PHASE I. Internal Validation of TrueAllele Genetic Calculator as an Expert Assistant for Reads and Review of Data from Reported Sexual Assault Evidence

Expert System and Version	Manufacturer	Instrument Platform	Kit
TrueAllele Genetic Calculator Analyze Module Build # 252, Version # 9 (8-July- 2010) VUIer Version # 3.3.3919.1 (16-Jul-2010) Server Version # 3.25.3768.1	**Cybergenetics, Corp.**	**Applied Biosystems 3130xl with Collection v. 3.0**	**Identifiler**

Purpose:

To conduct an internal validation for the use of the current version of the TrueAllele Genetic Calculator (TAGC) and higher as an expert assistant (1-2) to read and review reported sexual assault (RSA) evidence data processed on an AB 3130xl genetic analyzer with collection software version 3.0 or higher. All studies follow the SWGDAM Guidelines for an Internal Validation (Standard 8.7).

Background:

Software tools have been used in forensic DNA analyses to assist users in interpreting and reviewing profile data. Expert Systems (ES) have been defined by the FBI for single source DNA profiles from known and convicted offender samples and validations of such must comply with NDIS appendix B (3-7). Expert systems allow data to be uploaded to CODIS without manual intervention (6-7, Appendix A).

Software tools have subsequently been developed to assist DNA analysts in conducting mixture interpretations and deconvolutions. These systems are not expert systems, without manual intervention, but are considered Expert Assistants (EA) involving human interactions and review of all data and interpretations made by the computer. These EA's use quantitative information to calculate values for variables such as peak heights, sister allele balance, stutter and mixture ratios of contributors. The TrueAllele Genetic Calculator (TAGC) models peaks and applies mathematical principles to the data to

explain and objectively infer genotypes for the given DNA profile in the electropherogram. Each genotype comparison with a suspect produces a match information statistic reported as a likelihood ratio (LR). Several references are included to support the use of LR's and their application to forensic DNA analyses (8-13). The TAGC utilizes a probabilistic genotyping method that has been referenced in the SWGDAM Mixture Interpretation Guidelines as a suitable approach for addressing potential stochastic amplification (Section 3.2.2) (4).

The MSPCL has utilized quantitative information in assessment of DNA profile data for over a decade. When performing manual methods, the DNA analyst may use worksheets (with hand calculations) and/or excel spreadsheets to calculate sister allele peak balance and mixture weights to better deconvolute a mixture. This experience makes the MSPCL well suited to conduct the validation of an EA and compare/contrast the manual method with the TAGC method.

Introduction:

The TAGC will be used as a tool to assist in reading data from evidentiary samples in specified sexual assault cases. Specific studies were designed to deconvolute mixtures from samples whereby a known reference from the victim was obtained and used to deduce a contributor(s) foreign to the victim. Some cases may contain more than two donors and will be evaluated based on the number of contributors, associated mixture weights and probative value. Cases may or may not have a known reference from a suspect. Cases without suspects will be analyzed for potential CODIS upload while cases with suspects will be validated to include a statistical inference.

The use of the TAGC with manual review and evaluations to read data and interpret results from questioned evidence will be referred to as the TrueAllele caseworking (TACW) system. The TACW system is different than conventional manual methods and renders results that may, initially, be viewed as confusing to the user. The concept of using a manual calculator has the expectation of inputting the same numbers and deriving the exact same answer. For example, if an analyst has conducted an interpretation manually and derived the genotype(s) of a questioned sample, POPSTATS will give the same frequency of occurrence for that single source or mixture profile in replicate runs (i.e. same frequency for a MPE, MMPE or CPE/CPI statistic). The TA genetic calculator will not yield the exact same statistic from the data in the electropherogram because it is a fluid mechanism based on random sampling from the joint distribution of the data and probability model. It would be better to think of the TAGC as a measuring device (similar to a RT PCR machine or analytical balance) that will give a different quantitation value or weight if replicate runs were performed. As one would expect the cycle threshold and associated DNA concentration or weight of an object to vary between replicate runs, the TAGC is expected to yield different results in replicate runs. Measurements are accompanied by uncertainty, the degree dependent on the instrument and the data. As such, this validation will investigate these differences to define the expectations of the TACW system. An overview of the TACW system is contained in Appendix B JBS to check App B and is described briefly herein. Refer to Cybergenetics

user manuals and MSP protocols for in depth descriptions of the use of the TAGC and the TACW system.

Methods:

Quote from JFS Paper:
"Cybergenetics' TrueAllele Genetic Calculator uses a fully Bayesian model of the STR data generation process, based on genotype, mixture weight and data certainty probability distributions (equations (1), (4) and (5), respectively). The calculator accounts for PCR stutter, relative amplification, DNA degradation, and other experiment factors. Conditioning on the observed quantitative STR data, the TrueAllele computer explores the model's parameter space using MCMC statistical search to determine the posterior probability distribution for every variable (Appendix). Results for variables of interest, such as genotypes and mixture weights, are reported as probability distributions (2). Match rarity results are reported as LRs (7, 26). The TrueAllele Visual User Interface (VUIer™) program lets a user visually explore their STR data and computed results (e.g., genotypes, matches, mixture weights)" (11)

The TACW system encompasses five modules to process the data (14-19).

Module I. Analyze

Note that this module has been thoroughly validated as part of the Developmental Validation for TrueAllele Databanking (Appendix A)

- Data is taken directly from the 3130xl genetic analyzer as a .fsa file and analyzed to assess all peaks in the given sample. This module checks tracking and size standards for each file and assigns base pair size, peak height and allelic designations without the use of any thresholds (e.g. analytical nor stochastic).
- Quality control checks are used to evaluate all controls including negatives and positives (i.e. extraction and amplification) and allelic ladders
- The files are converted to a .gel file in preparation for the next module.

Module II. Data

- Data is presented for human review. This module may be used to assess the number of donors.
- Data is uploaded to a database for subsequent processing.
- Batches of requests can be uploaded to the database for all samples or selected samples on one .gel file/run at a time.
- Analyst may conduct an additional review of controls, if necessary.

Module III. Request

- Analyst reviews the data to decide what requests to ask the TAGC to perform based on the number of contributors.

- Information regarding the unknown profile and the non-probative standards will be used to pose the appropriate questions to deduce the unknown probative profile.
    - This validation is specific for RSA cases submitted with a victim reference. The victim profile will be used to subtract out their contribution in an effort to obtain the probative unknown profile foreign to the victim (note that most sexual assault evidence is derived from intimate swabs taken from the victim's body).
- Interpretation questions are requested for the TAGC to solve.
- Parameters are established for each request including the known donor(s), the number of unknown contributor(s) and the run time, or number of iterations to be utilized to resolve the profile data. More complex data may require a longer run time.
- The requests are submitted to the TAGC and run on a processor. Processing time will vary dependent on the settings.

Module IV. Review

- The computer results provide the user with inferred genotypes and associated likelihoods at each locus. Note that the genotypes of the questioned sample, Q, are inferred without any information from the known, K (e.g. suspect or other reference samples).
- Data are presented to the analyst for evaluation of the computer interpretation. The analyst can check the computer modeling, mixture weights, convergence measures and match information to evaluate interpretations and draw conclusions.
- Decisions must be made after assessment of the results and additional or other questions may be resubmitted for interpretation, if necessary.

Module V. Report

- Statistical analyses are presented to include probabilities for four racial groups (Caucasian, African American, Hispanic and Asian).
- Review of match data, application of theta values, and evaluation of the data for reporting purposes are made in this module.
- Reports are generated and can be exported as .txt files.

Procedure:

This validation was performed by the following individuals as employees of the MA State Police Crime Laboratory System:

EMPLOYEE	TITLE	UNIT
Joanne B. Sgueglia	Forensic Research Scientist, Laboratory Supervisor III	Research, Development & Training
Kimberly S. Harrington	DNA Analyst, Chemist II	DNA
Kathleen M. Gould	DNA Analyst, Chemist II	DNA
Erica L. Blais	Criminalist/DNA Analyst, Chemist II	Criminalistics/ DNA Hybrid
David Cassidy	Temporary Employee	Research, Development & Training

The validation team would like to acknowledge and thank Leanna Farnam, Lindsay Allgeier and Jennifer Rogean for their assistance. All studies utilized data sets previously reviewed manually using GeneMapper (GMID) version 3.2.1 software.

Data Sets

Three data sets were utilized to conduct the validations described herein.

I. Single Source Titration Set-refer to Appendices C and F

a. Five individuals were used to obtain a set of eight dilutions. Samples were collected via buccal swab (1) and blood on FTA paper (4).

II. Mixture Titration Set-refer to Appendix D

a. Two individuals were used to obtain two sets of nine dilutions for total concentrations of 2ng and 1 ng.

III. RSA Cases-refer to Appendix E

a. Twenty five sexual assault cases were identified to challenge the TACW system and to compare results obtained using previous manual methods. Fourteen cases included suspects and eleven were no-suspect cases.

b. Questioned sample types included intimate swabbings, or those taken from a person, to include orifice swabbings (e.g. vaginal, anal and oral swabs) and body swabs (e.g. penile, external genital, neck, breast and bite marks). Other cases contained samples from underpants, bras and condoms.

c. All cases were submitted with a victim reference but not all contained a suspect reference. Cases without a suspect were compared for CODIS upload profiles for subsequent searching of the convicted offender database.

d. Cases were solicited from the DNA unit to represent a sampling of routine sexual assaults encountered in actual casework.

e. Cases represented contained different statistical analyses to include straightforward single source statistics (Match Probability Estimate/MPE), two person mixtures with the victim subtracted out to identify an unknown of probative value (Modified Match Probability Estimate/MMPE) and more complex mixtures whereby major and minor components were unable to be separated and all genotypes were considered for each locus (combined probability of exclusion/inclusion/CPE/CPI) or were not considered to be intimate samples (20).

The titration sets were used for the sensitivity and precision studies, sections A and F, respectively. The RSA cases applied to all studies.

Case Studies and Comparisons

I. Each case was assessed for accuracy, precision, concordance and, most importantly, for the amount of information gained from the two processes.
   a. Accuracy-comparison of the following:
      i. Allelic designations—assess if correct allele calls and overall concordance of data between the two methods were obtained. Note that there are differences that occur that can be explained by the processes and inherent expectations of two different methods (e.g. more data or alleles are reported if a threshold is not employed using the TAGC).

      ii. Inferred genotypes-assess the results tables and CODIS upload data information.

      iii. Mixture weights (MW), if applicable.

   b. Precision-assessment of duplicate runs of the same data to evaluate reproducibility. Note that slight differences are expected when running the same data through the TAGC due to random sampling variation in the Markov Chain in each separate run (11).

   c. Concordance-compare and contrast the conclusions obtained using both methods. Assess if a match between a Q and K was consistent. Explanation of any observed differences (match with one method and not the other) and investigation of inconclusive data.

   d. Information gain- compare match scores between the two methods. A match score is the final statistic derived from the base ten logarithm of the LR, $\log_{10}(LR)$. The match score is the value reported using TACW (the FBI Caucasian database was used for all calculations). In order to compare the

TACW match score against the manual method, conversions were calculated using the following formulae in Table 1 (21-22):

Table 1: NRC II Equations from Chapter 4

Homozygote Genotype Frequency to account for subpopulations. (Equation 4.4a)	$P^2 + p(1 - p)\theta$
Heterozygote Genotype Frequency (Equation 4.1b)	$2pq$
Assumption of evidence and suspect from the same subgroup.	
Homozygote (Equation 4.10a)	$\frac{[2\bar{\theta} + (1 - \bar{\theta})p][3\bar{\theta} + (1 - \bar{\theta})p]}{(1 + \bar{\theta})(1 + 2\bar{\theta})}$
Heterozygote (Equation 4.10b)	$\frac{2\,[\bar{\theta} + (1 - \bar{\theta})p][\,\bar{\theta} + (1 - \bar{\theta})p]}{(1 + \bar{\theta})(1 + 2\bar{\theta})}$

i. The manual method uses a theta correction (θ) for homozygotes (4.4a) and no correction for heterozygotes (4.1b) while the TAGC uses a theta correction for both homozygotes and heterozygotes, 4.10a and 4.10b respectively (21-22).

ii. The manual result was converted to a match score by converting to the log of the frequency. Although, this would still not be a direct comparison to the TAGC score using 4.10 equations. Note that the 4.10 formulae result in more conservative estimates or lower match scores relative to the manual method. Therefore, all MSPCL statistics were converted to use the 4.10 formulae for direct comparison to the TAGC score (21-22).

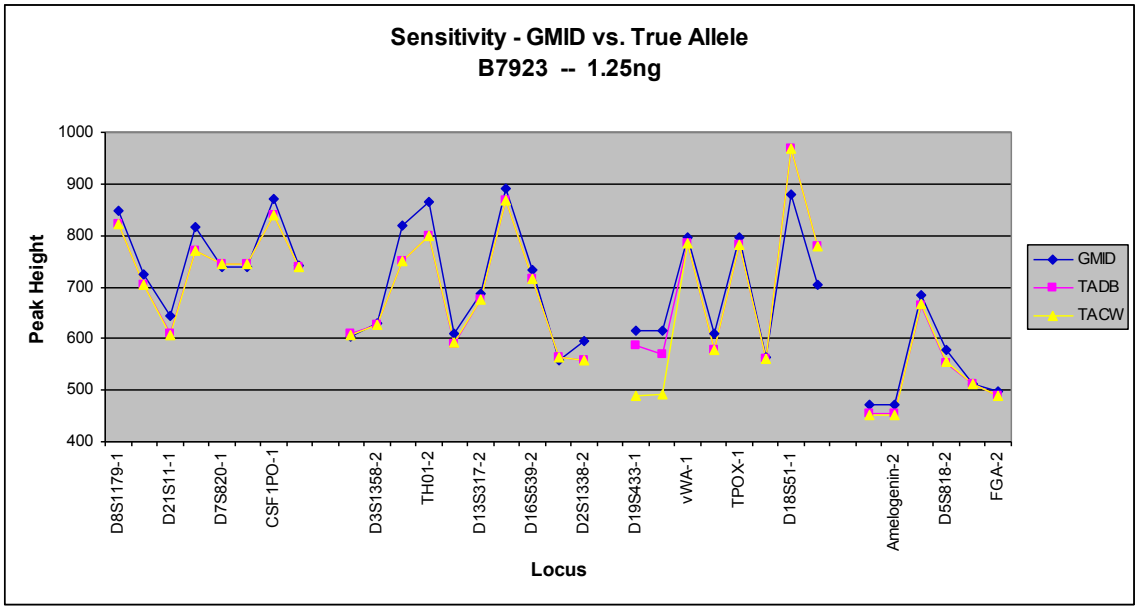
iii. The TAGC, unlike manual methods, incorporates the LR at every locus. The overall match score may be decreased due to results at particular loci. The LR at those loci may be very low or negative if the probability of the inferred genotype of the suspect is low or missing. Hence, all data is considered and the statistic will be more conservative if the data doesn't fit the model well.

A. Sensitivity Studies:

I. Peak Height Comparisons between GMID and TA softwares

Titration sets were utilized to compare peak heights obtained in relative fluorescent units (rfu) using GMID, TADB and TACW. Graphs representing the comparison of rfu values for each locus at every target were compiled (note that homozygote values were halved). The graph below (Figure 1) provides an example of the comparison for sample B7923 amplified at a target DNA concentration of 1.25 ng. The data illustrate no significance difference between the three systems throughout the range. In conclusion, TADB and TACW are comparable to GMID for measuring peak heights on a 3130xl. Therefore, data sized with TACW is reproducible and reliable. All other studies were conducted using the TACW system and associated peak heights.

Figure 1: Sensitivity Studies of TrueAllele Casework compared to TrueAllele Databank and GeneMapper ID.

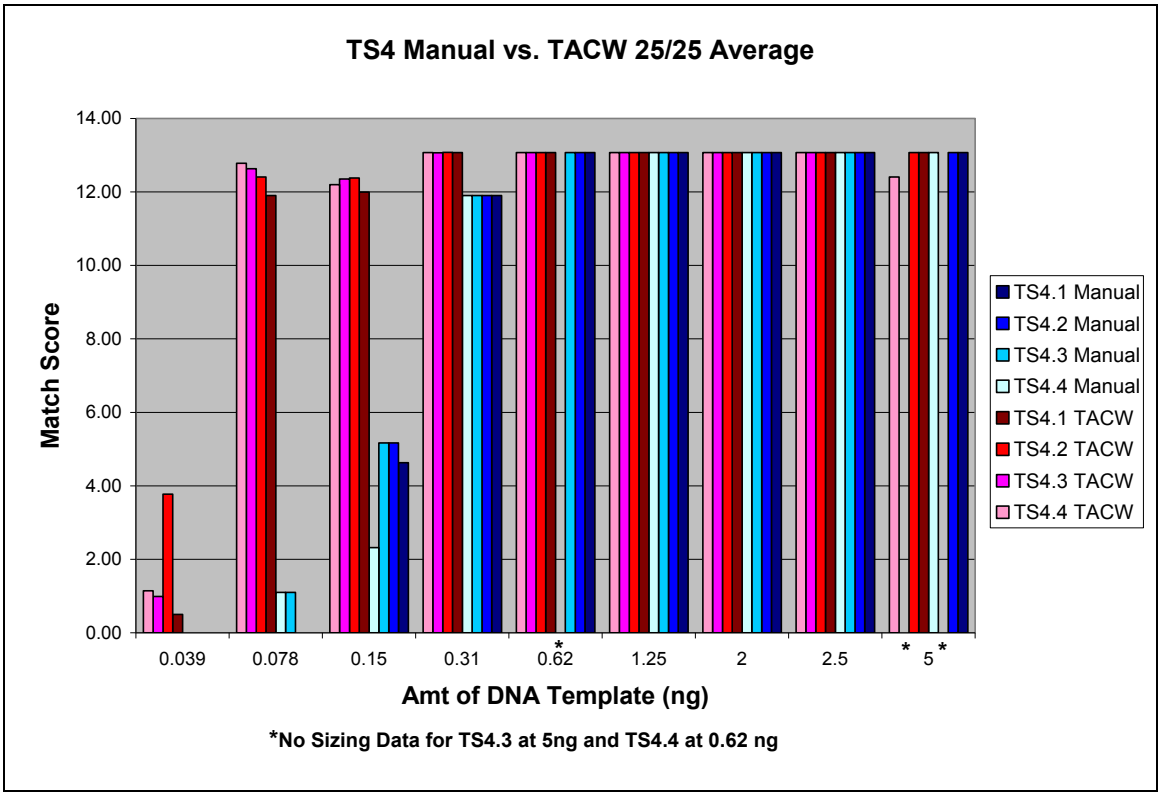


II. Single Source Titration Study

Five individual sources of DNA were amplified targeting 5, 2.5, 1.25, 0.62, 0.31, 0.15, 0.078 and 0.039 ng. Additionally, a 2 ng sample (not part of the serial dilution) was diluted separately to represent the target DNA concentration used in casework. Amplicons of the five titration sets (TS1-TS5) were run in quadruplicate (two runs on each Capillary Electrophoresis (CE), Appendix C) and results averaged. Note that all match scores were calculated using equations 4.10a and b from the National Research Council (NRC II, 1996, pp 114-115) as aforementioned. Results are evidenced in Figure 2 for TS4 below. Note that TS4 was selected as the most representative sample,

demonstrating peak heights at the expected signal of approximately 1500 rfu for heterozygote data at a target of 2 ng.

Figure 2: Sensitivity Study of Single Source Titration sets (TS4)—Comparison of Manual and TrueAllele Casework Results



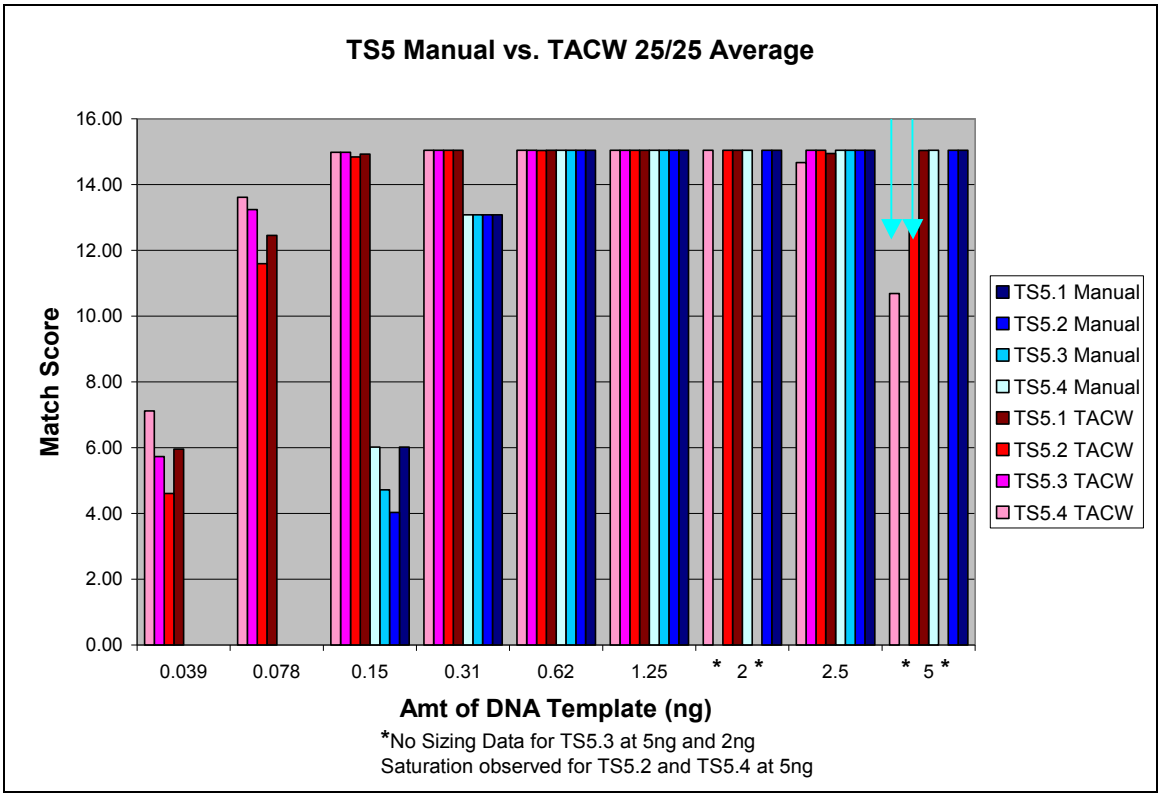
Results indicate that both the manual and TA methods produce similar match scores of 13 for single source samples with ample DNA (i.e. ranging from 5 ng to approximately 0.31 ng). This particular TS4 set contained 3 potential alleles below the instrument threshold of 165 rfu (1 @ D2 and both at FGA) at the 0.31 ng sample that accounts for the lower match score. Using target DNA less than about 0.31 ng indicates a significant decrease in the match score, or information gain, when employing manual methods. Additionally, there is more imprecision across manual runs that present as different match scores dependent on data called due to peak intensity and the use of thresholds. These data correlate with the reporting threshold set to obtain a partial profile using 0.15 ng of DNA. Targets below which will have very little information, if any, for statistical analyses (i.e. peaks evidenced in the gray zone or potential alleles below threshold are not used for statistical purposes).

The data presented clearly illustrate the TAGC is more sensitive than manual methods employing thresholds. Probabilistic genotyping does not use thresholds and is able to generate more data and hence better discriminating power from lower targets of DNA. Whilst manual methods yield limited or no information at less than or approximately 150 pg, the TACW system is able to yield substantially more information down to approximately 78 pg. However, once the data is very limited (i.e. approximately 39 pg)

the TAGC conservatively produced lower match statistics having greater variation. This is advantageous as too little DNA may be present to make comparative analyses and strong associations.

Larger amounts of DNA may also produce different match scores using the TAGC. Saturation observed at 5 ng of DNA template lowered the match score in two instances for TS5. An example is depicted in Figure 3 for TS5.2 and TS5.4 (light blue arrows).

Figure 3: Sensitivity Study of Single Source Titration sets (TS5)—Comparison of Manual and TrueAllele Casework Results



Saturation was observed in two of the automated runs whereby the match score was lowered from 15.04 generated with the three manual and one other automated run to 12.67 for TS5.2 and 10.69 for TS5.4. This occurs when the fluorescence detection system in the CE instrument is overloaded. Too much fluorescence from the DNA enters the camera and the signal produced is no longer in the linear range.

At the saturation point, the TAGC cannot accurately model the data. Genotype variation becomes more prevalent due to data distortion. Both the peaks with higher rfu signals and the baseline peak artifacts are affected. This distortion introduces uncertainty as the system must consider all data when inferring a genotype. The uncertainty in the genotype is reflected in the lower match score.

Additionally, accounting for population substructure by the use of co-ancestry coefficients or "theta" values greater than zero will lower match scores accordingly. The

theta value impacts a genotype that contains a 'rare' or 'off ladder' (OL) allele more than if the genotype is comprised of more common alleles. Laboratories normally use either 1% or 3% theta values. The match score will become lower as the theta value is increased (see Table 19). This mixture titration set was run using a theta of 3% and data containing OL alleles evidenced decreased match scores. The table below illustrates the differences in the match log (LR) observed at the D18 locus.

Sample #	Without Saturation Effects TS5.1 (5 ng)	With Saturation Effects TS5.4 (5 ng)
Genotype D18	16, 21	16, 21 16, 16 16, 21.1
0 %	2.950	2.789
1 %	2.463	1.693
3 %	2.062	0.931

At theta = 3%, the saturation effects due to uncertain genotypes results in decreasing the overall match log (LR) value by 1.13 log units.

III. Mixture Titration

Mixture titration sets were used to compare and contrast the information gained at different ratios/target DNA using manual and TACW methods.

a. Two mixture titration sets were run using the same two sources of DNA and ratios with varied total concentrations. The ratios were run at 1:1, 1:5, 1:10, 1:15 and 1:20 and vice-versa. Set 1 contained a total input of 2ng whilst Set 2 contained a total input 1ng. This assessment was made via the rfu values obtained after STR analyses, as these sets were stored from previous validation studies and were not prepared specifically for this study. Set 1 will be used for the majority of data analyses as Set 2 contained too little DNA for evaluation of the minor component beyond the 1:5 to 1:10 ratios (see mixture titration binder for additional information). The DNA target amounts for the minor component for both sets are listed in the tables below (Tables 2A and 2B):

Table 2A: Quantity in Mixture Set 1

Total = 2 ng DNA	DNA from Contributor A (ng)	DNA from Contributor B (ng)
1:20	0.095	1.905
1:15	0.125	1.875
1:10	0.182	1.82
1:5	0.333	1.67
1:1	1	1
5:1	1.67	0.333
10:1	1.82	0.182
15:1	1.875	0.125
20:1	1.905	0.095

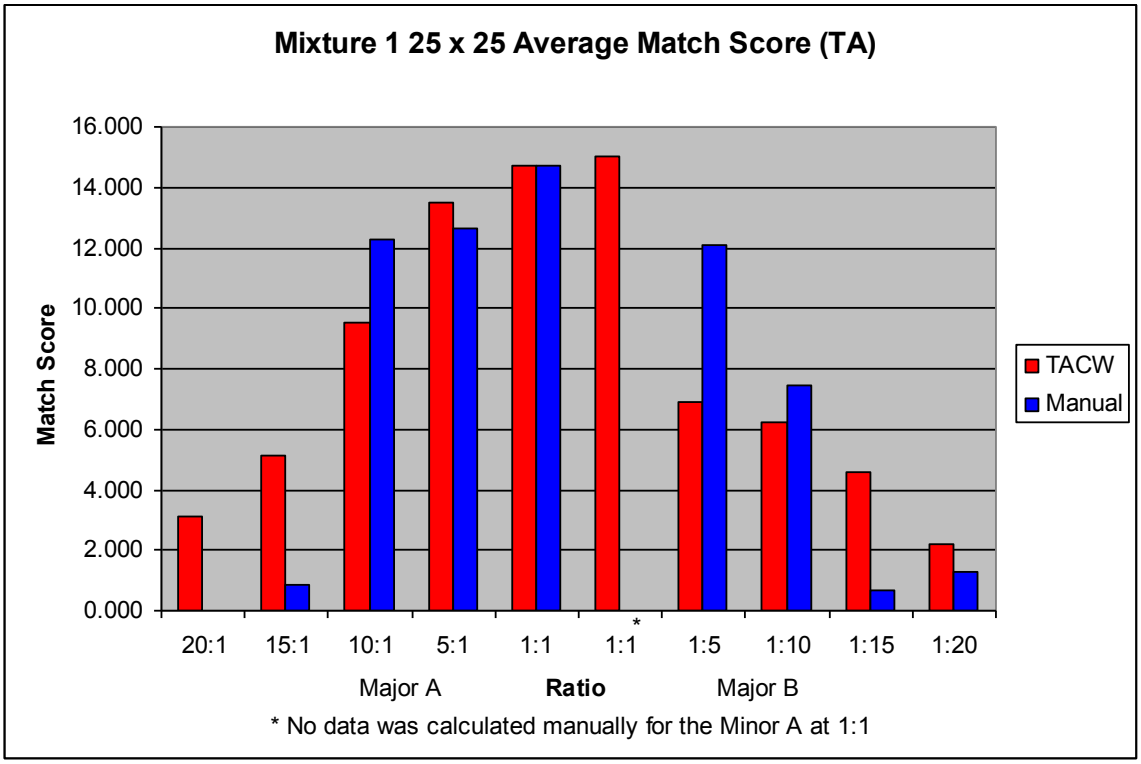
Table 2B: Quantity in Mixture Set 2

Total = 1 ng DNA	DNA from Contributor A (ng)	DNA from Contributor B (ng)
1:20	0.047	0.953
1:15	0.062	0.938
1:10	0.091	0.909
1:5	0.167	0.833
1:1	0.5	0.5
5:1	0.833	0.167
10:1	0.909	0.091
15:1	0.938	0.062
20:1	0.953	0.047

i. Although there are DNA targets equal to or less than 100 pg levels for the larger ratios, the TAGC did converge for the entire set of data. Somewhat higher convergence scores were obtained for some runs at the 1:15 and 1:20 ratios but were in the acceptable range. Hence, all data for this mixture set can be used for analyses (See Figures 33 and 34).

ii. The average match score from the two regular runs with the TACW system and the manual match score obtained from MMPE worksheet deconvolutions are depicted in the figure below (Figure 4).

