6-2001

Short Tandem Repeat-based Identification of Individuals and Parents

Martin Tracey
Department of Biological Sciences, Florida International University, traceym@fiu.edu

CROATIAN MEDICAL JOURNAL
CMJ

42(3):233-238,2001

FORUM

Short Tandem Repeat-based Identification of Individuals and Parents

Martin Tracey

Department of Biological Sciences, Florida International University, Miami, Fla, USA

Estimation of short tandem repeat (STR) multilocus genotype frequencies for the identification of individuals and estimation of allele frequencies for parentage assignment both depend on (a) testing a lot of loci, (b) high levels of polymorphism at each locus tested, and (c) independence among alleles. Independence is critical, because the estimation of multilocus genotype and gamete frequencies is based on multiplying individual allele frequencies to produce a composite frequency estimate. Independence among alleles at a locus is known as Hardy-Weinberg equilibrium, whereas allelic independence between loci is known as linkage equilibrium. The frequency at which individual identification may be declared is a matter of opinion, as there is no scientific way to specify certainty based on frequency estimates. Similarly absolute assignment of parentage is impossible in theory; in practice it is more difficult than individual identification, because only half as much information is available (gamete vs genotype frequency) and because mutation may confound parentage analysis.

Key words: *alleles; DNA fingerprinting; gene frequency; genetic markers; genetics, population; inbreeding; microsatellite repeats; paternity; polymorphism (genetics)*

The logic or algorithm of identification used in forensic DNA identification is identical to the algorithm used in a wide variety of identification strategies (1). Focusing on a specific characteristic, we describe the form of that particular characteristic. In describing the form seen we eliminate or exclude all others. For example, describing a person as either male or female eliminates the other gender in the same logical way that describing a tossed coin as either heads or tails eliminates the other possibility. The person described as male could be any one of approximately three billion males on the Earth, but he can not be a female.

Similarly, microbiologists use a Gram+/Gram– binary analysis as the first step in species identification for bacteria; when an unknown bacterium is typed as Gram+, a vast array of microbes remains to be searched by other characters before a specific species or strain of pathogen is identified. Still, all Gram– microbes are absolutely out of the picture once a sample is typed as Gram+; they are no longer a factor in the identification of the unknown. The Gram+/Gram– analysis is the logical equivalent of male/female or heads/tails analyses. The power of these binary decisions depends on the percentage of males and females, heads and tails, or microbes classified as Gram+ or Gram–. If the distribution is 50:50 we do not have to consider half of all microbes in our analysis of the unknown. Once the Gram stain analysis is completed, another test is performed on the Gram+ unknown to eliminate another portion of the microbial kingdom.

Using precisely the same logical algorithm in describing a specific person, we add additional information to our description once we have specified sex. We might describe the person as an African male. The combination of gender and continent-of-origin allows exclusion of all females and non-Africans. In addition it allows the calculation of the percentage of people still included in the class male Africans. This inclusion-in-one- group/exclusion-from-all-others analysis is universally applied in the biological sciences. Consider these widely used taxonomic characters: nucleus/no nucleus, backbone/no backbone, DNA/RNA, and so on. The identification of a particular species is based on binary inclusion into one group defined by a particular characteristic and frequency and exclusion from the other group.

Of course, we are not limited to just two groups. For example, chemists use the identical algorithm in identification, but ask the gas/liquid/solid question of an unknown in doing qualitative analysis. If the substance is a liquid, it is absolutely not a gas or solid; the question is still binary although there are more than two possible states. One more example will suffice to make the argument. Consider medical diagnosis. Perhaps the first step is to ask *Does the patient have a rash?* While there are surely wide variations in both the type and extent of rash, the question may be asked in binary form in the same logical progression used in the other areas. Following this algorithm we ask an inclusion/exclusion question and once it is answered we proceed to the next inclusion/exclusion question until we have asked enough questions to have an identification. The power of this algorithm is in the number of characteristics analyzed and the frequencies of the classes or states within each character.

