Clarification of Statistical Issues Related to the Operation of CODIS

Bruce Budowle[1], John V. Planz[2], Ranajit Chakraborty[3], Thomas F. Callaghan[1], and Arthur J. Eisenberg[2]

1. Laboratory Division, FBI, Quantico, VA

2. Center for Human Identification, UNTHSC, Ft. Worth, TX

3. Center for Genome Information, Department of Environmental Health, University of Cincinnati, College of Medicine, Cincinnati, OH

Introduction

DNA databases, such as CODIS (Combined DNA Index System), have proven to be extremely useful for developing investigative leads for a variety of crimes where there is no known suspect, for linking multiple cases committed by the same individual, and for identifying human remains attributed to a missing person (1-6). Such databases routinely assist law enforcement in solving crimes by more quickly identifying true perpetrators and preventing future crimes by these individuals. In addition, searching a 10-13 short tandem repeat (STR) profile derived from an evidence sample against a large repository of Offender DNA profiles in effect excludes an extremely large number of individuals as being the source of an evidence sample. The multi-tiered structure of CODIS database, as well as its diverse use, make the construction as well as maintenance of this database a complex operation.

As in any complex operation, issues arise that may need clarification. There are three areas where information contained within CODIS and/or the interpretation of candidate matches

obtained through CODIS searches may be misinterpreted or inappropriately applied. These are 1) using the large number of reference offender DNA profiles to improperly validate or refute the current assumptions for assessing the statistical significance of an evidence profile that cannot be excluded as originating from a particular individual(s); 2) using the (relative) frequency of observing a DNA profile in the offender database in lieu of the more appropriate random match probability (or likelihood ratio) for assessing the rarity of a DNA evidence profile in a particular case; and 3) using the current searching algorithms in an attempt to identify perceived highly probable kinship relationships between an evidence sample and a offender DNA profile.

Should the CODIS Database be used to Invalidate Current Statistical Practices of Assessing Significance of a DNA Match in Casework Analysis?

In the U.S., the number of Offender DNA profiles in National DNA Index System (NDIS) has increased to 3,528,903 as of August 2006 (3). Recently, there have been legal attempts to obtain the computerized records of DNA profiles contained within the Offender database to perform empirical studies with the stated goal of assessing the validity of the current practices for generating random match probability estimates (as described in the NRC II Report (7) and generally followed throughout the United States by DNA forensic science practitioners). The premise is that pairwise profile comparisons (with more than 3 million profiles in CODIS there will be greater than 4 trillion comparisons) can generate data to empirically verify or invalidate current statistical practices (8).

These efforts are misguided because the nature and design of CODIS make it an inappropriate and meaningless source of DNA data for assessing the rarity of any specific DNA

profile. No valid analyses using such a repository can be carried out regarding the reliability of current statistical practices because there are duplicate profiles and profiles from relatives (of varied and unknown kinship category) contained within the database, and the population data are heterogeneous. The structure of the CODIS database in actuality is even more complex. For example, the number of relative pairs and the relative composition of different population groups embedded in the database are not readily available, nor can they be precisely estimated. This is so because of anonymity features of the offender profiles to ensure privacy of subjects in the database prescribed by the legal mandates of creation and maintenance of the CODIS system. Therefore, any results obtained from studies assessing the number of observed and expected genotypes at a number of loci from the offender database would be virtually irrelevant and would be misleading for either supporting or refuting current forensic DNA statistical practices. Among the criteria that define the evaluation of the statistical legitimacy of using allele frequency estimates under the assumption of independence (which is not the actual current forensic practice) are that the databases contain little or no duplicate profiles or profiles of close relatives. Therefore, before any legitimate inferences could be drawn from such databases, it would be imperative to remove as many as possible of the duplicates or profiles contributed by close family members. It is widely known that the current offender DNA databases unavoidably contain duplicate profiles and profiles of relatives. Maintaining such profiles in an offender database does not compromise its use for developing investigative leads. However, identifying and resolving (and then removing) these profiles would be a monumental task and would not be readily possible to accomplish. Regardless, such expenditure of resources would gain little insight into the intricacies of forensic statistics.

