J Forensic Sci, July 2007, Vol. 52, No. 4
doi: 10.1111/j.1556-4029.2007.00456.x
Available online at: www.blackwell-synergy.com

Benoît Leclair,[1] Ph.D.; Robert Shaler,[2†] Ph.D.; George R. Carmody,[3] Ph.D.; Kristilyn Eliason,[1] B.Sc.; Brant C. Hendrickson,[1‡] M.Sc.; Thad Judkins,[1] B.Sc.; Michael J. Norton,[1§] B.Sc.; Christopher Sears,[4] Ph.D.; and Tom Scholl,[1,5‡] Ph.D.

Bioinformatics and Human Identification in Mass Fatality Incidents: The World Trade Center Disaster*

ABSTRACT: Victim identification initiatives undertaken in the wake of Mass Fatality Incidents (MFIs) where high-body fragmentation has been sustained are often dependent on DNA typing technologies to complete their mandate. The success of these endeavors is linked to the choice of DNA typing methods and the bioinformatic tools required to make the necessary associations. Several bioinformatic tools were developed to assist with the identification of the victims of the World Trade Center attacks, one of the most complex incidents to date. This report describes one of these tools, the Mass Disaster Kinship Analysis Program (MDKAP), a pair-wise comparison software designed to handle large numbers of complete or partial Short Tandem Repeats (STR) genotypes, and infer identity of, or biological relationships between tested samples. The software performs all functions required to take full advantage of the information content of processed genotypic data sets from large-scale MFIs, including the collapse of victims data sets, remains re-association, virtual genotype generation through gap-filling, parentage trio searching, and a consistency check of reported/inferred biological relationships within families. Although very few WTC victims were genetically related, the software can detect parentage trios from within a victim's genotype data set through a nontriangulated approach that screens all possible parentage trios. All software-inferred relationships from WTC data were confirmed by independent statistical analysis. With a 13 STR loci complement, a fortuitous parentage trio (FPT) involving nonrelated individuals was detected. Additional STR loci would be required to reduce the risk of an FPT going undetected in large-scale MFIs involving related individuals among the victims. Kinship analysis has proven successful in this incident but its continued success in larger scale MFIs is contingent on the use of a sufficient number of STR loci to reduce the risk of undetected FPTs, the use of mtDNA and Y-STRs to confirm parentage and of bioinformatics that can support large-scale comparative genotyping schemes capable of detecting parentage trios from within a group of related victims.

KEYWORDS: forensic science, DNA typing, polymerase chain reaction, mass fatality incidents, mass disaster, software, D3S1358, vWA, FGA, D8S1179, D21S11, D18S51, D5S818, D13S317, D7S820, THOI, TPOX CSF1PO, D16S539

The level of complexity of a victim identification initiative undertaken in the aftermath of any MFI can vary tremendously and is dependent upon a number of variables that generally include the number of victims, the extent of body fragmentation, the number and extent of deterioration of remains that can be recovered and the availability/reliability of reference biological material and records collected to help with the identification of the missing individuals. These incident-specific variables largely drive the selection of victim identification technologies. Historically, conventional procedures such as visual identification, matching dental/X-ray/fingerprint records, and personal effects found with recovered remains have been the mainstay of identification efforts for MFIs (1–3), largely due to their reliability, ease of implementation, speediness and cost effectiveness. These conventional procedures adequately and rapidly address the identification needs of most circumstances where little or no body fragmentation has occurred. In these situations, the more costly and time-consuming DNA typing technologies are invoked only in cases where victims cannot be visually identified and/or reference documentation required for conventional identification procedures is unavailable. In situations where some body fragmentation has occurred, DNA typing technologies can provide the data required for the re-association of fragmented remains. When only partial recovery of highly fragmented remains is anticipated, DNA typing often proves to be the only successful identification modality for many victims as it is possible for this technology to derive identity information from highly compromised samples, regardless of tissue type. Since the Spitzbergen air crash of August 1996 (4), the usefulness of DNA typing technologies in victim identification initiatives has been amply demonstrated in numerous MFIs.

Whereas MFIs are generally accidental in nature and result from transportation mishaps or severe climatic events striking inhabited areas, some incidents are deliberately instigated through armed conflicts (e.g., mass graves in Bosnia) or terrorism (e.g., 9/11

[1] Myriad Genetic Laboratories, Inc., 320 Wakara Way, Salt Lake City, UT 84108.

[2] Office of the Chief Medical Examiner, 520 First Avenue, New York City, NY 10016.

[3] Department of Biology, Carleton University, 1125 Colonel By Drive, Ottawa, ON, Canada K1S 5B6.

[4] Ananomouse Corporation, 25 Paul Street, Suite 3, Boston, MA 02472.

[5] Department of Pathology, University of Utah, 30 N 1900 E, Salt Lake City, UT 84112.

[†] Present address: The Pennsylvania State University, 107 Whitmore Lab, University Park, PA 16802.

[‡] Present address: Genzyme Genetics, 3400 Computer Drive, Westborough, MA 01581.

[§] Present address: School of Medicine, Case Western Reserve University, 10900 Euclid Avenue, Cleveland, OH 44106.

[*] Presented at the 20th Congress of the International Society for Forensic Genetic, September 9–13, 2003, in Arcachon, France.

Received 17 Jun. 2006; and in revised form 10 Oct. 2006; accepted 30 Dec. 2006; published 24 May 2007.

attacks). On September 11, 2001, the twin towers of the World Trade Center (WTC) in New York City collapsed after being hit by hijacked civilian airliners. Including the occupants of the airliners and the remaining occupants of the largely evacuated towers, 2749 individuals are considered to have perished in this incident. The dynamics of the towers' collapse led to severe body fragmentation of the victims. The rapid recovery of remains proved impossible to achieve as the process of sifting through an estimated 1.6 million tons of tower debris in search of remains lasted 10 months. Fires, set ablaze by the crash of the airliners and feeding on jet fuel and combustible material within the tower debris, burned throughout the recovery operation, subjecting many remains to temperatures estimated to have exceeded 1000°C in some areas. The water used to extinguish these fires and cool the debris field introduced moisture in a warm environment, enhancing bacterial decay of the remains. These combined site-related conditions inflicted much damage on the human remains trapped in tower debris, a large percentage of the 19,979 recovered remains yielding partial or no genotypic information during laboratory analysis.

The process of remains identification depended heavily on DNA typing technologies. Source attribution required that algorithms/software used for previous MFIs and new software created for the WTC incident be adapted to process large numbers of partial genotypes. To mitigate the risks associated with rapid and continuous software development throughout the life of a large and complex identification initiative, three different comparative genotyping software packages were developed in parallel to provide redundancy and confirmation of results by concordance (5–7). We report here on one of these software approaches and demonstrate the capabilities of large-scale comparative genotyping and kinship analysis for use in MFI victim identification on the scale of WTC. We also underscore the limitations of kinship analysis for MFIs with respect to scale and complexity.

Materials and Methods

Genotypic Data

In this report, the genotype derived from biological trace material recovered from a personal effect purported to have belonged to a victim is referred to as a "PE" genotype, the one derived from a reference biological sample obtained from a next-of-kin is referred to as a "NOK" genotype. In accordance with standard forensic naming conventions, genotypes derived from remains are referred to as Questioned or "Q" genotypes, those derived from PEs and NOKs are referred to as Known or "K" genotypes.

The development and testing of the software was accomplished with computer-generated STR genotypic data sets simulating large numbers of virtual families. STR genotypic data submitted for the purpose of identification were provided by the Office of the Chief Medical Examiner (OCME) of New York City and the New York State Police (NYSP) Forensic Laboratories as modified Common Message Format (CMF) files. For each Q genotypic data record, a sequential number was provided as a record name. For each K genotypic data record, an anonymized record name provided sample type (i.e., PE, NOK), purported biological relationship to the victim for NOKs, and a "RM" (Reported Missing) number referring to a single victim to allow the software to regroup all reference samples relevant to a given victim and support kinship analysis functionalities. STR data provided by the OCME and NYSP included all reportable alleles within the 13 core CODIS STR loci complement, with Promega's pentanucleotide repeat loci data appearing later in the initiative in some data sets.

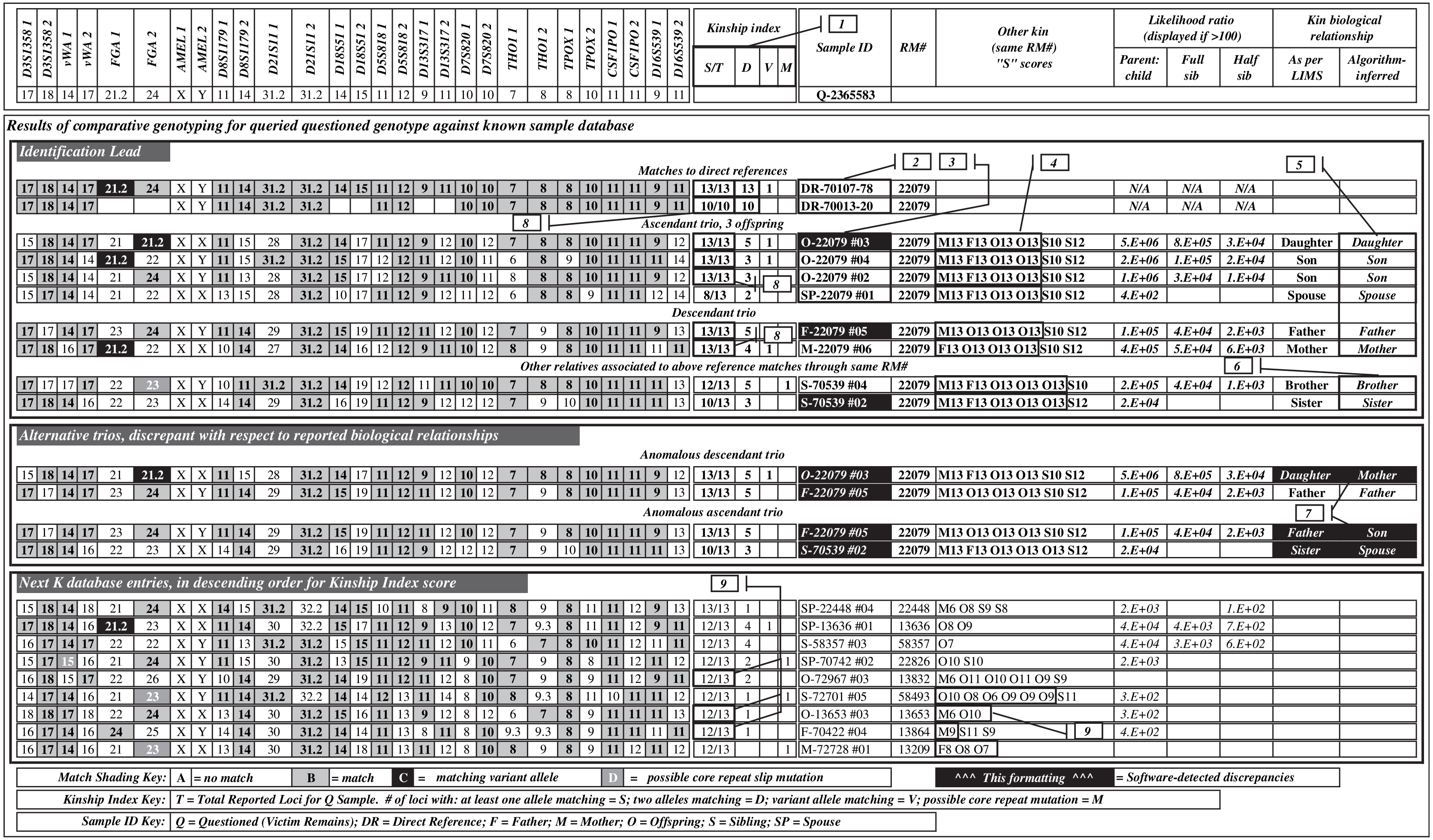
Genotype Comparison Software

Short tandem repeats, mitochondrial DNA, and single nucleotide polymorphism data were generated in this victim identification effort. STR DNA typing being the most discriminating platform, all successful DNA identifications were made with this type of data. The simple systematic pair-wise STR genotype comparison scheme that proved successful for the Swissair Flight 111 MFI (8,9) was the starting point for the development of algorithms for the Mass Disaster Kinship Analysis Program (MDKAP) used for the WTC incident. Briefly, this scheme permits the detection of: (1) direct matches of Q to PE genotypes, where alleles at nearly all tested loci are expected to match, with the possible exception of allele drop-outs in compromised Q samples; (2) parent – offspring (P/O) relationships where, according to Mendelian inheritance rules, compared genotypes are expected to share at least one allele at all loci being tested; (3) higher than average sharing of alleles between most siblings, by virtue of shared parentage. To facilitate P/O relationship detection, interrogated data sets were ranked on two tiers: (1) the number of loci at which at least one allele is shared (i.e., Single Match or "S" score) between query and database entry, (2) the number of loci at which both alleles are shared (i.e., Double Match or "D" score) between query and database entry. These basic scores are supplemented with a Variant score ("V") where matching alleles encountered at a frequency of <1% in a reference Caucasian population database (10) are tabulated. Finally, a Mutation score ("M") was incremented when a potential core repeat mutation [defined as an entry where at least one allele is shared at all loci except one locus where a match would be declared if the allele was offset by one repeat, the most common type of core slip mutation [http://www.aabb.org/Documents/Accreditation/Parentage_Testing_Accreditation_Program/rtannrpt04.pdf] was detected. After comparison with a queried Q genotype, genotypes in the interrogated K data set were ranked in decreasing order of score, on four tiers: S, D, V, and M. Fig. 1 presents an example of a typical score report available to OCME data reviewers.

