

FORENSIC BIOLOGY PROTOCOLS FOR FORENSIC STR ANALYSIS

Population Frequencies for STR's		
Status: Published		Document ID: 1151
DATE EFFECTIVE 11/05/2020	APPROVED BY Nuclear DNA Technical Leader	PAGE 1 OF 3

1 Population Frequencies for STR's

- 1.1 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.
- 1.2 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. Multiple statistics can be calculated.
 - 1.2.1 The profile frequency calculates the rarity of the observed profile in a particular population group. This can also be thought of as the probability that the DNA in a random sample chosen from a population of unrelated individuals would have the same profile as the DNA in the evidence sample, and is equal to the frequency of the evidence profile genotypes in the relevant population. Profile frequency estimates are calculated according to the National Research Council (NRC II) report entitled *The Evaluation of Forensic DNA Evidence* (National Academy Press 1996, pp. 4-36 to 4-37), following recommendation 4.1.
 - 1.2.2 The match probability, described by the sampling formulae, conditions the probability for the observed profile within the evidence in a particular population group given that it has already been observed at least once (i.e., matches an individual in question). This method is described in NRC II under recommendation 4.2.
 - 1.2.3 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.
- 1.3 Spreadsheets may be used to automate the calculation of the population specific genotype, profile frequency estimates and match probabilities. The spreadsheets, or links to the spreadsheets, are located in Qualtrax.
- 1.4 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, or the NIST 1036 Revised U.S. Population Dataset (July 2017). Y-STR haplotype frequency estimates are based on the US Y-STR database (release 4.2 February 18, 2017).
- 1.5 If both autosomal and Y-STRs statistics are calculated, the results are reported separately.

Controlled versions of Department of Forensic Biology Manuals only exist in the Forensic Biology Qualtrax software. All printed versions are non-controlled copies.

FORENSIC BIOLOGY PROTOCOLS FOR FORENSIC STR ANALYSIS

Population Frequencies for STR's		
Status: Published		Document ID: 1151
DATE EFFECTIVE 11/05/2020	APPROVED BY Nuclear DNA Technical Leader	PAGE 2 OF 3

2 Match Statistics for Autosomal STRs

- 2.1 For a match to an Identifier, Cofiler and/or Profiler Plus evidence profile, enter the evidence profile alleles in the worksheet of the POPSTATS spreadsheet. For Fusion body identification cases (ex. match between a remain to remain or a remain and a reference sample), the CODIS software POPSTATS module is used.
- 2.2 Off-ladder alleles are entered as decimals (for example, "12.2") or as ">" or "<" for values above or below the ladder, respectively.
- 2.3 For loci assigned a "Z" to indicate the possible presence of another allele, only one allele is entered in the calculation. 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.
- 2.4 If the comparison sample that matches an evidence profile has less loci with typing results than the evidence profile, a statistic sheet shall be filled out for the complete profile determined from the evidence, as well as a sheet capturing only the overlapping loci between the two profiles. The profile frequency calculated for the overlapping loci will be used for reporting.
- 2.5 The overall profile frequency estimate for each group is calculated by multiplying the individual locus genotype frequency estimates together.
- 2.6 In the profile frequency estimate, 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 $2p_i p_j$.
- 2.7 Within the POPSTATS spreadsheet, statistics are also calculated for match probabilities and for relatives using the formulas in the National Research Council Report (NRC II) Recommendations 4.2 and 4.4.
- 2.8 For each population, a $\theta = 0.03$ is applied within the formulas.
- 2.9 Calculations are retained in the case file for referral at a later date if necessary.
- 2.10 The most common profile frequency of the four population groups is reported to three significant figures, truncated. If the profile frequency is less than 1 in 6.80 trillion, the profile frequency is reported as 1 in greater than 6.80 trillion people.

FORENSIC BIOLOGY PROTOCOLS FOR FORENSIC STR ANALYSIS

Population Frequencies for STR's		
Status: Published		Document ID: 1151
DATE EFFECTIVE 11/05/2020	APPROVED BY Nuclear DNA Technical Leader	PAGE 3 OF 3

3 Frequency for Y STRs

- 3.1 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.
- 3.1.1 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”.
- 3.1.2 A haplotype that has been observed once in the Asian database would be reported as “1 in 196 Asians”.
- 3.1.3 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).
- 3.1.4 The 95% upper-bound confidence statistic from all ethnic groups is reported, truncated to three significant figures.
- 3.2 To calculate a statistic for Y-STR haplotypes or Y-STR mixtures, use the [Y-Mix Database Filter](#). Refer to the protocol [Usage of the ‘Y-Mix Database Filter’](#) for instructions.