Promega 15th Symposium on
Human Identification
October 6, 2004

Human Identification Software for Missing Persons, Scalable for a Mass Fatality Incident: Building on Lessons Learned Over the Course of a Major Disaster Victim Identification Project

HOWARD D. CASH† and MICHAEL J. HENNESSEY
Gene Codes Forensics, 775 Technology Drive, Suite 100A,
Ann Arbor, MI 48108, USA

A Disaster Victim Identification [DVI] project following a Mass Fatality Incident [MFI] has many similarities to the general problem of managing a missing persons identification effort. In fact, a broad missing persons program can be thought of as a mass fatality that takes place over an extended period of time. The forensic biologist is concerned with certain specific issues. Unidentified remains may be collected in varying degrees of biological compromise. Highly degraded samples may not be amenable to conventional short tandem repeat [STR] analysis and will require more sensitive testing such as mitochondrial DNA [mtDNA] sequencing or even Single Nucleotide Polymorphism [SNP] testing. Profiles generated from any or all of these techniques must then be compared to ante-mortem exemplars from the decedent or kinship exemplars from close relatives in order to make an identification.

The Mass-Fatality Identification System [M-FISys, pronounced like *emphasis*] is an extensive software system developed to support the Office of Chief Medical Examiner in the City of New York [OCME] following the World Trade Center [WTC] disaster of September 11, 2001. It combines STR, mtDNA and SNP profiles to support identification through both direct matching and kinship analysis. Samples can be tested multiple times and results combined to fill in missing loci with an audit trail back to each original analysis attempt, and all samples can be compared to each other simultaneously. Likelihood statistics are automatically generated for individual profiles as well as for complex pedigrees, and the analyst can annotate samples as the investigation progresses. Quality Control tools have been developed to detect non-obvious sources of error such as co-mingled remains and kinship swabs that result in inconsistent pedigrees. Extensive data browsers track each missing person and the exemplars available for identification. Tools are provided to augment the Administrative Review process needed to confirm the authenticity and chain of custody for those exemplars.

The next generation of M-FISys is designed to manage a large scale missing persons effort, but is engineered to be scalable for DVI if needed. For the WTC effort, OCME managers made a policy decision to largely restrict family exemplars to first order relatives [parents, children, full- and half-siblings] and to encode the familial relationship into the sample name. However, the current system can handle pedigrees of arbitrary complexity and does not rely on a particular sample naming convention to encode family relationships. Profiles can serve multiple roles, such as a personal exemplar from one decedent acting as a family reference for another, or one buccal swab being used as a kinship reference for multiple related decedents.

The WTC identification effort will be referred to as a source of important experience and lessons, but will not be discussed extensively in this paper or presentation. For more information on that effort, please see

http://www.promega.com/geneticidproc/ussymp13proc/contents/hennesseyrev1.pdf
and
http://www.bio-itworld.com/archive/091103/soul.html

† To whom correspondence should be addressed: howardc@genecodes.com

1 Background

Emergencies can oblige the community to respond to urgent needs with the development of innovative and technically pioneering solutions. Remedies and responses may be implemented on an emergency basis, but in an ideal world, they will be refined and extended for future use once the crisis has passed. After September 11, 2001, the first generation of the Mass-Fatality Identification System, *M-FISys* was developed in response to the disaster at the World Trade Center [WTC], to meet the needs of forensic biologists in New York City.[1] It is understandable that a new generation of tools might have been needed to manage the DNA data in a disaster with over 2,700 victims and approximately 20,000 highly compromised and co-mingled human remains:[2] Previously existing tools were not developed with this kind of application in mind. Furthermore, with so many scientific advances entering the forensic lab every year, there was little motivation to allocate the resources needed to create this kind of advanced software technology in the days before 9/11. But now that that technology has been developed and validated in one of the most complex forensic investigations in history, a failure to extend the system for conventional casework would be remiss.

M-FISys combines STR, mtDNA and SNP analysis (both direct matches to personal effects and kinship matches to family references), along with a rich collection of meta-data, Quality Control tools, calculations of likelihood statistics, and Administrative Review[3] functions. Integrating such a broad array of analysis tools was an ambitious undertaking; It was entered into with a design philosophy and goal of enabling a trained forensic biologist to organize and sort through large numbers of profiles and supporting data expeditiously, rather than building a program that would try to replace the scientist by autonomously making identifications. It can be used as a stand-alone application, but is more valuable in its client-server configuration, allowing several forensic scientists to work with the

---

[1] M-FISys was developed by Gene Codes Forensics, Inc. under a contract with the City of New York, but includes technology developed by Gene Codes Corporation over the past sixteen years. Under the terms of the contract, "All … middleware, new software and modifications of Sequencher [are] the exclusive intellectual property" of Gene Codes Forensics, Inc.

[2] Cash HD, Hoyle JW, Sutton AJ. "Development Under Extreme Conditions: Forensic Bioinformatics in the wake of the World Trade Center Disaster." Pacific Symp Biocomput. 2003;:638-53. PMID: 12603064

[3] World Trade Center and DNA Identifications: The Administrative Review, Mike Hennessey, *Thirteenth International Symposium on Human Identification – 2002*
http://www.promega.com/geneticidproc/ussymp13proc/contents/hennesseyrev1.pdf

data simultaneously.[4] Today, M-FISys continues to be enhanced to meet identification needs of the WTC effort after over three years of work. A separate initiative builds on that experience to develop equally advanced tools for managing broad missing persons identification efforts.

2 Differences Between Disaster Identification and a Missing Persons Systems

A missing persons program for a forensic biology lab has many similarities to certain Disaster Victim Identification [DVI] projects, as well as distinguishing differences. In both a missing persons investigation and a DVI response, the forensic scientist will try to match the DNA profiles of unidentified human remains to the reference samples of the person(s) reported missing or presumed killed. A disaster such as an aircraft crash can present the data analysts with a closed population set. This may allow certain identifications to be made by exclusion. The population of a massive disaster such as the WTC may never be provably closed, and an immense disaster over a wide area, such as earthquakes in Mexico City (1985, estimated 10,000 fatalities[5]) and Turkey (1999, more than 17,000 fatalities[6]) can result in a definitively open system. A broad missing persons program deals with an open system. and the data processing technology needs to recognize and support that fact.

A primary difference between a DVI project and a Missing Persons Program is that missing persons accumulate over an indeterminate period of time, whereas a disaster is thought of as a discrete (if not instantaneous) event. In practical terms, a missing persons system will continue to receive new input data indefinitely. In California alone, there are more than 2,100 unidentified remains (dating back to 1959) and more than 3,000 long-term missing persons (dating back to 1972).[7] There is currently no universally mandated protocol for sharing DNA profiles between all States in the USA. Until this is addressed, jurisdictional barriers will inevitably prevent some identifications from being made.