Figure 4: Comparison of Manual and TrueAllele Casework for Mixture Set 1



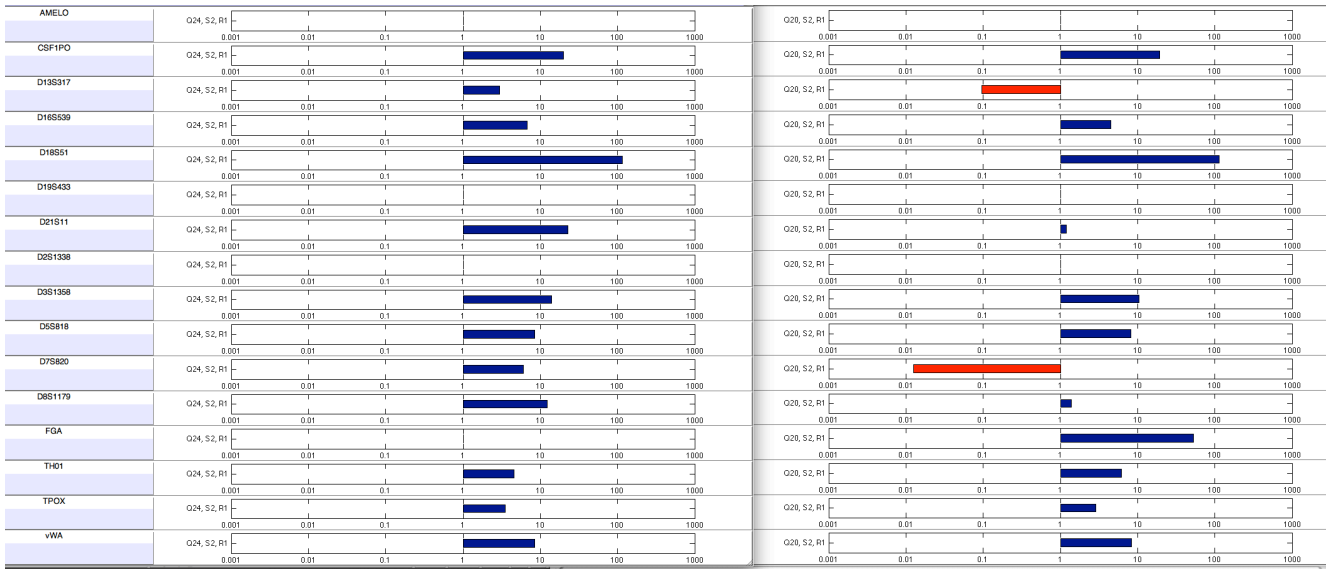
iii. The mixture set ratios were run in both directions, with donor A as the major on the left side and donor B as the major on the right side. The major donor on each side was run as the victim with the minor as the suspect for deconvolution to mimic casework scenarios. All ratios are read as Donor A : Donor B. The data exhibits differences dependent on which donor is the major and minor. In Figure 4 there are slightly higher match scores when deducing the minor on the left (donor B). It appears that the minor component A on the right has lower mixture weights than the minor component B on the right (See Figure 34). Differences in match score may also be indicative of the allelic combinations and overlap amongst the donors due to their respective genotypes. The frequency of various alleles/genotypes will affect the overall match scores. Although, after investigation of the rarity of component A and B, they were not significantly different (15.04 and 14.69, respectively). Further investigations to examine differences are described below.

iv. There are some significant differences observed at the 10:1 and 1:5 with the manual method having a higher match score than the TAGC. The differences observed at the 1:5 ratio were investigated at several loci. The TAGC had lower likelihoods for the suspect genotype (minor component A) due to the data presented. The peak heights for the given genotypes did not fit well and explain why the TAGC gave lower match scores. Examples were evidenced at the two loci that resulted in negative match

scores at D13 and D7. Both components A and B shared the same 11, 12 genotype at D13. One would have expected the 11 and 12 to be relatively similar, yielding a sister allele balance of over 70%. To the contrary, the 11 was 2994 rfu and the 12 was 1608 rfu, resulting in an intralocus balance of 54%. The TAGC, using the minor mix weight established to be about 13%, had a minor genotype of 11, 11 to be much more probable (96%) than an 11, 12 (3%). The manual method had both genotypes as possibilities and would give them an equal probability of 50% each. In this example, manual at 50% and TAGC at 3% would equate to match scores of 0.471 and -1.034 respectively. Several other examples of data (D7, D8 and D21) that did not fit the suspect genotype well account for an overall lower match score as displayed in the figure of LR's below (Figure 5). The MMPE calculation performed manually was converted to a match score using the TAGC and hence is depicted below using the Report module Display LR window interface for a direct comparison of the two methods.

Figure 5: Comparison of Manual and TACW Likelihood Ratios in the Report Module

MANUAL **TAGC**



v. The same loci that resulted in lower match scores at the 1:5 were assessed in the 1:10. The data presented in this mixture set did fit well and resulted in a match score that was not as disparate when compared to the manual score. The differences between the 1:5 and 1:10 for the affected loci are listed in Table 3. Although, a different locus, D3, did not fit the suspect genotype well and rendered a negative match score for the 1:10 but fit well in the 1:5.

Table 3: Suspect Genotype Probabilities

Locus	Suspect Genotype	TACW 1:5	TACW 1:10	Manual 1:5 and 1:10
D13	11, 12	3%	36%	50%
D7	8, 9	0.1%	50%	50%
D8	11, 13	6%	76%	50%
D3	17, 17	62.3%	7%	100%

vi. When investigating the differences observed at the 10:1 condition, the same scenario was evidenced. Three loci, D3, D5 and D8 yielded negative LR's with the TAGC with lower probabilities where the genotype of the suspect was not a good fit whereas manual methods would have much higher probabilities (e.g. D8 was 0.7% for TAGC whereas manual methods had 3 potential genotypes and gave the suspect genotype 1/3 or 33%). Therefore, the disparities in the graph evidencing higher manual match scores are due to the presentation of the data. The TAGC is objective and assesses the quantitative peak heights from the electropherogram and may or may not 'fit' well with the suspect genotype dependent on the amplification, stochastic effects and final signal intensity of each peak.

vii. There appeared to be an anomaly in that the manual 1:20 had a higher match score than the manual 1:15 (note this did not occur at the 20:1). Review of the data found differences occurred at the D21 and TH01 loci. There were peaks from the minor contributor that resulted in higher rfu's at the 1:20 than the 1:15 (31.2 in the stutter position at D21 and the 6 allele at TH01) such that the alleles were designated at the 1:20 and used for statistical analyses but would not have been used for statistics at the 1:15 (masked by stutter or a potential allele below threshold). This phenomenon has been observed in other instrument validations (3130xls: Ned, Xena, Jim and Rosalind) and may be attributed to peaks being somewhat higher if baselines are elevated. In comparing the electropherograms of the 20:1 and 1:20, the 1:20 did appear to have more noise and pull up effects compared to the 20:1.

viii. In comparing the automated to manual methods as the minor component is decreased, there are significant differences at the 20:1, 15:1 and 1:15 with the TAGC having a higher match score than manual methods. This is expected as there are very low DNA targets at these ratios that would be below established thresholds (note that manual thresholds yield partial profiles at approximately 150pg).

In summary, the TAGC meets the sensitivity requirements needed to deconvolute two person mixtures at various ratios. As expected, overall results indicate that less information is gained as the target level of the secondary contributor is decreased.

IV. Dynamic Range

Data can be assessed for a wide range of target inputs (i.e. rfu values) using the TAGC. The limitations are somewhat similar to manual methods occurring at the very high end when saturation exists and at the low end when a minor component is below approximately 5%. Once peaks are off scale the modeling is no longer linear and cannot develop accurate weights or peak distributions. Although, as stated above from the titration studies, the TAGC does not use thresholds and the modeling allows for the use of a significant amount of data on the low end. When peaks are closer to baseline the variation encompasses many possibilities and becomes more uncertain.

In investigating the dynamic range, questions arose regarding the need for enhancing signals and deciding which .fsa files should be analyzed for a given sample/item (e.g. 10 sec versus 20 sec injections). In working with Cybergenetics, considerations regarding the need to change the scale for our data were addressed. The differences in our 10 and 20 seconds were assessed and determined to have minimal impact and no scaling effects were observed. Hence, there was no need to change the scale as both sets of data could be run under the same conditions. The enhancement studies are described in the following section under optimization and calibrations.

B. Optimization/Calibration:

I. Enhanced Methods-Manual methods using thresholds may employ strategies to increase signals to obtain more information.

a. Injection Time

i. Manual methods use a standard injection time of 10 sec with the option to increase to 20 sec in an effort to obtain more data above thresholds.

ii. A study was conducted analyzing both the 10 and 20 sec injection data from breast swabs for case RSAV-08 (See comprehensive section p. 22). This sample contained a third trace contributor that would complicate analyses. Without evaluating the 3rd contributor, the data resembled close to a 1:1 mixture. TAGC had the minor weight as 43.2% when processed as a two person mixture (final 25/25 setting). This item was run at fast and long for 10 sec and run fast at 20 sec. There was no significant difference in any of the resultant CODIS upload profiles. Hence, the TAGC does not need to enhance the data if the peaks are already present in the 10 sec run. If no additional peak information is present, there is no need for enhanced injection time.

iii. In comparing the match scores, there was a slight increase in information gained for the 20 sec injection data with the TAGC (13.51 to 13.87). Although, there was a very significant difference in the match score obtained using the TAGC versus manual methods (over 10 log units or over a billion fold) due to the use of a CPE calculation for greater than 2 contributors (see section D on Information Gain).

iv. Current manual methods do support the use of combining data from separate injections to make a composite profile. Generally, the use of TAGC will not require additional injections unless additional peak information is detected.

b. Target Input

i. Manual methods use an optimal target DNA of 2 ng to achieve signals for heterozygotes at approximately 1500 rfu. If a mixture of DNA is present and the minor(s) are probative, amplification with more target DNA may be advantageous to enhance the minor(s) signal without saturation of the major component (usually the non-probative victim in sexual assault evidence).

ii. A study was conducted analyzing both a 2 ng and 4 ng amplification from an oral swab for case RSAV-05 (see comprehensive around p. 20). This study illustrates differences obtained using the TAGC at the D2S1338 locus between the 2 ng and 4 ng data sets. The 2 ng yielded a single genotype at 100% probability of a 17, 20 whereas the 4 ng yielded a more probable 17, 17 close to 100% (with a minor probability of a 17, 20). The suspect genotype was determined to be a 17, 20. The 4 ng did have a 17, 20 as an inferred genotype but the confidence interval had to be opened beyond 95% to 99.9%. In analyzing the data it was determined that the peak heights at that locus differed between the 2 ng and 4 ng amplification and caused the TAGC to use whatever quantitative peak height information was supplied. In conclusion, the inferred genotypes and associated likelihoods were derived from the data and do represent the quantitative information therein. Both the TAGC and manual method appear to be concordant in what best fits the quantitative data regardless of the actual suspect genotype obtained after deconvolution of the questioned item.

If more quantitative data (additional peak information) is presented to the TAGC a higher match score is obtained. For example, the match score for Case RSAV-05 oral swab more than doubled from 7 to 16.

iii. Future studies will investigate the possibility of combining data from separate amplifications to obtain more accurate modeling and inferred genotypes by employing a joint likelihood ratio. Current protocols do not

support the use of combining data from separate amplifications to make a composite profile.

c. Sample Concentration

i. Manual methods often use a micron filtration device to concentrate extracts that may contain limited quantities of DNA in an effort to optimize results. Concentration of the extract from larger volumes, ranging from 50ul up to 400 ul, to approximately 10 ul for amplification are performed to use all the DNA present in the PCR reaction.

ii. There was only one case in the study that contained both a non-concentrated and concentrated extract from a pair of underpants (RSAV-21). Note that this sample contained more than 2 contributors (i.e. a trace 3rd) and may not be the most suitable example to assess if additional information from concentrating a sample would be helpful. In fact, it may complicate the data if the third contributor was non-probative and the information from the secondary contributor was informative to the case. The primary and secondary resembled close to a 1:1 with TAGC calculating the secondary source at approximately 44%.

iii. This was a no suspect case and hence, no match score was obtained. Data was evaluated to determine the suitability for searching. Both samples gave inferred genotypes suitable for CODIS upload at 95% for all loci. Although, the concentrated sample gave more discriminating information whereas the non-concentrated sample resulted in search information containing more inferred genotypes (as evidenced in Table 4 the non-concentrated upload had 3 loci with 3 alleles whereas the concentrated sample only had 1 locus with 3 alleles). There was no upload information when the samples were assessed manually due to the presence of a trace tertiary contributor and too many inferred genotypes for CODIS (See Concordance section—Section E).

Table 4: CODIS Uploads for Concentrated and Non-concentrated Sample in Case RSAV-21

Best 25x25	V+1	Manual	TACW-95% Non-Concentrated			TACW-95% Concentrated		
Case	Marker	Allele	Allele 1	Allele 2	Allele 3	Allele 1	Allele 2	Allele 3
RSAV-21	D8S1179	No Upload	13	14		13	14	
RSAV-21	D21S11		30			30		
RSAV-21	D7S820		10	11+		8	10	11+
RSAV-21	CSF1PO		10	12		10	12	
RSAV-21	D3S1358		17			17		
RSAV-21	TH01		7	9.3		7	9.3	
RSAV-21	D13S317		12+	14		12+	14	
RSAV-21	D16S539		12			12		
RSAV-21	D2S1338		16	17+	23	16	17	
RSAV-21	D19S433		13+	14	14.2	13	14.2	
RSAV-21	vWA		17+	18		17	18	
RSAV-21	TPOX		8	9		8	9+	
RSAV-21	D18S51		14+	15		14	15	
RSAV-21	AMEL		X	Y		X	Y	
RSAV-21	D5S818		12	13		12	13	
RSAV-21	FGA		20+	21	23	20	23	

iv. As stated above, future studies will investigate the possibility of combining data from separate amplifications, which may encompass extracts that were concentrated and non-concentrated.

In summary, enhanced methods are only needed if additional information is presented to assist the TAGC. Generally, enhanced methods may not be needed for future data to be generated if the objective was to increase existing peak heights to raise them above thresholds. The TAGC yields similar data for different conditions as long as the data was evidenced in the dynamic range of the CE instruments. This system may not require additional injection times, more template or concentration of samples unless peaks are observed that were not under standard conditions. If more than one condition was run, it may behoove the laboratory to analyze all conditions to combine the data to yield more accurate inferred genotypes from several amplifications.

II. Parameters

a. Phase I (Evaluation of Settings: Fast, Long and Longer)

Initial studies were run with one or more parameters dependent on the level of complexity of the data. The number of contributors in the mixed profiles, the associated mixture weights of each contributor and which of the components is most probative are factors that must be considered in determining optimal data output.

i. Software Settings

A. The Setting tab dictates how many cycles the computer should perform when solving the questions. The run times can be adjusted dependent on the complexity of the data. There is an initial setting that allows the computer time to 'burn in' or how much time the system searches for the right region of mixture weights. There is a 'read out' that follows that determines how long the system gets to sample within that region. This allows the computer more time to complete the process of deconvoluting the mixture and better determine mixture weights and inferred genotypes. More information can be extracted from difficult data when there is more time for the system to sample. Note that the computer is solving the problem or question as it was posed by the operator. Different results will be obtained dependent on the number of contributors in the actual data.

B. The parameters provided with the initial version of software are approximately as follows (dependent on complexity of data):

Table 5: Settings for the initial set of runs in Phase I

Setting	Burn In	Read Out	Approximate Run Time (hours)
Fast	2,000	8,000	1.5
Long	10,000	40,000	7.5
Longer	50,000	50,000	16.0

C. Subsequent studies were run with a new setting provided to consider for all data using the following condition:

Table 6: Subsequent Setting for Phase II studies

Setting	Burn In	Read Out	Approximate Run Time (hours)
Regular	25,000	25,000	7.0

D. The initial study incorporated a 'level of complexity' scheme that was developed if various settings were to be considered. The initial write up of this scheme is contained in Appendix G and is no longer to be utilized and is provided for historical purposes and edification of the validation process under these previous settings. The new 'regular' setting is currently validated (see phase II below) and will be used for all levels of data.

b. Phase II (Evaluation/Performance of Regular Setting)

All data sets for the titration and case studies were rerun and evaluated at the regular setting. Each run was performed in duplicate and used for the precision study (Section F). Comparisons were made to the phase I data with regard to the match score obtained for suspect cases and the CODIS upload profiles for all cases. The binder for this study contains the comparisons for each case and can be referenced accordingly.

A summary table (Table 7) for the number of contributors and mixture weights obtained for all the RSA Cases/Items using the 'regular' setting is described below.

Table 7: Summary of mixture weights and number of contributors for the RSA Cases.

1 Contributor		
RSAV Case/Item # **Suspect –Red** **No suspect- Blue**	**Mixture Weight of Foreign Contributor**	**Comments**
13E2-sperm	N/A	Single source at 7 Loci
2 Contributors		
Major Suspect- **Percentages greater than 50%**		
11	84.4	
3	70.6	
18	69.2	
15	66.5	
13E	55.9	
17	51.7	
Minor Suspect **Percentages between 5% and 50%**		
19	39.9	
14	38.5	
7	34.5	
20	32.1	
25	31.9	
4	22.4	
5E2	17.4	
16	15.9	
Trace Suspect **Percentages approximately 5% or less**		
9	6.0	
2	4.7	
1	2.7	
6	3.6*	* average of longer runs

3 Contributors Percentage of probative contributor(s)		
8	44.5	Close to 1:1 with trace 3rd
21	43.2 or 4.6	Close to 1:1 with trace 3rd
10	33.0 (S1)	Approximate 65/35 with trace 3rd
12	36.3 or 14.2	Complex mixture may be >3
13E3	12.4 and 8.8	Two suspects
13E2-Epi	8.4 and 7.3	Two suspects
13E4	5.6 and 1.9	Two suspects
22	4.6 or 2.3	
24	20.5 or 2.1	
Greater than 3 Contributors		
23	26.1 (V +1)	May contain 4 sources
5E1	50.4 (V + 1)	May contain 4 sources

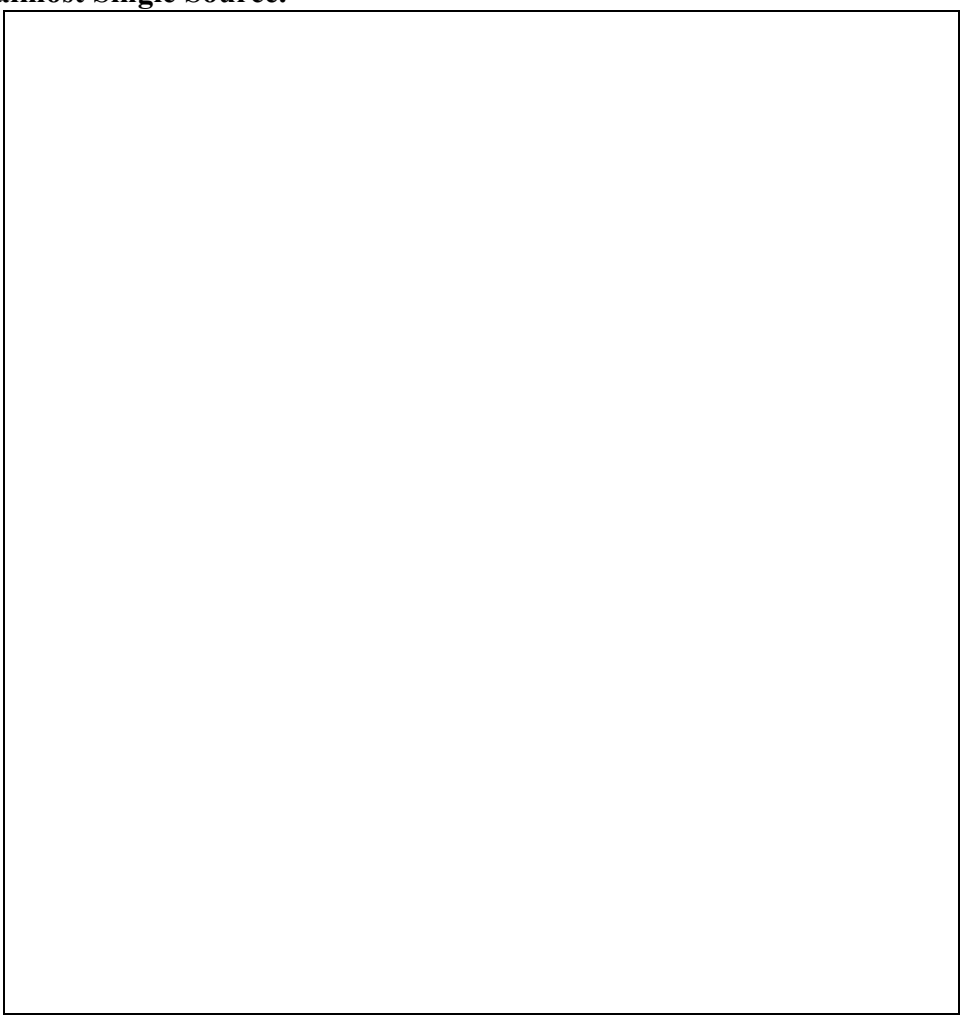
Cases were categorized after assessment of the number of contributors and mixture weights. Examples are depicted in the figures below, representing each level of complexity, to illustrate the amount of information gained and/or probative information deduced. All data was run using the 'regular' setting for each type of case classified as easy, intermediate or complex (see Table 16).

i. Easy

'Easy' cases can be described as one or two person mixtures that will yield a lot of information or relatively high match scores of over a trillion (average 13.35). Items in this category can be deduced to single source genotypes with high probabilities or almost single source with some loci containing more than one GT. These cases contain relatively certain data and the probability distributions normally contain fewer genotypes with higher probabilities.

Figure 6 below depicts an example of a probability distribution of inferred genotypes after deconvolution of the mixture using the victim reference. Case RSAV-17 consisted of a 52% contribution of the suspect that was deduced to a single source genotype at all but 1 locus (D19). This is represented by the blue bars going to 100% probability (all to the right up to 1) for the given single source genotype at each locus. The TAGC easily deduced the suspect genotype from a 1:1 mixture that would have been time consuming to accomplish manually.

Figure 6: Probability Distribution of Case RSAV-17, which is in the Easy Category and is almost Single Source.



ii. Intermediate

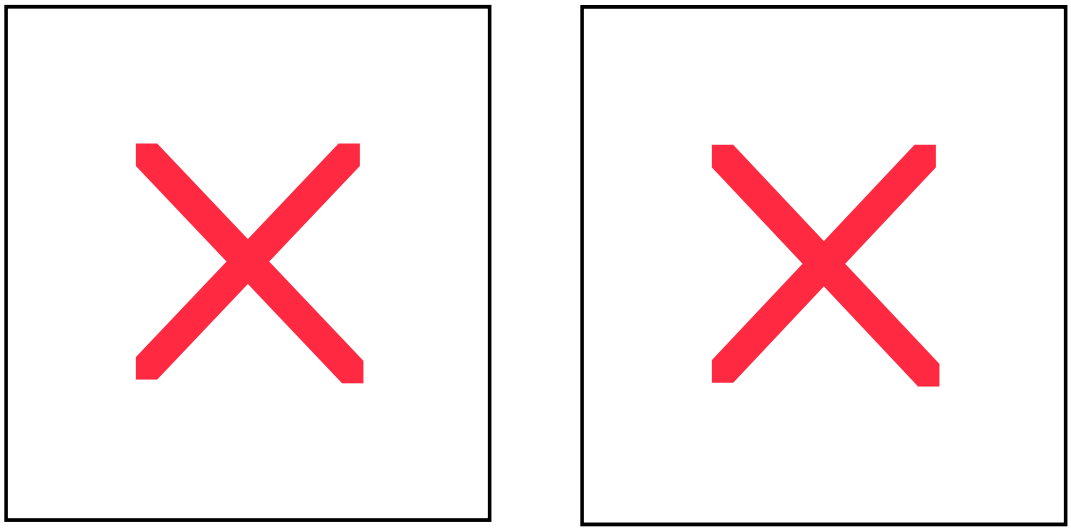
'Intermediate' cases can be described as two or three person mixtures that will yield a moderate amount of information or lower match scores of over a million (average 8.20), relative to easy cases. Items in this category may consist of two person mixtures with a lower minor component that may yield several genotypes at most loci. These cases contain some uncertain data and the probability distributions normally contain more genotypes with lower probabilities. As data becomes less straightforward it is best to run with more iterations to allow the TAGC to solve the problem. More accurate and reliable information should be obtained with more run time. The regular setting has been evaluated and produces comparable or better results for cases in this category.

The figures below correspond to the inferred genotypes derived from case RSAV-10 which consisted of a three person mixture with mixture weights of 64.4/33/2.5.

Figure 7A is the inferred genotype distribution for the secondary component of the mixture and Figure 7B is for the tertiary trace contributor.

Figure 7: Probability Distribution of Case RSAV-10, which is in the Intermediate Category

A. Contributor 3 (secondary contributor) **B.** Contributor 2 (tertiary contributor)



It is evident that the orange bars have high probabilities corresponding to a good deduction of the potential profile of the 33% secondary contributor. On the other hand, the blue bars reflect lower probabilities and many inferred genotypes at each locus corresponding to uncertainty in the data for the trace tertiary contributor. Data at this level (approximately 70 rfu) may be considered insufficient and may not be suitable for comparison or CODIS upload.

iii. Complex

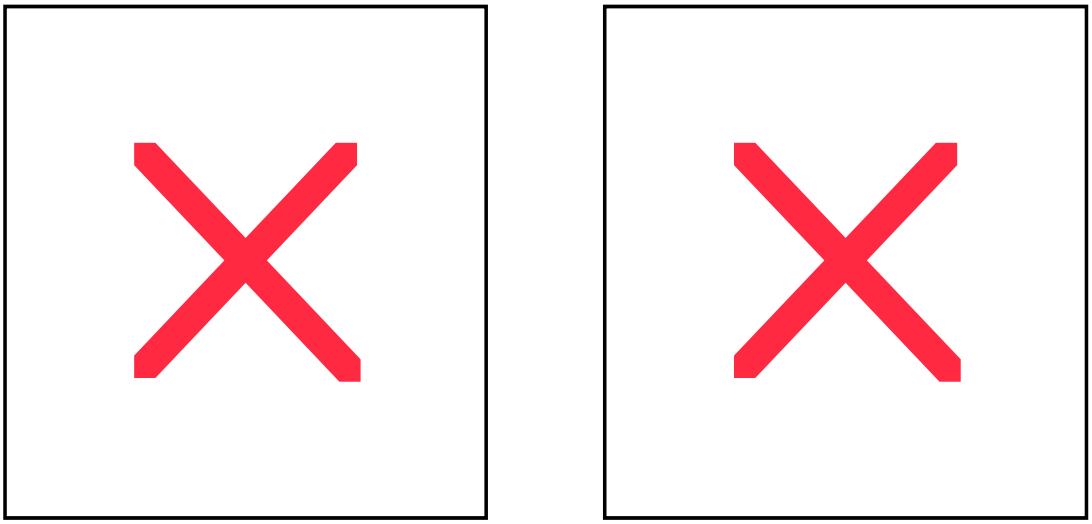
Complex data presents additional challenges to the system. There may be three or more donors and mixture weights that make it difficult to separate the components. Since the data was processed as 2 or 3 donors (V + 1 and/or V + 2), it may not accurately reflect the number of contributors and creates an inherent issue with allele overlap and the associated quantitative information used for the modeling. If the case is unduly complex, one may opt to rerun at the regular setting and/or use the longer setting to determine if the TAGC can provide a reliable solution.

The probability distribution for complex data is shown for case RSAV-22 in Figures 8A and 8B for the secondary and tertiary contributors respectively. The data was run as being assumed from three donors (V +2) and the TAGC calculated the mixture weights to be 93.1/4.6/2.3. Both minor contributors are trace and render more uncertainty, low probabilities and more possible genotypes for a contributor to this mixture. This was a no suspect case and therefore no

match score was obtained. If a suspect were to be identified for comparison, a low match score would be expected. This data would be considered too complex and uninformative for a CODIS upload under the given assumptions.

Figure 8: Probability Distribution of Case RSAV-22, which is in the Complex Category

A. Contributor 3 (secondary contributor) **B.** Contributor 2 (tertiary contributor)



In conclusion, the regular setting is acceptable as the parameter to run in every case.

C. Manual Evaluations

I. Data Assessment- is conducted prior to processing the request.

Manual input is required to assess the data in the data and/or request module to;

a. Develop strategies to ask questions regarding the number of contributors in request mode (See Rerun section—Section G).

II. Results Assessment- is conducted after the data is processed.

a. Convergence

i. To help assess the quality of the computer's variable sampling, we can view the graphed *History* of the computer's sampling of mixture weights, known as a Markov Chain.

ii. The weights modeled for each contributor are recorded and should settle on an answer by the end of the cycling process. This settling is called convergence. A convergence graph is the number of iterations or cycles

on the x axis versus the percentage weight on the y axis. Each contributor in the mixture is plotted in a different color. One can assess the plot to discern how well the TAGC has resolved the components. Note that the number of components or individuals is based on the operator assessment and the computer is solving the problem based on the question posed (e.g. a V + 1 question will assume there are two total donors in the mix).

iii. The settling of the data is represented by a *Convergence* statistic. If the chains have not settled by the end of cycling, have not moved at any point, or the convergence is over 1.20, then the computer has not sufficiently sampled the data to produce a reliable statistic.

iv. Additional runs at the same setting (e.g. duplicate or triplicate) and/or with the longer setting may assist in comparing data for reproducibility. If the data displays similar and acceptable values for convergence, mixture weight, variance/standard deviation and match score in the duplicate/triplicate run(s), then the results may be reliable. If it is not reproducible, the problem may not be solvable and may be considered inconclusive (see Rerun section—Section G).

v. Identifying convergence

It may sometimes be challenging to tell when a Markov chain is converged (14-19, 23). Below are several contrasting examples of properly and poorly converged Markov chains. When a sample has converged, the graph represents visually that the Markov chains have explored options and settled in on an answer.

