for forensic DNA testing laboratories and for convicted offender DNA databasing laboratories. Forensic Sci Communications. July 2000;2(3).
11. Roby RK, Christen AD. Validating expert systems: Examples with the FSS-i³TM expert systems software. Profiles in DNA. September 2007;10(2):13-5.
12. Product brochure: GeneMarker[®] HID STR human identity software. 2008.

Appendix A

Approach to Calibrating the Expert System for Internal Validation of GeneMapper *ID*



Appendix B

Overview of the Workflow for GeneMapper *ID* and GeneMarker *HID*

Steps	GeneMapper <i>ID</i> v3.2	GeneMarker <i>HID</i> v1.71
1. Import Data	<p>Open GMID and select Add Samples to Project from the File Menu.</p> <p>Navigate to the location of the samples and select.</p> <p>Click Add to List and Add when list is complete.</p>	<p>Open GMKR and select Open Data from the Run Wizard window.</p> <p>Select either Add for individual samples or Add Folder for a batch or run.</p> <p>Navigate to the location of the samples and select.</p> <p>Click Ok to load samples.</p>
2. View Raw Data	<p>(Raw data can be viewed at this time by selecting Raw Data from the View tab.)</p>	<p>Upon loading, GMKR displays the raw data for inspection prior to analysis. The synthetic gel can also be visualized at this point.</p>
3. Setup Analysis Specifics	<p>Under the Samples Tab, define the following parameters: panel, analysis method (previously set), size standard, sample types, and export (for CODIS applications).</p> <p>GMID allows for multiple panels and analysis methods to be used in one run.</p>	<p>In the run wizard, chose Auto Run to analyze with the previously used template settings, or click Run to chose another template/create new template settings.</p>
4. Analyze	<p>Green arrows appear by each sample indicating that they are ready to analyze. Click the green Run arrow in the toolbar to analyze the samples.</p>	<p>At the end of the Run Wizard, clicking Ok automatically initiates analysis.</p>
5. Inspect “Low Quality” Sizing Standards	<p>In the Samples Window, check for “Low Quality” Size Standards.</p> <p>Analyst must investigate any flag in the Sizing Quality (SQ) column.</p>	<p>Samples with poor Size Standard quality will be denoted by a yellow icon in the file tree and must be investigated.</p> <p>Click on the Size Calibration icon in the toolbar to visualize the Size Standard scoring and linearity.</p>
6. Inspect Ladder(s)	<p>With ladder(s) selected in the Samples Window, click on the Genotypes Tab.</p> <p>Select the Genotypes Plot icon (or click Analysis>Display Plots). Scroll through the ladders and verify correct alignment.</p>	<p>Select the ladder used for analysis from the Sample Tree (it will be bolded). Verify proper allele calls in the electropherogram and/or Peak Table.</p>
7. Examine Low Quality Genotypes/ Alleles	<p>Highlight the marker’s row in the Genotypes Window. Select the Genotypes Plot icon to see the electropherogram for that locus.</p> <p>You can also view the locus in the context of the sample by clicking the Samples Plot icon in the Samples Window. Raw data can also be visualized.</p>	<p>Scroll through the allele table and double click the marker(s) with the “Low Quality” flag(s). GMKR automatically pulls up the electropherogram for that locus.</p> <p>To isolate or layer dye colors, toggle using the Show Color icon in the tool bar. Zoom out to see the marker alleles in context of the sample, or click the Browse by All Color icon in the toolbar.</p>
8. Export Table	<p>Click the Export Table icon in the toolbar, which exports the displayed information from the Genotypes Table.</p> <p>The Export CODIS Table under File exports a table with CODIS specific information.</p>	<p>Click the Save Peak Table icon in the toolbar to export detailed information on all called peaks.</p> <p>Click the Save Table icon on the Allele Table to save the allele calls only.</p> <p>Electropherograms and Trace Data can be exported under the Tools tab.</p> <p>Export CODIS is found under the Applications tab.</p>

Appendix C: GeneMapper ID Expert System Protocol

High Throughput STR Data Analysis Using the Expert System Capabilities of GeneMapper® ID v3.2

Purpose: To expedite data analysis of single source reference/population samples amplified with Identifiler® and/or Yfiler® PCR Amplification Kit by using the expert system capabilities of GeneMapper® ID v3.2.

Validated parameters optimize the accuracy and efficiency of allele calling by the software program. Rule settings and thresholds alert the analyst of data that require manual intervention. When using an expert system, the two reviewing analysts are only required to manually review loci which are flagged with a red octagon (●). Passing loci, which are flagged with a green square (■), are accepted as correctly genotyped and require no analyst review.

Supplies


GeneMapper ID v3.2

Procedure

A. First Read: Importing and Analyzing Identifiler or Yfiler Data

Note: If this is the first time analysis is being performed on a particular computer, *Analysis Methods* must be established. Refer to Appendix A for guidance.



1. Launch the GeneMapper ID software.
2. Click **File** → **Add Samples to Project....**
3. Navigate through the file tree to the location of the data. Select the run folder(s) and click **Add To List>>**. The files will appear in the *Samples To Add* field; click **Add**. The samples now appear in the **Samples** tab.
4. In the left pane, expand the run folder(s). Click on the first sample. In the right pane, select the **Raw Data** tab. Establish the start and stop points for data analysis by inspecting a few samples throughout the run. Click **View** → **Samples** to return to the **Samples** tab.
5. Verify that the *Table Setting* field is set to display the *HID table*.
6. Select **Tools** → **GeneMapper Manager**. Click on the **Analysis Methods** tab and select *Expert System Identifiler/Expert System Yfiler*. Confirm all settings with Appendix A, Figures 1 or 2 accordingly. If necessary, adjust the *Partial Range* for analysis under the **Peak Detector** tab. Once all settings are confirmed, click **OK**.
7. Verify that the appropriate identifiers are chosen in the *Sample Type* field (i.e., Allelic Ladder, Sample, controls).

8. Establish the analysis parameters for the first sample in the project.
 - In the *Analysis Method* field, select *Expert System Identifier/Expert System Yfiler* from the dropdown menu.
 - Click in the *Panel* field. For Identifier data, double-click *AmpFLSTR_Panels_v1* and then double-click *Identifier_v1*. For Yfiler data, double-click *AmpFLSTR_Yfiler_Panel_v2* and then double click *Yfiler_v2*.
 - Click in the *Size Standard* field. Select *CE_G5_HID_GS500* from the dropdown menu.
9. Click in the column header for *Analysis Method* and drag across to the *Size Standard* header to select the three columns. Click **Edit → Fill Down** (or Ctrl+D) to populate the settings for the all samples in the project.
10. Click the Green run arrow  in the icon toolbar.
11. Save the project with an extension of “*.ALL” which denotes that all .fsa files for the run are in the project.