In the case of a MFI, it may be helpful to classify the magnitude of a disaster from the perspective of the team involved specifically with human identification.

---

[4] At the New York City Office of Chief Medical Examiner [OCME], M-FISys is typically used simultaneously by six to eight analysts and management personnel. The database architecture is based on Microsoft's SQL Server and can thus be expanded to support more analysts.

[5] "Earthquake Damage in Mexico City, Mexico, September 19, 1985," National Geophysical Data Center, (National Oceanic & Atmospheric Administration); http://www.ngdc.noaa.gov/seg/hazard/slideset/3/3_slides.shtml

[6] "Turks mourn on anniversary of earthquake," CNN, August 17, 2000; http://www.cnn.com/2000/WORLD/europe/08/17/turkey.quakeanni/index.html

[7] May 2003 letter from John Tonkyn, Convicted Offended DNA Databank Program and Missing Persons DNA Program, State of California Department of Justice.

The medical examiner in charge will be called upon to quickly assess the extent of the disaster and its concomitant demands on his or her resources. One tool that the DVI leadership can use to make this assessment is the Hennessey *Disaster Classification Matrix,*[8] a 2 x 2 grid where the horizontal axis represents working with the families and the vertical axis covers identifying the remains. The issues that comprise these areas of responsibility can be plotted on their respective axes, with increased distance from the origin representing a higher degree of difficulty.

Remains (Complex)		
Remains (Simple)		
	Families (Simple)	**Families** (Complex)

Examples of the kind of issues to plot on the matrix include the size of the victim population (and whether it is considered open or closed), the condition of the remains (levels of fragmentation and decomposition, presence of hazardous materials), the length of the recovery effort, the cultural diversity of the victims relative to the responding agency, and the proximity of the families to the disaster site. Each of these factors can be categorized as "simple" or "complex," based on the size of the problem relative to the agency's resources. For example, for "victim population," a dozen fatalities in a small town would be plotted as "complex," while, the same number of casualties in a larger urban area would probably be rated as "simple."

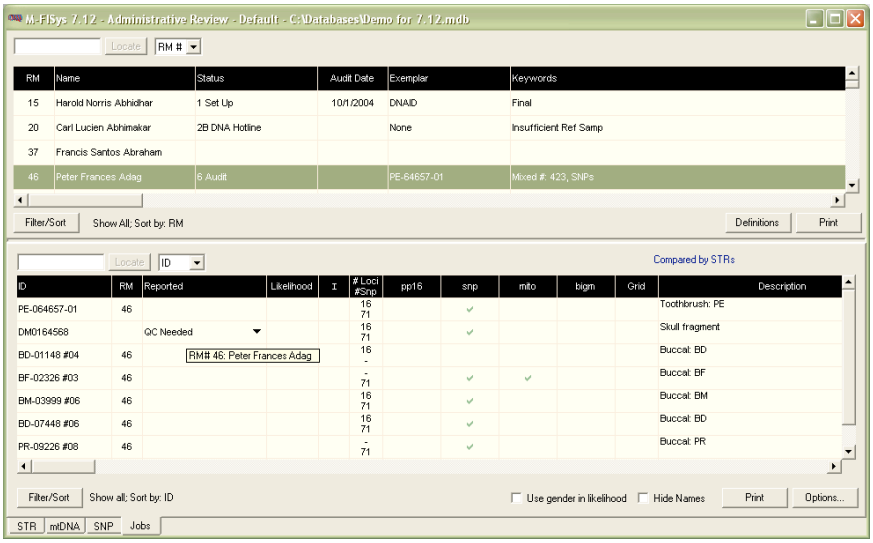
A missing persons case is triggered when either unidentified remains are recovered or a missing persons report is filed. That is, the ante mortem or post mortem data collection is usually the starting point, not the disappearance itself. Thus, while a DVI follows a linear path, missing persons cases have numerous starting points and often lack terminal events. This contrast with an MFI has distinct implications for both information handling strategies and for the setting in which the ante mortem collections take place

One agency is usually responsible for collecting both the ante mortem and post mortem data in a DVI. The opposite is usually the case in a missing persons investigation as the agency where the missing persons report was filed is often not the same jurisdiction where the remains are found. This may have a significant impact on data management: The integration of the ante mortem and post mortem

[8] MJ Hennessey and HD Cash, *The Disaster Classification Matrix: A Method for Quantifying Responses to Mass Fatality Incidents;* Work in preparation.

information is typically more manageable in a DVI precisely because it is located within one jurisdiction.

From a data management perspective, materially important information kept in an electronic system needs to be preserved like records kept in a laboratory notebook or entries into an accounting system. It is tempting to use technology to allow operators to modify entries when an error has been detected. However, it is helpful to remember that every piece of incorrect information started as data that someone believed was accurate. Wholesale changes of data are bound to introduce new errors in some instances, and it is important that an audit trail be kept of all modifications so that they can be reversed if needed, and justified if accurate. M-FISys requires that each operator log in with a password-protected account, and changes in data (correcting allelic drop out, marking contaminated samples as invalid, etc.) must be annotated and electronically signed. M-FISys is not a Laboratory Information Management System [LIMS] but the Administrative Review process is supported with tools to track all samples (remains and exemplars), analyses that have been performed on those samples and problems that have been encountered. These tools may detect data collection mistakes, but those errors must be corrected in a structured, formal way. A system that allows certain privileged users to override this discipline is doomed to exacerbate the very problems that that privileged user is trying to correct.



Finally, one of the most significant differences between a DVI and missing persons case is the crisis environment associated with the MFI. Learning a new set

of data management tools in the middle of an emergency creates avoidable delay, and adds stress to the experience of the forensic scientists who are very much on the front line. Surely it is helpful if the tools used in an MFI under crisis conditions are the same tools used by those scientists in their regular case work and missing persons management.

3 Software Design Issues

Certain criminal investigation and missing person identification databases are organized by design to take an unknown sample and compare it to all available reference profiles (e.g., known persons and unidentified crime-scene samples). M-FISys supports such unique-profile searches but adds the ability to perform a rigorous all-against-all comparison of all available profiles. This has the dual utility of aggregating multiple remains with the same profile (an important feature for working with an MFI such as an aircraft accident with highly fragmented bodies) and allowing all remains to be compared to all reference samples (e.g., all exemplars provided by families of reported missing persons with profiles from all recovered remains).