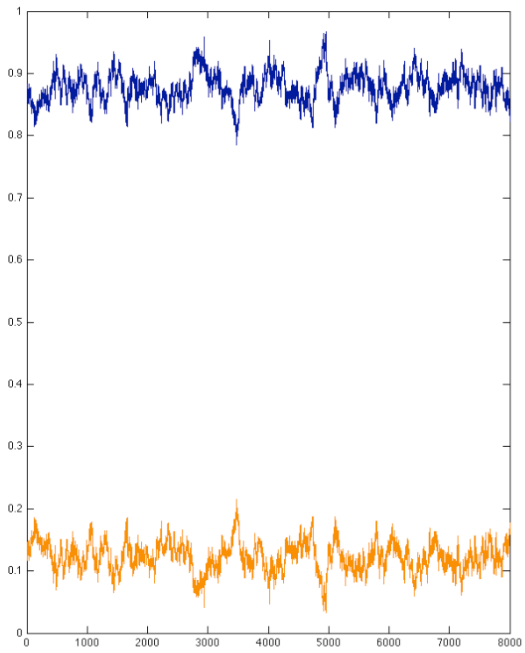


Figure 9. *Example of a converged Markov chain*

Poor convergence can indicate the need for further action. Often, the visual representation can identify when a Markov chain needs more time for sampling. When this occurs, the Markov chains look as if they have wandered and never settled on an answer.

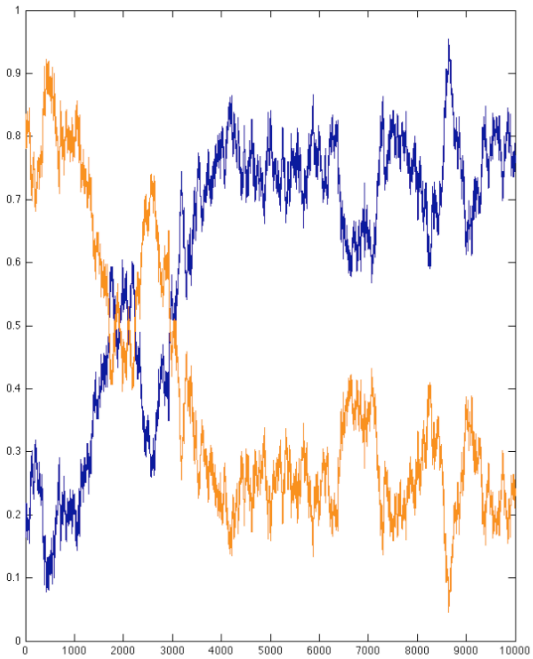


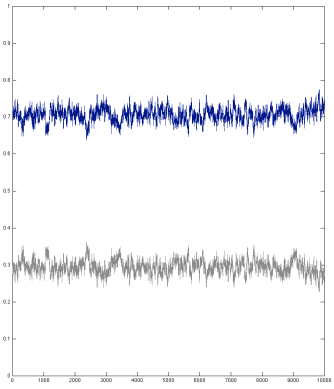
Figure 10. *Poor Convergence*. This chain needs more sampling time.

In Figure 10, the chain started to sample the right region, but needed a longer sampling time in order to converge on the correct answer.

vi. Case Examples-Generally, more divergent donors make for an easier deconvolution as evidenced in the approximate 70%:30% mixture below (Figure 11). If there is ample template DNA but the weights are similar to a 1:1 mixture, the separation of the components is more challenging (Figure 12, RSAV-13, vaginal swab). Once the data becomes more limited, due to lower targets of DNA (Figure 13, RSAV-01) or more than two contributors with two of the targets having similar weights (Figure 14, RSAV-13E3, right and left face cheek swabs, the problem is more difficult to solve. Additionally, presenting the TAGC with an incorrect number of donors may result in graphs that cannot converge on an answer (Figure 15, RSAV-05 E1).

Figure 11: Example of a 70:30 mixture.

RSAV-03 Convergence



Regular
25000/25000

Mixture Convergence: MSP...

Template	1	2
RSAV-03_E_V	1.000	1.000

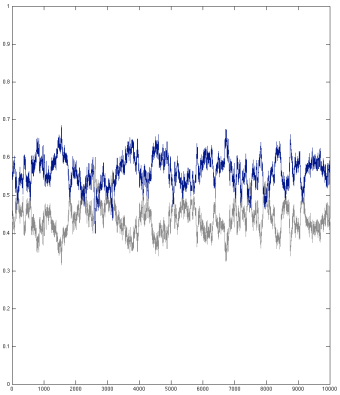
Figure 12: Example of a mixture weight of two persons close to 1: 1 (43.8%:56.2%)

RSAV-13 E Convergence (Vaginal swab)

Regular
25000/25000

Mixture Convergence: MSP

Template	1	2
RSAV-13_E	1.001	1.001



Longer
50000/50000

Mixture Convergence: MSP...

Template	1	2
RSAV-13_E	1.007	1.007

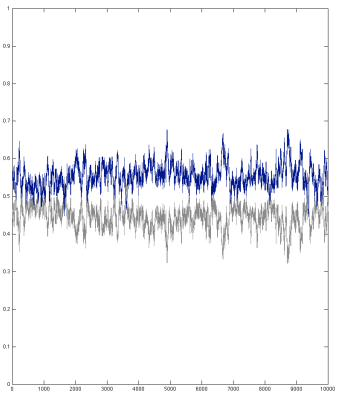


Figure 13: Example of a two person mixture with a very limited minor component (2.7%).

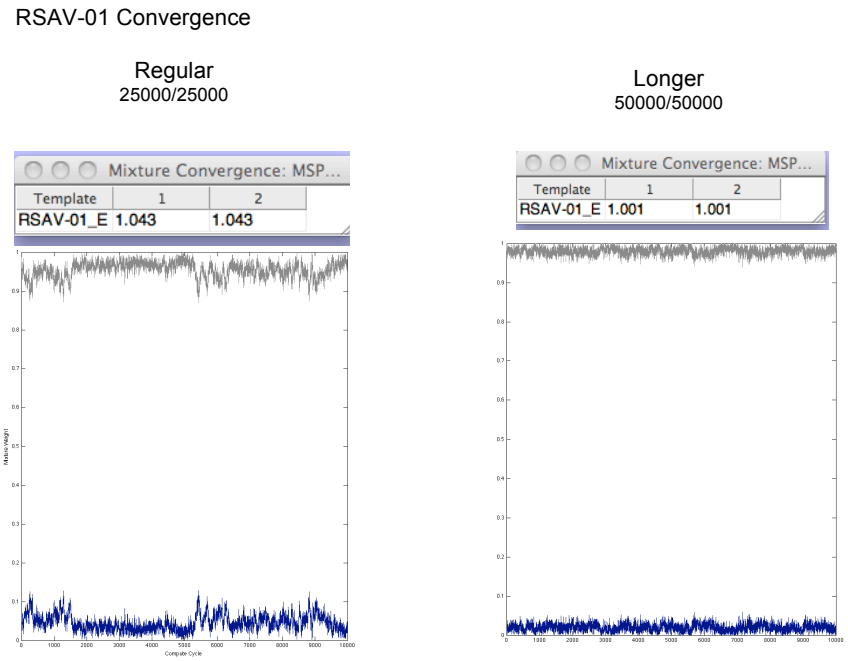
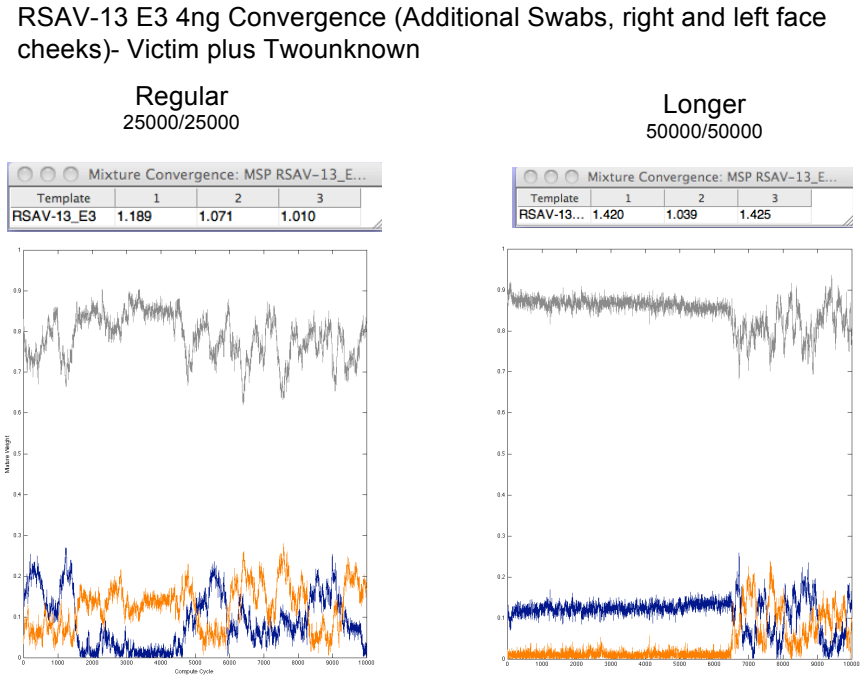


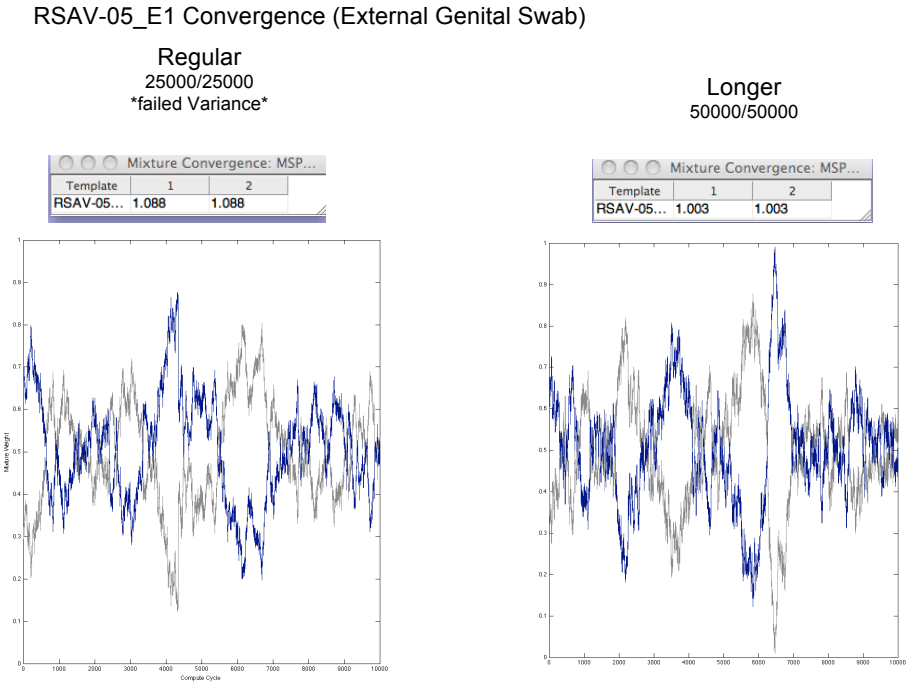
Figure 14: Example of a three person mixture with weights of 78.1%:12.4%:8.8%



Note the longer run had convergence scores above 1.2 for contributor 1 and 3.

The TAGC designates the known contributor (e.g. victim) in gray, and randomly (not according to mixture weight or the secondary and tertiary source) designates blue to a second contributor and orange to a third contributor.

Figure 15: Example of a mixture of at least 4 individuals that was run as two.



Although the convergence score is passing (less than 1.2), there were very high sampling areas. The blue and gray lines have traversed a large distance on the y axis and start and end in different areas/weights. The standard deviation of mixture weight is another parameter that should be assessed in addition to the convergence statistic.

b. Variance/Standard Deviation

The *Variance* table contains additional information such as the standard deviation, which represents the variation across the observed locus experiments. The variance measures the degree of spread or mixture weights that are being explored during the iterations. A low standard deviation means that the mixture weight is roughly the same across the loci. The variance should exhibit some spread as the TAGC needs to search and explore to settle in on a solution. If the computer did not search a wide enough area, the data should be evaluated more closely. Generally, standard deviations of less than 0.01 were observed for complex data in our set and may alert the user that such data may not be acceptable. Note that the convergence statistic may appear to be very good, approaching or reaching 1.0. Therefore, it is important to assess the convergence score and standard deviation simultaneously for a comprehensive evaluation of the data. It is somewhat counterintuitive as one may expect that the lower the variance, the tighter or more precise the data and the better the result. Although, in this instance the Markov Chain needs to

exhibit some variance to ensure it has considered a range of mixture weights to obtain more accurate answers.

In Figure 16, the top graphs depict data that appears to have a very good convergence statistic of 1 for all components but only sampled a small area as evidenced by the standard deviation, narrow histogram and tight chain. This data was run at the regular setting and would require a duplicate run at the same setting or consideration of a re-run at the longer setting, if necessary.

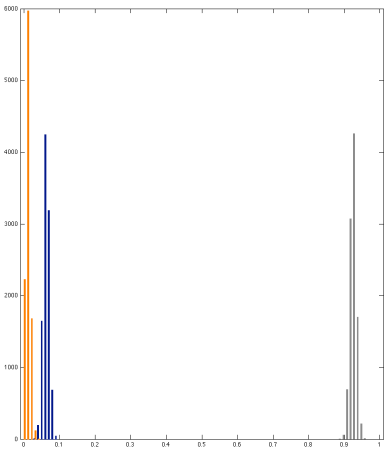
Figure 16: Examples of Good and Bad Convergence Data.

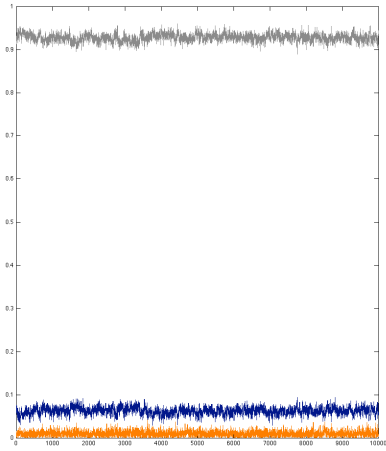
Variance STD too low, Histogram narrow curve, Chains Stuck

Template	Contrib	Weight	Stdev	95% Interval
RSAV-13...	1	0.928	0.009	[0.910, 0.946]
RSAV-13...	2	0.063	0.009	[0.045, 0.081]
RSAV-13...	3	0.010	0.007	[-0.004, 0.024]

Mixture Convergence: MSP RSAV-13_E...

Template	1	2	3
RSAV-13...	1.001	1.001	1.000





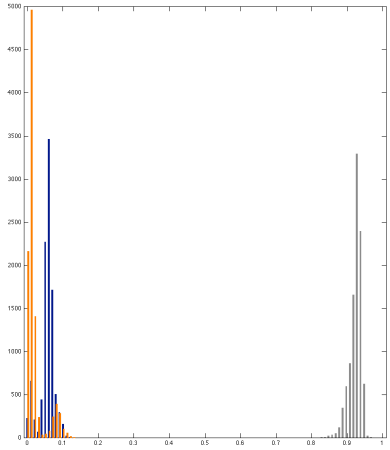
Variance STD, Convergence and Chains Are Acceptable

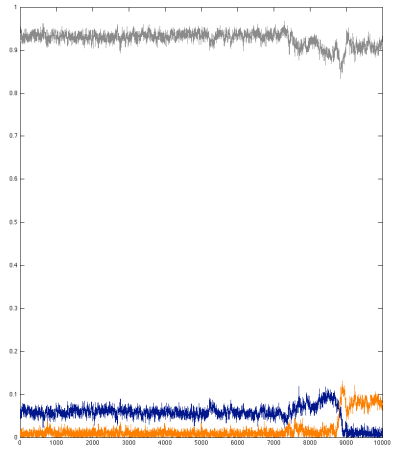
Mixture Variance: MSP RSAV-13_E4_V_2_25x25_run1.2

Template	Contrib	Weight	Stdev	95% Interval
RSAV-13...	1	0.926	0.016	[0.895, 0.957]
RSAV-13...	2	0.056	0.020	[0.017, 0.095]
RSAV-13...	3	0.019	0.031	[-0.042, 0.080]

Mixture Convergence: MSP RSAV-13_E...

Template	1	2	3
RSAV-13...	1.158	1.027	1.157





The bottom graphs (Figure 16) exhibit better sampling reflected by a higher standard deviation, wider histogram distribution and chains.

Additionally if the SD becomes too large, it may be a more difficult problem to solve as the TAGC is searching a very wide range of mix weights. Higher SD's may also be indicative of an incorrect assumption of the number of contributors. The modeling may not fit the data well when trying to quantitatively explain the peak weights with a different number of contributors. Our data exhibited higher SD's, greater than approximately 0.10, for some samples (see Table 8).

Table 8: RSA Cases with High Variance Standard Deviation.

Cases with High Variance/SD (>0.10)	# of Contributors	Mixture Weights (1:2 or 1:2:3)	Variance SD	Process	Level of Complexity
RSAV-07	2	60.8% : 32.8% : 6.4%	0.122: 0.030:0.125	V+2	Easy
RSAV-13 E3	3	78.1% : 12.4% : 8.8%	0.054:0.056:0.102	V+2	Intermediate
RSAV-21 E2 (concentrated)	3	60.5% : 39.5%	0.114	V+1	Intermediate
RSAV-21 E (non-concentrated)	3	62.2% : 37.8%	0.134	V+1	Intermediate
RSAV-12	3	59.6% : 40.4%	0.114	V+1	Complex
		40.6% : 47.6% : 11.8%	0.124:0.106:0.061	V+2	
RSAV-05 E1	>3	49.6% : 50.4%	0.124	V+1	Complex

Several of these cases were found to contain a different number of contributors than originally assumed. RSAV-07 (see p. 49) was determined to contain two donors but the example in the table was run using 3 donors, yielding a high SD. RSAV-12 was examined in depth and may appear to contain more than three donors. The table illustrates that high SD's may be obtained when running as a two or three person mixture. Case 13E3 had two minor donors that were fairly close in mixture weight to each other and needed to sample a wide area to resolve the components. Case 21 was determined to contain a trace third individual and is illustrated in the table as a mixture of two. Hence, larger standard deviations should alert the user to examine the data closely and conduct reruns as necessary.

In summary, manual evaluation requires a comprehensive assessment of the convergence graph and associated statistic along with the variance/SD to either accept the data or reject the particular run and consider a rerun, if necessary. Additional runs can then be assessed for reproducibility (see Rerun section—Section G).

D. Information Gain: Statistics-for cases with Suspects CODIS Upload Profiles-for No Suspect cases

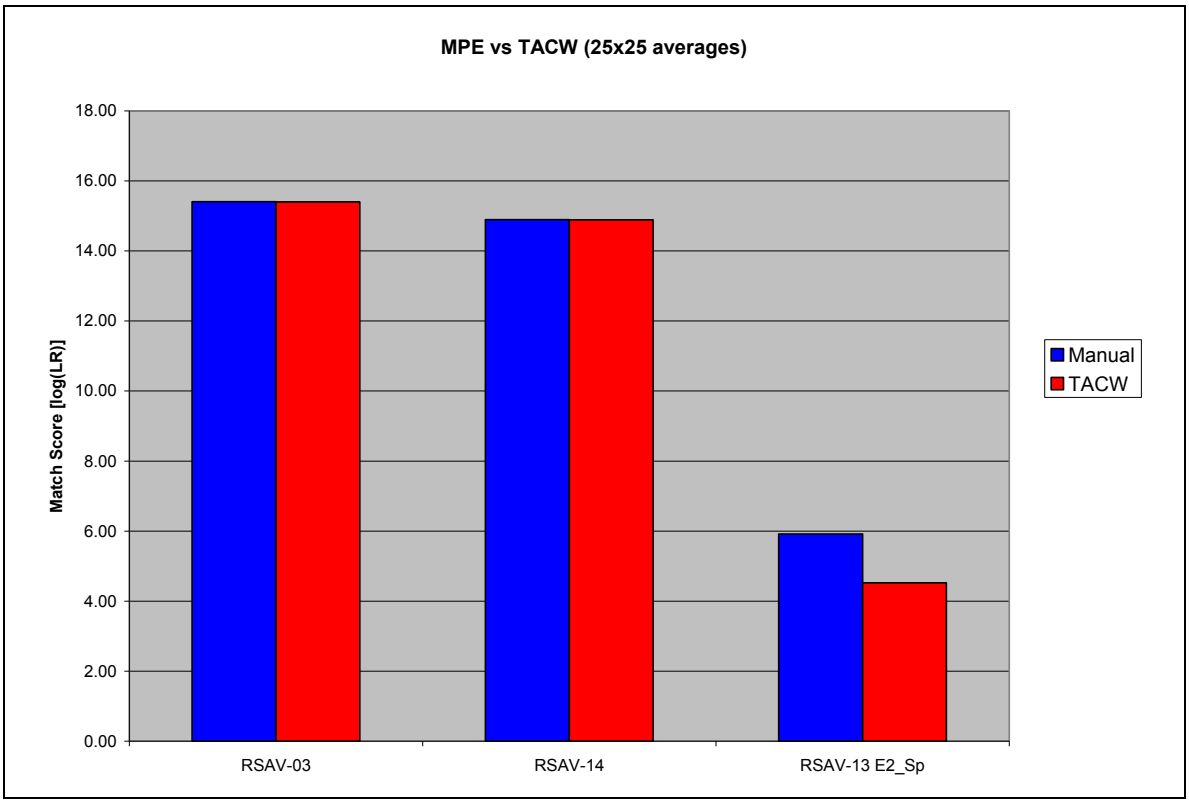
I. RSA Case Studies- Results from Cases with Suspects

a. Fourteen of the twenty five cases were submitted with a suspect. Eleven cases contained thirteen item comparisons with sufficient information for an inclusion and a statistical inference was generated using both manual and TAGC methods. The TAGC values were averaged from duplicate runs.

b. Comparisons of the information gain was subdivided into groups to reflect the three types of statistics performed manually:

i. MPE

Figure 17: Comparison of Manual MPE calculations to TrueAllele Casework Match Scores for RSA Cases



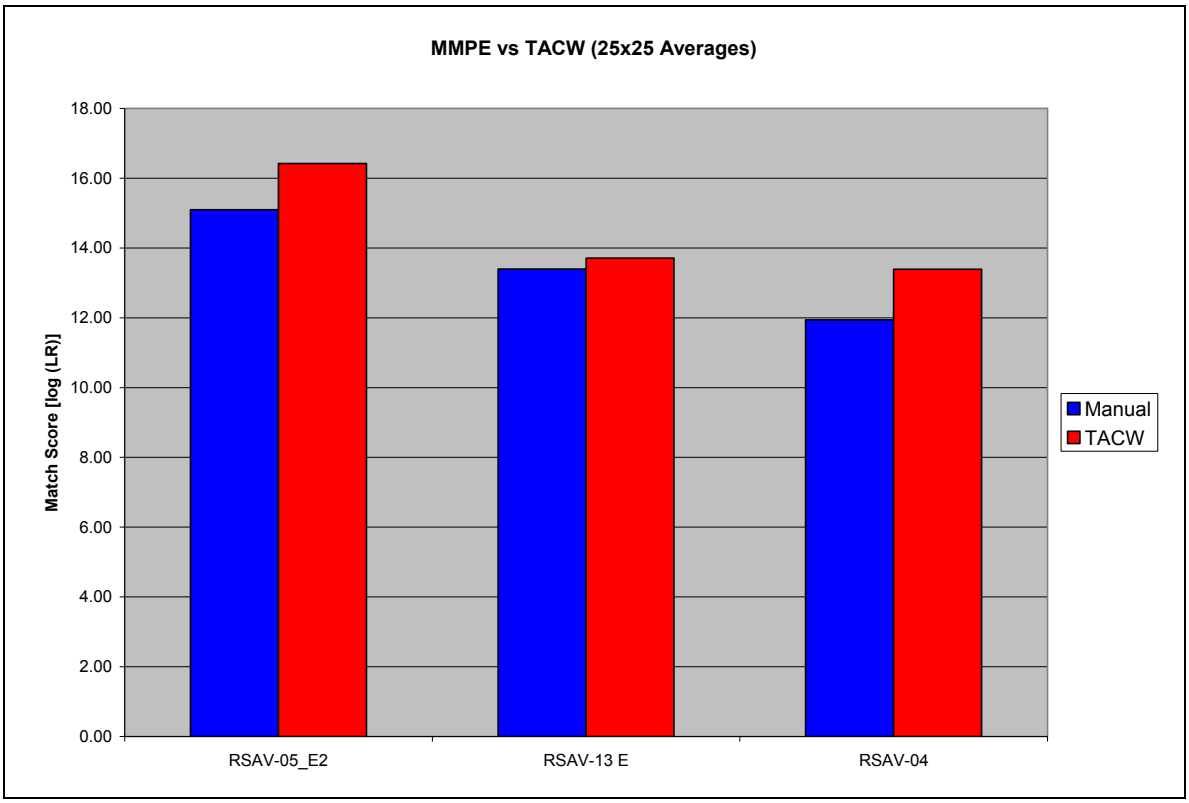
These bar graphs compare the match scores or information gained between the manual method (blue) and TAGC (red) for a single source sample (i.e. one genotype at each locus).

Cases RSAV-03 and RSAV-14: These cases contained full profile data for both contributors with no potential alleles below threshold or any drop out. Deconvolution using the victim reference resulted in a single source using both methods and hence no differences were observed. Although, the amount of time and potential for error performing hand calculations would support the use of the TAGC for these types of samples.

Case RSAV-13E2-sperm: This case was a single source partial profile at seven loci. The manual match score is higher than TACW by 1.4 log units. This difference occurred because the TAGC subtracted from the match score for genotypes with low probabilities or missing genotypes at other loci (pg 7- last paragraph iii in Case studies and Comparisons section). There were two loci, D13 and D21, exhibiting a very small second allele that may be due to sampling of sperm and under-representation of the second allele (see comprehensive/notes pp. 1).

ii. MMPE

Figure 18: Comparison of Manual MMPE Calculations to TrueAllele Casework Match Scores for RSA Cases



These bar graphs compare the match scores or information gained between the manual method (blue) and TAGC (red) for a deduced profile (i.e. one or more genotypes at each locus).

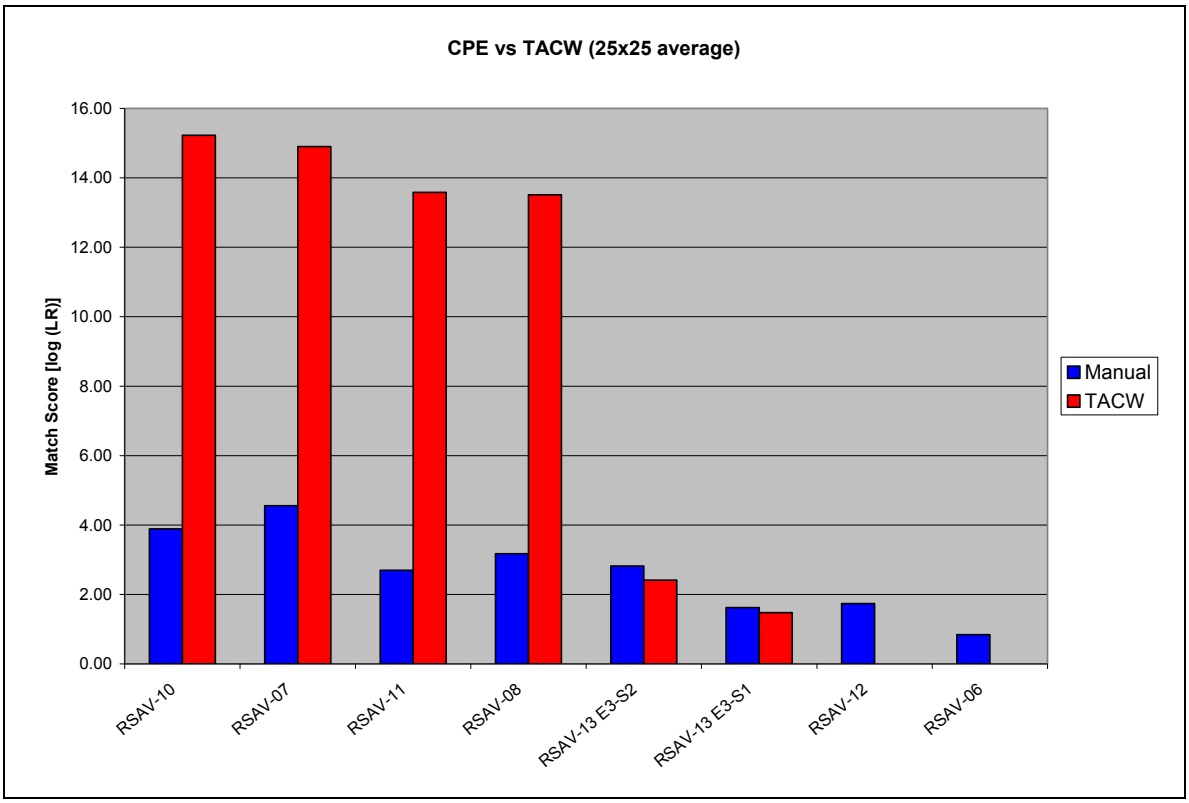
The manual method employed at MSP subtracts out the victim when the data supports a mixture of two individuals. Quantitative peak height data is utilized to deduce the contributor foreign to the victim by designating obligate alleles (p.p. 64, 65) and calculating sister allele balance and mixture weights. Manual methods are effective in narrowing the pool of potential genotypes but are laborious and not always consistent from analyst to analyst. These data illustrate the TAGC had a slight increase in match score and would be beneficial with regard to efficiency and consistency.