To attend the complexities of the WTC incident, the approach was complemented with several additional capabilities. These included: (1) a data collapsing routine that regrouped Q genotypes sharing identical/near-identical complete or partial data, (2) a composite genotype generator that produced more complete "virtual" genotypes out of groups of partial overlapping genotypes, (3) a parentage trio searching routine, (4) a consistency check feature that detected discrepancies between NOK self-reported biological relationships and kinship data, and (5) a likelihood ratio (LR) calculation routine for pair-wise relationships. Finally, algorithms were built to provide overnight processing capability for the anticipated 240 million pair-wise comparisons required per identification round (see Table 1), a 1400-fold increase over the number of comparisons required in the Swissair situation (9). These capabilities were coded in Visual Basic for Applications (VBA) and executed on a Microsoft Excel platform.

Genotypic Data Collapse

In order to return the recovered remains to families, morgue officials need the ability to re-associate remains, a task greatly facilitated by DNA typing. For the WTC incident, the regrouping of remains according to genotype also answered the crucial need to reduce the size of the genotypic data set, to hasten the kinship analysis process, to contain data processing time. In a context where



D3S1358 1	D3S1358 2	vWA 1	vWA 2	FGA 1	FGA 2	AMEL 1	AMEL 2	D8S1179 1	D8S1179 2	D21S11 1	D21S11 2	D18S51 1	D18S51 2	D5S818 1	D5S818 2	D13S317 1	D13S317 2	D7S820 1	D7S820 2	THO1 1	THO1 2	TPOX 1	TPOX 2	CSF1PO 1	CSF1PO 2	D16S539 1	D16S539 2	Kinship index				Sample ID	RM#	Other kin (same RM#) "S" scores	Likelihood ratio (displayed if >100)			Kin biological relationship	
																												S/T	D	V	M				Parent: child	Full sib	Half sib	As per LIMS	Algorithm-inferred
17	18	14	17	21.2	24	X	Y	11	14	31.2	31.2	14	15	11	12	9	11	10	10	7	8	8	10	11	11	9	11					Q-2365583							
Results of comparative genotyping for queried questioned genotype against known sample database																																							
Identification Lead																																							
Matches to direct references																																							
17	18	14	17	21.2	24	X	Y	11	14	31.2	31.2	14	15	11	12	9	11	10	10	7	8	8	10	11	11	9	11	13/13	13	1		DR-70107-78	22079		N/A	N/A	N/A		
17	18	14	17			X	Y	11	14	31.2	31.2			11	12			10	10	7	8	8	10	11	11	9	11	10/10	10			DR-70013-20	22079		N/A	N/A	N/A		
Ascendant trio, 3 offspring																																							
15	18	14	17	21	21.2	X	X	11	15	28	31.2	14	17	11	12	9	12	10	12	7	8	8	10	11	11	9	12	13/13	5	1		O-22079 #03	22079	M13 F13 O13 O13 S10 S12	5.E+06	8.E+05	3.E+04	Daughter	Daughter
17	18	14	14	21.2	22	X	Y	11	15	31.2	31.2	15	17	12	12	11	12	10	11	6	8	9	10	11	11	11	14	13/13	3	1		O-22079 #04	22079	M13 F13 O13 O13 S10 S12	2.E+06	1.E+05	2.E+04	Son	Son
15	18	14	14	21	24	X	Y	11	13	28	31.2	15	17	12	12	9	11	10	11	8	8	8	10	11	11	9	12	13/13	3			O-22079 #02	22079	M13 F13 O13 O13 S10 S12	1.E+06	3.E+04	1.E+04	Son	Son
15	17	14	14	21	22	X	X	13	15	28	31.2	10	17	11	12	9	12	11	12	6	8	8	9	11	11	12	14	8/13	2			SP-22079 #01	22079	M13 F13 O13 O13 S10 S12	4.E+02			Spouse	Spouse
Descendant trio																																							
17	17	14	17	23	24	X	Y	11	14	29	31.2	15	19	11	12	11	12	10	12	7	9	8	10	11	11	9	13	13/13	5			F-22079 #05	22079	M13 O13 O13 O13 S10 S12	1.E+05	4.E+04	2.E+03	Father	Father
17	18	16	17	21.2	22	X	X	10	14	27	31.2	14	16	12	12	9	11	10	12	8	9	8	10	11	11	11	11	13/13	4	1		M-22079 #06	22079	F13 O13 O13 O13 S10 S12	4.E+05	5.E+04	6.E+03	Mother	Mother
Other relatives associated to above reference matches through same RM#																																							
17	17	17	17	22	23	X	Y	10	11	31.2	31.2	14	19	12	12	11	11	10	10	7	8	8	10	11	11	11	13	12/13	5		1	S-70539 #04	22079	M13 F13 O13 O13 O13 S10	2.E+05	4.E+04	1.E+03	Brother	Brother
17	18	14	16	22	23	X	X	14	14	29	31.2	16	19	11	12	9	12	12	12	7	9	10	10	11	11	11	13	10/13	3			S-70539 #02	22079	M13 F13 O13 O13 O13 S12	2.E+04			Sister	Sister
Alternative trios, discrepant with respect to reported biological relationships																																							
Anomalous descendant trio																																							
15	18	14	17	21	21.2	X	X	11	15	28	31.2	14	17	11	12	9	12	10	12	7	8	8	10	11	11	9	12	13/13	5	1		O-22079 #03	22079	M13 F13 O13 O13 S10 S12	5.E+06	8.E+05	3.E+04	Daughter	Mother
17	17	14	17	23	24	X	Y	11	14	29	31.2	15	19	11	12	11	12	10	12	7	9	8	10	11	11	9	13	13/13	5			F-22079 #05	22079	M13 O13 O13 O13 S10 S12	1.E+05	4.E+04	2.E+03	Father	Father
Anomalous ascendant trio																																							
17	17	14	17	23	24	X	Y	11	14	29	31.2	15	19	11	12	11	12	10	12	7	9	8	10	11	11	9	13	13/13	5			F-22079 #05	22079	M13 O13 O13 O13 S10 S12	1.E+05	4.E+04	2.E+03	Father	Son
17	18	14	16	22	23	X	X	14	14	29	31.2	16	19	11	12	9	12	12	12	7	9	10	10	11	11	11	13	10/13	3			S-70539 #02	22079	M13 F13 O13 O13 O13 S12	2.E+04			Sister	Spouse
Next K database entries, in descending order for Kinship Index score																																							
15	18	14	18	21	24	X	X	14	15	31.2	32.2	14	15	10	11	8	9	10	11	8	9	8	11	11	12	9	13	13/13	1			SP-22448 #04	22448	M6 O8 S9 S8	2.E+03		1.E+02		
17	18	14	16	21.2	23	X	X	11	14	30	32.2	15	17	11	12	9	13	10	12	7	9.3	8	11	11	12	9	11	12/13	4	1		SP-13636 #01	13636	O8 O9	4.E+04	4.E+03	7.E+02		
16	17	14	17	22	22	X	Y	11	13	31.2	31.2	15	15	11	12	11	12	10	11	6	7	8	10	11	12	11	11	12/13	4			S-58357 #03	58357	O7	4.E+04	3.E+03	6.E+02		
15	17	15	16	21	24	X	Y	11	15	30	31.2	13	15	11	12	9	11	9	10	7	9	8	8	11	12	11	12	12/13	2		1	SP-70742 #02	22826	O10 S10	2.E+03				
16	18	15	17	22	26	X	Y	10	14	29	31.2	14	19	11	12	11	12	8	10	7	9	8	9	11	12	9	11	12/13	2			O-72967 #03	13832	M6 O11 O10 O11 O9 S9					
14	17	14	16	21	23	X	Y	11	14	31.2	32.2	14	14	12	13	11	14	8	10	8	9.3	8	11	10	11	11	12	12/13	1		1	S-72701 #05	58493	O10 O8 O6 O9 O9 O9 S11	3.E+02				
18	18	17	18	22	24	X	X	13	14	30	31.2	15	16	11	13	9	12	8	12	6	7	8	9	11	11	11	13	12/13	1			O-13653 #03	13653	M6 O10	3.E+02				
16	17	14	16	24	25	X	Y	14	14	30	31.2	14	14	11	13	8	11	8	10	9.3	9.3	8	9	11	11	11	11	12/13	1			F-70422 #04	13864	M9 S11 S9	4.E+02				
16	17	14	16	21	23	X	X	13	14	30	31.2	14	18	11	13	11	12	8	10	8	9	8	9	11	12	11	12	12/13			1	M-72728 #01	13209	F8 O8 O7					

Match Shading Key: A = no match; B = match; C = matching variant allele; D = possible core repeat slip mutation

^^^ This formatting ^^^ = Software-detected discrepancies

Kinship Index Key: T = Total Reported Loci for Q Sample. # of loci with: at least one allele matching = S; two alleles matching = D; variant allele matching = V; possible core repeat mutation = M

Sample ID Key: Q = Questioned (Victim Remains); DR = Direct Reference; F = Father; M = Mother; O = Offspring; S = Sibling; SP = Spouse

FIG. 1—*Example of a score report. (In order to protect the privacy of the individuals involved, sample names and identification numbers have been anonymized in all figures displaying genotypes throughout this manuscript.) This figure provides an example of a typical K score report for a queried Q genotype. There were over 2300 Q and 11,000 K genotypes in the respective databases at the end of the ID initiative. (1) Scored K genotypes are ranked in decreasing order of score, on four tiers: S, D, V and M. See Kinship Index key. (2) Sample name of DR samples reflect the batch and tube number. (3) O-22079 #03,'' "O" = Offspring, "22079" = original family #, sometimes same as RM#, "# 03" = third individual collected in this family's pedigree. (4) Consistency check: True ascendants and descendants of a victim are expected to display a "S" score of n out of n victim reportable loci, or n – 1/n if a potential core repeat slip mutation was detected at the nonmatching locus. (5) Only if parentage trio algorithm is successful at accounting for all alleles of the offspring of a trio is a relationship inference made. (6) Only if a parentage trio(s) is/are detected in the same family is a sibling inference made. (7) If the inference of the parentage trio detection algorithm does not concur with the reported relationship, results are highlighted. Note that the same individuals are involved in nonanomalous trios in the Identification Lead box. (8) Qualifying "S" scores for DR and purported ascendants/descendants. (9) Disqualifying "S" scores for purported ascendants/descendants.*

TABLE 1—*MDKAP performance statistics under different scenarios.*

	Incident variables				
	Swissair Flight 111				
	As processed in 1998	***processed today*[1]**	**WTC**	***Scenario #1*[2]**	***Scenario #2***
# of victims	229	*229*	2749	*50000*	*1.0E + 06*
# of typed remains	1278	*1278*	19979	*363387*	*7.3E + 06*
# NOKs	310	*310*	6854	*124664*	*2.5E + 06*
# PEs	45	*45*	4242	*77155*	*1.5E + 06*
	Number of pair-wise comparisons				
Remains data collapse (1)		*8.2E + 05*	2.0E + 08	*6.6E + 10*	*2.6E + 13*
Qs vs Qs (2)	2.6E + 04	*2.6E + 04*	3.8E + 06	*1.2E + 09*	*5.0E + 11*
Qs vs NOKs (3)	7.1E + 04	*7.1E + 04*	1.9E + 07	*6.2E + 09*	*2.5E + 12*
Qs vs PEs (4)	1.0E + 04	*1.0E + 04*	1.2E + 07	*3.9E + 09*	*1.5E + 12*
NOKs vs Qs (5)	7.1E + 04				
NOKs vs NOKs (6)		*4.8E + 04*	2.3E + 07	*7.8E + 09*	*3.1E + 12*
	Number of pair-wise comparisons performed for Parentage Trio detection				
Events without related victims					
Total trios, potential		*3.3E + 07*	1.9E + 11	*1.2E + 15*	*9.3E + 18*
Total trios searched (7)		*1.5E + 04*	2.6E + 06	*9.6E + 09*	*7.5E + 13*
Events with related victims					
Total trios, potential		*5.1E + 07*	2.2E + 11	*1.4E + 15*	*1.1E + 19*
Total trios searched (8)		*1.8E + 04*	4.4E + 06	*1.9E + 10*	*1.5E + 14*
	Computing load				
Total comparisons, events without related victims (1 + 3 + 4 + 5 + 6 + 7)		*9.6E + 05*	2.6E + 08	*9.4E + 10*	*1.1E + 14*
Total comparisons, events with related victims (1 + 2 + 3 + 4 + 5 + 6 + 8)	1.8E + 05	*9.9E + 05*	2.6E + 08	*1.0E + 11*	*1.8E + 14*
	Computing time (h)				
Events without related victims, w/MDKAP		*0.07*	19.7	*7193*	*8.3E + 06*
Events with related victims, w/MDKAP		*0.08*	20.1	*7989*	*1.4E + 07*
Events with related victims, w/Bloodhound (3)		*0.02*	*0.06*	*5.6*	*9.1E + 02*

MDKAP performance numbers were obtained from a conventional desktop personal computer.