In the early 1990s there were debates on the statistical practices used to assess the rarity of a DNA profile. One of the most notable criticisms was that the use of allele frequency data from heterogeneous databases is improper; because it would violate the basic assumptions of allelic independence (i.e., Hardy-Weinberg expectations) (9-12). The argument was that major population groups, such as Caucasian and African American, are composed of subgroups, and the allele frequencies (and thus the predicted genotype frequencies) may be quite different among the subgroups within a major population group. This would then cause the assumption that the alleles at a genetic marker are independent to be erroneous within the particular sample population data set. Thus, multiplying allele and genotype frequencies to derive an estimate of the rarity of a DNA profile would be inappropriate. These criticisms have been shown to be of minimal concern when following current forensic practices of compiling broad population group data sets, such as African American and Caucasian, and following the recommendations of the National Research Council (7, 13-20; note: only a few references are provided because they are too numerous to list). However, these concerns about population substructure apply even more critically to the use of an offender DNA database to verify the assumptions of allelic independence (for reasons stated above). Obviously, the compiled profiles in the data set are not from any single major population group; instead the DNA profiles derive from individuals from the various major population groups of unknown proportions. Indeed, CODIS qualifies as one of the most heterogeneous DNA profile databases available. It is comprised of individuals from many different population groups: African American, Asian, Caucasian, Hispanic, Native American, and Oceanian. Such population heterogeneity is an irresolvable issue because population affinity is not maintained with the DNA profile data. This lack of information

prevents the data being parsed into more homogeneous major population groups (i.e., the population partitions used by forensic scientists). Observed departures from expectations would have no relevance for questioning the reliability of statistical practices because 1) the very heterogeneous data set would be expected to violate the basic assumptions of independence; and 2) the data are not separated into populations as used by the DNA forensic community. The issue of having relatives (of unknown kinship types and their unknown proportions) also is a serious consideration; it is well known that inclusion of relatives in databases can cause discernable departure from expectations of multilocus matches based on standard population genetic assumptions, even after adjustments for embedded population substructure (21,22).

Better defined population data, those partitioned into major population categories, which are sufficiently abundant, would provide a more meaningful assessment of the validity of allelic independence and the current statistical practices used in forensic DNA analyses. Many studies on better defined population data sets already have addressed the validity of the basic assumptions of DNA forensic statistics and overwhelmingly support that the current statistical practices are reliable (7, 13-20; note: only a few references are provided because they are too numerous to list).

Even if the above complicated matters are ignored, analyzing the more than 3 million profiles in CODIS to test the assumption that the alleles among the various loci are statistically independent will shed little insight, because the use of the product rule is not strictly applied by United States forensic practitioners. The current practices for estimating the rarity of a DNA profile assume there are violations of the assumption of independence and use the coancestry coefficient described as "θ" in the calculations (7,23). Basically, undetected substructure in a

population is assumed and the value θ is used to adjust the product rule accordingly. Thus, tests demonstrating a violation of independence add little to refute the current practices since the assumption of independence is not made.

When conducting pairwise DNA profile comparisons using offender DNA profiles even if the population data set were relatively homogeneous, it is important to recognize that observing a number of profiles matching at, for example, 9 or more loci is predicted, is directly related to the basic principles used to generate a random match probability estimate, and is grounded in well-established probability theory (note: these profiles are not high stringency matches as defined for direct matching for CODIS purposes; they are exclusions. The discussion here is solely for addressing the current forensic practices for estimating a DNA profile frequency). The principle is the same as that for the well-known "birthday scenario." If asked "assuming the probability of having a birthday is independent and all birthdays are equally distributed, how many people would have to be gathered in a room for odds to be better than even that two people share the same birthday?" most people answer this question with numbers of 180 or more. In fact, only 23 people need to be gathered for odds to be better than even that two will share the same birthday, which could be any of the 365 possible birthdays (24). This result may seem counterintuitive; however, one must appreciate the question being asked. The question is not "what is the chance of a match between a specific person(s) and a specific birthday?" Instead, the question is "what is the chance that any two people might share any birthday, with the birthday not being specified?" The latter is based on the number of pairwise comparisons. When seeking a match of a specific birthday in a group of 23 people, there are only 23 comparisons made. For any two to share a birthday out of 23 people, there are 253 total

pairwise comparisons. The same principle applies to “matches” and “partial matches” in the offender database. In fact, this similarity becomes fairly obvious when one reviews the actual profiles of the “matches” or “partial matches” that had been reported from pairwise comparisons databases. These profiles are arbitrary ones, and not specified or previously targeted. The total number of pairwise comparisons, for example, for the CODIS database with approximately 3,000,000 profiles would be more than 4 trillion. Databases with only 100,000 samples would enable more than a few billion pairwise comparisons. Thus, with so many pairwise comparisons, matches at 9 or more of the 13 loci are expected (25). Observing such partial matches would not call into question the current forensic practices. Indeed, it would support the current statistical practices.

