Interpretation of Legacy DNA Data

Table of Contents

Interpretation of Legacy DNA Data

1 Introduction

These procedures describe the procedures for reinterpretion of data from a legacy amplification kit and/or reporting comparisons to previously reported DNA typing results generated using the AmpF*l*STR® Profiler Plus™, Profiler Plus™ *ID,* COfiler™, Identifiler® Plus, and/or MiniFiler™ PCR Amplification Kits that are no longer in use in the DNA Units. In addition, guidance for statistical calculations is included for data generated using legacy amplification kits for which the STRmix™ software for statistical analysis is not validated for use.

2 Scope

These procedures apply to DNA personnel who are authorized to reinterpret data from a legacy amplification kit and/or report comparisons to previously reported DNA typing results generated using an amplification kit that is no longer in use in the DNA Units.

3 Equipment

- PopStats (version 7.0 or higher)

4 Procedure

When comparisons to previously reported DNA typing results are conducted, it may be necessary to reevaluate the DNA data generated using an amplification kit or instrument no longer in use in the DNA units. The data in these instances are referred to as legacy data. Reinterpretation is defined as the need to reevaluate any of the allele calls or genotype calls, remove alleles (or entire loci) from statistical estimates, or change any of the assumptions used to make conclusions. The DNA examiner must be previously qualified and currently authorized in the interpretation of data from the legacy amplification kit and platform instrument model to perform reinterpretation of and/or comparison to legacy data from evidentiary samples. If necessary, the examiner performing reinterpretation should review the validations that support the legacy results.

4.1 Application of Peak Height Thresholds to Allelic Peaks

A. The Analytical Threshold (AT) is 50 relative fluorescence units (RFU) for the Profiler Plus™, Profiler Plus™ *ID*, COfiler™, Identifiler® Plus, and MiniFiler™ Amplification Kits.

B. The Stochastic Threshold (ST) is:
   - 200 RFU for the Profiler Plus™, Profiler Plus™ *ID*, and COfiler™ Amplification Kits
   - 200 RFU for the Identifiler® Plus (27 cycles) Amplification Kit
   - 300 RFU for the Identifiler® Plus (28 cycles) Amplification Kit
   - 300 RFU for the MiniFiler™ Amplification Kit

   1. With the exception of the sample types in 2 below, all the peaks at a given locus must be ≥ ST to be used for matching/statistical purposes. Peaks < ST

may only be used for purposes of exclusion and/or to establish the presence of a mixture of DNA.

2. For samples known or expected to be of single source origin (e.g., reference samples, alternate reference samples, bones) that display results consistent with having arisen from a single individual, the ST is applied to only those loci at which stochastic loss of information is possible (i.e., loci that display a single allelic peak < ST).

4.2 Peak Height Ratios

Peak height ratios (PHR) can be used to associate two alleles to a common source.

A. PHRs are calculated by dividing the peak height of the allele with the lower RFU value by the peak height of the allele with the higher RFU value, expressed as a percentage.

B. Table 1 through Table 4 describe the minimum expected PHR percentages for the applicable amplification kit. The PHRs for Profiler Plus™ *ID* and COfiler™ are dependent on the detection instrument (i.e., ABI Prism 310 or 3130xl) used. PHR guidelines are only applicable to allelic peaks that meet or exceed the ST.

AmpF*l*STR Profiler Plus™ and Profiler Plus™ *ID* Amplification Kit

Peak Height	D3S1358	vWA	FGA	D8S1179	D21S11	D18S51	D5S818	D13S317	D7S820
200-300 RFU	60%	60%	55%	60%	60%	55%	60%	60%	60%
301-1000 RFU	65%	65%	65%	65%	60%	60%	65%	65%	65%
Above 1000 RFU	70%	65%	65%	70%	65%	65%	70%	70%	70%

AmpF*l*STR COfiler™ Amplification Kit

Peak Height	D3S1358	D16S539	TH01	TPOX	CSF1PO	D7S820
200-300 RFU	60%	60%	60%	60%	60%	60%
301-1000 RFU	65%	65%	65%	60%	60%	65%
Above 1000 RFU	70%	75%	75%	75%	75%	70%

Table 1 - Minimum Expected Heterozygous Peak Height Ratio Guidelines for STR Loci Analyzed Using the Profiler Plus™, Profiler Plus™ *ID* and COfiler™ Kits (310 Data)

