DNA: Reporting DNA Typing Results

Statistical calculations for STR analysis

When a probative association between an evidence profile and a reference profile is made, a statistic is calculated to give weight to the association. When the evidence and reference samples in a case are analyzed with a current kit and a legacy kit, the comparison between samples will only be performed on the DNA markers common to both. Statistics will be calculated in the method appropriate to the kit that was used to type the evidence samples.

Conclusions

The five types of conclusions for reporting STR analysis are inclusion, exclusion, inconclusive, uninterpretable, and no results.

Calculating Likelihood Ratio (LR) for PowerPlex® Fusion 6C results

The LR provides a statistical measurement of the strength of support for one hypothesis over another. LRs are calculated from the revised NIST 1036 allele frequencies published on the NIST STRbase website (strbase.nist.gov/NISTpop.htm). LRs are calculated for four major U.S. population groups: African-American, Asian, Caucasian, and Hispanic populations.

The LR is represented by the following equation:

$$LR = \frac{\sum_{j} wjPr\,(Sj \mid H1)}{\sum_{u} wuPr\,(Su \mid H2)}$$

where $\sum$ = *sum of values of a set of genotypes (j or u)*
w = *weight of a set of genotypes (j or u)*
Pr = *probability*
S = *set of genotypes (j or u)*
$\mid$ = *given*
H_1 = *the hypothesis of one theory, often the prosecution*
H_2 = *the hypothesis of an opposing theory, often the defense*

Continued on next page

DNA: Reporting DNA Typing Results, Continued

Calculating Likelihood Ratio (LR) for PowerPlex® Fusion 6C results (continued)

The point estimate for the LR is calculated using the population genetic model described by Balding and Nichols in NRC II (Recommendation 4.2):

$$\text{Heterozygotes: } \frac{2[\theta + (1-\theta)pi][\theta + (1-\theta)pj]}{(1+\theta)(1+2\theta)}$$

$$\text{Homozygotes: } \frac{[2\theta + (1-\theta)pi][3\theta + (1-\theta)pi]}{(1+\theta)(1+2\theta)}$$

where $\theta = 0.01$
pi = allele frequency of allele i
pj = allele frequency of allele j

Allele probabilities are assigned as

$$\frac{\text{xi} + \frac{1}{k}}{\text{Na} + 1}$$

where x_i = *number of observations of allele i in a database*
N_a = *number of alleles in the database*
k = number of allele designations with non-zero observations in the database

The case circumstances may indicate that an alternate LR would be appropriate. For example, a brother of the person of interest is alleged to have contributed to the DNA results. The LRs appropriate for relative scenarios are automatically calculated, and will be printed for inclusion in the case file. They may be included in the report under certain case circumstances.

Continued on next page

DNA: Reporting DNA Typing Results, Continued

Conservative calculations – Fusion 6C

The following features were implemented to ensure that the LRs are conservatively calculated:

- Balding and Nichols formulas for the calculation of heterozygous and homozygous genotype probabilities
  - Take into account variation due to subpopulations and do not require Hardy-Weinberg equilibrium or linkage equilibrium.
  - Additionally, the use of theta (θ) accounts for the amount of co-ancestry within a population. Theta is set conservatively at 0.01.
- Factor of N! LR
  - Allows for an individual to be included as any one of the observed contributors in a mixture. There are N! ways to combine the different contributor probabilities in a mixture of N people.
- 99% lower Highest Posterior Density (HPD)
  - Accounts for sample variation within a database.
  - The HPD represents the interval most likely to contain the true value (e.g. LRs).
  - Within the HPD, a point is chosen at the lower end such that 99% of the time, the point estimate LR will be greater than the chosen value.
- 1-sided MCMC uncertainty at the 99th quantile
  - Accounts for the variation in genotype weights between STRmix™ runs by treating the genotype weights as distributions and resampling from the distribution during the HPD calculation.

