JOURNAL OF FORENSIC SCIENCES

J Forensic Sci, January 2011, Vol. 56, No. S1
doi: 10.1111/j.1556-4029.2010.01631.x
Available online at: onlinelibrary.wiley.com

PAPER

CRIMINALISTICS

Jianye Ge,[1,2] *Ph.D.; Bruce Budowle,*[1,2] *Ph.D.; and Ranajit Chakraborty,*[1,2] *Ph.D.*

Choosing Relatives for DNA Identification of Missing Persons

ABSTRACT: DNA-based analysis is integral to missing person identification cases. When direct references are not available, indirect relative references can be used to identify missing persons by kinship analysis. Generally, more reference relatives render greater accuracy of identification. However, it is costly to type multiple references. Thus, at times, decisions may need to be made on which relatives to type. In this study, pedigrees for 37 common reference scenarios with 13 CODIS STRs were simulated to rank the information content of different combinations of relatives. The results confirm that first-order relatives (parents and fullsibs) are the most preferred relatives to identify missing persons; fullsibs are also informative. Less genetic dependence between references provides a higher on average likelihood ratio. Distant relatives may not be helpful solely by autosomal markers. But lineage-based Y chromosome and mitochondrial DNA markers can increase the likelihood ratio or serve as filters to exclude putative relationships.

KEYWORDS: forensic science, DNA identification, missing person, pedigree likelihood ratio, short tandem repeats, posterior odds, Y chromosome, mitochondrial DNA, simulation

Over the past two decades, DNA forensic analyses of evidence from criminal and civil investigations have been extremely useful for personal identification. After the successes in identifying living persons in such cases as paternity testing, efforts shifted toward applying DNA-based identification to war victims in mass graves, missing soldiers, and missing person from mass disasters (1–13). When direct reference samples, such as personal items, from missing individuals are not available, identifications can be made indirectly based on ranking of likelihood ratios (LRs) constructed from a comparison of the probability of observing DNA profiles of remains of a presumed relationship with profiles from reference samples of the alleged family members versus the hypothesis that the remains are unrelated to the family (14–23). Brenner (17) described a way to calculate posterior odds for identification and set 99.9% as a degree of confidence. The 99.9% confidence was also advocated by the DNA Commission of the International Society for Forensic Genetics (12) and those who worked on identifying the human remains from the terrorist attack on the World Trade Center (10). There are two ways to increase the power of identification: (i) type more markers and (ii) type more relatives. The number of markers that can be typed will be limited by the quality and quantity of DNA derived from a remains. If there is sufficient DNA, then a large battery of genetic markers are available to assist in making an identification. In many cases, the quality and quantity of DNA is poor. Increasing the number of reference relatives can increase the chances of identifying remains and particularly for challenged samples. In some cases, the number of relatives can be quite large, but in others, the number of available relatives is very limited. Typing all relatives of a large pedigree can be costly and may not be necessary to reach a certain threshold for identification. Because there are information and cost factors regarding the selection and number of relatives, respectively, typed, some selection criteria should be considered to guide identity testers. The probabilities of identity with certain combinations of relatives are more powerful than are other combinations.

In this study, the 37 most common relative combination scenarios in missing person identifications were selected and using the 13 CODIS short tandem repeats (STRs) as genetic profile data large numbers of pedigrees (e.g., 1,000,000) were simulated for each scenario. The distribution of LRs for each scenario was evaluated to first confirm the well-known single relative reference scenarios and second to determine the most informative combinations of relatives for identifying an unknown person. Thus, guidance is given on which and how many relatives should be selected and typed for kinship analyses for identification so that efficiency can be optimized under the constraints of limited resources. The LRs were calculated by a novel software program MPKin (24,25), which is based on the Elston-Stewart algorithm (26) and jointly considers DNA profiles from all available family members and missing persons/remains. MPKin was validated using several full DNA profile pedigrees with the help from International Commission on Missing Persons.

Identification Principle

To evaluate whether a missing person (*MP*) belongs to a family pedigree (*P*), usually one or more family reference persons from the putative pedigree are typed. Identification is assessed basically by comparing two alternative hypotheses: H_p: *MP* is the specific

[1]Institute of Investigative Genetics, University of North Texas Health Science Center, 3500 Camp Bowie Blvd., Fort Worth, TX 76107.
[2]Department of Forensic and Investigative Genetics, University of North Texas Health Science Center, 3500 Camp Bowie Blvd., Fort Worth, TX 76107.

