



# Source Attribution of a Forensic DNA Profile

**CURRENT ISSUE**

**BACK ISSUES**

**SEARCH  
ALL ISSUES**

**JOURNAL  
DESCRIPTION**

**EDITORS**

**INSTRUCTIONS  
FOR AUTHORS**

**HANDBOOK OF  
FORENSIC SERVICES**

**LINKS**

**LAB HOME**

**FBI LIBRARY**

**Bruce Budowle**  
Senior Scientist  
Federal Bureau of Investigation  
Quantico, Virginia

**Ranjit Chakraborty**  
Professor  
Human Genetics Center  
University of Texas  
School of Biomedical Sciences  
Houston, Texas

**George Carmody**  
Professor  
Department of Biology  
Carleton University  
Ottawa, Canada

**Keith L. Monson**  
Research Chemist  
Forensic Science Research Unit  
Federal Bureau of Investigation  
Quantico, Virginia

*Read about . . .*

[Abstract](#)

[Introduction](#)

[Criterion for Unrelated Individuals](#)

[Table 1: Random Match Probability](#)

[Thresholds for Source Attribution at Various  
Population Sizes and Confidence Levels](#)

[Criterion for Related Individuals](#)

[Profiles From Individuals From the Same Population](#)

[Conclusion](#)

[References](#)

## **Abstract**

A sufficient number of highly polymorphic genetic markers can be typed from DNA derived from forensic biological samples such that in many cases the reciprocal of the random match probability exceeds the world population many fold. The

magnitude of these estimates has approached the point where it is unlikely that two unrelated individuals carry the same type, and probabilities for related individuals, excluding identical twins, carrying the same type are remote. Once the rarity of a multiple locus profile is estimated (which, depending upon the scenario, may include the parameter  $\theta$  describing relatedness and/or conditional probabilities for relatives), objective criteria can be used to report that with reasonable scientific certainty a particular individual is the source of an evidentiary sample. Approaches are described to assess source attribution of an evidentiary profile and are illustrated for selected confidence levels and population sizes.

## Introduction

There are a large number of genetic markers available today for human identity testing. The FBI Laboratory routinely types 13 short tandem repeat (STR) loci. These loci, which also comprise the core loci for CODIS, are

- CSF1PO
- FGA
- THO1
- TPOX
- VWA
- D3S1358
- D5S818
- D7S820
- D8S1179
- D13S317
- D16S539
- D18S51, and
- D21S11.

By typing these STR loci, the random match probability for a multiple locus profile will be exceedingly small (Committee on DNA Forensic Science, National Research Council 1996 [hereinafter, NRC II Report 1996]; Chakraborty et al. 1999). The average random match probability for unrelated individuals for the 13 STR loci is less than one in a trillion, even in populations with reduced genetic variability (such as Apaches; see Chakraborty et al. 1999).

In many forensic cases processed routinely today, a sufficient number of highly polymorphic markers are used so that the reciprocals of random match probabilities exceed the world population many fold. The magnitude of these estimates has approached the point where, if objective criteria are met, it may be appropriate to assert source attribution, that is, that with reasonable scientific certainty, a particular individual is the source of an evidentiary sample. However, source attribution should not be confused with "uniqueness." When deducing source attribution, there often is little need to establish that a DNA profile is found in only one person in the entire world. Instead, source attribution should be considered in the context of the case, and rarely would the entire world's population be considered as the pool of potential contributors of an evidence sample (for a discussion on source attribution see DNA Advisory Board 2000 [hereinafter, DAB 2000]).

The following describes an approach to assess whether a particular multiple locus DNA profile can be considered "unique" within the context of a case. Collateral issues of Hardy–Weinberg expectations, linkage equilibrium, the parameter  $\theta$  describing relatedness, confidence levels of single or multiple locus frequencies, or the probability of the same genotype in a relative have not been addressed. These are discussed elsewhere (NRC II Report 1996).

[Back to the top](#)

## Criterion for Unrelated Individuals

An evidentiary profile may be considered unique in the context of a forensic identity case if it originated from only one person (excluding identical twins) in a population of  $N$  individuals. To develop criteria to assess the question of uniqueness, let  $p_x$  equal the random match probability for a given evidentiary profile  $X$ . The random match probability is calculated using the NRC II Report (1996) Formulae 4.1b and 4.4a for general population scenarios or formula 4.10 under the assumption that the contributor and the accused could only come from one subgroup. The value of  $\theta$  is 0.01 (NRC II Report 1996; Chakraborty et al. 1999; Budowle et al. in press), except for estimates for isolated subgroups, where 0.03 is used (NRC II Report 1996; Budowle 2000). The frequency is then increased by a factor of 10 to produce a more conservative estimate (NRC II Report 1996). Then

$$(1 - p_x)^N$$

is the probability of not observing the particular profile in a sample of  $N$  unrelated individuals (i.e., it is unique). We require that this probability be greater than or equal to a  $1 - \alpha$  confidence level,

$$(1 - p_x)^N \geq 1 - \alpha$$

or

$$p_x \leq 1 - (1 - \alpha)^{1/N}$$

Specifying a  $(1 - \alpha)100\%$  confidence level of 95% or 99% (i.e., an  $\alpha$  of 0.05 or 0.01, respectively) enables determination of the random match probability threshold to assert with a specific degree of confidence (95% or 99%) that the particular evidentiary profile is unique within a population of  $N$  unrelated individuals.