Peak Height	All Profiler Plus™ *ID* and COfiler™ Loci
200-499 RFU	50%
500-999 RFU	60%
1000 RFU and above	70%

Table 2 - Minimum Expected Heterozygous Peak Height Ratio Guidelines for STR Loci Analyzed Using the Profiler Plus™ ID and COfiler™ Kits (3130xl Data)

Peak Height	All Identifiler® Plus STR Loci (27 cycles)
200-499 RFU	50%
500-999 RFU	60%
1000 RFU and above	70%

Table 3 - Minimum Expected Heterozygous Peak Height Ratio Guidelines for STR Loci Analyzed Using the Identifiler® Plus (27 cycles) Amplification Kit (3130xl Data)

Peak Height	All MiniFiler™ STR Loci
300-999 RFU	50%
1000 RFU and above	60%

Table 4 - Minimum Expected Heterozygous Peak Height Ratio Guidelines for STR Loci Analyzed Using the MiniFiler™ Amplification Kit

C. Because reference samples and human remains (e.g., bones, teeth, tissue samples) are attributable to a single individual, PHR assessments are generally not used in their interpretation.
    1. If necessary, however, major contributor types of these samples amplified using Identifiler® Plus (27 cycles) or MiniFiler™ can be assessed using the peak height ratios described.
    2. Major contributor types of these samples amplified using Identifiler® Plus (28 cycles) can be assessed generally using a PHR of 50%.

4.3 Stutter Percentages

The kit-specific stutter percentage guidelines provided in Table 5 are estimates (Average + 3 standard deviations [SD]) of the maximum expected relative stutter values for each locus in the specified amplification kits. These values are expressed as a percentage relative to the source allelic peak height (i.e., % stutter). Although usually observed as N-4, stutter may occur at other locations (e.g., N-8, N+4).

AmpF*l*STR® Profiler Plus™ *ID* Amplification Kit

Locus	D3S1358	vWA	FGA	D8S1179	D21S11	D18S51	D5S818	D13S317	D7S820
% Stutter	11	12	12	13	11	14	10	11	13

AmpF*l*STR® COfiler™ Amplification Kit

Locus	D3S1358	D16S539	TH01	TPOX	CSF1PO	D7S820
% Stutter	11	10	16	13	8	9

AmpF*l*STR® Identifiler® Plus (27 or 28 cycles) Amplification Kit

Locus	D8S1179	D21S11	D7S820	CSF1PO	D3S1358	TH01	D13S317	D16S539
% Stutter	11	12	10	11	13	6	12	11

Locus	D2S1338	D19S433	vWA	TPOX	D18S51	D5S818	FGA
% Stutter	13	13	13	7	15	11	13

AmpF*l*STR® MiniFiler™ Amplification Kit

Locus	D13S317	D7S820	D2S1338	D21S11	D16S539	D18S51	CSF1PO	FGA
% Stutter	11	11	13	13	13	15	11	12

Table 5 - Maximum Expected Stutter Percentage Guidelines for STR Loci Analyzed Using the Profiler Plus™ ID (includes Profiler PlusTM), COfiler™, Identifiler® Plus, and MiniFiler™ Kits

4.4 Interpretation of DNA Typing Results for Single Source Specimens

A sample is generally considered to have originated from a single individual if one or two alleles are present at all loci for which typing results were obtained and the PHR for all heterozygous loci are within the empirically determined values.

A. For single source specimens, when two peaks at a heterozygous locus each fall into a different peak height category, the PHR will be evaluated relative to the category that corresponds to the peak height of the shorter peak.

B. A sample that displays a heterozygous peak height imbalance at one or two loci, but for which no other results indicate the presence of a mixture, is generally considered a single-source specimen.[1]

C. Samples in which three allelic peaks[2] are observed at a locus, without any other indications of a mixture, may be concluded to be single-source. This conclusion should be based on the relative peak heights of the three peaks and the size range of

[1] Peak height imbalances may be seen in the results from a single individual due to stochastic amplification, elevated stutter, primer binding site variants that result in attenuated amplification of one allele of a heterozygous pair, or tri-allele patterns in which two copies of an allele are present within the genotype.
[2] Observed tri-allele patterns are recorded at http://cstl.nist.gov/biotech/strbase/tri_tab.htm.

the alleles that occur at that locus (i.e., larger loci are less likely to display minor contributor alleles for lesser amounts of DNA template).

