International Congress Series 1288 (2006) 471–473

www.ics-elsevier.com

Monte Carlo Bayesian identification using STR profiles

Donald I. Promish

United States Department of Defense, United States Navy, 68 Richardson Street, Burlington, Vermont 05401-5026, United States

Abstract. The method demonstrated here enables an investigator to analyse individual culprits' STR profiles quickly without depending on "reference groups", to estimate group substructure as a composite of local substructures, and to estimate the minimum profile needed for a specific case.

Keywords: Group membership; Singular group; STR profile; Substructure; Bayes; Monte Carlo; Stratified random sample

1. Introduction

The Monte Carlo Bayesian (MCB) method has several operational features worth noting. (a) The method is case-specific. Both evaluation of and adjustment for substructure are automatic, and they depend only on the culprit STR profile at issue. (b) The method accommodates variation in prior probabilities according to the investigator's judgment regarding non-profile evidence. (c) The method produces probabilities not likelihood ratios. (d) The method does not rely on "reference group" allele frequency data. The investigator can use the method when she/he lacks either knowledge of, or immediate access to, suitable frequency data.

2. Method

The MCB method, in the form of a computer program, iteratively applies Bayes' theorem to stratified random sample arrays comprising specimens taken, in part, from 10 discrete, equal-sized homozygosity ranges, called "demes". The specimens' allele frequency distributions are modelled by normal density functions whose standard

E-mail address: DonaldPromish@cs.com.

0531-5131/
doi:10.1016/j.ics.2005.08.014

deviations correspond to deme homozygosity and whose means are the sampling random variables. In addition to the 10 demes, an 11th array element, a "singular group" representing a suspect with a matching profile may be included. The program, during each iteration, evaluates each deme (and the suspect's singular group) for its ability to produce the profile. The likelihood of the profile, given the specimen, is merely an intermediate result during the MCB procedure.

The MCB program develops a collection of probability sets on the sample arrays. By taking the average of this collection, the program calculates the set of probabilities that the culprit is either the suspect (a member of the singular group) or somebody else (a member of a non-suspect deme).

Each MCB computation for this article consisted of 500 iterations on a MicroSoft Excel spreadsheet, and took less than 20 s. The software is available from the author on request by post.

3. Results

Typical results of a Monte Carlo Bayesian STR profile analysis are shown below. The profile of the subject, labelled C096, was taken from among the hundreds of profiles available in Ref. [1].

Subject: C096			
Profile:			
Locus	(Alleles)	Locus	(Alleles)
D3S1358	(16, 18)	VWA	(17, 17)
FGA	(21, 22)	D8S1179	(13, 15)
D21S11	(30, 31)	D18S51	(15, 18)
D5S818	(12, 12)	D13S317	(12, 13)
D7S820	(11, 12)	CSF1PO	(11, 12)
TPOX	(8, 11)	TH01	(8, 8)
D16S539	(12, 13)		

Local substructure probabilities for Subject C096, using uninformative ("flat") prior:

Homozygosity interval	Posterior probability
0.0–0.1	0.000
0.1–0.2	0.003
0.2–0.3	0.068
0.3–0.4	0.133
0.4–0.5	0.179
0.5–0.6	0.180
0.6–0.7	0.171
0.7–0.8	0.099
0.8–0.9	0.094
0.9–1.0	0.071

"Cold hit" probability, given a match between Subject C096 and a known individual, that the subject is the known individual: 1.000... to 7 decimal places. This is based solely on the estimated world population in 2050 that sets the prior probability at 1/(10 billion).

Table 1
Minimum probability, given a match, that the culprit is the suspect

Number of loci in culprit's profile	"Cold hit" prior: 10^{-10}	"Minimum probable cause" prior: 0.50
	Posterior probabilities	Posterior probabilities
13	0.92	1.000... to 8 decimal places
12	0.81	1.000... to 7 decimal places
11	0.64	1.000... to 7 decimal places
10	0.45	1.000... to 7 decimal places
9	0.23	1.000... to 6 decimal places
8	0.08	1.000... to 5 decimal places
7	0.01	1.000... to 4 decimal places
6	–	1.000... to 4 decimal places
5	–	1.000
4	–	0.999
3	–	0.996
2	–	0.982
1	–	0.901
0	–	0.500

"Minimum probable cause" probability, given a match between Subject C096 and a known individual, that the subject is the known individual: 1.000... to at least 30 decimal places. This is based on non-profile evidence that sets the prior probability at 0.500....

The size of an STR profile will, in conjunction with other evidence, affect the investigator's ability to identify a culprit with a suspect. Table 1 shows the *least* one can expect under two conditions of non-profile evidence strength. Note that this minimum probability depends only on the number of loci in the culprit's profile, and not on which specific loci they are.

Acknowledgments

Humble gratitude to Professor Judith E. Stone, of the University of Vermont, for relentless guidance in matters of written English.

Deepest thanks to Karen L. Ayres, PhD., of the University of Reading, for generous counsel and encouragement.

Highest admiration to several anonymous reviewers (and to one anonymous editor) for their patience and support over several years.

References

[1] B. Budowle, T.R. Moretti, Genotype profiles for six population groups at the 13 CODIS short tandem repeat core loci..., Forensic Science Communications, July 1999, Volume 1, Number 2; http://www.fbi.gov/hq/lab/fsc/backissu/july1999/budowle.htm, dnaloci.txt.