Offender databases are excellent tools for providing investigative leads (1-6). However, they are extremely poor for inferences regarding the assumptions of current forensic statistical practices. The current statistical practices are well established and grounded on numerous population studies readily available in the scientific literature (13-20; note: only a few references are provided because they are too numerous to list). The DNA profiles in offender databases are not well-suited for quality population statistics studies, because they are not properly annotated for population affinity, are not parsed out into major population categories, and duplicates and relatives reside in the database. Any results obtained from population genetic and forensic statistics validity studies using the data in its current form would be meaningless. Lastly, to attempt to resolve any matching or partially matching profiles would require obtaining names of individuals within the database and investigating the relationships of these people when no

criminal investigation is underway; such disclosure of names is beyond the intent of the use of the offender database and is a violation of Federal law.

Statistical Assessment of a Matching Profile Derived from a Database Search

The significance of observing a matching profile found by searching a offender database of "N" individuals versus the estimate of the rarity of a DNA profile for assessing the weight of the evidence in a case was described by the NRC II Report (7) and clarified by the DNA Advisory Board (26). Two different questions distinguish these scenarios: (1) What is the rarity of the DNA profile? and (2) What is the probability of finding such a DNA profile in the database searched? The different questions will produce different answers for the same profile because they address different issues. The first question, which addresses the random match probability, is always of interest to the fact finder and forms the foundation for addressing the second question.

Consider a comparison is made of DNA profiles derived from evidence and a reference sample from a suspect and there is a failure to exclude the individual as the source of the sample. In this case, the weight of the evidence is determined using a statistical assessment, such as the random match probability with modifications (7) to convey how common or rare is the observed DNA profile. In other words, the statistics used are designed to answer the first question, stated above. In contrast, when a DNA profile from a crime scene sample matches a single profile in a felon DNA database, the NRC II Report (7) recommended a formula multiplying N times p, where N is the size of the database and p is the random match probability (RMP). This formula conveys the probability of finding the DNA profile in the database searched and may have

investigative value. This formulation answers the second question stated above and was not intended to be used to estimate the frequency of observing an evidentiary DNA profile in an unrelated randomly selected individual.

Using the database search calculation as a true RMP would give a false impression that more people share the profile than is reasonably possible. Suppose a RMP was estimated as 1 in 1 million and the size of the database searched is N = 1 million; using the NRC II formula Np, the value would be 1. If the value of 1 is then used to convey rarity of the evidence DNA profile to the fact finder, it would erroneously imply that 100% of the population carries the profile. Conveying such is not particularly useful to a jury that is being asked to consider a body of evidence and to decide whether or not a defendant has been proven guilty beyond a reasonable doubt. It grossly understates the value of the evidence when it is being considered in the context of the case.

The Np calculation was not recommended to replace the true RMP, and any assertion of such erroneously cites the language of NRC II Report (7). On page 40 of the NRC II Report the following is stated: "If one wishes to describe the impact of the DNA evidence under the hypothesis that the source of the evidence sample is someone in the database, then the likelihood ratio should be divided by N" (7,27). Thus, the formula was not intended to supersede the random match probability estimate; it addresses a different issue.

Familial Searches and Partial Matches are not Equivalent

The identification of matching DNA profiles between a forensic sample and a felon sample in CODIS is referred to as a "high stringency match." The software currently in use in

CODIS for high stringency matching requires that the genotypes in the files be identical at all loci between the two samples. Because of the potential for allele dropout (28,29) and limitations in identifying some obligate alleles in mixtures (for entry into CODIS), a "moderate stringency match" search also is employed using CODIS software. For a moderate stringency candidate match, the following profile associations per locus, where A_iA_j represents a heterozygote and A_iA_i represents a homozygote, are recognized:

1. $A_iA_j = A_iA_i$ or A_jA_j

2. $A_iA_i = A_iA_j$

For example, a 15,16 type can be associated with a 15,15 or a 16,16; and a 16,16 type can be associated with any heterozygote that contains allele 16 (such as 16,17; 16,18; 16,19...). As few as one locus up to all loci in the two profiles can be moderate for the profiles to be designated as a moderate stringency candidate match. The following are a few examples of two profiles that would be designated as moderate stringency matches:

Scenario I - although an extreme example, all loci in one profile are heterozygous and in the corresponding profile all loci are homozygous (or pseudohomozygous).