This is the logic of forensic DNA identification of individuals by genotype and of parents by allelic types. It is also the logic of classical serological testing, but here an insufficient number of tests are available for routine identification. Large percentages of the population may be excluded by some serological assays, but we do not often have enough binary tests to claim identification. The power of forensic DNA analysis is in the polymorphisms at the short tandem repeat (STR) loci and the number of loci used (2). Each single STR genotype or allele identified in an evidence sample is compared to known samples and classified as a match or nonmatch, an inclusion or exclusion. The nonmatch or exclusion is a certainty.

The match or inclusion must be quantified, usually by estimating the STR genotype or allele frequency in the relevant population of potential matches or inclusions. This is done by use of the tools of population genetics, and the frequency and methods used to argue individual identification or parentage remain, and will remain, moot (3,4).

Individual Identification

Short Tandem Repeat Match or Inclusion

A vaginal swab is genotyped as a 12,10 at the D16S539 locus and this is compared to four suspects who have been genotyped from buccal swabs. Three are excluded, because they have D16S539 genotypes, different from the swab genotype. Suspect two, on the other hand, is a D16S539 12,10; he matches or is included in the group of 12,10 people. Checking our database we see that approximately 5% of all people are 12,10 at this locus. Ninety-five percent of all people will be excluded as potential sources of this vaginal swab DNA pattern at this locus, but one person in twenty will match. This is clearly a very useful, polymorphic locus, but it does not provide identification.

There are a lot of people left in the pool of possible sources. Considering the Earth as the pool, we see that there are 0.05 × 6 billion or approximately three hundred million people in this class of people who are D16S539 12,10. Still, 95% of all people have been eliminated by this single evaluation of an STR genotype.

Independent Short Tandem Repeat Match or Inclusion

The second locus genotyped off the vaginal swab DNA is THO1 and a 9.3,8 type is recorded. Suspect two, who matched at D16S539, is a 9.3,8 at this locus; we can combine the two loci to estimate the odds of a match at both D16S539 and THO1. The data base frequency of the 9.3,8 genotype is approximately 6%. Since the two loci are independent, we can estimate the joint frequency as:

f(D16S539/12,10 = 0.05) × f(THO1/9.3,8 = 0.06) = 0.003.

Approximately one person in 333 (1/0.003) will be a 12,10 at D16S539 and a 9.3,8 at THO1. Each of these loci is a powerful excluder, because more than 90% of all people are excluded by each test; the individual locus polymorphism accounts for this power of exclusion. But the two loci in combination exclude more than 99% of all people as potential sources of the DNA on the vaginal swab.

The true power of DNA testing lies in polymorphism at the individual loci *and* the number of loci tested. To estimate multilocus genotype frequencies in this manner, the alleles at each locus must be inherited independently (Hardy-Weinberg equilibrium) and the alleles among loci must be inherited independently (linkage equilibrium). These two independence conditions must hold because we are multiplying individual allele frequencies to produce a composite estimate. The situation is analogous to the estimation of the frequency with which we expect to draw the ace of spades five times in a row from a fair deck of cards. Each draw must be fair and there must be no correlation among draws.

Hardy-Weinberg Equilibrium

When Gregor Mendel formulated his rules of inheritance, he allowed for continuous selfing among F_1 progeny of two pure lines and he showed that heterozygosity is lost under this breeding scheme. The next logical step would have been to estimate the genotypic changes seen under random mating (5). Mendel did not take this step. Following the rediscovery of Mendelism, a number of scientists studied the behavior of alleles and genotypes in populations, but British mathematician G. H. Hardy and German physician W. Weinberg are most frequently credited with the proof that allele and genotype frequencies are in equilibrium after a single generation of random mating (5). Assuming infinite population size, lack of selection, mutation and migration, as well as random mating, we can show that allele and genotype frequencies do not change (Table 1).

The logic of this proof table is straightforward, if tedious. First we estimate the random mating frequencies for all possible pairings; then we use Mendelian ratios to estimate the frequencies of progeny genotypes. For the AA × Aa matings the Mendelian ratio is 1:1, and half of the mating frequency is distributed to AA and half to Aa. Once all of these estimates have been generated, we sum the progeny genotype classes and find that the frequencies are identical to those of the parental generation, $p^2 + 2pq + q^2$.