B. Quality Control Checks (Reported in worksheet Y:\R&D Worksheets\QC for Batch Processing)

1. Verify the quality of the internal sizing standards. While still in the **Samples** tab, quickly check that all samples contain green squares (■) in the **SQ** column. If any sample contains a red **SQ** icon (●), inspect the size standard. If the sizing issue cannot be resolved, the sample should be removed from the GeneMapper Project, and the .fsa file should be moved to a cherry picking folder for the batch.

Note: Do not delete any .fsa files. All .fsa files should be archived.

2. Inspect the allelic ladders for proper bin alignment. In the **Samples** tab, highlight all ladders (Ctrl+select). Click on the **Sample Plots** icon . Select only one dye at a time, and verify that all ladder alleles are correctly called (confirm with the allelic ladder figures in Appendix B).
3. Inspect the negative control and reagent blank. In the **Samples** tab, select all negative controls and reagent blanks (Ctrl+select). Click on the **Sample Plots** icon . No peaks should be present. If any peaks appear, attempt to identify the cause. If true alleles are present, consult with a supervisor.
4. Verify the control profile. Click on the Genotypes tab. In the *Table Setting* field select *Control Check*. Any locus in the control that is not concordant with the designated profile will have a red icon in the **CC** column. Manually inspect these loci.




Note: If this is the first time analysis is being performed on a particular computer, the Control Check Table Setting must be established. Refer to Appendix C for guidance.

Note: In GeneMapper *ID*, the default positive control profiles are 9947A and 007. If using 9948, it is necessary to change the control profile in the software. Appendix D describes the procedure for verifying/changing the control profiles.

5. Click **File** → **Save....**


C. First Read: Manual Review of Low Quality Loci


Note: If this is the first time analysis is being performed on a particular computer, specific *Table Settings* must be established. Refer to Appendix C for guidance.

1. While in the analyzed project (named “*.ALL”), select **File**→ **Save Project As....** Name the project as desired with an extension of “*.First.INT”, where *INT* is the analyst’s initials.
2. Click on the **Genotypes** tab.
3. For Yfiler data analysis only: In the *Table Setting* field, select *DYS385 passing*. Select **Edit** → **Select All**. Select the **Sample Plots** icon . Right-click on the red **GQ** (●) icon and override the genotype quality for all DYS385 samples in this window. These are only flagged because there are two alleles detected and can be accepted as accurate allele calls. Close the window when finished.
4. In the *Table Setting* field, select *Low Quality*. Only the loci which are flagged red (●) in the **GQ** column will appear.
5. Select **Edit** → **Select All**. Click on the **Genotype Plots** icon . Review every locus and make the appropriate allele calls. If necessary, return to the **Samples** tab and select the **Sample Plots** icon  to view the locus in the context of the entire sample (e.g., to diagnose spectral pull up peaks). Once changes are made, the **GQ** icon turns green.

Note: Some loci may not require changes (e.g., those with off scale peaks or low peak height ratio). If the genotype can be confirmed, right-click on the red **GQ** (●) icon and override the genotype quality.

Note: Once a change is made at a locus, the **GQ** icon turns to green; it is important to not edit a locus unless it is entirely interpretable. If any genotype is uninterpretable, leave the **GQ** icon red (●).




6. Export the first reads and delete incomplete/uninterpretable samples. It is important that the following steps be performed in order. All tables should be exported as Tab- and Comma- delimited Text (*.txt, *.csv).
 - Click on the **Genotypes** tab. In the *Table Setting* field, select *Passing*. Click the **Export Table** icon . In the *File name* field, name the batch with an extension of “*.partial and full.INT”, where *INT* is the analyst’s initials. Navigate to the appropriate batch folder, and open the *GMID tables* folder. Click **Export Table**.


- In the Genotypes tab, change the *Table Setting* field to *Low Quality*. Any samples with uninterpretable loci appear with a red GQ icon (●). Click **Edit → Select All**. Click on the **Samples** tab. The highlighted samples contain uninterpretable loci. Note these sample names and add them to the cherry picking list for retesting. Select **Edit → Delete from Project**.
 - Click on the **Genotypes** tab, and change the *Table Setting* field to *Passing*. Click the **Export Table** icon . In the *File name* field, type the batch name with an extension of “*.full.INT”, where *INT* is the analyst’s initials. Navigate to the appropriate batch folder, and open the *GMID tables* folder. Click **Export Table**.
7. Return to the **Samples** tab and click **File → Save....**
 8. Close GeneMapper *ID*.

D. Second Read: Importing the project

1. Launch the GeneMapper *ID* software.
2. Click **File → Open Project...**
4. Locate the project with an extension of “*.ALL*”.
5. Click **File → Save Project As....** Name the project with an extension of “*.*Second.INT*”, where *INT* is the analyst’s initials.

E. Second Read: Manual Review of Low Quality Loci

1. Click on the **Genotypes** tab.
 2. For Yfiler data analysis only: In the *Table Setting* field, select *DYS385 passing*. Select **Edit → Select All**. Select the **Sample Plots** icon . Right-click on the red **GQ** (●) icon and override the genotype quality for all DYS385 samples in this window. These are only flagged because there are two alleles detected and can be accepted as accurate allele calls. Close the window when finished.
 3. In the *Table Setting* field, select *Low Quality*. Only the loci which are flagged red (●) in the **GQ** column appear.
 4. Select **Edit → Select All**. Click on the **Genotype Plots** icon . Review every locus and make the appropriate allele calls. If necessary, return to the **Samples** tab and select the **Sample Plots** icon  to view the locus in the context of the entire sample (e.g., to diagnose spectral pull up peaks).
- Note:** Some loci may not require changes to verify the genotype (e.g., those with off scale peaks or low peak height ratio). If the genotype can be confirmed, right-click on the red **GQ** (●) icon and override the genotype quality.

- Note:** If any genotype is uninterpretable, leave the **GQ** icon red (●), and verify that the sample was added to the cherry picking list.
- Once all Low Quality Loci have been reviewed and edited, change the *Table Setting* field to *Passing*. All **GQ** icons should be green (■).
 - Click the **Export Table** icon . In the *File name* field, enter the batch name with an extension of “*.second.INT”, where *INT* is the analyst’s initials. Navigate to the appropriate batch folder, and open the *GMID tables* folder. Click **Export Table**.
 - Close GeneMapper *ID*.

F. Concordance Check Between Manual Reads

The exported *Passing* tables will be checked for concordance between the first and second read. The two analysts will meet and resolve any discrepancies that exist. If the two analysts cannot resolve a discrepancy, a supervisor will be consulted.

G. Interpretation Guidelines for STR Analysis

Note: This is a guide for STR interpretation. These guidelines are to be used by the analyst as an aid in making allele calls since it is not possible to develop a set of rules that address each result.