Locus	Offender	Forensic	Forensic[a]
D3S1358	15,16	15,15	15
vWA	17,18	17,17	17
FGA	21,22	22,22	22
D8S1179	13,14	13,13	13
D21S11	29,30	30,30	30
D18S51	14,17	14,14	14

D5S818	11,12	11,11	11
D13S317	11,12	11,11	11
D7S820	10,11	10,10	10
CSF1PO	11,12	11,11	11
TPOX	8,11	8,8	8
TH01	6,9.3	9.3,9.3	9.3
D16S539	11,12	12,12	12

a) In the database, a locus that is either a homozygote or a pseudohomozygote is displayed with only one allele

Scenario 2 - approximately half the loci in one profile are heterozygous and these correspond to homozygous (or pseudohomozygous) loci in the other profile.

Locus	Offender	Forensic	Offender[a]	Forensic[a]
D3S1358	15,16	15,15	15,16	15
vWA	17,18	17,17	17,18	17
FGA	21,22	22,22	21,22	22
D8S1179	13,14	13,13	13,14	13
D21S11	29,30	30,30	29,30	30
D18S51	14,17	14,14	14,17	14
D5S818	11,12	11,11	11,12	11
D13S317	11,11	11,12	11	11,12
D7S820	10,10	10,11	10	10,11
CSF1PO	11,11	11,12	11	11,12
TPOX	8,8	8,11	8	8,11
TH01	9.3,9.3	6,9.3	9.3	6,9.3
D16S539	12,12	11,12	12	11,12

a) In the database, a locus that is either a homozygote or a pseudohomozygote is displayed with only one allele

In both Scenarios 1 and 2 all loci between the two profiles meet the moderate stringency criterion; i.e., one locus in one profile is heterozygous and the same locus in the other profile is homozygous and the two profiles at that locus share a common allele.

Scenario 3 - also is considered a moderate stringency candidate match. Some loci (the first four listed) are high stringency matches and the rest meet a moderate stringency match criterion. In this scenario 9 loci are a moderate stringency match.

Locus	Offender	Forensic	Offender[a]	Forensic[a]
D3S1358	15,16	15,16	15,16	15,16
vWA	17,18	17,18	17,18	17,18
FGA	21,22	21,22	21,22	21,22
D8S1179	13,14	13,14	13,14	13,14
D21S11	29,30	30,30	29,30	30
D18S51	14,17	14,14	14,17	14
D5S818	11,12	11,11	11,12	11
D13S317	11,12	11,11	11,12	11
D7S820	10,11	10,10	10,11	10
CSF1PO	11,11	11,12	11	11,12
TPOX	8,8	8,11	8	8,11
TH01	9.3,9.3	6,9.3	9.3	6,9.3
D16S539	12,12	11,12	12	11,12

a) In the database, a locus that is either a homozygote or a pseudohomozygote is displayed with only one allele

If only one locus showed a moderate stringency and all other loci were at a high stringency match, the two profiles would still be classified as a moderate stringency candidate match. However, such a scenario would likely be considered a high stringency match with one locus demonstrating allele drop out.

"Partial matches," as shown in the scenarios above, are a very small subset of moderate stringency candidate matches. Because CODIS is designed to facilitate obtaining direct matches, partial matches constitute exclusions. However, some may seek to use moderate stringency search algorithms with hopes of finding investigative leads to identify the sources of evidentiary material through kinship or familial inferences. The premise is that close relatives, i.e., parent-offspring and sib-sib, would share more alleles in common than unrelated individuals. Therefore,

when there is no high stringency match obtained via a CODIS search, a moderate stringency candidate match may associate an evidence profile to a relative of the true source of the evidence profile. Indeed, a moderate stringency match does meet the general criterion for a potential relative (often favoring a parent-offspring relationship) being the source, because there is one allele in common at all loci. However, caution should be taken before proceeding with such a proposition.

