

# FORENSIC BIOLOGY PROTOCOLS FOR FORENSIC STR ANALYSIS

POPULATION FREQUENCIES FOR STR's		
DATE EFFECTIVE 06-20-2016	APPROVED BY NUCLEAR DNA TECHNICAL LEADER	PAGE 1 OF 5

## Population Frequencies for STR's

To interpret the significance of a match between genetically typed samples, it is necessary to know the population distribution of alleles at the loci that were typed. If the STR alleles of the relevant evidence sample are different from the alleles of a subject's reference sample, then the subject is "excluded," and cannot be the donor of the biological evidence being tested. An exclusion is independent of the frequency of the alleles in the population.

If the subject and evidence samples have the same alleles, then the subject is "included", or is a "match", and could be the source of the evidence sample. The random match probability, or the probability that another, unrelated, individual would also match the evidence sample, is equal to the frequency of the evidence profile genotypes in the relevant population. Population frequencies are estimated separately for the Asian, Black, Caucasian and Hispanic populations. Additional population frequencies may be used for other population groups. If a source contains more than one frequency for a single population group, then the highest frequency is used for calculations. Allele frequencies are used for all calculations. Profile frequency estimates are calculated according to the National Research Council report entitled *The Evaluation of Forensic DNA Evidence* (National Academy Press 1996, pp. 4-36 to 4-37).

Spreadsheets are used to automate the calculation of the population specific genotype and profile frequency estimates. The spreadsheets are located in the "POPSTATS" subdirectory on the network and explanations for their use are included with the spreadsheets.

STR population frequency estimates are based on the OCME STR database, and the Population Data in the AmpF/STR® Identifiler™ PCR Amplification Kit User's Manual (2001) Population Data, Applied Biosystems, Foster City, California.

### I. Random Match Probability for Autosomal STRs

- A. Enter the evidence profile alleles in the Identifiler worksheet of the POPSTATS spreadsheet. Off-ladder alleles can be entered as decimals (for example, "12.2") or as ">" or "<" for values above or below the ladder, respectively.
- B. For loci assigned a "Z" to indicate the possible presence of another allele, only one allele is entered in the calculation spreadsheet. In this manner, the locus is not treated as a true homozygote whose statistical values are determined by squaring the allele frequency ( $p^2$ ). Rather "Z" loci utilize the probability only of the one assigned allele ( $2p$ ), which allows the second allele to be anything.

# FORENSIC BIOLOGY PROTOCOLS FOR FORENSIC STR ANALYSIS

POPULATION FREQUENCIES FOR STR's		
DATE EFFECTIVE 06-20-2016	APPROVED BY NUCLEAR DNA TECHNICAL LEADER	PAGE 2 OF 5

- C. The overall profile frequency estimate for each group is calculated by multiplying the individual locus genotype frequency estimates together.
- D. In the standard scenario, homozygote genotype frequencies are estimated for each population using the formula  $p^2+p(1-p)\theta$  for  $\theta = 0.03$  and heterozygote genotype frequencies are estimated using the formula  $2p_i p_j$ .
- E. Genotype and profile frequencies are also estimated for isolated populations (i.e., “evidence and subject from the same subgroup (isolated village)”) and for relatives using the formulas in the National Research Council Report.
- F. For each population, the overall profile frequency estimate under the standard scenario of  $\theta = 0.03$  unless there is reason to suspect that the “evidence DNA and subject are from the same subgroup” or a relative of the subject left the biological sample.
- G. Calculations and allele frequencies are retained in the case file for referral at a later date if necessary.

## II. Frequency for Y STRs

- A. The frequency for a Y STR haplotype is estimated by counting the number of times the haplotype occurs in each of the population databases and dividing by the total number of individuals in the database.
  - 1. A haplotype that has not been previously observed in the Asian database, which includes 196 individuals, would be reported as “less than 1 in 196 Asians”.
  - 2. A haplotype that has been observed once in the Asian database would be reported as “1 in 196 Asians”.
  - 3. A haplotype that has been observed 5 times in the Asian database is reported as “1 in 39 Asians” (5 in 196 is equal to 1 in 39).