The M-FISys control panel divides interaction with the program into functional sections, based largely on how forensic staff might typically be assigned in an extended project. This allows new team members to quickly become effective contributors without having to learn the entire system. The main buttons access functions that might be used on a regular basis, while "Admin" buttons access additional functions that would be used rarely or would be used once in a typical session and can subsequently be ignored (such as loading a batch of new data or creating an account for a new user).

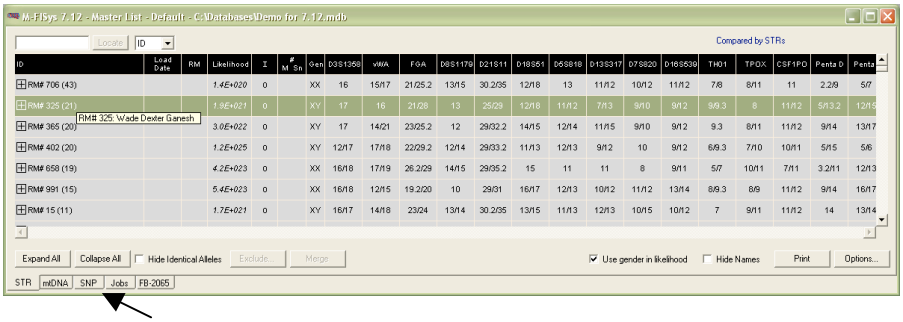
An important objective of the M-FISys design is to minimize hand entry of data, where human error can introduce problems that will be hard to detect by automated review. STR profiles can be read directly into M-FISys electronically without keying in information by hand.[9] mtDNA profiles can be loaded directly from Sequencher[10] and SNP profiles from electronic tables prepared by Orchid

[9] In the case of the WTC effort at the OCME, initial data was exported from CODIS and loaded electronically into M-FISys. When the CODIS step was retired from the WTC process, data was transferred from the GenoTyper™ program from Applied Biosystems (Foster City, CA). Elaine Mar, OCME Criminalist / Supervisor, World Trade Center DNA Identification Unit. Personal communication.
[10] Sequencher ™ (Gene Codes Corporation, Ann Arbor, MI) is a widely used program for general DNA sequencing, first introduced in 1991. In 1997, a specialized *"forensic build"* of Sequencher was developed for mitotyping at the request of the US Armed Forces DNA Identification Laboratory

Biosciences. STR profiles can also be exchanged between M-FISys and CODIS. Work is underway in collaboration with Doug Hares and Deborah Polanskey of the FBI Laboratory Division to establish a standard for exchanging mtDNA between Sequencher and a yet-to-be-released version of CODIS that is expected to begin supporting mtDNA data.

The extensive functionality in M-FISys is a double-edged sword: Information overload is a problem that is always considered during the design process. While a major disaster might require that all techniques available be brought to bear on the identification process, a Missing Persons program in a particular jurisdiction might focus on basic techniques unless a specific case requires a more aggressive laboratory approach. Furthermore, different interpreting analysts within a forensic lab may have different areas of expertise when reviewing DNA profiles. M-FISys allows STR, mtDNA and SNP profiles to be reviewed independently, while maintaining referential integrity between different assays performed on the same sample.

STRs are the most commonly used DNA profile, and the master match index defaults to an STR-based view. The first three tabs in the lower-left corner of the window (see arrow) allow the analyst to move between the STR-centric, mtDNA-centric and SNP-centric views of the data.



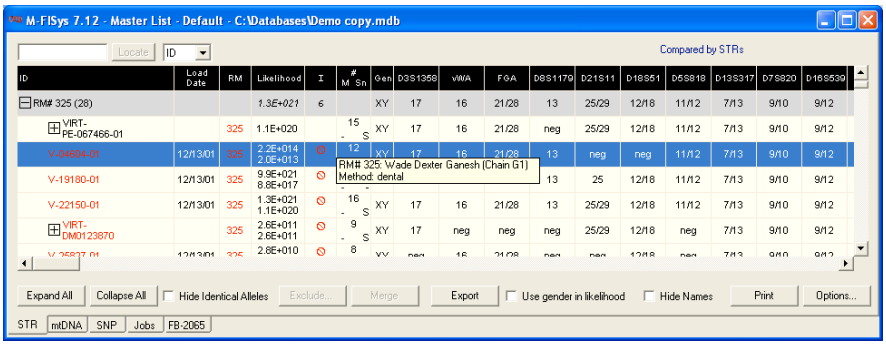
To orient the reader, each line represents one victim (missing person) with an associated RM number (RM stands for "Reported Missing"). The highlighted profile shows a simulated STR profile for a fictitious person named Wade Dexter Genesh, RM #325. The profile includes the 13 core loci (plus gender) used typically by US law enforcement, plus the Penta-D and Penta-E loci from Promega's PowerPlex® 16 multiplex STR system.[11] (M-FISys can be configured to support other STR markers as well, such as Y-STRs and loci used by Interpol/FSS.)

[AFDIL]. In this paper, the term Sequencher refers to the *forensic build* of Sequencher unless otherwise noted. Sequencher screen images in this paper are all taken from the *forensic build*.

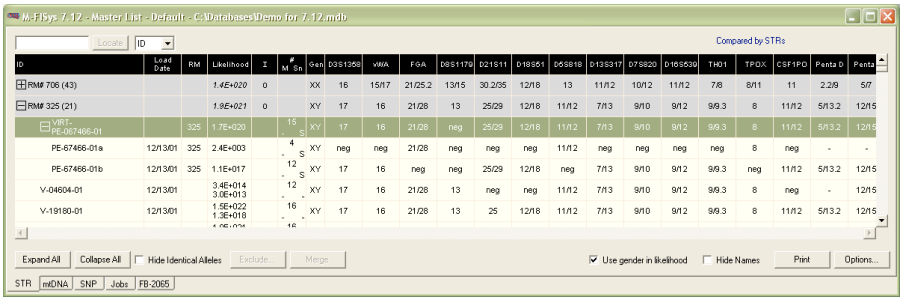
[11] Promega Corporation, Madison, WI; www.promega.com

The identifier in the left margin reads "RM# 325 (21)". The number '21' in parentheses indicates that this line represents an aggregate of twenty-one separate items with consistent profiles, for example, a reference exemplar plus 20 individually collected remains from the same person.

A mouse click on the turnstile discloses the substituent samples with their profiles.