Bloodhound performance benchmarks were calculated for the following configuration: a 512 cluster of Dell PowerEdge 1855 servers equipped with dual Intel EM64T 3.6 GHz processors and four gigabytes of memory. Figures appearing in italics result from a simulation exercise.

[1] The Swissair scenario was simulated by running MDKAP with a virtual data set matching the number of genotypes handled during the Swissair MFI.

[2] Scenarios #1 and 2 are simulated data sets scaled up directly from the incident variables and the total number of comparisons obtained with the WTC data set.

[3] Bloodhound was loaded with virtual data sets according to the specifications of the different scenarios and run. The benchmark for Scenario #2 was extrapolated from the other scenarios.

partial genotypes were common, any pair of remains were considered to originate from the same contributor when their STR genotypes shared enough alleles for the random match probability (RMP) to equal or exceed 10^8 (11). Pairs of genotypes with RMP values between 10^4 and 10^8 were still regrouped under the same consensus genotype, but these assignments were flagged as tentative. Regardless of the calculated RMP values, if a given genotype was found to "collapse" under more than one consensus group, cross-references would appear at all matching consensus genotypes (see Fig. 2). Data reviewers were thus alerted to the uncertainty surrounding the source attribution of flagged remains. The most complete genotype (or a composite, see below) of each group was considered the consensus for that set of remains, and this consensus only was used as query against the K database for kinship/direct matches. Remains failing to be assigned to a group were queried individually against the K database.

Composite Q Genotype Generator

Significant gains in matching capability were accrued by allowing composite Q genotypes to be assembled when sufficient overlap between partial genotypes within a consensus group was detected. Composite genotypes were created when the RMP value for the donor and acceptor genotypes was $\geq 10^8$ (see Fig. 3*a*).

Parentage Trio Searching Routine

As no pair of WTC victims shared a P/O relationship, only parentage trio scenarios with two known contributors out of three could be observed within the WTC data set: mother + father + victim, herein referred to as the "descendant" scenario; and offspring + spouse + victim, herein referred to as the "ascendant" scenario. Pair-wise comparisons were used to locate parentage trios within the data set, as shown in Fig. 4. Regardless of whether a suspected F2 contributor was a victim or a NOK, the ability of F1 contributors to account, according to Mendelian inheritance rules, for all alleles encountered in an F2 contributor's genotype constituted the definition of a parentage trio. If a parentage trio was located, the involved NOK entries along with all other members of the same family were prioritized for group display at the top of the results page (see Fig. 1) to facilitate review of the trio data.

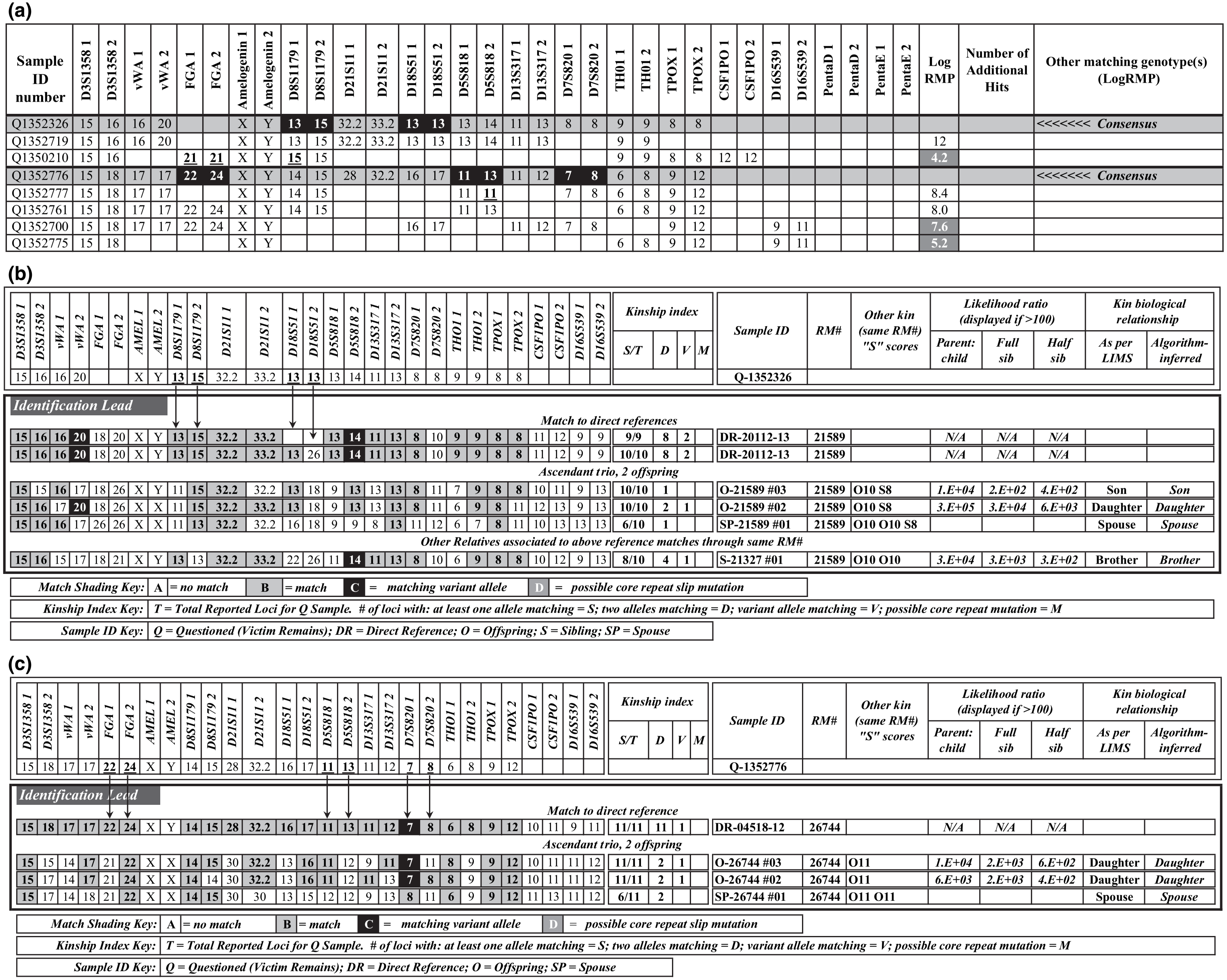
Sample ID number	D3S1358 1	D3S1358 2	vWA 1	vWA 2	FGA 1	FGA 2	Amelogenin 1	Amelogenin 2	D8S1179 1	D8S1179 2	D21S11 1	D21S11 2	D18S51 1	D18S51 2	D5S818 1	D5S818 2	D13S317 1	D13S317 2	D7S820 1	D7S820 2	TH01 1	TH01 2	TPOX 1	TPOX 2	CSF1PO 1	CSF1PO 2	D16S539 1	D16S539 2	PentaD 1	PentaD 2	PentaE 1	PentaE 2	Log RMP	Number of Additional hits	Other matching genotype(s) (LogRMP)
Q235494	16	16	17	18	21	23	X	Y	13	13	29	32.2	12	20	11	12	11	11	10	11	6	9.3	8	11	11	12	10	12	9	14	12	17			<<<<<<< *Consensus*
Q235498	16	16	17	18	21	23	X	Y	13	13	29	32.2	12	20	11	12	11	11	10	11	6	9.3	8	11	11	12	10	12	9	14	12	17	22		
Q235542	16	16	17	18	21	23	X	Y	13	13	29	32.2	12	20	11	12	11	11	10	11	6	9.3	8	11	11	12	10	12	9	14	12	17	22		
Q235506	16	16	17	18	21	23	X	Y	13	13	29	32.2			11	12	11	11	10	11	6	9.3	8	11	11	12	10	12	9	14			17		
Q235467	16	16	17	18	21	23	X	Y	13	13	29	32.2	12	20	11	12	11	11	10	11	6	9.3	8	11	11	12	10	12					17		
23 "LogRMP = 17" entries deleted for the purpose of this figure																																			
Q235541	16	16	17	18	21	23	X	Y	13	13	29	32.2	12	20	11	12	11	11	10	11	6	9.3	8	11	11	12	10	12					17		
Q235545	16	16	17	18	21	23	X	Y	13	13	29	32.2	12	20	11	12	11	11	10	11	6	9.3	8	11	11	12	10	12					17		
Q235470	16	16	17	18	21	23	X	Y	13	13	29	32.2	12	20	11	12	11	11	10	11	6	9.3	8	11			10	12					16		
Q235530	16	16	17	18	21	23	X	Y	13	13	29	32.2	12	20	11	12	11	11	10	11	6	9.3	8	11			10	12					16		
Q235468	16	16	17	18	21	23	X	Y	13	13	29	32.2	12	20	11	12	11	11	10	11	6	9.3	8	11			10	12					16		
Q235518	16	16	17	18	21	23	X	Y	13	13	29	32.2	12	**12**	11	12	11	11	10	11	6	9.3	8	11	11	12	10	12					15		
Q235508	16	16	17	18	21	23	X	Y	13	13	29	32.2	12	20	11	12	11	11			6	9.3			11	12	10	12					15		
Q235480	16	16			21	23	X	Y	13	13	29	32.2	12	20	11	12	11	11	10	11			8	11	11	12	10	12					15		
Q235501	16	16	17	18	21	23	X	Y	13	13	29	32.2			11	12	11	11	10	11	6	9.3	8	11	11	12	10	12					14		
Q235513	16	16	17	18	21	23	X	Y	13	13	29	32.2			11	12	11	11	10	11	6	9.3	8	11	11	12	10	12					14		
Q235471	16	16	17	18	21	23	X	Y	13	13	29	32.2	12	20	11	12	11	11			6	9.3					10	12					14		
Q235522	16	16	17	18	21	23	X	Y	13	13	29	32.2			11	12	11	11			6	9.3	8	11	11	12	10	12					13		
Q235492	16	16	17	18	21	23	X	Y	13	13	29	32.2	12	20	11	12	11	11			6	9.3	8	11									13		
Q235548	16	16	17	18	21	23	X	Y	13	13	29	32.2	12	20	11	12	11	11			6	9.3	8	11									13		
Q235466	16	16			21	23	X	Y	13	13			12	20	11	12	11	11	10	11			8	11	11	12	10	12					13		
Q235453	16	16			21	23	X	Y					12	20			11	11	10	11	6	9.3	8	11	11	12	10	12					12		
Q235465	16	16	17	18	21	23	X	Y	13	13	29	32.2			11	12	11	11			6	9.3	8	11			10	12					12		
Q235462	16	16	17	18	21	23	X	Y	13	13	29	32.2			11	12	11	11			6	9.3	8	11			10	12					12		
Q235510	16	16	17	18	21	23	X	Y	13	13					11	12	11	11	10	11	6	9.3	8	11			10	12					12		
Q235521	16	16					X	Y	13	13	29	32.2			11	12	11	11	10	11	6	9.3	8	11	11	12	10	12					12		
Q235457	16	16					X	Y	13	13			12	20	11	12	11	11	10	11			8	11	11	12	10	12					11		
Q235525	16	16	17	18			X	Y	13	13	29	32.2			11	12	11	11	10	11	6	9.3	8	11	12	12	12	12					11		
Q235491	16	16	17	18	21	**21**	X	Y	13	13	29	32.2			11	12	11	11	10	11	6	9.3	8	11									11		
Q235496	16	16	17	18			X	Y	13	13	29	32.2			11	12	11	11	10	11			8	11			10	12					11		
Q235493	16	16	17	18	21	23	X	Y	13	13	29	32.2			11	12	11	11			6	9.3	8	11									11		
Q235523	16	16	17	18	21	23	X	Y	13	13	29	32.2			11	12	11	11			6	9.3	8	11									11		
Q235634	16	16	17	18	21	23	X	Y	13	13	29	32.2			11	12	11	11			6	9.3	8	11									11		
Q235479	16	16	17	18			X	Y	13	13	29	**29**			11	12	11	11			6	9.3	8	11			10	12					10		
Q235512	16	16			21	23	X	Y	13	13	29	32.2			11	12	11	11	10	10	6	**6**	8	11									9.3		
Q235455	16	16					X	Y	13	13							11	11	10	11	6	9.3	8	11	11	12	10	12					9.1		
Q235532	16	16	17	18			X	Y	13	13	29	32.2			11	12	11	11			6	9.3	8	11									8.9		
Q235488	16	16	17	18	21	23	X	Y	13	13					11	12	11	11			6	9.3	8	11									8.8		
Q235499	16	16					X	Y	13	13	**32.2**	32.2			11	12			10	11	6	9.3	8	11			10	12					8.8		
Q235529	16	16	17	18	21	23	X	Y	13	13	29	32.2			11	12	11	11															8.5		
Q235546					21	23	X	Y			29	32.2					11	11	10	11			8	11			10	12					7.9		
Q235524	16	16	17	18			X	Y	13	13	29	32.2			11	12	11	11					8	11									7.8	2	447(7); 1510(6)
Q235503	16	16	17	18			X	Y	13	13					11	12	11	11			6	9.3	8	11									7.2	4	447(6); 638(6); 1325(6); 1510(6)
Q235535	16	16			21	23	X	Y	13	13					11	12			10	11			8	11									6.7	1	1325(4)
Q235473	16	16					X	Y	13	13	29	32.2			11	12	11	11															5.7	4	447(5); 963(5); 1325(4); 1510(4)
Q235504	16	16	17	18			X	Y	13	13	29	32.2											**11**	11									5.7	1	1510(4)
Q235461	16	16			21	23	X	Y	13	13									10	11													5.0	1	447(4)
Q235540	16	16	17	18			X	Y	13	13					11	12																	4.2	2	447(4); 1510(4)