# FORENSIC BIOLOGY PROTOCOLS FOR FORENSIC STR ANALYSIS

POPULATION FREQUENCIES FOR STR's		
DATE EFFECTIVE 06-20-2016	APPROVED BY NUCLEAR DNA TECHNICAL LEADER	PAGE 3 OF 5

- B. For Y-STR haplotypes, use the US Y-STR database to estimate haplotype frequencies.
- Using Internet Explorer, navigate to [www.usystrdatabase.org](http://www.usystrdatabase.org)
  - Enter the Y-STR alleles from the profile into the drop-down boxes on the screen.
  - To specify a value not listed in the drop-down box, enter the value in the text box next to the drop-down box.
  - The following value types are allowed:
    - Standard ladder allele such as "12"
    - Off-ladder allele value such as "12.2"
    - Off-ladder low- or high-value such as "<15" or ">21"
    - Null allele: enter "0" if the sample is believed to contain a legitimate null allele, for example, due to a primer binding site mutation.
    - No data: "\*" is the default value. Loci with \* are treated as wild cards.
  - In the "Search By Ancestry" box, select "All".
  - Click "Search"
  - Scroll down for the results. The website reports the number of times the haplotype was observed in the database, the observed frequency of the haplotype, and the upper bound of the 95% confidence interval. These values are reported for each of the populations in the database (African American, Asian, Caucasian, Hispanic, and Native American) and for all of the populations combined.
  - Adjust the margins of the page by selecting "Page Setup" from the printer menu at the top of the page and changing the top and bottom margins to 0.5, then choosing "OK".
  - Print the screen by selecting "Print" from the printer menu at the top of the page and selecting a printer.
  - Verify on the printout that the Y-haplotype alleles were correctly entered into the website.

# FORENSIC BIOLOGY PROTOCOLS FOR FORENSIC STR ANALYSIS

## POPULATION FREQUENCIES FOR STR's

DATE EFFECTIVE 06-20-2016	APPROVED BY NUCLEAR DNA TECHNICAL LEADER	PAGE 4 OF 5
------------------------------	---	----------------

11. Report the 95% upper-bound confidence statistic from all ethnic groups, and round down to three significant figures.
12. If both autosomal and Y-STRs are typed, the results are reported separately.

### III. Combined Probability of Inclusion (CPI) for Mixtures

The combined probability of inclusion (CPI) is defined as the probability that a randomly selected individual would be a contributor to a mixture of labeled DNA alleles. In other words, it is the expected frequency of individuals who could be included as potential contributors to the mixture because all of their alleles are labeled in the evidence profile.

CPI can only be used if all of the following circumstances are met:

- When the evidence sample contains a non-deducible mixture.
- When the alleles of the associated known sample are labeled at all of the conclusive loci in the evidence sample.

A conclusive locus is a locus with concordant or repeating alleles. If an evidentiary sample is amplified more than once, loci with concordant alleles (HT-DNA samples) or repeating alleles (LT-DNA samples) are determined. Loci that are designated as “NEG” (for negative) or “INC” (for inconclusive) are not used in the CPI calculation. To avoid the possibility of bias, the determination to deem a locus inconclusive in the evidence profile must be made prior to viewing the comparison sample profile.

Although CPI can be calculated at any point, for efficiency and workflow, CPI is calculated (if necessary) after the DNA profile of the comparison sample(s) is determined to be included in the evidence sample. The CPI is calculated for informative samples. If RMP values have been generated, the CPI may not need to be calculated. The CPI is reported in the evidence report.

The comparison is based on the previously determined allele calls. If any of the alleles of a comparison sample are missing from the evidence profile at conclusive loci, CPI is not appropriate.

# FORENSIC BIOLOGY PROTOCOLS FOR FORENSIC STR ANALYSIS

POPULATION FREQUENCIES FOR STR's		
DATE EFFECTIVE 06-20-2016	APPROVED BY NUCLEAR DNA TECHNICAL LEADER	PAGE 5 OF 5

## A. Computing CPI

1. Open CPI worksheet named "CPI.xls"
2. In cells A9 through P9 of the Data Entry worksheet, enter each allele that is labeled in the evidence profile at conclusive loci, up to 10 alleles per locus. Alleles should be separated by commas and/or spaces. A profile from a PG sheet may be pasted into cells A9 through P9. **All alleles that are labeled at conclusive loci in all amplifications must be entered.**
3. Press the blue "Run CPI macro" button. The CPI for the Black, Caucasian, Hispanic, and Asian populations appears at the bottom of the Results worksheet.
4. Print the results by selecting File > Print while in the Results worksheet. The printout will include the alleles entered and the results.

### Note:

Off-ladder alleles may be entered in either 15.x format or as "<" or ">". 5/2N will be used as the frequency for an off-ladder allele.

## B. Interpretation

Results are presented for each of the four populations: Black, Caucasian, Hispanic, and Asian. The probability of inclusion is stated in the report.

Combined Probability of Inclusion is the expected frequency of individuals who are carrying only alleles that are labeled in the mixture in question, and if tested could potentially be included as contributors to this mixture. It is the expected frequency of individuals who could be included as potential contributors to the mixture because they do not carry any alleles that are not labeled in the evidence profile.