- True alleles that fall outside the range of the allelic ladder are designated as off-ladder (OL) by the analysis software. If a true allele is one repeat or less outside of the allelic ladder, then it is to be manually called for population samples. Alleles will reported with less-than (<) or greater-than (>) signs preceding the relevant lower or upper bound allele of the ladder.
- Microvariants that fall between the rungs of the ladder for a given marker category may be designated as off-ladder (OL) by the analysis software if no virtual bin has been defined. When it is possible to unambiguously determine the proportions of the partial repeat, the allele should be reported with the **(nominal allele).X** format, where X = .1 for one base, X = .2 for two bases and X = .3 for three bases.
- The general shape and appearance of a peak should be well-defined so as to discriminate between a true allelic peak and artifacts (amplification artifacts, electrophoresis artifacts, kit artifacts...).

A. Artifacts associated with amplification

- PCR artifacts may include minus A and stutter peaks. Minus A peaks are one basepair shorter than the true allele and can be identified as a split or shoulder on the true peak. Minus A is often seen in samples that have been amplified with excessive template DNA. An identified minus A peak can be removed by an analyst.

- Stutter peaks are one repeat unit shorter than the true allele (four basepairs shorter for tetranucleotide loci). They are characterized by their distinctive position and decreased intensity as compared to their true allelic counterparts. Stutter percentages vary between loci and between alleles within the same locus. Within a locus, stutter ratios tend to increase with allelic size. If a peak is detected within a stutter position that exceeds the GeneMapper *ID* default stutter filter ratio, the related locus should be evaluated in the context of the entire profile to assess whether or not a mixture is present. Due to instrument detection limits, stutter percentages may be falsely augmented in overamplified data.

B. Electrophoretic artifacts

- Spikes can occur when air bubbles or urea crystals are present within the polymer. Spikes have atypical peak morphology with a similar peak width at the baseline and at the maximum peak height. Spikes are usually seen in all dye colors at the same position. They are not reproducible and may be resolved by re-injecting the sample.
- Pull-up is an effect caused by spectral overlap of component dyes. It can be observed as a bleed through peak in one fluorescent color that corresponds within one or two scan data points of a true allele in a different fluorescent color. Pull-up data are generally reproducible and a result of off-scale peaks generated through over-amplification. If pull-up is observed, especially in samples that do not contain off scale data, then spectral calibrations for the affected dye sets should be performed.
- Dye artifacts occur when fluorescent dyes detach from their respective primers and migrate independently during capillary electrophoresis. They are typically broad peaks consisting of only a single dye color.

Appendix A (of protocol)

Establishing the “Expert System Identifier” and “Expert System Yfiler” Analysis Methods

Note: This procedure is required at the initial setup of the software; subsequent Identifier and Yfiler data analyses will use the Analysis Method established here. However, it is important to verify the analysis parameters shown in Figures 1 and 2 before using the Expert System Analysis Methods.

1. Launch the GeneMapper *ID* software.
2. Select the **Tools** menu→**GeneMapper Manager**.
3. Select the **Analysis Methods** tab and click **New....**
4. Select the **HID** radio button.
5. Under the **General** Tab, type “Expert System Identifier” in the *Name* field. The *Description* and *Instrument* fields can remain empty.
6. Select the tabs for **Allele**, **Peak Detector**, **Peak Quality**, and **Quality Flags** one at a time, and verify that the parameters are set according to Figure 1.

Note: Under the **Peak Detector** tab, the partial range for analysis will vary. Consult the raw data to establish a scan data point range which spans from the 75 bp peak through the 450 bp peak of the LIZ (orange) labeled internal sizing standard signal.

7. Repeat steps 1-6 to create the “Expert System Yfiler” Analysis Method. Select the tabs for **Allele**, **Peak Detector**, **Peak Quality**, and **Quality Flags** one at a time, verifying that the parameters are set according to Figure 2.

Figure 1- Parameter Settings for *Expert System Identifiler* Analysis Method

The figure displays four screenshots of the 'Analysis Method Editor - HID' dialog box, showing different tabs and their respective parameter settings.

Allele Tab

Bin Set: AmpFLSTR_Bins_v1

☒ Use marker-specific stutter ratio if available

Marker Repeat Type:	Tri	Tetra	Penta	Hexa
Cut-off Value	0.0	0.075	0.0	0.0
MinusA Ratio	0.0	0.2	0.0	0.0
MinusA Distance	From 0.0	0.25	0.0	0.0
	To 0.0	1.5	0.0	0.0
Minus Stutter Ratio	0.0	0.0	0.0	0.0
Minus Stutter Distance	From 0.0	3.25	0.0	0.0
	To 0.0	4.75	0.0	0.0
Plus Stutter Ratio	0.0	0.0	0.0	0.0
Plus Stutter Distance	From 0.0	0.0	0.0	0.0
	To 0.0	0.0	0.0	0.0

Amelogenin Cutoff: 0.0

Buttons: Range Filter..., Factory Defaults, OK, Cancel

Peak Detector Tab

Peak Detection Algorithm: Advanced

Ranges

Analysis	Sizing
Partial Range	Partial Sizes
Start Pt: 4150	Start Size: 75
Stop Pt: 25000	Stop Size: 450

Smoothing and Baseline

Smoothing: ☐ None ☒ Light ☐ Heavy

Baseline Window: 51 pts

Size Calling Method

☐ 2nd Order Least Squares
☐ 3rd Order Least Squares
☐ Cubic Spline Interpolation
☒ Local Southern Method
☐ Global Southern Method

Peak Detection

Peak Amplitude Thresholds:

B: 50 R: 50
G: 50 O: 50
Y: 50

Min. Peak Half Width: 2 pts
Polynomial Degree: 3
Peak Window Size: 15 pts

Slope Threshold

Peak Start: 0.0
Peak End: 0.0

Buttons: Factory Defaults, OK, Cancel

Peak Quality Tab

Signal level

Homozygous min peak height: 200.0
Heterozygous min peak height: 100.0

Heterozygote balance

Min peak height ratio: 0.6

Peak morphology

Max peak width (basepairs): 1.5

Pull-up peak

Pull-up ratio: 0.05

Allele number

Max expected alleles: 2

Buttons: Factory Defaults, OK, Cancel

Quality Flags Tab

Quality weights are between 0 and 1.