FIG. 2—*Examples of a Q data collapsing. All remains attributable to a given victim are re-grouped. The consensus genotype appears at the top against light gray background. All matching remains appear below the consensus genotype along with their respective log (RMP). Discrepancies with the consensus are highlighted with black background. Any log (RMP) below eight is highlighted in dark gray. If a genotype is found to match against another consensus, then a flag in the last column indicates the relevant genotype number followed by the associated log (RMP) in brackets.*

Under the "descendant" scenario and assuming reference samples were available for both parents, each parent was expected to appear on the list of potential P/O relationships to the relevant victim. The examination of all trios involving the victim and every possible combination of two NOK entries within the list of potential P/O relationships to the victim was expected to locate the productive parentage trio. Under the alternative "ascendant" scenario and assuming reference samples were available for offspring and spouse (or multiple spouses in the case of re-marriages), only offspring genotypes were expected to appear on the list of potential P/O relationships to the relevant victim. The third member of the trio, the spouse, was located in the NOK data set through her/his P/O relationship(s) to the offspring, not the victim. To this end, a list of potential P/O relationships within the NOK data set was created for each NOK prior to initiating a round of Q versus K queries. To detect an ascendant parentage trio, the software searched for a productive trio among all combinations involving the victim, each potential offspring and any other NOK with a potential P/O relationship to each potential offspring (see Fig. 4). This processing scheme reduced the number of tested trios by >64,000 fold (see Table 1) through, among other means, the elimination of ascendant-type trios where the purported surviving spouse and offspring do not share a potential P/O relationship.

Consistency Check Feature

The establishment of potential P/O relationships underpins much of the kinship matching capabilities of this approach, the parentage trio-searching algorithm in particular. However, as demonstrated with the Swissair data, one out of every 2000 pair-wise comparisons between 13-loci STR genotypes is expected to generate a kinship score consistent with a P/O relationship where no such relationship actually exists (9), a situation we refer to herein as fortuitous kinship associations (FKA). With an average of 26 million Q versus NOK/NOK versus NOK pair-wise genotype comparisons to perform per identification round, it was anticipated that some 13,000 FKAs would be observed, an average of five per queried Q genotype. In order to avoid generating incorrect inferences of biological relationship, a tool to distinguish FKAs from genuine kinship associations was developed. In practice, most FKAs can be distinguished from the genuine kinship association when multiple first-degree relatives (e.g., parents, offspring) are available as references in the family of the tested NOK. For instance, if a given NOK is found to show a potential P/O relationship to a Q genotype, then a potential P/O relationship should also be observed between the Q genotype and other available siblings (if the victim is an anticipated parent) or mother/father (if the victim is an anticipated offspring) of the NOK,



(a)

Sample ID number	D3S1358 1	D3S1358 2	vWA 1	vWA 2	FGA 1	FGA 2	Amelogenin 1	Amelogenin 2	D8S1179 1	D8S1179 2	D21S11 1	D21S11 2	D18S51 1	D18S51 2	D5S818 1	D5S818 2	D13S317 1	D13S317 2	D7S820 1	D7S820 2	TH01 1	TH01 2	TPOX 1	TPOX 2	CSF1PO 1	CSF1PO 2	D16S539 1	D16S539 2	PentaD 1	PentaD 2	PentaE 1	PentaE 2	Log RMP	Number of Additional Hits	Other matching genotype(s) (LogRMP)
Q1352326	15	16	16	20			X	Y	13	15	32.2	33.2	13	13	13	14	11	13	8	8	9	9	8	8											<<<<<<<< Consensus
Q1352719	15	16	16	20			X	Y	13	15	32.2	33.2	13	13	13	14	11	13			9	9											12		
Q1350210	15	16			21	21	X	Y	15	15											9	9	8	8	12	12							4.2		
Q1352776	15	18	17	17	22	24	X	Y	14	15	28	32.2	16	17	11	13	11	12	7	8	6	8	9	12											<<<<<<<< Consensus
Q1352777	15	18	17	17			X	Y	14	15					11	11			7	8	6	8	9	12									8.4		
Q1352761	15	18	17	17	22	24	X	Y	14	15					11	13					6	8	9	12									8.0		
Q1352700	15	18	17	17	22	24	X	Y					16	17			11	12	7	8			9	12			9	11					7.6		
Q1352775	15	18					X	Y													6	8	9	12			9	11					5.2		

(b)

D3S1358 1	D3S1358 2	vWA 1	vWA 2	FGA 1	FGA 2	AMEL 1	AMEL 2	D8S1179 1	D8S1179 2	D21S11 1	D21S11 2	D18S51 1	D18S51 2	D5S818 1	D5S818 2	D13S317 1	D13S317 2	D7S820 1	D7S820 2	THO1 1	THO1 2	TPOX 1	TPOX 2	CSF1PO 1	CSF1PO 2	D16S539 1	D16S539 2	Kinship index				Sample ID	RM#	Other kin (same RM#) "S" scores	Likelihood ratio (displayed if >100)			Kin biological relationship	
																												S/T	D	V	M				Parent: child	Full sib	Half sib	As per LIMS	Algorithm-inferred
15	16	16	20			X	Y	13	15	32.2	33.2	13	13	13	14	11	13	8	8	9	9	8	8									Q-1352326							
Identification Lead																																							
Match to direct references																																							
15	16	16	20	18	20	X	Y	13	15	32.2	33.2			13	14	11	13	8	10	9	9	8	8	11	12	9	9	9/9	8	2		DR-20112-13	21589		N/A	N/A	N/A		
15	16	16	20	18	20	X	Y	13	15	32.2	33.2	13	26	13	14	11	13	8	10	9	9	8	8	11	12	9	9	10/10	8	2		DR-20112-13	21589		N/A	N/A	N/A		
Ascendant trio, 2 offspring																																							
15	15	16	17	18	26	X	Y	11	15	32.2	32.2	13	18	9	13	13	13	8	11	7	9	8	8	10	11	9	13	10/10	1			O-21589 #03	21589	O10 S8	1.E+04	2.E+02	4.E+02	Son	Son
15	16	17	20	18	26	X	X	11	15	32.2	33.2	13	18	9	13	13	13	8	11	6	9	8	11	12	13	9	13	10/10	2	1		O-21589 #02	21589	O10 S8	3.E+05	3.E+04	6.E+03	Daughter	Daughter
15	16	16	17	26	26	X	X	11	13	32.2	32.2	16	18	9	9	8	13	11	12	6	7	8	11	10	13	13	13	6/10	1			SP-21589 #01	21589	O10 O10 S8				Spouse	Spouse
Other Relatives associated to above reference matches through same RM#																																							
15	16	15	17	18	21	X	Y	13	13	32.2	33.2	22	26	11	14	11	13	8	10	6	9	8	8	10	12	9	13	8/10	4	1		S-21327 #01	21589	O10 O10	3.E+04	3.E+03	3.E+02	Brother	Brother

Match Shading Key: A = no match B = match C = matching variant allele D = possible core repeat slip mutation

Kinship Index Key: T = Total Reported Loci for Q Sample. # of loci with: at least one allele matching = S; two alleles matching = D; variant allele matching = V; possible core repeat mutation = M

Sample ID Key: Q = Questioned (Victim Remains); DR = Direct Reference; O = Offspring; S = Sibling; SP = Spouse

(c)

D3S1358 1	D3S1358 2	vWA 1	vWA 2	FGA 1	FGA 2	AMEL 1	AMEL 2	D8S1179 1	D8S1179 2	D21S11 1	D21S11 2	D18S51 1	D18S51 2	D5S818 1	D5S818 2	D13S317 1	D13S317 2	D7S820 1	D7S820 2	THO1 1	THO1 2	TPOX 1	TPOX 2	CSF1PO 1	CSF1PO 2	D16S539 1	D16S539 2	Kinship index				Sample ID	RM#	Other kin (same RM#) "S" scores	Likelihood ratio (displayed if >100)			Kin biological relationship	
																												S/T	D	V	M				Parent: child	Full sib	Half sib	As per LIMS	Algorithm-inferred
15	18	17	17	22	24	X	Y	14	15	28	32.2	16	17	11	13	11	12	7	8	6	8	9	12									Q-1352776							
Identification Lead																																							
Match to direct reference																																							
15	18	17	17	22	24	X	Y	14	15	28	32.2	16	17	11	13	11	12	7	8	6	8	9	12	10	11	9	11	11/11	11	1		DR-04518-12	26744		N/A	N/A	N/A		
Ascendant trio, 2 offspring																																							
15	15	14	17	21	22	X	X	14	15	30	32.2	13	16	11	12	9	11	7	11	8	9	9	12	10	11	11	12	11/11	2	1		O-26744 #03	26744	O11	1.E+04	2.E+03	6.E+02	Daughter	Daughter
15	17	14	17	21	24	X	X	14	14	30	32.2	13	16	11	12	11	13	7	8	8	9	9	12	10	11	11	12	11/11	2	1		O-26744 #02	26744	O11	6.E+03	2.E+03	4.E+02	Daughter	Daughter
15	17	14	18	21	22	X	X	14	15	30	30	13	15	12	12	9	13	8	11	6	9	9	12	11	13	11	12	6/11	2			SP-26744 #01	26744	O11 O11				Spouse	Spouse

Match Shading Key: A = no match B = match C = matching variant allele D = possible core repeat slip mutation

Kinship Index Key: T = Total Reported Loci for Q Sample. # of loci with: at least one allele matching = S; two alleles matching = D; variant allele matching = V; possible core repeat mutation = M

Sample ID Key: Q = Questioned (Victim Remains); DR = Direct Reference; O = Offspring; SP = Spouse

