Recommendations from the SWGDAM Ad Hoc Working Group on Familial Searching

Executive Summary

To inform future planning with respect to familial searching at the national level, the Scientific Working Group on DNA Analysis Methods (SWGDAM) empanelled scientific experts and practitioners independent of the Federal Bureau of Investigation (FBI) to form the Familial Searching Ad Hoc Working Group. Within the DNA community, familial searching is understood to be a deliberate search of a DNA database conducted for the intended purpose of potentially identifying close biological relatives of the source of the unknown forensic profile obtained from crime scene evidence. While a handful of states now conduct familial searches of their State DNA databases, this type of searching is not currently performed at the National DNA Index System (NDIS).

The Working Group conducted its evaluation in accordance with the following assumptions: (1) familial searching at NDIS would utilize the 13 CODIS core loci for the present time and only DNA records with the complete 13 core loci would be searched; (2) only single source forensic samples would be searched; and (3) the offender/arrestee DNA records at NDIS would not contain Y-STR data. In its review, the Working Group emphasized that a relative of the perpetrator would need to be in the database in order to be identified in the search and even then, the relative may not be detected in the list of candidates generated by the familial search.

The Working Group considered the relevant scientific literature and designed/conducted studies with the participation of three states that currently perform familial searching. The studies included familial searches of artificial profiles against the State offender databases of California, Virginia, and Wyoming as well as modeled databases, to determine rankings that will occur and the length of the candidate list that would need to be reviewed in order to reach the true relative. The combination of the three State DNA Index System (SDIS) databases allowed the committee to experiment with a database that is approximately one-fifth the size of NDIS. Additionally, the combined SDIS databases were rescaled to 10 million profiles and simulated 10 million person offender databases were searched to more directly test databases the size of NDIS. More importantly, the searches were performed with a relative present in the database and thus the results represent the most positive outcome for familial searches.

Recommendations

The Ad Hoc Working Group provided the following recommendations to the questions posed by the FBI's CODIS Unit through SWGDAM. The Ad Hoc Working Group recognizes that the operational goal of familial searching is to minimize the false positive associations for a tolerable

level of false negatives. In other words, the Working Group acknowledges that some true relatives may be missed when minimizing investigations of unrelated individuals whose presence on the list is adventitious. The Working Group also considered that as the size of the database increases, familial search comparisons will have more opportunities for false associations given the large ratio of unrelated to related individuals.

Is kinship matching (producing a ranked list of candidates based upon kinship statistics) more efficient at detecting relatives than counting the number of alleles shared?

Recommendation 1.1: The effectiveness of a familial searching method can be evaluated by its ability to determine true first-degree relatives (i.e., parent, child, full siblings) among the DNA records of unrelated individuals exhibiting similar genetic profiles by chance alone. Because the kinship Likelihood Ratio (LR) approach takes into account the allele frequencies as well as the number of alleles that match, it is more effective in identifying the true relative than counting the number of shared alleles. The Familial Searching Ad Hoc Working Group endorses and recommends the use of kinship LRs as the preferred method for familial searching.

If the number of false positives generated prior to finding a true match is inversely related to the likelihood of sibship, does this suggest that many true siblings would not be found in large databases? If so, is there an optimal database size range for performing familial searching?
Can we establish the number of ranked candidates (kinship matching) that would require investigation in order to ascertain a "true" relative when searching a database with over 10 million DNA profiles? If so, what is that number?

Recommendation 1.2: The combined SDIS values from the study demonstrate that there is approximately a 1 in 5 chance that a first-degree relative will be ranked #1. Actual NDIS values would be expected to be lower. Since the true relative is not always ranked as the #1 candidate in familial searches, especially as the size of the database being searched increases, ranked lists of offenders should be reviewed. Therefore, all investigative efforts should not be focused on the #1 ranked candidate.

Recommendation 1.3: To the extent that additional filters or thresholds are available, the ranked lists can be further distilled to reduce the number of false positive associations. Possible filters include testing additional genetic sites (typically includes autosomal STR or Y-STR loci) and evaluation of metadata (e.g., age, gender, population group) and/or geography of the offender and the crime.

Recommendation 1.4: It would be difficult to establish a threshold ranking for the review of ranked lists for familial searching at the national level where the offender database contains over

10 million offender DNA records. A threshold ranking could be established at which there is a 50% or greater probability of including a true first-degree relative within the ranked list. However, the additional filters of metadata and geography are not available at the national level since this data is not collected and stored at NDIS. Moreover, the availability of Y-STR testing for filtering would also be left to each individual state as validation and resources permit. Considering the current assumptions, the lack of a threshold ranking, and inconsistent availability of additional filters to NDIS participants, the Familial Searching Ad Hoc Working Group does not recommend performing routine familial searching at the national level at this time. Once the additional CODIS core loci are implemented, the Working Group recommends that these issues be reconsidered. Meanwhile, the data accompanying these recommendations may provide guidance to states considering routine familial searching based upon the kinship LRs.

Introduction

In April, 2012, the SWGDAM Chair empanelled a group of scientific experts and practitioners to look into familial searching and provide input on three specific questions posed by the Federal Bureau of Investigation (FBI) Laboratory's Combined DNA Index System (CODIS) Unit. These experts include two scientists from academia, three experts from forensic DNA laboratories that are engaged in familial searching, and the Vice Chair of SWGDAM.[1]

At this inaugural meeting, the FBI's CODIS Unit provided background information on the SWGDAM Partial Match recommendations as well as the information on familial searching available at the CODIS Unit's Internet web page.[2] The FBI wanted a group separate and independent of its agency to provide the scientific perspective on questions that would inform future planning for familial searching on a national scale.

The Working Group Members were briefed by the CODIS Unit on CODIS operations and searching. The three questions posed to the Ad Hoc Working Group were as follows:

- Is kinship matching (producing a ranked list of candidates based upon kinship statistics) more efficient at detecting relatives than counting the number of alleles shared?
- If the number of false positives generated prior to finding a true match is inversely related to the likelihood of sibship, does this suggest that many true siblings would not be found in large databases? If so, is there an optimal database size range for performing familial searching?
- Can we establish the number of ranked candidates (kinship matching) that would require investigation in order to ascertain a "true" relative when searching a database with over 10 million DNA profiles? If so, what is that number?

Additionally, the following assumptions were proposed with the understanding that the Ad Hoc Group may modify these:

- National searching would require using the 13 CODIS core loci

- Only single source forensic samples would be searched

- Only complete 13 CODIS core loci profiles would be searched[3]

- No Y-STR data is associated with the offender/arrestee data at NDIS

Background

As part of the law enforcement DNA databases and CODIS program, a standardized set of loci was identified by the Federal Bureau of Investigation to ensure that the DNA records contributed to the national, state and local DNA databases would be compatible. The CODIS software used by the federal, state and local laboratories that participate in the National DNA Index System searches DNA records that contain a numeric representation for the two alleles at each of the 13 CODIS core loci: CSF1PO, FGA, TH01, TPOX, vWA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, and D21S11.[4]

At the national level (NDIS), the Forensic Index of crime scene DNA records is searched against the Convicted Offender and Arrestee Indexes containing the DNA records of known offenders and arrestees collected in accordance with Federal and State laws. These DNA records are searched at moderate stringency in which, to return a candidate match, each of the alleles from the forensic and offender/arrestee DNA records must match but the target forensic DNA record and offender/arrestee DNA record may have a different number of alleles.

What is familial searching? For the purposes of our discussions, the definition of familial searching adopted by the FBI was followed:

> "Familial searching is an additional search of a law enforcement DNA database conducted after a routine search has been completed and no profile matches are identified during the process. Unlike a routine database search which may spontaneously yield partial match profiles, familial searching is a deliberate search of a DNA database conducted for the intended purpose of potentially identifying close biological relatives to the unknown forensic profile obtained from crime scene evidence. Familial searching is based on the concept that first-order relatives, such as siblings or parent/child relationships, will have more genetic data in common than unrelated individuals. Practically speaking, familial searching would only be performed if the comparison of the forensic DNA profile with the known offender/arrestee DNA profiles has not identified any matches to any of the offenders/arrestees."[5]

As noted in the description above, familial searching is only performed when the routine search of the crime scene (forensic) DNA record against the known offender/arrestee DNA records does not yield any direct matches. In those instances, some jurisdictions may perform a second search, using specially designed software, to identify if any of the offender/arrestee DNA records may be from a relative of the putative perpetrator (represented by the crime scene or forensic DNA record).[6] Another important consideration for familial searching is that a relative of the perpetrator must be in the database in order to be identified in the search and even then, the relative may not be produced in the list of candidates generated by the familial search. While familial searching may not be appropriate for all cases that do not yield a direct match in NDIS, a

handful of jurisdictions in the United States are now performing familial searches in cases for which no suspects have been identified through the routine DNA database searches. Familial searching is not currently performed at the national level (NDIS).

During the 112th Congressional Session, Congressman Schiff introduced a bill entitled "Utilizing DNA Technology to Solve Cold Cases Act of 2011" (H.R. 3361) that would have required the Attorney General to adopt policies and procedures for the FBI to conduct familial searches in Federal investigations; for States to request familial searches; and to protect the privacy interests of those persons identified in familial searches. The bill also specified criteria for conducting familial searches as well as for the state requests for familial searches. The House Subcommittee on Crime, Terrorism and Homeland Security held hearings on the bill on April 25, 2012.[7] While this proposal was not enacted before the end of the 112th Session, the Conference Report (House Report 112-284[8]) for the 2012 Federal Fiscal Year Appropriations Bill contained the following provision:

> "DNA programs – The conferees encourage the FBI to undertake activities to facilitate familial DNA searches of the Combined DNA Index System database of convicted offenders and work with the National DNA Index System (NDIS) Procedures Board to consider the establishment of procedures allowing familial searches only for serious violent and sexual crimes where other investigative leads have been exhausted. The procedures should provide appropriate protections for the privacy rights of those in the NDIS database."