Allelic Independence/Hardy-Weinberg Justifies Rare Genotype Frequencies

Often STR genotype frequencies for a single locus are less than 0.01 and the composite genotype frequency over twelve or more loci is less than 1.0×10^{-15}. People are very easily confused by these astronomically low frequencies, especially when confronted by the fact that the data bases used to estimate allele frequencies are often considerably less than 500 people or 1,000 alleles. Common sense sug-

Table 1. Hardy-Weinberg equilibrium[a]

		Offspring		
Matings	Mating frequency	AA	Aa	aa
AA x AA	$(p^2)(p^2) = p^4$	p^4		
AA x Aa	$2(p^2)(2pq) = 4p^3q$	$2p^3q$	$2p^3q$	
Aa x Aa	$4p^2q^2$	p^2q^2	$2p^2q^2$	p^2q^2
AA x aa	$2p^2q^2$		$2p^2q^2$	
Aa x aa	$4pq^3$		$2pq^3$	$2pq^3$
aa x aa	q^4			q^4
Total	100 %	p^2	$2pq$	q^2

[a]Defining the AA, Aa, and aa genotypes as p^2, $2pq$, and q^2 in the parental generation, respectively, we may predict the frequencies for all possible random matings by multiplying male and female frequencies. Then, Mendelian expectations are used to partition offspring for each mating. Finally, the offspring classes are totaled and we see that the frequencies have not changed; they are in equilibrium.

gests to judges, jurors, and others that frequencies converting to odds of one in a quadrillion should be based on databases of quadrillions.

The response to this misperception is simple. As long as alleles are inherited independently, the Hardy-Weinberg applies at each locus, and linkage equilibrium will apply among loci. This means that composite genotype frequencies may be estimated by multiplying allele frequencies to estimate individual locus genotypes and composite genotypes. The relevant estimation error is, then, only the allele frequency estimation error (6,7).

The number of genotypes increases exponentially with the number of alleles at a locus according to the formula:

$$GENOTYPES = n(n+1)/2$$

where *n* is the number of alleles. For a locus with two alleles, we must estimate allele frequency for two alleles to generate frequency estimates for three genotypes using the Hardy-Weinberg Law. If p = frequency of allele 1, and q = frequency of allele 2, then the two homozygotes are simply p^2 and q^2, whereas the estimator of heterozygote frequency is 2pq. If there are more than two alleles, the efficiency of genotype frequency estimation increases dramatically. For a locus with six alleles, 21 genotypes will be produced and we must estimate six allele frequencies. We have three times as many allele frequencies to estimate as in the two allele case, but we have seven times as many genotypes.

A similar exponential increase is seen when we turn to estimating the frequency of composite or multilocus genotypes. Here the independence among alleles is usually treated as linkage equilibrium and it was, theoretically, a problem because Hardy-Weinberg equilibrium is achieved in a single generation of random mating after a perturbation. Linkage equilibrium, on the other hand, approaches equilibrium at a rate dependent on the recombination rate; for unlinked loci departures from linkage equilibrium are halved each generation:

$$D_t = (1-r)^t D_0.$$

The maximum initial value of disequilibrium D_0 value depends on the founder population allele frequencies; with allele frequencies of ½, $D_0 = +0.25$ or –0.25. Under these circumstances, approximate linkage equilibrium between loci on different chromosomes will be reached in four generations. Fortunately, there is scant evidence of any significant departures from linkage equilibrium (7).

The major sources of sampling error in estimating the frequency of multilocus genotypes are sampling error and the possible variation among geographic and ethnic populations. Fortunately, there are only occasional significant differences seen among different populations sampled. The proportion of departure from Hardy-Weinberg expectation attributable to population subdivision is F_{ST} G_{ST} or theta in most of the literature. Estimates are almost always less than 0.01 when populations are compared, especially if the populations are classed in the same major racial group (8).