FIG. 3—*Composite genotype generator. The top panel displays two consecutive consensus groups. See Fig. 2 for shading scheme. The middle and bottom panel display the score report for the consensus groups shown in Panel A. Underlined alleles in the query genotype are "filled-in" alleles. In the middle panel, D8S1179 filled-in alleles are confirmed in a DR genotype. DR data for D18S51 and D7S820 data suggest allele drop-outs in the queried genotype. In the bottom panel, FGA, D5S818 and D7S820 filled-in alleles are confirmed in the DR genotype.*

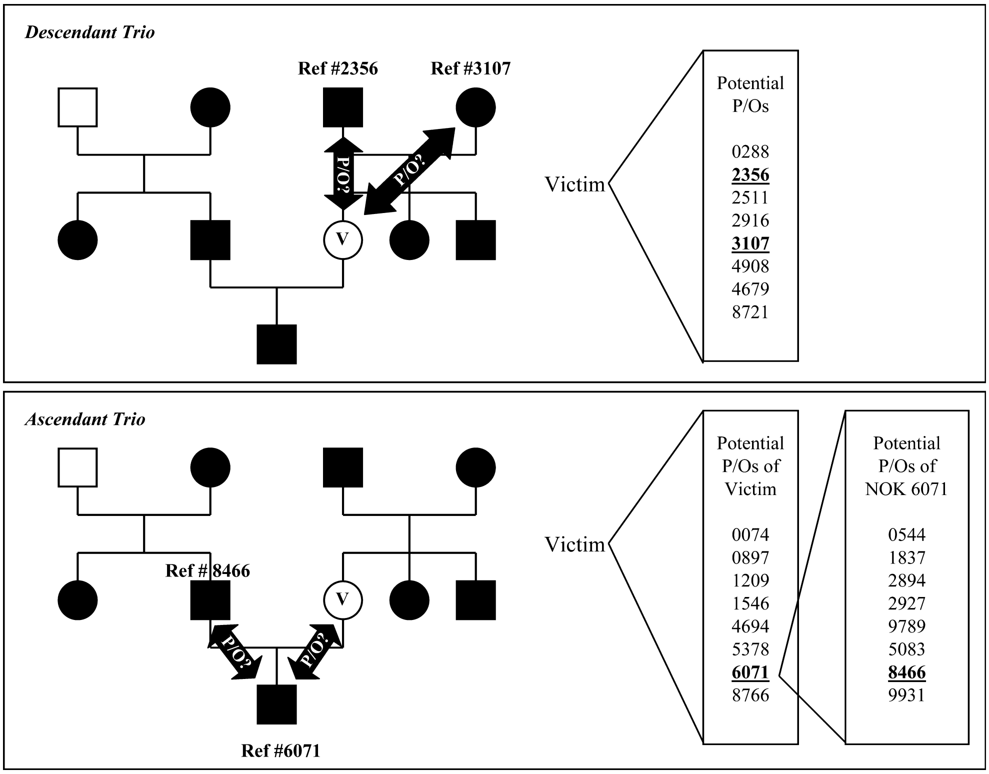


FIG. 4—*Locating a parentage trio with pair-wise comparisons. The top panel displays a "descendant" parentage trio. Both parents can be found on a single listing of potential P/O relationships to the victim. The bottom panel displays an "ascendant" parentage trio. Only the offspring can be found on a listing of potential P/O relationships to the victim. The spouse will be found on a separate listing of potential P/O relationships to the offspring. Both descendant and ascendant scenarios are explored for every trio.*

if the Q genotype is truly genetically linked to the tested family. To provide this consistency check, the S score for every other family member [i.e., same Reported Missing (RM) no.] of a high S score NOK is displayed within a concatenated string (see Fig. 1) on every result page. For the 45 top scoring NOKs of every Q versus K query, the consistency check verified that the self-reported relationships for the NOKs and their relatives were consistent with the calculated S scores. This feature automatically retrieved S scores for any relative regardless of their ranking position in the 10,000 + genotype-ranking stack in the K database. The capability to detect FKAs was essential in the ability of the software to detect false parentage trios (FPT).

Likelihood Ratio Calculations

The performance of the kinship index scheme used in this software was evaluated by comparison to calculated LRs (see Fig. 1). For every Q query, once the software had ranked the K database according to the kinship index score, the KinTest program (G.C.) calculated LRs for parent:child, full sibling and half-sibling using frequencies from a U.S. Caucasian database [see reference (9) for a discussion on the choice of reference databases] for the 45 best scoring Ks (i.e., the number of entries displayed on the result page).

Automated Processing

Sub-processes were linked and executed in the following order (Fig. 5): (1) conversion of CMF file format to tabular format for incoming genotypic data, (2) import of STR data, (3) removal of duplicate entries, (4) removal of genotypes with less than 12 alleles, (5) collapse of victim's genotypes, (6) creation of a consensus victims' genotype listing, (7) NOK database prescreen for potential P/O relationships within the NOK data set, (8) query of each consensus victim's genotype against the K database, (9) recording of the results for later retrieval, (10) establishment of a priority list to allow data reviewers to locate quickly the most promising identification leads. To avoid repeated reviewing of data associated with completed identifications from previous identification rounds, an electronic list of remains considered by the OCME as identified was provided with each new data set, and MDKAP flagged relevant consensus genotypes accordingly in the priority list. Entries related to an identified victim were not removed from the data set to allow for any eventual inconsistency to be detected if newly added data challenged a prior match. The entire linked process ran unattended overnight on incremental data sets.

Results

Confirming the Suitability of the Search Algorithm

The program used to process genotypic data during the Swissair MFI victim identification initiative was tested in a simulation environment to assess whether the approach could maintain the sensitivity required to make all Swissair identifications despite a large increase in the size of the genotypic data set. To that end, the Swissair data was supplemented with 12,000 random 13-STR loci genotypes virtually generated to reflect Caucasian allele frequencies. All Swissair identifications could still be carried out against this much larger background (data not shown). Parentage trios involving descendants were easily detected without the routine developed for MDKAP. Parentage trios involving ascendants (i.e., offspring + spouse + victim) could be detected and manually verified

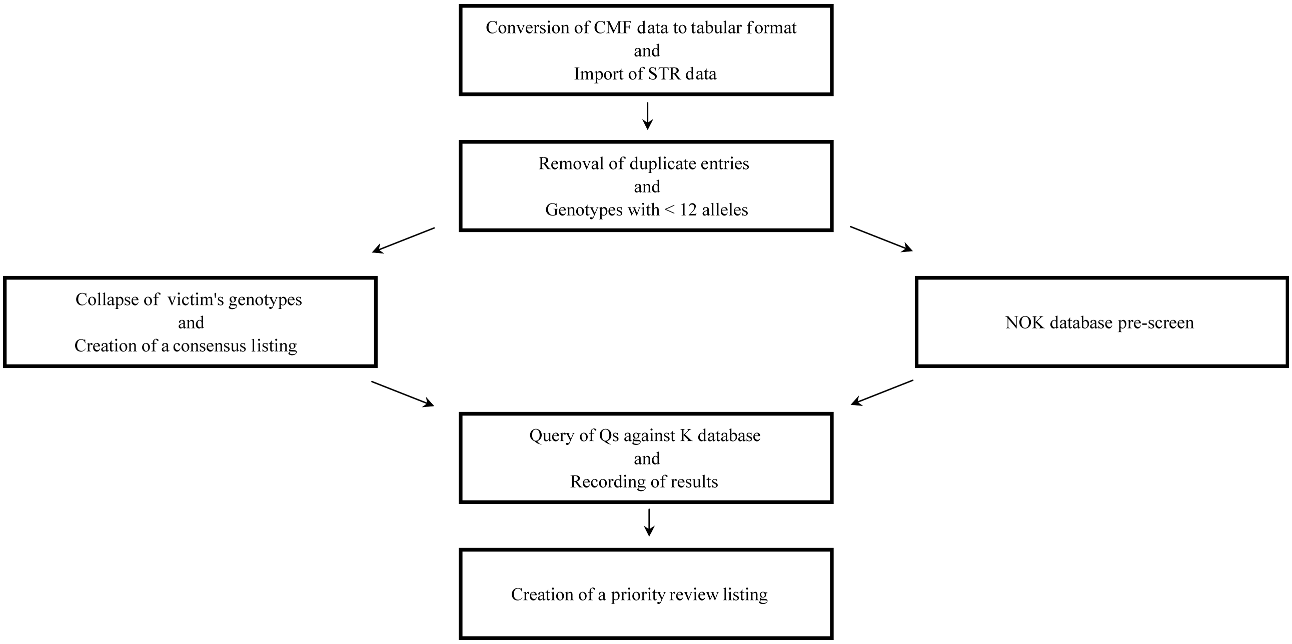

Conversion of CMF data to tabular format
and
Import of STR data
Removal of duplicate entries
and
Genotypes with < 12 alleles
Collapse of victim's genotypes
and
Creation of a consensus listing
NOK database pre-screen
Query of Qs against K database
and
Recording of results
Creation of a priority review listing

FIG. 5—*Flowchart of data processing scheme.*

only through the use of NOK sample accessory information, and were dependent on the accuracy of this information. It was apparent that the much larger WTC data set would require accelerated algorithms to maintain processing times within operational requirements as well as many supplemental routines to take full advantage of the information content of the genotypic data set.

Genotypic Data Collapse

The body fragmentation sustained by the victims of WTC produced an average of 10 Q samples per victim, some collapsed data sets regrouping over 200 remains. Collapsing the data set to a smaller number of consensus genotypes achieved two goals: regrouping of remains for return to the families and containment of processing time. Figure 2 displays an example of collapsed data consisting of 70 remains originating from a single victim. As a first step, the algorithm ranks genotypes according to the number of reported alleles. The most complete genotype is then used to identify within the data set other genotypes that share enough alleles with the query for the RMP to exceed the set threshold. Allele drop-out was observed even with almost complete genotypes. Genotypes that were assigned to a consensus group were still tested against all other consensus groups to confirm the assignment. As shown in Fig. 2, partial genotypes were often found to match under more than one consensus group when the RMP value fell under the 10^8-threshold set by the OCME. The ability to collapse the Q data set yielded a five to sevenfold reduction in the number of queries to the K database.