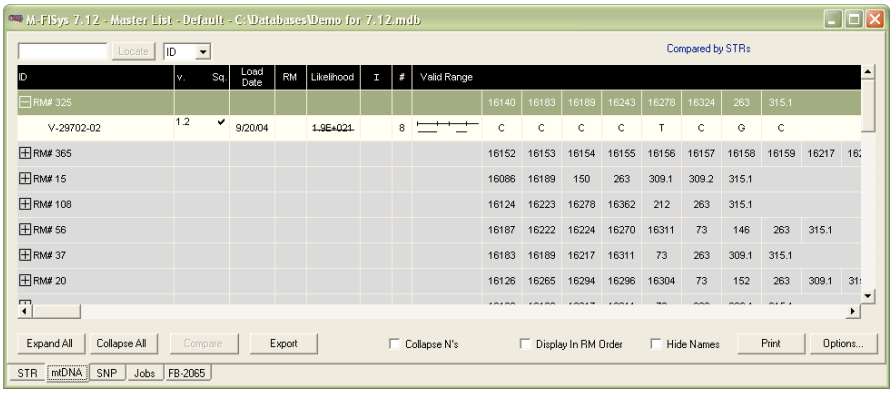
The first line in the aggregate (VIRT-PE-067466-01) is a "virtual" profile[12] of a Personal Effect [PE]. In this case, it represents two attempts to extract a DNA profile from a toothbrush that belonged to the missing person. Clicking on the turnstile for this sample would expose the individual attempts to type the item. You can see that even after two attempts, it was not possible to collect a full profile. The D8S1179 locus is listed as 'neg' in the composite or "virtual" profile.



[12] Cash HD, Hoyle JW, Sutton AJ. "Development Under Extreme Conditions: Forensic Bioinformatics in the wake of the World Trade Center Disaster." Pacific Symp Biocomput. 2003;:638-53. PMID: 12603064

All the other samples begin with 'V–', indicating that they are victim samples,[13] and each sample shows two different values in the likelihood column. In the highlighted row (V–0460401), the upper likelihood (2.2×10^{14}) is a measure of how unique this profile is based on a defined pop stats (allele frequency) database. The lower number (2.0×10^{13}) only counts those loci that overlap with the exemplar. Since the personal effect has no data at the D8S1179 locus, only the probative loci are combined using the product rule and this likelihood value is slightly lower.

Clicking on the mtDNA tab brings forward the mtDNA-centric view, with the same individual (RM#325) selected.



The program follows the accepted convention for displaying the mitochondrial sequence as a list of the nucleotides and base positions where the sequence varies from a reference sequence (the reference, such as the Anderson Sequence, can be specified). An insert, such as the common extra "C" after base position 315 is listed as "315.1 C". For matching purposes, the program tolerates errors in nomenclature for equivalent variants such as the extra C in the poly-cytosine region being reported as "314.1 C".

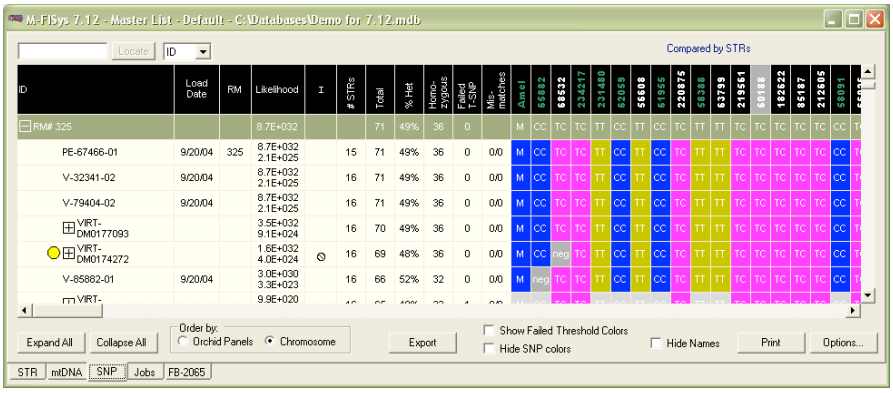
Raw mtDNA sequence data can be viewed and reviewed by an analyst trained in mitotyping. In the image below, the left window shows a highlighted variation from the rCRS[14] reference. At position 263, this individual types as "G" whereas the rCRS sequence has an "A". To the upper right, this difference is highlighted in the aligned sequences (the Reference Sequence is outlined), and raw

[13] The naming conventions, such as "V-" prefixes for victim samples, are not required by the program. They are conveniences for the purposed of our internal Software Quality Assurance process and are maintained here to make examples easier to describe in the narrative.
[14] rCRS is the Revised Cambridge Reference Sequence, the most commonly used mtDNA reference sequencing in general use in forensic biology labs, from "Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA" by Andrews, et. al., 1999

electropherograms of the forward and reverse sequencing reads are shown below. M-FISys presents this review of raw data by launching the Sequencher program on the same computer.

The SNP-centric view of the data shows profiles as a list of well characterized, bi-allelic SNPs. This data is simulated but uses the same panel of 70 SNPs developed by Orchid Biosciences for the OCME's WTC DNA Identification Unit.[15] Just like STR profiles, these samples can be compared one at a time against the database or a complete all-against-all comparison can be performed. Match statistics are provided, but the color coding of loci as TT homozygotes, CC homozyotes, or TC heterozygotes makes it easy to see the accuracy and fidelity of the matches with the naked eye.

[15] Dr. Robert Giles, Orchid Biosciences; Dallas, TX. Personal Communication.