D. Single-source samples, including single-source fractions from differentially extracted samples, may display a stutter peak(s) that exceeds the guidelines, generally at only one locus. A peak in a stutter position that exceeds the expected stutter percentage may be interpreted as a stutter peak.

4.5 Interpretation of DNA Typing Results for Mixed Specimens

4.5.1 General Mixture Interpretation Guidelines

Mixtures are generally declared if three or more alleles are present at one or more loci and/or the heterozygous peak height ratios for three or more loci are below expectations. Mixtures can be categorized as distinguishable or indistinguishable. Generally, a multi-locus DNA profile from a mixed sample that contains one or more true minor contributors can be expected to display at least one allelic peak from the minor contributor(s) in a non-stutter position relative to the major contributor.

A. Because locus-specific parameters (e.g., stutter expectations, heterozygous peak height ratios, mutation rate, tri-allelic profile frequencies) may not permit conclusive allelic assignments at a given locus, the classification of any profile as a mixture must be based on an evaluation of the profile in its entirety.

B. The minimum number of contributors to a sample should be estimated, generally by selecting a locus that exhibits the greatest number of allelic peaks. Relative peak heights and possible allelic peak(s) in the stutter position should be considered when determining the minimum number of contributors.

C. For any pair-wise comparison of peak heights, if two peaks fall into different peak height categories, the PHR will be evaluated relative to the category that corresponds to the peak height of the higher intensity peak.

D. For a mixed sample, including mixed fractions from differentially extracted samples, any peak(s) in a stutter position that exceeds its corresponding expected stutter percentage cannot be interpreted as a stutter peak with certainty and must be concluded to be a possible allelic peak(s).

E. A sample for which two or more stutter peaks exceed their corresponding expected stutter percentages is generally considered a mixture.

4.5.2 *Deduced Profiles Determined by Subtraction of Expected DNA Typing Results*

A. Typing results from conditional reference specimens may be subtracted from the other mixture results to facilitate identification of the foreign alleles. For example, in an apparent two-person mixture, a locus has alleles “12, 14, 15, 19” and alleles “12, 14” are attributable to the conditional reference specimen, the extrinsic alleles “15, 19” are attributable to a different individual.

B. If sharing of alleles among the donor of the conditional reference specimen and an additional individual is suspected, any separation of each individual’s alleles must be based on quantitative differences in allelic peak heights at a given locus. For example, given alleles “12, 13, 14” with respective peak heights of 800, 1000, and 200 RFU, although the peak height ratio of alleles “12” and “13” is within expectations (80%), it is possible that the “13” allele is shared by both individuals. If the DNA typing results from the conditional reference specimen are “12, 13” then the possible genotypes for the unknown contributor to this mixture result are either “13, 14” or “14, 14.” Using this assignment strategy, the genotype for the unknown contributor cannot be further refined.

C. The extrinsic alleles may effectively constitute a single-source profile (i.e., there is one DNA contributor in addition to the individual from whom the specimen was taken) or a mixture profile (i.e., there are multiple DNA contributors in addition to the individual from whom the specimen was taken). Regardless, the remaining alleles should be assessed using mixture guidelines for PHR assessments (i.e., use the taller peak to determine the PHR category) and stutter (i.e., consider elevated stutter peaks as possibly allelic).

D. This approach can be used when another known individual can be expected to have contributed biological material to the mixed specimen (e.g., consensual partner). In such situations, the strategies given for the subtraction of the DNA typing results of a single conditional reference specimen from any extrinsic alleles present should be applied for both conditional reference specimens (e.g., victim and consensual partner).

E. This approach can be applied to evidentiary items from which DNA is isolated by means of a differential extraction. The strategies given for the separation of the typing results of a conditional reference specimen(s) from any extrinsic alleles present in an appropriate mixed question specimen may be applied to a mixed result obtained from either the female and/or male fractions. In such situations, the single-source or major contributor typing results from one fraction may be used as a conditional reference specimen(s) to the complementary fraction.