PowerPlex® Fusion 6C conclusions

Conclusions regarding a person of interest's contribution to a sample will be made based upon the lowest 99% 1-sided lower HPD interval LR returned across the four population groups. Inclusions will be reported as, *"it is at least [lowest LR] times more likely to obtain the DNA results if [H_1 proposition] than if [H_2 proposition]."*

NOTE: On rare occasion, a large discrepancy can be seen between the point LR that is generated and the 99% 1-sided lower HPD interval LR. This is usually due to the difficulty of the HPD modeling when a very rare allele is present. In these cases, the 99% 1-sided lower HPD interval LR will not be reported. With documented DNA Technical Leader approval, the lowest Factor of N! LR returned across the four population groups may be reported

Continued on next page

DNA: Reporting DNA Typing Results, Continued

PowerPlex® Fusion 6C conclusions (continued)

- **Inclusion** – An individual is **included** when the interpretation of the evidence profile returns an LR ≥ 100 in support of an individual being a contributor to a DNA sample.

Statistics will be reported for all inclusions except for associations made between the profile derived from an intimate sample and the individual from whom the sample was collected. Depending on the case scenario, statistics may not be provided for associations made between an individual and items that can reasonably be expected to have his or her DNA present.

- **Exclusion** – An individual is **excluded** when:
  - the interpretation of the evidence profile returns an LR ≤ 0.01 which is support of an individual not being a contributor to a DNA sample
  - the individual is visually excluded from a single source evidence sample

Exclusions will be reported whether the exclusion was based on a visual review or from an LR. The case notes will include how each individual was excluded from each sample.

- **Inconclusive**
  - An individual's contribution is **inconclusive** when the interpretation of the evidence profile returns an LR between 0.01 and 100 upon comparison to the individual's DNA profile.

- **Uninterpretable**
  - A profile is **uninterpretable** when it is unsuitable for interpretation with STRmix™, either due to a low level of DNA, an uncertain number of contributors, or the presence of more than 4 contributors.

- **No results** – A finding of **no results** is reported when non-artifactual fluorescent signal greater than or equal to the detection threshold is not observed.

Conclusions regarding sex of contributor(s)

Conclusions for sex of contributor(s) are based on an assumption of typical DNA results and do not consider mutations such as duplications in Y-STR loci. If atypical data becomes apparent, these conclusions are subject to change.

Continued on next page

DNA: Reporting DNA Typing Results, Continued

PowerPlex® Fusion 6C conclusions (continued)

- For single-source samples:
  - If there is a Y allele at Amelogenin and/or alleles at any of the three Y-STR loci (DYS391, DYS570, and DYS576), then the contributor can be reported as male.
  - If there is not a Y allele or alleles at the Y-STR loci, and the X allele is greater than 450 rfus, the profile can be reported as female. If the X allele is less than 450 rfus and/or there are indications of alleles at the Y-STR loci, then the sex should not be reported.
- For mixture samples:
  - If there is a Y allele at Amelogenin and/or alleles at any of the Y-STR loci that fully explain the assumed number of contributors, the number of males can be reported.
    - For example, a 3-person mixture with 3 alleles at one or more Y-STR loci can be reported as coming from three males.
  - If there is a Y allele at Amelogenin and/or alleles at any of the Y-STR loci that partially explain the assumed number of contributors, then one of the following two options may be pursued:
    - the minimum number of males can be reported
      - For example, a 3-person mixture with 1 allele at one or more Y-STR loci can be reported as coming from at least one male.
    - based on overall quality of the profile and balance among contributors, if the sex determination is obvious for one or more contributors, then the sex determination can be made
      - For example, a 3-person mixture with a mixture proportion of 90:5:5 and a large X peak compared to the Y peak can be reported as coming from at least one female, at least one male.
  - If there is not a Y allele or alleles at the Y-STR loci, then one of the following two options may be pursued:
    - the sex should not be reported
    - based on the overall quality of the profile and balance among contributors, if a female sex determination is obvious for one or more contributors, then the determination can be made.
      - For example, a 2-person mixture with a 95:5 mixture proportion and a large X peak can be reported as coming from at least one female.

The sex of any assumed contributors can be considered when assessing the sex of the remaining contributors.

Continued on next page

DNA: Reporting DNA Typing Results, Continued

Calculating frequencies for legacy kits results

Frequency estimates for evidence typed with legacy kits (Profiler, Cofiler, Identifiler, Identifiler Plus) will use empirical values tabulated from data in the following references:

- Budowle, B., et al., "Population data on the thirteen CODIS core short tandem repeat loci in African Americans, U.S. Caucasians, Hispanics, Bahamians, Jamaicans, and Trinidadians," *Journal of Forensic Sciences*, 1999, 44(6) pp. 1277-1286.
- Budowle, B., "Genotype Profiles for Five Population Groups at the Short Tandem Repeat Loci D2S1338 and D19S433," *Forensic Science Communications*, 2001, 3(3).
- Moretti, T., et al., "Erratum," *Journal of Forensic Sciences*, 2015, 60(4) pp. 1114-1116.