Case RSAV-05E2: The sperm fraction of the oral swab contained a minor profile (17.4%) that was deduced to a single genotype at 6 loci (D7, CSF, TH01, D13, D2 and TPOX) with the TAGC that the manual method could not isolate to one genotype. This accounts for the higher match score for TAGC relative to manual.

Cases RSAV-13E and RSAV-04: These cases contained suspect profiles that were deduced to single source genotypes with TAGC that were not completely single source with manual methods for some loci.

iii. CPE/CPI

Figure 19: Comparison of Manual CPE Calculations to TrueAllele Casework Match Scores for RSA Cases



These bar graphs compare the match scores or information gained between the manual method (blue) using all genotypes at a locus (i.e. generally more than one genotype at each locus unless it was a single allele/ homozygote) and TAGC (red).

There are significant differences observed when comparing the CPE/CPI manual calculations to the TACW match scores.

Cases RSAV-10 and RSAV-07: Differences in these cases can be attributed to the indication of a third contributor. Manual methods do not allow for an MMPE calculation if there are more than two donors. Case 10 contained potential alleles below threshold at two loci (FGA and D2), whereas Case 7 was not deduced due to an indication of a third contributor based on a peak that could not be differentiated from stutter (See comprehensive notes p.7)

Case RSAV-11: This case consisted of a bite mark that was categorized and run as a two person mixture containing a major suspect at about 85% (although there is an indication of a trace third individual, it would not have significantly impacted the profile of the major donor). **In manual methods, the CPE was calculated because the stat was done on the minor profile** (See comprehensive notes p. 7). The TAGC resulted in a better deconvolution and higher match score accordingly.

Case RSAV-08: This case contained a mixture of at least two individuals with an indication of a third contributor, where the two major contributors were close to 1:1 (minor at 44.5%), which deduced well with the TAGC. Manually, the profile was not deduced as MSP laboratory guidelines do not support this with a third individual present and therefore, manual calculations were deferred to a CPE statistic.

Case RSAV-13E3 swabbing of R and L face cheeks: This case was run as V + 2 and compared to two suspects. The data comprised mixture weights of 78.1%/12.4%/8.8% with the victim as the major contributor. Both suspect 1 and suspect 2 were compared to the profile deduced as the secondary contributor (12%) and the tertiary contributor (9%). This was a relatively complex case as a three person mixture with two minor components. The two minor components were fairly close in weight and had overlap of alleles amongst the three individuals (i.e. V, S1 and S2), after review of the known genotypes (refer to rerun section G for case run with one unknown by using each suspect separately to deconvolute the other to simulate a consensual donor). There were not many loci exhibiting distinct information to separate the two suspects. This may account for why both S1 and S2 were potentially included to the same secondary contributor to the evidentiary sample.

In both instances the TAGC had slightly lower match scores than the manual CPE as evidenced in Figure 19. The manual method included both S1 and S2 as potential contributors to the overall mixture. Two different CPE/CPI calculations were performed, one for each suspect. The final statistic yielded 2.82 for S2 and 1.62 for S1. This difference occurred because four loci were not used for the statistical analyses for S1.

Case RSAV-12: This case resulted in a low CPE statistic that manually encompassed seven loci that were considered to have all alleles present while there was no statistic with TACW. This case was run as V + 2 since it appeared to consist of 3 sources that cannot be totally segregated. Due to close mixture weights, the comparison to the suspects and subsequent statistics are negative values (-3.34 for contributor 2 and -3.66 for contributor 3 for longer V+2 run). The overall data and inferred genotypes may appear inclusive at many loci but reflect a low probability. Several loci have very low probabilities that result in the overall likelihood ratios not supporting any inclusions.

After investigating further, the electropherogram derived from this bra sample, given the known victim profile, may contain more than three donors. The ratio of the victim and suspect do not remain consistent from locus to locus and may explain why the deconvolution by the TAGC was not supportive of a match of the suspect to any of the unknown profiles (See pp. 63-64 for further discussion).

Case RSAV-06: This case presents a small contribution of a secondary source. There was only 1 allele called manually (D8) with four other loci containing a potential allele below threshold. A CPE was calculated manually resulting in a low match score (0.84 or 1 in 7). The TAGC resulted in negative scores in duplicate longer runs due to the uncertain data and did not support an inclusion with a statistic (see p. 58).

The comparisons of the CPE/CPI statistic relative to the TAGC using quantitative information and the victim to derive the unknown contributor are significant. This is expected since the CPE/CPI considers all genotypic combinations as potential contributors to a mixture, including even those that would not be included dependent on the data at the locus. For example, a two person mixture at a locus where no bands are shared and both contributors are in equal proportions and heterozygous would exhibit 4 alleles (all with similar peak heights). A CPE/CPI calculation would consider the four homozygous types in the calculation even though no such contributor genotype would be possible. The CPE/CPI calculation results in a much lower match score and loss of information. The match scores obtained using the TAGC are higher by approximately 10 log units, which equates to over ten billion times rarer or more likely than the probability obtained manually using the CPE/CPI.

II. RSA Case Studies- Results from Cases without Suspects

a. Eleven of the twenty five cases were submitted without a suspect. All eleven items were evaluated for subtraction of the victim to deduce a profile for CODIS Upload.

b. Initial comparisons of the information gain could not be made using match scores since no suspect profile was available to calculate a probability. A system was devised to evaluate the upload profiles based on the number of alleles and combination of genotypes at each locus. These criteria proved useful to create upload strategies using the most discriminating loci and removing loci that contained too many alleles or inferred genotypes.

c. CODIS upload profiles could be created based on SDIS (mandate of a minimum of 6 core loci) or NDIS (mandate of a minimum of 10 core loci) rules and to comply with the 4 x 4 rule.

d. CODIS upload system
   i. **Confidence Intervals:** The use of a confidence interval (CI) is useful to limit the number of allele pairs for upload to CODIS. Studies conducted evaluated the use of a 95% and 99% CI to eliminate the less probable allele pairs contained in the probability distribution to determine the allele set (see RSA Cases-No Suspects Binder). This strategy would focus on allele pairs with a higher likelihood of contributing to the sample and lead to a better search mechanism. The validation has investigated the application of a confidence interval and has determined that a 95% interval would be best for a preliminary search of the database. If CODIS uses a moderate stringency search a mismatch would be allowed at each locus. The TAGC could then be evaluated to check if the allele pair was present in the remaining 5% of inferred genotypes. The potential candidate match profile can then be compared to obtain a statistical inference. Hence, there may be a difference in the inferred genotype(s)

that would be searched from a larger full data set (e.g. 99.99%) that could be used for a match statistic.

ii. Case RSAV-16 with no suspect produced a mixture of approximately 83.5%:16.5% (longer) from a bite mark swab. The original runs conducted using the old settings of fast, long, and longer did not allow for a CODIS upload for all 15 loci. This study used the longer run upload information for comparative analyses to the manual method (as that was expected to yield the best data at the time).

iii. The data for both the manual and TACW methods are depicted below using the assumption of two donors (run as V + 1).

Manual Data:

Table 9: Results table from report for RSAV-16

Sample Description	**Case # Item #**	**D8S1179**		**D21S11**		**D7S820**		**CSF1PO**		**D3S1358**		**TH01**		**D13S317**		**D16S539**		**D2S1338**	
Bite mark swab	RSAV-16 1-1-07.1	8 12 **13 14**		**28 29** 31.2		**9** 10 **11**		**10** 12 **13**		**14 15** 16 18		**6 8** 9.3		**11 12**		**11** 12 **13**		**18** 20 **23** 25	
KBS Victim	RSAV-16 1-1-02.1.1	13	14	28	29	9	11	10	13	14	15	6	8	11	12	11	13	18	23

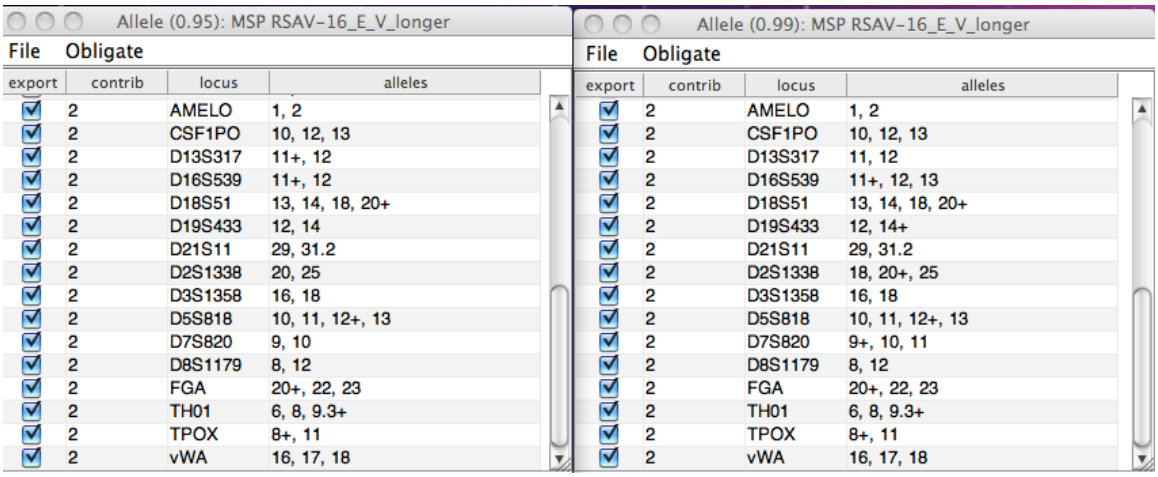
Sample Description	Case # Item #	**D19S433**		**vWA**		**TPOX**		**D18S51**		**AMEL**		**D5S818**		**FGA**	
Bite mark swab	RSAV-16 1-1-07.1	**12** 14 **14.2**		**17 18**		**8** 11 **12**		**13** 14 **18** 20		**X** Y		10 **11** 12 **13**		**19** 20 22 **23**	
KBS Victim	RSAV-16 1-1-02.1.1	12	14.2	17	18	8	12	13	18	X	X	11	13	19	23

Table 10: CODIS Upload Profile for RSAV-16

D8S1179	**D21S11**	**D7S820**	**CSF1PO**	**D3S1358**	**TH01**	**D13S317**	**D16S539**
8,12	**▼ 28,29,31.2**	**9,10**	**▼ 10,12,13**	**16,18**	**▼ 6,8,9.3**	**11,12**	**11,12**
D2S1338	**D19S433**	**vWA**	**TPOX**	**D18S51**	**AMEL**	**D5S818**	**FGA**
20,25	**▼ 12,14, 14.2**	**17,18**	**8,11**	**14,20**	**X,Y**	**10,12**	**20,22**

▼= obligate allele

Figure 20: TACW Data at 95% and 99 % for RSAV-16



Allele (0.95): MSP RSAV-16_E_V_longer

File Obligate

export	contrib	locus	alleles
☑	2	AMELO	1, 2
☑	2	CSF1PO	10, 12, 13
☑	2	D13S317	11+, 12
☑	2	D16S539	11+, 12
☑	2	D18S51	13, 14, 18, 20+
☑	2	D19S433	12, 14
☑	2	D21S11	29, 31.2
☑	2	D2S1338	20, 25
☑	2	D3S1358	16, 18
☑	2	D5S818	10, 11, 12+, 13
☑	2	D7S820	9, 10
☑	2	D8S1179	8, 12
☑	2	FGA	20+, 22, 23
☑	2	TH01	6, 8, 9.3+
☑	2	TPOX	8+, 11
☑	2	vWA	16, 17, 18

Allele (0.99): MSP RSAV-16_E_V_longer

File Obligate

export	contrib	locus	alleles
☑	2	AMELO	1, 2
☑	2	CSF1PO	10, 12, 13
☑	2	D13S317	11, 12
☑	2	D16S539	11+, 12, 13
☑	2	D18S51	13, 14, 18, 20+
☑	2	D19S433	12, 14+
☑	2	D21S11	29, 31.2
☑	2	D2S1338	18, 20+, 25
☑	2	D3S1358	16, 18
☑	2	D5S818	10, 11, 12+, 13
☑	2	D7S820	9+, 10, 11
☑	2	D8S1179	8, 12
☑	2	FGA	20+, 22, 23
☑	2	TH01	6, 8, 9.3+
☑	2	TPOX	8+, 11
☑	2	vWA	16, 17, 18

6 loci with ≥ 3 alleles

9 loci with ≥ 3 alleles

Table 11: Comparison of Manual and TACW Allele Tables for RSAV-16 longer

			Manual		TACW-95%					TACW-99%			
tem	Case	Marker	Allele		Allele 1	Allele 2	Allele 3	Allele 4		Allele 1	Allele 2	Allele 3	Allele 4
	RSAV-16	**D8S1179**	8,12		No Upload					No Upload			
	RSAV-16	**D21S11**	28,29,31.2+										
	RSAV-16	**D7S820**	9,10										
	RSAV-16	**CSF1PO**	10,12+,13										
	RSAV-16	**D3S1358**	16,18										
	RSAV-16	**TH01**	6,8,9.3+										
	RSAV-16	**D13S317**	11,12										
	RSAV-16	**D16S539**	11,12										
	RSAV-16	**D2S1338**	20,25										
	RSAV-16	**D19S433**	12,14+, 14.2										
	RSAV-16	**vWA**	17,18										
	RSAV-16	**TPOX**	8,11										
	RSAV-16	**D18S51**	14,20										
	RSAV-16	**AMEL**	X,Y										
	RSAV-16	**D5S818**	10,12										
	RSAV-16	**FGA**	20,22										

iv. There is no upload for case 16 with TACW for either the 95% or 99% CI data as both would fail the CODIS 4 x 4 rule (i.e. 6 and 9 loci containing 3 or more alleles, respectively). Hence, one must examine the data and perform an evaluation to consider a manual modification for a search strategy. See section B (p. 44) below for a re-evaluation of case 16.

v. **Manual Modification**: Data illustrate that many profiles may not be suitable for upload if all 15 loci are considered. Some loci have too many alleles and inferred genotypes that would not be allowed according to the

CODIS 4 x 4 rule. A manual modification can be made by assessing the loci and only using those loci that have acceptable data. Loci that have been derived to a single source genotype would be the most discriminating. Loci with more potential genotypic combinations could be sorted and rated to differentiate the most useful data. Those loci with too many alleles and inferred genotypes would not be included in the upload profile in an effort to create a useful profile for the database search. The modification for case 16 is illustrated below as an example of an intermediate level case under section B (p. 44).

v. **Evaluation Graphs**: Graphs of the number of alleles and number of genotypes were created to evaluate profile data for CODIS uploads. These graphs plot the data to compare and contrast run conditions to evaluate if a longer run time produces a more discriminating profile for CODIS upload. All data was plotted from the TAGC generated allele table and genotype table (see RSA Case Binders). These data plot the allele or genotype category from most to least discriminating on the x-axis and the number of loci on the y-axis. In assessing a profile, it is evident that the more loci with better discrimination would be best for searching. If most loci are not very discriminating, the profile may not be suitable for upload.

vi. The allele and genotype categories are defined in the table below rated from category 1 as most discriminating to the least discriminating:

Table 12: Allele and Genotype Categories.

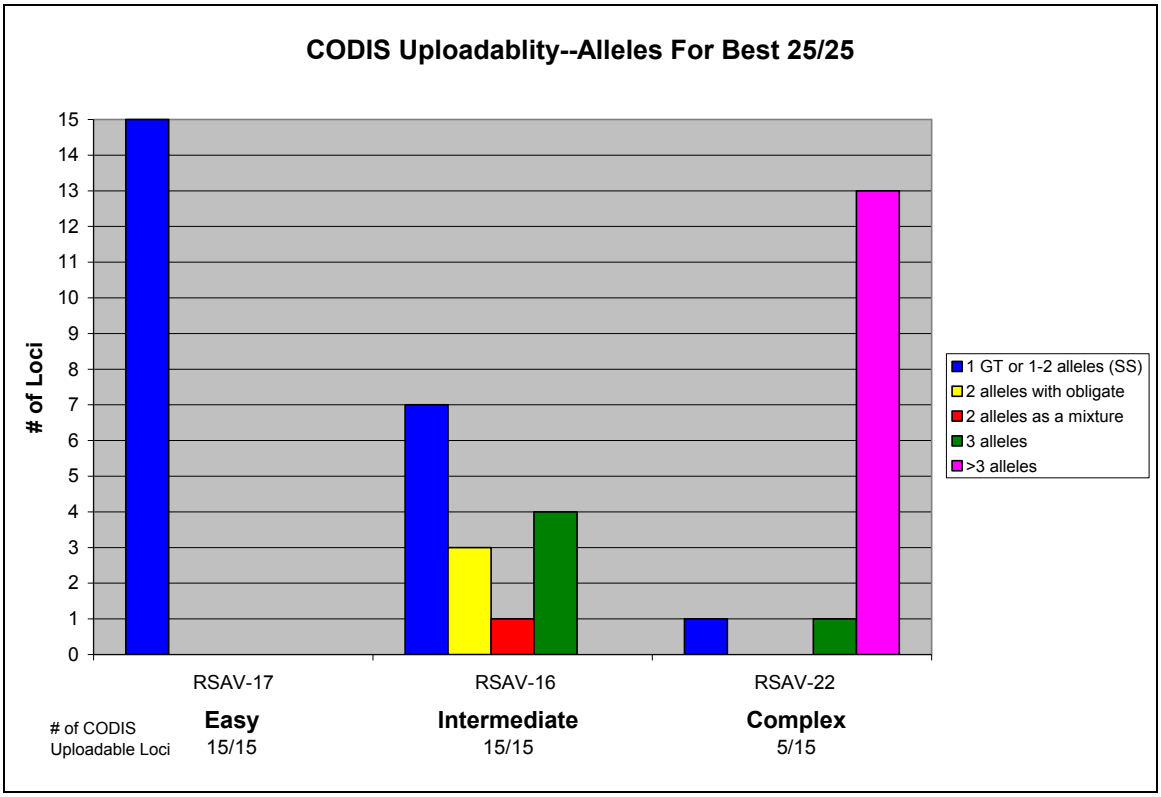
Category Number	Allele Categories
1 (blue)	1 or 2 alleles (1 GT)
2 (yellow)	2 alleles with an obligate
3 (red)	2 alleles as a mixture
4 (green)	3 alleles
5 (pink)	Greater than 3 alleles

Category Number	Genotype (GT) Categories – number of genotypes
1 (blue)	1 genotype
2 (red)	2 genotypes
3 (yellow)	3 genotypes
4 (green)	Greater than 3 genotypes

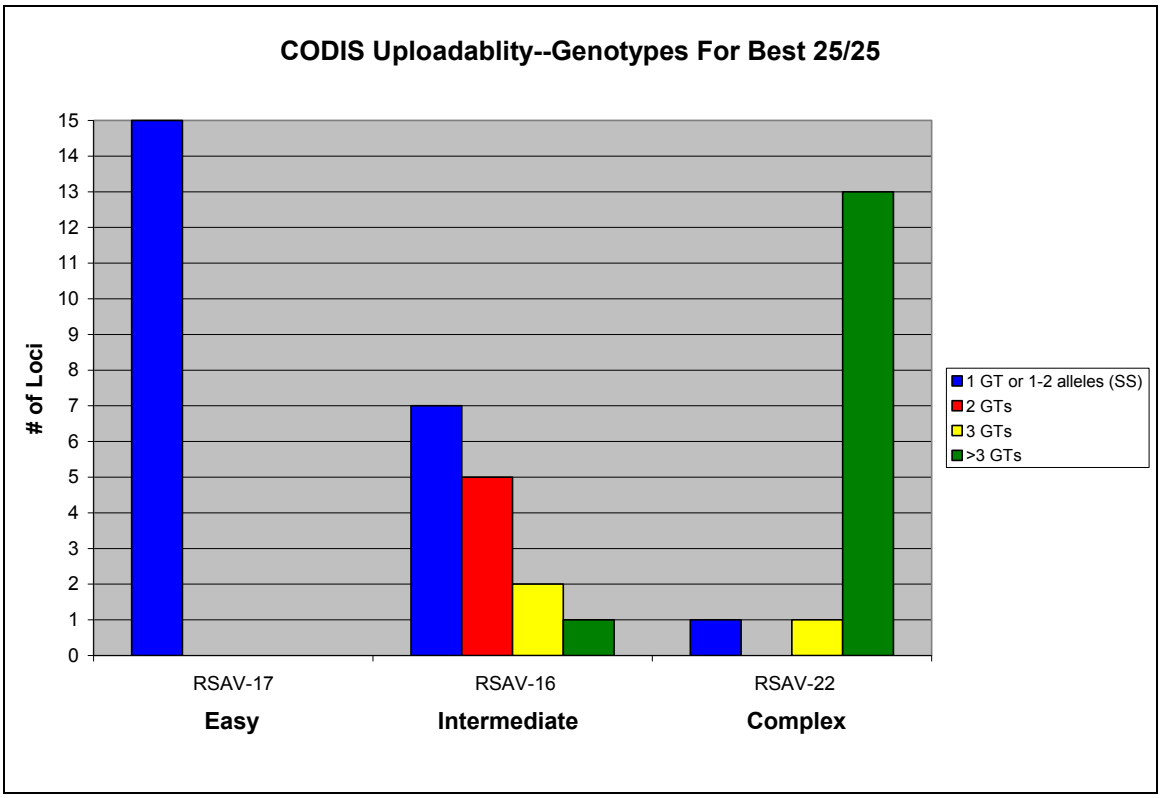
vii. The following graphs represent examples of cases for various levels of complexity.

Figure 21: Allele and Genotype Tables for Comparison in example cases for each of the Levels of Complexity

A.



B.



A. Easy: Case RSAV-17
   i. This data represents a mixture weight close to 1: 1 that was easily deconvoluted to a single source genotype after subtraction of the victim. Cases at the easy level produce good search profiles. The blue bars exhibit very discriminating profile information for a CODIS search as each locus is deduced to a single source profile (see Figure 6).

B. Intermediate: Assessment and modification of case 16 from above
   i. Case 16 contained a 16% minor that was unexpectedly somewhat difficult to deconvolute, as this would normally be a fairly simple two person mixture based on the weights of the components. After further investigation of the electropherogram, this mixture may contain DNA from greater than two sources (especially if another donor could be a family member's DNA, such as a child, that was on the body of the victim). This may explain why the data presented more inferred genotypes for certain loci. Only 6 loci could be deduced to single source genotypes with the longer setting. The allele graph below depicts the data for each locus at the regular and longer setting. Note that the 25/25 setting, run after changing the parameters (see phase II above), was the most discriminating and would allow for all 15 loci to be searched.

   ii. Loci may move from one category to another if run at a different setting dependent on the data presented (e.g. D18 and D5 moved to the greater than 3 allele category when run at longer). Both loci appeared to contain two additional minor alleles foreign to the victim, which consistently yielded the highest probability using both settings. The data examined at these loci exhibited more inferred genotypes at the longer setting. The additional genotypes in both instances contained heterozygote genotypes consisting of the larger minor allele and an allele masked by the victim. Although these additional genotypes had smaller probabilities, they preclude the use of these loci in the CODIS upload table below (Table 13) and result in a less discriminating profile for searching (13 of 15 loci). This may be due to the number of contributors and giving the TAGC more time to investigate the problem, reveals the data is not the best fit. It renders weaker data which may reflect an incorrect assumption of the number of donors and/or represent differences due to the variation and random sampling inherent in the particular run.

Figure 22: Allele Table for RSAV-16—Comparison of the Regular and Longer Settings

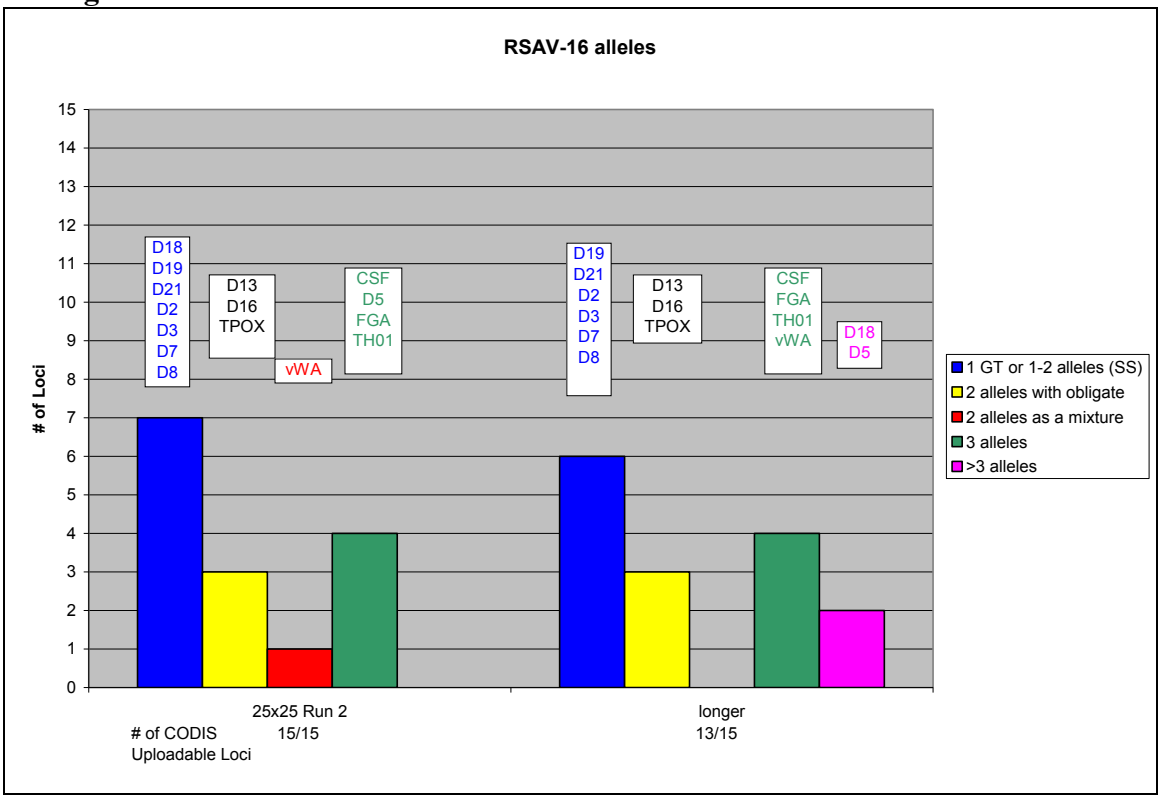


Table 13: Example of the modified longer CODIS Upload with 13/15 loci uploadable for RSAV-16

			Manual		TACW-95%			
Item	Case	Marker	Allele		Allele 1	Allele 2	Allele 3	Allele 4
	RSAV-16	**D8S1179**	8,12		8	12		
	RSAV-16	**D21S11**	28,29,31.2+		29	31.2		
	RSAV-16	**D7S820**	9,10		9	10		
	RSAV-16	**CSF1PO**	10,12+,13		10	12	13	
	RSAV-16	**D3S1358**	16,18		16	18		
	RSAV-16	**TH01**	6,8,9.3+		6	8	9.3+	
	RSAV-16	**D13S317**	11,12		11+	12		
	RSAV-16	**D16S539**	11,12		11+	12		
	RSAV-16	**D2S1338**	20,25		20	25		
	RSAV-16	**D19S433**	12,14+, 14.2		12	14		
	RSAV-16	**vWA**	17,18		16	17	18	
	RSAV-16	**TPOX**	8,11		8+	11		
	RSAV-16	**D18S51**	14,20					
	RSAV-16	**AMEL**	X,Y		X	Y		
	RSAV-16	**D5S818**	10,12					
	RSAV-16	**FGA**	20,22		20+	22	23	

iii. Intermediate level cases do not result in a single source genotype but provide acceptable data for a CODIS search.

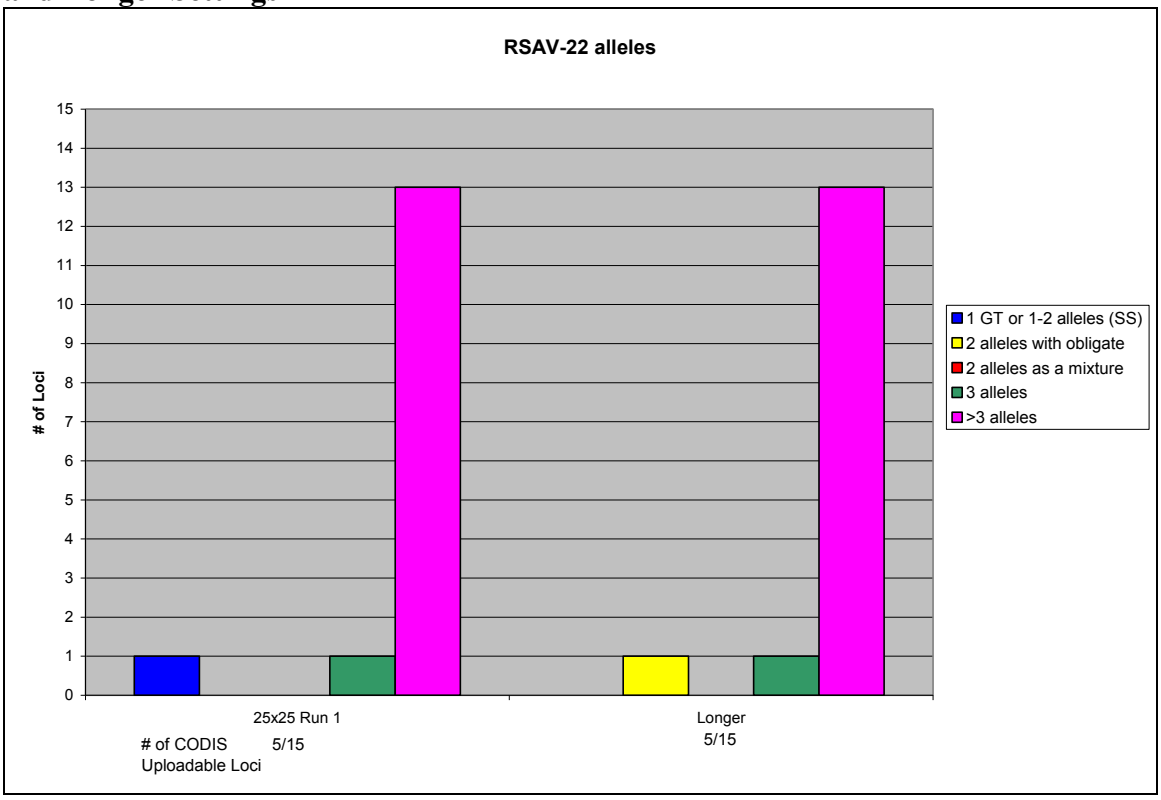
C. Complex--Case RSAV-22

i. This case contains a major of the victim and a minor(s) at low level. It is difficult to distinguish if it is a two person mixture (minor at 5.7%) or potentially a three person mixture (minors at 4.6% and 2.3%). This low level minor was not suitable for upload using TACW as only 5 loci were acceptable.

ii. The allele graph (Figure 23) reflects a large number of loci with greater than 3 alleles in pink. This would not provide enough information for either SDIS or NDIS searches.

iii. Manual methods did upload at all loci (9 loci plus amelogenin had potential alleles below threshold).