4.5.3 *Deduced Single-Source Profiles Determined from Distinguishable Mixtures*

A. Determination of the genotype of individual contributors may be limited to only some loci.

B. The major contributor can be determined to be heterozygous at a locus if the two alleles of greatest amplitude at a given locus meet PHR expectations and if no pair-wise comparison(s) of the other allelic peak heights at the locus meet PHR expectations.

C. The major contributor can be determined to be homozygous at a locus if the allelic peak that displays the greatest height is not in PHR with any other allelic peak when this peak is considered as both a single allelic dose (i.e., total peak height) and considered as two hypothetical allelic doses with heights at the extremes of the appropriate PHR expectation range.

1. The purpose of this assessment is to consider the tallest peak observed at a locus in terms of two separate amplifications of individual gametic contributions of the same allele. By considering the observed peak in this way, the potential homozygous genotype is effectively considered as simply a special type of heterozygote (i.e., maternal and paternal contributions of a like allele) whose allelic pair displays the minimum allowable PHR. Each dose of this hypothetical pair is then compared to the other allelic peak heights at the locus to determine if either or both are in PHR expectation with any other observed allele.

2. To determine the peak heights of the two hypothetical homozygous doses of the tallest allele detected at the locus, start with the equation:

$$H_{\text{Peak A}} = H_{\text{Observed}} / (1 + PHR_{\text{Minimum}})$$

where $H_{\text{Peak A}}$ is the calculated height of the first hypothetical dose, H_{Observed} is the observed peak height of the tallest allele detected at the locus, and PHR_{Minimum} is the minimum expected PHR value for the half-height of the tallest allele prescribed by Tables 6 and 7. Then use the equation:

$$H_{\text{Peak B}} = H_{\text{Observed}} - H_{\text{Peak A}}$$

where $H_{\text{Peak B}}$ is the calculated height of the second hypothetical dose.

3. For example, given a potential distinguishable mixture result at a locus with the alleles “12, 13, 14, 15, 16” with respective peak heights of 78, 800, 2500, 149, and 200 RFU, while the peak height ratio of allele “14” is not within PHR expectation with any other allelic peak based on its H_{Observed}, this peak must also be evaluated as a pair of individual hypothetical homozygous doses. Using the equations above, allelic peak “14” has a half-height of 1250 RFU

which requires the use of a minimum expected heterozygous peak height ratio of 70% (i.e., 0.70) which yields one dose of peak height:

$$H_{Peak\ A} = H_{Observed} / (1 + PHR_{Minimum}) = 2500\ RFU / (1 + 0.70) = 1470\ RFU$$

and a second dose of height:

$$H_{Peak\ B} = H_{Observed} – H_{Peak\ A} = 2500\ RFU – 1470\ RFU = 1030\ RFU$$

Given this hypothetical homozygous dose pair, because Peak B is in PHR expectation (77%) with the observed “13” allele (the second tallest observed allelic peak), the genotype “14, 14” is not the only possible genotype for the major contributor.

4. The DNA typing results attributed to the minor contributor must also meet PHR expectations in order to be used in a random match probability calculation.

4.6 Application of Statistics

A. The random match probability (RMP)[3,4] may be used to calculate the multi-locus genotype frequency for single-source[5] or deduced single-source[6] (i.e., major contributor, minor contributor, deduced extrinsic contributor) matches. The combined probability of inclusion (CPI) may be used to calculate the multi-locus genotype frequency for a mixture. The CPI is used for indistinguishable mixtures. The CPI may be used to represent the minor contributor(s) in a distinguishable mixture, where the minor contributor(s) cannot be refined to a single genotype.

B. No statistics are calculated for exclusionary or inconclusive conclusions.

C. Each DNA association must be clearly and properly qualified with either a statistic or a qualitative statement. A qualitative statement not based on a statistical calculation should be limited to situations in which the presence of an individual’s DNA on an item is reasonably expected. The provenance of the sample must be established in the case record when statistics are not calculated.