In practice,  $p_x$  is calculated for each major population group residing in the geographic area where the crime was committed. When there is no reason to believe that a smaller population group is relevant, the FBI sets  $N$  to 260 million, the approximate size of the U.S. population. For smaller, defined populations,  $N$  is based on census values or other appropriate values determined by the facts of the case. (It should be noted that  $N$  can be configured to the context of the specific case, which could be a sample as small as two unrelated individuals to that of an entire town, city, state, or country. Alternatively,  $N$  can be set by laboratory policy.) The source attribution formula advocated here is simple and exceedingly likely to be conservative, because  $N$  (set at 260 million) is substantially larger than the size of a population of potential sample contributors that would inhabit the area where a crime was committed. Moreover, because an upper confidence limit of 95% (or 99%) of the estimate of the random match probability is used, the probability of not observing a particular profile in a population of  $N$  unrelated individuals typically will be underestimated. As a consequence, the degree of certainty of uniqueness is likely to be larger than  $100(1 - \alpha)\%$ .

[Table 1](#) displays the maximum values of  $p_x$  which would support an assertion of source attribution, given various population sizes and confidence levels. For example, at an  $N$  approximately the size of the U.S. population (i.e., 260,000,000) a random match probability less than  $3.9 \times 10^{-11}$  will confer at least 99% confidence that the evidentiary profile is unique in the population.

[Back to the top](#)

## Criterion for Related Individuals

Although the previous approach enables assessment regarding source attribution for unrelated individuals, in some cases, it may be relevant to consider that a relative of the suspect may be in the pool of potential contributors to the sample. Of course, this is moot if the suspect does not have the prerequisite relative(s) or the relative(s) could not have had access to the crime scene. If a relative had access to a crime scene and there is reason to believe that he or she could have been a contributor of the evidence, then a reference sample should be taken from the relative (under such circumstances there should be sufficient probable cause to obtain a sample). Thus, in the United States, rarely should there be a need to calculate the probability that a relative carries the same type as the accused, as typing the relative would resolve identity issues.

When a suspected relative cannot be typed, the conditional probability that the relative has the same DNA profile as the accused (limited to those loci typed in the evidence) can be calculated (NRC II Report 1996; Li and Saks 1954). The current core 13 STR loci should be more than sufficient to resolve the question of whether or not a relative carries the same DNA profile as the accused. Chakraborty and colleagues (Chakraborty et al. 1999) reported that among African Americans, Chinese, and Caucasians, the most common conditional probability for a 13 STR locus profile is expected to occur with a frequency no more than one in 40,000 among full siblings. For rape cases where semen is the evidence material, this value decreases by one half. For more distantly related relatives, for example first cousins, the most common conditional probability is less than one in a billion. Balding (1999) recently suggested that 11 STR loci would be sufficient to assert uniqueness (at a 99.9% confidence level) even when considering brothers. Thus, source attribution should be possible routinely for scenarios where relatives of the suspect cannot be typed with typing results of the suspect from 11–13 of the CODIS STR loci.

[Back to the top](#)

## Profiles From Individuals From the Same Population

In most cases there is little or no evidence a priori regarding the ethnic make-up of the individual who deposited the evidentiary sample (or to which subgroup the true perpetrator belongs; Budowle et al. 1992). In most cases, the ethnicity of the suspect is irrelevant (NRC II Report 1996; Budowle et al. 1992). Moreover, the ethnicity of a suspect usually is not easily defined (Budowle et al. 1992). However, Balding and Nichols (1994; 1997) contend that it is appropriate in most cases to assume that the suspect and true perpetrator are from the same subgroup and estimates should be conditioned on the suspect's profile and subgroup.

If conditional probabilities are deemed appropriate in some cases, then  $N$  (otherwise set at 260 million) becomes limited to the size of the population of potential contributors of the evidentiary sample who belong to a specific subgroup. Thus,  $N$  becomes substantially smaller by several orders of magnitude. Further, the use of a larger  $\theta$  value (particularly for the conditional probability approach) obliges the use of a much smaller  $N$ . Thus, even with a higher value for  $\theta$  (with the concomitant reduction in  $N$ ), the threshold for evaluating source attribution for this conditional

probability is several orders of magnitude larger than described previously for unrelated individuals. Thus, we contend that the current formula for the threshold for source attribution for unrelated individuals, although simplified, is conservative and generally applicable for practical purposes.