Figure 23: Allele Table for Complex Case RSAV-22—Comparison of the Regular and Longer Settings



iv. The allele graphs proved to be more useful for selecting which loci to use for CODIS uploads while the genotype graphs give an indication of the probability distribution and certainty or uncertainty of the answer.

v. Selection of Loci for Upload - The loci chosen for upload using these evaluation graphs would be identified from more discriminating (left) to less discriminating (right). Strategies may be incorporated to choose the number of loci dependent on upload to SDIS and/or NDIS. Selection of loci may encompass more than one upload profile. Different profiles may be entered simultaneously to be searched to yield the best potential matches. If there are several loci in the same allele/genotype category it may be best to use the loci deemed to be more discriminating.

vi. The final stage of the no suspect study investigated if a CODIS hit was made on any of the cases. Three cases resulted in hits (1 case to case and 2 to convicted offender) and the upload data for both manual and TACW methods has been compared to the suspect/questioned profiles. Cases RSAV-13, RSAV-17, RSAV-20 were relatively simple deconvolutions consisting of two person mixtures with suspect weights of 55.9%, 52.2% and 32.1%. All searches hit to 1 candidate with the original manual upload data. The differences in the discrimination of the upload information at all 15 loci between the two methods are summarized in the table below:

TABLE 14: RSA Cases with CODIS Hits to the Convicted Offender Database or Case to Case Hits

Hits	Manual	TAGC at 95%
RSAV-13 Case to case	13 - single source 2 - obligates	15 - single source
RSAV-17	12 - single source 3 -not searched	15 - single source
RSAV-20	10 - single source 4 – obligates 1-additional allele and obligate	14 - single source 1 – obligate

vii. The upload tables for case RSAV-20 in Table 15 illustrates the differences in the genotype data deduced manually and with TAGC. There are more potential genotypes for database searching using the manual method that could yield more adventitious candidates.

Table 15: DNA Profile Uploaded to CODIS for RSAV-20

			Manual		TACW-95%			
Item	Case	Marker	Allele		Allele 1	Allele 2	Allele 3	Allele 4
best 25x25	RSAV-20	D8S1179	11, 13		11	13		
	RSAV-20	D21S11	28, 32.2		28	32.2		
	RSAV-20	D7S820	10, 12		10	12		
	RSAV-20	CSF1PO	10, 12		10	12		
	RSAV-20	D3S1358	14 16+		14	16		
	RSAV-20	TH01	7, 8		7	8		
	RSAV-20	D13S317	12, 13		12	13		
	RSAV-20	D16S539	9, 12		9	12		
	RSAV-20	D2S1338	19+, 25		19	25		
	RSAV-20	D19S433	14, 15		14	15		
	RSAV-20	vWA	16+, 17		16+	17		
	RSAV-20	TPOX	8 11+		8	11		
	RSAV-20	D18S51	12, 18		12	18		
	RSAV-20	AMEL	X, Y		X	Y		
	RSAV-20	D5S818	10 11 12+		11	12		
	RSAV-20	FGA	21, 22		21	22		

In summary, the automated method resulted in a better deconvolution than manual deductions. The deduced profiles were concordant with the candidate profiles, illustrating the TAGC method to be accurate and reliable. The more discriminating data (i.e. single source genotypes) provided using the TACW system will prove to be more useful for CODIS searches, especially as database size increases (SDIS or NDIS).

E. Concordance-

I. Compare and contrast the conclusions and data obtained using both manual and TACW Methods.

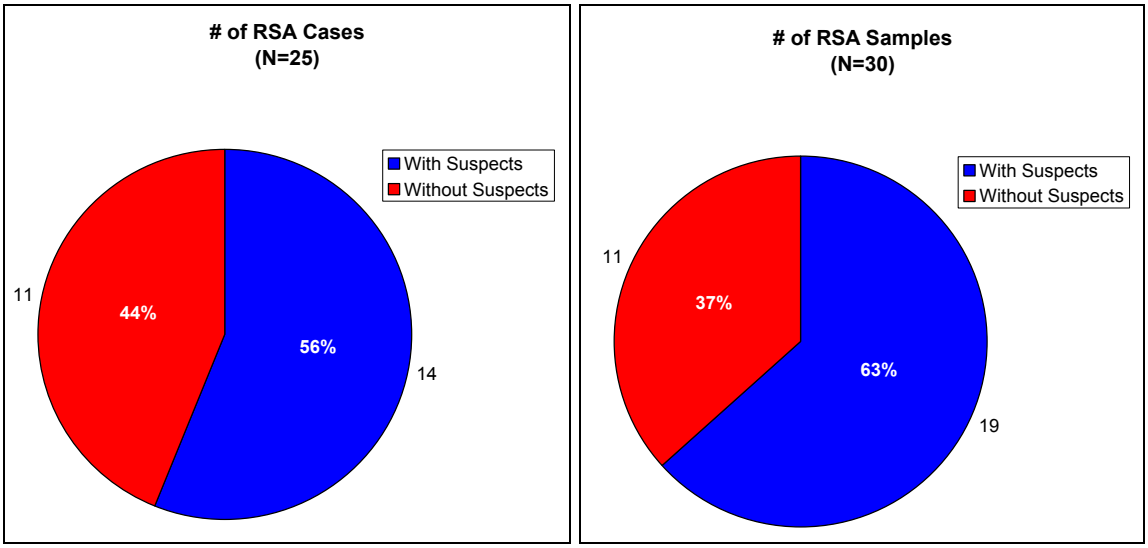
a. Metrics for Data Set

i. The final data set of 25 cases consisted of 14 suspect cases and 11 no suspect cases. A total of 30 items were assessed (Case RSAV-05 had an additional item and Case RSAV-13 had 4 additional items) with 19 and 11 items for suspect and no suspect cases respectively.

ii. All of the allelic designations were accurate, containing no miscalled alleles, for the entire data set throughout the study. There were a few instances whereby the manual method, using GMID, had an 'allele.X' designation while the TAGC, using the analyze module, made an allele call, without the .X.

Figure 24: Total Number of Cases and Samples.

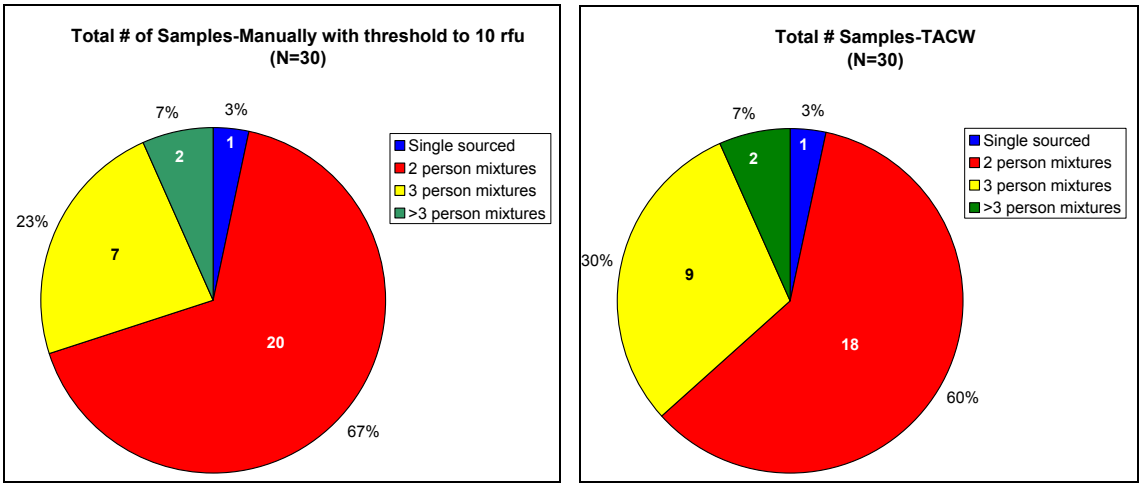
A. **B.**



A. The number of contributors for the data set is depicted in the graphs below (Figure 25), illustrating the majority of samples tested were two person mixtures. Data was evaluated using both manual methods (GMID and threshold lowered to 10 rfu) and automated methods (TAGC baseline is set at 10 rfu). The manual method was modified to investigate below threshold to more accurately reflect the number of donors to make more appropriate requests. Note that assessing the number of donors in GMID may not be recommended routinely as signal to noise issues become arduous and problematic.

B. A comparison of the two methods yielded similar results overall but does not reflect the case differences that existed. Cases RSAV-07 and RSAV-11 were run as two person mixtures with TACW and were three person mixtures manually. Both cases were explained in other sections as being due to a possible stutter peak for case 7 and an indication of a 3rd trace in case 11. A trace third was indicated in case 8 and three donors were apparent in case 13 after knowledge of the known suspects.

Figure 25: Sample Distribution by Number of Contributors



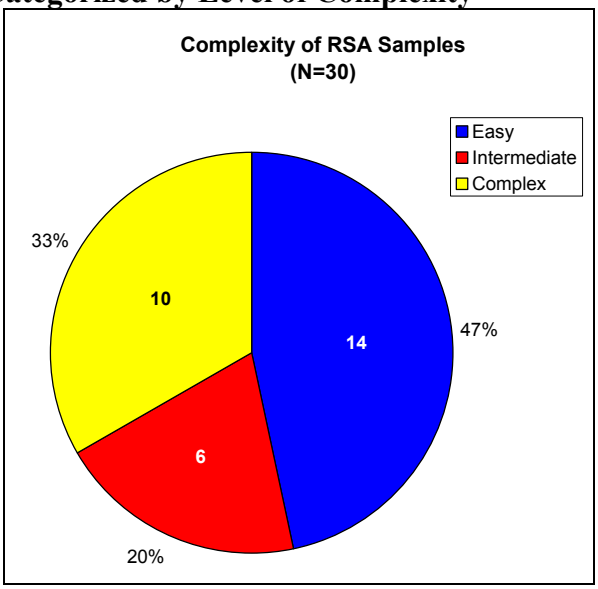
C. Mixture weights—the mixture weights obtained using the TACG for all cases are referenced in Table 7. There were no comparative analyses of mixture weights as manual methods do not specify weights of individual components. Manual methods may give a range of potential mixture weight for a minor contributor to a two person mixture when conducting an MMPE.

D. In assessing the number of contributors and associated mixture weights, all cases/items were categorized according to the level of complexity (see Figure 21) in Table 16 and Figure 26 below.

Table 16: Cases categorized by Level of Complexity

Level of Complexity	Suspect cases (RSAV #/Item)	Count	No Suspect Cases (RSAV#)	Count	Total Count
Easy	03, 04, 05E2, 07, 11, 13E, 13E2-Sp, 14	**8**	15, 17, 18, 19, 20, 25	**6**	**14**
Intermediate	08, 10, 13E3	**3**	16, 21, 24	**3**	**6**
Complex	01, 02, 05E1, 06, 09, 12, 13E2-Epi, 13E4	**8**	22, 23	**2**	**10**

Figure 26: RSA Cases Categorized by Level of Complexity



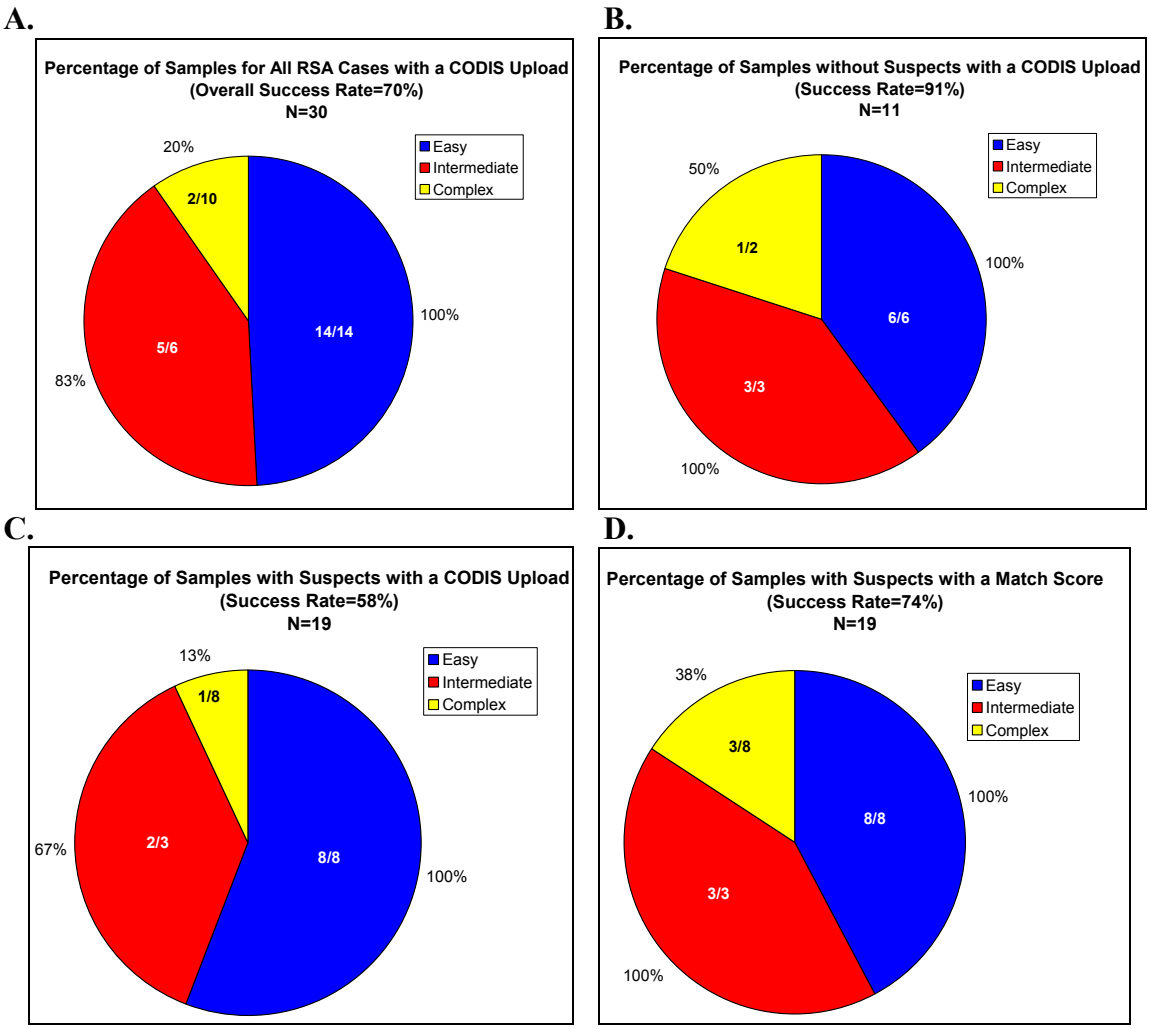
E. Success rates, using the TAGC, were obtained for cases in each of the three categories. As expected, success rates are higher for the easy cases and decline as the data becomes more complicated (Figure 27).

F. Easy—cases with two contributors and disparate mixture weights (without containing low target DNA) normally result in a full deconvolution to render a single source profile of the suspect. Cases in this category have the expectation of being solved (i.e. 100% success rate).

G. Intermediate—adding more complication to mixtures of two or more individuals with similar or low mixture weights results in lower successes and more uncertainty. Cases in this category may result in lower match scores or less definitive profiles for CODIS uploads.

H. Complex—the more complicated mixtures resulted in lower success rates. Some cases could not be solved, match scores could not be obtained or resultant profiles were not suitable for upload.

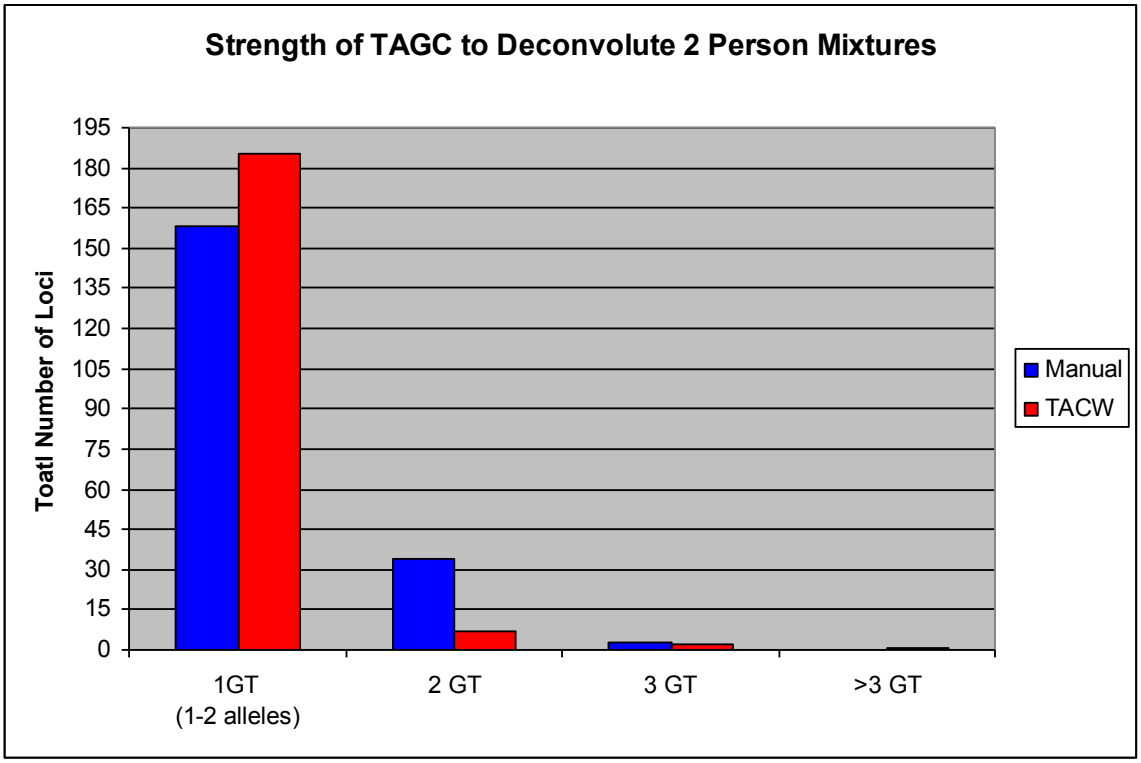
Figure 27: Success Rates with the TAGC (CODIS Uploads and Match Scores)



b. Strength of Deconvolution for Two Person Mixtures

An overall assessment of the manual and automated methods strength to deconvolute a two person mixture was conducted. Thirteen cases (both with and without suspects) contained both manual and 95% TAGC upload data for this study. The total number of loci compared for all items was 195. The figure below (Figure 28) clearly illustrates the TAGC was able to deduce to a single source genotype more effectively than manual means at 27 loci (185 vs. 158) or in approximately 14% of the data set (27/195). Therefore, there are less loci with two or more genotypes, which results in better discrimination and more useful deduced profiles for CODIS uploads. This deconvolution power is an enormous improvement and provides CODIS users with an excellent tool to triage potential associations to convicted offenders. Searches will result in less adventitious hits, thereby reducing the time and labor associated with further investigations into false candidate matches (false positives).

Figure 28: Strength of Deconvolution of Mixtures with two Contributors



c. Strength of Deconvolution for Three Person Mixtures with a Third Trace Component

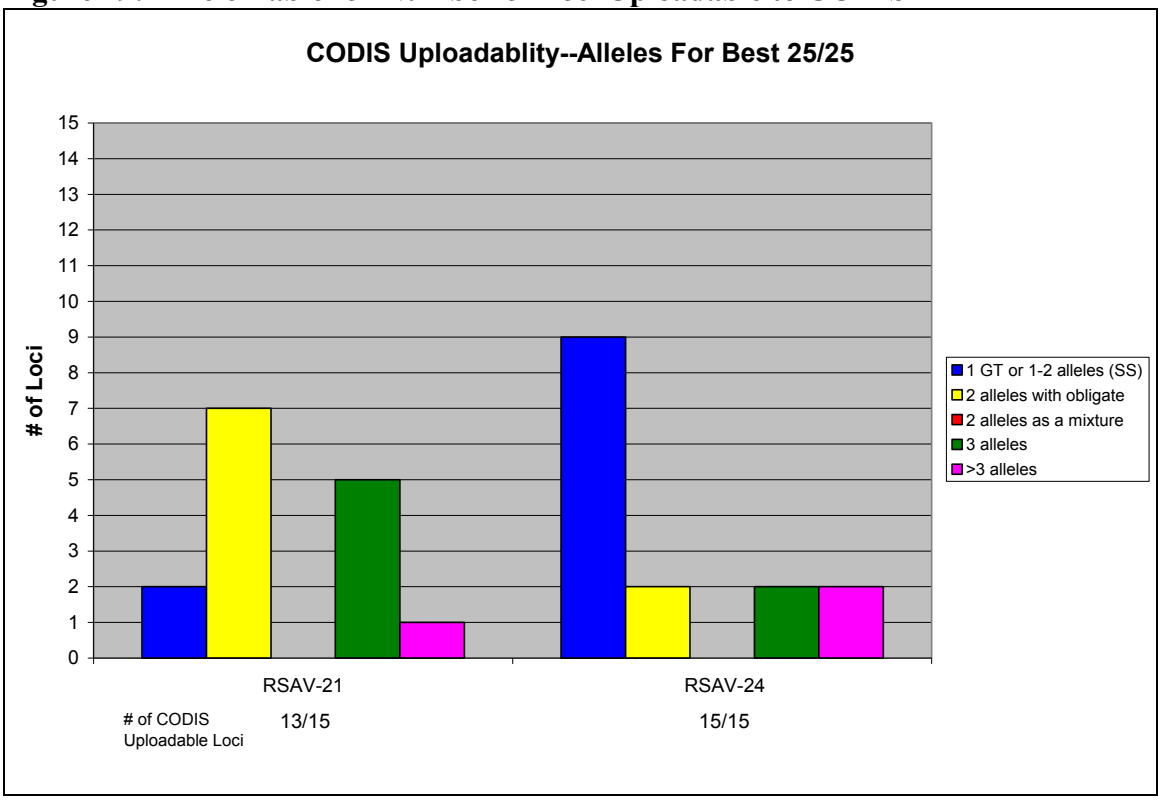
i. There were four cases consisting of 3 persons with a 3rd trace component as evidenced in the table below (Table 17). These cases had no manual upload for two reasons; 1) No MMPE is calculated for mixtures with more than two donors, 2) The data was assessed and would not pass the CODIS 4 x 4 rule. The table illustrates the number of loci containing too many alleles. This is an underestimate as the count is without consideration of the loci that contained potential alleles below threshold.

Table 17: RSA Cases with 3 Person Mixtures that have CODIS Uploads with TACW, but Not Manually.

RSA Case #	# Loci with ≥3 alleles Manually—Breaks CODIS 4 x 4 rule	Mixture weight of the trace 3rd contributor	TACW Match Scores Average 25/25
RSAV-08	7	1.2%	13.51
RSAV-10	13	2.5%	15.22 (S1) -19.80 (S2)
RSAV-21	8	4.6%	N/A
RSAV-24	8	2.1%	N/A

ii. Although the presence of a trace third contributor is inhibitory (per protocol) for manual means, it is not an issue when employing the TAGC. The TAGC deconvolutes the secondary contributor regardless of the presence of a trace third and is able to yield informative data for comparison to a suspect or for a CODIS upload. The match score after a comparison to a suspect is shown in the table for cases RSAV-08 and RSAV-10 (which was well over 10 orders of magnitude higher than the manual CPE as evidenced in Figure 19). The information derived for upload for the no suspect cases RSAV-21 and RSAV-24 is depicted in the figure below (Figure 29), which allowed for 13 and 15 loci respectively.

Figure 29: Allele Table for Number of Loci Uploadable to CODIS



iii. Case RSAV-22 was the only case that was uploaded manually that was not using the TAGC. This case may contain a trace third individual and when run as V + 2 was not suitable for upload as both the second and third contributor were low level (4.6% and 2.3%).

iv. The advantage of using the TAGC on this type of evidentiary sample has a significant impact on effectuating the use of the CODIS database and solving crimes that would otherwise remain unsolved.

II. Conclusionary Statements

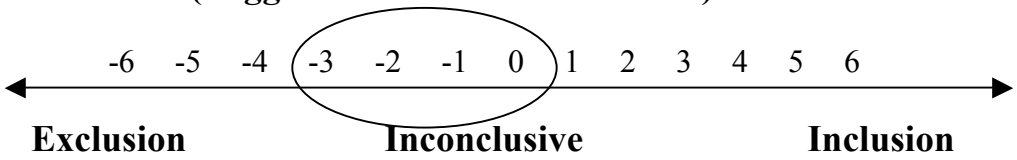
a. INCLUSION/EXCLUSION/INCONCLUSIVE----USE OF MATCH SCORE TO DETERMINE CONCLUSION

i. Introduction—The TACW system does not draw a conclusion as to whether a comparison of a known sample to a questioned sample results in an inclusion, exclusion or is deemed inconclusive. The TAGC models the data and provides a probability distribution of inferred genotypes. If a suspect is provided for comparison, a match score is obtained.

ii. A manual assessment of the match score and the number of loci with hits and misses must be conducted to determine the overall conclusion to be drawn. One must consider the level of complexity of the data and the results obtained to determine if the suspect is included, excluded or can neither be included nor excluded.

iii. Inclusions—reasonably high match scores (e.g. above 6 which equates to a million times more likely) are indicative of a match. Lower match scores (e.g. less than 2-3) may need to be investigated closely by the analyst to determine if an inclusion is supported. One may need to check each locus individually and investigate the genotypes and associated likelihood ratios. Certain loci may have data that does not fit well and may contain negative match scores (which lowers the overall likelihood ratio accordingly). If any loci contain misses, those loci need to be evaluated to determine if there is a scientific explanation for the genotype being absent. For example, a profile may exhibit degradation and/or inhibition that may cause allelic or locus drop out that may explain the negative match score for that particular locus. The comprehensive profile and associated match scores at each locus will assist in drawing a conclusion. A gray area appears to exist when the match score is 0 or where the probability of the suspect versus a random man is equal. Slightly negative or slightly positive data will need to be investigated carefully, using the data complexity and assessments to determine how to report the data.

Match Score: (Suggested Possible Conclusions)



iv. Exclusions-A comparison that contains many misses and negative match scores is indicative of exclusion. Case RSAV-04 was submitted with a victim, suspect and 15 other reference samples for comparison.

Both manual and TACW methods excluded all 15 other profiles as being contributors to the questioned evidence. The TAGC match score values for these exclusions ranged from -14.18 to -24.55 with an average of -21.19.

v. Inconclusive-Data in the gray zone (circled area on line graph) may be deemed inconclusive after a comprehensive evaluation of the data. Most cases in this study did not fall into this category. Comparison of cases that were complex and created differences and/or disconcordance between the manual and automated methods were evaluated closely and described in section b below.

vi. Additional cases have been identified for subsequent phases of the validation that will use profiles that were deemed inconclusive manually to evaluate the results obtained with the TAGC.

b. Complex Case Comparisons and Associated Differences/Disconcordance

The complex cases presented various challenges for interpretation including mixtures of two or more individuals, low level minor components and similar ratios of minor components that are difficult to deconvolute. These cases highlight some differences obtained between the manual and automated methods. Examples of each type of scenario are described below.

Table 18: Summary of Statistical Data for both methods.

Complex Cases								
Inconclusive vs. Inclusion No inclusion statistic with TACW			***Inconclusive vs. Inclusion*** No manual inclusion statistic			***Inconclusive vs. Exclusion*** No statistic with either manual or TACW		
RSA Case #	**Manual Match Score**	**TACW Match Score**	**RSA Case #**	**Manual Match Score**	**TACW Match Score**	**RSA Case #**	**Manual Match Score**	**TACW Match Score**
RSAV-06	0.84	-2.68‡	RSAV-02	N/A	0.31	RSAV-01	N/A	-5.73
RSAV-12	1.74	-5.77 (cont 2) -4.69 (cont 3)	RSAV-13 E2-Epi	N/A 2.63 (S1)*	2.07 (S1, cont 2) 1.33 (S1, cont 3) -1.91 (S2, cont 2) -1.22 (S2, cont 3)	RSAV-05 E1	N/A	-10.04
			RSAV-13 E4	N/A 0.44 (S1)*	2.66 (S1, cont 2)^ -4.32 (S1, cont 3)^ -4.82 (S2, cont 2)^ -5.95 (S2, cont 3)^	RSAV-09	N/A	-6.60 (S1)^ -15.03 (S2)^

KEY: ‡ Average Match score of Longer setting

* No Manual statistics were originally calculated in the case file (N/A); however for this validation the manual match score was calculated.

^ From best 25/25. Only one run passed.

S1=Suspect 1, S2=Suspect 2, cont=contributor

These cases present more difficult data that may exhibit drop out of alleles/loci. Manual methods often consider drop out, especially at larger loci, and do not include such loci in the statistical analyses. Although, an overall inclusion may still be reported for the remaining loci for which a statistic was allowable using the CPE/CPI (e.g. all alleles are accounted for with no potential alleles below threshold). Note that this phase of the validation did not assess the use of the current versions degradation function.