Received 2 Sept. 2009; and in revised form 9 Nov. 2009; accepted 27 Dec. 2009.

member of the putative pedigree and H_d: *MP* is unrelated to the known reference members of the putative pedigree.

The LR is calculated based on a ratio of the probabilities of the DNA evidence under each hypothesis, represented by the general expression:

$$LR = \frac{\Pr(G_{MP}, G_P|H_p)}{\Pr(G_{MP}, G_P|H_d)} \quad (1)$$

where G_{MP} refers to the DNA profile of the missing person (from remains) and G_P represents the joint DNA profiles of all typed family members in the pedigree. H_p is favored if the LR is >1, and when the LR is <1, H_d is better supported. Usually, very high LRs are required to confirm the relationship although prior odds need to be considered as well to provide posterior odds of the putative relationship. For H_p, usually, the position of *MP* in *P* is fixed. If no prior information of *MP* is provided to specify H_d, *MP* may be regarded as not related to anyone in *P*. The known members of *P* are confirmed by consistency of their genetic data and their reported biologic relationships prior to any comparison with an unknown sample.

In many mass disaster and missing person cases, LRs of each unknown sample with each putative pedigree will be calculated; an identification is made with both the calculated LRs and prior information if a defined threshold is met. Other meta-data may confirm or refute the putative genetic relationship.

Simulation Study and Results

Thirty-seven common reference scenarios (ranging from a single relative as a reference sample to combinations of relatives for kinship analysis) were selected. The pedigrees consisting of DNA profiles for each scenario were simulated using the Caucasian population data on the 13 CODIS STR loci (27) assuming no population substructure and mutation (Table 1). To generate simulated data, the alleles of founders (i.e., individuals without parents in the pedigree) were randomly assigned according to the allele frequencies of each locus (27), and each locus was treated independently. Founders transmitted with equal probability a single allele at each locus to his/her offspring. One million pedigrees were simulated for most scenarios. As a result of the computational complexity of the LR calculation, only thousands or tens of thousands of pedigrees were simulated for some complex scenarios. Logarithm base 10 of the LR, $Log_{10}(LR)$, was calculated by comparing the probabilities of observing a DNA profile under two hypotheses: the missing person belongs to the pedigree under a specified relationship or the missing person is unrelated to the pedigree. Table 1 shows the mean and variance of the $Log_{10}(LR)$ distributions of reference scenarios, as well as the 5th, 1st, and 0.1th percentiles of the distributions (i.e., 95%, 99%, and 99.9% confidence level, respectively) and number of simulations in each scenario. As already well known, for pedigrees containing a single reference relative (Fig. 1), informativeness of relationships can be ranked as follows:

TABLE 1—*Means, variances, 5th, 1st, and 0.1th percentiles of $Log_{10}(LR)$ of relative scenarios.*