[Back to the top](#)

## Conclusion

Procedures have been described for determining when it is reasonable to report that the suspect with a matching profile is the source of the evidence. The inequality  $p_x$  also sets an approximate upper bound for  $p_x$  as  $\alpha/N$  for reasonable values of  $\alpha$  (i.e., 0.10 to 0.01). Source attribution of evidence does not require that the profile be unique, but instead that there is reasonable scientific certainty regarding the source of the evidence. Obviously, the level of significance attached to a computation (see [Table 1](#)) leading to a statement of source dictates how small is the chance that the profile is carried by another individual in the population of  $N$  individuals (decreasing as the confidence level increases). Of course, because of ancillary information (though not quantifiable by genetic principles) in the overwhelming number of cases, the consequence of another individual carrying the same DNA profile would have little effect on the assertion that the suspect is the source of the evidence.

Source attribution can be assessed for unrelated individuals and, when appropriate, for a relative conditioned on the suspect's DNA profile. The size of the population and the appropriate confidence level to use become policy decisions for the laboratory or can be considered on the basis of the circumstances of the case. For most cases, excluding the subgroup scenario, the FBI Laboratory sets  $N$  at the approximate size of the U.S. population for the population of unrelated individuals. Thus,  $N$  will be larger than any one subgroup in the United States and substantially larger than the population of potential contributors of the evidentiary material. The size of  $N$  for the special case where only a subgroup is to be considered will be determined on a case-by-case basis. In all cases, a minimum confidence level of 99% will be used by the FBI Laboratory.

[Back to the top](#)

## References

- Balding, D. J. When can a DNA profile be regarded as unique?, *Science & Justice* (1999) 39:257–260.
- Balding, D. J. and Nichols, R. A. DNA profile match probability calculations: How to allow for population stratification, relatedness, database selection and single bands, *Forensic Science International* (1994) 64:125–140.
- Balding, D. J. and Nichols, R. A. Significant genetic correlations among Caucasians at forensic DNA loci, *Heredity* (1997) 78:583–589.
- Budowle, B. *CODIS STR Population Data*. American Academy of Forensic Sciences, Reno, Nevada, 2000.
- Budowle, B., Defenbaugh, D. A., and Keys, K. M. Genetic variation at nine short

tandem repeat loci in Chamorros and Filipinos from Guam, *Legal Medicine* (in press).

Budowle, B., Monson, K. L., and Wooley, J. R. The reliability of statistical estimates in forensic DNA typing. In: *DNA Identification*. P. R. Billings, ed. Cold Spring Harbor Press, New York, 1992, pp. 79–90.

Committee on DNA Forensic Science, National Research Council. *An Update: The Evaluation of Forensic DNA Evidence*. National Academy Press, Washington, DC, 1996.

Chakraborty, R., Stivers, D. N., Su, B., Zhong, Y., and Budowle, B. The utility of STR loci beyond human identification: Implications for the development of new DNA typing systems, *Electrophoresis* (1999) 20:1682–1696.

DNA Advisory Board. Statistical and population genetics issues affecting the evaluation of the frequency of occurrence of DNA profiles calculated from pertinent population database(s) (approved February 23, 2000), *Forensic Science Communications* (July 2000). Available at:  
[www.fbi.gov/programs/lab/fsc/current/dnastat.htm](http://www.fbi.gov/programs/lab/fsc/current/dnastat.htm)

Li, C. C. and Sacks, L. The derivation of joint distributions and correlation between relatives by the use of stochastic matrices, *Biometrics* (1954) 10:347–360.

[Back to the top](#)

FORENSIC SCIENCE COMMUNICATIONS    JULY 2000    VOLUME 2    NUMBER 3

[CURRENT ISSUE](#)

[BACK ISSUES](#)

[SEARCH  
ALL ISSUES](#)

[JOURNAL  
DESCRIPTION](#)

[EDITORS](#)

[INSTRUCTIONS  
FOR AUTHORS](#)

[LINKS](#)

[HANDBOOK OF  
FORENSIC SERVICES](#)

[LAB HOME](#)

[FBI  
LIBRARY](#)