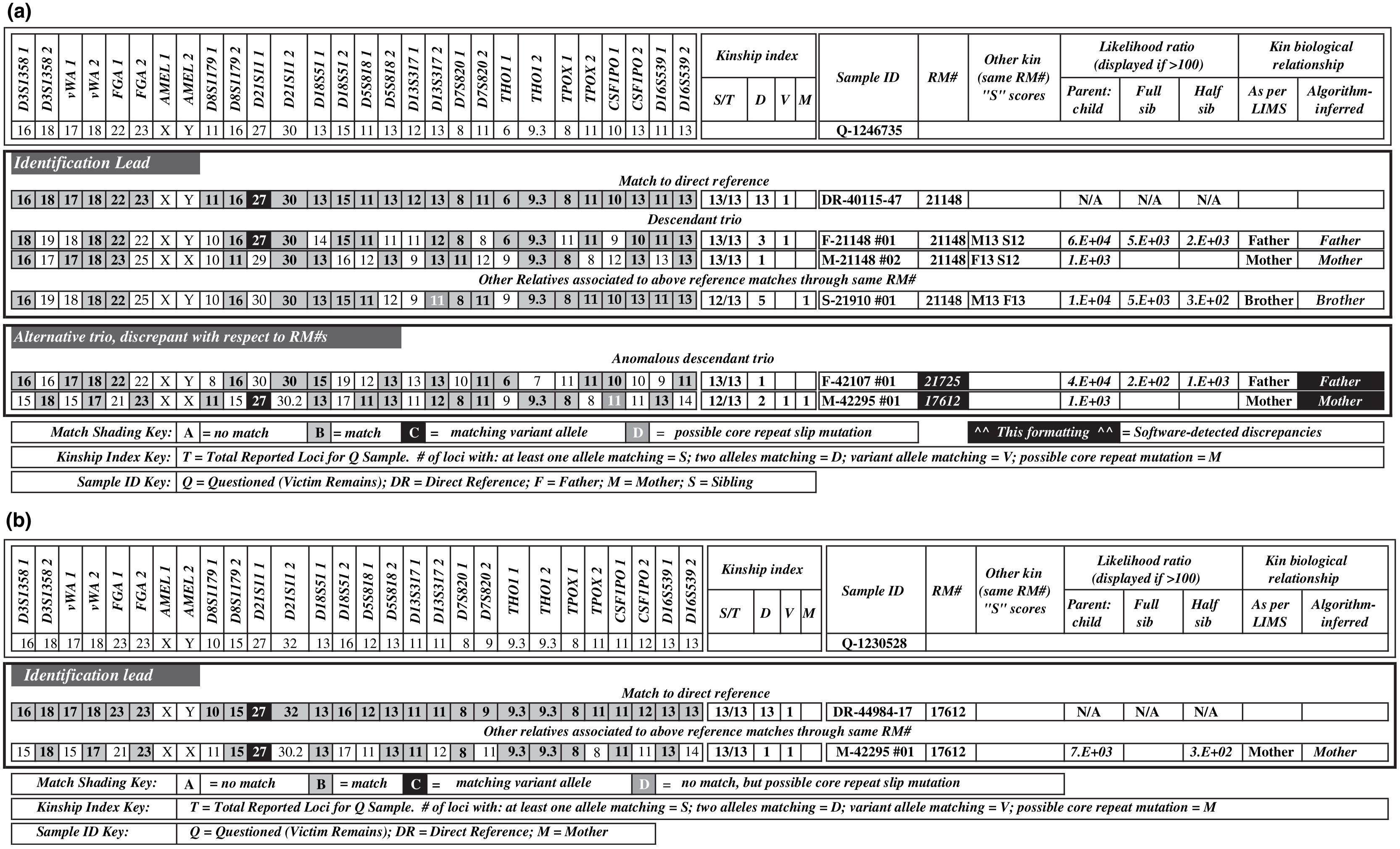
Virtual Genotype Generator

Figure 3*a* provides two examples of composite genotypes. Allelic designations displayed against a black-background in the consensus genotype (gray background) were missing in the original remains genotype and were filled in as the RMP value between contributing and receiving partial genotypes exceeded the 10^8 threshold. This rule prevented the FGA 21,21 alleles of Q1350210 from being integrated into the composite genotype in the first example. In this particular case, the victim was identified with both PEs and NOKs, which confirmed that the victim was FGA 18, 20, not 21,21 (see Fig. 3*b*). The Q1350210 genotype is likely derived from remains of another victim (this genotype was not found to match with any other consensus genotype, as documented by the absence of an entry under the "Other matching genotypes…" header), or the FGA 21,21 is a single drop-in allele, a distinct possibility considering that some remains are known to have co-mingled (12). The requirement to meet the 10^8 threshold also prevented the CSF1PO 12,12 from being integrated into the composite genotype, a threshold enforcement that proved to be correct as a later comparison to a PE genotype documented an allele drop-out at that locus. In this example, the "filled-in" alleles were confirmed in the PE genotypes.

Both examples demonstrate how allele drop-outs are detected and handled within a rules-based automated system. The D8S1179 and D5S818 loci of the first and second examples respectively suggest that allele drop-out has occurred in one of the contributing genotypes (i.e., the homozygous allele is one of the alleles of the heterozygous genotype). In these cases, the virtual genotype incorporates the encountered heterozygous genotype as allele drop-outs are more common than allele drop-ins. In both cases, perfect matches to PEs confirmed the heterozygous genotype at the drop-out locus of the original most complete genotype (see Fig. 3*b,c*). This is the only condition that allowed for data gap filling to proceed when a discrepancy was encountered at a locus.

Parentage Trio Searching Routine

Figure 3*b,c* showcases ascendant parentage trios. Under that scenario, the offspring is expected to display a score consistent with a potential P/O relationship to the victim, and the spouse is expected to display a score consistent with a potential P/O relationship to the offspring. For each K genotype, the algorithm-inferred kin biological relationship column on the reports displays the relationship consistent with the detected parentage trio. Discrepancies



(a)

D3S1358 1	D3S1358 2	vWA 1	vWA 2	FGA 1	FGA 2	AMEL 1	AMEL 2	D8S1179 1	D8S1179 2	D21S11 1	D21S11 2	D18S51 1	D18S51 2	D5S818 1	D5S818 2	D13S317 1	D13S317 2	D7S820 1	D7S820 2	THO1 1	THO1 2	TPOX 1	TPOX 2	CSF1PO 1	CSF1PO 2	D16S539 1	D16S539 2	Kinship index				Sample ID	RM#	Other kin (same RM#) "S" scores	Likelihood ratio (displayed if >100)			Kin biological relationship	
																												S/T	D	V	M				Parent: child	Full sib	Half sib	As per LIMS	Algorithm-inferred
16	18	17	18	22	23	X	Y	11	16	27	30	13	15	11	13	12	13	8	11	6	9.3	8	11	10	13	11	13					Q-1246735							
Identification Lead																																							
Match to direct reference																																							
16	18	17	18	22	23	X	Y	11	16	27	30	13	15	11	13	12	13	8	11	6	9.3	8	11	10	13	11	13	13/13	13	1		DR-40115-47	21148		N/A	N/A	N/A		
Descendant trio																																							
18	19	18	18	22	22	X	Y	10	16	27	30	14	15	11	11	11	12	8	8	6	9.3	11	11	9	10	11	13	13/13	3	1		F-21148 #01	21148	M13 S12	6.E+04	5.E+03	2.E+03	Father	Father
16	17	17	18	23	25	X	X	10	11	29	30	13	16	12	13	9	13	11	12	9	9.3	8	8	12	13	13	13	13/13	1			M-21148 #02	21148	F13 S12	1.E+03			Mother	Mother
Other Relatives associated to above reference matches through same RM#																																							
16	19	18	18	22	25	X	Y	10	16	30	30	13	15	11	12	9	11	8	11	9	9.3	8	11	10	13	11	13	12/13	5		1	S-21910 #01	21148	M13 F13	1.E+04	5.E+03	3.E+02	Brother	Brother
Alternative trio, discrepant with respect to RM#s																																							
Anomalous descendant trio																																							
16	16	17	18	22	22	X	Y	8	16	30	30	15	19	12	13	13	13	10	11	6	7	11	11	10	10	9	11	13/13	1			F-42107 #01	21725		4.E+04	2.E+02	1.E+03	Father	Father
15	18	15	17	21	23	X	X	11	15	27	30.2	13	17	11	13	11	12	8	11	9	9.3	8	8	11	11	13	14	12/13	2	1	1	M-42295 #01	17612		1.E+03			Mother	Mother

Match Shading Key: A = no match B = match C = matching variant allele D = possible core repeat slip mutation ^^ This formatting ^^ = Software-detected discrepancies

Kinship Index Key: T = Total Reported Loci for Q Sample. # of loci with: at least one allele matching = S; two alleles matching = D; variant allele matching = V; possible core repeat mutation = M

Sample ID Key: Q = Questioned (Victim Remains); DR = Direct Reference; F = Father; M = Mother; S = Sibling

(b)

D3S1358 1	D3S1358 2	vWA 1	vWA 2	FGA 1	FGA 2	AMEL 1	AMEL 2	D8S1179 1	D8S1179 2	D21S11 1	D21S11 2	D18S51 1	D18S51 2	D5S818 1	D5S818 2	D13S317 1	D13S317 2	D7S820 1	D7S820 2	THO1 1	THO1 2	TPOX 1	TPOX 2	CSF1PO 1	CSF1PO 2	D16S539 1	D16S539 2	Kinship index				Sample ID	RM#	Other kin (same RM#) "S" scores	Likelihood ratio (displayed if >100)			Kin biological relationship	
																												S/T	D	V	M				Parent: child	Full sib	Half sib	As per LIMS	Algorithm-inferred
16	18	17	18	23	23	X	Y	10	15	27	32	13	16	12	13	11	11	8	9	9.3	9.3	8	11	11	12	13	13					Q-1230528							
Identification lead																																							
Match to direct reference																																							
16	18	17	18	23	23	X	Y	10	15	27	32	13	16	12	13	11	11	8	9	9.3	9.3	8	11	11	12	13	13	13/13	13	1		DR-44984-17	17612		N/A	N/A	N/A		
Other relatives associated to above reference matches through same RM#																																							
15	18	15	17	21	23	X	X	11	15	27	30.2	13	17	11	13	11	12	8	11	9.3	9.3	8	8	11	11	13	14	13/13	1	1		M-42295 #01	17612		7.E+03		3.E+02	Mother	Mother

Match Shading Key: A = no match B = match C = matching variant allele D = no match, but possible core repeat slip mutation

Kinship Index Key: T = Total Reported Loci for Q Sample. # of loci with: at least one allele matching = S; two alleles matching = D; variant allele matching = V; possible core repeat mutation = M

Sample ID Key: Q = Questioned (Victim Remains); DR = Direct Reference; M = Mother

FIG. 6—*True and fortuitous parentage trios (FPT). The top panel displays a score report where a pair of fortuitous parents (dark gray background) was detected. Two Direct References and the true mother and father of the victim document the identity of the victim. The middle panel confirms that the fortuitous mother shown in the top panel is linked to another victim. The true offspring of the fortuitous father was never found.*

between the relationship entered in the OCME tracking system and the one inferred by the parentage trio detection algorithm would be highlighted with a background shade.

Figure 6*a* demonstrates the effectiveness of the parentage trio-searching algorithm. For this particular victim, a PE genotype confirmed a detected parentage trio involving two NOKs, the PE and NOK genotypes sharing the same family number. The algorithm that locates parentage trios makes use of a nontriangulated approach (e.g., a method by which a list of obligate alleles derived from the genotypes of the two surviving members of a parentage trio is not used to locate the third member of the trio within a large genotype data set) where all possible trios are tested. Although computationally intensive, this strategy allows for the detection of all trios including those where NOK self-reported biological relationships to the victim appear incorrect according to the OCME tracking system, situations that would otherwise escape detection with a triangulated approach. This strategy will cause for Fortuitous Parentage Trios (FPT) to be detected, apparent parentage trios where no such biological relationship actually exists.

There were three types of detected FPTs. The vast majority of detected FPTs in the WTC data set involved related individuals. As fewer alleles are encountered within family pedigrees, it is possible for offspring to substitute for a parent to one of their siblings and still produce parentage trios where all alleles are accounted for in the individual being considered an offspring. A typical example is shown in Fig. 1 under the "Alternative trios, discrepant with respect to reported biological relationships" heading. This type of event is common, especially when numerous offspring are available for analysis, regardless of the scope of the incident. Except for one situation, the remainder of detected potential FTPs for the WTC dataset proved to be real trios but with incorrectly reported family number or biological relationships (see Fig. 8): an administrative review of these specific case files resolved the detected discrepancies. Finally, FPTs involving nonrelated individuals are expected to be very rare, and only one was detected in the 1.9×10^{11} possible trios in the WTC data set (Fig. 6*a*). In this case, while the true parentage trio was detected and confirmed as described above, a FPT involving the same victim and a mother/father pair unrelated to each other was detected and flagged. This FPT was dismissed as the putative mother was eventually linked to her true offspring through a direct match to a PE (Fig. 6*b*), but no match to other Q data was ever made for the putative father.

The identification shown in Fig. 6*a* demonstrates the ease with which the true parentage trio could be distinguished from a fortuitous one with the use of reported family numbers and biological relationships. Although the WTC victim identification initiative was dealt many complicating factors, its kinship analysis situation was ideal in one respect: none of the WTC victims shared a P/O relationship, no parentage trio ever included more than one victim, no assembly of family pedigrees had to be executed from within the victim's genotypic data set. These conditions facilitated the identification of all potential FPTs. The WTC situation was in sharp contrast with most MFIs, especially transportation accidents, where the list of victims often includes many partial or entire families, making it more difficult in those circumstances to assess the validity of detected trios that include more than one victim. For example, if the five individuals depicted in Fig. 6*a* were set in the context of another MFI of the same scale as WTC's, where: (1) all five individuals were among the victims; (2) PEs were unavailable for the offspring (e.g., all useful PEs destroyed in the incident); (3) the remains of one or both of the actual parents were never recovered; under these conditions, the identification process could have easily been confounded. A similar scenario applied to the situation depicted in Fig. 1 could have led to identity switches among family members within an otherwise correctly identified family. The successful use of kinship analysis is therefore dependent on the circumstances of the MFI and performs best under conditions where some PEs or clear identification information from another modality on specific Q samples are available to clarify ambiguous pedigrees.