One may believe that a moderate stringency match is strong evidence for there being a kinship relationship. To support this contention, a kinship index (KI) might be calculated for the two profiles. In Scenarios 1 and 2, the KI favoring parentage or full sibship versus unrelated are 1135 and 4.23, respectively, based on Caucasian population data. For Scenario 3, the KI favoring parentage or full sibship versus unrelated are 1792 and 358, respectively, based on Caucasian population data. Using African American population data, however, the results are markedly different. In Scenarios 1 and 2, the KI favoring parentage or full sibship versus unrelated are 135,500 and 212, respectively. For Scenario 3, the KI favoring parentage or full sibship versus unrelated are 180,500 and 2,155, respectively. The reason for the larger KIs in African American is that the alleles in the scenarios are the most common alleles in the Caucasian population.

These KI results seem compelling. However, they are taken out of context. There was no consideration of the number of samples searched (i.e., greater than 3 million people). With that many profiles, there is a high probability that two unrelated people could have moderate stringency matching profiles. Recent studies show that in some cases a number of unrelated men cannot be excluded as a potential biological father of a child and at times the Paternity Index (PI) is greater than that for the true biological father. A test familial database was created from

paternity cases (from the Center for Human Identification, UNTHSC) containing 38,244 individuals (12,836 biological fathers, 11,113 biological mothers, and 14,295 children to serve as unidentified person samples). Searches were performed to detect associations between individuals that could not be excluded as a possible parent-offspring relationship. The searches required a single allele match at each of the 13 STR loci, i.e., the equivalent to a low stringency search. The search returned 174,807 matches meeting the requirement of sharing at least one allele at all loci which is consistent with a possible parent–offspring relationship. Of these matches, 21,805 represent the true parent-offspring matches and 152,908 (note: that these numbers do not sum to exactly 174, 807 because a few individuals are entered more than once under different paternity cases) are fortuitous parent-offspring matches in which a false father or mother was associated with an individual classified as a child. Over 400 of these fortuitous matches had paternity indexes (PIs) in excess of 1 million, with the greatest false parentage association yielding a PI of 914 million. A sibship search would be less stringent, as the requirement of a shared allele at each locus would not be required. On average approximately 25% of the loci between full sibs would be expected not to share any alleles.

Poetsch et al (30) performed a similar study with a smaller data set (based on 13-15 STR loci) and found similar results. They compared 336 children with 348 unrelated men and observed 26 pairs in which there were no non-excluding STR loci. This empirical value was close to the predicted number of fortuitous matches (23.4 or approximately 0.02% of comparisons). Of the 26 pairs, 19 of them had identified the biological father (the other 7 pairs were from cases where the alleged father was excluded). In these 19 pairs, three unrelated men had PIs higher than the true biological father. Using this empirical observation, the number of

fortuitous non-excluded profiles in the CODIS database would be approximately 600 and 95 pairs would have greater PIs than true fathers (note that this is an extrapolation assuming a homogeneous database, a criterion that CODIS does not meet; therefore the values are meant only to provide context). With fewer than 13 loci, the number of non-excluded unrelated profiles would increase substantially. If a sibling relationship also is considered, the number of potential profiles would be immense. Currently, the participants of CODIS cannot address so many candidate matches and remain functional.

In addition to the fact that the KIs may be misleading, reliance on a moderate stringency match as providing strong evidence for a kinship relationship (again favoring the parent-offspring scenario with one allele in common at each locus) ignores the many candidate profiles that would meet the criterion of a potential kinship relationship are not identified by current CODIS searches. Consider the D13S317 locus type 12,13. This type could only moderately match types 12,12 and 13,13. Thus, types 8,12; 9,12; 10,12; 11,12; 12,14; 12,15; 8,13; 9,13; 10,13; 11,13; 13,14; and 13,15 would not be included as candidate moderate stringency matches. Because on average the majority of loci in a profile are heterozygous, the overwhelming majority of profiles that meet the criterion of sharing an allele in common at each locus is not identified.

Another exacerbating factor in missing potential relatives is that the moderate stringency search algorithms do not allow for the phenomenon of mutation. If a mutation occurs for the allele transmitted from a true biological parent and its offspring, it is possible they will show no alleles in common at that locus. Even though 12 of the 13 core CODIS loci would meet the moderate stringency criterion, the 13th locus would eliminate the profile as a candidate and it would never be registered. If a mutation did occur and a one or two locus mismatch was allowed,

there will be an extreme number of fortuitous matches with PIs greater than the true biological father.

Lastly, a one allele per locus matching criterion (of which a moderate stringency search finds a very limited subset) favors a parent-offspring scenario over a sib-sib relationship. Again, many potential kinship relationships that may be plausible would not be identified under the current CODIS searching algorithms.