Quality Flag Settings

Spectral Pull-up	0.8	Control Concordance	1.0
Broad Peak	0.8	Low Peak Height	0.3
Out of Bin Allele	0.8	Off-scale	0.8
Overlap	0.8	Peak Height Ratio	0.3

PQV Thresholds

Pass Range: Low Quality Range:

Sizing Quality: From 0.75 to 1.0 From 0.0 to 0.25
Genotype Quality: From 0.75 to 1.0 From 0.0 to 0.75

Buttons: Factory Defaults, OK, Cancel

Figure 2- Parameter Settings for *Expert System Yfiler* Analysis Method

Analysis Method Editor - HID

General | **Allele** | Peak Detector | Peak Quality | Quality Flags

Bin Set: AmpFLSTR_Yfiler_Binset_v2

☐ Use marker-specific stutter ratio if available

Marker Repeat Type :	Tri	Tetra	Penta	Hexa
Cut-off Value	0.12	0.12	0.12	0.12
MinusA Ratio	0.2	0.2	0.2	0.2
MinusA Distance	From 0.25	From 0.25	From 0.25	From 0.25
	To 1.5	To 2.0	To 1.5	To 1.5
Minus Stutter Ratio	0.15	0.15	0.0	0.0
Minus Stutter Distance	From 2.25	From 3.25	From 4.25	From 5.25
	To 3.75	To 4.75	To 5.75	To 6.75
Plus Stutter Ratio	0.1	0.0	0.0	0.0
Plus Stutter Distance	From 2.25	From 0.0	From 0.0	From 0.0
	To 3.75	To 0.0	To 0.0	To 0.0

Amelogenin Cutoff: 0.0

Range Filter... Factory Defaults

OK Cancel

Analysis Method Editor - HID

General | Allele | **Peak Detector** | Peak Quality | Quality Flags

Peak Detection Algorithm: Advanced

Ranges

Analysis: Partial Range
Sizing: Partial Sizes

Start Pt: 4000
Stop Pt: 17000

Start Size: 75
Stop Size: 450

Smoothing and Baseline

Smoothing: ☐ None ☒ Light ☐ Heavy

Baseline Window: 51 pts

Size Calling Method

☐ 2nd Order Least Squares
☐ 3rd Order Least Squares
☐ Cubic Spline Interpolation
☒ Local Southern Method
☐ Global Southern Method

Peak Detection

Peak Amplitude Thresholds:

B: 50 R: 50
G: 50 O: 50
Y: 50

Min. Peak Half Width: 2 pts
Polynomial Degree: 3
Peak Window Size: 15 pts

Slope Threshold

Peak Start: 0.0
Peak End: 0.0

Factory Defaults

OK Cancel

Analysis Method Editor - HID

General | Allele | Peak Detector | **Peak Quality** | Quality Flags

Signal level

Homozygous min peak height: 100.0
Heterozygous min peak height: 50.0

Heterozygote balance

Min peak height ratio: 0.4

Peak morphology

Max peak width (basepairs): 1.5

Pull-up peak

Pull-up ratio: 0.05

Allele number

Max expected alleles: 1

Factory Defaults

OK Cancel

Analysis Method Editor - HID

General | Allele | Peak Detector | Peak Quality | **Quality Flags**

Quality weights are between 0 and 1.

Quality Flag Settings

Spectral Pull-up	0.8	Control Concordance	1.0
Broad Peak	0.8	Low Peak Height	0.3
Out of Bin Allele	0.8	Off-scale	0.8
Overlap	0.8	Peak Height Ratio	0.3

PQV Thresholds

Pass Range: Low Quality Range:

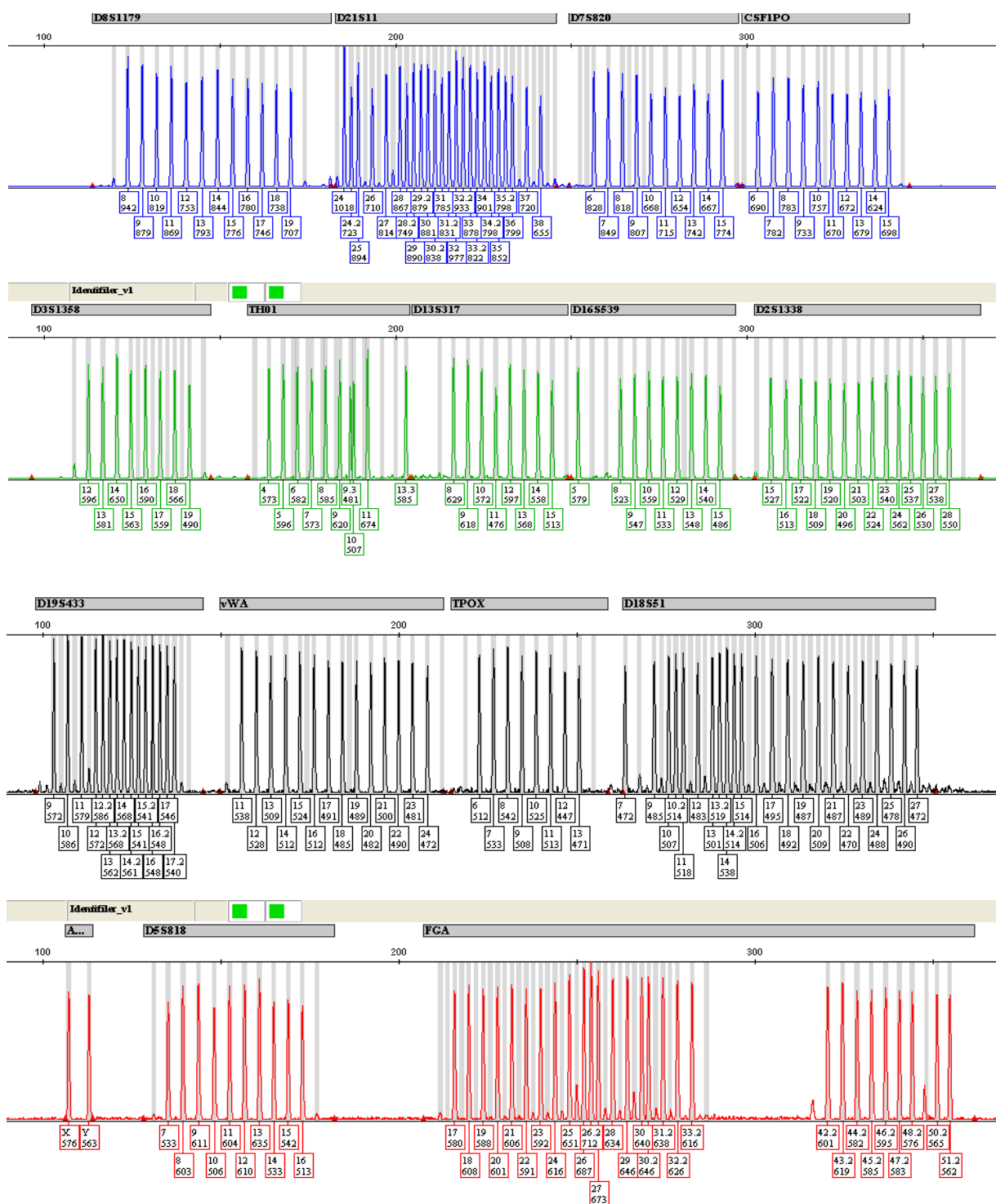
Sizing Quality: From 0.75 to 1.0 From 0.0 to 0.25
Genotype Quality: From 0.75 to 1.0 From 0.0 to 0.75