Relative scenarios	Mean	Variance	5%	1%	0.1%	Simulations
3 children + spouse	**12.4199**	3.72	**9.31**	**8.054**	**6.646**	1,000,000
4 children	**10.4023**	3.2167	**7.481**	**6.216**	4.725	1,000,000
Both parents	**10.2561**	2.0865	**8.071**	**7.337**	**6.5855**	1,000,000
2 spouses + 2 children (1 each)	**10.1768**	3.3654	**7.303**	**6.258**	5.141	1,000,000
2 children + spouse	**10.1122**	3.4419	**7.197**	**6.133**	4.9875	1,000,000
3 children	**9.057**	3.2728	**6.14**	4.958	3.6355	1,000,000
1 parent + 3 fullsibs	**8.8683**	2.9959	**6.07**	4.878	3.504	1,000,000
1 child + 1 parent + spouse	**8.8089**	2.6358	**6.285**	5.356	4.369	1,000,000
Spouse + 1 child + 1 child with 2nd spouse	**8.8088**	2.6381	**6.288**	5.35	4.36	1,000,000
4 fullsibs	**8.5438**	3.419	5.493	4.147	2.593	1,000,000
1 fullsib + 1 child + spouse	**8.1415**	3.1616	5.314	4.231	3.046	1,000,000
1 parent + 2 fullsibs	**7.9474**	3.2484	5.03	3.839	2.5135	1,000,000
3 fullsibs	**7.5199**	3.74	4.346	2.985	1.4155	1,000,000
1 child + 1 parent	**7.2595**	2.2328	4.925	4.03	3.0285	1,000,000
2 children	**6.9823**	2.8332	4.334	3.346	2.304	1,000,000
1 fullsib + 1 child	**6.5897**	3.4195	3.628	2.457	1.15	1,000,000
1 fullsib + 1 parent	**6.3159**	3.3385	3.39	2.248	1.01	1,000,000
1 child + spouse	**6.1687**	1.2632	4.497	3.935	3.353	1,000,000
2 fullsibs	5.877	3.9238	2.653	1.339	−0.116	1,000,000
1 halfsib + 1 parent (not the parent of the halfsib)	5.5361	2.4912	3.028	2.056	0.9969	1,000,000
1 uncle + 1 parent (they are not related)	5.5344	2.4926	3.028	2.047	0.9628	1,000,000
1 grandchildren + 1 child (they are uncle/nephew)	4.6249	1.6739	2.623	1.885	1.084	1,000,000
1 parent/1 child	4.086	1.1195	2.484	1.915	1.3155	1,000,000
1 halfsib + 1 fullsib	3.9493	3.7025	0.8514	−0.3731	−1.719	1,000,000
1 fullsib	3.4193	3.6201	0.3731	−0.7899	−2.05	1,000,000
2 uncles (they are not related)	1.9935	2.0261	−0.2634	−1.168	−2.2525	10,000
2 grandchildren (who are cousins)	1.707	1.5767	−0.3177	−1.1385	−2.0365	50,000
2 halfsibs (2 halfsibs are also halfsibs)	1.6348	1.3162	−0.2141	−0.9612	−1.8105	1,000,000
2 halfsibs (2 halfsibs are fullsibs)	1.4454	1.1333	−0.2844	−1.034	−1.901	1,000,000
2 grandchildren (who are fullsibs)	1.4449	1.1318	−0.2847	−1.031	−1.9115	1,000,000
2 uncles (who are fullsibs)	1.4442	1.1325	−0.2869	−1.03	−1.9075	1,000,000
1 grandparent/grandchild	0.9154	0.8947	−0.5804	−1.164	−1.79	1,000,000
1 uncle/nephew	0.9149	0.8938	−0.5797	−1.16	−1.792	1,000,000
1 halfsib	0.9138	0.8929	−0.5793	−1.162	−1.799	1,000,000
2 cousins (they are also cousins)	0.4691	0.4605	−0.5709	−0.9808	−1.4165	10,000
2 cousins (they are fullsibs)	0.3661	0.3637	−0.5539	−0.8964	−1.3425	25,000
1 cousin	0.2485	0.2607	−0.5054	−0.7674	−1.039	1,000,000

Bold numbers (i.e., greater than 6) indicates informative identification.

- Parent/child
- Fullsib
- Grandparent/grandchild, uncle/nephew, halfsib
- First cousin

This part of the study was performed to establish that the simulations were providing reliable data on relationship and information content for kinship analysis. The parent or child is the most preferred reference among single relative scenarios with all LRs greater than one. The fullsib scenario has the highest variance of LRs among all single reference relative scenarios, owing to the wide distribution of identity-by-descent (IBD) alleles with fullsibs (i.e., 1/4, 1/2, and 1/4 for IBD = 0, 1, and 2, respectively). Grandchild, uncle, and halfsib essentially have the same distributions. The majority of LRs (i.e., 99.5%) with a first cousin scenario (cousin in short) were <100, which is consistent with the known relationship of that of a cousin, and that such genetic data (i.e., STRs) typically do not provide much information for missing person identification.

Because the data are consistent with the known power of single source relatives of various relationships, multiple reference relative pedigrees were evaluated. Again as expected, the scenarios with a higher number of closer relatives generally gave higher LRs. LRs with two children are expected to be on average 12 times greater than those for two fullsibs, which are 60 times more than those for single parents (Fig. 2, Table 1). The two parents scenario gives more than 1000 times greater LRs than one parent + one child and two children scenarios, mostly because genotypes of parents are (assumed to be) independent but genotypes of children are dependent on each other and on the missing person and parent of the missing person. The one parent + one child scenario has slightly greater average LRs and lower variance than the two children scenario. Hence, parents are more informative than children in multiple relative scenarios. This observation is consistent with the practices in paternity testing.