*Source Attribution of a  
Forensic DNA Profile*

**CURRENT ISSUE**

**BACK ISSUES**

**SEARCH  
ALL ISSUES**

**JOURNAL  
DESCRIPTION**

**EDITORS**

**INSTRUCTIONS  
FOR AUTHORS**

**HANDBOOK OF  
FORENSIC SERVICES**

**LINKS**

**LAB HOME**

**FBI LIBRARY**

**Table 1**

**Random Match Probability Thresholds for Source  
Attribution at Various Population Sizes and  
Confidence Levels**

Sample Size (N)	Confidence Levels			
	0.90	0.95	0.99	0.999
2	$5.1 \times 10^{-2}$	$2.5 \times 10^{-2}$	$5.0 \times 10^{-3}$	$5.0 \times 10^{-4}$
3	$3.5 \times 10^{-2}$	$1.7 \times 10^{-2}$	$3.3 \times 10^{-3}$	$3.3 \times 10^{-4}$
4	$2.6 \times 10^{-2}$	$1.3 \times 10^{-2}$	$2.5 \times 10^{-3}$	$2.5 \times 10^{-4}$
5	$2.1 \times 10^{-2}$	$1.0 \times 10^{-2}$	$2.0 \times 10^{-3}$	$2.0 \times 10^{-4}$
6	$1.7 \times 10^{-2}$	$8.5 \times 10^{-3}$	$1.7 \times 10^{-3}$	$1.7 \times 10^{-4}$
7	$1.5 \times 10^{-2}$	$7.3 \times 10^{-3}$	$1.4 \times 10^{-3}$	$1.4 \times 10^{-4}$
8	$1.3 \times 10^{-2}$	$6.4 \times 10^{-3}$	$1.3 \times 10^{-3}$	$1.3 \times 10^{-4}$
9	$1.2 \times 10^{-2}$	$5.7 \times 10^{-3}$	$1.1 \times 10^{-3}$	$1.1 \times 10^{-4}$
10	$1.1 \times 10^{-2}$	$5.1 \times 10^{-3}$	$1.0 \times 10^{-3}$	$1.0 \times 10^{-4}$
25	$4.2 \times 10^{-3}$	$2.1 \times 10^{-3}$	$4.0 \times 10^{-4}$	$4.0 \times 10^{-5}$
50	$2.1 \times 10^{-3}$	$1.0 \times 10^{-3}$	$2.0 \times 10^{-4}$	$2.0 \times 10^{-5}$
100	$1.1 \times 10^{-3}$	$5.1 \times 10^{-4}$	$1.0 \times 10^{-4}$	$1.0 \times 10^{-5}$
$1 \times 10^3$	$1.1 \times 10^{-4}$	$5.1 \times 10^{-5}$	$1.0 \times 10^{-5}$	$1.0 \times 10^{-6}$
$1 \times 10^5$	$1.1 \times 10^{-6}$	$5.1 \times 10^{-7}$	$1.0 \times 10^{-7}$	$1.0 \times 10^{-8}$
$1 \times 10^6$	$1.1 \times 10^{-7}$	$5.1 \times 10^{-8}$	$1.0 \times 10^{-8}$	$1.0 \times 10^{-9}$

<b>1 × 10<sup>7</sup></b>	1.1 × 10 <sup>-8</sup>	5.1 × 10 <sup>-9</sup>	1.0 × 10 <sup>-9</sup>	1.0 × 10 <sup>-10</sup>
<b>5 × 10<sup>7</sup></b>	2.1 × 10 <sup>-9</sup>	1.0 × 10 <sup>-9</sup>	2.0 × 10 <sup>-10</sup>	2.0 × 10 <sup>-11</sup>
<b>2.6 × 10<sup>8</sup></b>	4.1 × 10 <sup>-10</sup>	2.0 × 10 <sup>-10</sup>	3.9 × 10 <sup>-11</sup>	3.9 × 10 <sup>-12</sup>
<b>1 × 10<sup>9</sup></b>	1.1 × 10 <sup>-10</sup>	5.1 × 10 <sup>-11</sup>	1.0 × 10 <sup>-11</sup>	1.0 × 10 <sup>-12</sup>
<b>5 × 10<sup>9</sup></b>	2.1 × 10 <sup>-11</sup>	1.0 × 10 <sup>-11</sup>	2.0 × 10 <sup>-12</sup>	2.0 × 10 <sup>-13</sup>

[Back to the top](#)

[Back to the text](#)

FORENSIC SCIENCE COMMUNICATIONS JULY 2000 VOLUME 2 NUMBER 3

[CURRENT ISSUE](#)

[BACK ISSUES](#)

[SEARCH  
ALL ISSUES](#)

[JOURNAL  
DESCRIPTION](#)

[EDITORS](#)

[INSTRUCTIONS  
FOR AUTHORS](#)

[LINKS](#)

[HANDBOOK OF  
FORENSIC SERVICES](#)

[LAB HOME](#)

[FBI  
LIBRARY](#)