Conclusion

A number of issues have arisen due to the application of CODIS for developing investigative leads or because of legal proceedings. This paper addresses several topics that required some elucidation. First, CODIS is an excellent investigative tool, but not a good source for validating statistical practices for estimating DNA profile frequencies, because of the heterogeneous nature of the database and the existence of an unknown number of duplicates and relatives in the database. Voluminous already published population studies demonstrate the validity of current forensic practices for estimating the rarity of a DNA profile. Second, the RMP (or LR) is always meaningful in describing the rarity of a DNA profile and forms the basis for estimating the probability of observing a specified DNA profile in a database of "N" individuals. The RMP and the database search probability address different questions. Therefore, different values are expected. Such differences do not infer that there is any discord. Third, using a moderate stringency candidate match as an investigative lead for identifying a close relative in most cases is not likely to be successful. CODIS currently is not designed for the purposes of identifying suspects through kinship analysis. However, the attempts to apply current partial

matches to familial searches suggest that discussion is needed on proper requirements and/or practices for familial searching. It is important to frame criteria to enable successful and efficient familial searching of forensic DNA databases. Topics in this regard include evaluating existing or developing effective familial searching software; use of additional autosomal loci and lineage based genetic markers, such as Y STRs, for more efficient searching and resolving fortuitous candidates; and establishing thresholds for selecting candidate matches from familial search candidate lists.

Acknowledgments

We would like to thank Frederick Bieber and Rock Harmon for their thoughtful comments.

This is publication number 07-01 of the Laboratory Division of the Federal Bureau of Investigation. Names of commercial manufacturers are provided for identification only, and inclusion does not imply endorsement by the Federal Bureau of Investigation.

References

1. Budowle, B., Moretti, T.R., Niezgoda, S.J., and Brown, B.L. (1998) CODIS and PCR-based short tandem repeat loci: Law enforcement tools. In: Second European Symposium on Human Identification 1998, Promega Corporation, Madison, Wisconsin pp 73-88.

2. Budowle, B. and Brown, B.L. (2001) The use of DNA typing for forensic identification. Forensica 1(1):9-37.

3. CODIS. Available at http://www.fbi.gov/hq/lab/codis/index1.htm.

4. Martin P.D. (2004) National DNA databases: practice and practicability. A forum for discussion. Prog. Forensic Genet. 10:1–8.

5. Werrett D.J. (1997) The national DNA database. Forensic Sci. Int. 88:33–42.

6. Werrett, D.J. and Sparkes, R. (1998) 300 matches per week -- the effectiveness and future development of DNA intelligence databases. In: Ninth International Symposium on Human Identification. Genetic Identity Conference Proceedings, available at http://www.promega.com/geneticidproc/ussymp9proc/.

7. National Research Council Report (1996) The Evaluation of Forensic DNA Evidence, Washington, D.C., National Academy Press.

8. People of the State of Illinois v Juan Luna, In The Circuit Court Of Cook County, County Department - Criminal Division, No. 02 CR 15430, 2006.

9. Lander, E.S. (1989) Population genetic considerations in the forensic use of DNA typing. In Banbury Report 32: DNA Technology and Forensic Science. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, pp. 143-156.

10. Lander, E.S. (1991) Invited editorial: Research on DNA catching up with courtroom application. Amer. J. Hum. Genet. 48:819-823.

11. Lewontin, R.C. and Hartl, D.L. (1991) Population genetics in forensic DNA typing. Science 254:1745-1750.

12. Shields, W.M. (1992) Problems and solutions associated with matching and generating inclusion probabilities. In: Proceedings of The Third International Symposium on Human Identification, Scottsdale, AZ, Promega Corporation, pp 1-50.

13. Alshamali, F., Alkhayat, A., Budowle, B., and Watson, N.D. (2005) STR population diversity in nine ethnic populations living in Dubai. Forens. Sci. Int. 152:267-279.

14. Ang, H.C., Sornarajah, R., Lim, S.E.S., Syn, C.K.C., Tan-Siew, W.F., Chow, S.T., and Budowle, B. (2005) STR data for the 13 CODIS loci in Singapore Malays. Forensic Sci. Int. 148(2-3):243-245.

15. Budowle, B., Chidambaram, A., Strickland, L., Beheim, C.W., Taft, G.M., and Chakraborty, R. (2002) Population data at 13 STR loci for three Native Alaska population groups. Forens. Sci. Int. 129(1):51-57.