Spouses are typically unrelated. An LR can increase several orders of magnitude if the spouse is included in scenarios with the corresponding children as reference samples, e.g., in reconstruction cases. LRs with a spouse are expected to be more than 100 times higher than those of scenarios without a spouse and if only one child is available. The LRs with a spouse could be higher with multiple children; one more child increases the LRs by two magnitudes on average with an accompanying slightly higher variance as expected (Fig. 3). The scenario with two spouses + two children gives comparable LRs as the both parents scenario, although the variance is larger. The spouse + two children scenario has a similar distribution as two spouses + two children (one child for each spouse) scenario (Table 1). The average LRs of one parent + fullsib(s) scenario are apparently less than those of a spouse + children (Fig. 3). The average LR of one fullsib + one child is slightly greater than that in one fullsib + one parent because of the closer genetic dependence between fullsib and parent. Genetic dependence can be measured by the kinship coefficient; the smaller the kinship coefficient values, the less genetic dependence there is between relationships. It would take about seven fullsibs + one parent to obtain similar LRs on average as that of both parents, but only two children + one spouse are required to achieve similar LRs (Fig. 3). Four children or 10 fullsibs can also yield comparable LRs (Fig. 4).

It was interesting to notice that three fullsibs, one fullsib + one parent, one fullsib + one child, and two children scenarios yielded higher LRs than one child + spouse, one common paternity testing scenario (Table 1). However, these four scenarios have much higher variance than that of one child + spouse, which leads to higher probability of false exclusions for identifications. Some scenarios' rankings are worth noting: (i) two halfsibs (two halfsibs are also halfsibs) > two halfsibs (two halfsibs are fullsibs); (ii) two uncles (who are not related) > two uncles (who are fullsibs); (iii) two grandchildren (who are cousins) > two grandchildren (who are fullsibs); and (iv) two cousins (who are also cousins) > two cousins (who are fullsibs). Because halfsibs, uncle/nephew, and grandparent/grandchild have the same IBD distribution, the LR distributions of two halfsibs, uncles, or grandchildren (who are fullsibs) are identical. But if the two relatives are not fullsibs, less genetic dependence between these relatives gives a higher average LR (i.e., two unrelated uncles > two grandchildren as cousin > two halfsibs as halfsib for average LR).

The same simulations as above were performed for one parent, one fullsib, and one halfsib scenarios but with 15 Caucasian STRs (i.e., 13 CODIS STR + the loci D2S1338 and D19S433; Table 2). The average LR of the one parent scenario is almost sevenfold higher with these two extra STRs, and the increment is less than twofold for the halfsib scenario. As already known, closer relatives will provide higher LR increments. Additional markers can provide higher LRs and more reliable identifications.

Discussion

DNA-based analysis often is an essential part of missing person identification. In many cases, identity is inferred from reference