Multilocus Frequencies for Different Populations

The theoretically possible departures from Hardy-Weinberg and linkage equilibrium have been very difficult to find in databases; either because departures do not exist or, more likely, because they are trivial. In any case, STRs permit identification of each person on the planet (9), and the differences among databases are inconsequential when a dozen or more STR loci are analyzed.

Four U.S. population databases were used to calculate an estimated multilocus genotype frequency for the genotype listed in the first column of Table 2. The average – an illegitimate value – is presented in the sixth column and these average values were used to calculate the multilocus frequency, which is given in the last column. Locus D8S1179 is the best locus; it excludes more than 99% of all individuals, all who are not D8S1179 (9,10). But the combination of the two least powerful loci, D3S1358 and vWA, excludes more than 98% on average. The power of STR analysis is based in the number of loci tested, and the precise frequency at which different people will accept that the data justify a conclusion of individual identification will remain undefinable.

Parental Identification

The use of genetic systems to analyze parentage began shortly after the description of ABO blood groups, which were also used in forensic exclusion. The scientific value of parentage analysis is, strictly speaking, exclusion; one can never absolutely prove that there is no other man or woman in the world who could be the father or mother. Still, the assignment of parentage may be one of the most certain truths of life if enough alleles are analyzed. Occasionally, as in the paternity suit against Charlie Chaplin, the courts have ignored the scientific certainty of exclusion in favor of another perceived good. Chaplin was ordered to pay child support, although the genetic proof that he could not have been the child's father was available to the court (10). In the absence of mutation, the presence in a child of an allele not seen in the alleged parent is absolute proof that the alleged parent is not the true parent. Still mutation is a fact (11) and one might argue parentage over a number of genetic tests showing exclusion by insisting on the possibility that they were all mutations. Formally, then, the frequency of parentage under this condition of mutation would be equal to the mutation rate raised to the power of the number of tests, say $(10^{-3})^{15}$ for a 15-allele test of parentage.

By a similar argument, parentage is assigned when alleles are shared between the child and the alleged parent, although it is formally possible that another individual who carries all or most of the shared alleles could be the true parent.

Short tandem repeat loci offer the opportunity to study enough loci to satisfy all but the most absurd requirements for exclusion or proof of parentage, because the loci used are highly polymorphic and there are a great many loci available for testing parentage (12,13).

Principles of Parentage Analysis

The basis of parentage analysis is very simple: a child must receive, in the absence of mutation, one allele matching each parent. For example, in a simple case, a mother who is a genotype 10,12 and a father who is a 9,14 may produce children of the following types: 9,10 9,12 10,14, and 12,14. If the specific genotype of a child is known, say it is 10,14, and the genotype of the parent contributing one allele, usually the mother, is also known, we may identify the allele that must have come from the other questioned parent. In this case, the mother is known to be a 10,12; the egg must have been a 10. The true father must have contributed a sperm carrying the 14 allele. When the true father is known to be a heterozygote or an alleged father is identified as a heterozygote, he has a 50:50 chance of fathering a child with each of his alleles. In this case, a heterozygous male with a 14 allele will produce a child carrying his 14 allele ½ of the time. If we switch focus from a specific male to the entire population of possible fathers, we can say only that the true father must carry the 14 allele. Thus the chance of randomly drawing the true father from the population of possible fathers is simply the frequency of the 14 in this population. The analysis of parentage is formally equivalent to the inclusion/exclusion algorithm used in forensic identification and the examples discussed in the introduction.

The two possibilities: (a) true father is identified, or (b) father is randomly drawn from the population of possible fathers, may be tabulated as a Punnett square in the first case, and as part of a tabulation of all possible matings and children for a multiallelic locus in the second case. The full table for the biallelic or binomial case is presented in Table 3. For the Punnett square or true father case, we write the frequency of eggs across the top of the square and the frequency of sperm along the left side of the square. The 50:50 probability that a specific child will be produced by this heterozygous man may be seen by examining the column of sperm allele frequencies, where 50% are 9 and 50% are 14:

	½ 10	½ 12
½ 9	¼ 9,10	¼ 9,12
½ 14	¼ 10,14	¼ 12,14

or by summing the progeny frequencies in the 9 and 14 rows to see that half of the children carry a 9 and half carry a 14.