The TAGC calculates likelihood ratios for all loci. Some loci may have very small probabilities resulting in negative LR's, especially with the use of a 3% theta value. The overall negative likelihood may result in a random man being more likely to contribute to the evidentiary sample than the suspect in question. Note: Future versions of TAGC will allow for the removal of loci that do not contain information that have been assessed apriori (i.e. for loci determined to exhibit drop out that were assessed examining the evidentiary stain prior to comparison with the suspect).

This validation seeks to evaluate the information derived from the TAGC to make a determination of inclusion, exclusion or inconclusive for current reporting methods. Some data contain the suspect genotype included at every locus in the probability distribution but renders an overall negative likelihood match score. This will be referred to as an inclusion with no supporting statistic (will this be inconclusive or use a separate 'policy' type of statement—get from Bruce H.--).
Email of verbiage from Bruce: Discuss with Amy
In such cases, where the analyst has good reason to believe the probative reference standard is included, but we don't have a statistical model to support the inclusionary statement, we may use:
"DNA from more than one (two, three) individual(s) was obtained from the [list items (#s)]. The DNA profile obtained from this item does not satisfy the Laboratory's inclusionary reporting criteria. No further conclusions can be made regarding this item."

The statistical analyses will be evaluated at theta values of zero, 1% and 3% as part of the validation study.

i. Inclusion Versus Inconclusive: Supported Statistic With Manual But Not With TAGC.

Case RSAV-06: The duplicate 25/25 runs had acceptable convergence statistics but low SD's (i.e. 0.009) that did not appear to sample well. Duplicate runs at the longer setting (50/50?) were used to evaluate this minor component (average of 3.6%/ approximately 29-79 rfus). Both runs contained the suspect genotype at every locus but resulted in negative LR's at 8 and 5 loci for longer run 1 and 2, respectively. The overall average LR obtained was negative at approximately – 2.68 and 0.12, using a theta value of 3% and 1% respectively. If no theta was applied, a low positive match score of 2.12 was obtained. The suspect cannot be excluded from the inferred genotypes but is not accompanied by a supportive LR. ?? INCLUSION W/O STAT or Inconclusive ??

The manual match score, calculated using two loci, was 0.84 which equates to approximately 1 in 7.

Was Y STR considered on this case??

Case RSAV-12: Duplicate runs were conducted at regular and longer settings. One run from each setting was used for evaluation. The data was run as V + 2 and hence evaluated as being a three contributor mixture (secondary at approximately 37% and tertiary at approximately 13%). Although it appears there may be more donors (see pp. 79-80). Two loci, vWA and FGA, contained 6 and 5 alleles respectively and the model did not fit the data well when the genotypes of the victim, suspect and other were reviewed in the explain window.

Comparison of the suspect to contributor two resulted in exclusion, as the TAGC yielded no match data. The average overall match score was negative at -3.8.

Comparison to contributor three contained the suspect genotype at every locus but resulted in approximately 6 loci yielding negative LR's. The overall average LR obtained was negative at approximately – 3.48 for theta at 3%. The values remained negative for theta at 1% and without theta (-1.79 and -0.72, respectively). Given the number of contributors to the mixture is most likely greater than three and that the match score is negative without theta, this comparison results in an inconclusive finding.

The manual match score, using 7 loci, was 1 in 379 (1.74 match score)

Was Y STR considered on this case??

ii. Inconclusive Versus Inclusion – Supported Statistic With TAGC But Not With Manual.

Case RSAV-02: This case was considered a two person mixture run at V + 1. Although, after comparison with the suspect, there were several alleles (D21 32.2 at 75 rfu and vWA 14 at 50) that did not originate from either the victim nor suspect and the mixture may contain a trace third component. There were duplicate regular runs and a longer run. The longer run will be used to evaluate the data. The minor component was at 4.5% (rfus approximately 60-170) and all suspect genotypes were in the probability distribution for all loci. Three of the loci resulted in negative LR's with an overall LR of 1.75. This LR with a theta of 3% equates to a weak inclusion with a match score less than 1 (0.31).

The manual method reported this data as inconclusive. This may have been due to analyst discretion and was a conservative interpretation, possibly due to the number of contributors and assessment of drop out. When reviewed for the validation it appears that a CPE/CPI statistic may have been allowable at several loci up to potentially six loci.

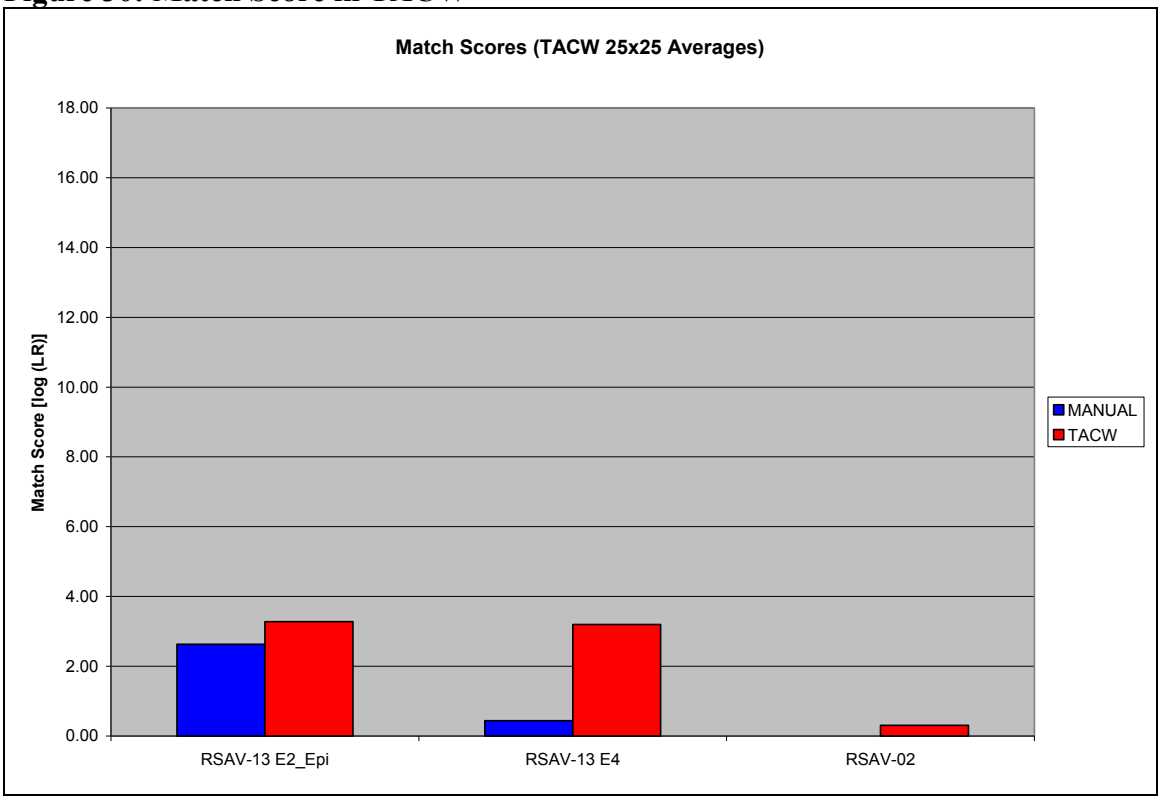
A. Supplemental information from case 13 for comparative analysis

Case RSAV-13: This case contained two suspects and two items, 13E2-epi and 13E4, that had duplicate regular runs and one longer run performed for evaluation. All 3 runs were assessed for item 13E2-epi while one 25/25 was not used for item 13E4. Both items were run as V + 2 and resulted in the victim as the major donor (approximately 85%). The other two minor components were compared to both suspects. Neither item had a manual statistic performed during the case testing in the operations laboratory (note that a different item was reported with a statistic for case 13).

Retrospectively, a manual statistic was obtained as part of the validation for matches to S1. These items had a CPE/CPI calculated manually and would have been included in the section regarding information gain (see Figure 19) to compare the manual and automated statistics (see graph below, Figure 30).

Manual and TAGC were concordant, including suspect 1, for both items to the minor contributors (the minor contributors ranged from approximately 6-8% each).

Figure 30: Match Score in TACW



These items are described herein with regard to comparisons to suspect 2 to illustrate how the theta value affects complex data and may be used to assist interpretations. Comparison of suspect 2 was considered inconclusive for both items with the manual method. The TAGC contained all suspect 2 genotypes for all loci to both contributor 2 and 3 for item 13E2-epi in all three runs. Negative LR's were obtained for each run for both contributors, ranging from two to six loci (see Table 19 below). The match scores were assessed for all comparisons without theta and for theta at 1% and 3%.

Table 19: Effect of Theta on Match Scores.

Item 13E2-epi	Run 1 (25/25)	Run 2 (25/25)	Run 3 (50/50)
S2 comparison to **Contributor 2**			
Theta = 0	2.69	1.99	2.06
Theta = 1%	1.29	0.54	0.61
Theta = 3%	- 1.44	- 2.38	- 1.67
Negative Loci	2	3	4
S2 comparison to **Contributor 3**			
Theta = 0	2.04	0.96	1.81
Theta = 1%	0.63	- 0.31	0.43
Theta = 3%	- 0.80	- 1.64	- 1.69
Negative Loci	3	6	3

It is evident from Table 19 that match scores decrease as the theta value is increased. The match scores were negative for all three runs to both contributors using a theta value of 3%. If no theta value was applied the match scores consistently support a weak inclusion.

Evaluation of item 13E4 was similar to 13E2-epi yielding somewhat of a weaker inclusion statistic (average of 1.24) for S2 to contributor 3 without the application of theta.

Given all genotypes of the suspect are included in the probability distribution for both contributors, it would not seem appropriate to exclude. Although, dependent on the theta value, there may not be a supportive statistic to accompany the inclusion.
?? INCLUSION W/O STAT or Inconclusive ??

iii. Inconclusive Versus Exclusion

Case RSAV-01: Data derived from a penile swab (note that the probative evidence was victim female DNA and nomenclature is vice-versa to maintain consistency with all other case scenarios) was assessed for duplicate regular runs and one longer run using a V + 1 request. Although, after comparison to the suspect, there was one D16 10 allele that did not belong to the victim nor suspect and may be indicative of a trace third individual, contamination or a mutation. The case remained categorized with the two person mixtures as the potential for a third component was not addressed.

The minor component was averaged to be approximately 3.1 % with peak heights ranging from 39 to 82 rfu. In reviewing the manual data for the minor profile, stochastic effects were observed. The blue loci (e.g. D8), which are usually the most sensitive, did not exhibit any minor contribution unless in stutter positions (D21 and CSF). Yet, other less sensitive loci (e.g. TH01) had called alleles. The manual conclusion was inconclusive due to insufficient DNA for comparative analyses.

The TAGC had no data in the Match window and was assessed to determine if there were any loci that contained misses. The 99% distribution had two loci, D8 and D18, which contained the alleles of the suspect but did not have them as heterozygous together matching the suspect genotype. The 100% distribution had the genotype, albeit at a very low probability, at D8 but did not for D18. In review of the LR's for each locus, only 3 were slightly positive. The overall match scores for all three runs at each of the three theta settings (0, 1% and 3%) were negative ranging from – 0.62 to – 6.76. The average match score for the 3% setting (used for the validation study) was – 5.55. Review of all data using the TAGC led to the determination of exclusion.

Case RSAV-05 E1: This case presented as a complex mixture with many potential alleles below threshold that may consist of at least four individuals (7 alleles at a locus). It was originally run as a two person mixture with TACW but was not analyzed further as it was considered unsuitable for this study. It was considered inconclusive manually.

Case RSAV-09: One of the duplicate regular runs and one longer run at V + 1 were used for evaluation. Note there was no indication of greater than 2 contributors prior to comparison to both reference samples. Retrospectively, there were alleles in the questioned stain that did not originate from either suspect. No additional requests were made.

The minor component of approximately 6% contained peak heights ranging from 24rfu (Amel Y) to 134 rfu. This item appeared to contain a low level of minor represented at the blue and green loci but rendered no additional information from the victim genotype and associated stutter peaks in either yellow or red loci with the exception of the Y at amelogenin. Comparison of the suspects to the minor component yielded no data. S1 had 6 misses and 6 positive LR's whilst S2 had 3 misses and only 2 positive LR's. Match scores for both runs at all three theta values rendered negative values ranging from – 4.50 to – 15.03. This data clearly supports exclusion of both suspects to the minor donor.

Manual methods determined the mixture contained insufficient DNA for comparison and was reported as inconclusive.

The table below summarizes the findings for these eight complex suspect cases. Case 12 is the only case that was considered inconclusive with TAGC and inclusive manually. Case 6 was inclusive with both methods but TAGC was without a supportive statistic. On the other hand, five cases reported as inconclusive manually that resulted in one weak inclusion, two inclusions without a supportive statistic and two exclusions. Lastly, one

case was too complex for interpretation. Therefore, the TAGC will allow users to further investigate what is currently reported as inconclusive manually to give a better indication/conclusionary statement of inclusion or exclusion.

Table 20: Conclusions of Inconclusive Data.

Summary of Conclusions								
Inconclusive vs. Inclusion No inclusion statistic with TACW			***Inconclusive vs. Inclusion*** No manual inclusion statistic			***Inconclusive vs. Exclusion*** No statistic with either manual or TACW		
RSA Case #	**Manual Conclusion**	**TACW Conclusion**	**RSA Case #**	**Manual Conclusion**	**TACW Conclusion**	**RSA Case #**	**Manual Conclusion**	**TACW Conclusion**
RSAV-06	Inclusion	IWS	RSAV-02	INC	Weak inclusion	RSAV-01	INC	Exclusion
RSAV-12	Inclusion	INC	RSAV-13 E2-Epi	Inclusion (S1) INC (S2)	Inclusion (S1) IWS (S2)	RSAV-05 E1	INC	ND
			RSAV-13 E4	Inclusion (S1) INC (S2)	Inclusion (S1) IWS (S2)	RSAV-09	INC	Exclusion

KEY: IWS=Inclusion without a Stat
INC=Inconclusive
ND=Not Determined
S1=Suspect 1
S2=Suspect 2

Summary:

i. Caution should be exercised when data presents at approximately 5% or less. The minors in the cases discussed in this section ranged from about 3-6% with rfu values from 100 to 200. It is expected that lower minor components will contain higher levels of uncertainty and more inferred genotypes with lower probabilities. It may be advisable to check the peak heights of minor components to correlate the target level of DNA and recognize the stochastic effects associated with such samples. All information reviewed comprehensively should assist the user in making conclusions for reporting purposes.

ii. The more complex data illustrates some differences between the manual and automated methods. The manual methods are subjective and are dependent on stochastic and analytical thresholds, interpretation of various injections and the use of analyst discretion often based on experience and qualifications. Manual methods do not attempt deconvolution for mixtures of greater than two individuals as the laboratory uses the CPE/CPI method for complex mixtures. Comparisons are made of the suspect(s) to the mixture without any assumption of the number of donors. This accounts for the differences observed for RSAV-12 whereby the suspect was included in the mixture without accounting for the number of donors but was rendered inconclusive if there were an assumption of only three donors with TAGC. It is uncertain if the suspect would be included with TAGC if the mixture consists of four individuals (this scenario has not been

addressed at this phase of the validation). The question being asked and the comparisons made are different for the manual versus the TAGC methods. Manual is asking if the suspect is a contributor to the overall alleles/genotypes whereas TAGC is deconvoluting the mixture into its separate components (assumption of the number of donors as dictated by the request) and comparing the suspect(s) to each of the individual components with associated inferred genotypes and likelihood ratios. Hence, one may expect to see some differences in interpretation when asking different questions, making different assumptions and comparing different profile data.

iii. The TAGC offers an objective method of providing inferred genotypes of the evidentiary sample without any knowledge of the suspect genotypes. Once a comparison is made to the suspect genotypes, a probability distribution can be reviewed to determine if the suspect is included at every locus. The probability will give an indication of how likely that genotype was to contribute to the evidence prior to knowledge of the suspect. A likelihood ratio is provided as a quantitative assessment to assist the user in making a qualitative statement for reporting purposes.

III. Difference in Designation of Obligate Alleles:

Case RSAV-05 item E2: This item was used to illustrate several comparisons of manual to automated methods at various loci (see comprehensive—MMPE Cases section p.2). One such comparison for the designation of obligate alleles is explained below.

OBLIGATE ALLELE (denoted with a +): The designation of an obligate allele differs between the manual and TACW methods. Manually, the obligate allele is defined as the allele that is foreign to the victim or obligatory to that of the unknown contributor, and is assigned regardless of quantitative peak height information. The TACW system designates an obligate allele based on the common allele present in the inferred genotypes which is based on peak height information. The oral swab data for the TH01 locus illustrates this difference between the CODIS upload request performed manually versus TACW system in Table 21 below.

Table 21: Differences in Obligate Allele Designation for Manual Methods and TACW.

TH01 (25x25)	Manually	TACW
Alleles evidenced in oral swab	6 (515 rfu), 7 (3232 rfu), 8 (2524 rfu)	6 (468 rfu), 7 (2984 rfu), 8 (2339 rfu)
Victim Profile of 12, 12	Used to subtract out victim and assess foreign allele(s) of the unknown contributor.	Used to perform modeling and calculate mixture weights with victim type to infer genotypes of the unknown contributor.
Obligate Allele	6	7
CODIS Upload	6+, 7, 8	6, 7+, 8
Inferred Potential Genotypes	6, 7 6, 8	6, 7 7, 8
Suspect Genotype	6, 7	6, 7

Although the manual and TAGC designate obligate alleles differently, both methods included the suspect's genotype in this example. In this instance, the TAGC did not have the 6, 8 as a possibility in the inferred genotypes and would therefore not search for a candidate with a 6, 8 genotype at TH01. Likewise, the manual method does not have a 7, 8 as a possibility and would therefore not search for such a candidate. The suspect's genotype is a 6, 7 which would be included both manually and with TACW.

All 25 cases were examined for obligate allele differences and only this case and case 23 had such occurrences. Case RSAV-23 had three loci containing obligate alleles assigned differently than manual means. This case was considered to contain four donors but was run as two donors and would, therefore, have varying peak heights that may account for the disparities. The details on these occurrences are contained in Appendix I.

Careful examination of all the affected loci determined that the inherent differences would not have caused any missed CODIS hits using one method over the other. Many situations resulted in the suspect genotype being allele 1, allele 2 whereas one method denoted the obligate on allele 1 and the other on allele 2. Again, the methods are different and hence different strategies may be employed but the overall results are similar. Additionally, if there are more inferred genotypes derived using the TAGC, without an allele in common, there will be no assignment of an obligate allele and the search would not miss the candidate.

IV. Difference in Mixture Weight Determinations:

Case RSAV-05 item E2: This item was also used for mixture weight comparisons of manual to automated methods at various loci.

Mixture weights can vary significantly when calculating them at each locus independently, using four allele loci and amelogenin. Manual calculations for the minor

male profile ranged from 12 % to 28% and make it difficult to eliminate potential genotypes.

There were five loci used to calculate MW manually. The results compared to the MW obtained using the TAGC are in the following Table:

Table 22: Mixture Weight Comparison of the Oral Swab

Mixture Weight of Minor Male (%)		
Locus	**Manually**	**TAGC**
D21	13	17.4
D3	14	
D18	28	
Amelogenin	12	
FGA	21	
Average	17.6	
Standard Deviation	6.8	1.5
95% CI/ 2 SD Range	3.99 to 31.2	14.4 to 20.4
Range Difference	27.21	6.0
There is a 4.5 fold difference in the range established manually compared to that of the TAGC.		

F. Precision

I. Titration Sets

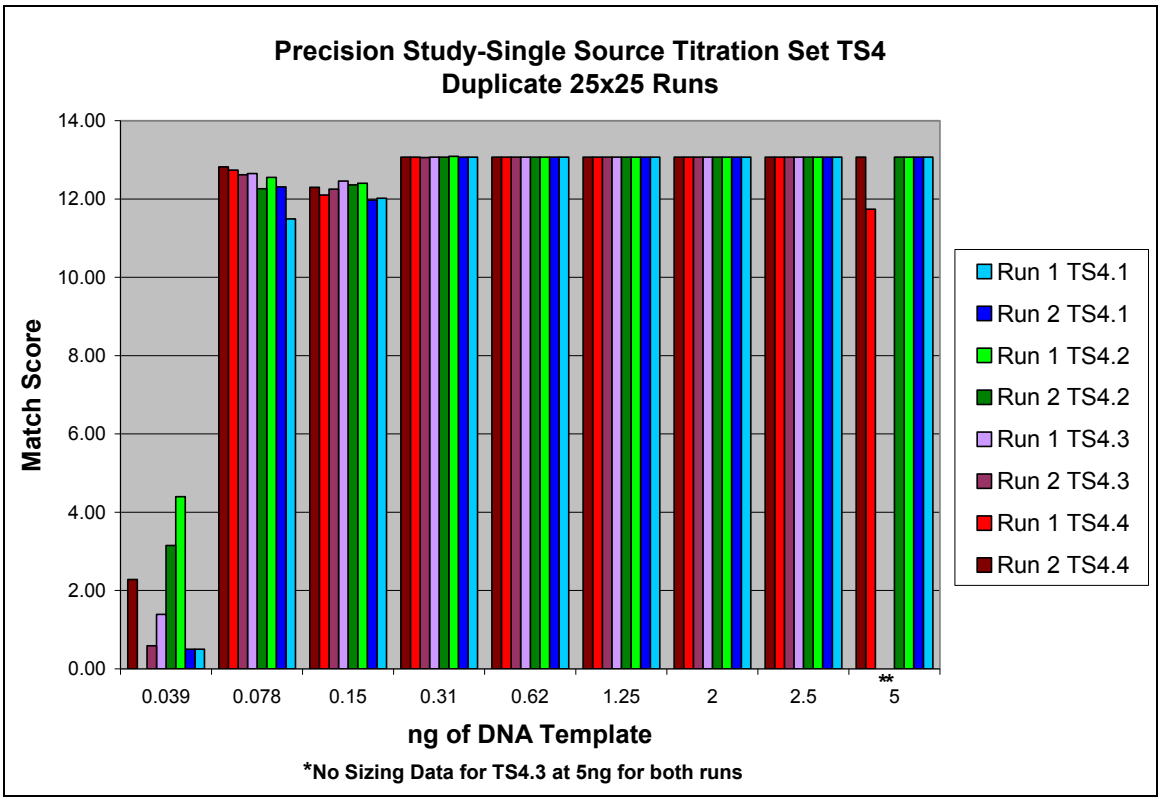
a. Single Source

i. The five sources of DNA used for the titration sets exhibit differences in the amount of target DNA based on the rfu values obtained on the CE. Some variation is expected as quantitation may not result in exact amounts from different individuals biological fluids. There is also expected variation when using different capillaries, arrays, etc. during the electrophoresis process. The amount of DNA ranged from more to less concentrated for sources TS2, TS5, TS3, TS4 and TS1 respectively. Hence, there may be some differences between sources but the overall trends were the same.

ii. The match scores were graphed in Figure 31 for each titration source over all DNA targets run in duplicate at the regular setting. The TACW runs exhibited no significant variation with 150 pg or more DNA. Minor variation was evidenced at 78 pg. More variation was observed when analyzing 39 pg. This would be expected as it represents the DNA from only 5-6 diploid cells.

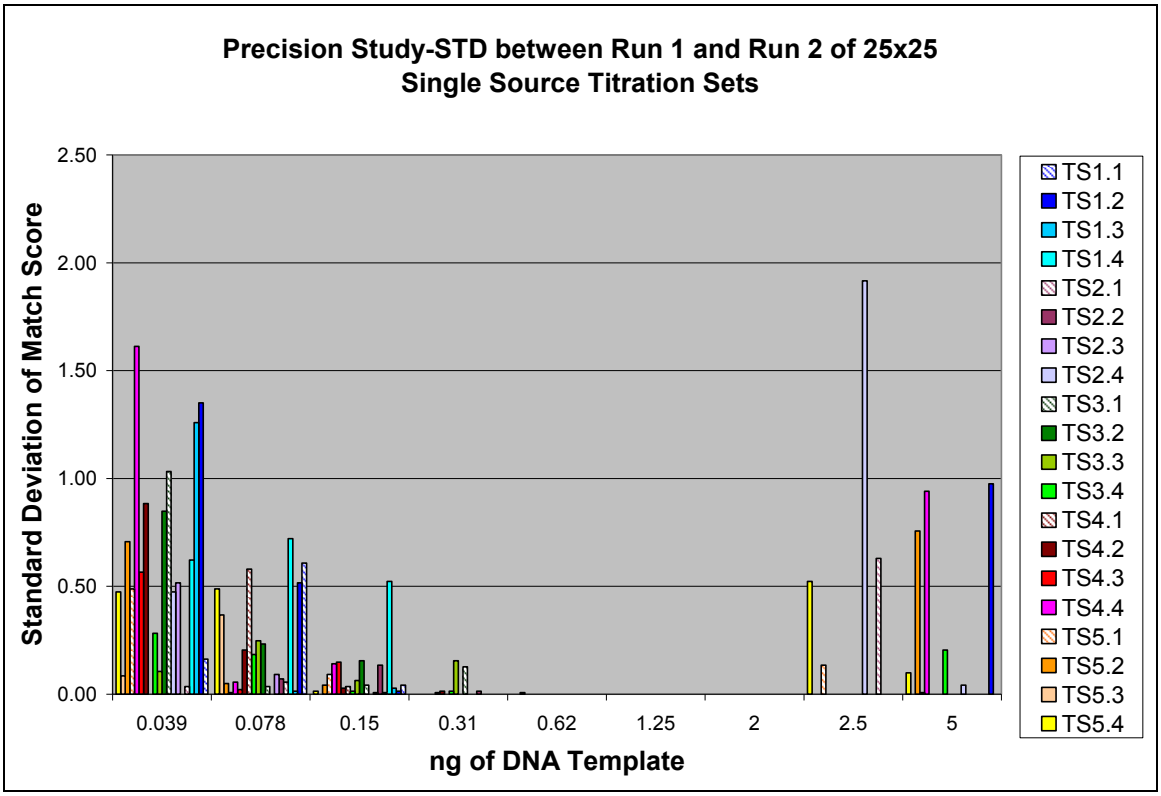
iii. These results demonstrate that lower peak heights contain higher levels of uncertainty and likewise exhibit higher standard deviations between runs.

Figure 31: Precision Study: Duplicate Regular runs of Single Source Titration set TS4.



iv. Figure 32 displays the variation as the standard deviation of the match score over the range of DNA. The trend is illustrated by the higher to lower SD from 39 pg to 150pg and then no variation until larger target DNA of 2.5ng to 5 ng are reached. There are also more runs exhibiting the variation consistently at the low end and only some of the runs exhibiting variation at larger targets.

Figure 32: Precision Study: Standard Deviation Between Duplicate Runs of Single Source Titration Sets.



In summary, the precision or reproducibility of the match score for single source samples is very good for samples with ample target. There is more variation when small amounts of target DNA are present. Additionally, if large amounts of target DNA are present the linearity of the data is affected and data may be less reproducible.

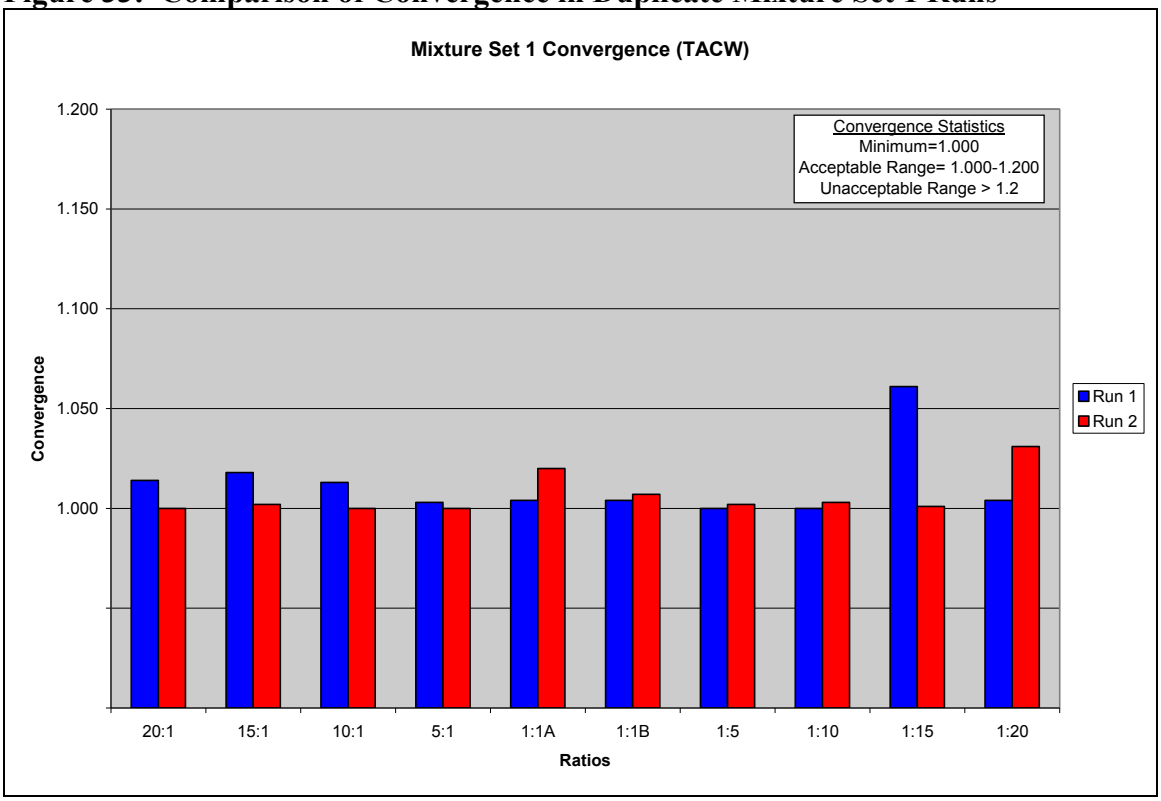
b. Mixture Titrations

The single source titration data could not be used to study parameters that would affect mixtures. Hence the mixture titration sets were used to study reproducibility of data for convergence statistics, mixture weights and standard deviations. All ratios were run in duplicate with the regular setting. Most conditions did not require a rerun, with the exception of the 20:1 in both sets. The following data is described for mixture set 1.

i. Convergence, Mixture Weights and Standard Deviations:

A. The duplicate runs for all data have acceptable convergence statistics of less than 1.2 as shown below (Figure 33).

Figure 33: Comparison of Convergence in Duplicate Mixture Set 1 Runs



B. The precision between runs was very good with the largest difference seen at the 1:15 ratio. In reviewing the duplicate runs at this condition, run 1 and 2 had similar mixture weights of 6.7% and 5.6%, respectively (Figure 34). On the other hand, the match scores for run 1 was more than three-fold that of run 2 (6.834 vs. 2.280) as seen in Figure 36. Note that a third run was not conducted at this time. Although, it may be considered in the future to evaluate if one of the duplicate runs was an outlier.