Frequency estimates are calculated for at least three major population groups, generally Caucasian, African American, and Hispanic. Additional population/ethnic groups known to be relevant to the case for which data is available may also be calculated, if deemed appropriate or if requested.

Based on the interpretation of the profile, the analyst may apply one of the following statistical calculations:

- Random Match Probability (RMP)
  - The RMP estimates the probability that a profile from a random person in the population is consistent with the profile from the evidence sample.
- Combined Probability of Inclusion (CPI)
  - The CPI calculation estimates the frequency that a randomly selected person would be included as a possible contributor to an observed mixture.

Continued on next page

DNA: Reporting DNA Typing Results, Continued

Random match probability – legacy kits

Depending on the mixture type, either a "restricted" or an "unrestricted" random match probability may be applied.

- The "restricted" RMP is conditioned on the number of contributors and with consideration of quantitative peak height information and inference of contributor mixture ratios. A "restricted" approach will limit the genotypic combinations of possible contributors.
- The "unrestricted" RMP is also conditioned on the number of contributors, but is performed without consideration of quantitative peak height information or inference of contributor mixture ratios.

Both use the following formulas:

$2pq$	Heterozygote genotype frequency
$p^2 + p(1-p)\theta$	Homozygote genotype frequency
$2p - p^2$	Obligate allele with dropout *('2p')*

where p = the frequency of allele p
q = the frequency of allele q
θ = homozygote correction factor (see *Conservative calculations – legacy kits*)

The appropriate calculation to estimate the frequency of all genotypes that include an obligate allele (with a frequency of p) is *'2p.'* The laboratory will use the expanded *'2p'* formula, $2p - p^2$. If the '*2p*' formula is used more than once at a locus, the frequency of one of the duplicated heterozygote genotypes will be removed.

- When the interpretation for a locus or a profile is inconclusive, that locus or profile will not be used for statistical analysis.
- When the interpretation at a locus includes only one genotype, the appropriate formula above is used to calculate the genotype frequency.
- When the interpretation at a locus includes more than one genotype, the RMP is the sum of the individual frequencies for the genotypes included following mixture interpretation. Adding the frequencies of each genotype provides a frequency of A genotype or B genotype.

The frequencies calculated for all of the individual loci are then multiplied together, using the product rule, to give the estimated probability of the profile as a whole. If a number greater than one is generated by adding the individual frequencies together, round the number to 1.0.

Continued on next page

DNA: Reporting DNA Typing Results, Continued

Combined Probability of Inclusion (CPI) – legacy kits

CPI is applied to mixture profiles where the contributions of individual donors cannot be resolved. There are no assumptions about the number of contributors when using CPI. Loci with alleles below the stochastic threshold may not be used for statistical purposes.

The probability of inclusion (PI) calculation provides an estimate, at a locus, of the portion of the population that has a genotype that is represented in the mixed profile, and therefore would be included as a possible contributor to the mixed profile.

• Example: If evidence includes three alleles (A_1, A_2, A_3) at a locus, then:

$$PI = (a_1 + a_2 + a_3)^2$$

$$PI = (a_1\,a_1 + a_2\,a_2 + a_3\,a_3 + 2a_1\,a_2 + 2a_1\,a_3 + 2a_2\,a_3)$$

In this example, the only genotypes that would be included in the PI calculation and as possible contributors to the mixture would be:

- A_1A_1
- A_2A_2
- A_3A_3
- A_1A_2
- A_1A_3
- A_2A_3

PI at each locus is first determined and then PIs from all of the included loci are multiplied together, using the product rule, to give the combined probability of inclusion (CPI).

A locus that contains a single allele will not be included in a PI calculation if allele dropout is suspected. If the analyst concludes that there is no allele dropout, and all contributors are represented by the single allele (all homozygous), then the allele can be included in a PI calculation.