Factory Defaults

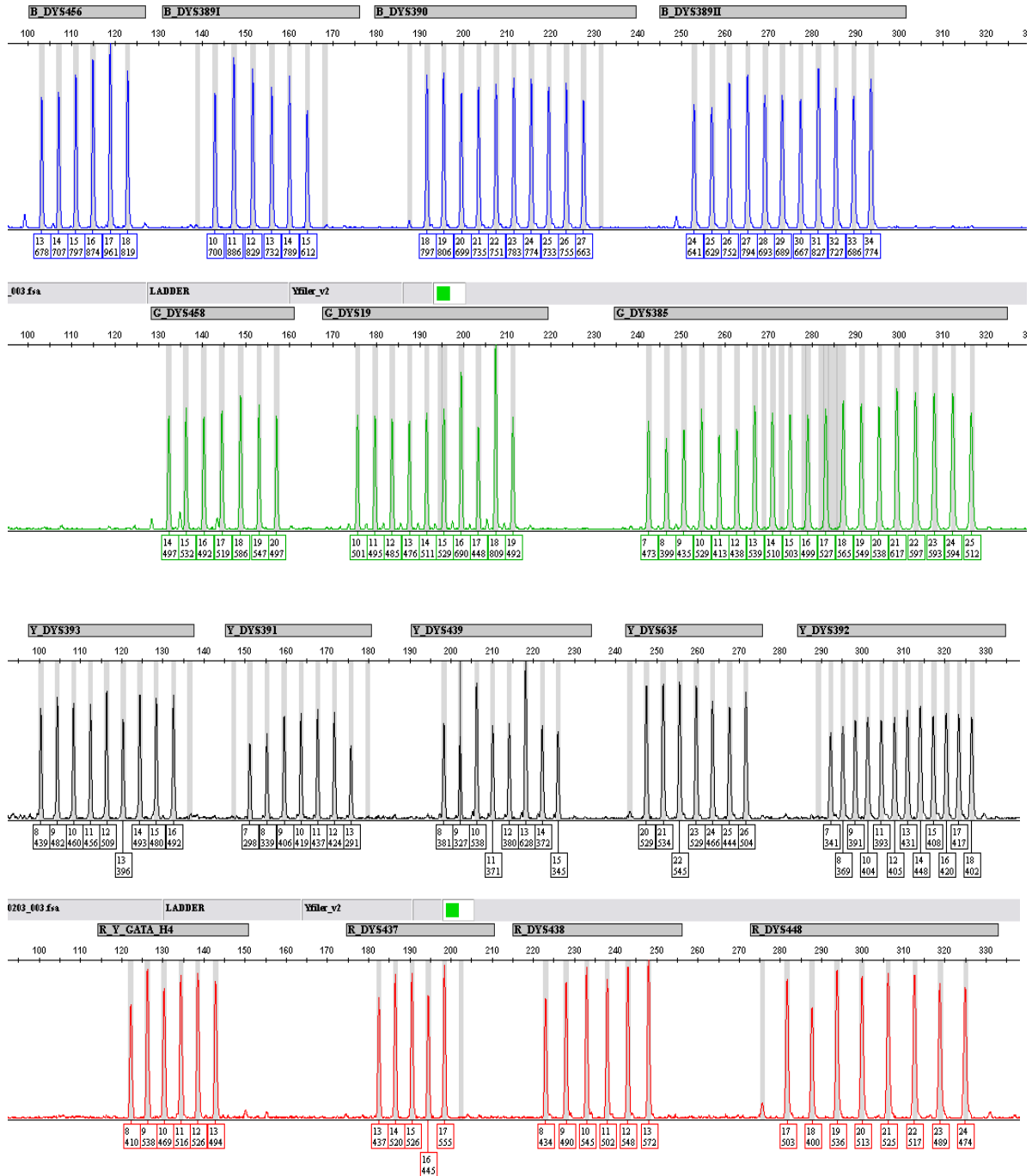
OK Cancel

Appendix B (of protocol)

Identifiler Allelic Ladder



Yfiler Allelic Ladder



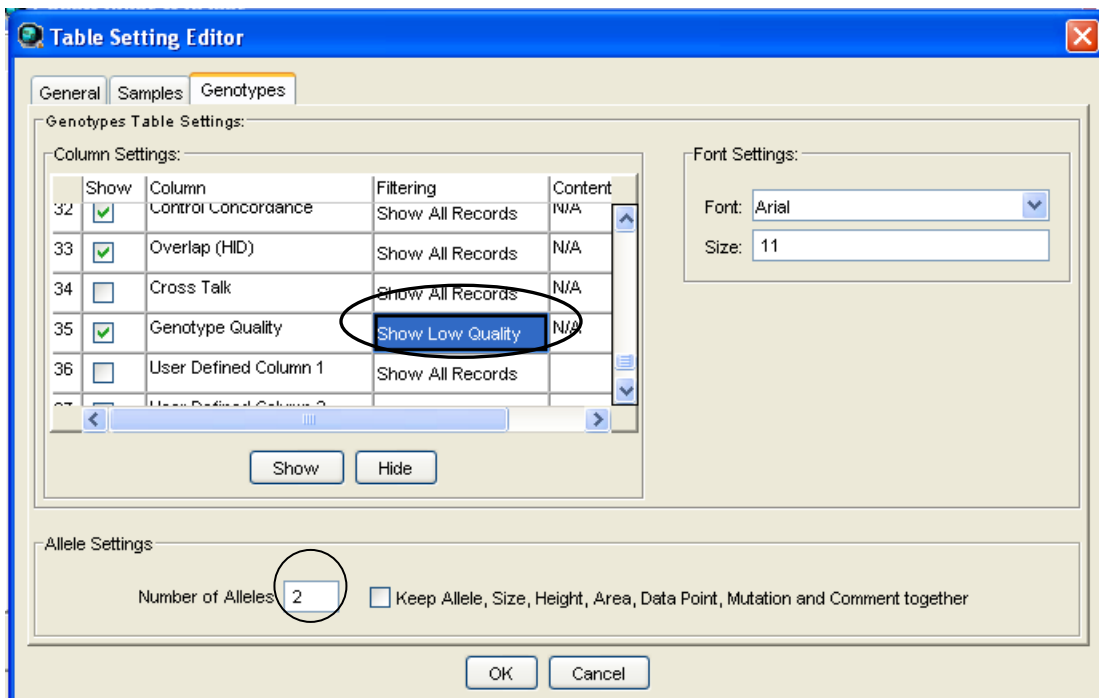
Appendix C (of protocol)

Establishing the *Low Quality, Passing, Control Check, and DYS385 Passing Table Settings*

Note: This procedure is required at the initial setup of the software; subsequent Identifiler and Yfiler data analyses will use the Table Settings established here.