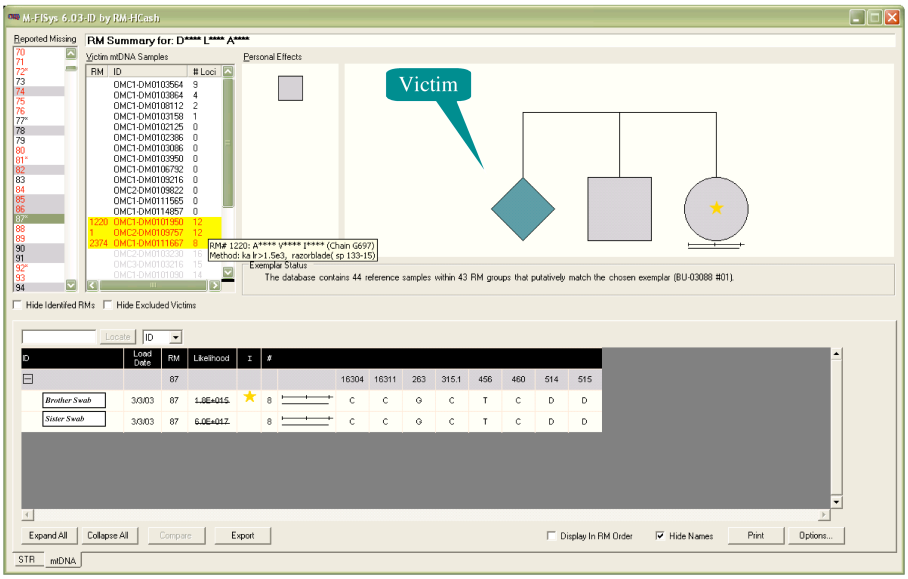


4 An Example of Integrating Data Views: Using mtDNA As a Screening Tool

For forensic Missing Persons databases, some in the scientific community advocate a protocol whereby mtDNA profiles would be used as a first-pass screening tool for data matching purposes, followed by more commonly used STR testing for potential matches. It is well established that mtDNA profiles of the hypervariable HV1 and HV2 regions lack the discriminating power of STR profiles at the 13 loci used in the United States, but in highly compromised remains, mtDNA may be more readily extracted in usable form. We offer no contribution to the debate on the efficacy of using mtDNA as an initial screen, but tools have been developed in M-FISys to support this approach, comparing STR profiles based on candidates derived from mtDNA matching.

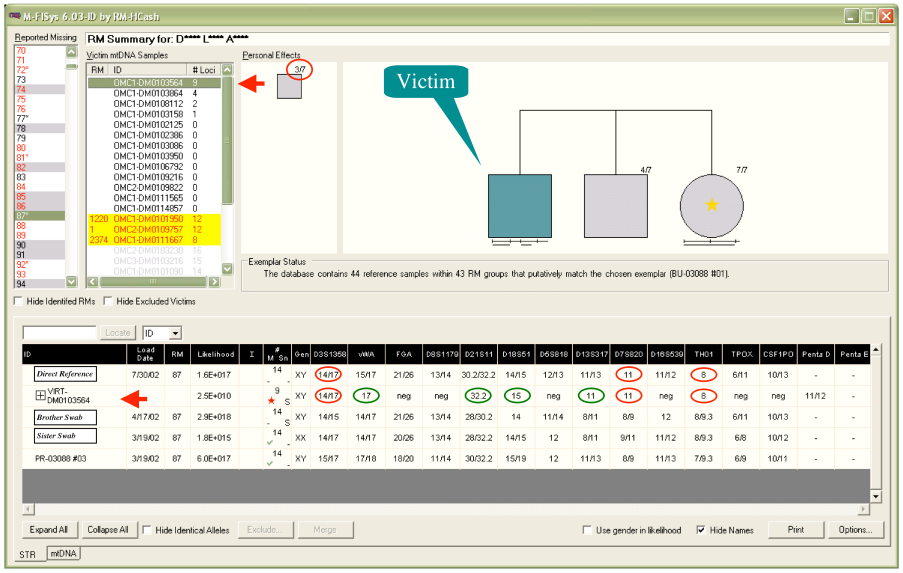
In the figure below, missing person number 8887 has been selected in the left most column. The next column to the right lists all physical remains that have been tested and shown to have consistent mtDNA profiles. Since mtDNA sequences of the HV1 and HV2 regions are inherently less discriminating than complete STR profiles, we would expect some of the mito matches to be merely coincidental. In fact, some of the adventitious matches are from remains that have already been shown (by STRs or other methods) to be other individuals, such as RMs 6620, 6666 and 6674, highlighted in yellow.[16] The mitotypes for the two available siblings are shown at the bottom of the window. Note that there is also a personal effect available for this decedent.

[16] It is a convention in M-FISys to "yellow flag" items with data conflicts by highlighting them in yellow. In this case, a search is being done for mitotypes that match RM 8887. Samples that match by mitotype but have otherwise been assigned to other RM identifications are not excluded from the display, but merely highlighted. This allows the QC team to continually search for errors in prior work.



Like the master list, this window has tabs at the bottom that allow a forensic scientist to move back and forth between STR—, mtDNA— and SNP—centric views (in this case, no SNP profiles are available so there are only two tabs in the lower left corner).

In the next view, the user has clicked on the STR tab. We are looking at the STR data for the same case at the bottom of the window. Of all the candidate mtDNA matches, the top item (OMC1-DM0163564) is selected and its STR profile is shown (red arrow), along with the profiles of the siblings and the personal effect. Note that only three out of seven overlapping loci match between the partial profile available from the human remains, and the personal effect. The matching loci are circled in red. But the remaining four overlapping loci (circled in green) show that in each case, a homozygous value in the tested tissue matches one allele in the heterozygous exemplar. Considering the degraded conditions that lead to the partial profile, it is reasonable to aggressively re-test the sample to see if the missing loci may simply be examples of allelic dropout. Neither the mtDNA nor the STR profiles gave enough information to make this a strong investigative lead, but by combining the two techniques, a new positive match is possible.



5 General Kinship Analysis

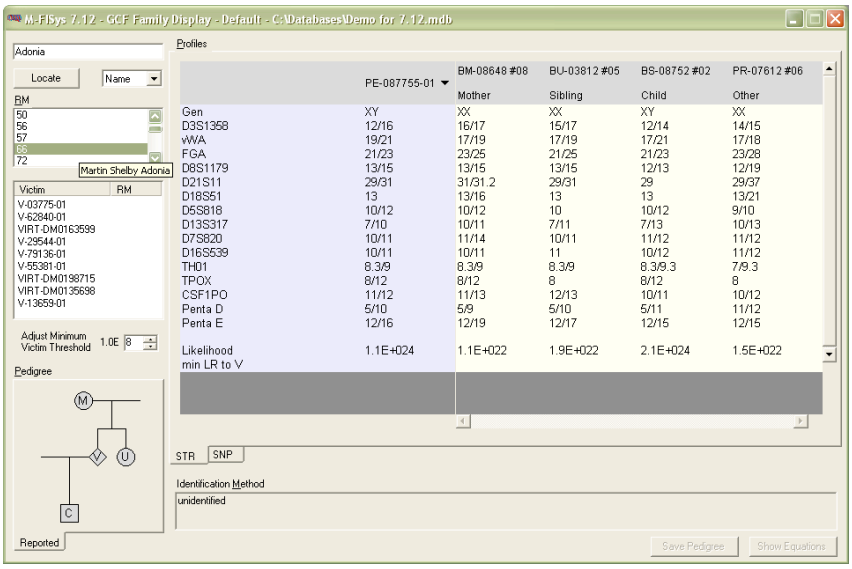
For simplicity, most technologies above have been illustrated using direct matches between remains, or between remains and personal exemplars. The missing persons implementation of M-FISys can perform sophisticated kinship analyses on arbitrarily complex pedigrees.