The frequency of randomly drawn possible fathers may be seen by writing the possible matings that will produce a specific progeny genotype, say the 10,14 discussed above:

Since the frequency of the 14 allele is the sum of the 14 homozygote frequency plus half all the possi-

Table 2. Short tandem repeat inclusion frequencies for a selection of U.S. populations in North Carolina (NC) and Florida (Fla)[a]

	Population genotype frequencies					
	U.S. Caucasians		U.S. African ancestry			
Locus	NC	Fla	NC	Fla	"average"	freq$_{Cum}$
FGA(21,24)	0.05	0.04	0.03	0.04	0.04±0.01	–
PENTAE(7,14)	0.02	0.03	0.02	0.02	0.02±0.01	0.0008
D13S317(12,13)	0.06	0.07	0.14	0.14	0.10±0.04	8.0×10^{-5}
D16S539(10,12)	0.03	0.04	0.04	0.06	0.04±0.01	3.2×10^{-6}
CSF1PO(9,13)[b]	0.004	0.003	0.004	0.004	0.004±0.001	1.3×10^{-8}
D21S11(28,30)	0.09	0.11	0.09	0.08	0.09±0.01	1.0×10^{-9}
D18S51(14,15)	0.05	0.05	0.02	0.02	0.04±0.02	4.6×10^{-11}
D8S1179(9,10)	0.003	0.004	0.001	0.002	0.002±0.001	9.2×10^{-14}
D3S1358(15,16)	0.12	0.14	0.21	0.19	0.16±0.04	1.5×10^{-14}
D7S820(9,12)[c]	0.05	0.04	0.02	0.02	0.03±0.02	4.4×10^{-16}
D5S818(11/.01)	0.11	0.17	0.06	0.05	0.10±0.06	4.4×10^{-17}
VWA(16,17)	0.12	0.13	0.11	0.10	0.120.01	5.3×10^{-18}
THO1(8,10)	0.003	0.004	0.003	0.005	0.004±0.001	2.1×10^{-20}
TPOX(8,10)	0.07	0.07	0.05	0.07	0.060±.01	1.3×10^{-21}
PENTAD(10,14)[d]	0.02	0.01	0.005	0.004	0.01±0.01	1.3×10^{-23}

[a] Genotypic frequencies for 15 loci from four populations, along with an average, are shown. The final column is the cumulative frequency (freq$_{Cum}$) for the multilocus genotype.
[b] Over five loci the odds of coincidental match are 1 in 78 million.
[c] Over 10 loci the odds of coincidental match are 1 in 2.2 quadrillion.
[d] Over 15 loci the odds of coincidental match are 1 in 78 sextillion.

ble 14 allele heterozygotes, the frequency of men capable of fathering a child with a 14 allele is just the allele frequency.

Likelihood that the True Father is the Alleged Father vs Some Random Man

As we have seen above, a heterozygous man, with an allele matching that seen in the child, will contribute this allele to half of his progeny. Thus, this man will produce a child with the required allele ½ of the time. If this alleged father is not the true father, then the chance that a randomly selected male will carry the required allele is the frequency of that allele. The relative frequencies are ½ and p_{14} in our example and a commonly used measure of the evidence favoring paternity for the alleged father is the ratio of these two values, known as the *paternity index* (PI):

PI = (1/2)/p_{14}.

If p_{14} is 10% in the relevant population data base, the presence of a matching 14 allele in the child and the alleged father is evidence five times more favorable to the assignment of paternity to the alleged father than to some other random man.

