Paternity Exclusion by DNA Markers: Effects of Paternal Mutations


ABSTRACT: In parentage testing when one parent is excluded, the distribution of the number of loci showing exclusion due to mutations of the transmitting alleles is derived, and it is contrasted with the expected distribution when the exclusion is caused by non-paternity. This theory is applied to allele frequency data on short tandem repeat loci scored by PCR analysis, and VNTR data scored by Southern blot RFLP analysis that are commonly used in paternity analysis. For such hypervariable loci, wrongly accused males should generally be excluded based two or more loci, while a true father is unlikely to be excluded based on multiple loci due to mutations of paternal alleles. Thus, when these DNA markers are used for parentage analysis, the decision to infer non-paternity based on exclusion of paternity should be based on exclusionary events at two or more loci has a statistical support. Our approach places a reduced weight on the combined exclusion probability. Even with this reduced power of exclusion, the probability of exclusion based on combined tests on STR and VNTR loci is sufficiently large to resolve most paternity dispute cases in general populations.

KEYWORDS: forensic science, paternity testing, STR, VNTR, germline mutation

Introduction

It is now well known that in parentage testing, DNA markers are more efficient than the classical markers, both in terms of exclusionary power for wrongly accused men and inclusion probability for identification of the true biological father (1-5). As a consequence, the paternity testing laboratories in the US and elsewhere are in the process of supplementing and/or replacing the traditional serologic markers with DNA markers. As discussed earlier (7,8), specifically, we show that under certain justifiable assumptions, the distribution of the number of loci that would exclude a randomly accused man can be computed for every mother-child pair, which in turn can be contrasted with the expected distribution when exclusions occur only by mutations of paternal alleles. Using population data on eleven tandem repeat (STR) loci (HUMRENA4[ACAG], HUMFESPPS[AAAT], HUMFABP[AAT], HUMCD4[AAAAAG], HUMCSF1PO[AGAT], HUMTHO1[AAATG], HUMPL2A1[AAT], HUMF13A01[AAAG], HUMCYBR[AAAT], HUMMPNRA[AAAG], and HUMLIPOL[AAAT]), we show a clear dichotomy of the distribution of the number of loci at which exclusions are found for excluded random men, and the one expected from mutations for the STR loci. In contrast, with the possibility of mutations, a single-locus exclusion could become frequent enough for the battery of VNTR probes used in RFLP analysis for paternity testing. Thus, the inference of true exclusion, based on excluding a parent when at least two exclusionary events using these polymorphic markers as a replacement of the standard serological markers for resolving paternity disputes. As discussed by Alford et al. (5) and Pena and Chakraborty (2), DNA testing can also be used to resolve paternity from hairs, paraffin-included blocks of tissue, and even exhumed materials in cases of deceased parents; these are situations in which the classical serological markers are difficult, if not impossible, to use. The high efficiency of DNA markers, of course, has its attendant caveats. This efficiency arises from their comparatively higher number of segregating alleles and consequently higher per-locus heterozygosity. These biological features of DNA markers are caused by a higher rate of mutation at DNA loci in comparison with the traditional markers. In particular, the DNA markers that are most efficient for personal identification purposes are those whose polymorphism is due to variation of the copy numbers of short tandemly repeated nucleotide sequences in specific regions of the genome (called the Variable Number of Tandem Repeat, VNTR, loci). Jeffreys et al. (6) reported empirical data on the rate of mutations at such loci, indicating that the mutation rate at such loci increases with per-locus variability. As a consequence, mutations may become a major factor in deciding whether or not sporadic discordances of genotypes between a child and its putative father could truly be due to mutations, and not due to non-paternity.