At the March 2013 Congressional Hearing on the FBI's budget, Congressman Schiff stated his intention to again introduce legislation on familial searching and questioned the FBI Director on the need for such legislation and the FBI's progress in implementing familial searching at the national level.[9] Director Mueller reiterated the need for statutory authorization to conduct familial searching at the national level and noted that technical/scientific recommendations on familial searching are expected during the summer.

Familial Search Studies

The Ad Hoc Working Group reviewed the published literature, including the California Department of Justice study performed during the planning and implementation of their familial searching program. California performed a validation study of their proposed likelihood ratio (LR) approach to familial searching. California tested their State DNA Index System (SDIS) offender database for rankings that will occur and the length of the candidate list that would need to be reviewed in order to reach the authentic relative. California describes their searching methodology as using "STR-based LR calculations, analytical thresholds, and subsequent Y-STR analyses to perform familial searches on an ~1 million profile offender database, attempting to identify potential first-degree relatives to a test/evidence sample while severely limiting the

possibility of generating a false lead. Due to practical limitations, we accepted as given that our system would not identify every relative in the database. In our examination of 100 artificially created families, the simple application of our analytical thresholds removed approximately 99.96% of unrelated offenders from further consideration."[10] At the time of the study, approximately 71% of the California felon database was typed for the 15 loci of the Identifiler® STR kit, the remaining 29% was typed for just the 13 CODIS core loci. California's familial search procedure included 96% of the fathers and 72% of the full siblings when the relative was among offenders typed at 15 loci. When the relative was among the smaller pool of offenders typed with only the 13 CODIS core loci, California's procedure included 93% of the fathers and 61% of the full siblings. These inclusion rates were obtained by examining the top ranked comparisons up to a maximum of 168 offenders (a practical limit established for follow-up Y-STR testing of highly ranked offender samples based upon Cal DOJ Data Bank protocols).

To assist the Ad Hoc Working Group in its discussions, a three-part study consisting of test familial searches was performed to examine the efficacy of familial searching a database of 10 million DNA profiles (the approximate number of the offender/arrestee DNA records in NDIS). The study followed the assumptions provided to the Ad Hoc Working Group (13 CODIS core loci, single source, no Y-STR or additional genetic data). This three-part study was comprised of searches of actual offender databases, as well as modeled databases.

1. Because of the complexity of a real database (unknown population structure, relatives, duplicates), tests using actual databases were performed to look for practical deviations from the models. Such tests could not be performed on NDIS, so 200 artificial evidence profiles were searched against the SDIS offender databases of California, Virginia and Wyoming, three states that have familial searching programs. The database sizes of these states provide a range that is representative of the variation seen across the nation, so search results were evaluated separately and in combination. The combination of the three SDIS databases allowed the committee to experiment with a database that is approximately one-fifth the size of NDIS.

2. The above study was replicated using the same set of 200 artificial evidence profiles searched against simulated 2.13 million person offender databases. This enabled comparisons between the three-state SDIS data and the results expected from theoretical models.

3. The second study was replicated using simulated 10 million person offender databases to more directly test searches of NDIS. For this study, the number of artificial evidence search profiles was expanded to 1,000.

Additional details of the studies are provided in the Appendix.

Ad Hoc Working Group's Responses to the Three Questions Posed by the FBI

> **Is kinship matching (producing a ranked list of candidates based upon kinship statistics) more efficient at detecting relatives than counting the number of alleles shared?**

Among the methods that could be used to perform familial searching are kinship matching, counting shared alleles, or a combination of both. Kinship matching has been described as "the comparison of the joint probabilities of the forensic and offender profiles given that the donors are related versus they are unrelated."[11] The counting of shared alleles compares the number of alleles that are the same between the forensic DNA record and the offender/arrestee DNA record. Jurisdictions that count the number of shared alleles for purposes of familial searching generally establish a minimum number of the shared alleles (such as 20 out of the 26 alleles) that will initiate review as a possible relative.

What will determine the effectiveness of the familial searching method will be its ability to determine the true relative among the DNA records of unrelated individuals exhibiting similar genetic profiles by chance alone. For example, at any given locus, full siblings may not share any locus types. Similarly, two unrelated individuals may share the same locus types by chance alone.[12]

Based upon the literature[13] and state experiences with familial searching, the Ad Hoc Working Group endorses and recommends the use of kinship LRs as the preferred method for familial searching. Considering that the number of alleles evaluated is limited to the 13 CODIS core loci for the immediate future, the counting of shared alleles, without additional filters, would rarely provide meaningful information for familial searches of a database that contains over 10 million offender/arrestee DNA records. Because the kinship LR approach takes into account the allele frequencies as well as the number of alleles that match, it is more effective in identifying the true relative than the method of only counting the number of shared alleles.

> **If the number of false positives generated prior to finding a true match is inversely related to the likelihood of sibship, does this suggest that many true siblings would not be found in large databases? If so, is there an optimal database size range for performing familial searching?**

> **Can we establish the number of ranked candidates (kinship matching) that would require investigation in order to ascertain a "true" relative when searching a database with over 10 million DNA profiles? If so, what is that number?**

The questions on optimal database size and ranked lists are interrelated and the discussion that follows is applicable to both issues.

The rankings for the searches performed by selected states that are currently performing familial searching (California, Virginia and Wyoming) are contained in Table 1. These familial searches were performed with a relative present in the database and thus the results represent the most positive outcome for such searches.

According to the Bureau of Justice Statistics' (BJS) report on inmates "an estimated 46% of jail inmates in 2002 had a family member who had been incarcerated in a prison or jail. About 31% had a brother; 19% a father; 9% a sister; and 7% a mother who had been incarcerated."[14] Therefore, unlike the studies performed by the Ad Hoc Working Group which deliberately placed one or more relatives in the database, a real-life perpetrator may not have a relative in the actual SDIS or NDIS databases. If the percentages reported by BJS for inmates are representative of current offender databases, there will be a close (first-degree) relative in the database approximately half of the time. The rankings contained in Table 1 (which represent searches when a relative is present in the database) would then be lowered by 50% or more to be consistent with the BJS findings.

Table 1: Summary of Rankings by State/NDIS
All Results Assume the Presence of a Relative in the Database

Database	Relationship	% Ranked #1	% Ranked #1-10	% Ranked #1-25	% Ranked #1-50	% Ranked #1-100
WY (19,300)[a]	Parent-Child	72.8	97.9	100	100	100
	Full Siblings	56.8	80	87.5	91.8	94.8
	Two Relatives	85.7	99.2	99.9	100	100
VA (356,000)[a]	Parent-Child	35.8	71.5	84.5	91.1	97.2
	Full Siblings	35	57	63.5	69.8	77.0
	Two Relatives	55.3	86.2	92.8	96.6	99.0
CA (1,780,000)[a]	Parent-Child	18.5	48	60.4	71.4	82.1
	Full Siblings	22.8	40.5	50.5	55.8	63.0
	Two Relatives	33.7	66.9	79.1	86.1	92.1
Comb. SDIS (2,130,000)[a]	Parent-Child	16.9	45.2	57.6	68.1	79.6
	Full Siblings	21.8	38.2	47.8	54.2	61.0
	Two Relatives	31.5	64.3	76.6	84.2	91.0
NDIS (Combined SDIS rescaled to	Parent-Child	[b]	28.9	38.5	46.5	56.2

10,000,000)						
	Full Siblings	[b]	30.5	34.5	38.5	46.5
	Two Relatives	[b]	47.8	57.7	65.5	75.3

[a]   The number in parentheses ( ) under the State are estimates of the number of unique profiles searched based upon the average percentage of duplicates observed in individual searches.

[b]   The scaling formula is incapable of adjusting rank 1.

For purposes of Table 1, duplicates were filtered out of the search output. The estimated unique profiles for Wyoming are 19,300; for Virginia are 356,000 profiles; and for California are 1,780,000 profiles. The estimated combined total of the three states performing the searches is 2,130,000 profiles. Additionally, an NDIS search was approximated by rescaling the results for this combined number of offenders to a total of 10,000,000 using the formula in Curran and Buckleton (2008).[15]

The rankings demonstrate that, as the size of the database of offenders increases, the percentage of true relatives (parent/child, full siblings, two relatives) ranked as #1 decreases. For example, Wyoming, with the smallest number of unique profiles in this study (19,300), was able to identify the full sibling in the #1 ranked position approximately 56.8% of the time. California, having the largest number of unique profiles (1,780,000) was able to identify the full sibling in the #1 ranked position approximately 22.8% of the time. As the size of the database being searched increases, there is an increase in the number of false positives being generated by the search.[16]

With that in mind, the Ad Hoc Working Group recognized that each potential relative on the ranked list should be reviewed but also acknowledged the burden and resource implications for follow-up on large lists. As a practical matter, since the true relative is not always ranked as the #1 candidate in familial searches, especially as the size of the database being searched increases, ranked lists of offenders will have to be reviewed.   In fact, using the combined SDIS values in Table 1, there is approximately a 1 in 5 chance that a first-degree relative will be ranked #1. (Prior case experience at the SDIS level has revealed that the #1 ranked candidate was not always the true relative.) Actual NDIS values would be expected to be lower. To the extent that additional filters or thresholds are available, these lists can be further distilled to reduce the number of false positives.[17] One method employed by states currently performing familial searching is that of testing additional STR or Y-STR loci.[18] The ability to employ this additional filter to search results in states having a large number of offenders (e.g., more than 300,000) in their databases will assist in the evaluation of the ranked list. Another filter that has been successfully used by the United Kingdom in familial searching is that of geography.[19] In

addition to using metadata (e.g., age, gender, population group) for additional filters, the United Kingdom refines its ranked list by geographic location of the crime and offender.