Consistency Check

The consistency check provides useful confirmation of valid parentage trios but its best use is in situations where parentage trios are not available and pair-wise comparisons are the only path to identification. Figure 7 provides an interesting example of the usefulness of this tool and underscores the caution required in the interpretation of results from pair-wise comparisons in the context of large data sets containing many incomplete genotypes, many incomplete pedigrees. Early into the identification initiative, the Q genotype in Fig. 7 displayed a tentative P/O relationship with the mothers of families nos. 33327 and 37533 (fathers were not available for sampling). Parent:child LRs were 1×10^4 and 2×10^5 respectively, favoring the second family. In addition, three NOKs from the second family, reporting a sibling relationship to their missing relative, shared between 13 and 17 alleles out of a possible 26 and significant LR values with this victim, again favoring the second family as a kinship association. Those odds later proved misleading as a perfect match between the Q genotype and that of a PE submitted for the victim of the first family was detected. Much later in the identification initiative, samples from offspring of victim no. 37533 (second family) were finally submitted and proved not to meet the P/O expected S score for them to be related to the queried Q genotype. The consistency check allowed for the discrepancies in S scores for the offspring of victim no. 37533 to be brought to the attention of the data reviewer. This particular example demonstrates the benefits in securing samples from all P/O relationships to the victims as early as possible in the identification process.

Figure 8 features an example of an incorrectly reported NOK biological relationship. Two separate male individuals reported a father relationship to this victim and both generated a parentage trio with the reported mother of the victim. The second individual displayed a core repeat slip mutation and high sharing of alleles with the victim, making him less likely to be the true father. An administrative review of the case proved the assumption correct as the first individual was the actual father, an outcome consistent with the calculated LRs.

Likelihood Ratios

On score reports, LRs were calculated to allow for an assessment of the performance of the kinship index as means to rank NOKs for genetic relatedness to a queried genotype. Both kinship index and LRs were run in parallel for the duration of the identification initiative. Figure 1 provides a typical example, demonstrating that discrete changes in ranking order would be observed if LRs were used for sorting but without affecting the conclusion.

Computational Workload

Table 1 summarizes computing statistics of MDKAP with WTC data for smaller and larger events under two scenarios: (1) the victims are unrelated; (2) the victims' data set includes families. A very small increase in workload (2–3%) is associated with the presence of related individuals among the victims. However, the workload increases exponentially with the scale of the event. The

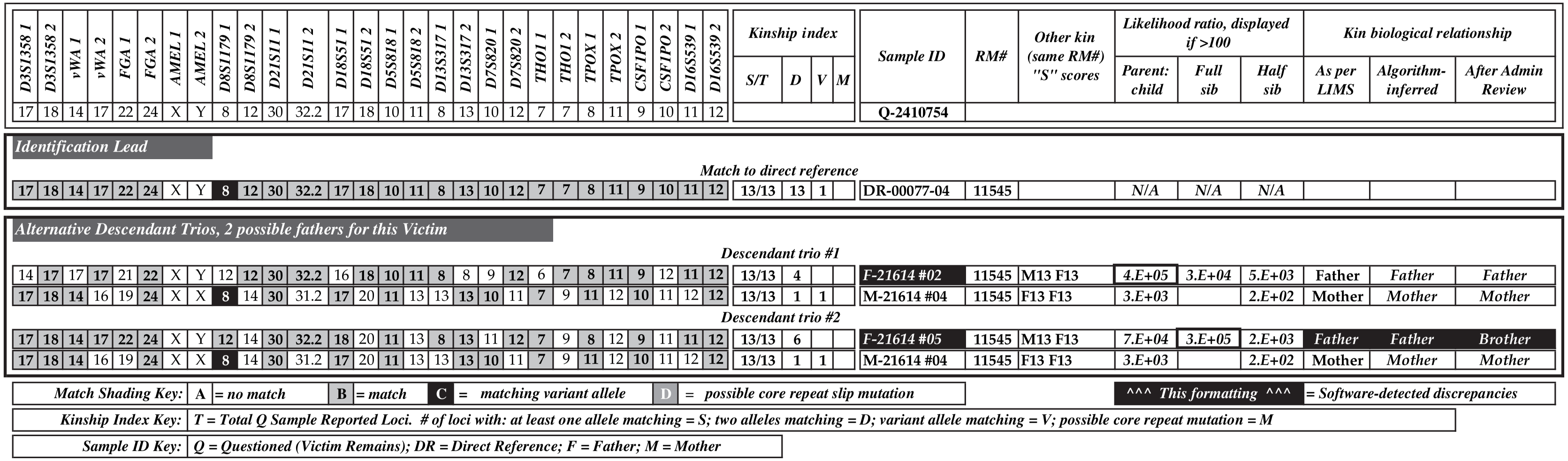
D3S1358 1	D3S1358 2	vWA 1	vWA 2	FGA 1	FGA 2	AMEL 1	AMEL 2	D8S1179 1	D8S1179 2	D21S11 1	D21S11 2	D18S51 1	D18S51 2	D5S818 1	D5S818 2	D13S317 1	D13S317 2	D7S820 1	D7S820 2	THO1 1	THO1 2	TPOX 1	TPOX 2	CSF1PO 1	CSF1PO 2	D16S539 1	D16S539 2	PentaD 1	PentaD 2	PentaE 1	PentaE 2	Kinship index				Sample ID	RM#	Other kin (same RM#) "S" scores	Likelihood ratio (displayed if >100)			Kin biological relationship	
																																S/T	D	V	M				Parent: child	Full sib	Half sib	As per LIMS	Algorithm-inferred
16	17	14	18	21	24	X	Y	11	13	32.2	32.2	12	18	11	12	11	12	9	10	6	8	11	11	10	11	11	11	9	9	12	15					Q-2361733							
On early data sets																																											
K database entries, best leads																																											
Family #1																																											
16	**17**	18	**18**	19	**24**	X	X	**13**	15	29.2	**32.2**	15	**18**	12	**12**	**11**	**12**	10	**10**	**6**	6	8	**11**	**10**	10	**11**	**11**					13/13	3			M-61443 #02	33327		1.E+04	4.E+02	3.E+02	Mother	
Family #2																																											
15	**16**	16	**18**	**24**	25	X	X	**11**	**13**	30	**32.2**	**12**	**18**	**11**	11	**11**	**12**	**9**	**10**	**6**	7	10	**11**	11	**11**	**11**	13					13/13	4			S-51533 #03	37533	M12 S12 S13	8.E+03	2.E+03	3.E+02	Sister	
16	18	**14**	17	**21**	25	X	Y	9	**11**	31.2	**32.2**	**12**	13	**11**	11	**11**	11	8	**10**	**6**	6	10	**11**	9	**11**	**11**	13					13/13				S-51533 #02	37533	M12 S12 S13	2.E+03			Brother	
15	18	**14**	**18**	**21**	**24**	X	X	**11**	11	**32.2**	**32.2**	**12**	12	**11**	11	**11**	**12**	8	**9**	**6**	7	**11**	**11**	11	**11**	**11**	12					12/13	5		1	M-51533 #04	37533	S13 S12 S13	2.E+05	4.E+04	1.E+03	Mother	
14	**18**	16	**18**	**21**	**24**	X	X	9	**11**	30	**32.2**	**12**	13	10	**11**	**11**	**12**	**9**	**10**	**6**	7	10	**11**	9	**11**	9	**11**					12/13	3		1	S-51533 #01	37533	M12 S12 S13	1.E+03			Sister	
On late data sets																																											
Identification Lead																																											
Match to direct reference																																											
16	**17**	**14**	**18**	**21**	**24**	X	Y	**11**	**13**	**32.2**	**32.2**			**11**	**12**	**11**	**12**	**9**	**10**	**6**	**8**	**11**	**11**	**10**	**11**	**11**	**11**	**9**	**9**	**12**	**15**	14/14	14			DR-33327-01	33327		N/A	N/A	N/A		
Other Relatives associated to above reference matches through same RM# (Family #1)																																											
16	**17**	18	**18**	19	**24**	X	X	**13**	15	29.2	**32.2**	15	**18**	12	**12**	**11**	**12**	10	**10**	**6**	6	8	**11**	**10**	10	**11**	**11**					13/13	3			M-61443 #02	33327	H10 H12	1.E+04	4.E+02	3.E+02	Mother	*Mother*
16	18	15	**18**	19	26	X	X	8	**13**	29	29.2	15	**18**	**12**	13	8	**11**	**10**	11	**6**	9.3	8	8	**10**	10	**11**	13	**9**	12	11	**15**	12/15				H-33327 #04	33327	M13 H10				Half-Sister	
16	18	15	**18**	**24**	26	X	Y	8	15	29	29.2	14	15	**12**	13	**11**	**12**	**10**	11	**6**	9.3	8	8	**10**	13	**11**	**11**	12	13	11	**15**	10/15	2			H-33327 #03	33327	M13 H12				Half-Brother	
Family #2 eliminated, low "S" scores for 3 offspring																																											
15	**16**	16	**18**	**24**	25	X	X	**11**	**13**	30	**32.2**	**12**	**18**	**11**	11	**11**	**12**	**9**	**10**	**6**	7	10	**11**	11	**11**	**11**	13					13/13	4			S-51533 #03	37533	M12 O9 O9 O11 S12 S13	8.E+03	2.E+03	3.E+02	Sister	
16	18	**14**	17	**21**	25	X	Y	9	**11**	31.2	**32.2**	**12**	13	**11**	11	**11**	11	8	**10**	**6**	6	10	**11**	9	**11**	**11**	13					13/13				S-51533 #02	37533	M12 O9 O9 O11 S12 S13	2.E+03			Brother	
15	18	**14**	**18**	**21**	**24**	X	X	**11**	11	**32.2**	**32.2**	**12**	12	**11**	11	**11**	**12**	8	**9**	**6**	7	**11**	**11**	11	**11**	**11**	12					12/13	5		1	M-51533 #04	37533	O9 O9 O11 S13 S12 S13	2.E+05	4.E+04	1.E+03	*Mother*	*No P/O*
14	**18**	16	**18**	**21**	**24**	X	X	9	**11**	30	**32.2**	**12**	13	10	**11**	**11**	**12**	**9**	**10**	**6**	7	10	**11**	9	**11**	9	**11**					12/13	3		1	S-51533 #01	37533	M12 O9 O9 O11 S12 S13	1.E+03			Sister	

Match Shading Key: A *= no match* B *= match* C *= matching variant allele* D *= no match, but possible core repeat slip mutation* ***^^^ This formatting ^^^*** *= Software-detected discrepancies*

Kinship Index Key: *T = Total Reported Loci for Q Sample. # of loci with: at least one allele matching = S; two alleles matching = D; variant allele matching = V; possible core repeat mutation = M*

Sample ID Key: *Q = Questioned (Victim Remains); DR = Direct Reference; M = Mother; O = Offspring; S = Sibling; H = Half-Sib*

FIG. 7—*The Consistency Check feature and detection of potential miscalls. In this example, two potential identities were considered for this victim: RM nos. 33327 and 37533. On early data, LRs favored the second family, no. 37533. On later data sets, a direct reference for RM no. 33327 and offspring to RM no. 37533 became available and confirmed the victim as RM no. 33327, of the first family.*



D3S1358 1	D3S1358 2	vWA 1	vWA 2	FGA 1	FGA 2	AMEL 1	AMEL 2	D8S1179 1	D8S1179 2	D21S11 1	D21S11 2	D18S51 1	D18S51 2	D5S818 1	D5S818 2	D13S317 1	D13S317 2	D7S820 1	D7S820 2	THO1 1	THO1 2	TPOX 1	TPOX 2	CSF1PO 1	CSF1PO 2	D16S539 1	D16S539 2	Kinship index				Sample ID	RM#	Other kin (same RM#) "S" scores	Likelihood ratio, displayed if >100			Kin biological relationship		
																												S/T	D	V	M				Parent: child	Full sib	Half sib	As per LIMS	Algorithm-inferred	After Admin Review
17	18	14	17	22	24	X	Y	8	12	30	32.2	17	18	10	11	8	13	10	12	7	7	8	11	9	10	11	12					Q-2410754								
Identification Lead																																								
Match to direct reference																																								
17	18	14	17	22	24	X	Y	8	12	30	32.2	17	18	10	11	8	13	10	12	7	7	8	11	9	10	11	12	13/13	13	1		DR-00077-04	11545		N/A	N/A	N/A			
Alternative Descendant Trios, 2 possible fathers for this Victim																																								
Descendant trio #1																																								
14	17	17	17	21	22	X	Y	12	12	30	32.2	16	18	10	11	8	8	9	12	6	7	8	11	9	12	11	12	13/13	4			F-21614 #02	11545	M13 F13	4.E+05	3.E+04	5.E+03	Father	Father	Father
17	18	14	16	19	24	X	X	8	14	30	31.2	17	20	11	13	13	13	10	11	7	9	11	12	10	11	12	12	13/13	1	1		M-21614 #04	11545	F13 F13	3.E+03		2.E+02	Mother	Mother	Mother
Descendant trio #2																																								
17	18	14	17	22	24	X	Y	12	14	30	32.2	18	20	11	13	8	13	11	12	7	9	8	12	9	11	11	12	13/13	6			F-21614 #05	11545	M13 F13	7.E+04	3.E+05	2.E+03	Father	Father	Brother
17	18	14	16	19	24	X	X	8	14	30	31.2	17	20	11	13	13	13	10	11	7	9	11	12	10	11	12	12	13/13	1	1		M-21614 #04	11545	F13 F13	3.E+03		2.E+02	Mother	Mother	Mother

Match Shading Key: A = no match B = match C = matching variant allele D = possible core repeat slip mutation ^^^ This formatting ^^^ = Software-detected discrepancies

Kinship Index Key: T = Total Q Sample Reported Loci. # of loci with: at least one allele matching = S; two alleles matching = D; variant allele matching = V; possible core repeat mutation = M

Sample ID Key: Q = Questioned (Victim Remains); DR = Direct Reference; F = Father; M = Mother

FIG. 8—*The Consistency Check feature and detection of mis-reported relationships. Two separate male individuals (i.e., different genotypes) reported a father relationship for this victim. Considering that, for WTC, the most frequent error in reporting relationship was to describe the relationship of the victim to the NOK instead of the opposite, the second individual (F-21614# 05) appeared more likely to be an offspring. An administrative review of the case proved the assumption correct.*

processing contingencies would have been very different if the collapse of the WTC towers had trapped the normal weekday occupancy of 50,000. The last scenario in Table 1 with a million casualties is an example of the possible death toll of a tsunami hitting a large coastal city, or of a nuclear strike against a large metropolitan area, and showcases the need for powerful bioinformatics applications to handle such large data sets.