A 1991 recommendation by the AABB suggests that when hypervariable single-locus probes are used for paternity determination, exclusion of paternity should be based on exclusionary events at 2 or more loci. While this suggestion is intuitively reasonable, it should accompany a rigorous statistical validation. The purpose of this study is to provide such a validation, using a theory proposed earlier (7,8). Specifically, we show that under certain justifiable assumptions, the distribution of the number of loci that would exclude a randomly accused man can be computed for every mother-child pair, which in turn can be contrasted with the expected distribution when exclusions occur only by mutations of paternal alleles. Using population data on eleven tandem repeat (STR) loci (HUMRENA4[ACAG], HUMFESPPS[AAAT], HUMFABP[AAT], HUMCD4[AAAAAG], HUMCSF1PO[AGAT], HUMTHO1[AAATG], HUMPL2A1[AAT], HUMF13A01[AAAG], HUMCYBR[AAAT], HUMMPNRA[AAAG], and HUMLIPOL[AAAT]), we show a clear dichotomy of the distribution of the number of loci at which exclusions are found for excluded random men, and the one expected from mutations for the STR loci. In contrast, with the possibility of mutations, a single-locus exclusion could become frequent enough for the battery of VNTR probes used in RFLP analysis for paternity testing. Thus, the inference of true exclusion, based on excluding a parent when at least two exclusionary events
(as suggested by the AABB), appears reasonable for the battery of single-locus VNTR probes. Finally, the implications of these findings in selecting a battery of DNA typing loci for paternity analysis are addressed to revisit the question of relative utility of multiple numbers of single locus probes versus multilocus DNA fingerprint methods in parentage analysis (9,10).

Theory

Analysis without invoking mutations

For \( L \) unlinked loci, let \( P_E(l) \) denote the locus-specific probability of exclusion for the \( l \)-th locus \((l = 1, 2, \ldots, L)\). In other words, \( P_E(l) \) is the chance that a random alleged father would be excluded based on DNA typing at the \( l \)-th locus. It is well documented that this probability can be computed at a locus-level averaged over all possible mother-child pair (11-13), or for each possible combinations of mother-child genotype pairs (14). For example, for a multi-allelic codominant locus, the average exclusion probability is given by

\[
P_E(l) = 1 - 2a_2 + a_3 + 3(a_2a_3 - a_3) - 2(a_2^2 - a_4)
\]  

where \( a_i \) is the sum of \( r \)-th power of all allele frequencies at a locus (12,13). In contrast, for a specific mother-child pair, the exclusion probability for the same locus takes the form

\[
P_E(l) = \begin{cases} 
(1-p_i)^2 \\
(1-p_i-p_j)^2
\end{cases}
\]

where the first event refers to the case where the child can receive a single specific allele (say, \( A_i \)) from the biological father, and the second event occurs when both the mother and child are heterozygous for the same pairs of alleles (say \( A_i \) and \( A_j \)). Thus, for any given \( i \) and \( j \), the first event takes place with probability \( p_i(1 - p_i + p_j^2) \) and the second with probability \( p_j(1 - p_i^2) \), for all \( i = 1, 2, \ldots, k \) and \( j > i \) (13,14). The implicit assumptions in deriving equations (1) and (2) are that the population is in approximate Hardy-Weinberg equilibrium at the locus, and that the allele frequencies in the male and female gene pool of the population are equal. When the mother and father are presumably from different populations, the respective expressions are given in (13). It is important to note that most of the discussions of utility of genetic markers in the context of parentage analysis relate to the average probability of exclusion offered by the locus (equation 1). However, in reality the observed exclusion rates should use (2), since these probabilities may be substantially different from the average for the locus as a whole.

When \( L \) such loci are used in paternity analysis, the combined probability of exclusion is given by

\[
P_E(C) = 1 - \prod_{i=1}^{L} (1 - P_E(l))
\]

which may again be computed by replacing the locus-specific values for an average mother-child pair (equation 1), or for the specific mother-child genotype pair observed at the loci (equation 2). The additional assumption in deriving equation (3) is that the alleles at different loci segregate independently of each other (i.e., the loci are under gametic phase equilibrium (15,16)). Also implicit in equation (3) is the notion that a random male is excluded when at least one locus shows exclusion of paternity.

