



Anexo I: Soporte estadístico sobre la valoración estadística de cromosoma Y (aprobado en la reunión del Pleno de la CNUFADN de 11 de diciembre de 2023)

La estimación de la frecuencia de un perfil de cromosoma Y es diferente a la que se realiza en marcadores autosómicos, dada la práctica ausencia de recombinación en el cromosoma Y, lo que hace inviable el uso de la regla del producto para el cálculo de la frecuencia de un haplotipo dentro de una población. Es decir, la frecuencia de un haplotipo de cromosoma Y debe ser estimada a partir de una base de datos poblacional que refleje lo más adecuadamente posible la realidad de nuestra población.

Establecido como consenso que la base de datos a usar sea YHRD (Y-Chromosome STR Haplotype Reference Database; <https://yhrd.org/>) y que la población de referencia a usar sea la metapoblación de Europa occidental, tenemos varias opciones, que pasamos a detallar, para la estimación de la frecuencia de un haplotipo.

Método de conteo (“counting method”)

Consiste en determinar el número de veces (n) que aparece un determinado haplotipo en una base de datos que tenga un tamaño N . La frecuencia del haplotipo será:

$$p = \frac{n}{N}$$

Obviamente nos interesará que el tamaño de la base de datos (N) sea lo más grande posible, pero incluso en esta circunstancia nos encontraremos con la ausencia en la base de datos de muchos de los haplotipos buscados con lo que la frecuencia estimada de ese haplotipo sería 0.

A fin de evitar esto se suele usar una modificación de este método.

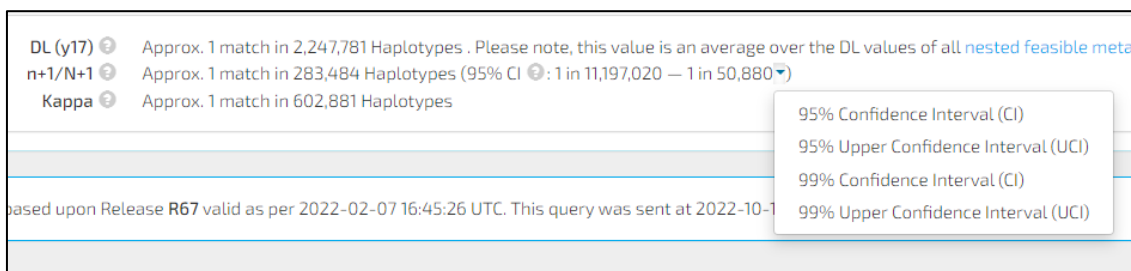
Método de conteo aumentado (“augmented counting method”)

Consiste en añadir el haplotipo buscado tanto en el numerador como en el denominador de la ecuación, lo que se denomina corrección de Balding y Nichols [1]. Esto es lo recomendado por la Comisión de ADN de la ISFG [2]. Con ello conseguimos que n sea mayor de 0 y, por tanto, la probabilidad del haplotipo sea distinta de 0:

$$p = \frac{n + 1}{N + 1}$$

Tenemos varias opciones para ajustar más adecuadamente (desde un punto de vista estadístico) y de una forma más conservadora (si $n=0$ la frecuencia del haplotipo pasa a ser de $1/N$ a aproximadamente $3/N$ [3]) el valor este valor y una de ellas es el uso de intervalos de confianza (normalmente 95%). Uno de los métodos más utilizados para el cálculo de intervalos de confianza en distribuciones binomiales es el llamado método exacto de Clopper y Pearson [4].

En YHRD podemos utilizar diferentes intervalos de confianza:



The screenshot shows a table with three rows of statistics:

DL (y17)	Approx. 1 match in 2,247,781 Haplotypes . Please note, this value is an average over the DL values of all nested feasible meta
n+1/N+1	Approx. 1 match in 283,484 Haplotypes (95% CI : 1 in 11,197,020 — 1 in 50,880)
Kappa	Approx. 1 match in 602,881 Haplotypes

Below the table, there is a dropdown menu with the following options:

- 95% Confidence Interval (CI)
- 95% Upper Confidence Interval (UCI)
- 99% Confidence Interval (CI)
- 99% Upper Confidence Interval (UCI)

At the bottom of the screenshot, there is a timestamp: "based upon Release R67 valid as per 2022-02-07 16:45:26 UTC. This query was sent at 2022-10-1"

El método de conteo aumentado presenta como ventaja que es de fácil explicación en el informe pericial (o ante los Tribunales de Justicia) pero presenta como inconvenientes principales que dicho método es muy conservador y depende del tamaño de la base de datos.

Este es el método recomendado por UK Forensic Regulator (<https://www.gov.uk/government/publications/y-str-profiling>).