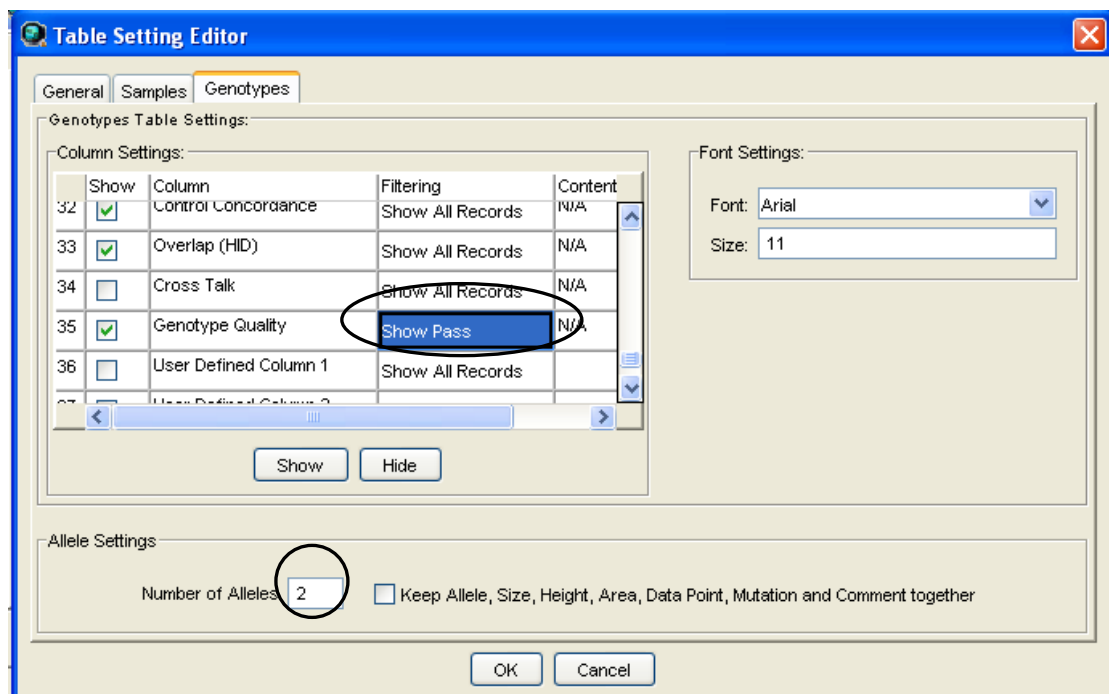
A. *Low Quality* Table Settings

1. Launch the GeneMapper *ID* software.
2. Select the **Tools** menu → **GeneMapper Manager**.
3. Select the **Table Settings Editor** tab and click **New...**
4. Under the **General** tab in the *Name* field, enter “Low Quality.”
5. Select the **Genotypes** tab. In the *Column Settings* field, scroll down to Field 35, titled *Genotype Quality*. In the *Filtering* column, select *Show Low Quality* from the dropdown menu as shown below.
6. In the *Allele Settings* field, enter “2” for *Number of Alleles*.
7. Click **OK**.



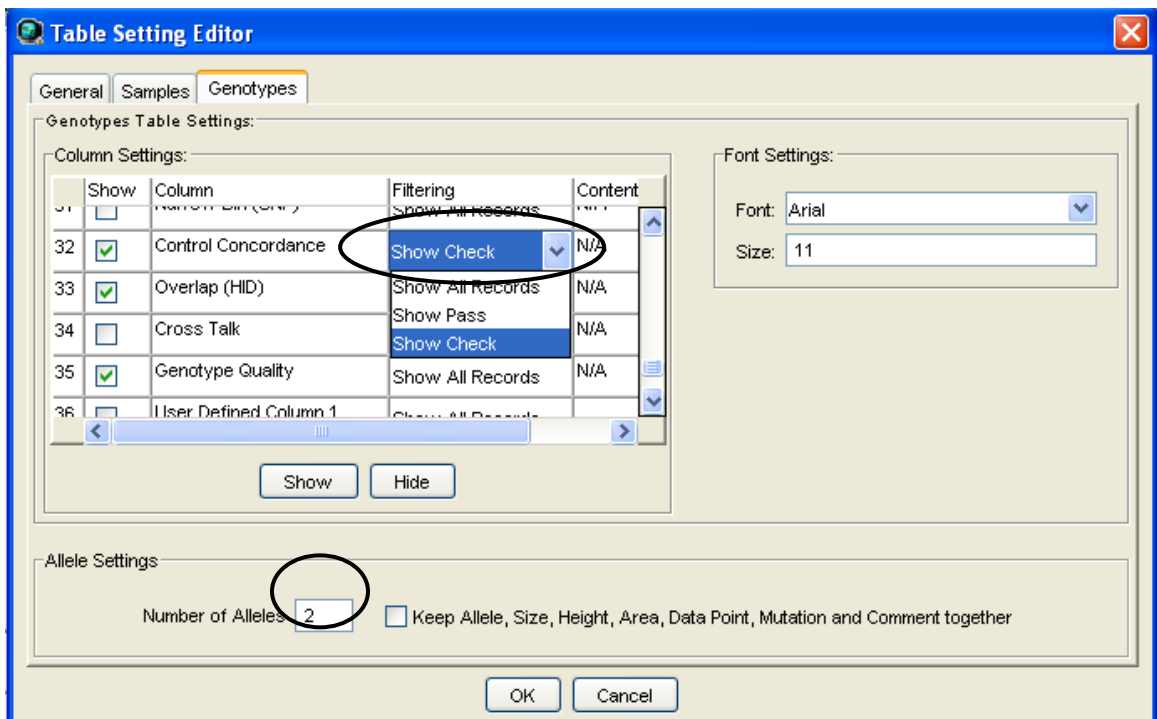
B. *Passing* Table Settings

1. Launch the GeneMapper *ID* software.
2. Select the **Tools** menu → **GeneMapper Manager**.
3. Select the **Table Settings Editor** tab and click **New...**
4. Under the **General** tab in the *Name* field, enter “*Passing*.”
5. Select the **Genotypes** tab. In the *Column Settings* field, scroll down to Field 35, titled *Genotype Quality*. In the *Filtering* column, select *Show Pass* from the dropdown menu as shown below.
6. In the *Allele Settings* field, enter “2” for *Number of Alleles*.
7. Click **OK**.



C. Control Check Table Settings

8. Launch the GeneMapper *ID* software.
9. Select the **Tools** menu → **GeneMapper Manager**.
10. Select the **Table Settings Editor** tab and click **New...**
11. Under the **General** tab in the *Name* field, enter “Control Check.”
12. Select the **Genotypes** tab. In the *Column Settings* field, scroll down to Field 32, titled *Control Concordance*. In the *Filtering* column, select *Show Check* from the dropdown menu as shown below.
13. In the *Allele Settings* field, enter “2” for *Number of Alleles*.
14. Click **OK**.