D. Composite DNA profiles for samples may be used for matching/statistical purposes. To reasonably ensure that a profile compiled from genetic information derived from separate extractions, amplifications, and/or injections has arisen from the same individual, the resultant profile must 1) be compiled from different items from a common source (e.g., replicate vaginal swabs), multiple cuttings of the same

[3] The sex typing results from the amelogenin locus are not included in the random match probability calculations.
[4] Any triallelic locus is not included in the statistical calculations, but may be used for exclusionary purposes.
[5] For samples determined to be single-source, loci that do not meet PHR expectations may be used for matching/statistical purposes.
[6] For deduced single-source samples, loci that do not meet PHR expectations may not be used in the RMP calculation.

evidentiary stain, or cuttings from different stains of the same grouping on a given evidence item and 2) demonstrate concordance at the available redundant loci.

E. RMPs and CPIs[7] are always calculated using four general United States population groups (African American, Caucasian, Southeastern Hispanic, and Southwestern Hispanic).

1. Additional RMPs or CPIs are calculated generally based on the geographic location of the collected samples.
    i. Native American (i.e., Apache, Minnesota Native American, and Navajo) or Caribbean (i.e., Trinidadian) population databases are used for specimens that potentially originate from these populations. They are appropriate for use regardless of the specific Native American or Caribbean population group in the case scenario.
    ii. Statistics for cases originating from Puerto Rico do not require the use of the Caribbean population databases.
    iii. Statistics for all cases originating from Alaska will be calculated using the Native Alaskan (i.e., Athabaskan, Inupiaq, and Yupik) population databases. Native American (i.e., Apache, Navajo, Minnesota Native American) population databases are also required if the specimens originate from an Indian Reservation.
    iv. The Chamorro and Filipino population databases are used for cases originating from the U.S. territories of Guam and the Commonwealth of Northern Mariana Islands (e.g., Saipan)

2. The allele frequency distributions for the African American (which includes samples from African American, Bahamian, and Jamaican populations), Caucasian, Southeastern Hispanic, Southwestern Hispanic, Apache, Navajo, Trinidadian, Chamorro, and Filipino populations are published.[8] The allele frequency distributions for the Minnesota Native American population and the Native Alaskan populations (i.e., Athabaskan, Inupiaq, and Yupik) are included in the applicable DNA procedure (i.e., BIO-570).

3. Other published allele frequency distributions for the calculation of RMPs or CPIs may be used in cases potentially involving other defined, dissimilar population groups.

F. The single most common statistic obtained across all populations used in the calculation will be reported.

---

[7] All RMPs and CPIs will be reported as "1 in …" values truncated to two significant figures.
[8] Moretti TR, Moreno LI, Smerick JB, Pignone ML, Hizon R, Buckleton JS, Bright J-A, Onorato AJ. Population data on the expanded CODIS core STR loci for eleven populations of significance for forensic DNA analyses in the United States. Forensic Science International: Genetics (2016) 25: 175-181.

4.7 Suggested Reporting Language

The results and/or conclusions for specimens subjected to DNA analysis will generally be reported in narrative form. The formatting and administrative information required in a Laboratory Report are described in the FBI Laboratory level 1 documents and applicable DNA procedure (i.e., BIO-500). Reporting language should follow the guidance in the appropriate DNA interpretation procedure (i.e., BIO-570) whenever possible. Statistics should be reported as described below.

4.7.1 Inclusion (Match)[9] Conclusions

A. Matches to Single Source Specimens

A match is declared when, upon comparison of a reference specimen to a single-source specimen, the profiles are found to be concordant at all loci for which allelic results were obtained. Single-source match conclusions that fit the following scenarios should be reported as outlined:

RMP (the single most common RMP is reported) :

Male DNA is present in item 1(1). Item 1(1) was interpreted as originating from one individual. The DNA profile from item 1(1) matches SMITH. The random match probability calculated for item 1(1) is approximately 1 in 1.2 million.[A]

The associated endnotes represented by the superscript letters are listed in section C below.

B. Inclusions to Mixed Specimens

A mixture inclusion is declared when, upon comparison of the DNA profile from a reference specimen to a mixed specimen, the genotype of the reference specimen is concordant with one of the possible genotypes present in the mixture. It is noted that one or more of the reference specimen's alleles may not be detectable in the mixture [e.g., due to stochastic loss of an allele(s) or a peak that is below the stutter threshold]. Mixed specimens that fit the following scenarios should be reported as outlined:

Mixture conclusions are generally preceded with a statement regarding the number of potential contributors:

DNA from two or more individuals was obtained from item 1.