In the window below, all available missing persons are listed near the top left corner in a scrollable list by RM number. This can be searched by a number of criteria, and in this case we have searched for reference to a fictitious missing person named Martin Shelby Adonia, RM #66. In the pedigree drawing, the missing person is in the node marked with a V (for "Victim") and there are three family exemplars available: Mother, full sibling and a son.[17] There is a fourth family profile available and it is listed in the right-most column as PR-07612#06. In this simulation, we are assuming that we cannot confirm the relationship of that person to the victim.

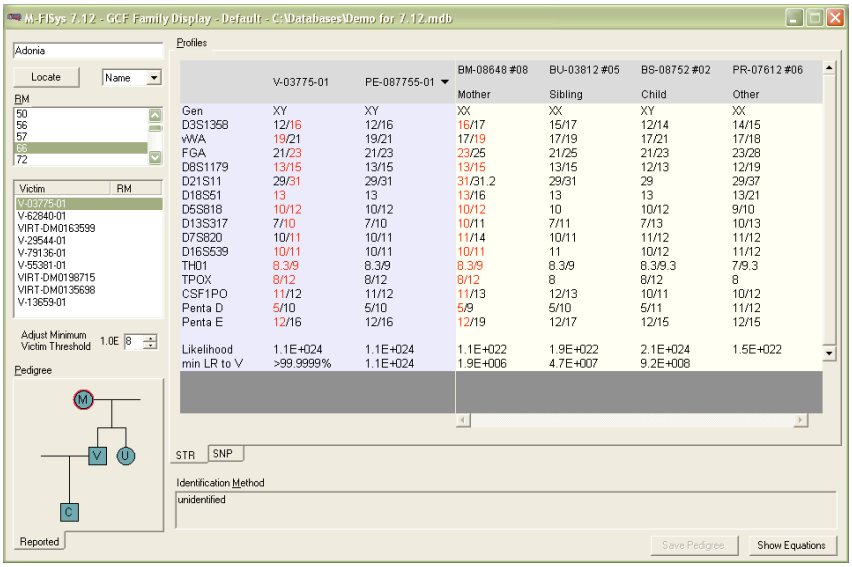
Once RM #66 has been selected, the 'Victim List,' immediately below, is populated with all of the candidate profiles that can fit into the kinship specified.

[17] The program alerts the operator if a reference pedigree is internally inconsistent. For example if the reported full sibling in this pedigree had a profile inconsistent with being a child of the reported mother, the program would alert the user to the conflict.

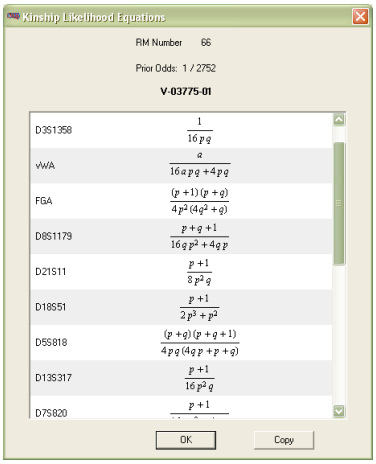
The forensic scientist can click on any of the candidates and review the kinship. Below each profile is a likelihood value. Initially, this is not related to the kinship likelihood ratio, but rather just shows a measure of the rarity of that individual profile in the specified population.



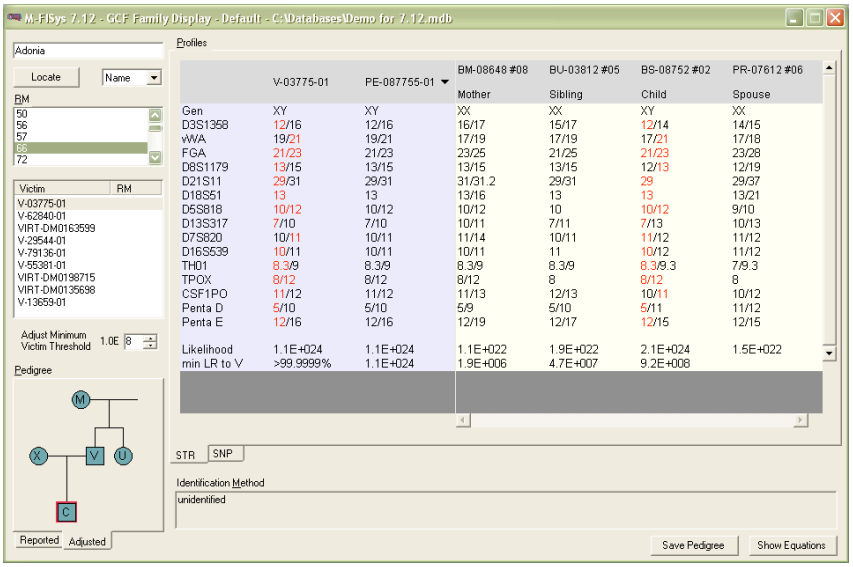
Once a sample is selected from the list of candidate matches, a great deal of additional information becomes available. In the next view, sample V—0377501 has been chosen. The mother node has been clicked, so all alleles shared between the postulated mother and victim are highlighted in red. In this case, there happens to also be a personal effect reference available (PE-0877555-01).



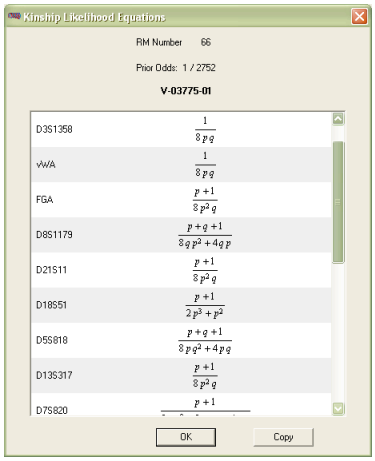
There is now a second row of likelihoods under the individual profiles. Under each kinship exemplar is the likelihood ratio [LR] of the specified relationship to the proposed victim. For example, the LR that profile BM-08648#09 is from the biological mother of selected victim sample V—0377501 is 1.9×10^6. Under the victim sample is a posterior probability (expressed as a percentage to distinguish it) that this victim fits the entire pedigree. If there is a need to see how this "overall" likelihood was calculated, the user can click on the Show Equations button in the lower right corner of the window. In this case, we are using prior odds of 1/2752.



Not all cases are so trivial and may require some expert manipulation by a geneticist to tease out the most valuable information. Recall that in our example there was one profile where the relationship was not known. By inspection, a forensic biologist might hypothesize that the swab came from the wife of the victim. The profile can be assigned to the relationship of victim's spouse. This will update the pedigree drawing and change label on the altered profile column from "other" to "spouse."