Discussion

Over the last decade, the use of DNA typing in numerous victim identification initiatives has firmly established the technology as a powerful human identification modality in the wake of MFIs. DNA is a robust molecule, and identity information can be derived from minute amounts of biological material/tissue, attributes that make DNA typing ideal for the identification of victims of high body fragmentation incidents. Given adequate processing resources and light-to-moderate remains decay, the generation of genotypes from remains recovered from most MFIs no longer constitutes a major logistical obstacle. As well, the comparative genotyping component of such incidents can be a straightforward process if PEs are readily available for the vast majority of victims, the source attribution of PEs is not an issue, and both parents, spouse and offspring are available to provide genotypic references for kinship analysis. The reality of such events in the field, however, rarely meets these ideal criteria. More often than not, for most high-fragmentation MFIs such as air crashes, a number of possible complicating circumstances (see Table 2) can lead to significant additional challenges that make DNA-based identifications more difficult. The design of any comparative genotyping software must take into account extensive variation in the limitations imposed by the circumstances of different events.

With respect to comparative genotyping, the unique circumstances of the WTC incident represented a significant departure from previous situations. First, the number of victims and recovered remains were both 10-fold higher than what had been encountered in previous jetliner mishaps, the computational requirements for the WTC situation being exponentially increased. Second, the WTC incident was considered "open" as the list of missing individuals was not accurate. Third, a large proportion of Q genotypic data was made up of partial genotypes, making it more difficult for these partial genotypes to exceed the RMP threshold during comparison with complete PE genotypes. Fourth, except for one pair of siblings, the victims were not related, thereby eliminating the requirement for the reconstruction of family pedigrees from within the victims' genotype pool, the only aspect of the incident that facilitated DNA identification work.

TABLE 2—*Complicating circumstances encountered in many mass fatality incidents.*

Remains recovery is partial
A proportion of recovered remains has incurred significant thermal/chemical/bacterial decay
Many of the most probative PEs are unavailable (e.g., traveling with their owner and lost in the incident)
A proportion of available PEs recovered from the victims' residences carry biological trace material from an individual other than the anticipated victim
Partial or complete families may be among the victims thereby making pedigree reconstruction from within the victim's data set necessary
Many potential contributors to parentage trios involving older victims may be predeceased
Many victims may have few NOKs that can be used as genotypic references

Pair-wise comparisons are the mainstay of computing efforts aimed at matching Q and K genotypes in MFI victim identification. A direct match to a PE genotype provides the most direct route to identification, assuming the Q genotype holds enough alleles for the RMP to clear the set threshold, and the PE's source attribution is deemed reliable or has been confirmed through kinship analysis. In the absence of a PE for a given victim, a parentage trio detectable through pair-wise comparisons is the next best path although it provides evidence of biological relationship, not identity. If the available NOKs for a given victim do not allow for parentage trio analysis to be used, then any number of NOKs displaying P/O relationships may provide an identification lead through adequate LRs. In the absence of P/O relationships, then multiple second-degree relatives from both paternal and maternal ancestry (9) might provide an identification lead. The use of pedigree analysis packages (6) becomes useful in the last two situations. Finally, regardless of whether PEs and/or which NOKs are available as biological references for a given victim, all pair-wise comparisons being considered for a given Q sample must be consistent with the reported relationship to the victim. Pair-wise comparisons are also used to collapse the Q genotype data set, reducing large number of remains genotypes to a smaller data set. An RMP threshold value is used to allow for reconstruction of "virtual" Q genotypes to maximize the use of available data and benefit situations where little Q data is available.

The MDKAP was designed to capture this logic into a rules-based system suitable for the WTC incident and for future large-scale events involving related victims. The premise that most identification leads can be secured without the use of complex statistical tools was proven correct in the context of the Swissair Flight 111 MFI victim identification initiative (9). This screening concept was further expanded to include an array of additional capabilities and has provided victim identification capability for the WTC incident. MDKAP sorts reference samples from a genotypic data set according to basic rules of genetic association to remains genotypes and derives identification leads that can be further evaluated with appropriate statistical tools. Although triangulation approaches could have significantly reduced computing requirements, they were not employed in order to: (1) maintain the capability to detect inaccuracies in NOK self-reported biological relationships to the victims, thereby decreasing the potential for missed identifications, (2) preserve the ability to locate parentage trios from within the victims data set in future MFIs that may include families among the victims. All identification leads produced by MDKAP were confirmed upon subsequent statistical evaluation.

The several hundreds of millions of pair-wise comparisons performed with each incremental data set were successfully handled overnight with a VBA application on a desktop computer. With pair-wise comparisons schemes, as computational requirements grow exponentially with the number of victims, the size of some future events may eventually exceed the capacity of such a platform, as shown in Table 1. A C++ version operating in a parallel processing environment has been written (13): up to 1 million genotypes can be processed under the algorithms described in this paper while maintaining a practical execution time.

Considering the escalating computational workload imposed by nontriangulated parentage trio searches for MFIs involving related victims, the task of identifying these victims through DNA analysis would be greatly facilitated if every victim had a personal reference sample stored away. However, except for situations involving military personnel where the collection of personal reference samples may be mandatory, it is impractical to expect this type of personal reference sample storage to ever become common practice in the general population. Retrieval of probative PEs from the victims' residences will remain a valuable source of reference material but this type of reference material may prove unrecoverable in certain large-scale incident scenarios. Events like tsunamis and nuclear blasts impacting residential areas could lead to the death of large numbers of families as well as concomitant destruction/dispersal of nearly all usable PEs for these victims. In these situations, if the level of body fragmentation is severe, DNA identification would have to rely heavily on kinship analysis, making use of any remaining reference NOKs living away from the devastated area.

In more general terms, it is reasonable to anticipate that, for most incidents, a combination of PEs from a variable proportion of the victims and kinship analysis with available relatives will provide the available data processing path. Fortuitous Parentage Trios (FPT) were detected in both the Swissair and WTC events, and larger scale events involving related victims would be vulnerable to FPT nondetection. Beyond hardware/software enhancements, these larger-scale events may require additional nuclear STRs as well as Y-STRs and mtDNA to help confirm relationships in detected parentage trios. The bioinformatics components of such events can be simulated *in silico* to address issues of scale and complexity, and the processing outcome used to delineate the genetic technology enhancements required to meet the challenge of future larger scale events.

Despite the challenging condition of many remains recovered from the WTC disaster site, DNA typing generated data to support a large number of identifications. Along with other software, MDKAP was successfully adapted to afford the efficient use of recovered STR genotypic data and generate a maximum number of identification leads from the available STR data. Kinship analysis has proven successful in this incident but its continued success in larger-scale MFIs is contingent on the use of a sufficient number of STR loci to reduce the risk of undetected FPTs, the use of mtDNA and Y-STRs to confirm parentage and of bioinformatics that can support large-scale comparative genotyping schemes capable of detecting parentage trios from within a group of related victims.

Acknowledgments

The authors wish to acknowledge the members of the Biology sections of the Office of the Medical Examiner of New York City and of the New York State Police for their support during development of MDKAP, Julia Reid for advice on statistical issues, Drs. Chantal J. Frégeau and Mecki Prinz for critical reading of the manuscript, the members of the KADAP group for their support and constructive feedback on performance and feature specifications, and the National Institute of Justice for providing the necessary logistics and financial support for the KADAP group meetings.

References

1. McCarty VO, Sohn AP, Ritzlin RS, Gauthier JH. Scene investigation, identification, and victim examination following the accident of Galaxy 203: disaster preplanning does work. J Forensic Sci 1987;32(4):983–7.
2. Ludes B, Tracqui A, Pfitzinger H, Kintz P, Levy F, Disteldorf M, et al. Medico-legal investigations of the Airbus, A320 crash upon Mount Ste-Odile, France. J Forensic Sci 1994;39(5):1147–52.
3. Hutt JM, Ludes B, Kaess B, Tracqui A, Mangin P. Odontological identification of the victims of flight AI. IT 5148 air disaster Lyon-Strasbourg 20.01.1992. Int J Legal Med 1995;107(6):275–9.
4. Olaisen B, Stenersen M, Mevåg B. Identification by DNA analysis of the victims of the August 1996 Spitsbergen civil aircraft disaster. Nat Genet 1997;15(4):402–5.

5. Cash HD, Hoyle JW, Sutton AJ. Development under extreme conditions: forensic bioinformatics in the wake of the World Trade Center disaster. Pac Symp Biocomput 2003;8:638–53.
6. Brenner CH, Weir BS. Issues and strategies in the DNA identification of World Trade Center victims. Theor Popul Biol 2003;63(3):173–8.
7. Leclair B. Large-scale comparative genotyping and kinship analysis: evolution in its use for human identification in mass fatality incidents and missing persons databasing. In: Doutremépuich C, Morling N, editors. *Progress in forensic genetics, Volume 10. Proceedings of the 20th Congress of the International Society for Forensic Genetics; 2003 Sept 9–13; Arcachon, France*. Amsterdam, The Netherlands: Elsevier Science, 2004;42–4.
8. Leclair B, Frégeau CJ, Bowen KL, Borys SB, Elliott J, Fourney RM. STR DNA typing and human identification in mass disasters: extending kinship analysis capabilities. In: Sensabaugh GF, Lincoln PJ, Olaisen B, editors. *Progress in forensic genetics, Volume 8. Proceedings of the 18th Congress of the International Society for Forensic Haemogenetics; 1999 Aug. 17–21; San Francisco (CA)*. Amsterdam, The Netherlands: Elsevier Science, 1999;91–3.
9. Leclair B, Frégeau CJ, Bowen KL, Fourney RM. Enhanced kinship analysis and STR-based DNA typing for human identification in Mass Fatality Incidents: the Swissair Flight 111 disaster. J Forensic Sci 2004;49(5):939–53.
10. Budowle B, Moretti TR, Baumstark AL, Defenbaugh DA, Keys KM. 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 1999;44(6):1277–86.
11. Biesecker LG, Bailey-Wilson JE, Ballantyne J, Baum H, Bieber FR, Brenner C, et al. DNA identifications after the 9/11 World Trade Center attack. Science 2005;310(5751):1122–3.
12. Budimlija ZM, Prinz MK, Zelson-Mundorff A, Wiersema J, Bartelink E, MacKinnon G, et al. World Trade Center human identification project: experiences with individual body identification cases. Croat Med J 2003;44(3):259–63.
13. Ananomouse Corporation, *Bloodhound software program*. Available at: http://www.ananomouse.com.

Additional information and reprint requests:
Benoît Leclair, Ph.D.
Myriad Genetic Laboratories, Inc.
320 Wakara Way
Salt Lake City, UT 84108
E-mail: bleclair@myriad.com