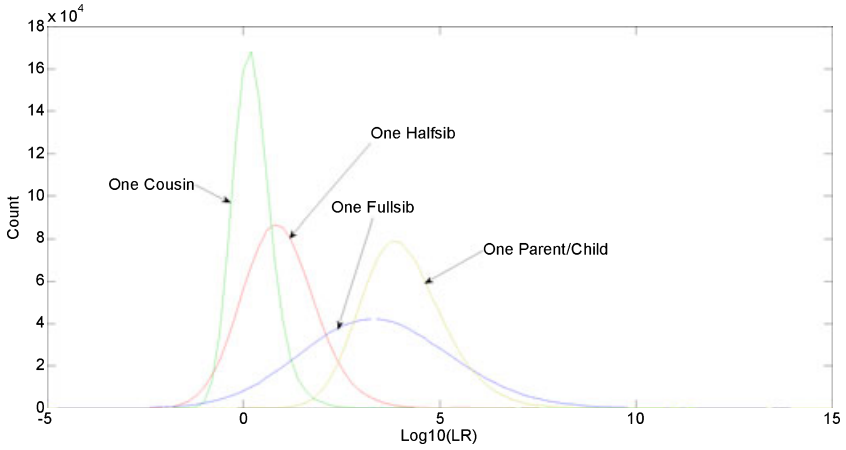

FIG. 1—*Log_{10}(LR) distributions of single relative scenarios.*

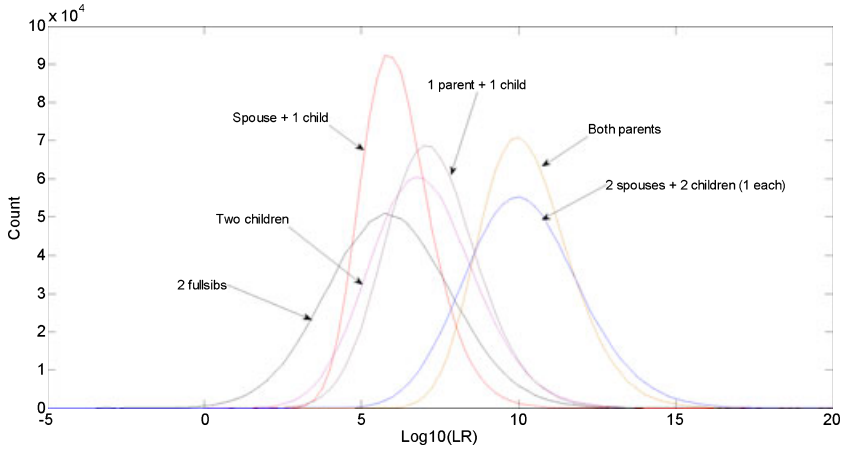

FIG. 2—*Log_{10}(LR) distributions of some common reference scenarios.*

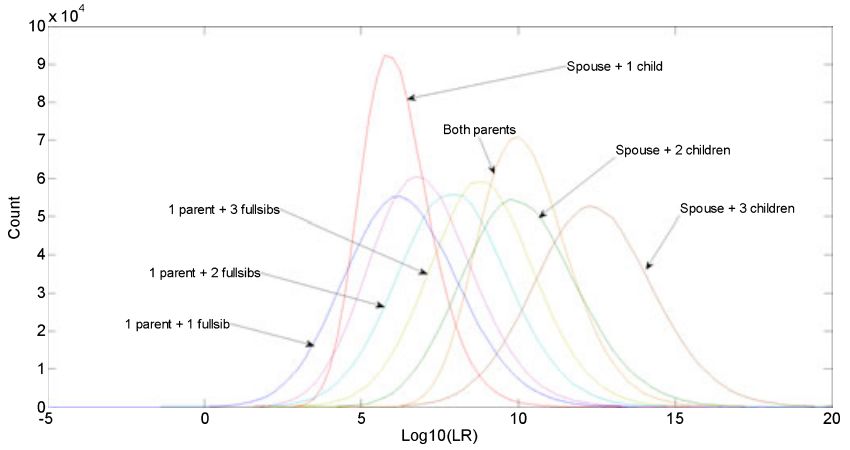

FIG. 3—*Log_{10}(LR) distributions of 1 parent + fullsibs, 1 spouse + child(ren) and both parents.*

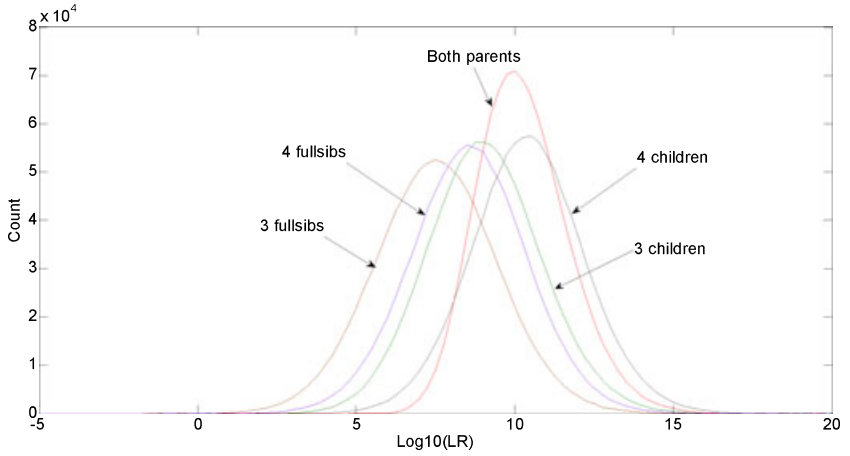

FIG. 4—*Log_{10}(LR) distributions of high number of fullsibs and children versus both parents.*

relatives by kinship analysis. Choosing the most informative relatives can impact positively identification and can reduce the cost by minimizing superfluous testing. We simulated 37 common reference pedigree scenarios in missing person identifications with 13 CODIS STR data. This study initially confirmed by simulation that when a single family reference samples are all that are available generally genetically closer relatives will yield higher LRs, in the order of parents > children > fullsibs > halfsibs = uncles/nephews = grandparents/grandchildren > cousins. These relationships are well known in kinship analyses (28,29). However, an identification based solely using a single reference relative often may be inadequate because large proportions of LRs are small (i.e., <1000). For

TABLE 2—*Means and variances of three single relative scenarios with 13 or 15 CODIS STRs (i.e., 13 core loci + D2S1338 and D19S433).*