At the bottom of the pedigree drawing, note that there are now two tabs, allowing access to both the original REPORTED pedigree and the user-ADJUSTED pedigree. By highlighting the Child node, C, the alleles contributed by the proposed victim (father) are highlighted in red. The remaining alleles are all consistent with PR-07612#06, now labeled "Spouse," being the mother of that same child. The posterior probability does not change appreciably in this example, but the equations used to calculate that probability are slightly different. In cases with few kin references, this ability to reassign mislabeled kinship profiles can be the difference between a correct identification and none at all.



6 Assigning Multiple Roles to Samples

In many disasters, it is an expected tragedy that more than one member of a family may be killed. This situation may not be as common in a missing persons database, but certainly, it can be expected to occur. This creates a situation where 1) identification pedigrees for related decedents may overlap and 2) a sample may

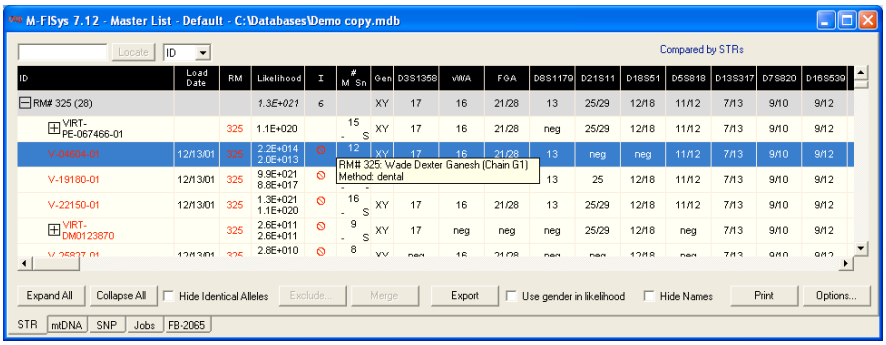
serve multiple *roles*. For our purposes, the "role" for a profile can be one or more of the following

- Remains (identified or unidentified)
- Direct Reference to a missing person
- Kinship Reference to a missing person

For example, if a mother and child are both among the missing persons, then a direct reference for the mother (for instance, a DNA profile from her toothbrush) can also be used as a maternal kinship reference for the child. If two brothers are listed as missing and one is identified by non DNA methods such as dental matching or fingerprints, then a DNA profile from that person's blood can also serve the role of a kinship reference for his sibling.

In some historical cases, the issue of multiple roles was obviated by duplicating samples. For example, this approach was used in some cases during the WTC effort. An example of this would be taking multiple swabs from the same mother to be used as exemplars for each of her missing children.[18] For use in missing persons databases, we have established a process for designating multiple roles for the same profile.

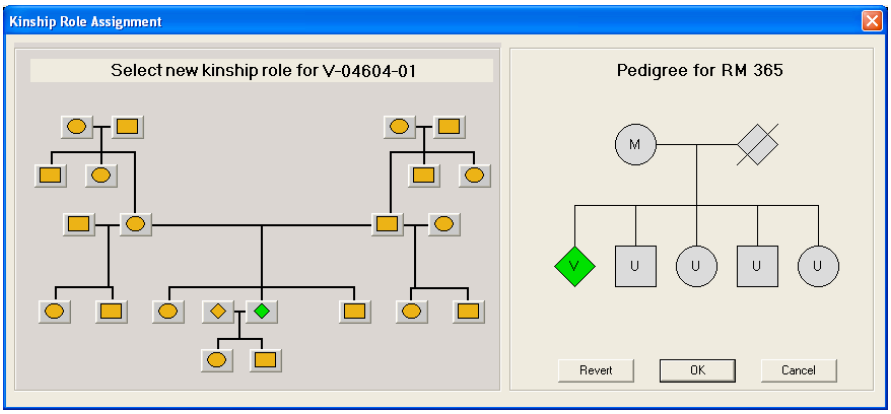
In the case below (simulated), the remains labeled V-04604-01 were originally identified through dental records as missing person #325, Wade Ganesh. It happens that Wade is also the father of missing person #365.



Thus, the post mortem STR profile from these remains can be assigned the additional role of Father exemplar to missing person #365. Sample V-04604 is

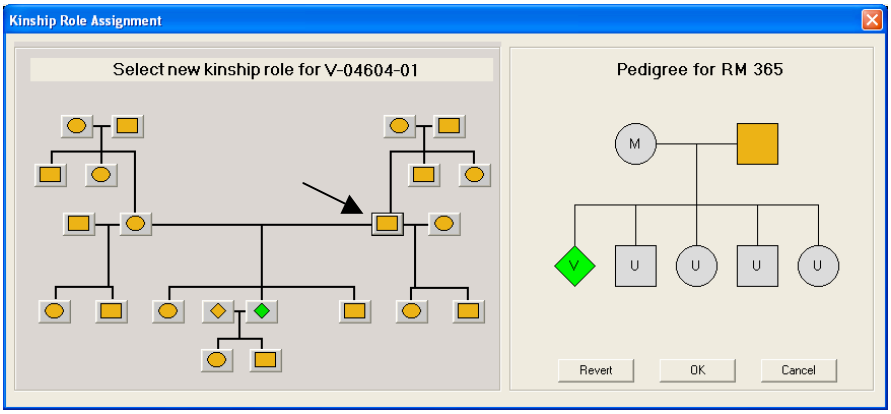
[18] Michael J. Hennessey, presentation at *13th International Symposium on Human Identification – 2002*

selected and a new Kinship role is specified, saying that the new relationship will be with missing person 365. The following relationship editor appears.

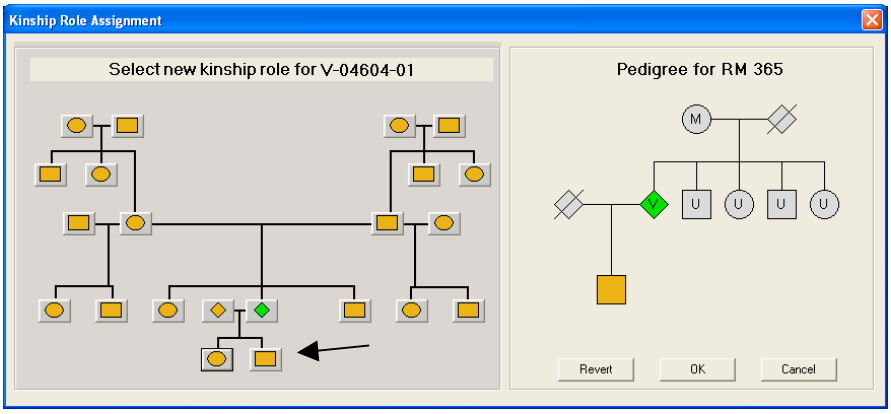


On the right, we show all the relatives that are currently available for person 365. Note that the Father, also a decedent, is not yet available for typing.