Regardless of the specific formula used, when the \( P_E(l) \) values are available for \( L \) unlinked loci \((l = 1, 2, \ldots, L)\), the logic of derivation for the distribution of the number of loci with respect to which exclusions would be found for a random alleged male was first given by Chakraborty and Schull (7), and used in other contexts in Chakraborty (8). A mathematically equivalent, and computationally simpler, version of their logic may be described as follows. Let \( Q_m^r \) be the probability that a randomly tested male shows exclusions based on exactly \( r \) of the first \( m \) loci tested \((r = 0, 1, 2, \ldots, m; m = 1, 2, \ldots, L)\). The probability \( Q_m^r \) for \( r = 0, 1, 2, \ldots, L \) may be computed by using the recurrence relationship

\[
Q_{m+1}^r = Q_m^r[1 - P_E(m + 1)] + Q_{m-1}^r P_E(m + 1)
\]

for \( m = 1, 2, \ldots, L - 1; \) and \( r = 1, 2, \ldots, m \) with the boundary conditions that for any \( m = 1, 2, \ldots, L \), we have

\[
Q_0^m = \prod_{i=1}^{m} [1 - P_E(l)]
\]

and

\[
Q_0^m = \prod_{i=1}^{m} P_E(l)
\]

Note that the values of \( Q_0^0, Q_1^1, Q_2^2, \ldots, Q_L^L \), computed from recursive use of equation (4) with boundary conditions (5) and (6) also give other important values that can be reported for paternity test results. For example, the combined power of exclusion based on \( L \) loci, when paternity exclusion is based on one or more locus-specific exclusions is \( P_E(C) = 1 - Q_0(L) \), while combined power of exclusion of \( L \) loci when two or more locus-specific exclusions would be considered as the criterion of exclusion is \( P_{E^2}(C) = 1 - Q_0^2 - Q_1^2 \). Now, note that

\[
Q_1^1 = \sum_{l=1}^{L} P_E(l) \prod_{l'=1}^{l-1} [1 - P_E(l')]
\]

Thus, the combined exclusionary power for \( L \) loci based on two or more exclusions, \( P_{E^2}(C) \), becomes

\[
P_{E^2}(C) = 1 - Q_0 \left[ 1 + \sum_{l=1}^{L} \left( \frac{P_E(l)}{1 - P_E(l)} \right) \right]
\]

Furthermore, the mean and variance of the number of loci showing exclusions are given by

\[
\sum_{l=1}^{L} P_E(l)
\]

and

\[
\sum_{l=1}^{L} P_E(l)[1 - P_E(l)]
\]
inclusion of X-linked or Y-linked loci also does not affect such computations (17), although for Y-linked loci \( P_e(l) \) will be zero for a female child. For a X-linked locus, likewise, it would be zero for a male child. In equations (7)–(9), we further assume that the loci are mutually independent to each other (i.e., no linkage disequilibrium exists between loci).

**Analysis Invoking Mutations**

Thus far we have assumed that any discordance between parental and child’s genotypes is due to non-paternity, and that the mother is always correctly identified. When hypervariable DNA markers are used, these assumptions are not necessarily valid, since for tandem repeat loci, mutation rates are occasionally high enough to produce genotypic discordances between parents and offspring even when the family relationship is correct. Rothman et al. (18) addressed the question of assigning the probability of paternity in the presence of mutations. However, the emphasis of their approach was to provide estimates of mutation rates in the presence of non-paternity in the population, whereas we are interested in the question of determining how often exclusion at one or more loci are to be expected in the presence of mutations. In this section, we assume that the mother and alleged father are the true biological parents of the tested child. Furthermore, we assume that the maternal allele is faithfully transmitted to the child, so that no mutation has occurred during the maternal meiosis. Under this scenario, for a specific mother-child pair, the probability that a true father will be excluded due to mutation of the paternal allele(s) is given by

\[
P_e(\mu) = \begin{cases} 
\frac{\mu_k(1-p_k)}{p_i} & \text{when both parents are heterozygous} \\
\frac{(\mu_i + \mu_j)(1-p_i - p_j)}{p_i + p_j} & \text{when one parent is heterozygous and the other is homozygous} 
\end{cases}
\]

where, as in equation (2), the first probability applies when the child has received the \( i \)-th allele from the biological father, and the second probability applies when the mother and the child are heterozygous for the same pair of alleles (say, \( A_i \) and \( A_j \)). The probability of mutation to the \( i \)-th allele is \( \mu_i \). In general, \( \mu_i \) can be assumed to be \( \mu k \), where \( \mu \) is the overall mutation rate for this locus and \( k \) is the number of allele at the locus.