C. The difference in match score appeared to be attributable to the D16 and TH01 loci. The D16 minor genotype of 12, 12 (approximately 240 rfu), found in a stutter position (although greater than 18% stutter), had a probability of 35% in run 1 and only 6% in run 2. This difference resulted in a match score of 0.5 and – 0.3 respectively. The TH01 locus did not evidence a significant difference in probability (14% to 17%) but did cause the match score to go from a slight positive value (0.099) to a negative score (-1.18). The disparity in this case was recognized when evaluating the match score and would have been difficult if it were a no-suspect case based on just the convergence, mixture weight and standard deviation.

D. In order to further investigate the reproducibility between runs an assessment of the probability of each genotype was undertaken (see section starting on p. 73).

E. The associated mixture weights exhibit very good precision (Figure 34) with only the 1:15 ratio demonstrating a difference greater than 1% (1.1% as evidenced in Figure 35).

Figure 34: Comparison of Mixture Weights in Duplicate Mixture Set 1 Runs

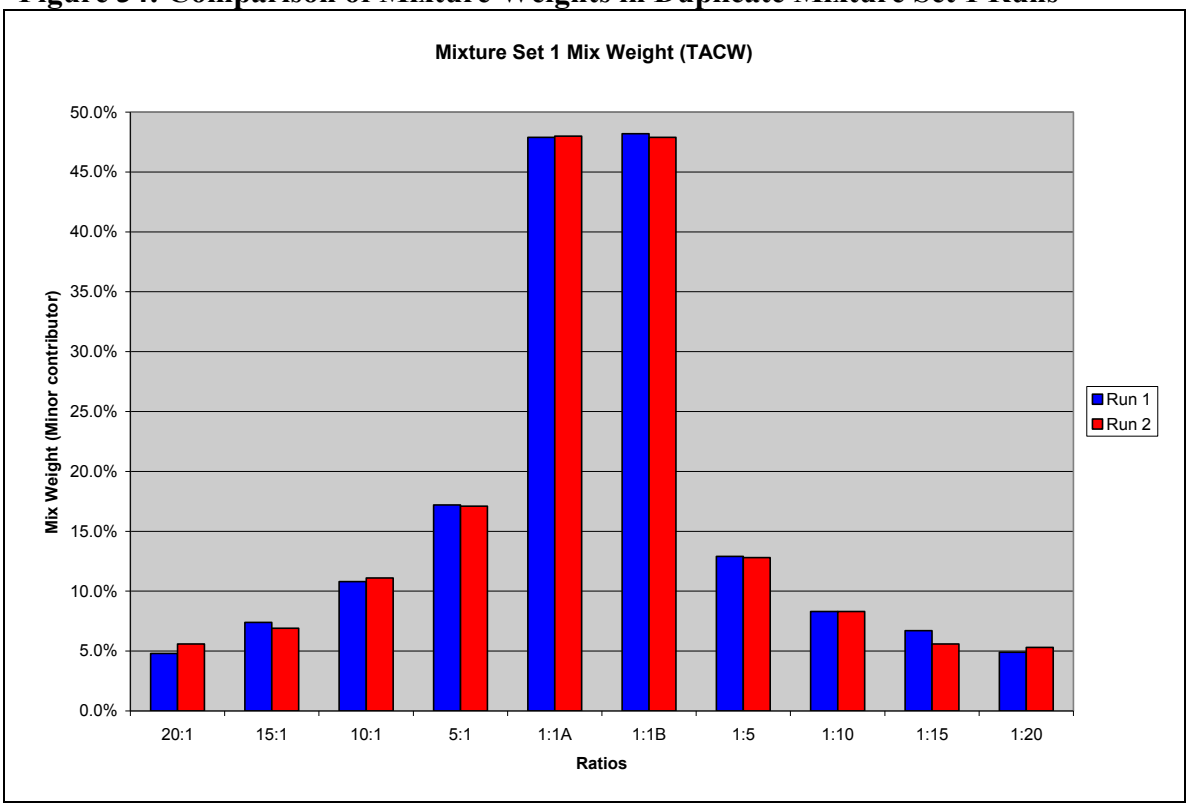
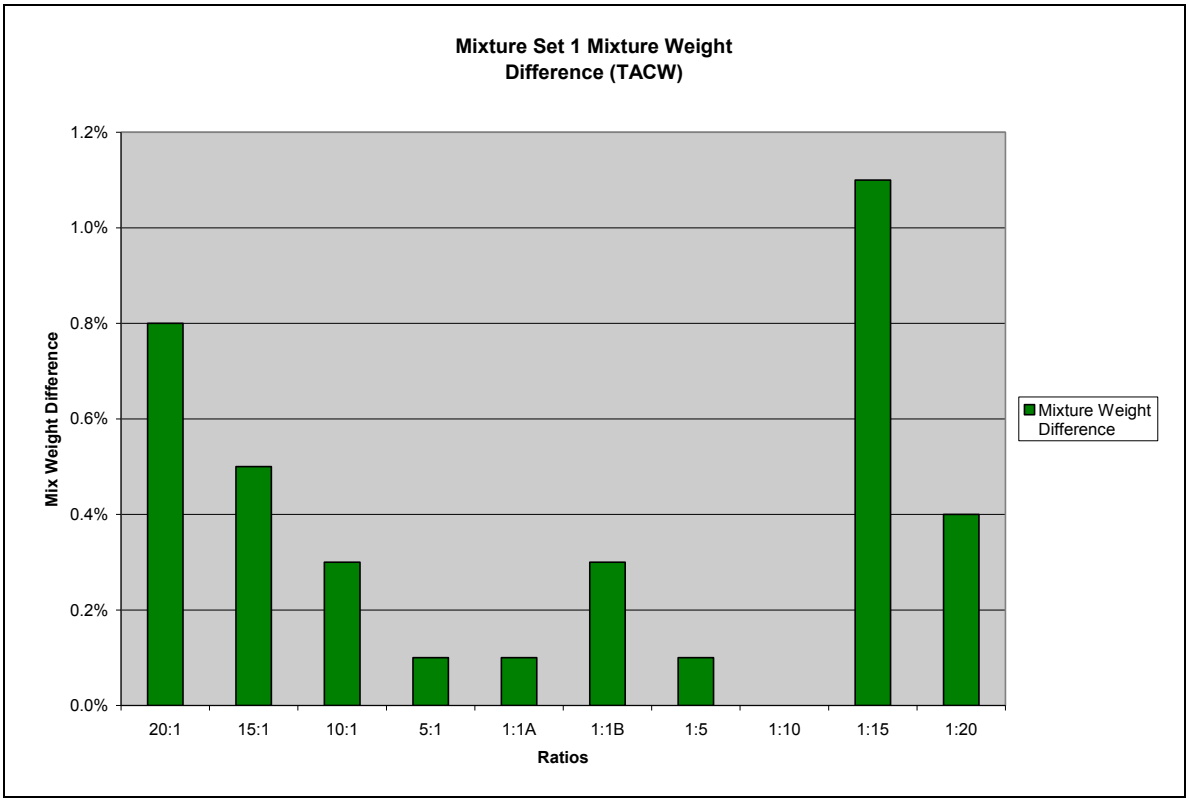


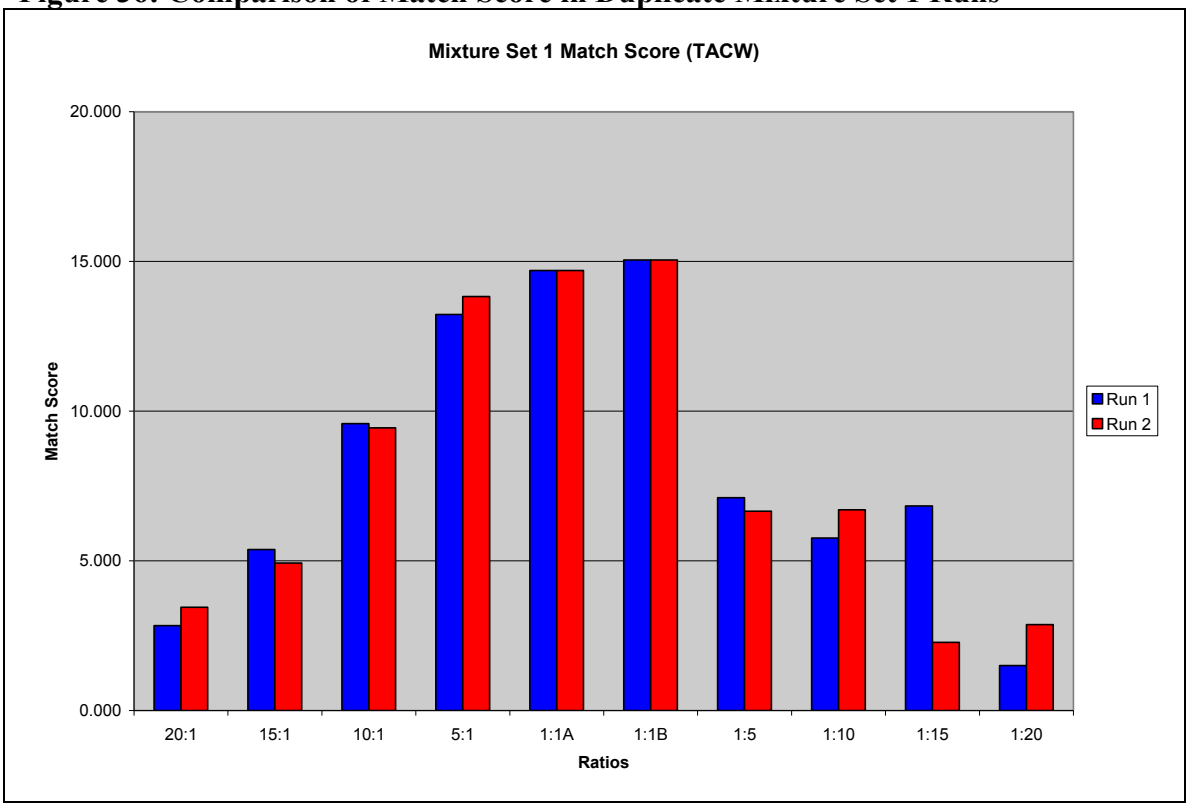
Figure 35: Comparison of Mixture Weight Differences in Duplicate Mixture Set 1 Runs.



ii. Match Scores

A. The match scores for the duplicate runs at each condition are in Figure 36. Overall the match scores were reproducible across the data set with the exception of the 1:15 (as discussed on p. 69). The TAGC provides reproducible data for two person mixtures across a wide range with a 2 ng input of DNA. This is the optimal target DNA used by the laboratory and the ratios encompass the dynamic range of the CE instruments. The TAGC assessed data below approximately 150 rfu's that would not normally be used for a statistic due to the thresholds that are set for manual interpretations.

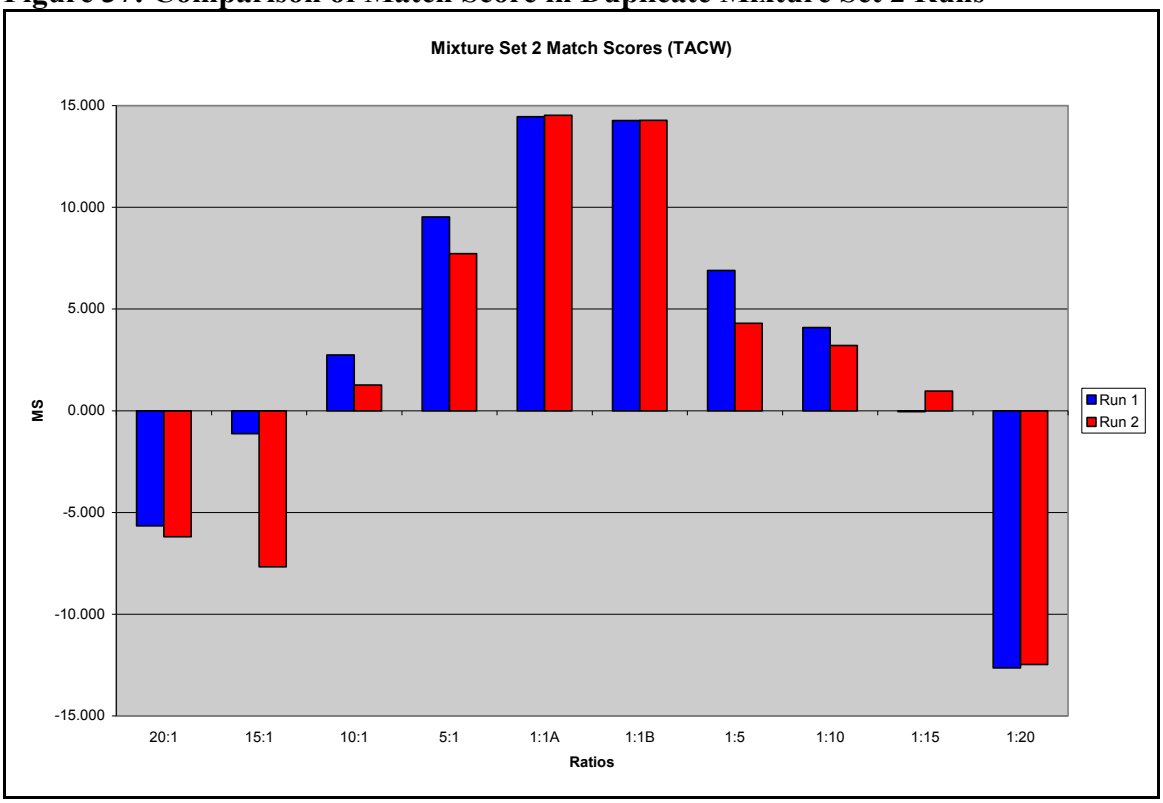
Figure 36: Comparison of Match Score in Duplicate Mixture Set 1 Runs



B. Although Mixture set 2 was not used due to the lower targets of DNA (1ng total) for most comparisons, it is useful to examine affects of decreasing amounts of DNA. Oftentimes, one may amplify samples with less than the optimal target DNA due to limited quantities from crime scene samples. Mixture set 2 had good convergence and standard deviation for all conditions with the exception of the 20:1 which had high standard deviations (> 0.1) in two of three runs.

C. The match scores obtained for Mixture set 2 are evidenced in Figure 37. The data illustrate a trend from more to less DNA target of the minor component whereby match scores are reproducible, become less reproducible and then are reproducible again. Duplicate runs exhibit fairly consistent data when there is ample amount of the minor component. As the minor component decreases, the standard deviation or differences between the runs increases with the largest difference observed at 1:15. Although, once there is very little to no contribution from the minor source at 1:20, the data appears to become more reproducible with lower standard deviations. This may be attributable to the fact that once there is so little data, it is consistently poor or uninformative, that it will remain as such in duplicate runs.

Figure 37: Comparison of Match Score in Duplicate Mixture Set 2 Runs

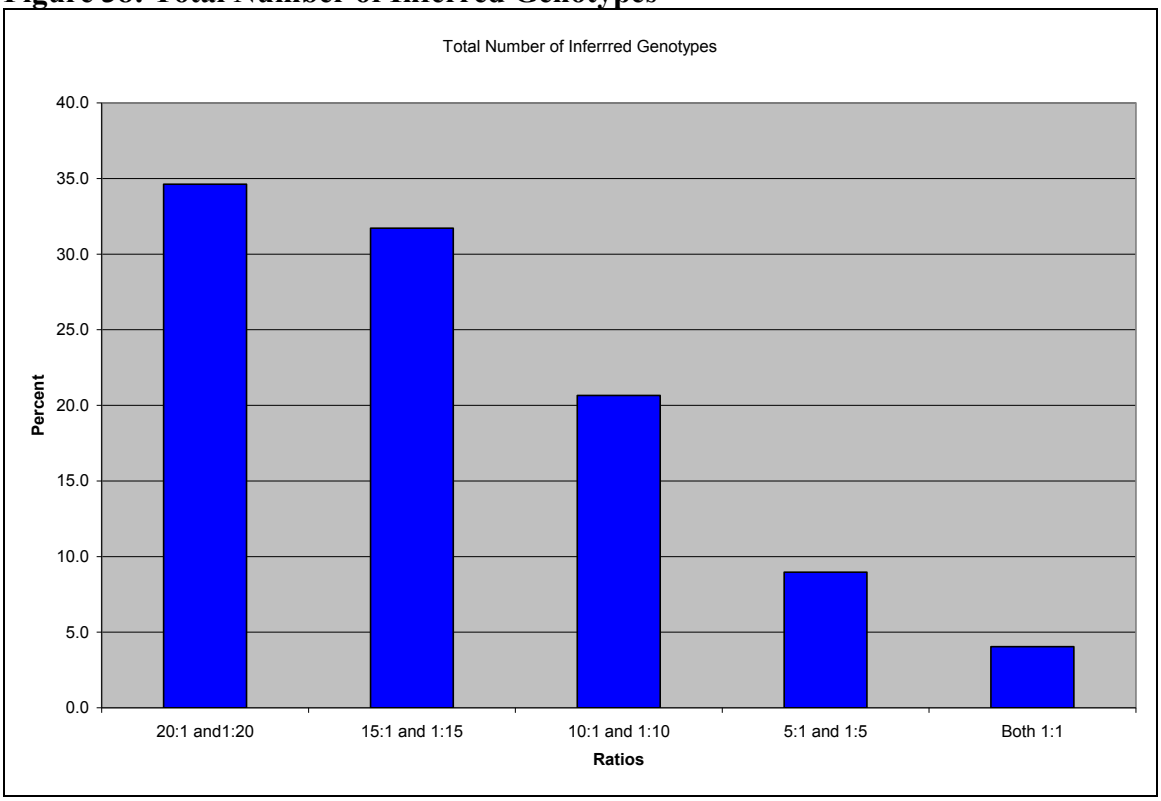


iii. Reproducibility of Genotypes and Associated Probabilities

A. A study was conducted using the duplicate runs from mixture set 1 to evaluate uncertainty and reproducibility by assessing the number of inferred genotypes and probability differences across and within the ratio conditions, respectively.

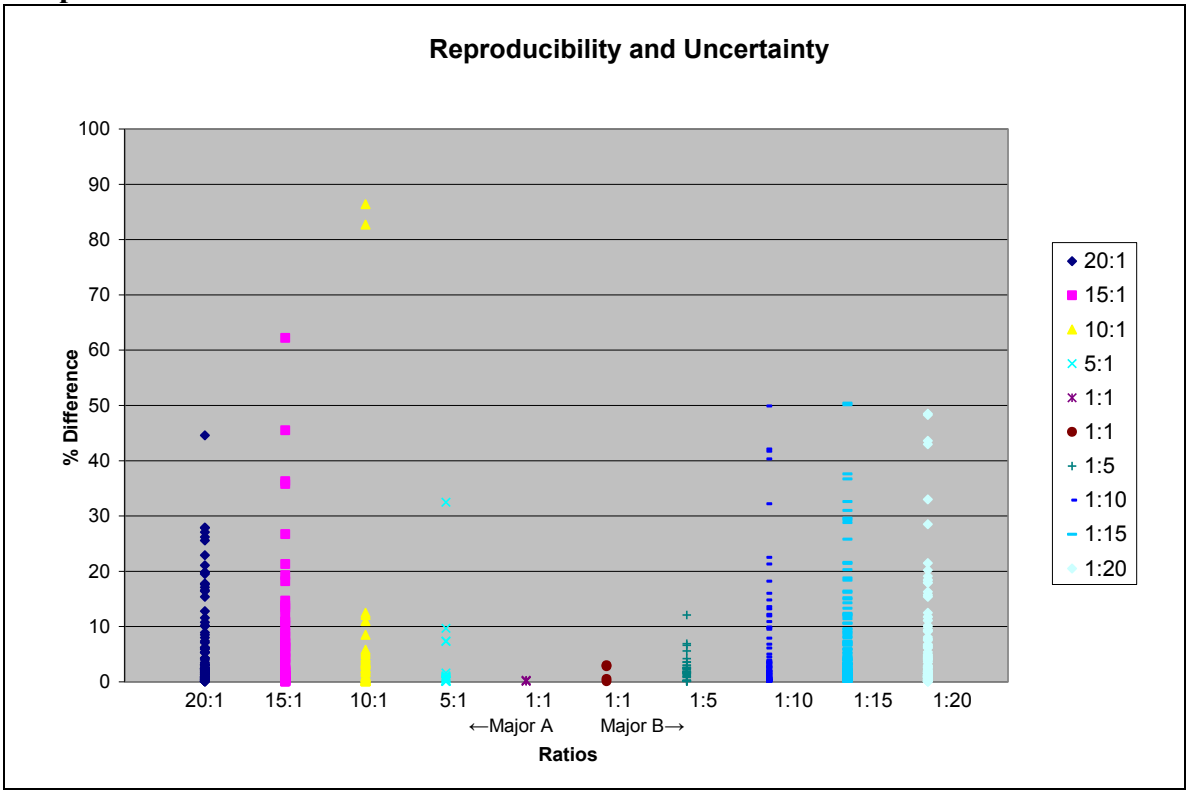
B. The number of inferred genotypes was added for both the 20:1 and 1:20, 15:1 and 1:15, etc. to determine the number of genotypes expected when the minor component ranges from a lesser to greater portion of the mixture (i.e. approximately 4.7 % at 20:1 or 1:20 up to 50% at 1:1). The graph in Figure 38 illustrates the trend that more genotypes are inferred when there is less DNA from the minor component. The total number of genotypes found at the 20 conditions accounts for a greater percentage than the 15 conditions, etc.

Figure 38: Total Number of Inferred Genotypes



C. Duplicate runs within each condition were assessed to investigate differences in genotype probabilities (i.e. comparison of the probabilities for the same genotype in the 20:1 between run 1 and run 2). This difference is what was previously discussed as accounting for the difference in match score between runs when compared to a suspect at the 1:15 condition (see p. 69). The data below is not a comparison of only the suspect genotype. The probability of each inferred genotype at all loci for every mixture ratio was compared. The difference in the reported probabilities was graphed in Figure 39.

Figure 39: Reproducibility and Uncertainty: Genotype Probability Differences in Duplicate Runs—Mixture Set 1



D. The graph clearly illustrates that more certain data, with ample amounts of the minor component (1:1, 1:5), are very reproducible with almost no differences observed in the majority of the data.

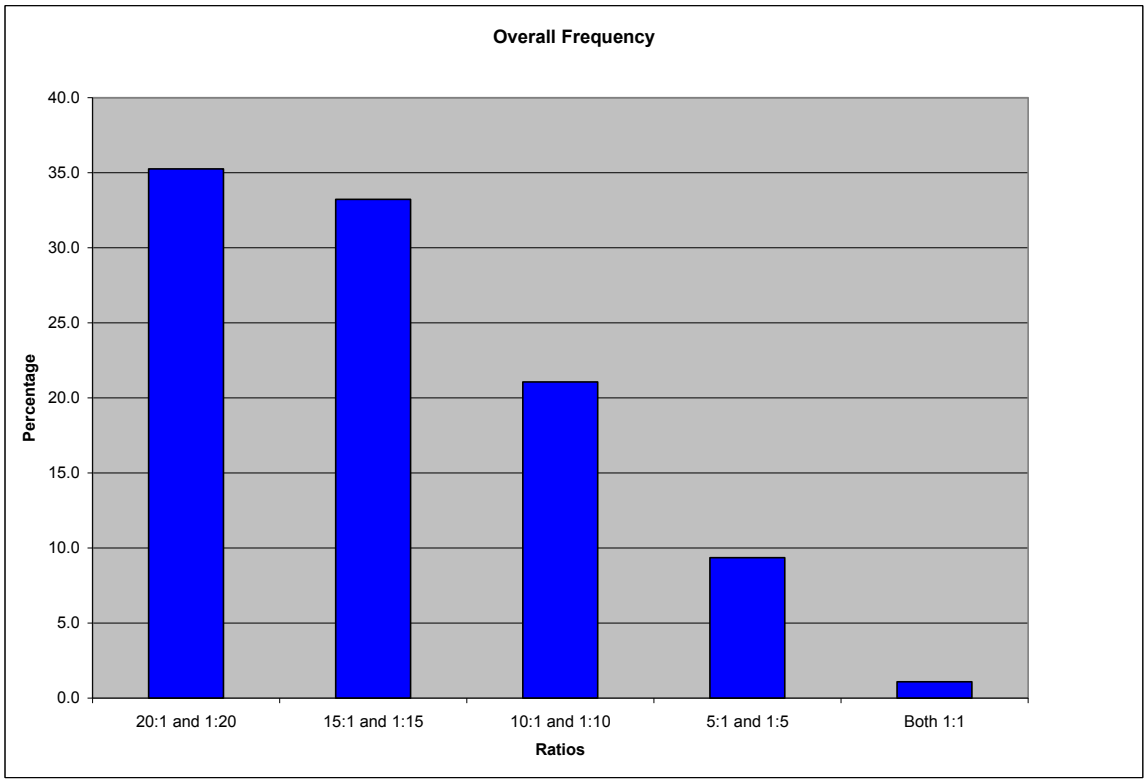
E. On the other hand, as the target DNA of the minor component decreases and less information is available from the minor component, larger probability differences occur between duplicate runs and are observed more frequently.

F. There were two data points, evidenced in yellow, that demonstrated differences greater than 80% (86.4% and 82.7%). Both were found at the D8 locus in the 10:1 condition whereby the major genotype switched from a 15, 15 in run 1 to a 11, 15 in run 2. The second run also contained an inferred genotype to include a 10, 15 whereby the 10 was in a stutter position but the first run did not**.** This is because there is more uncertainty and more variation in the inferred genotypes when there is low level DNA.

G. The frequency or distribution showing the number of occurrences with percentage differences at each of the conditions is evidenced in Figure 40 (e.g. number of data points with differences at 20:1 and 1:20 divided by the total data points for the entire set

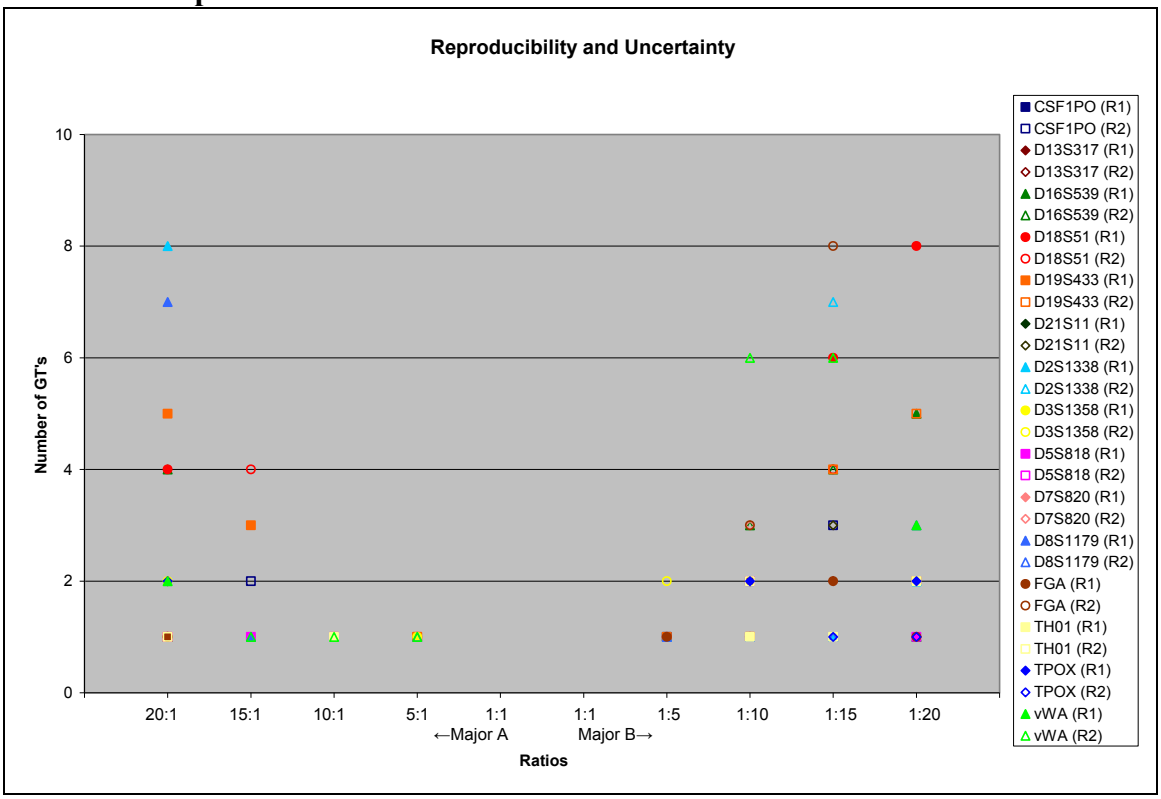
accounted for over 35% of the data). The trend was linear, displaying there are less differences in probabilities observed when there is more DNA in the minor component.

Figure 40: Overall Frequency



H. Additionally, there were instances whereby one run contained inferred genotypes whereas the other run did not. The data demonstrative of the presence of a genotype in one run and absence in the other is depicted in Figure 41.