Possible matings			Frequency	Progeny frequencies	
				10,14	12,14
10,12	×	14,14	$2p_{10}p_{12} \times p_{14}^2$	$p_{10}p_{12}p_{14}^2$	$p_{10}p_{12}p_{14}^2$
	×	14, i	$2p_{10}p_{12} \times 2p_{14}p_i$	$2p_{10}p_{12}p_{14}p_i$	$2p_{10}p_{12}p_{14}p_{i.}$

Probability of Paternity

Assuming only that the child has a father, we may split the pool of possible fathers into the alleged father and all other possible fathers, that is all men carrying the required allele seen in the child. The probability of paternity (PP) for the alleged father is, in this formulation, the probability that he fathered this child (AF) divided by the sum of the probability that he fathered the child (AF) plus the probability that some other male fathered the child (O):

PP = (AF)/[(AF) + (O)] × 100%.

For the above example:

PP = (1/2)/[(1/2) + (1/10)] = 0.83 × 100% = 83%.

Independence

The power of STR analysis resides in the polymorphism seen at individual loci and in the number of loci tested. The frequencies may be multiplied to estimate the expected frequency of a multilocus haplotype contributed to the child by the questioned parent, but this multiplicative estimation is only legitimate if the alleles at a locus and among loci are inherited independently, that is, Hardy-Weinberg and linkage equilibrium must be established for data bases used in parentage analysis. The 15-allele parentage analysis in Table 3 uses, arbitrarily, the first allele from the forensic analysis presented in Table 2. This allows comparison of the power of forensic and parentage analysis for this one example.

For the forensic example, the five locus genotype frequency was 1.3×10^{-8}; the odds of selecting another matching genotype were 1 in 78 million. The five locus haplotype frequency used in this example of parentage analysis is 1.9×10^{-5}, and the associated odds are 1 in 51,000. The genotypic analysis is clearly far more powerful than the haplotype analysis used in parentage analysis. This is obvious, because we use only a single allele in calculating haplotype frequencies; it is, however, a fact too often ignored when parentage analysis is reported as a match over a number of loci.

Problem of Mutation and False Exclusion

While false exclusion, based in mosaicism, etc, is a formal possibility in genotypic analysis, it is not

Table 3. Short tandem repeat allele frequencies and parentage indices for a selection of populations[a]

	Population allele frequencies						
	U.S. Caucasians		U.S. African ancestry				
Locus	NC	Fla	NC	Fla	"average"	PI	PI_{CUM}
FGA(21)	0.18	0.18	0.11	0.12	0.15	3.4	-
PENTAE(7)	0.15	0.18	0.10	0.14	0.14	3.5	11.9
D13S317(12)	0.30	0.25	0.43	0.38	0.34	1.5	17.5
D16S539(10)	0.05	0.06	0.11	0.14	0.09	5.6	97.2
CSF1PO(9)[b]	0.02	0.03	0.04	0.03	0.03	16.7	1620.0
D21S11(28)	0.17	0.18	0.24	0.18	0.19	2.6	4207.8
D18S51(14)	0.14	0.14	0.08	0.07	0.11	4.7	19571.2
D8S1179(9)	0.01	0.02	0.01	0.01	0.01	40.0	$7.8\text{x}10^5$
D3S1358(15)	0.25	0.28	0.30	0.31	0.23	2.2	$1.7\text{x}10^6$
D7S820(9)[c]	0.18	0.12	0.10	0.10	0.12	4.0	$6.7\text{x}10^6$
D5S818(11)	0.32	0.40	0.24	0.23	0.30	3.4	$2.3\text{x}10^7$
VWA(16)	0.21	0.23	0.27	0.27	0.25	2.0	$4.6\text{x}10^7$
THO1(8)	0.11	0.10	0.24	0.23	0.17	2.9	$1.4\text{x}10^8$
TPOX(8)	0.49	0.58	0.34	0.40	0.45	1.1	$1.5\text{x}10^8$
PENTAD(10)[d]	0.13	0.12	0.10	0.10	0.11	4.4	$6.7\text{x}10^8$

[a] Allele frequencies, which match between alleged parent and child for 15 loci from four U.S. populations in North Carolina (NC) and Florida (Fla), and an average are shown. The paternity index (PI) or the ratio of the probability of parentage for the alleged individual divided by the probability that another randomly drawn individual is the true parent, is presented for each locus, as well as cumulatively (PI_{CUM}).