Continued on next page

DNA: Reporting DNA Typing Results, Continued

Conservative calculations – legacy kits

The following concepts were implemented to ensure that the frequency estimates are conservatively calculated:

- Correction factor for homozygotes
  - To account for non-random mating, θ is applied to the calculation for a homozygote genotype. Empirical studies have shown that a conservative value for θ is 0.01. A θ value of 0.03 is applied when calculating a homozygote genotype for an isolated sub-population such as a Native American population.
- Five-event minimum allele frequency
  - A five-event minimum allele frequency is used for rare alleles. For each individual allele, an observed allele count less than five is raised to five. This modified allele count is converted to a frequency and used for all subsequent genotype calculations.

Legacy kit conclusions

- **Inclusion** – An individual is **included** when:
  - Alleles in the reference profile are detected in an evidence profile.
  - Alleles in the reference profile are included as being a reasonable genotype identified from the true contributor.
    - An individual may still be included even if some of the loci were determined to be inconclusive.

Statistics will be reported for all inclusions except for associations made between the profile derived from an intimate sample and the individual from whom the sample was collected. Depending on the case scenario, statistics may not be provided for associations made between an individual and items that can reasonably be expected to have his or her DNA present.

- **Exclusion** – An individual is **excluded** when:
  - Alleles in the reference profile are not detected in an evidence profile and there is no scientific explanation for the non-match.
  - Alleles in the reference profile are not included as being a reasonable genotype identified from the true contributor and there is no scientific explanation for the non-match.
  - The statistical result obtained for an evidence profile is 1 in 1 and the alleles in the reference profile are not detected in the evidence profile.

Continued on next page

DNA: Reporting DNA Typing Results, Continued

Legacy kit conclusions (continued)

- **Inconclusive** – A profile is **inconclusive** when none of the loci are suitable for comparison to reference profiles. This could be due to a low level of DNA or an uncertain number of contributors.

- **No results** – A finding of **no results** is reported when non-artifactual fluorescent signal greater than or equal to the detection threshold is not observed.

Generating paternity statistics

The CODIS *Popstats* software will be used to generate the Combined Paternity Index (CPI) and Probability of Paternity (PP) from the revised NIST 1036 allele frequencies on the NIST STRbase website (strbase.nist.gov/NISTpop.htm). CPI and PP are calculated for four major U.S. population groups: African-American, Asian, Caucasian, and Hispanic populations. The lowest CPI returned across the four population groups will be reported.

Popstats compares the child and alleged parent profiles at each locus and automatically computes the following values:

- *Parentage Index*
- *Probability of Exclusion*
- *Probability of Parentage*

This laboratory will only be entering the mother as the 'known parent' and the alleged father as the 'alleged parent.' Therefore, the *Parentage Index* generated will be the reported *Paternity Index* and the *Probability of Parentage* will be the reported *Probability of Paternity*. *Probability of Exclusion* will not be reported.

In order to perform parentage calculations using *Popstats*, three profiles must be entered:

- mother
- child, and
- suspected father

Continued on next page

DNA: Reporting DNA Typing Results, Continued

Generating paternity statistics (continued)

The following procedure is used to obtain results.

Step	Action
1	Log on to CODIS and open the *Analyst Workbench* program. Select **Popstats**.
2	Select **Parentage Calculation**.
3	Enter the profiles of • Biological parent (mother) • Child/product of Conception • Alleged parent (suspected father)
4	Click on **Calculate** and the window will display the *Parentage Statistics*.

The **Paternity Index (PI)** is a likelihood ratio based on two conditional probabilities:

$$PI_{locus} = \frac{P\text{ (that the alleged father passed an allele to the child)}}{P\text{ (a randomly selected man passed an allele to the child)}}$$

The *Paternity Index* reflects how many more times likely it is to observe a particular set of alleles under the hypothesis that the alleged father is the biological father compared to the hypothesis that a randomly selected man is the biological father. The *Paternity Index* is based on the assumption that the randomly selected man has a similar ethnic background to the alleged father.

Formula tables for the *Paternity Index* are listed in the Parentage *Formula Table* in the *Popstats* software. The exact formula for the *Paternity Index* depends on the obligate paternal allele and the genotype of the alleged father. Obligate paternal alleles are alleles that the biological father is required to have based on the relationship between the mother and the child.

The **Combined Paternity Index (CPI)** for a DNA profile is calculated using the Product Rule by multiplying the individual PIs for each locus tested.