For familial searching at the national level where the offender database contains over 10 million offender DNA records,[20] it would be difficult to establish a reasonable threshold for the review of ranked lists. For example, if the threshold for review of ranked lists at the national level was determined to be 100, it would need to be understood that in approximately 25% to 50% of the searches, a relative (if one or more relatives are even present in the database) would not be listed in the first 100 candidates. Moreover, the additional filters of metadata and geography are not available at the national level since this data is not collected and stored at NDIS.[21]   The availability of Y-STR testing for further filtering would also be left to each individual state as validation and resources permit, which will necessarily create disparate treatment of the ranked lists by the states. We are also aware of jurisdictional issues relating to the legal authority to perform familial searching, and that this impacts the scientific implementation of familial searching at the national level.[22] The Group also acknowledges that the implementation of additional core loci will increase the feasibility of successful familial searching at the national level.[23]

Appendix on the Familial Searching Studies[24]

Purpose:

To evaluate the efficacy of familial searching on a national level given the current 13 CODIS core loci, simulations were performed that examined the efficiency of familial searching NDIS-sized databases. However, given the unknown but expected complexity of a real database (population structure; relatives; duplicates), tests using the actual NDIS database would be desirable in part to look for suggestions of practical deviations from the models.* At present, familial searching studies using NDIS cannot be performed. By combining test results from multiple states, experiments could be performed on a combined SDIS database that approaches one fifth of the total NDIS count. The California Department of Justice Bureau of Forensic Services, the Virginia Department of Forensic Science, and the Wyoming State Crime Laboratory agreed to participate in this study.

*Acknowledging that, due to the limited number of search profiles, these may not provide statistically significant results.

Method and Design:

Modeling the STR profiles of the perpetrator (the search profile), his family, and unrelated offenders

STR profiles were created in Microsoft® Excel® for 200 artificial families using the 13 CODIS core loci. The families consisted of a perpetrator, his father and mother, two full siblings, his spouse, and his two children. Each family was randomly assigned to population groups in proportion to the population breakdown of prisoners in the United States[25] (see Table 2 below).

Table 2: Population Groups Used in Study

Database:	Reference:	Expected Proportions	Study Proportions
African American	JFS 1999 44(6):1277-1286[26]; FSC 2001 3(3)[27]	0.36	0.335
Caucasian	JFS 1999 44(6):1277-1286; FSC 2001 3(3)	0.401	0.4
Hispanic	JFS 1999 44(6):1277-1286; FSC 2001 3(3)	0.185	0.195
Han Chinese	FSI (2005) 147; pp 89–91[28]	0.054	0.07

In creating the STR profiles, the subpopulation concepts of Balding and Nichols (1995)[29] were applied to the selection of all founder alleles (i.e., the alleles used to create the profiles of the father, mother, perpetrator's spouse, and comparison offender). In this approach, the probability of selecting the next allele (P_i) is affected by the population frequency of that allele (p_i), how many of that allele have been selected to that point (m_i), the total number of alleles that have been selected to that point (m), and the probability that two randomly selected alleles are identical by descent (θ):

Eq. 1

$$\frac{m_i\theta + (1 - \theta)p_i}{1 + (m - 1)\theta} = \frac{m_i\theta}{m\theta + (1 - \theta)} + \frac{(1 - \theta)p_i}{m\theta + (1 - \theta)}$$

On the right hand side of Equation 1, the first part represents the proportion of the time the next allele is expected to be identical by descent with an allele of type P_i. This takes into account the number of times that particular allele has been observed in the tested individuals so far (m_i) and all of the possible outcomes:

- $m\theta$ is the probability the next allele is identical by descent (IBD) with an allele already observed; and
- $1 - \theta$ is the probability the next allele is not IBD with any of the alleles already observed.

Eq. 2

$$\frac{m_i\theta}{m\theta + (1 - \theta)}$$

Equation 2 was applied at $\theta = 0.01$ to the selection of alleles of founders (father, mother, spouse, and offender) when the individuals were from the same population group. The father, mother, and spouse were always assigned to the same randomly selected population group, while the offender was randomly assigned to any of the four groups.

When modeling founder alleles, a random number would be compared to the value resulting from Equation 2 to see if the next allele would/wouldn't be IBD with one previously selected. When the next allele was not identical by descent with a previous founder allele, it would be assigned in proportion to allele frequencies in the cognate population group.

Once the founder profiles were established, the perpetrator, his two full siblings, and his two children were created according to Mendelian principles. Allele 1 was randomly selected from paternal alleles. Allele 2 was randomly selected from maternal alleles. Mutations were not simulated.[30]

SDIS searches

The participating states provided the following information regarding their familial searching process (see Table 3 below): (1) source of their familial searching software; (2) the perpetrator-SDIS comparison metrics captured by their software [e.g., profile(s); SDIS sample identifier; Parent-Child likelihood ratios (PI); Full Sibling likelihood ratios (SI)[31]; locus count; alleles shared]; (3) citations to the population databases used for their PI and SI calculations; (4) the theta level applied (e.g., 0; 0.01) to likelihood ratios; (5) the mutation calculation approach or formula; and (6) any thresholds applied to the likelihood ratios.

Table 3: State Familial Searching Information

State	Software Source	Frequency Databases	Source	Theta	Mutation (PI)	Minimum Allele Frequency	LR Thresholds for Study[a]
WY	In house (Partially adapted from Denver)[b]	African American, Caucasian, Hispanic[c]	JFS 2001 46(3):453-489[32]	0.01	1 locus: Ayres, FSI 114 (2000) using rates from STRbase.[33] 2+ loci: 0 (Exclusion)	5/2N	1.0 across all databases
VA	Denver[b]	African American, Caucasian, Hispanic	In house[d]	0	0.001	0.01	Parent-Child: 100 Full Siblings: 33
CA	In house	African American, Caucasian, Hispanic	JFS 1999 44(6):1277-1286	0	0 (Exclusion)	5/2N	1.0 across all databases
Simulations	In house	African American, Caucasian, Hispanic	JFS 1999 44(6):1277-1286	0	AABB 2008[34] Paternal mutation rates	5/2N	None

[a] All agencies incorporate the database size and apply the 0.1 and 1.0 SWGDAM thresholds when conducting actual casework familial searches.
[b] Originally developed by the Denver Police Department, Crime Laboratory Bureau.

c   Wyoming uses FBI (University of North Texas Health Science Center) Southwest Hispanic and Florida Southeast Hispanic databases. For ranking purposes in this study, the minimum of the two LRs was assigned to "Hispanic."

d   Available at http://www.dfs.virginia.gov/wp-content/uploads/2013/07/210-D400-Forensic-Biology-Section-Procedures-Manual-Section-III.pdf (210-D400 Forensic Biology Section Procedures Manual, Section III).

The individual SDIS database laboratories conducted familial searches using the 200 perpetrator profiles. PIs and SIs were calculated for all perpetrator-offender comparisons across all population groups (African American, Caucasian, and Hispanic). The states created lists of results that included all SDIS comparisons with likelihood ratios exceeding the thresholds listed in Table 3 above.

The SDIS results were filtered to remove duplicates (see Table 4 below).

Table 4: Filtering for Duplicates*

State	SDIS N: Total with Duplicates	Average % Duplicates in FS Lists	Unique Offenders: Maximum in FS List	Unique Offenders: Minimum in FS List	Unique Offenders: Average
WY	19,338	0.0588%	656	57	274
VA	356,363	0.161%	6,881	456	2,349
CA	2,053,383	13.2%	132,222	8,335	37,821
Comb.	2,429,084	12.3%	136,373	9,098	40,444

*"Duplicates" may include monozygotic twins

SDIS search results (Wyoming, Virginia, California, Combined)

PIs and SIs were calculated for all simulated family members except the spouse (i.e., perpetrator vs. father; perpetrator vs. mother; etc.) across all population groups. The perpetrator vs. relative kinship likelihood ratios were added to the respective lists of LRs from the database searches.

Comparisons ranked true relatives within each SDIS search using LRs calculated according to each state's approach (database, coancestry corrections, mutation approaches). Each relationship was ranked by the minimum LR across three population databases (African American, Caucasian

and Hispanic) and ranked by the PI or SI, respective to the true relationship. The rank represents where the true relationship sat within the SDIS or NDIS list of comparisons to unrelated people. For example, a rank of 10 for the father of the perpetrator means that 9 unrelated people had profiles that gave higher Parent-Child LRs when compared to the perpetrator. The ranking results were evaluated for each state individually as well as combined. In addition, an estimate of these rankings in a 10 million offender database (NDIS) was plotted by rescaling the combined results using the formula in Curran and Buckleton (2008).

The ranks versus the proportion of the 200 perpetrators with a family member included up to that rank were calculated and graphed. For example, at rank 10, the graph would show the proportion of relatives that ranked anywhere from 1 to 10. The cumulative proportion included at a given rank was then averaged for each family member of a similar relationship:
- The father, mother, and children of the perpetrator are all ranked by PI, so those results were averaged for the graphs below.
- Similarly, the full siblings are ranked by SI, so those results are averaged.

Rankings were also examined for instances where two family members were present in the offender database. Here, the rank represents the highest rank of the two relatives of the perpetrator (e.g., a person ranked 1st is in a higher ranking than a person ranked 10th). The following combinations were included: father and mother; father and full sibling; father and perpetrator's child; mother and full sibling; mother and perpetrator's child; two full siblings; full sibling and perpetrator's child; and the perpetrator's two children. The graphs use the average cumulative percentages across all combinations.

These results are also summarized in Table 1.

