

User Directions

To search for a haplotype, enter the alleles of interest for each marker by using the drop-down menu. To specify a value not listed in the drop-down menu, use the corresponding text boxes. Enter any of the following value types:

Standard ladder allele such as "14"

Off-ladder allele value such as "17.2"

Off-ladder-low or off-ladder-high values such as "<15" or ">17"

Null allele: enter "0" as the allele designation when the sample is believed to contain a legitimate null allele type (due to mutation).

Enter "*" to find matches for any allele at a particular locus (default). To search for only one allele at DYS385, manually enter the desired allele followed by a comma and an "*", e.g. "11,*" to return all haplotypes containing an 11 at this locus.

If only one allele is detected, please enter that allele in the first column only. Do not enter the same allele designation into multiple columns for any locus.

If multiple alleles are to be entered, please select the smallest allele in the first box for the marker of interest or manually enter the alleles in numerical order separated by commas, such as "12,13,14".

Select the ancestry or geographical location of interest from the appropriate boxes. Multiple selections can be made by holding down the Ctrl key. A list of the ancestral groups within the database is provided in the Database Descriptive Statistics section.

Select the "search" button to query the database for the chosen haplotype.

Frequency and Confidence Interval Calculations

The basis for the haplotype frequency estimation is the counting method. The application of a confidence interval corrects for database size and sampling variation.

If the haplotype has not been previously observed in the database, the formula used for calculating the upper 95% confidence limit is:

$$1 - (0.05)^{1/n} \quad n = \text{database size}$$

If the haplotype has been observed in the database, the formula used for calculating the 95% confidence limit is that of an exact confidence interval (Clopper and Pearson; Weir and Buckleton). Let the number of samples in the database be n and the number of matches x . This gives a sample proportion p .

Clopper and Pearson seeks a value p , such that

$$\sum_{k=0}^x \binom{n}{k} p_0^k (1 - p_0)^{n-k} = 0.05$$

References:

1. Holland, M.M. and T.J. Parsons, *Mitochondrial DNA Sequence Analysis – Validation and Use for Forensic Casework*. Forensic Science Review (1999). **11**(1): p. 21-50.
2. Clopper, C.J. and E.S. Pearson, *The use of confidence or fiducial intervals illustrated in the case of the binomial*. Biometrika (1934). **26**: p. 404-413.
3. Weir, B. and J. Buckleton, Personal Communication, (2010).