Averaging over all mother-child genotypes pairs, the probability that the true biological father will be excluded due to mutation of paternal alleles is given by

\[
\bar{P}_e(\mu) = \mu - \frac{\mu k}{k} [2 + a_2 - 3a_3 + 3a_4 - 2a_5]
\]

The distribution of the number of loci showing exclusions where they are caused by mutations of paternal alleles when the alleged father is truly the biological father is also derived by the same theory, where the \( P_e(l) \) values are replaced by equations (10) or (11). Equations (4)–(9) hold, giving a contrast of the distributions of number of loci showing exclusions by non-paternity versus by mutations of paternal alleles.

**Application to Paternity Testing Cases Typed by Minisatellite VNTR and STR Loci**

**STR Loci**

Edwards et al. (19) and Hammond et al. (20) characterized the allele frequency distributions at 13 short tandem repeat (STR) loci in four major human population groups of the USA (Caucasians, Afro-Americans, Hispanics, and Asians). These authors also estimated the locus-specific mutation rates at these loci. Of these two are X-linked loci \([\text{HUMHPR}1\text{[AGAT]}_1, \text{HUMHPR1}[AG/GC]]\), and hence, they are not generally used in parentage testing. While the mutation rates estimates given by these authors are indirect, good agreement of these estimates are found for loci where direct estimates are available (21). Table 1 presents the locus-specific exclusion probabilities at the 11 autosomal STR loci using the Caucasian allele frequencies from these surveys. Two points are worth noting from these computations. First, for each locus, the locus-specific probability of excluding a random male for a specific mother-child pair can vary substantially depending on the genotypes observed in the mother-child pair. This is evident from the differences of minimum and maximum values of exclusion probabilities. Second, even the most likely mother-child genotype pair can have exclusion probabilities that are different from the average for the locus. Of course, all of the exclusion probabilities (minimum, most likely, maximum, or locus-specific average) are increasing functions of the heterozygosity at the locus. In these computations, mutations of the maternal allele are not considered. The same conclusions also hold when the true biological father is excluded because of mutations of the transmitted paternal alleles (see lower part of Table 1). Of course, the absolute paternity exclusion probabilities are all smaller by several orders of magnitude, if mutation is the only source of exclusion.

In Fig. 1 we show the dichotomy of the distributions of the number of loci with respect to which exclusions are to be found when the exclusion is truly due to non-paternity (blank histograms) versus when they are caused by mutation alone (shaded histograms). In panel (a) of Fig. 1, we plotted these distributions with locus-specific average exclusion probabilities \( (P_e(l) \) and \( \bar{P}_e(\mu) \)), while in panel (b), the same computations are done for the most-likely values of \( P_e(l) \) and \( \bar{P}_e(\mu) \) for each locus. For both panels, the dichotomy of the distributions is obvious; by non-paternity we expect exclusions at little more than 5 locus for this battery of 11 markers (mean = 5.28 with s.d. 1.61), while mutations alone will produce far less than one-locus exclusion (mean = 0.004 with s.d. = 0.0192). In fact, with mutations alone, the probability of finding exclusions at two or more loci is less than \( 10^{-7} \). This is true for the average locus-specific exclusion rates, as well as for the most likely genotypes for the mother-child pairs. The low mutation rates at the STR loci, furthermore, suggests that even one-locus exclusion is likely to be due to non-paternity, and may not be caused by mutation at the locus where exclusion is observed.

**RFLP VNTR Loci**

To examine whether mutation rates higher than the ones for the STR loci affect the above conclusions, we also applied the theory to VNTR loci where mutation rates are known to be more frequent than those at STR loci. Southern blot RFLP analysis of seven VNTR loci (D1S7, D2S44, D4S139, D5S110, D10S28, D14S13, and D17S79) constitute the most common battery of RFLP VNTR markers for parentage testing. Since discrete alleles at these loci cannot be identified by the Southern blot RFLP analysis, paternity exclusion probabilities, or paternity index computations at such loci are done by binned definition of alleles. Using the fixed bin allele frequencies of these loci in Caucasians ((22) and Dr. Bruce Budowle, personal communication), we computed the locus-specific average, minimum, maximum, and most-likely paternity exclusion probabilities that are shown in Table 2. Olaisen et al.
(23) and Eisenberg et al. (24) estimated mutation rates at these VNTR loci through direct observations on discordances of parental and children's genotype data. With these estimated mutation rates, we evaluated the rates of excluding a true father (average, minimum, maximum, and most likely) when exclusion is caused by mutations of paternal alleles at these loci. As in the case of the STR loci, here also we observe that the exclusionary chance due to mutations (combined Pe(bt)) for 7 loci for a random mother-child pair is 0.999,996). Data presented in Table 2 also shows that even paternity (combined Pe(C) for 7 loci for a random mother-child pair is 0.0580) is far less likely than those expected due to non-paternity versus exclusion due to mutations (combined Pe(C) for 7 loci for a random mother-child pair is 0.0004 for a random mother-child pair. The appreciable probability of single-locus exclusion due to mutation, mainly arise from the hypermutability of the D1S7 locus. Nevertheless, the chance of exclusions at 2 or more loci is 0.0005 due to mutations, even when the hypermutable D1S7 locus is included in the analysis.