[b] The cumulative haplotype frequency over five loci is $1.9\text{x}10^{-5}$. Over five loci the odds of drawing this haplotype are 1 in 51,000.

[c] The cumulative haplotype frequency over 10 loci is $1.4\text{x}10^{-10}$. Over 10 loci the odds of drawing this haplotype are 1 in $6.9\text{x}10^9$.

[d] The cumulative haplotype frequency over 15 loci is $9.1\text{x}10^{-14}$. Over 15 loci the odds of drawing this haplotype are 1 in $1.1\text{x}10^{13}$.

frequently encountered. Mutation rates for many STR loci are fairly high and meiotic mutations will be detected in the children. One way to handle mutation is to use an estimate of the mutation rate as the probability that the alleged father contributed this nonmatching allele. For example, assume that we have a four allele match between an alleged parent and a child for FGA, PENTAE, D13S317, and D16S539 as in Table 3.

On the other hand, CSF1PO does not match. Since all matching loci are heterozygous in the alleged parent, we may write the probability that this specific individual produced the gamete that produced this child as the product of the four allelic segregations, $(1/2)^4$ times the mutation rate estimate for CSF1PO, say it is 10^{-3}. The denominator of the PI is just the haplotype frequency over these five loci:

$$PI = [(1/2)^4 \times 10^{-3}]/[(0.15)(0.14)(0.34)(0.09)(0.03)$$
$$= 6.25 \times 10^{-5} / 1.9 \times 10^{-5}$$
$$= 3.24.$$

The data still favor the hypothesis of parentage for the alleged parent, but incorporating the mutant allele makes the case much less convincing. The method does, however, have the distinct advantage of being fairly realistic, and estimated rates are available for some STR loci (e.g., *http:// www.cstl.nist.gov/ div831/strbase/mutation.htm*).

References

1 Duncan GT, Tracey ML. Serology and DNA typing. In: Eckert WG, editor. Introduction to forensic science. 2nd ed. London: CRC Press; 1996.

2 Butler JM. Forensic DNA typing: biology and technology behind STR markers. 1st ed. London: Academic Press; 2001.

3 Weir BS. DNA match and profile probabilities: comment on Budowle et al (2000) and Fung and Hu (2000). Forensic Science Communications 2001;3:1-3. Available at: *http://www.fbi.gov/hq/lab/fsc/back issue*. Accessed: March 29, 2001.

4 Reilly P. Legal and public policy issues in DNA forensics. Nat Rev Genet 2001;2:313-7.

5 Provine WB. The origins of theoretical population genetics. 1st ed. Chicago (IL): The University of Chicago Press; 1971.

6 Chakraborty R. Sample size requirements for addressing the population genetics issues of forensic use of DNA typing. Hum Biol 1992;64:141-59.

7 Committee of the National Research Council. The evaluation of forensic DNA evidence. Washington (DC): National Academy Press; 1996.

8 Budowle B, Chakraborty R. Population variation at the CODIS core short tandem repeat loci in Europeans. Legal Med. In press 2001.

9 Crow JF. Two centuries of genetics: a view from halftime. In: Lander E, Page D, Lifton R, editors. Annual Review of Genomics and Human Genetics (Vol 1). Palo Alto (CA): Annual Reviews; 2000. p. 21-40.

10 Ayala FJ, Black B. Science and the courts. American Scientist 1993;81:203-9.

11 Holtkemper U, Rolf B, Hohoff C, Forster P, Brinkmann B. Mutation rates at two human Y-chromosomal microsatellite loci using small pool PCR techniques. Hum Mol Genet 2001;10:629-33.

12 Connor JM, Ferguson-Smith MA. Essential medical genetics. 3rd ed. Oxford: Blackwell Scientific Publications; 1991.

13 Gelehrter TD, Collins FS, Ginsburg D. Principles of medical genetics. 2nd ed. London: Williams and Wilkins; 1998.

Received: April 11, 2001
Accepted: May 4, 2001

Correspondence to:
Marty Tracey
Biological Sciences
Florida International University
University Park 11200 8th St.
Miami, FL 33199, USA
traceym@fiu.edu