[9] As used here, "match," "inclusion," "included," and "cannot be excluded," (as well as any other variations on such language), are used synonymously and are not intended to capture differing degrees of similarity between two DNA profiles or to imply increased/decreased meaning to a particular set of DNA profiles that are found to be indistinguishable.

When the results could have originated from a single individual or a mixture, based on the nature of the results, the number of potential contributors may be expressed as:

DNA from at least one individual was obtained from item 1.

<u>Major/minor contributor RMP</u>:

SMITH is potentially the major contributor of the DNA obtained from item 1(1). The random match probability calculated for the major contributor to item 1(1) is approximately 1 in 7.8 billion.[A]

<u>Combined probability of inclusion</u>:

SMITH and JONES cannot be excluded as potential contributors to this mixture. The combined probability of inclusion calculated for item 1 is approximately 1 in 99,000.[B]

When the CPI estimate results in a frequency which is equivalent to approximately "1 in every 1 individuals", this estimate should be reported as noted above, but also with the percentage (i.e., 67%, rounded to two significant figures)[10] on which the "1 in 1" estimate is based.

1 in 1, which includes approximately 72% of the population.

<u>Unknown major with a CPI on the minor</u>:

A major contributor can be discerned from item 1(1) and is suitable for comparison purposes. SMITH is excluded as a potential major contributor of the DNA recovered from item 1; however, SMITH cannot be excluded as a potential minor contributor to this mixture. The combined probability of inclusion calculated for item 1(1) is approximately 1 in 5,400.[B]

<u>Mixture results in which reference profiles can be subtracted out, the sample is intimate, and the provenance of the sample is established in the case record:</u>

Male and female DNA was obtained from item 1(1). Item 1(1) was interpreted as originating from two individuals, one of whom is JONES. The DNA profile unlike JONES obtained from item 1(1) matches SMITH. The random match probability calculated for the DNA profile unlike JONES obtained from item 1(1) is approximately 1 in 780 billion.[A]

---

[10] The composite frequency (e.g., 6.829E-01) is converted into a percentage, which is rounded to two significant figures. For example, in the traditional manner, 67.49% is rounded down to 67%, and 67.51% is rounded up to 68%. However, when the decimal value is exactly 0.50%, rounding is performed according to the "round half up" rule, whereby, for example, 67.50% is rounded *up* to 68%.

Male and female DNA was obtained from item 1. Item 1 was interpreted as originating from at least three individuals, one of whom is JONES. SMITH is potentially the major contributor of the DNA unlike JONES obtained from item 1. The random match probability calculated for the major contributor to the DNA profile unlike JONES obtained from item 1 is approximately 1 in 98 trillion.[A]

A final statement may be included if the reference samples can or cannot account for all the typing results:

It is noted that SMITH and JONES can/cannot account for all of the DNA typing results obtained from item 1.

C. Associated Endnotes for Inclusionary Reporting Language

[A] *The random match probability is defined as the chance of selecting an unrelated individual at random having an STR profile matching the DNA obtained from the evidence item. The uncertainty associated with a random match probability has been empirically demonstrated to be less than 10-fold in either direction. Calculations were performed using the African American, Caucasian, Southeastern Hispanic, and Southwestern Hispanic populations. The most common random match probability is reported.*

[B] *The combined probability of inclusion is defined as the chance of selecting an unrelated individual at random who could be a potential contributor to the mixture of DNA obtained from the evidence item. Calculations were performed using the African American, Caucasian, Southeastern Hispanic, and Southwestern Hispanic populations. The most common combined probability of inclusion is reported.*

4.7.2 Inconclusive Results

An inconclusive conclusion is declared when the DNA typing results are declared not suitable for matching/statistical purposes (i.e., suitable for exclusionary purposes only), and the comparison does not result in an exclusion. A single-locus or multi-locus DNA typing result can be declared not suitable for matching/statistical purposes when the peak height of an allele(s) falls below the ST. It is noted that inconclusive may also refer to individual locations whose peaks cannot be resolved. If peaks cannot be resolved at a locus, that locus should not be used for comparison purposes.

A. Inconclusive Results in a Single-Source Specimen

1. Loci are deemed not suitable for matching/statistical purposes based on the ST of the amplification kit used and the sample type analyzed.