	13 CODIS STRs		13 CODIS + 2 STRs	
Scenarios	Mean	Variance	Mean	Variance
One parent/child	4.086	1.1195	4.9048	1.3348
One fullsib	3.4193	3.6201	4.0845	4.3641
One halfsib	0.9138	0.8929	1.1072	1.0835

instance, with a single parent, about 14.5% of LRs are less than 1000. According to previous studies (10,12), a minimum of 99.9% is suggested as the posterior probability (i.e., 1000 is the minimum posterior odds threshold) for rendering an identification. In cases where there are multiple unknown persons to identify the posterior odds = (LR) (prior odds), an LR of 10^6 is required for a one thousand missing person pool (the prior probability would be 1/1000); the prior odds will differ depending on the number of missing persons in a particular case) (17,18). The study by Brenner (17) was inaccurate for the probability of identity for some relationship scenarios likely due to a very limited number of simulations carried out (i.e., 30–50 simulations). With a million simulations, better predictions of LR distributions were obtained.

In our simulations, only 4.6% of LRs of the single parent scenario are greater than 10^6. Hence, in situations where there are multiple unknown victims, such as mass disasters, multiple reference relatives are highly recommended and necessary to achieve sufficient information for identification even with 13 STR loci. Table 1 shows the calculated 5th, 1st, and 0.1th percentiles of each scenario. Only both parents and three children + spouse have more than 99.9% of LRs greater than 10^6. Four children, two spouses + two children (one each), and two children + spouse can yield LRs greater than 10^6 with 99% confidence. Several more scenarios (bolded in 5th percentile column) may also be informative if the confidence level is lowered to 95% (however, false positives will increase). With both mean and variance considered, these scenarios are the most reliable for identification. Typing more markers can allow more scenarios to obtain 99% or even 99.9% of LRs $>10^6$. Unfortunately, the quantity and quality of DNA samples from human remains may not be sufficient in many cases to enable additional marker typing. If more typing can be carried out, we recommend analyzing more genetic markers in concert with selecting the most informative members of a pedigree to manage the costs.

Another important issue in missing person cases is that the relationships within the putative pedigrees should be validated by kinship analysis tools (e.g., MPKin) before identifying missing persons. If any false relationship was found, the individual(s) with false relationship should be removed from the reference pedigree.

Y Chromosome and Mitochondrial DNA

In cases where close relatives (i.e., father, sons, etc.) may not be available or the number of relatives is limited, the LR calculated based on distant or limited number of relatives will often be insufficient to draw any meaningful conclusion for identification of an unknown person. Additionally, while this study did not include it, population substructure effects can reduce the LRs by several orders of magnitude (29–31). However, lineage-based systems, such as the markers residing on the Y chromosome and mitochondrial DNA (mtDNA) genome, can be used as additional tools to increase the LR and may be the only informative markers for multigenerational comparisons.

The Y chromosome and the mtDNA genome are only inherited in a paternal and a maternal lineage, respectively, and have been very useful in identification cases (32,33). Y chromosome and/or mtDNA haplotypes are commonly used as filters to eliminate genetically improbable pedigree relationships for the unknown person before assessing relationships with autosomal markers. Autosomal STRs biologically assort independently of mitochondrial sequences and Y-chromosomal haplotypes. Budowle et al. (34) and Walsh et al. (35) further statistically demonstrated that forensic markers on the Y chromosome and mtDNA genome are independent with the forensically selected autosomal STR loci. Therefore, the LRs calculated for each of these three marker groups can be directly multiplied together. The discrimination power of current multiplex Y-STR systems used for identification can reach 0.999 with 10 Y-STRs and greater than 0.9999 with 16 Y-STRs (31,35). The LRs are expected on average to be at least 1000 times higher (currently limited by reference population size) if a Y haplotype match was found between a missing person and putative family members. Even with a one-step mutation in the Y haplotype, the LRs can be doubled (assuming a mutation rate is 0.002; [36]). The mtDNA haplotype has a similar discrimination power also depending on the length of sequence used and the size of the reference population database. If both mtDNA and Y haplotypes are combined, the LRs on average could be 1,000,000 times higher (some isolated populations would yield lower LRs).

Recommendations

For single relative reference samples, the selection and value for identification is well known. But because multiple reference samples are required to obtain sufficient power for many identification cases, these recommendations may be helpful for selecting the most likely combination of reference relatives to on average yield the highest LRs for true pedigree relationships and the best probability of excluding false relationships. These relationships should be considered when approaching indirect analyses because the alternate approach of using more genetic markers to achieve sufficient power is not always viable with challenged biologic samples or cost may limit the number of reference samples that can be typed. Based on the simulation results, the following guidelines are recommended in choosing relatives for missing person identification.