Continued on next page

DNA: Reporting DNA Typing Results, Continued

Generating paternity statistics (continued)

The **Probability of Paternity (PP)** is based upon Bayes Theorem. The probability of paternity tests the hypothesis that the alleged father is the biological father by incorporating a prior probability that the alleged father is the true biological father.

$$PP = \frac{\text{CPI (prior probability)}}{\text{CPI (prior probability)} + [1-(\text{prior probability})]}$$

The laboratory uses a prior probability set to a neutral value of 0.5 which simplifies the formula to:

$$PP = \frac{CPI}{(1+CPI)}$$

For example, a probability of paternity of 99% reflects a 99% probability that the hypothesis that the alleged father is the biological father is correct and a 1% probability that this hypothesis is incorrect.

Paternal mutation rates and mean power of exclusion
In cases where there is a paternal mutation, *Popstats* requires a mutation rate and mean power of exclusion to be entered for that locus. The current mutations rates and mean powers of exclusion for each population group can be found in the AABB Annual Report Summary for Testing.

Paternity conclusions

- **Inclusion** – An individual is **included** when:
  - No more than three obligate paternal allele mismatches are identified after comparing the reference profile of the child to the reference profile of the alleged father. The maximum of three mismatches for an inclusion applies exclusively to the interpretation of paternity samples.

- **Exclusion** – An individual is **excluded** when:
  - Four or more obligate paternal allele mismatches are identified after comparing the reference profile of the child to the reference profile of the alleged father. The requirement of four or more mismatches for an exclusion applies exclusively to the interpretation of paternity profiles.

Continued on next page

DNA: Reporting DNA Typing Results, Continued

Haplotype statistics for Y-STR analysis

The *Y Chromosome Haplotype Reference Database* (www.yhrd.org) consisting of anonymous Y-STR profiles from various population/ethnic groups will be used for reporting the significance of a Y-STR inclusion. The website also has the statistical formulas that are used to calculate frequency estimates.

The search of the database provides the number of times a specific haplotype is observed in the database. The basis for the haplotype frequency estimation is the counting method. The application of the upper bound of a confidence interval corrects for sampling variation uncertainty. Typically, upper bound frequency estimates for African American, Asian, Caucasian, and Hispanic populations will be used for reporting purposes.

The following procedure is used to calculate haplotype statistics.

Step	Action
1	Go to yhrd.org. Click **Search Database**.
2	Select whether you will be searching using an export file or will be manually entering the haplotype.
3	Select **PowerPlex Y23** and enter the alleles. Click **Search**.
4	Click **+Add feature to this Report** and select **National Database (with Subpopulations, 2014 SWGDAM-compliant)**. Click the X in the *Worldwide* box to close it.
5	Review the results. Right click and select **Print** to print the statistical results. Report the appropriate frequency results from the *Observed* section.

Continued on next page

DNA: Reporting DNA Typing Results, Continued

Y-STR conclusions

- **Inclusion** – An individual is **included** when:
  - The reference haplotype is the same as a single-source evidence haplotype.
  - The reference haplotype is consistent with a deduced evidence haplotype.
    - An individual may still be included even if some of the loci were determined to be inconclusive.

Statistics will be reported for all inclusions except for associations made between the profile derived from an intimate sample and the individual from whom the sample was collected. Depending on the case scenario, statistics may not be provided for associations made between an individual and items that can reasonably be expected to have his DNA present.

NOTE: Patrilineal male relatives will have the same haplotype (barring genetic mutations). The possibility that a close relative of the suspect is a potential contributor to an evidence haplotype should be considered.

Reports for Y23 inclusions of single source profiles or a contributor deduced from a mixture should indicate that the statistics were arrived at using the online YHRD database calculations.

- **Exclusion** – An individual is **excluded** when:
  - The reference haplotype is not the same as (single-source) or consistent with (deduced) the evidence haplotype and there is no scientific explanation for the non-match.
  - A single-source haplotype cannot be deduced from a mixture, but alleles in the reference profile are not included.

- **Uninterpretable** – A profile or component is **uninterpretable** when none of the loci are suitable for comparison to reference profiles.

- **No results** – A finding of **no results** is reported when non-artifactual fluorescent signal greater than or equal to the detection threshold is not observed.

Reporting Y-STR DNA results

The DNA typing results will be recorded on the PowerPlex® Y23 Results Table found in *DNA Manager*, and a printed copy will be included in the case file as a notes page. A table of the results will not be included in the report.