2. If all loci are declared not suitable for matching/statistical purposes, then the DNA profile may be used for exclusionary purposes only. Any comparison not resulting in an exclusion should be declared inconclusive and reported as follows:

   The STR typing results for item 1 may only be used for exclusionary purposes.[C] *SMITH is excluded as a potential contributor. JONES is inconclusive with regards to the comparison to item 1.*[D]

   Inconclusive results that are consistent with the individual whose DNA is expected to be present should be reported as follows:

   The STR typing results obtained for item 1 may only be used for exclusionary purposes;[C] *however, no DNA unlike SMITH was obtained from item 1.*

   The associated endnotes represented by the superscript letters are listed in section C below.

B. Inconclusive Results in a Mixed Specimen

1. In a distinguishable mixture, the allelic peaks for the major and minor contributor(s) are evaluated separately. For a given potential contributor, a locus cannot be used for matching/statistical purposes if any allele is below the ST. For example, if the mixed specimen exhibits a major contributor type “15, 16” and the minor peaks are “12, (13)” where the “13” allele is <ST, the alleles of the major contributor are interpretable, but the minor contributor is not suitable for matching/statistical purposes at this locus. If all loci attributable to the major or minor contributor are not suitable for matching/statistical purposes, any comparison not resulting in an exclusion is declared inconclusive with regard to the respective contributor. This scenario may be reported as follows:

   DNA from two or more individuals was obtained from item 1. A major contributor can be discerned and is suitable for matching purposes. The STR typing results for the minor contributor(s) to item 1 may only be used for exclusionary purposes.[C] *SMITH is excluded as a potential major contributor; however, SMITH is inconclusive with regards to the comparison to the minor contributor to item 1.*[D]

2. For an indistinguishable mixture, all allelic peaks are considered collectively for the purpose of determining loci that are not suitable for matching/statistical purposes. If any locus contains an allelic peak below the ST, the locus is declared not suitable for matching/statistical purposes. For example, at a given locus, if peaks "(12), 13, 14, 15" are detected with peak "12" <ST, this locus would be deemed not suitable for matching/statistical purposes. If all loci exhibit one or more allelic peaks that are below the ST, then the entire mixture is not suitable for matching/statistical purposes. Any comparison not resulting in an exclusion is declared inconclusive and reported as follows:

   DNA from two or more individuals was obtained from item 1. These results may only be used for exclusionary purposes.[C] *SMITH is inconclusive with regards to the comparison to item 1.*[D]

C. Associated Endnotes for Inconclusive Reporting Language

   [C] *When the potential exists that not all of the genetic information in a biological sample has been detected, the results are not suitable for matching purposes; however, they may be used for exclusionary purposes. For STR typing results to be used for matching purposes, sufficient DNA quality and/or quantity is necessary.*

   [D]*A comparison is inconclusive when the reference sample can be neither included nor excluded as a potential contributor.*

5 CALCULATIONS

The calculations described are generally accomplished with the aid of the FBI Popstats software.

5.1 Calculation of Random Match Probabilities

A. For a heterozygous locus A with alleles Ai and Aj, the single-locus genotype frequency[11] (P_{ij}) is twice the product of the frequencies of occurrence of the allele Ai (p_i) and allele Aj (p_j) of the heterozygous type.

   As an example, for the heterozygous type "22, 24", in which the frequency of occurrence for the "22" allele is 0.2250 and the "24" allele is 0.1861, the single-locus genotype frequency estimate is as follows:

   $$P_{ij} = 2\, p_i\, p_j = (2)(0.2250)(0.1861) = 0.0837$$

[11] The term frequency is used customarily to refer to a relative frequency, or proportion. The lowercase "p" refers to an allelic frequency, and the uppercase "P" refers to a genotypic frequency.

B. For a homozygous locus, the single-locus genotype frequency (P_{ii}) is calculated using the following expression:

$$P_{ii} = p_i^2 + [p_i(1-p_i)\ \Theta]$$

where p_i represents the frequency of occurrence of allele A_i, and the parameter Θ is an estimate of the effects of population subdivision. A Θ value of 0.01 is used for the African American, Caucasian, Southeastern Hispanic, Southwestern Hispanic, Chamorro, Filipino, and Trinidadian populations. A Θ value of 0.03 is used for Native American populations (i.e., Apache, Minnesota Native American, and Navajo) and Native Alaskan populations (i.e., Athabaskan, Inupiaq, and Yupik).