D. DYS385 Passing Table Settings

1. Launch the GeneMapper *ID* software.
2. Select the **Tools** menu → **GeneMapper Manager**.
3. Select the **Table Settings Editor** tab and click **New....**
4. Under the **General** tab in the *Name* field, enter “DYS385.”
5. Select the **Genotypes** tab. In the *Column Settings* field, scroll down to Fields 6, 19, 24-27, 29 and 33. In the *Filtering* column, make the appropriate selections and entries, using the following figures as guides.
6. In the *Allele Settings* field, verify that “2” is entered for the *Number of Alleles*.
7. Click **OK**.

Table Setting Editor

General Samples **Genotypes**

Genotypes Table Settings:

Column Settings:

Column	Filtering	Content
4 Run Name	Show All Records	
5 Panel	Show All Records	
6 Marker	Show Records Con...	DYS385
7 Dye	Show All Records	N/A
8 Allele	Show All Records	

Show Hide

Font Settings:

Font: Arial

Size: 11

Allele Settings:

Number of Alleles: 2 ☐ Keep Allele, Size, Height, Area, Data Point, Mutation and Comment together

OK Cancel

Table Setting Editor

General Samples **Genotypes**

Genotypes Table Settings:

Column Settings:

Column	Filtering	Content
18 Omit From Clustering (SNPle)	Show All Records	N/A
19 Off-scale	Show Pass	N/A
20 Sharp Peak (M)	Show All Records	N/A
21 One Basepair Allele (M)	Show All Records	N/A
22 Single Peak Artifact (M)	Show All Records	N/A

Show Hide

Font Settings:

Font: Arial

Size: 11

Allele Settings:

Number of Alleles: 2 ☐ Keep Allele, Size, Height, Area, Data Point, Mutation and Comment together

OK Cancel

Table Setting Editor

General Samples **Genotypes**

Genotypes Table Settings:

Column Settings:

	Column	Filtering	Content
22	Single Peak Attract (M)	Show All Records	N/A
23	Split Peak (M)	Show All Records	N/A
24	Out of Bin Allele	Show Pass	N/A
25	Peak Height Ratio	Show Pass	N/A
26	Low Peak Height	Show Pass	N/A

Font Settings:

Font: Arial

Size: 11

Allele Settings:

Number of Alleles: 2 ☐ Keep Allele, Size, Height, Area, Data Point, Mutation and Comment together

OK Cancel

Table Setting Editor

General Samples **Genotypes**

Genotypes Table Settings:

Column Settings:

	Column	Filtering	Content
25	Peak Height Ratio	Show Pass	N/A
26	Low Peak Height	Show Pass	N/A
27	Spectral Pull-up	Show Pass	N/A
28	Allele Number	Show Check	N/A
29	Broad Peak	Show Pass	N/A

Font Settings:

Font: Arial

Size: 11

Allele Settings:

Number of Alleles: 2 ☐ Keep Allele, Size, Height, Area, Data Point, Mutation and Comment together

OK Cancel

Table Setting Editor

General Samples **Genotypes**

Genotypes Table Settings:

Column Settings:

	Column	Filtering	Content
32	Control Concordance	Show All Records	N/A
33	Overlap (HID)	Show Pass	N/A
34	Cross Talk	Show All Records	N/A
35	Genotype Quality	Show All Records	N/A
36	User Defined Column 1	Show All Records	

Font Settings:

Font: Arial

Size: 11

Allele Settings:

Number of Alleles: 2 ☐ Keep Allele, Size, Height, Area, Data Point, Mutation and Comment together

OK Cancel

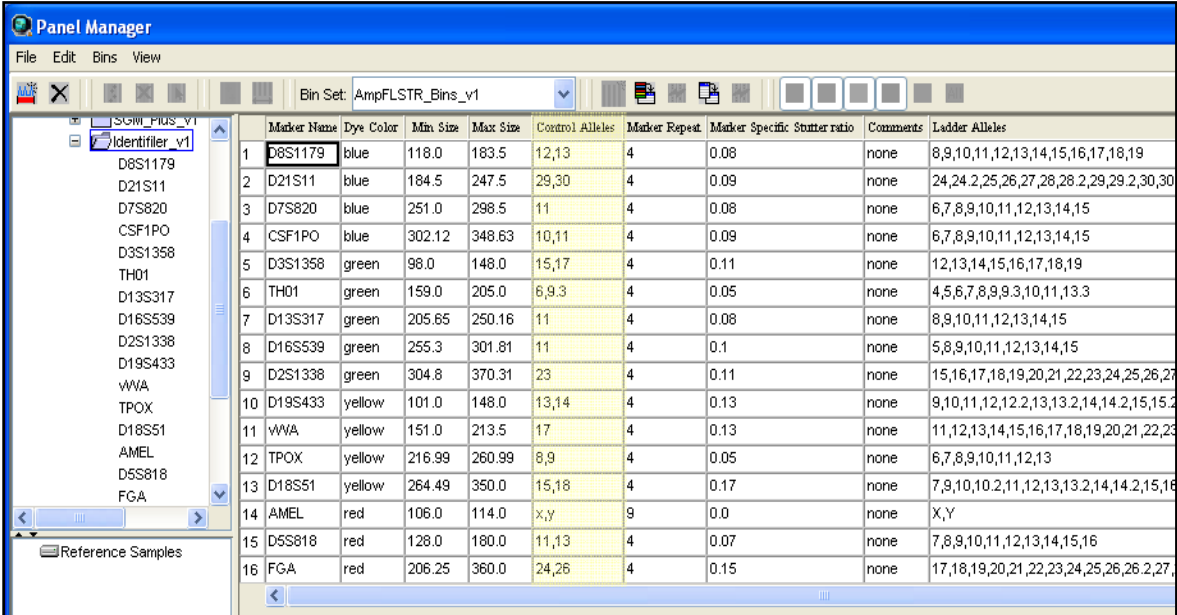
Appendix D (of protocol)

Verifying/Changing the Positive Control Profiles

A. Verifying/Changing the Identifiler control profile

1. Select the **Tools** menu→ **Panel Manager**.
2. Double-click *AmpFLSTR_Panels_v1*.
3. Click *Identifiler_v1*. In the right pane, the current control profile is displayed in the *Control Alleles* column. Verify that the profile is correct.
4. To change the profile alleles, select the field for each marker and type the correct allele(s). Heterozygous alleles should be separated by a comma. Homozygous alleles are only entered once. Figure 1 shows the control profile for 9948.
5. Click **Apply** to save any changes and **OK** to exit. Click **Cancel** if the profile is correct and no changes were necessary.