- Parents are the preferred relatives, and both parents of the missing person should be typed when possible. If both parents are typed, all other relatives, including fullsibs, may not be necessary.
- Children are the second preferred relatives. Type as many children as possible or until the unknown genotype of the missing person can be reconstructed. In absence of parents, if the missing person is of male sex, sons are preferred because of the same Y chromosome shared between father and sons; otherwise, sons and daughters are equivalent.
- Even if a child is available, the spouse of the missing person (i.e., the father/mother of the child) should be considered for typing, if he/she is available.
- Fullsibs are the third preferred relatives. If the missing person is of male sex, brothers are preferred compared to sisters, because missing person and brothers share both Y chromosome and mtDNA, and it is reasonable to type less relatives with the same discrimination power owing to economical reasons in some conditions; otherwise, brothers and sisters are equivalent.
- All other distant relatives, such as grandparents/grandchildren, halfsibs, uncles/aunts, and cousins, only provide limited identification

capabilities based on autosomal markers, but their Y chromosomes and mtDNA can be used to increase LR or filter out false relationships.
- Less genetic dependence between reference relatives provides a higher LR on average. This is practiced routinely for standard paternity cases where two unrelated parents are sought. But the concept can be applied to extended pedigrees as well. For example, two biologically unrelated uncles can be more informative than two related uncles.
- With limited number of relatives, type as many as markers as possible.

Conflict of interest: The authors have no relevant conflicts of interest to declare.