From both sets data (STR and VNTR) we observe that when hypervariable single locus probes are used for exclusion of paternity, exclusions at multiple loci cannot be explained by the high mutation rates at such loci. For VNTR loci, occasional single-locus exclusions may be caused by mutations. Irrespective of the use of specific mother-child genotype data, or the average exclusionary power of the loci, the suggestion of excluding a male from paternity by at least two-locus exclusions will avoid the possibility of excluding a true father. This suggestion is obviously a very conservative approach for the STR loci, since even one-locus exclusion by mutation at such loci can occur rarely in a population (P < 0.0004). Thus, we provide a rigorous statistical validation of the AABB prescription that exclusion of paternity should be based on when two or more loci show evidence of excluding the accused male. The above calculations also justify the avoidance of the hypermutable D1S7 locus for parentage analysis (which is generally the practice in US laboratories).

Discussions and Conclusions

The numerical illustrations, shown above, are for the allele frequency data from the US Caucasian populations. Since allele frequencies at STR and VNTR loci may be different for other populations, one should consider the general applicability of our
CHAKRABORTY AND STIVERS • EFFECT OF MUTATION ON PATERNAL EXCLUSION 675

1.0
0.8
0.6
0.4
0.2
0.0
0 1 2 3 4 5 6 7 8 9 10 11
NUMBER OF LOCI

(a)
(b)

FIG. 1—The distribution of the number of loci showing exclusions due to nonpaternity (blank bars) and due to mutations of parental alleles (shaded bars) for 11 autosomal short tandem repeat loci (see text for a list of loci). Allele frequency data for the loci are from US Caucasian populations (19,20).

1.0
0.8
0.6
0.4
0.2
0.0
0 1 2 3 4 5 6 7 8 9 10 11
NUMBER OF LOCI

(a)
(b)

FIG. 2—The distribution of the number of loci showing exclusions due to nonpaternity (blank bars) and due to mutations of parental alleles (shaded bars) for 7 VNTR loci (see text for a list of loci). Allele frequency data for the loci are from US Caucasian populations (22).