States' Familial Search Results[35]

Observed Ranks: SDIS
Wyoming, Virginia, California, Combined, NDIS Estimate

$$y = 0.145\ln(x) + 0.1198$$
$$R^2 = 0.9976$$

$$y = 0.1205\ln(x) + 0.0043$$
$$R^2 = 0.9796$$
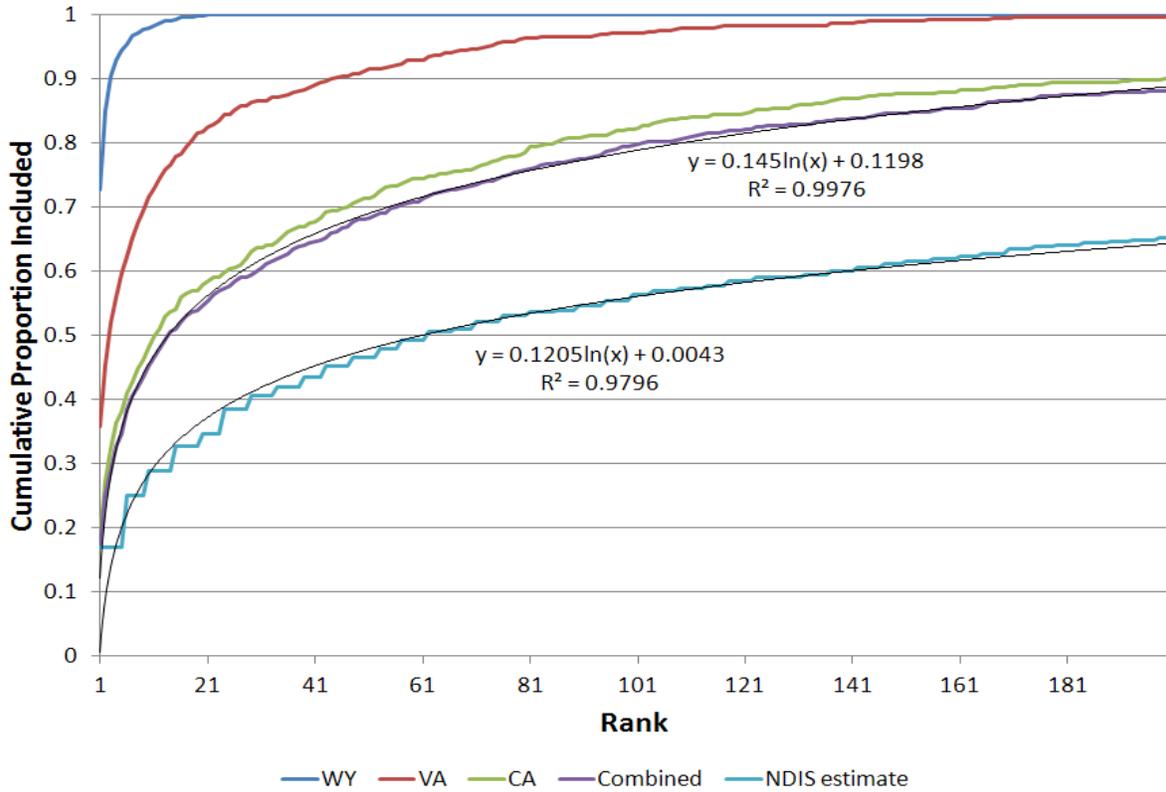
Legend: WY, VA, CA, Combined, NDIS estimate

Figure 1: Parent/Child LR - Average

Ranking of the true relationship within the LR list

Observed Ranks: SDIS

Wyoming, Virginia, California, Combined, NDIS Estimate

$y = 0.0924\ln(x) + 0.1814$
$R^2 = 0.9954$

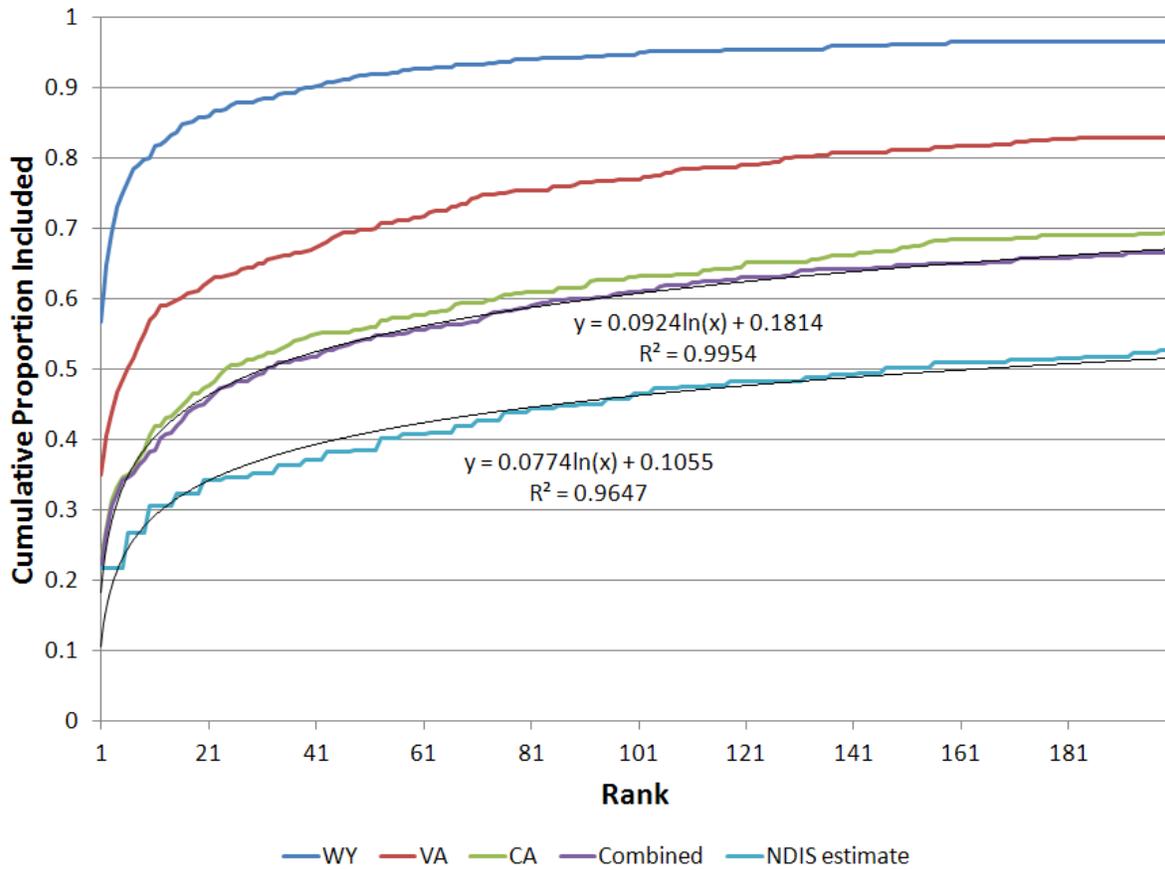$y = 0.0774\ln(x) + 0.1055$
$R^2 = 0.9647$

Figure 2: Full Siblings LR – Average

Ranking of the true relationship within the LR list

Observed Ranks: SDIS

Wyoming, Virginia, California, Combined, NDIS Estimate

$y = 0.1075\ln(x) + 0.4056$
$R^2 = 0.9786$

$y = 0.1211\ln(x) + 0.1868$
$R^2 = 0.9876$
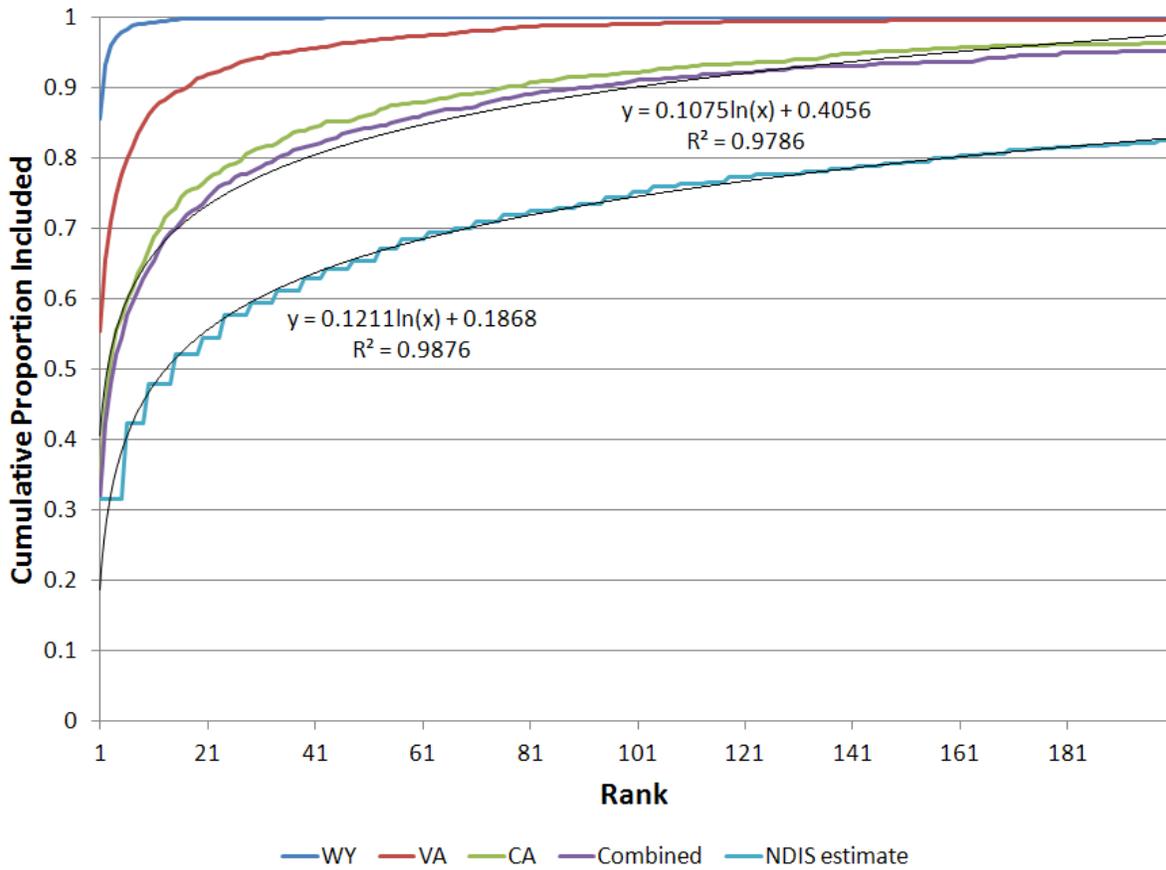
Legend: WY, VA, CA, Combined, NDIS estimate

Figure 3: Two Relatives – Average

Ranking of the true relationship within the LR list

Simulated SDIS (2.13 million) and NDIS (10 million) searches

The profiles of the 200 perpetrators' families were used to perform simulated familial searches using databases of the following sizes:

- 2.13 million, representing the size of the combined SDIS study (2,429,084 offenders) less the approximately 12.3% of duplicated profiles; and
- 10 million, representing the full size of NDIS.