References

1. Clayton TM, Whitaker JP, Maguire CN. Identification of bodies from the scene of a mass disaster using DNA amplification of short tandem repeat (STR) loci. Forensic Sci Int 1995;76(1):7–15.
2. Olaisen B, Stenersen M, Mevag B. Identification by DNA analysis of the victims of the August 1996 Spitsbergen civil aircraft disaster. Nat Genet 1997;15(4):402–5.
3. Hsu CM, Huang NE, Tsai LC, Kao LG, Chao CH, Linacre A, et al. Identification of victims of the 1998 Taoyuan Airbus crash accident using DNA analysis. Int J Legal Med 1999;113(1):43–6.
4. Lorente JA, Entrala C, Alvarez JC, Arce B, Heinrichs B, Lorente M, et al. Identification of missing persons: the Spanish ''Phoenix'' program. Croat Med J 2001;42(3):267–70.
5. Huffine E, Crews J, Kennedy B, Bomberger K, Zinbo A. Mass identification of persons missing from the break-up of the former Yugoslavia: structure, function, and role of the International Commission on Missing Persons. Croat Med J 2001;42(3):271–5.
6. Meyer HJ. The Kaprun cable car fire disaster—aspects of forensic organization following a mass fatality with 155 victims. Forensic Sci Int 2003;138(1–3):1–7.
7. 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.
8. Marchi E. Methods developed to identify victims of the World Trade Center disaster. Am Labor 2004;36:30–6.
9. Alonso A, Martin P, Albarran C, Garcia P, Ferandez de Simon L, Iturralde MJ, et al. Challenges in DNA profiling in mass disaster investigations. Croat Med J 2005;46(4):540–8.
10. Budowle B, Bieber FR, Eisenberg AJ. Forensic aspects of mass disasters: strategic considerations for DNA based human identification. Legal Med 2005;7(4):230–43.
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. Prinz M, Carracedo A, Mayr WR, Morling N, Parsons TJ, Sajantila A, et al. DNA Commission of the International Society for Forensic Genetics (ISFG): recommendations regarding the role of forensic genetics for disaster victim identification (DVI). Forensic Sci Int Genet 2007;1(1):3–12.
13. Lee J, Scott P, Carroll D, Eckhoff C, Harbison S, Ientile V, et al. Recommendations for DNA laboratories supporting Disaster Victim Identification (DVI) Operations—Australian and New Zealand consensus on ISFG recommendations. Forensic Sci Int Genet 2008;3(1):54–6.
14. DNA Advisory Board. Statistical and population genetic issues affecting the evaluation of the frequency of occurrence of DNA profiles calculated from pertinent database(s). Forensic Sci Commun 2000;2(3), http://www2.fbi.gov/hq/lab/fsc/backissu/july2000/dnastat.htm (accessed November 12, 2010).
15. Gornik I, Marcikic M, Kubat M, Primorac D, Lauc G. The identification of war victims by reverse paternity is associated with significant risk of false inclusion. Int J Legal Med 2002;116(5):255–7.
16. Biruš I, Marciki M, Lauc D, Dijan S, Lauc G. How high should paternity index be for reliable identification of war victims by DNA typing? Croat Med J 2003;44(3):322–6.
17. Brenner CH. Reuniting El Salvador families, http://dna-view.com/ProBusqueda.htm (accessed March 16, 2009).
18. Brenner CH, Weir BS. Issues and strategies in the identification of World Trade Center victims. Theor Popul Biol 2003;63(3):173–8.
19. Leclair B, Fregeau 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.
20. Dzijan S, Primorac D, Marcikic M, Andelinovic S, Sutlovic D, Dabelic S, et al. High estimated likelihood ratio might be insufficient in a DNA-lead process of war victim's identification. Croat Chem Acta 2005;78(3):393–6.
21. Buckleton J, Triggs CM, Clayton T. Disaster victim identification, identification of missing persons, and immigration cases. In: Buckleton J, Triggs CM, Walsh SJ, editors. Forensic DNA evidence interpretation. Boca Raton: FL: CRC Press, 2005; 395–437.
22. Brenner CH. Some mathematical problems in the DNA identification of victims in the 2004 tsunami and similar mass fatalities. Forensic Sci Int 2006;157(2):172–80.
23. Leclair B, Shaler R, Carmody GR, Eliason K, Hendrickson BC, Judkins T, et al. Bioinformatics and human identification in mass fatality incidents: the world trade center disaster. J Forensic Sci 2007;52(4):806–19.
24. Ge J, Wang T, Birdwell DJ, Chakraborty R. Further remarks on: ''Paternity analysis in special fatherless cases without direct testing of alleged father'' (FSI 146S (2004) S159–S161) and remarks on it (FSI 163 (2006) 158–160). Forensic Sci Int 2007;3:e6–8.
25. Elston RC, Stewart J. A general model for the genetic analysis of pedigree data. Hum Hered 1971;21:523–42.
26. Ge J, Budowle B, Chakraborty R. DNA identification by pedigree likelihood ratio accommodating population substructure and mutations. Investig Genet 2010;1:8, doi:10.1186/2041-2223-1-8.
27. Butler JM, Schoske R, Vallone PM, Redman JW, Kline MC. Allele frequencies for 15 autosomal STR loci on U.S. Caucasian, African American, and Hispanic populations. J Forensic Sci 2003;48(4):908–11.
28. Li CC, Sacks L. The derivation of joint distribution and correlation between relatives by the use of stochastic matrices. Biometrics 1954;10:347–60.
29. Thompson EA. The estimation of pairwise relationships. Ann Hum Genet 1975;39:173–88.
30. Kracun SK, Curic G, Birus I, Dzijan S, Lauc G. Population substructure can significantly affect reliability of a DNA-led process of identification of mass fatality victims. J Forensic Sci 2007;52(4):874–8.
31. Ge J, Budowle B, Chakraborty R. US forensic Y-chromosome short tandem repeats database. Leg Med 2010;12(6):289–95.
32. Gusmão L, Butler JM, Carracedo A, Gill P, Kayser M, Mayr WR, et al. DNA Commision of the International Society for Forensic Genetics (ISFG): an update of the recommendations on the use of Y-STRs in forensic analysis. Forensic Sci Int 2006;157(2):187–97.
33. Carracedo A, Bär W, Lincoln PJ, Mayr W, Morling N, Olaisen B, et al. DNA Commission of the International Society for Forensic Genetics: guidelines for mitochondrial DNA typing. Forensic Sci Int 2000;110(2):79–85.
34. Budowle B, Ge J, Aranda X, Planz J, Eisenberg A, Chakraborty R. Texas population substructure and its Impact on estimating the rarity of Y STR haplotypes from DNA evidence. J Forensic Sci 2009;54(5):1016–21.
35. Walsh B, Redd A, Hammer M. Joint match probabilities for Y chromosomal and autosomal markers. Forensic Sci Int 2008;174(2):234–8.
36. Budowle B, Ge J, Aranda X, Planz J, Eisenberg A, Chakraborty R. Mutation rates at Y chromosome short tandem repeats in Texas populations. Forensic Sci Int: Genetics 2009;3(3):179–84.

Additional information and reprint requests:
Jianye Ge, Ph.D.
Institute of Investigative Genetics
Department of Forensic and Investigative Genetics
University of North Texas Health Science Center at Ft. Worth
3500 Camp Bowie Blvd
Ft. Worth, TX 76107
E-mail: Jianye.Ge@unthsc.edu