<table>
<thead>
<tr>
<th>Locus</th>
<th>Heterozygosity/mutation rate</th>
<th>Probability of exclusion</th>
<th>Prob. M-C pair for most likely PE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Average</td>
<td>Minimum</td>
<td>Maximum</td>
</tr>
<tr>
<td>D1S7</td>
<td>0.9454</td>
<td>0.8893</td>
<td>0.7123</td>
</tr>
<tr>
<td>D2S44</td>
<td>0.9261</td>
<td>0.8503</td>
<td>0.5914</td>
</tr>
<tr>
<td>D4S139</td>
<td>0.8990</td>
<td>0.7978</td>
<td>0.4597</td>
</tr>
<tr>
<td>D5S110</td>
<td>0.9301</td>
<td>0.8588</td>
<td>0.5944</td>
</tr>
<tr>
<td>D10S28</td>
<td>0.9430</td>
<td>0.8842</td>
<td>0.6889</td>
</tr>
<tr>
<td>D14S13</td>
<td>0.8986</td>
<td>0.8026</td>
<td>0.3944</td>
</tr>
<tr>
<td>D17S79</td>
<td>0.7995</td>
<td>0.6044</td>
<td>0.2632</td>
</tr>
<tr>
<td>D1S7</td>
<td>5.20 × 10⁻²</td>
<td>1.98 × 10⁻²</td>
<td>5.37 × 10⁻¹</td>
</tr>
<tr>
<td>D2S44</td>
<td>3.40 × 10⁻⁴</td>
<td>8.29 × 10⁻⁵</td>
<td>5.36 × 10⁻³</td>
</tr>
<tr>
<td>D4S139</td>
<td>1.60 × 10⁻³</td>
<td>2.69 × 10⁻⁴</td>
<td>2.33 × 10⁻²</td>
</tr>
<tr>
<td>D5S110</td>
<td>8.00 × 10⁻⁴</td>
<td>1.89 × 10⁻³</td>
<td>7.20 × 10⁻²</td>
</tr>
<tr>
<td>D10S28</td>
<td>5.50 × 10⁻⁴</td>
<td>1.94 × 10⁻⁴</td>
<td>3.94 × 10⁻³</td>
</tr>
<tr>
<td>D14S13</td>
<td>7.20 × 10⁻⁴</td>
<td>6.36 × 10⁻⁵</td>
<td>7.44 × 10⁻³</td>
</tr>
<tr>
<td>D17S79</td>
<td>4.80 × 10⁻⁴</td>
<td>3.99 × 10⁻⁵</td>
<td>1.22 × 10⁻²</td>
</tr>
</tbody>
</table>

NOTE: For the upper half, the second column represents the heterozygosity for binned alleles, whereas for the lower half, the mutation rates at the loci are listed (see text for source).
results to parentage testing involving mothers and alleged fathers of differing racial origin. Analysis of allele frequency data at these loci from worldwide populations (25,26) show that the effect of inter-population allele frequency differences can be examined by choosing the most disparate population databases. The results of such an analysis are given in Table 3, where the basic summary statistics (i.e., exclusion probabilities based on at least one-locus and at least two-loci exclusions, caused by non-paternity as well as mutations), are compared for allele frequencies taken from the US Caucasian and US Black populations ((22) and Dr. Bruce Budowle, personal communication for the VNTR loci; and (19, 20) for the STR loci). Comparison of these statistics for these two databases indicate that the conclusion regarding the reliability of two or more exclusions by non-paternity is valid for the Black databases as well. Computations shown in this table also addresses another important issue raised by Pena (9) about the adequacy of single-locus probes alone for parentage analysis. Table 3 data shows that even with the stringency of deciding non-paternity based on two or more loci exclusions, for a battery of 18 VNTR and STR probes, the combined exclusion probability is 0.9999999, which is sufficiently large enough to resolve almost all paternity dispute cases. This probability is hardly affected by allele frequency differences between populations. Even when the hypermutable locus D1S7 is deleted, the combined probability of exclusion based on two or more locus exclusions does not reduce appreciably (0.9999996). Thus, while the multi-locus probes provide unique visual reliability of parentage determinations, combination of several single-locus probes provides almost as much accuracy as that of multi-locus probes. Use of single-locus probes has another distinct advantage, at least in societies where paternity dispute cases are litigated in a customarily adversarial jurisdiction system. The underlying assumptions, namely independence of alleles within and across loci, are readily testable for databases with single-locus probes, while such assumptions are somewhat difficult to test for multilocus probe data. Furthermore, when isolated loci show exclusions, the statistical consideration of whether such exclusions could arise from mutations of true paternal alleles can be based on the paternity index; the logic of this is discussed in the literature (27).