For the NDIS portion, 800 additional families were modeled for simulated searches for a total of 1,000.

During the simulated searches part of this study, all offenders in the database were modeled *de novo* for each perpetrator family. Each offender was randomly assigned to a population group in proportion to the population breakdown of prisoners in the United States. The assignment of offender alleles depended upon the population group assigned to that person. If the offender was from the same population group as the father, mother, and spouse, then the probability that offender allele 1 was IBD with one of the family's alleles was dictated by Equation 2. If not IBD, offender allele 1 was selected at random in proportion to his cognate population allele frequencies. The same applies to offender allele 2, except the probability that allele 2 was IBD with a previously observed allele was dictated by the father, mother, and spouse's alleles as well as offender allele 1. If the offender was from a population group different from the perpetrator's family, offender allele 2 may still be IBD with offender allele 1. If it was not IBD with any allele, it was selected at random. Note: The alleles observed for one offender were not applied to the modeling of any other offender.

Ranking and graphing of the perpetrator's family members were performed as described above for the SDIS searches. The combined SDIS results from above were included for comparison.

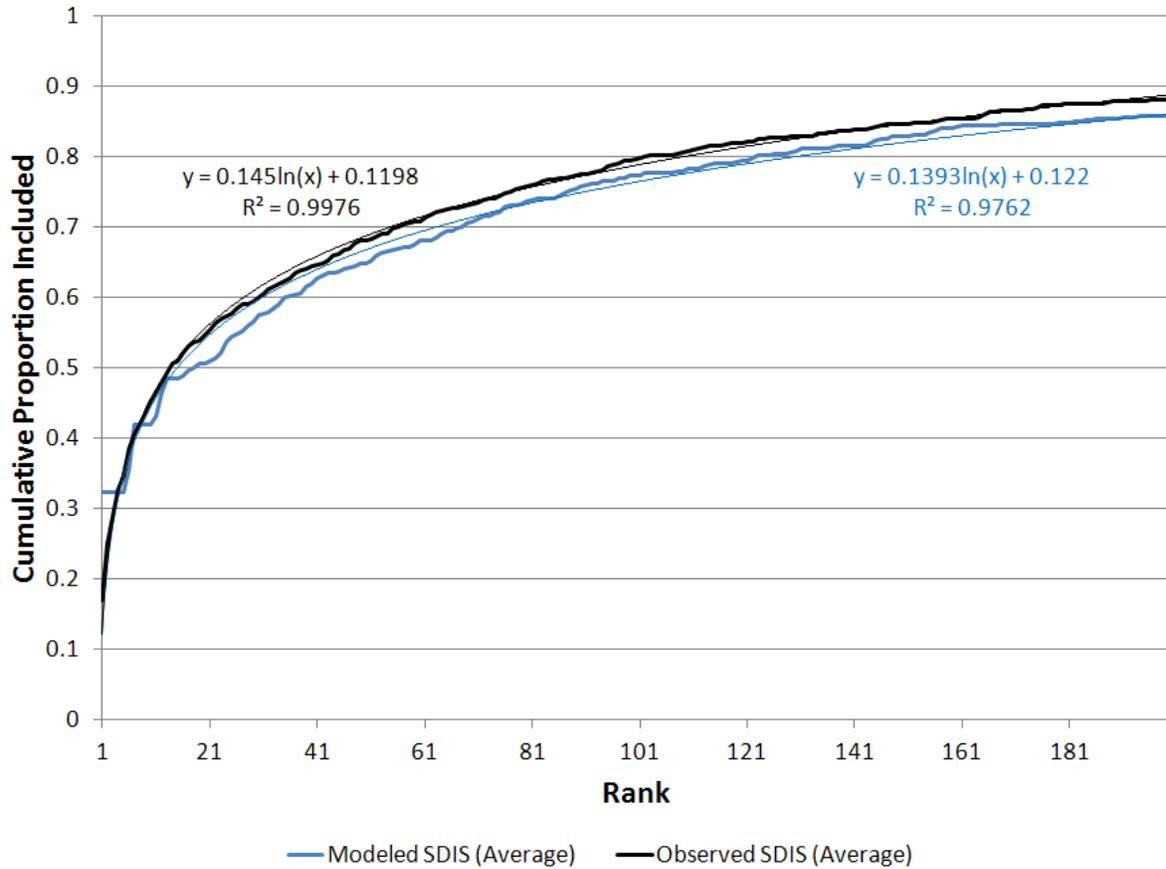**Observed vs. Modeled Ranks:**
Combined SDIS (2.13 million)

Figure contents:
- Y-axis: Cumulative Proportion Included (0 to 1)
- X-axis: Rank (1 to 181)
- $y = 0.145\ln(x) + 0.1198$
- $R^2 = 0.9976$
- $y = 0.1393\ln(x) + 0.122$
- $R^2 = 0.9762$
- Legend: Modeled SDIS (Average), Observed SDIS (Average)

Figure 4: Parent/Child LR- Average

Ranking of the true relationship within the LR list

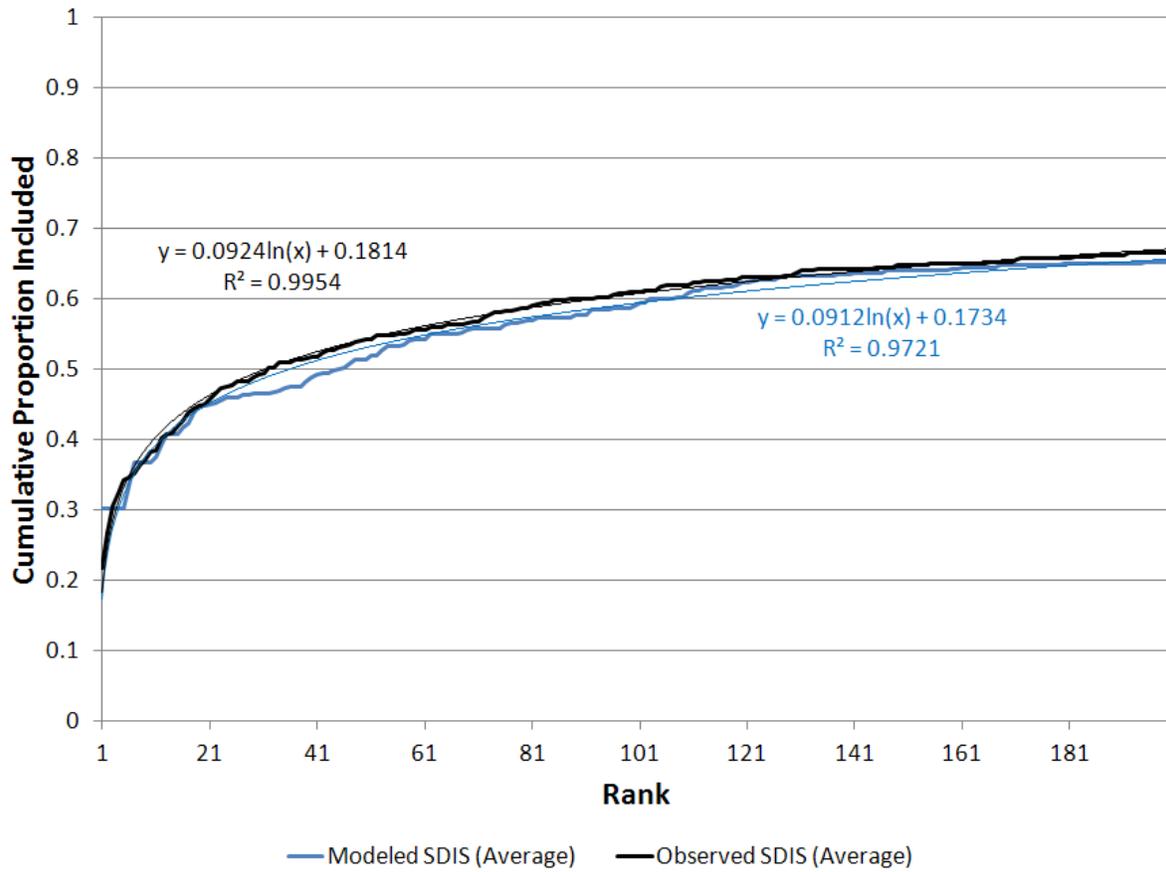**Observed vs. Modeled Ranks:
Combined SDIS (2.13 million)**

$y = 0.0924\ln(x) + 0.1814$
$R^2 = 0.9954$

$y = 0.0912\ln(x) + 0.1734$
$R^2 = 0.9721$

—— Modeled SDIS (Average)   —— Observed SDIS (Average)