Figure 1- The Identifiler profile for 9948

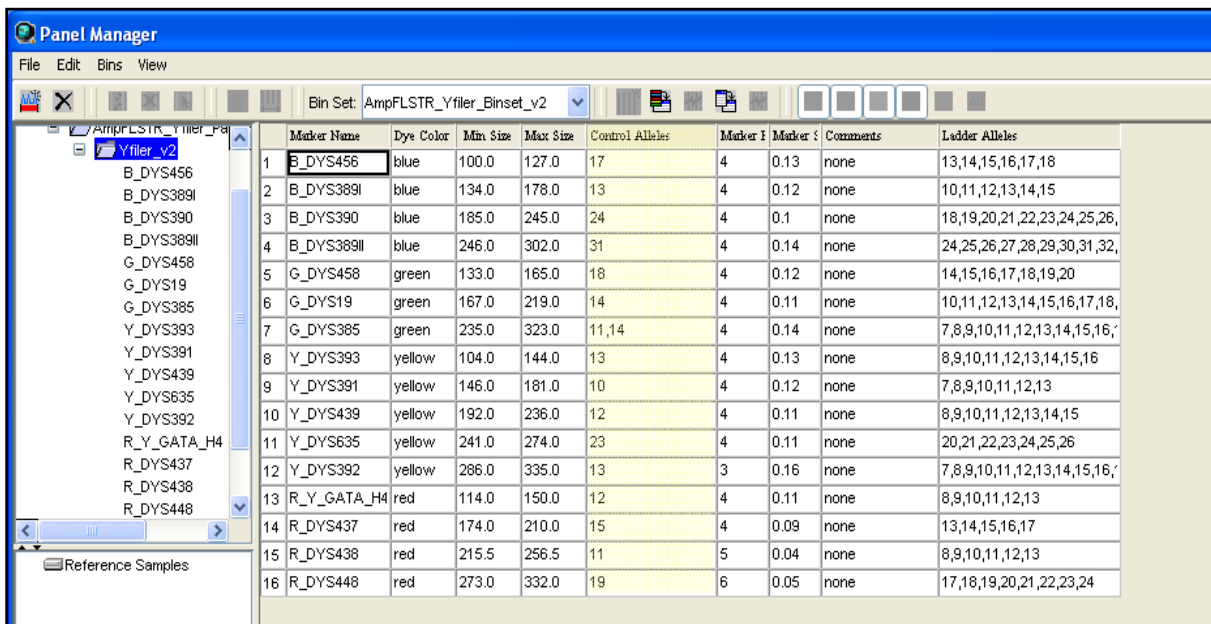


Marker Name	Dye Color	Min Size	Max Size	Control Alleles	Marker Repeat	Marker Specific Stutter ratio	Comments	Ladder Alleles
1 D8S1179	blue	118.0	183.5	12,13	4	0.08	none	8,9,10,11,12,13,14,15,16,17,18,19
2 D21S11	blue	184.5	247.5	29,30	4	0.09	none	24,24.2,25,26,27,28,28.2,29,29.2,30,30
3 D7S820	blue	251.0	298.5	11	4	0.08	none	6,7,8,9,10,11,12,13,14,15
4 CSF1PO	blue	302.12	348.63	10,11	4	0.09	none	6,7,8,9,10,11,12,13,14,15
5 D3S1358	green	98.0	148.0	15,17	4	0.11	none	12,13,14,15,16,17,18,19
6 TH01	green	159.0	205.0	6,9,3	4	0.05	none	4,5,6,7,8,9,9.3,10,11,13,3
7 D13S317	green	205.65	250.16	11	4	0.08	none	8,9,10,11,12,13,14,15
8 D16S539	green	255.3	301.81	11	4	0.1	none	5,8,9,10,11,12,13,14,15
9 D2S1338	green	304.8	370.31	23	4	0.11	none	15,16,17,18,19,20,21,22,23,24,25,26,27
10 D19S433	yellow	101.0	148.0	13,14	4	0.13	none	9,10,11,12,12.2,13,13.2,14,14.2,15,15.2
11 vWA	yellow	151.0	213.5	17	4	0.13	none	11,12,13,14,15,16,17,18,19,20,21,22,23
12 TPOX	yellow	216.99	260.99	8,9	4	0.05	none	6,7,8,9,10,11,12,13
13 D18S51	yellow	264.49	350.0	15,18	4	0.17	none	7,9,10,10.2,11,12,13,13.2,14,14.2,15,16
14 AMEL	red	106.0	114.0	X,Y	9	0.0	none	X,Y
15 D5S818	red	128.0	180.0	11,13	4	0.07	none	7,8,9,10,11,12,13,14,15,16
16 FGA	red	206.25	360.0	24,26	4	0.15	none	17,18,19,20,21,22,23,24,25,26,26.2,27

B. Verifying/Changing the Yfiler control profile

6. Select the **Tools** menu→ **Panel Manager**.
7. Double-click *AmpFLSTR_Yfiler_Panel_v2*.
8. Click *Yfiler_v2*. In the right pane, the current control profile is displayed in the *Control Alleles* column. Verify that the profile is correct.
9. To change the profile alleles, select the field for each marker and type the correct allele(s). Heterozygous alleles should be separated by a comma. Homozygous alleles are only entered once. Figure 2 shows the control profile for 9948.
10. Click **Apply** to save any changes and **OK** to exit. Click **Cancel** if the profile is correct and no changes were necessary.

Figure 2- The Yfiler profile for 9948



Marker Name	Dye Color	Min Size	Max Size	Control Alleles	Marker ID	Marker Size	Comments	Ladder Alleles
1 B_DYS456	blue	100.0	127.0	17	4	0.13	none	13,14,15,16,17,18
2 B_DYS389I	blue	134.0	178.0	13	4	0.12	none	10,11,12,13,14,15
3 B_DYS390	blue	185.0	245.0	24	4	0.1	none	18,19,20,21,22,23,24,25,26,
4 B_DYS389II	blue	246.0	302.0	31	4	0.14	none	24,25,26,27,28,29,30,31,32,
5 G_DYS458	green	133.0	165.0	18	4	0.12	none	14,15,16,17,18,19,20
6 G_DYS19	green	167.0	219.0	14	4	0.11	none	10,11,12,13,14,15,16,17,18,
7 G_DYS385	green	235.0	323.0	11,14	4	0.14	none	7,8,9,10,11,12,13,14,15,16,
8 Y_DYS393	yellow	104.0	144.0	13	4	0.13	none	8,9,10,11,12,13,14,15,16
9 Y_DYS391	yellow	146.0	181.0	10	4	0.12	none	7,8,9,10,11,12,13
10 Y_DYS439	yellow	192.0	236.0	12	4	0.11	none	8,9,10,11,12,13,14,15
11 Y_DYS635	yellow	241.0	274.0	23	4	0.11	none	20,21,22,23,24,25,26
12 Y_DYS392	yellow	286.0	335.0	13	3	0.16	none	7,8,9,10,11,12,13,14,15,16,
13 R_Y_GATA_H4	red	114.0	150.0	12	4	0.11	none	8,9,10,11,12,13
14 R_DYS437	red	174.0	210.0	15	4	0.09	none	13,14,15,16,17
15 R_DYS438	red	215.5	256.5	11	5	0.04	none	8,9,10,11,12,13
16 R_DYS448	red	273.0	332.0	19	6	0.05	none	17,18,19,20,21,22,23,24