The green node marked "V" is the *victim* or *missing person,* 365. The left side of this window can be thought of as simply a palette for specifying the new relationship role for sample V-04604-01. By clicking on the father node (indicated by the arrow in the picture below), the profile will be added as a father exemplar for sample 365, as shown in the right side of the same window.



Had a different node been selected from the palette, for instance the victim's son (indicated by the arrow in the next figure) the pedigree for missing person #365 would have updated differently, as shown.



7 Summary

It is said that necessity is the mother of invention. Tools developed for the New York City Medical Examiner in the wake of the World Trade Center disaster necessarily advanced the state of the art in forensic human identification. Information management tools were only one part of that advance, but they are a critical component that can be redeployed for use in other investigations. The effort to identify 2,749 victims has constantly been challenged with new problems to solve. Many of the solutions are applicable to the general problem of identifying John Doe remains and matching forensic profiles to missing persons exemplars.

We have not tried to describe every aspect of M-FISys, nor to share every lesson learned over the course of this project. When asked to extend our work to build a Missing Persons identification program that can be scaled up for a disaster, we recognized certain key elements as basic requirements. These included

- Handling multiple testing technologies (STR, mtDNA and SNPs)
- Combining those technologies so that the information is accessible, without overwhelming the analyst.
- Supporting an administrative review of meta data
- Keeping an electronic audit trail of analyst initiated edits
- Collapsing matching identical profiles to reduce the amount of data being reviewed
- Providing access to primary laboratory data
- Supporting a protocol of screening matches on the bases of mtDNA profiles
- Performing complex kinship analysis, and
- Allowing profiles to serve more than one identification role.

We know that there will be future disasters, be they natural or man-made. And even today there are missing persons whose remains have been found around the world but who cannot yet be named. One can only hope that out of the disaster of 9-11 comes new knowledge and technology that will bring some comfort to future families touched by tragedy by bringing their loved ones home.

Acknowledgments

We express our thanks to Bob Shaler, Georger Carmody and Bruce Budowle for reviewing and commenting on this manuscript. Thanks to the many contributors including the primary process analysis, development and support team: Tracy Beeson, Debra Cash, Nama Eden, Peter Fattori, Lucy Hadden, Dan Harrington, Eric Heikkila, Jonathan Hoyle, Jeff Ingalls, Anna Khizhnyak, Tom Kubit, Anna Korn, Simon Mercer, Peter Metz, Judy Nolan, Gregory Poth, David Relyea, John Snyder, Matt Smith, Francis Sullivan, Amy Sutton, our *Extreme Programming* coach Bill Wake, and Glenda Wilson. Other truly major contributors have included staff at Gene Codes Corporation, Gene Codes Forensics, the New York State Police FIC, AFDIL, Bode Technology Group, Orchid Biosciences and the OCME too numerous to name. Special thanks to Vickie Bair who did not get the job she signed up for, the OCME's *WTC DNA Identification Unit* Supervisor Elaine Mar and her tireless data-analysis team, Jeanine Baisch and Bob Giles at Orchid, George Carmody (again) for his always generous technical assistance on short notice, and to the many inspiring families we have met along the way who lost loved ones on September 11.

References

1. S. Anderson et al, *Sequence and Organization of the Human Mitochondrial Genome* (Nature 290: 457-465, 1981).
2. R.M. Andrews, et al, *Reanalysis and Revision of the Cambridage Reference Sequence for Human Mitochondrial DNA* (Nat Genet 23(2): 147, 1999)
3. P. Awadalla et al, *Linkage Disequilibrium and Recombination in Hominid Mitochondrial DNA* (Science 286: 2524-2525, 1999).
4. D. Balding and P. Donnelly, *Inferring Identity from DNA Profile Evidence* (Proc. Natl. Acad. Sci. USA 92: 11741-11745, 1995).
5. J. Ballantyne et al, *DNA Technology and Forensic Science* (Cold Spring Harbor Laboratory Press, 1989).
6. H. Baum, *PCR Statistics: STR13.xls* (personal communication, Office of Chief Medical Examiner, 6/14/99).

7. CH Brenner, *Symbolic Kinship Program*. (Genetics. 1997 Feb;145(2):535-42) PMID: 9071605
8. B. Budowle et al, *Population Data on the Thirteen CODIS Core Short Tandem Repeat Loci in African-American, U.S. Caucasians, Hispanics, Bahamians, Jamaicans, and Trinidadians* (J. Forensic Sci. 44: 1277-1286, 1999).
9. J. Butler, *Forensic DNA Typing* (Academic Press, 2001).
10. G. Carmody, *KinTest - CODIS 13 Core Loci* (personal communication, 2001).
11. HD Cash, JW Hoyle, AJ Sutton, *Development Under Extreme Conditions: Forensic Bioinformatics in the wake of the World Trade Center Disaster* (Pacific Symp Biocomput. 2003;:638-53. PMID: 12603064)
12. I. Evett and B. Weir, *Interpreting DNA Evidence* (Sinauer Associates, 1998).
13. Federal Bureau of Investigations, *National DNA Index Systems (NDIS) Procedures Manual* (1999).
14. Federal Bureau of Investigations, *State DNA Database Statutes* (1999).
15. M. Hennessey, *World Trade Center and DNA Identifications: The Administrative Review* (*Thirteenth International Symposium on Human Identification – 2002)* http://www.promega.com/geneticidproc/ussymp13proc/contents/hennesseyrev1.pdf
16. A. Jeffreys et al, *Hypervariable Minisattellite Regions in Human DNA* (Nature 314: 67-72, 1985).
17. A. Jeffreys et al, *Individual Specific "Fingerprints" of Human DNA* (Nature 316: 75-79, 1985).
18. R. Jensen, Mass Fatality and Casualty Incidents (CRC Press, LLC, 2000)
19. L. Kirby, *DNA Fingerprinting* (Stockton Press, 1990).
20. National Institute of Justice, *The Future of Forensic DNA Testing* (2000).
21. National Research Council, *The Evaluation of Forensic DNA Evidence* (National Academy Press, 1996).
22. National Research Council, *DNA Technology in Forensic Science* (National Academy Press, 1992).
23. N. Rudin and K. Inman, *Forensic DNA Analysis* (2nd Edition, CRC Press, 2002).
24. R. Saferstein, *Forensic Science Handbook -Volume 1* (2nd Edition, Prentice Hall, 1982, 2002).