Figure 5: Full Siblings LR – Average

Ranking of the true relationship within the LR list

**Observed vs. Modeled Ranks:
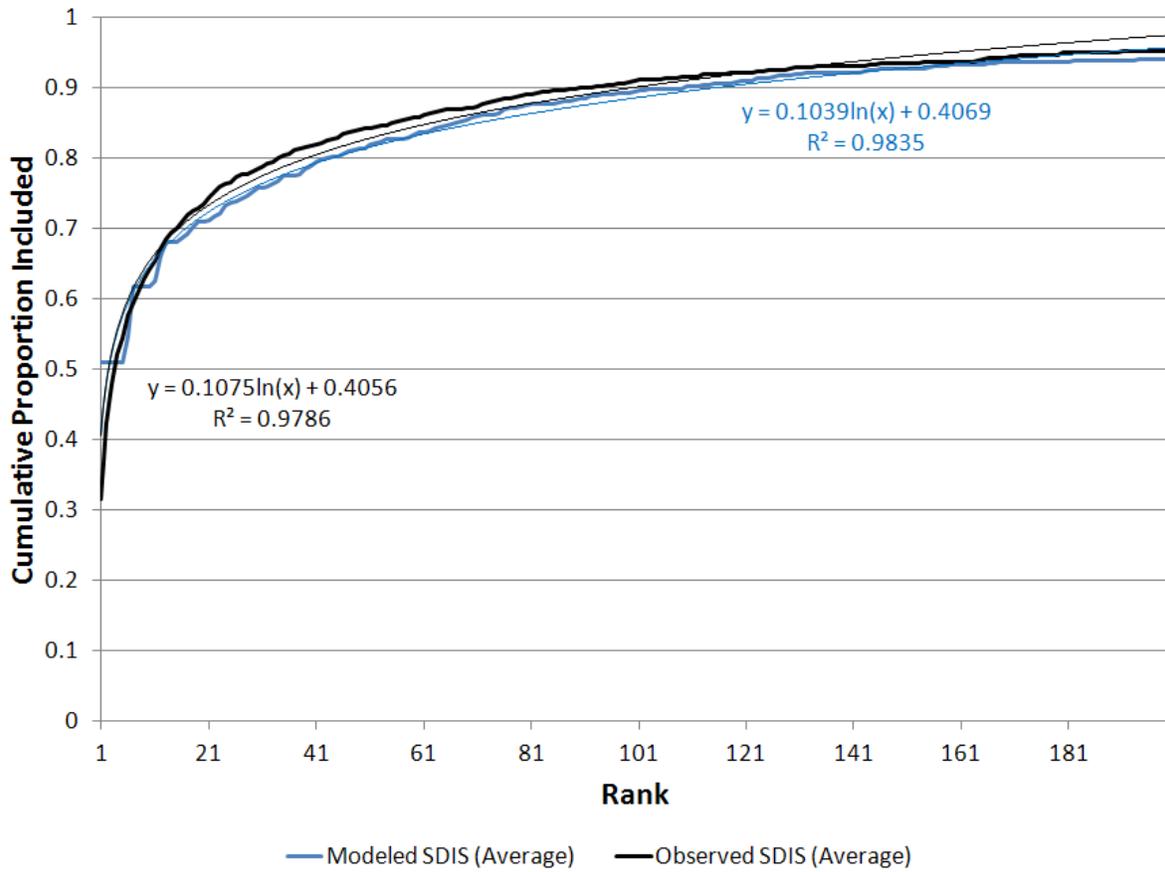Combined SDIS (2.13 million)**

$y = 0.1039\ln(x) + 0.4069$
$R^2 = 0.9835$

$y = 0.1075\ln(x) + 0.4056$
$R^2 = 0.9786$

Modeled SDIS (Average)   Observed SDIS (Average)

Figure 6: Two Relatives – Average

Ranking of the true relationship within the LR list
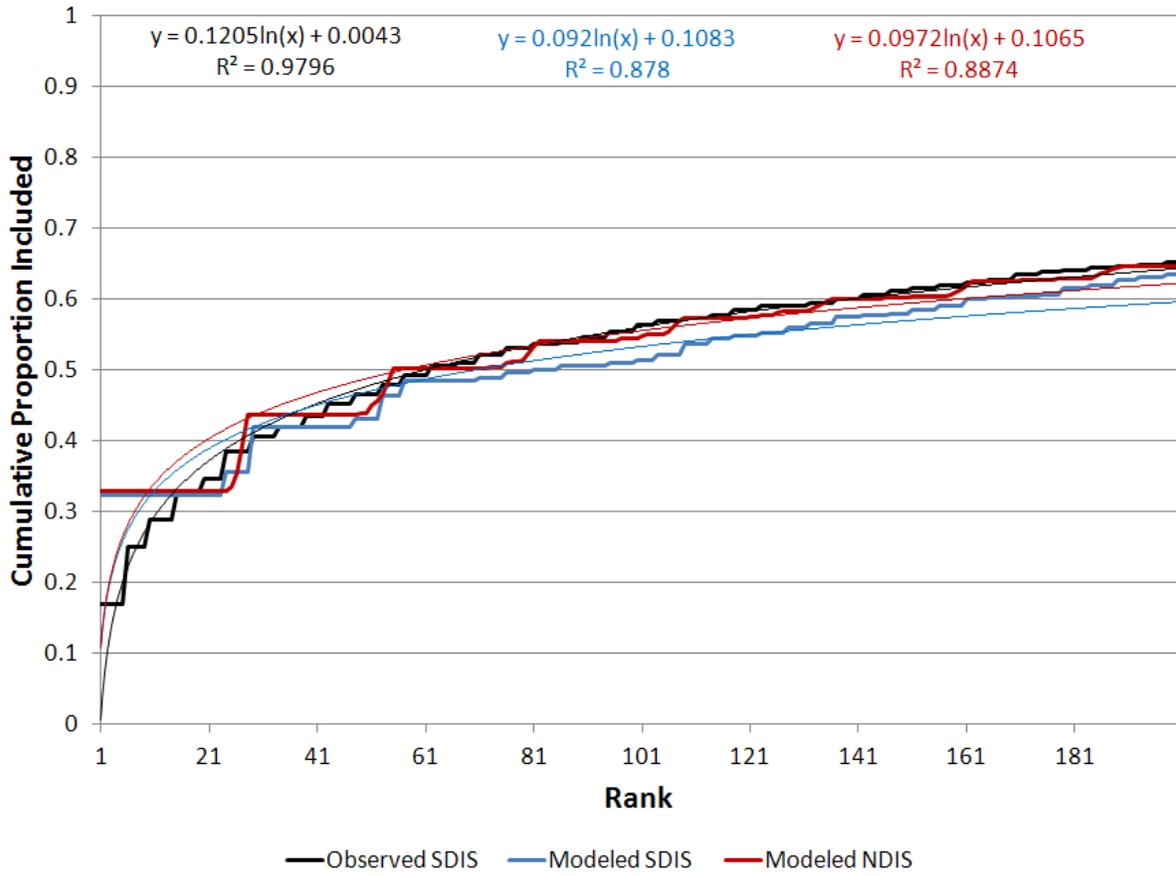
Observed vs. Modeled Ranks:
NDIS (10 million)

$y = 0.1205\ln(x) + 0.0043$
$R^2 = 0.9796$

$y = 0.092\ln(x) + 0.1083$
$R^2 = 0.878$

$y = 0.0972\ln(x) + 0.1065$
$R^2 = 0.8874$

Cumulative Proportion Included

Rank

—— Observed SDIS —— Modeled SDIS —— Modeled NDIS

Figure 7: Parent/Child LR – Average

Ranking of the true relationship within the LR list

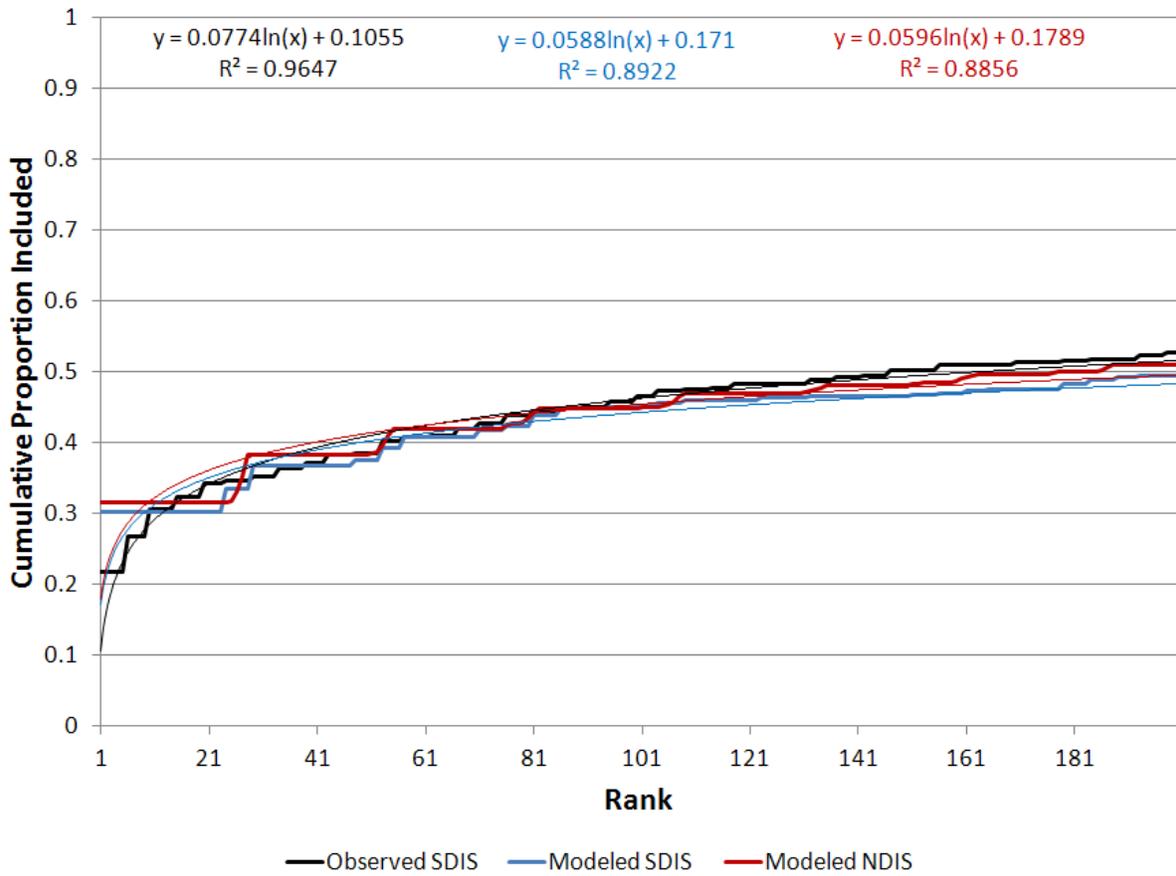**Observed vs. Modeled Ranks:**

NDIS (10 million)

$y = 0.0774\ln(x) + 0.1055$
$R^2 = 0.9647$

$y = 0.0588\ln(x) + 0.171$
$R^2 = 0.8922$

$y = 0.0596\ln(x) + 0.1789$
$R^2 = 0.8856$

Cumulative Proportion Included

Rank

—— Observed SDIS   —— Modeled SDIS   —— Modeled NDIS

Figure 8: Full Siblings LR – Average

Ranking of the true relationship within the LR list

Observed vs. Modeled Ranks:

NDIS (10 million)