Método kappa

Este método tiene en cuenta la proporción de haplotipos únicos (k) (singletons) existentes en la base de datos [5].

$$p = \frac{(n + 1)(1 - k)}{N + 1}$$

$$k = \frac{\alpha}{N}$$

k = Proporción de haplotipos únicos en la base de datos

a = Número de haplotipos únicos (singletons) en la base de datos

Si k es cercano al 100% este método no es aplicable.

Este es el método recomendado a usar por el grupo polaco de la ISFG [6]

Método discreto de Laplace

Este método [7] tiene en consideración que los diferentes haplotipos de cromosomas Y existentes en la población descienden de un reducido número de ascendientes y que si nos remontáramos a unos cientos de generaciones atrás podríamos llegar hasta un número manejable (computacionalmente) de ancestros (clusters). Obviamente, los descendientes de un ancestro tendrán el mismo cromosoma Y (salvo mutación, la cual puede ocurrir independientemente en cada locus).

A partir del perfil de cromosoma Y de un ancestro, la probabilidad del perfil de un descendiente viene dado por una distribución discreta de Laplace considerando el modelo evolutivo de Fisher-Wright (evolución neutral, tamaño poblacional finito, generaciones no superpuestas, no mutación, no migración, apareamiento aleatorio, misma tasa de reproducción en todos los varones...) y el modelo mutacional de Ohta-Kimura (stepwise mutation model) [8].

Es necesario estimar el número de ancestros (clusters), el porcentaje de la población que desciende de cada ancestro y los haplotipos de los ancestros considerando tasas de mutación, la varianza del suceso reproductivo, la tasa de crecimiento poblacional, el número de generaciones desde el ancestro... Todo ello significa que es un modelo de gran exigencia computacional por lo que únicamente es aplicable al haplotipo mínimo y al kit Yfiler en YHRD (bajas tasas de mutación y número de dimensiones manejables computacionalmente). Además, no es aplicable a haplotipos que incluyan alelos nulos, alelos intermedios y duplicaciones, tampoco es aplicable a perfiles parciales (muestras degradadas...) y únicamente es aplicable al considerar metapoblaciones.

Este es el método recomendado a usar por parte de la ISFG [2], Alemania [9] y Filipinas [10].

Bibliografía

- [1] DJ Balding, RA Nichols. DNA profile match probability calculation: how to allow for population stratification, relatedness, database selection and single bands. *Forensic Science International* 64(2-3): 125-140 (1994).
- [2] L Roewer, MM Andersen, J Ballantyne, JM Butler, A Caliebe, D Corach, ME D'Amato, L Gusmão, Y Hou, P de Knijff, W Parson, M Prinz, PM Schneider, D Taylor, M Vennemann, S Willuweit. DNA commission of the International Society of Forensic Genetics (ISFG): Recommendations on the interpretation of Y-STR results in forensic analysis. *Forensic Science International Genetics* 48: 102308 (2020).
- [3] MM Andersen, DJ Balding. Assessing the Forensic Value of DNA Evidence from Y Chromosomes and Mitogenomes. *Genes* 12(8): 1209 (2021).
- [4] CJ Clopper, ES Pearson. The Use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial. *Biometrika*, 26, 404-413 (1934).
- [5] CH Brenner. Fundamental problem of forensic mathematics – The evidential value of a rare haplotype. *Forensic Science International Genetics* 4: 281-291 (2010).
- [6] K Rębała, W Branicki, R Pawłowski, M Spólnicka, T Kupiec, A Parys-Proszek, M Woźniak, T Grzybowski, M Boroń, M Wróbel, M Ciesielka, A Ossowski, R Jacewicz. Recommendations of the Polish Speaking Working Group of the International Society for

Forensic Genetics on forensic Y chromosome typing. *Archiwum Medycyny Sądowej i Kryminologii* 70 (1): 1–18 (2020).

- [7] MM Andersen, PS Eriksen, N Morling. The discrete Laplace exponential family and estimation of Y-STR haplotype frequencies. *Journal of Theoretical Biology* 329: 39-51 (2013).
- [8] M Kimura, T Ohta. Stepwise mutation model and distribution of allelic frequencies in a finite population. *Proceedings of the National Academy of Sciences USA* 75(6): 2868-2872 (1978).
- [9] S Willuweit, K Anslinger, G Bäßler, M Eckert, R Fimmers, C Hohoff, M Kraft, C Leuker, G Molsberger, U Pich, S Razbin, T Rothämel, H Schneider, PM Schneider, M Templin, M Vennemann, A Wächter, V Weirich, H Zierdt, L Roewer. Gemeinsame Empfehlungen der Projektgruppe „Biostatistische DNA-Berechnungen“ und der Spurenkommission zur biostatistischen Bewertung von Y-chromosomalen DNA-Befunden. *Rechtsmedizin* 28: 138-142 (2018).
- [10] J Russell, B Rodríguez, RP Laude, MCA de Ungria. An integrated system for forensic DNA testing of sexual assault cases in the Philippines. *Forensic Science International Synergy* 3: 100133 (2021).