16. Budowle, B., Nhari, L.T., Moretti, T.R., Kanoyangwa, S.B., Masuka, E., Defenbaugh, D.A., and Smerick, J.B. (1997) Zimbabwe black population data on the six short tandem repeat loci - CSF1PO, TPOX, THO1, D3S1358, VWA, and FGA. Forens. Sci. Int. 90:215-221.

17. Budowle, B., Shea, B., Niezgoda, S., and Chakraborty, R. (2001) CODIS STR loci data from 41 sample populations. J. Forens. Sci. 46(3):453-489.

18. Henke, L., Aaspollu, A., Biondo, R., Budowle, B., Drobnic, K., van Eede, P.H., Felske-Zech, H., Fernandez de Simon, L., Garafano, L., Gehrig, C., Luckenbach, C., Malik, N., Muche, M., Parson, W., Primorac, D., Schneider, P.M., Thomson, J., and Vanek, D. (2003) Evaluation of the STR typing kit PowerPlex™ 16 with respect to technical performance and population genetics: a multicenter study. In: Progress in Forensic Genetics 9,(Brinkmann, B., and Carracedo, A., eds.), Elsevier, Amsterdam, pp. 789-794.

19. Lim, S.E.S., Tan-Siew, W.F., Syn, C.K.C., Ang, H.C., Chow, S.T., and Budowle, B. (2005) Genetic data for the 13 CODIS STR loci in Singapore Indians. Forensic Sci. Int. 148(1):65-67.

20. Syn, C.K.C., Chuah, S.Y., Ang, H.C., Lim, S.E.S., Tan-Siew, W.F., Chow, S.T., and Budowle, B. (2005) Genetic data for the 13 CODIS STR loci in Singapore Chinese. Forensic Sci. Int. 152:285-288.

21. Teshima, K.M., Lee, H.S., and Chakraborty, R. (2002) Population substructure effects on the extent of genotype sharing: A tool for detecting the presence of relatives in databases. Amer. J. Hum. Genet. 71:A570.

22. Lee, H., Hwang, J.J., and Chakraborty, R. (2003) Inclusion of relatives in a sample produces intra-and inter-locus association of alleles without biasing allele frequency estimates. Amer. J. Hum Genet. 73(5):619.

23. Weir, B.S. and Cockerham, C.C. (1984) Estimating F-statistics for the analysis of population structure. Evolution 38: 1358-1370.

24. Birthday scenario. http://en.wikipedia.org/wiki/Birthday_paradox

25. Weir, B.S. (2004) Matching and partially-matching DNA profiles. J. Forensic Sci. 49(5):1009-1014.

26. DNA Advisory Board (2000) Statistical and population genetics issues affecting the evaluation of the frequency of occurrence of DNA profiles calculated from pertinent population database(s). Forensic Sci. Comm. 2(3) JULY 2000, available at http://lab.fbinet.fbi/library/fsc/backissu/july2000/index.htm.

27. Crow, J.F. (1998) The 1996 NRC Report: Another look. In: Ninth International Symposium on Human Identification. Genetic Identity Conference Proceedings, available at http://www.promega.com/geneticidproc/ussymp9proc/.

28. Budowle, B., Masibay, A., Anderson, S.J., Barna, C., Biega, L., Brenneke, S., Brown, B.L., Cramer, J., DeGroot, G.A., Douglas, D., Duceman, B., Eastman, A., Giles, R., Hamill, J., Haase, D.J., Janssen, D.W., Kupferschmid, T.D., Lawton, T., Lemire, C., Llewellyn, B., Moretti, T.,

Neves, J., Palaski, C., Schueler, S., Sgueglia, J., Sprecher, C., Tomsey, C., Yet, D. (2001) STR primer concordance study. Forens. Sci. Int. 124:47-54.

29. Budowle, B. and Sprecher, C.J. (2001) Concordance study on population database samples using the PowerPlex™ 16 Kit and AmpFlSTR® Profiler Plus™ Kit and AmpFlSTR® COfiler™ Kit. J. Forens. Sci. 46(3):637-641.

30. Poetsch, M., Ludcke, C., Repenning, A., Lutz, F., Malyusz, V., Simeoni, E., Lignitz, E., Oehmichen, N., and von Wurmb-Schwark, N. (2006) The problem of single parent/child paternity analysis – practical results involving 336 children and 348 unrelated men. Forens. Sci. Int. 159:98-103.