$y = 0.1211\ln(x) + 0.1868$
$R^2 = 0.9876$

$y = 0.0881\ln(x) + 0.3136$
$R^2 = 0.8995$

$y = 0.0903\ln(x) + 0.311$
$R^2 = 0.8965$

Cumulative Proportion Included

Rank

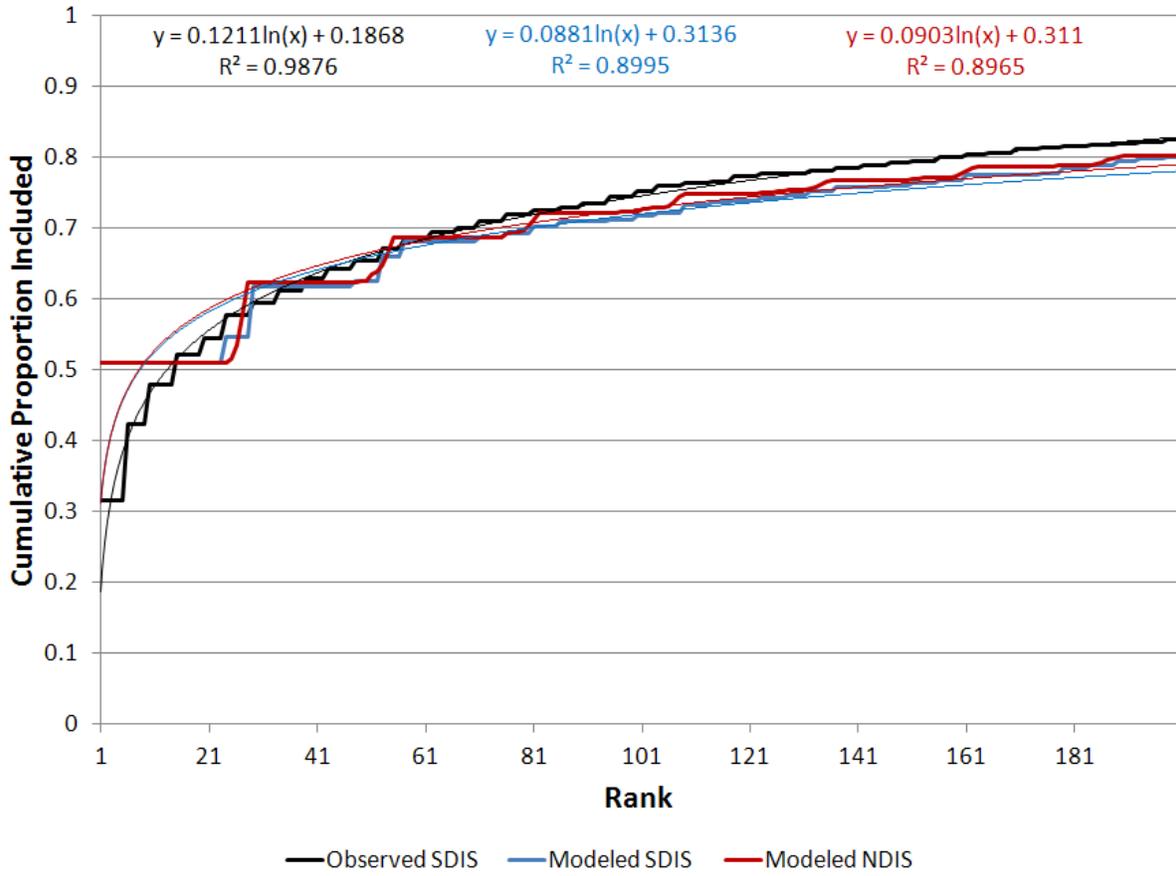—Observed SDIS —Modeled SDIS —Modeled NDIS

Figure 9: Two Relatives LR – Average

Ranking of the true relationship within the LR list

Additional Study Considerations Regarding Uncertainties in SDIS/NDIS Databases

The actual and modeled study data provide "guidance" of what would be found in the implementation of the recommendations (listed in the Executive Summary) in SDIS/NDIS familial searches. Accordingly, the data may not represent precise predictions as a result of the following uncertainties:

- Information on the racial composition and whether there are first-degree (or other distant) relatives of convicted offenders/arrestees are not available for the SDIS/NDIS databases. Thus, the computations for the racial compositions in this report may not completely represent the actual racial composition of the SDIS/NDIS databases. The minimum kinship statistic (i.e., likelihood ratios over three and four populations) was utilized to address these uncertainties.

- Similarly, allele frequencies for the relevant subpopulations in the SDIS/NDIS offender/arrestee databases were taken from the scientific literature (see Table 2), which may not exactly represent the allele frequencies in the actual SDIS/NDIS databases. This uncertainty would also have a relatively minor impact on the prediction of familial search results based on minimum kinship index (LR) computations using allele frequency data from multiple subpopulations.

- Demographic data reported by the United States' Bureau of Justice Statistics show the prevalence of first-degree relatives (parent-offspring and full siblings) among jail inmates in the United States. SDIS/NDIS data do not provide any detailed information on which pairs of CODIS records represent first-degree biological relatives. Further, no information exists in the CODIS system with regard to the presence of other biological relatives (e.g., half-siblings, uncle-niece/nephews, grandparent-grandchild, etc.). The LR computations in this report are based on comparisons of hypotheses of relatedness (parent-child and full siblings) versus unrelatedness (random individuals). The ranked lists of candidate offenders in the SDIS/NDIS databases may include relationships other than parent-child and full siblings. While individually such scenarios may be rare, they may exist when the database size is large (e.g., over 10 million) in actual implementation of the LR-based protocol recommended in this report.[36]

- The population substructure adjustment (i.e., θ-adjustment) made in this report is based on the model of Balding and Nichols (1995). The allele frequency structure under this model assumes a finite population of constant size which is at mutation-drift equilibrium, with mutations occurring under the infinite allele model (i.e., each mutation produces an allele not observed previously in the population). Population genetic studies of almost all extant human populations suggest that the assumption of mutation-drift equilibrium may not be strictly valid. Also, it is well-documented that mutations of the CODIS STR loci do not adhere to the infinite allele model. In contrast, there is substantial empirical

evidence showing that mutations at the CODIS STR loci occur by a pattern of forward-backward stepwise mutation (AABB 2008). Further, it is known that population substructure effects on loci evolving under the forward-backward stepwise mutation model are weaker than those for loci evolving under the infinite allele model (such as the model of Balding and Nichols).[37] Hence, the studies in this report use a conservative model of population substructure (i.e., higher values of θ), bringing the LRs of true relatives and unrelated individuals closer together than would be expected for the CODIS STR loci in the actual SDIS/NDIS databases.

- The success rates of familial searches predicted in this report are based on LR statistics that rely on the standard hypothesis testing paradigm. Under this paradigm, false positives of familial searches are the type I errors (rejecting the null hypothesis of no relationship while it is true) and the false negatives are the type II errors (accepting the null hypothesis of no relationship when it is false). As the size of the database increases, familial search comparisons will have more opportunities for type II errors given the large ratio of unrelated to related individuals. The LR approach maximizes the statistical power (complement of false negatives) for a given level of type I error (false positives). The operational goal of familial searching should be just the reverse, i.e. minimize false positives for a tolerable level of false negatives. In other words, accept that some true relatives may be missed but minimize investigations of unrelated individuals whose presence on the list is adventitious. When candidate ranked profiles are filtered by e.g., typing of additional loci and/or Y-STR typing, the actual false positive rates would be greatly reduced.

- Finally, in this report, the rankings of candidate profiles (for all but the topmost ranked one) were rescaled to an NDIS level by applying the approximation of Curran and Buckleton (2008). Success rates of familial searches rescaled in this manner may also be somewhat uncertain, since actual allele frequencies and race/ethnicity affiliations of the SDIS/NDIS records are not available.

All of these issues suggest that irrespective of the threshold based on the length of the candidate list, filtering the list (e.g., typing additional loci or Y-STRs, and/or review of vital records/demographic information available to law enforcement) is critical before any familial search information is released.

[1] The SWGDAM Ad Hoc Working Group on Familial Searching includes Dr. Ranajit Chakraborty, University of North Texas Health Science Center; Dr. Kenneth Kidd, Yale University; Mr. Brad Jenkins, Virginia Department of Forensic Sciences; Mr. Gary Sims and Mr. Steven Myers, California Department of Justice Jan Bashinski DNA Laboratory, and Mr. Russell Vossbrink, Scottsdale Police Department (prior to Mr. Vossbrink, Dr. Philip Kinsey was the SWGDAM Vice Chair and participated in the Group's meetings).

[2] Available at http://www.fbi.gov/about-us/lab/biometric-analysis/codis/familial-searching.

[3] It was clarified that the additional core loci proposed for CODIS should not be considered as available for familial searching at this time because these additional core loci efforts are in the initial stages of validation and implementation will occur in the future (at least 3 years out). *See* http://www.fbi.gov/about-us/lab/biometric-analysis/codis/planned-process-and-timeline-for-implementation-of-additional-codis-core-loci.

[4] Available at http://www.fbi.gov/about-us/lab/biometric-analysis/codis/codis-and-ndis-fact-sheet.

[5] Familial searching is not to be confused with the occurrence of partial matches resulting from the current routine moderate stringency matching of the database leading to a candidate match, which because of the number of shared alleles involved in the match, indicate to the forensic examiner that a relative of the offender/arrestee may be the putative perpetrator. *See* Familial Searching at http://www.fbi.gov/about-us/lab/biometric-analysis/codis/familial-searching. *See also* Steinberger, E., Sims, G. (2008) *Finding criminals through the DNA of their relatives – familial searching of the California offender DNA database*, Prosecutors Brief, Vol. XXXI, p. 28-32.

[6] Whether offender, arrestee, or both offender and arrestee DNA records are searched for familial searching purposes is determined by each jurisdiction; for example, California searches only convicted offender DNA records (http://www.denverda.org/DNA_Documents/Policies/CA%20FS%20Policy.pdf); Colorado and Virginia search offender and arrestee DNA records (http://www.denverda.org/DNA_Documents/Policies/CO%20FS%20Policy.pdf and http://www.denverda.org/DNA_Documents/Policies/VA%20FS%20Policy.pdf).