In the above theory we used the locus-specific exclusion probabilities assuming that the mother’s genotypes are also known. Occasionally, paternity testing laboratories face situations where the mother is not available for testing (called deficiency cases). The above theory still holds; the only change required is in the evaluation of locus-specific exclusion probabilities. The average power of exclusion at a multi-allelic codominant autosomal locus, when the mother is not tested, is given by Garber and Morris (28), although an algebraic reduction of their formula (see the last equation of these authors) gives the expression $1 - 4a_2 + 4a_3 - 3a_4 + 2a_5$, where, as in equation (1), $a_i$ represents the sum of $r$-th power of all allele frequencies at a locus. If an expression for $P_e(C)$ is required for a specific genotype of the child (but the mother is untested), equation (2) can be used without any change, since the obligatory paternal alleles consist of the collection of all alleles in the child’s genotype. Analogous changes can be made for loci involving dominance relationships between alleles, although such situations are uncommon for DNA markers. However, DNA markers may occasionally demonstrate non-detectable alleles in RFLP analysis (29–32) with single-locus minisatellite probes, and PCR analysis of microsatellite CA-repeat loci (33,34). These situations are parallel to each locus of the HLA system, for which Chakravarti and Li (12) have given the required formula for exclusion probability. Although they showed that the occurrence of “non-detectable” alleles reduces the exclusion probability, the reduction is not critical for DNA markers for two reasons. First, even at the STR loci the number of segregating alleles are large enough to make the reduction trivial. Second, when the silent alleles are disregarded and all detectable alleles are assumed codominant, the computed allele frequencies become conservative (i.e., larger than expected). These offset the bias caused by over-reporting the exclusion probabilities. For example, when there are 11 equally frequent alleles at a locus, one of which is non-detectable, Chakravarti and Li’s (12) computed average exclusionary power (with at least one-locus showing exclusion) at the locus is 0.775, whereas if each of the codominant equi-frequent 11 alleles are detectable, the average exclusion probability at the locus becomes 0.813.

In summary, the analyzes presented in this work suggest that a battery of single-locus probe RFLP markers involving 4 to 6 VNTR loci, along with PCR analysis of the STR loci, proves sufficiently reliable for parentage analysis. The reliability of inferring non-paternity is further increased when the two-or-more locus exclusion is required to establish non-paternity, since a person who is truly not the father of a child will be rarely excluded on the basis

<table>
<thead>
<tr>
<th>Loci</th>
<th>Number of loci</th>
<th>Population</th>
<th>$P_e(C)$</th>
<th>$P(C)$</th>
<th>Mean (s.d.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>VNTR 7</td>
<td>US Caucasians</td>
<td>0.9999957</td>
<td>0.999838</td>
<td>5.687</td>
<td>(1.004)</td>
</tr>
<tr>
<td>STR 11</td>
<td>US Caucasians</td>
<td>0.9999994</td>
<td>0.99968</td>
<td>6.023</td>
<td>(0.903)</td>
</tr>
<tr>
<td>STR 11</td>
<td>US Blacks</td>
<td>0.999446</td>
<td>0.993162</td>
<td>5.281</td>
<td>(1.605)</td>
</tr>
<tr>
<td>STR 11</td>
<td>US Blacks</td>
<td>0.998957</td>
<td>0.985847</td>
<td>4.976</td>
<td>(1.605)</td>
</tr>
<tr>
<td>Combined 18</td>
<td>US Caucasians</td>
<td>$\approx$ 1.0</td>
<td>0.9999999</td>
<td>10.969</td>
<td>(1.893)</td>
</tr>
<tr>
<td>Combined 18</td>
<td>US Blacks</td>
<td>$\approx$ 1.0</td>
<td>0.9999999</td>
<td>10.999</td>
<td>(1.841)</td>
</tr>
</tbody>
</table>

Note: $P_e(C) =$ probability of exclusion based on at least one-locus exclusions for a random mother-child pair; $P(C) =$ probability of exclusion based on two or more loci exclusions for random mother-child pairs; mean = average number of loci for which the male will be excluded; s.d. = the standard deviation of the number of loci with respect to which the male is excluded.
of only one of such a system of loci, and exclusions at two or more loci are very unlikely to be caused by mutations of paternally transmitted alleles. With the stringency of exclusions at two or more loci, the loss of exclusionary power for the battery of markers is not severe enough to reduce the power of DNA testing. Furthermore, not using the D157 locus for parentage analysis is supported by statistical evidence. Inter-population alleles frequency differences do not effect any of these conclusions.

Acknowledgments

This work was supported by grants GM 41399 and GM 45861 from the National Institutes of Health. We thank Dr. A. Eisenberg for comments and suggestions, and Dr. Bruce Budowle was kind enough to supply the VNTR databases used in this analysis.

References


Address requests for reprints or addition information to Ranajit Chakraborty, Ph.D. Human Genetics Center School of Public Health The University of Texas-Houston Health Science Center P.O. Box 20334 Houston, Texas 77225