As an example, for the homozygous type "22, 22" in which the frequency of occurrence for the "22" allele is 0.2250, the homozygous expression would yield the following single-locus genotype frequency estimate:

$$P_{ii} = p_i^{\ 2} + [p_i(1-p_i)\ \Theta]$$

$$= (0.2250)^2 + [(0.2250)(1-0.2250)(0.01)] = 0.0524$$

C. A single-locus genotype frequency ($P_{LOCUS(n)}$) obtained for a matching single-source sample can be multiplied by the frequency(ies) found for the same sample at other matching STR loci to obtain a combined multi-locus genotype frequency ($P_{PROFILE}$) estimate as follows:

$$P_{PROFILE} = P_{LOCUS1} \times P_{LOCUS2} \times \ldots . P_{LOCUS(n)}$$

5.2 Calculation of the Combined Probability of Inclusion (CPI)

A. For an indistinguishable mixture, only those loci for which all alleles detected are above the ST may be included in this calculation. However, if all alleles detected at a locus are above the ST, but there is an expectation of dropout (e.g., it is a larger locus, and smaller loci have exhibited alleles below the ST), those loci will be excluded from the calculation.

The frequencies of occurrence (e.g., p_1, p_2, p_3, …, p_n) of all the alleles present at the matching locus (A_n) are summed and then this sum is squared, as follows:

$$P_{LOCUS} = (p_1 + p_2 + \ldots p_n)^2$$

to yield the probability of an unrelated individual having a single-locus genotype that would be observed within the mixture result obtained from that locus (P_{LOCUS}).

B. The single-locus probabilities (e.g., P_{LOCUS1}, P_{LOCUS2}, …, P_{LOCUSn}) for all loci (A_n) for which the alleles are at or above the ST are then multiplied to yield the probability of an unrelated individual having a multi-locus genotype that would be observed within the mixture result obtained from that specimen. This value is the combined probability of inclusion (CPI).

$$CPI = P_{LOCUS1} \times P_{LOCUS2} \ldots \times P_{LOCUS(n)}$$

C. To calculate the combined probability of exclusion (CPE), the CPI is subtracted from 1. The CPE may be multiplied by 100% to express it as the percentage of the population excluded as a donor to the mixture of DNA obtained from the specimen.

$$CPE = (1 - CPI) \times 100\%$$

5.3 Minimum Allele Frequencies

A. A minimum allele frequency ($p_{MINIMUM}$) estimation is substituted for the calculated frequency for rare alleles at an STR locus. For purposes of using a minimum allele frequency, a rare allele at a given locus is defined as one that is found to occur ≤ 4 times in a population database. The use of a minimum allele frequency establishes a lower limit for the frequency of an allele at a locus and ensures that the frequency of occurrence of a rare allele is not underestimated.

B. The minimum allele frequency ($p_{MINIMUM}$) for the rare allele at a locus is established using the following equation:

$$p_{MINIMUM} = 5/2N$$

where N is the number of individuals making up a particular database (i.e., the number of individuals whose DNA types were included in the database for a given locus).

C. The minimum allele frequency will be used for microvariants and other rare alleles that are not included in the Allele Frequency Databases.

6 LIMITATIONS

- These procedures describe the parameters that should be used in conjunction with the appropriate interpretation and reporting procedures of the *DNA Procedures Manual*. Refer to the *DNA Procedures Manual* for additional information regarding the limitations of interpretation.
- It is not possible to anticipate the nature of all potential DNA typing results or the nature of the evidentiary specimens from which they may be obtained. These procedures do not exhaust the possible list of the results that may be encountered by the Examiner, nor the conclusions that an Examiner may render, based on his/her interpretation of those results. For those results not specifically described, conclusions should be drawn using the procedures given for the results above that are similar in concept and/or origin.

7 References

2015 FBI Population Data for the expanded CODIS core STR loci. Available at: https://www.fbi.gov/about-us/lab/biometric-analysis/codis/expanded-fbi-str-2015-final-6-16-15.pdf.

8 Revision History

Revision	Issued	Changes
00	07/01/2022	Reformatted DNA 230-2 into new template and assigned new Doc ID. Added information on population groups based on geographic location.