[7] Additional information on H.R. 3361 is available at http://thomas.loc.gov/cgi-bin/query/z?c112:H.R.3361.IH:. There was no further action on H.R. 3361 before the 112th Congressional Session adjourned.

[8] *See* Conference Report to Accompany H.R. 2112, 112th Cong., 1st Session, November 14, 2011, H. Rept. 112-284 (Washington: GPO, 2011) p. 238.

[9] *See* Budget Hearing – Federal Bureau of Investigation, House Committee on Appropriations, March 19, 2013 available at http://appropriations.house.gov/calendararchive/eventsingle.aspx?EventID=323252.

[10] Myers, S.P., Timken, M.D., Piucci, M.L., Sims, G.A., Greenwald, M.A., Weigand, J.J., Konzak, K.C., Buoncristiani, M.R. (2011) *Searching for first-degree familial relationships in California's offender DNA database: Validation of a likelihood ratio-based approach,* Forensic Sci. Int. Genet. 5(5):493-500, doi:10.1016/j.fsigen.2010.10.010.

[11] Ge, J., Budowle, B. (2012) *Kinship index variations among populations and thresholds for familial searching*. PLoS One 7(5); e37474, doi:10.1371/journal.pone.0037474

[12] Weir, B.S., Anderson, A.D., Hepler, A.B.(2006) *Genetic relatedness analysis: Modern data and new challenges*. Nat Rev Genet 7(10):771-80.

[13] Ge, J., Chakraborty, R., Eisenberg, A., Budowle, B. (2011) *Comparisons of familial DNA database searching strategies*, J Forensic Sci. 56(6): 1448–56, doi: 10.1111/j.1556-4029.2011.01867.x; Rohlfs, R.V., Fullerton, S.M., Weir, B.S. (2012) *Familial Identification: Population structure and relationship distinguishability*. PLoS Genet

8(2): e1002469. doi:10.1371/journal.pgen.1002469; Reid, T.M., Baird, M.L., Reid, J.P., Lee, S.C., Lee, R.F.(2008) *Use of sibling pairs to determine the familial searching efficiency of forensic databases*, Forensic Sci. Int. Genet. 2(4):340-2, doi: 10.1016/j.fsigen.2008.04.008; Balding, D.J., Krawczak, M., Buckleton, J.S., and Curran, J.M. (2013) *Decision-making in familial database searching: KI alone or not alone?* Forensic Sci. Int. Genet. 7(1): 52-4; doi: 10.1016/j.fsigen.2012.06.001; Curran, J.M. and Buckleton, J.S. (2008) *Effectiveness of familial searches*, Sci. Justice 48(4):164-7, doi: 10.1016/j.scijus.2008.04.004; Hicks, T., Taroni, F., Curran, J., Buckleton, J., Castella, V., and Ribaux, O. (2010) *Use of DNA profiles for investigation using a simulated national DNA database: Part II. Statistical and ethical considerations on familial searching*, Forensic Sci. Int. Genet. 4 (5): 316-22, doi: 10.1016/j.scigen.2009.11.006.

[14] U.S. Bureau of Justice Statistics, *Special Report: Profile on Jail Inmates 2002*, (July 2004, NCJ 201932) available at http://www.bjs.gov/content/pub/pdf/pji02.pdf.

[15] Curran, J.M. and Buckleton, J.S. (2008) *Effectiveness of familial searches*, Sci. Justice 48(4):164-7, doi: 10.1016/j.scijus.2008.04.004.
.
[16] O'Connor, K.L. (2011) *Familial searching of forensic DNA databases*, available at http://www.cstl.nist.gov/strbase/pub_pres/OConnor_ICFIS2011.pdf.

[17] Bieber, F., Brenner, C.H., Lazer, D. (2006) *Finding criminals through the DNA of their relatives*, Science 312(5778):1315-6, doi:10.1126/science.1122655.

[18] California Department of Justice (2008) *DNA Partial Match Policy*; Colorado Bureau of Investigation (2009) *DNA Familial Searching Policy*; and Texas (2012) *CODIS Partial Matches and Familial Searches* available at http://www.denverda.org/dna/Familial_DNA_Database_Searches.htm. The ability to analyze additional and Y-STR loci is an issue of implementation for many laboratories as not all laboratories currently perform Y-STR analysis. Analysis of the additional and Y-STR loci would encompass additional resources for staff, supplies, etc… *See also* O'Connor, K.L., Butts, E., Hill, C.R., Butler, J.M., Vallone, P.M. (2010) *Evaluating the effect of additional forensic loci on likelihood ratio values for complex kinship analysis*, Proceedings of the 21st International Symposium on Human Identification available at http://www.cstl.nist.gov/strbase/pub_pres/OConnor_Promega%20Proc%202010_Additional%20Loci%20Kinship.pdf.

[19] O'Connor, K.L. (2011) *Familial searching of forensic DNA databases*, available at http://www.cstl.nist.gov/strbase/pub_pres/OConnor_ICFIS2011.pdf.; Maguire, C.N., McCallum, L.A., Jones, K.E., Storey, C.L. (2009) *Developing a likelihood ratio approach to 'familial searching' of a DNA database using the advanced functionality of FSS-IBD,* 20th International Symposium on Human Identification Poster Abstracts available at http://www.promega.com/products/pm/genetic-identity/ishi-conference-proceedings/20th-ishi-poster-abstracts/.

[20] As of July 2013, NDIS contained 10,477,600 offender and 1,578,800 arrestee DNA records for a total of over 12 million offender/arrestee DNA records; *see* NDIS Statistics at http://www.fbi.gov/about-us/lab/biometric-analysis/codis/ndis-statistics.

[21] *See Frequently Asked Questions (FAQs) on the CODIS Program and National DNA Index System* (Under CODIS DNA Databases) at http://www.fbi.gov/about-us/lab/biometric-analysis/codis/codis-and-ndis-fact-sheet.

[22] The District of Columbia and Maryland currently prohibit familial searching of their DNA databases; see D.C. Code § 22-4151 (2009) ("DNA collected by an agency of the District of Columbia shall not be searched for the purpose of identifying a family member related to the individual from whom the DNA sample was acquired.") and Maryland Public Safety Code 2-506 (2008) ("*Limitation of search of statewide DNA data base.-* A person may not perform a search of the statewide DNA data base for the purpose of identification of an offender in connection with

a crime for which the offender may be a biological relative of the individual from whom the DNA sample was acquired.")

[23] Hares, D.R.(2012) *Expanding the CODIS core loci in the United States*. Forensic Sci. Int. Genet. 6(1): e52-4, doi:10.1016.fsigen.2011.04.012; Hares, D.R. (2012) *Addendum to expanding the CODIS core loci in the United States*. Forensic Sci. Int. Genet. 6(5):e135; doi:10.1016.j.fsigen.2012.01.003.

[24] Research design and modeling, California searching and compilation of all states' data by Steven Myers, California Department of Justice, with familial search study results contributed by the Virginia Department of Forensic Science (Brad Jenkins) and Wyoming State Crime Laboratory (Scott McWilliams).

[25] U. S. Bureau of Justice Statistics, *Special Report: Profile on Jail Inmates 2002*, (July 2004, NCJ 201932) available at http://www.bjs.gov/content/pub/pdf/pji02.pdf.

[26] Budowle, B., Moretti, T.R., Baumstark, A.L., Defenbaugh, D.A., Keys, K.M. (1999) *Population data on the thirteen CODIS core short tandem repeat loci in African Americans, U.S. Caucasians, Hispanics, Bahamians, Jamaicans, and Trinidadians*, J. Forensic Sci. 44(6):1277-86.

[27] Budowle, B. (2001) *Genotype profiles for five population groups at the short tandem repeat loci D2S1338 and D19S433*, Forensic Science Communications Vol. 3, No. 3 available at http://www.fbi.gov/about-us/lab/forensic-science-communications/fsc/july2001/index.htm/budowle1.htm.

[28] Wang, Z.Y., Yu, R.J., Wang, F., Li, X.S., Jin, T.B. (2005), *Genetic polymorphisms of 15 STR loci in Han population from Shaanxi (NW China),* Forensic Sci. Intl. 47(1):89-91.

[29] Balding, D.J., and Nichols, R.A. (1995) *A method for quantifying differentiation between populations at multi-allelic loci and its implications for investigating identity and paternity*, Genetica 96(1-2):3-12 (Erratum in Genetica 2008, 133(1):107.

[30] Paternal meiotic mutations occur at a known frequency of approximately 1 in 1,000 per locus. Parent-child false negatives attributable to paternal meiotic mutations may be offset by the individual being detected in the full sibling ranking list.

[31] Kinship testing likelihood ratio formulae incorporating subpopulation corrections can be found in Ayres, K.L. (2000) *Relatedness testing in subdivided populations*. Forensic Sci. Intl. 114(2):107-15; doi:10.1016/S0379-0738(00)00289-9.

[32] Budowle, B., Shea, B., Niezgoda, S., Chakraborty, R. (2001) *CODIS STR loci data from 41 sample populations*. J Forensic Sci. 46(3):453–89.

[33] Available at http://www.cstl.nist.gov/strbase/mutation.htm.

[34] American Association of Blood Banks Relationship Testing Standards Program Unit, *Annual Report Summary for Testing in 2008*. Available at http://www.aabb.org/sa/facilities/Documents/rtannrpt08.pdf.

[35] The log trendlines in Figures 1- 9 are based on ranks 1-200. This is typically the region of interest for familial searching. Trendlines selected for different ranges would provide different results, particularly the y-intercept.

[36] Rohlfs, R.V., Murphy, E., Song, Y.S., Slatkin, M. (2013) *The influence of relatives on the efficiency and error rate of familial searching*. PLoS One 8(8): e70495. doi:10.1371/journal.pone.0070495.

[37] Jin, L., and Chakraborty, R. (1995) *Population substructure, stepwise mutations, heterozygote deficiency and their implications in DNA forensics*. Heredity 74:274-285; doi:10.1038/hdy.1995.41.