Figure 41: Reproducibility and Uncertainty: Genotypes present in one run and absent in a duplicate run.



I. The majority of genotypes had very small probabilities and may have resulted in one run and not the other due to the random sampling between runs. There were 215 occurrences investigated with about 69% at a probability of less than 0.01 and 25% between 0.01 and 0.05.

J. One instance was at 70.3%. This point was at the 1:20 condition and contained a genotype to include a stutter position whereas the duplicate run did not contain the stutter peak in the set of inferred genotypes.

K. In reviewing the instances found with probabilities between 5 and 40% (11 data points) the genotype differences were attributable to peaks in stutter positions, masked under one of the victim/major alleles or consisting of very small rare alleles (less than 20 rfu). Hence, the presence of a genotype in one run and absent in another run is to be expected when employing probabilistic genotyping methods.

L. This phenomenon is similar to the PCR process in that large targets of DNA produce very consistent and reliable genotypes and lower DNA targets exhibit stochastic effects that result in variable

genotypic data in different amplifications of the same target amount. Low template DNA analysis is conducted by performing amplifications in triplicate and checking for concordance in two of the three runs. Likewise, it may behoove the analyst to run uncertain data more times to assess reproducibility and make appropriate interpretations.

In summary, results from this precision study indicate that additional runs of the same sample would be useful for assessment of more uncertain data. It may behoove the analyst to rerun the sample under the same conditions to evaluate the answer. If subsequent runs give different answers then the data may be more uncertain and be inconclusive whereas similar answers may assist in making a more definitive conclusion of inclusion or exclusion.

Note that the differences between runs were assessed for automated methods but were not for manual methods. There are expected differences for manual mixture interpretations since analysts have various levels of experience and may use discretion when performing interpretation and accompanied statistical analyses (p. 83).

G. Reruns

In conducting the validation it became evident that reruns would assist the user to consider alternate hypotheses of the number of donors, a longer run time and/or to check reproducibility.

I. Number of Donors:

a. Determining the number of donors in a DNA mixture is not always straightforward and must be assumed based on the number of alleles and use of quantitative peak height information. The user makes the request accordingly and then, after review of the modeled data based on the assumption, may have to pose another hypothesis or alternate number of donors to the TAGC. Several cases exemplified the difficulty encountered in making an assumption of the number of contributors.

b. The number of contributors that are assumed to be present in a mixture can have a very significant affect on the conclusion. Oftentimes a mixture assumed to be from two individuals may exclude an individual (i.e. there are obligate alleles foreign to the victim that do not originate from the suspect). Alternatively, that same mixture considered as a three person mixture, may include the individual in question (i.e. all alleles are in the mixture with additional obligate alleles from a third person).

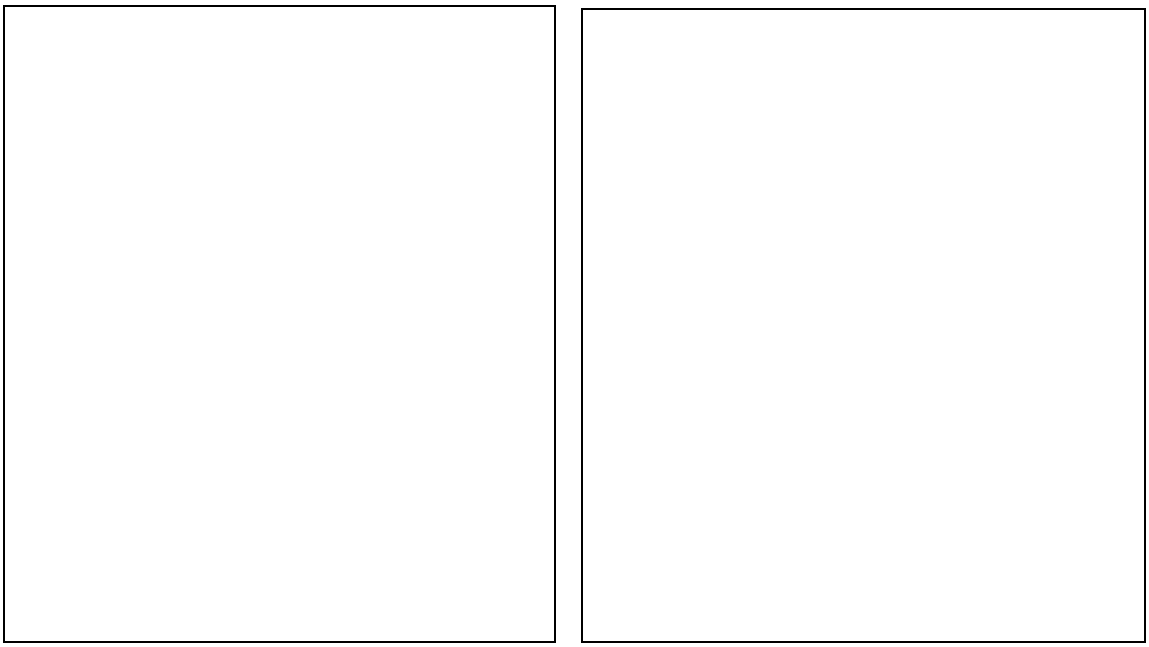
c. Some cases were run initially as V + 1 and then run additionally requesting a V + 2. There were 3 cases (8, 10, and 21) that contained trace third contributors that were investigated to make comparative analyses of the two requests and accompanying match scores. Appendix J illustrates that match scores of a suspect to the secondary contributor are not significantly different. This study, using the regular setting, concluded it is recommended to make the request as V + 2 if a trace third component is apparent in the mixture.

d. Other cases presented data that may call for another request to be made after inspection of the modeling and use of the explain window. Case 12 was discussed previously (pp. 38, 58) and did not display a 'good fit' when run as V + 2. Three donors weights (with the victim as a given) did not seem to account for the peak data in the electropherogram derived from this bra. Use of the explain window (18) enabled an investigation of the mixture containing both the victim and suspect genotypes with a tertiary 'other' component. The hypothesis that the mixture was comprised of the victim, suspect and other did not fit the data (see Figure 42). Hence, there may be a different number of donors or the suspect may not be a contributor to the mixture.

Figure 42: Explain window.

RSAV-12 **V+2**

FGA **vWA**



These figures are from the Explain window in the Review module and represents how the inferred genotypes for the three contributors best fit the data with the proportion of the peak heights. Gray, blue and orange boxes overlayed on top of the electropherograms show what scenarios give the best fit to the data. The sizes of the boxes represent the proportion of the mixture

weights for each contributor. Gray= Known Victim, Blue=second contributor and Orange=third contributor.

Table 23: RSAV-12 Genotypes from Explain Windows

RSA Case # / Locus	Known Victim (Gray)	Known Suspect 1 (Blue)	Unknown Other (Orange)
RSAV-12 Mix weights	*49.5%*	*36.3%*	*14.2%*
RSAV-12 FGA	20,23	24,24	22,22
RSAV-12 vWA	14,18	16,17	15,19

FGA: This locus had 5 called alleles and did not fit the data well under the scenario of genotypes depicted in the table. The blue bar far exceeds the peak and the orange bar is significantly under the peak. The victim genotype of 20, 23 is not modeled well under this scenario either.

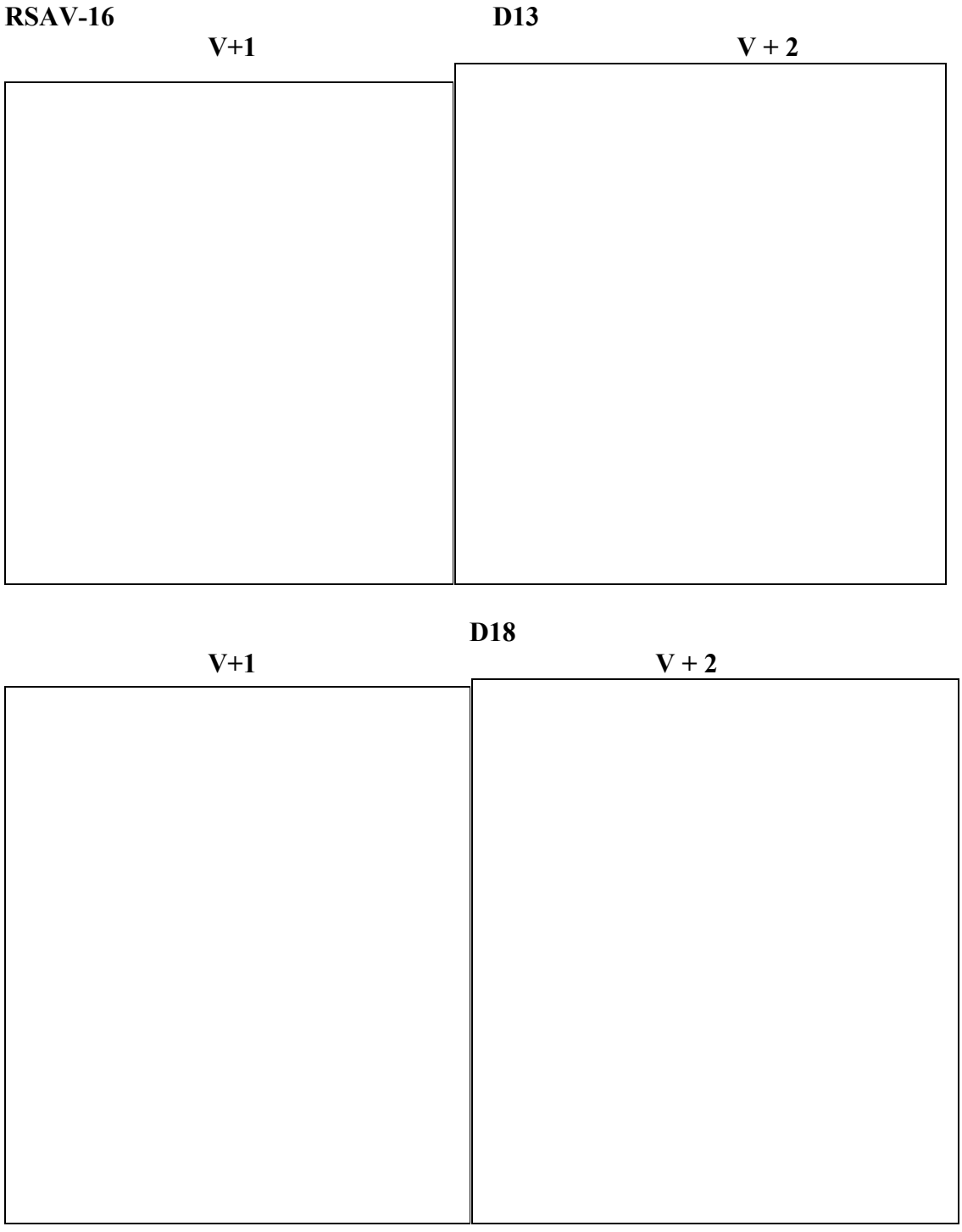
vWA: This locus contained 6 alleles and was the most informative to investigate a potential three person mixture containing the victim, suspect and other. It is clear that the quantitative peak height information attributed to the three heterozygote genotypes proposed in the model do not fit. The blue bars are much higher than the peaks and the orange bars are much lower.

e. Case RSAV-16 was assessed thoroughly in the explain window. This case was discussed earlier (pp. 40-46) and demonstrated the possibility of a third contributor, possibly a relative of the victim. This was a no-suspect case with a secondary contributor deduced for upload, but when the data was presented as a two person mixture, it did not fit well (see Figure 43). The data not fitting the pattern would result in a lower match score than would have been expected for a two person mixture with a 16% minor donor (e.g. case 5E2 deduced to a single source at 14 loci whereas case 16 contained only 7 single source loci). It would have been interesting to obtain elimination samples from family members that could have been present on the bite mark swabbing to assist with a three person deconvolution to assess if the data would render a better fit. This may then allow for a better representation of the data and potentially a more discriminating upload profile for searching.

f. The figure below (Figure 43) demonstrates the modeling at two loci, D13 and D18, for both V + 1 and V + 2. The two donor scenario exhibits extra DNA in the 11 allele at D13. The hypothetical three donor scenario was modeled using the explain window by assigning a child as a homozygote sharing the parent allele. Although the child may be heterozygous (and another allele could be present that does not overlap the victim), this was an assumption made for demonstrative purposes. The three person scenario is a better fit of the data

and may assist in explaining why this sample was not deconvoluted as well as expected when run as a two person mixture with a 16% minor component.

Figure 43: Explain window.



g. The D18 locus appeared to contain a very small 19 allele that was not accounted for in the two person mixture. Given the scenario of a child of the victim (mother) being present, this 19 allele may be explained as potentially coming from the father.

Table 24: RSAV-16 Genotypes from Explain Windows

RSA Case # / Locus	Process	Known Victim	Unknown Suspect-Secondary contributor	Unknown Suspect/Other-Tertiary contributor
RSAV-16 Mix weights	*V+1*	*84.1%*	*15.9%*	
RSAV-16 Mix weights	*V+2**	*80.3%*	*18.6%*	*1.2%*
RSAV-16 D13	V+1	11,12	11,12	
RSAV-16 D13	V+2	11,12	11,12	11,11
RSAV-16 D18	V+1	13,18	14,20	
RSAV-16 D18	V+2	13,18	14,20	13,19

*Note that the V + 2 run did exhibit a low SD (< 0.01) and the Markov chains may have been stuck.

h. Forensic samples are variable in nature and may often be comprised of unexpected mixtures. Initial assumptions may need to be reevaluated and alternative scenarios considered, especially to address CODIS uploads. CODIS approaches can be implemented by running various requests and using different search strategies.

II. Number of Known Donors:

a. Submittal of an elimination sample(s) that may be present in a mixture is important. Sexual assault cases are sometimes accompanied by the consensual partner of the victim to assist with interpretations. An exercise was conducted with case 13 item E3 to simulate a mixture with a consensual donor present. This face swabbing was consistent with a mixture of the victim and two suspects. The original runs were requested using just the victim as a known. The data was rerun using the V + S1 (e.g. assume S1 to be a husband) in deconvoluting S2 and vice-versa. All runs were in duplicate and the data for average match score is presented.

b. The data presented previously from the 2 unknown runs are compared to the one unknown runs in the table below (Table 25). All were run as three person mixtures. The data clearly demonstrate that a higher match score can be obtained from a difficult three person mixture if two of the donors are known.

Table 25: Match Score comparison of two unknowns and one unknown

RSAV-13 E3	V + S + one unknown		Victim + 2 unknowns	
	Suspect 1	Suspect 2	Suspect 1	Suspect 2
Average Match Score	5.02	6.32	1.48	2.41
Mixture weights	81.1/10.2/**8.8**	82.7/**9.9**/7.4	78.1/12.4/8.8	

c. In this scenario, running the mixture with two knowns (i.e. with the husband reference) increases the information gained (LR/match score). The TAGC was able to better deconvolute this mixture by using more information. Data can be rerun, with different requests of the number of donors, if reference samples are submitted at a later date to better interpret the evidence.

III. Longer Run Time:

a. The majority of sexual assault cases could be performed with the TAGC at the regular setting. Easy two person mixtures, with ample DNA target from the secondary contributor do not require any extended run time. Intermediate cases containing a third contributor do not generally require a longer run time unless there is limited target DNA or contributors with fairly equal targets with neither of known origin (e.g. case 13). Complex cases may require longer run times.

b. Longer run times give the computer more time to sample difficult mixtures in an effort to better separate components and establish mixture weight percentages. Future studies are recommended in later phases to run mixture titration sets contrived of three persons to investigate when the longer setting is needed.

Overall Summary

In summation, it is clear that significant differences exist between manual and automated methods due to the random sampling and fluidity of probabilistic genotyping. As aforementioned in the introduction, the TAGC is a tool that can be assimilated to an instrument that performs a measurement that may yield different answers in each run. As such, additional measurements/runs may be necessary when the data contains lower targets of DNA or uncertain mixture components. There are expected differences obtained with regard to match score, mixture weights and standard deviations.

The reason for some of these differences may be explained by a comparison of the methodologies with regard to interpretation and statistics. Manual methods vary between analysts at the data reads and interpretation stage. Analyst discretion is used to make inferences regarding:

- determination of the number of contributors
- deconvolution of major and minor components
- editing of artifacts
- designation of potential alleles below threshold
- calculation of ratio/percentage of donors
- application of sister allele balance and stutter filter percentages
- acceptance of mixture weight ranges
- consideration of DNA target, degradation and drop out
- assessment of which loci will be used for statistics

If analysts are concordant in evaluating each of the above considerations, the inferred genotype(s) and the subsequent statistic would be identical (i.e. same interpretation = same statistic). If analyst discretion led to differences in inferred genotype(s) and/or which loci to be used or method applied (e.g. MPE, MMPE or CPE/CPI), a different statistic would result (i.e. different interpretation = different statistic). These differences are what lead to inconsistencies due to the discretionary assessments and subjective nature of manual interpretations.

On the other hand, automated methods combine the interpretation and statistical analyses such that the final statistic, or match score, may be different dependent on the entire process of inferring genotypes using probabilities from random sampling of the same data. The modeling is processing many variables simultaneously including peak heights and variation around each peak, baseline and artifacts, sister allele balance, stutter, and mixture weights of the components. The computer will perform thousands of iterations to try to model the data to resolve the mixture (similar to performing many amplifications of a particular DNA target). If the model fits the data very well the probability of that particular genotype/profile will be high and the accompanying statistic will be more discriminating with an accompanying higher match score (similar to the reproducibility expected when amplifying a larger target of DNA). If the data does not fit the model well the probability of that genotype will be decreased and the accompanying LR will be lowered (similar to the decreased profile information and reproducibility expected when amplifying a smaller target of DNA). Run to run variability is to be expected as the process is random and each modeling of the data is independent (similar to each PCR reaction being independent of the other). Manual methods generally encompass examining one capture of an amplification by assessment of a single electropherogram. The TAGC uses that single electropherogram and calculates variations around the data in each iteration/trial to simulate conducting many PCR amplifications for the model to explore. For a request without using the victim as the known (e.g. oneunknown or twounknown), the computer applies the mathematical formulae for the questioned samples without any knowledge of the reference samples and objectively develops a set of inferred genotypes with associated probabilities for subsequent comparative analyses to a suspect.

As evidenced throughout this validation, more certain data yields discriminating and reproducible statistics whereas less certain data may yield less discriminating and reproducible match scores (section F, precision). If data is less certain, reruns are

recommended for assessment (section G, reruns). Match scores were averaged as part of this validation to make appropriate comparisons to manual methods. Operationally, one may opt to use averages, select the more conservative statistic, or provide all match scores from each run. Each case requires manual review of the data for reporting when using the TAGC as an expert assistant.

Conclusions

This extensive internal validation provides data to verify the use of the TrueAllele software as an expert assistant using file information derived from an ABI 3130xl with collection software version 3.0. The TAGC has proven to generate data that is accurate and reliable. The validation studies undertaken have compared and contrasted the data sets and results from the manual and automated methods. The differences between the data sets were explainable and resolvable; allowing the user to be confident that interpretation and review using the new probabilistic genotyping method will provide comparable or better results to the manual method. The TAGC demonstrated better deconvolution of two and three person mixtures. This tool will assist the user to conduct searches for CODIS upload cases and/or result in more information or better discrimination potential for comparisons with suspects.

The current version (Analyze Module Build # 252, Version # 9 (8-July- 2010); VUIer Version # 3.3.3919.1 (16-Jul-2010); Server Version # 3.25.3768.1) of the TAGC software is hereby respectfully submitted for approval as an expert assistant for review of sexual assault cases, including single sourced items or DNA mixtures of two or three individuals.

Check with Amy to keep here or separate

JBS to tell operations RSAV-22. TELL AMY: Note discrepancy (manual) at D21 uploaded as 30.2, 32.2 and not 30, 32.2

Quality Assurance Parameters for the TrueAllele Casework Validation for Phase I Reported Sexual Assault Samples

Quality assurance standards are described herein as they apply to the validation of the TrueAllele casework system using the TrueAllele Genetic Calculator from Cybergenetics. This document is provided to comply with standard 8.3.2 and has referenced other applicable standards (quoting of the standards are italicized).

Developmental Validation

Standard 8.2 The TACW system has undergone developmental validation as addressed in several peer reviewed publications (10, 11, 23, 33).

Internal Validation

Standard 8.3 The validation herein has included studies to address those in **standard 8.3.1** *known and evidentiary samples, reproducibility and precision, sensitivity and stochastic effects, mixture interpretations and contamination assessment.* This written validation serves as documentation of the studies and provides a comprehensive summary for each section.

Standard 8.7 is the primary standard applicable to software validations. It states: ***Modifications to software, such as an upgrade, shall require a performance check prior to implementation. New software or significant software changes that may impact interpretation or the analytical process shall require a validation prior to implementation.*** The validation is for new software that will impact DNA interpretations using a quantitative, probabilistic genotyping approach.

Standard 8.3.2 Internal validation shall define quality assurance parameters and interpretation guidelines.

QA Parameters:

- TACW is a tool for mixture interpretation that uses the final .fsa file from the CE instruments and therefore, no 'wet' techniques were encompassed for extraction, quantitation, amplification or separation. Below are the parameters that may still apply to this 'dry' validation of a new software program.
- Facilities (6.1)
    - 6.1.1 All validation studies comply with security measures in place for casework samples. All data was kept on a separate, secure server for data analyses.
- Analytical Procedures
    - 9.5.5 A NIST Traceable sample, EXP10-FTA, was run.
    - Interpretation of Data (9.6)
        - 9.6.1 Controls are assessed in the analyze module.
        - 9.6.2 Statistical analyses for a given population follow the NRC recommendations and use the FBI database.
        - 9.6.4 The TACW system will be used according to protocol XXXX and as such, will *follow a documented procedure for mixture interpretation that addresses major and minor*

contributors, inclusions and exclusions, and policies for the reporting of results and statistics.

- Equipment Calibration and Maintenance (10.1)
    - The TAGC was configured by Cybergenetics and the current version has been thoroughly validated internally. Subsequent versions will undergo performance checks as required.
- Documentation/Reports (11.1)
    - The interpretive information derived from the TAGC will be kept electronically for each item analyzed in appropriate folders on secure drives within the DNA unit or the Maynard facility. The data, in hard or electronic format, will comply to retain *sufficient documentation for each technical analysis to support the report conclusions such that another qualified individual could evaluate and interpret the data.*
    - 11.3 Confidentiality: Electronic data from .fsa files of samples used for this study do fall under the category of evidence and hand written records of the sample information are contained in the validation binders. Information was redacted or encoded, as necessary. ??Amy--what about any mention for when operational--??
- Review of Validation (5.2.3)
    - (5.2.3.2.1) The technical leader has documented review of this validation by initialing each page of the summary. Check w/ Amy
- Review (12.1)
    - 12.1.1 Talk to Amy about tech review qualifications
    - Technical Review Documentation (12.2)—all elements under this standard will be addressed for cases processed using the TAGC, including specific review of all electronic data to support conclusions (12.2.2).

Check with Amy about PT's ,15.2.2 for audit of validation

Green below is from the IDD validation---talk about the TACW one now with Amy

Standards to be addressed in the future, as the validation transitions into DBX operations.

NDIS Validation—check this with Amy and Sid

Standard 8.3.4 (Database ONLY) For inclusion into NDIS of profiles reviewed by an expert system, the expert system shall be validated in accordance with applicable NDIS operational procedures.

Standard 9.6 (Database) The laboratory shall have and follow written guidelines for the interpretation of data. An NDIS approved and internally validated Expert System may be used to complete the data interpretation process.

Standard 10.2.2 The following critical equipment requires quarterly recertification:

Standard 10.2.2.1 Expert systems approved for use at NDIS

Competency Testing

Standard 8.4 (Database) Before the introduction of a methodology into the database laboratory, the analyst or examination team shall successfully complete a competency test(s) to the extent of his/her participation in database analyses.

References

1. TrueAllele® Casework System (Cybergenetics, Pennsylvania) http://www.cybgen.com/systems/casework.html
2. Solving DNA Mixtures with a Visual Calculator. *Forensic Magazine.* (2008/2009) Dec/Jan, 5(6):32.
3. Roby, R.K. (2008) Expert systems help labs process DNA samples. *NIJ Journal* 260:16-19.
4. Scientific Working Group on DNA Analysis Methods. SWGDAM Interpretation Guidelines for Autosomal STR Typing by Forensic DNA Testing Laboratories. *SWGDAM Interpretation Guidelines for Autosomal STR Typing.* January 14, 2010.
5. FBI Laboratory. (2005) APPENDIX B: Guidelines for Submitting Requests for Approval of an Expert System for Review of Offender Samples. *National DNA Index System (NDIS) DNA Data Acceptance Standards*.
6. Massachusetts State Police National DNA Index System (NDIS) B Validation, Validation of TrueAllele Databank v 2.9 Expert System, Federal Bureau of Investigation and NDIS submitted and approval letter, May 2, 2008.
7. Massachusetts State Police DNA Methods Manual. *DNA-26 STR Data Interpretation of known DNA Samples & Technical Review utilizing TrueAllele® Databank.* Version 1.0, August 31, 2009.
8. Good, I. J. (1950) Probability and the weighing of evidence. London: Griffith.
9. Gill, P., Brenner, C. H., Buckleton, J. S., Carracedo, A., Krawczak, M., Mayr, W. R., Morling, N., Prinz, M., Schneider, P. M. and Weir, B. S. (2006) DNA commission of the International Society of Forensic Genetics: Recommendations on the interpretation of mixtures. *Forensic Science International* 160:90-101.
10. Perlin, M.W., Kadane, J.B. and Cotton, R.W. (2009) Match likelihood ratio for uncertain genotypes. *Law, Probability and Risk*. 8(3):289-302.
11. Perlin, M.W., Legler, M.M., Spencer, C.E., Smith, J.L., Allan, W.P., Belrose, J.L., and Duceman, B.W. Validating TrueAllele® DNA mixture interpretation. *Journal of Forensic Sciences.* 2011; 56 (November):in press.
12. Perlin, M.W. and Duceman, B.W. (2010) Casework validation of genetic calculator mixture interpretation (A77). AAFS 62nd Annual Scientific Meeting, 2010 February 22-27; Seattle, WA. *American Academy of Forensic Sciences*. p. 62-3.

13. Belrose, J.L. and Duceman, B.W. (2010) New York State Police validation of a statistical tool for genotype inference and match that solves casework mixture problems (A79). AAFS 62nd Annual Scientific Meeting, 2010 February 22-27; Seattle, WA. *American Academy of Forensic Sciences*. p. 64.
14. Cybergenetics. (2010) *TrueAllele® VUIer™ Getting Started.* TrueAllele® Casework System Users Manual.
15. Cybergenetics. (2010) *TrueAllele® Technology Analysis Module: Quality Assurance.* TrueAllele® Casework System Users Manual.
16. Cybergenetics. (2010) *TrueAllele® VUIer™ Data Module: Uploading Data.* TrueAllele® Casework System Users Manual.
17. Cybergenetics. (2010) *TrueAllele® VUIer™ Request Module: Asking Questions.* TrueAllele® Casework System Users Manual.
18. Cybergenetics. (2010) *TrueAllele® VUIer™ Review Module: Getting Answers.* TrueAllele® Casework System Users Manual.
19. Cybergenetics. (2010) *TrueAllele® VUIer™ Report Module: Reporting Results.* TrueAllele® Casework System Users Manual.
20. Massachusetts State Police DNA Methods Manual, *DNA-19 Interpretation Guidelines for Forensic STR DNA Analysis.* Version 3.1, March 26, 2009.
21. National Research Council. (1996) *The Evaluation of Forensic DNA Evidence*. National Academy Press, Washington, D.C.
22. Balding, D.J. and Nichols, R.A. (1994) DNA profile match probability calculation: how to allow for population stratification, relatedness, database selection and single bands. *Forensic Science International.* 64:125-140.
23. Curran, J. (2008) A MCMC method for resolving two person mixtures. *Science & Justice* 48:168-177.
24. Perlin, M.W. Explaining the likelihood ratio in DNA mixture interpretation. *Proceedings of Promega's Twenty First International Symposium on Human Identification*. San Antonio, TX, 2010.
25. Perlin, M.W., Coffman, D., Crouse, C.A., Konotop, F. and Ban, J.D. Automated STR data analysis: validation studies. *Proceedings of Promega's Twelfth International Symposium on Human Identification*. Biloxi, MS, 2001.
26. Massachusetts State Police. 2010 NIJ Efficiency Grant Proposal.
27. Sgueglia, J.B. DNA Mixture Interpretation: Principles and Practice in Component Deconvolution and Statistical Analysis. *Proceedings in American Academy of Forensic Sciences*, Annual Scientific Meeting, Feb 18-23, 2008, Washington, D.C.
28. Koehler, J.J. (2001) When are People Persuaded by DNA Match Statistics. *Law and Human Behavior*, 25: 493-513.
29. Krane, D.E., Ford, S., Gilder, J.R., Inman, K., Jamieson, A., Koppl, R., Kornfield, I.L., Risinger, D.M., Rudin, N., Taylor, M.S. and Thompson, W.C. (2008) Sequential unmasking: a means of minimizing observer effects in forensic DNA interpretation. *J. Forensic Sci.* 53(4):1006-1007.
30. Perlin, M.W. and Cotton, R.W. (2010) Three match statistics, one verdict (A78). AAFS 62nd Annual Scientific Meeting, 2010 February 22-27; Seattle, WA. *American Academy of Forensic Sciences*. p. 63.

31. Kadash, K., Kozlowski, B.E., Biega, L.A. and Duceman, B.W. (2004) Validation study of the TrueAllele automated data review system. *J Forensic Sci.* 49(4):660-667.
32. Perlin, M. W. (2006) Scientific validation of mixture interpretation methods. *Promega's Seventeenth International Symposium on Human Identification*, Nashville, TN. Available at http://www.promega.com/geneticidproc/ussymp17proc/oralpresentations/Perlin.pdf
33. Perlin, M.W. and Sinelnikov, A. (2009) An information gap in DNA evidence interpretation. *PLoS ONE*. 4(12):e8327.