SEDNA: A computer program for semiparametric estimation of densities and match probabilities in DNA forensic identification and paternity cases

D. Alonso[1], R. Cao[1], A. Carracedo[2] and E. Valverde[3]
[1]Dept. of Mathematics, Faculty of Computing Sciences, University of A Coruña, Spain
[2]Institute of Legal Medicine, Faculty of Medicine, University of Santiago de Compostela, Spain
[3]PharmaGen, S.A., C/ Calera, 3, Tres Cantos, Madrid, Spain

Recent approaches to estimate match probabilities in VNTR loci (Devlin et al. 1992; Berry et al. 1992; Evett et al. 1992) have made use of density functions. In this communication we present a computer program, based on a model previously reported, which shows very good practical results. This model relies on a semiparametric estimation of density functions (Valverde et al. 1992) and a subsequent calculation of the probability of a match between 2 bands by means of a reformulation of the Bayes theorem in terms of the conditional density function. The model has been further extended to the comparison of two-banded profiles, taking account of the correlation observed in the measurement errors of each pair of bands. This method has been implemented in C language, resulting in the so-called SEDNA program.

SEDNA uses an initial database containing information with measured fragment lengths and their correspondent true values to calibrate the experimental error. This adjustment can also be made from data of allelic controls analyzed repeatedly. A variety of databases pertaining to different loci, different restriction enzymes or different populations can be handled by SEDNA. This allows computation of the semiparametric density estimation of the fragment length as well as match probabilities. The program facilitates the use of information about relatives in paternity cases and also produces graphical outputs.

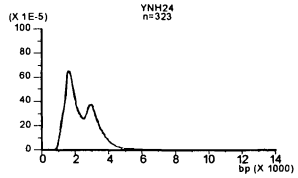

Figure 1. Plot of the density function of fragment lengths for the VNTR polymorphism HaeIII/YNH24 in the Spanish population.

To assess the practical performance of this method we carried out an experiment, similar to that described by Evett et al. (1992), using a data set of 229 individuals analyzed in duplicate. From this experiment we extracted information about comparisons both between persons and within persons. In the total of 26335 comparisons made using three probes (MS31, MS43a and YNH24) we didn't find any incorrect matches or non- matches. It is worth noting in this setting that previous calibration of the experimental error has a remarkable influence on the probability values, however this effect does not lead to an incorrect assignment of matches or non-matches (Figure 2).

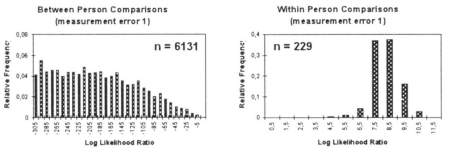

Figure 2A. Measurement error = 0,92 %. In the between person comparisons, 19975 (76.5 %) of 26106 values tested were outwith log likelihood ratio capacity of computer, therefore assumed = 0.

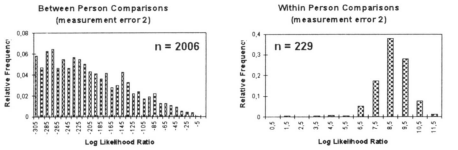

Figure 2B. Measurement error = 0,745 %. In the between person comparisons, 24100 (92.3 %) of 26106 values tested were outwith log likelihood ratio capacity of computer, therefore assumed = 0.

PROGRAM FUNCTION

SEDNA is a computer software package (programmed in C language) that carries out several calculations concerning the estimation of the experimental error, as well as the density function of the fragment length (which is also plotted) and match probabilities in either one-allele (disputed paternity) or two-allele (forensic identification) cases.

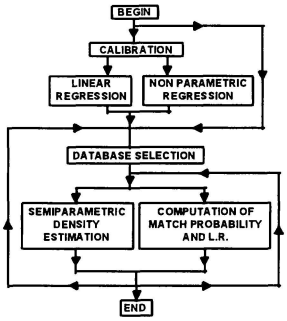

Figure 3. Flow diagram of SEDNA.

A number of different methods are available in the program to compute the estimated variance of the experimental error. Among them we emphasize the non parametric regression and the linear regression approximations to address that problem. This calibration for the program is only executed whenever the data concerning true and observed fragment lengths are updated for a particular laboratory.

The second part of the program is devoted to the estimation itself. A semiparametric estimation of the density function is computed and plotted for the user's selected database. Furthermore, given the observed fragment lengths the match probability is computed by the program in either one-allele or two-allele situations. An important technical problem in this part of the program relates to the achievement of fast computation by the kernel method for estimating the density function. This has been addressed by using the fast Fourier transform algorithm.

REFERENCES

Berry, DA; Evett, IW and Pinchin, R (1992) Statistical inference in crime investigations using deoxyribonucleic acid profiling. Appl Statist 41:499-531

Devlin, B; Risch, N and Roeder K (1992) Forensic inference from DNA fingerprints, J Am Stat Assoc 87:337-350

Evett, IW; Scranage, JK and Pinchin R (1992) An efficient statistical procedure for interpreting DNA single locus profiling data in crime cases. J Forensic Sci Soc 32:307-326

Valverde, E; Cabrero, C; Cao, R; Rodríguez-Calvo, MS; Díez, A; Barros, F; Alemany, J and Carracedo, A (1993) Population genetics of three VNTR polymorphisms in two different Spanish populations. Int J Leg Med 105:251-256