Journal Pre-proof

Disaster victim identification by kinship analysis: the Lampedusa October 3rd, 2013 shipwreck

Barbara Bertoglio, Pierangela Grignani, Paola Di Simone, Nicolò Polizzi, Danilo De Angelis, Cristina Cattaneo, Agata Iadicicco, Paolo Fattorini, Silvano Presciuttini, Carlo Previderè

PII: S1872-4973(19)30214-5

DOI: https://doi.org/10.1016/j.fsigen.2019.102156

Reference: FSIGEN 102156

To appear in: *Forensic Science International: Genetics*

Received Date: 10 May 2019

Revised Date: 14 August 2019

Accepted Date: 27 August 2019

Please cite this article as: Bertoglio B, Grignani P, Di Simone P, Polizzi N, De Angelis D, Cattaneo C, Iadicicco A, Fattorini P, Presciuttini S, Previderè C, Disaster victim identification by kinship analysis: the Lampedusa October 3rd, 2013 shipwreck, *Forensic Science International: Genetics* (2019), doi: https://doi.org/10.1016/j.fsigen.2019.102156

Disaster victim identification by kinship analysis: the Lampedusa October 3rd, 2013 shipwreck

Barbara Bertoglio[1,2], Pierangela Grignani[1], Paola Di Simone[3], Nicolò Polizzi[3], Danilo De Angelis[2], Cristina Cattaneo[2], Agata Iadicicco[4], Paolo Fattorini[5], Silvano Presciuttini[6], Carlo Previderè[1]

[1] Dipartimento di Sanità Pubblica, Medicina Sperimentale e Forense, Sezione di Medicina Legale e Scienze Forensi, Università di Pavia, Pavia, Italy

[2] LABANOF, Dipartimento di Scienze Biomediche per la Salute, Sezione di Medicina Legale, Università degli Studi di Milano, Milan, Italy

[3] Gabinetto Regionale Polizia Scientifica, Laboratorio di Genetica Forense, Palermo, Italy

[4] Ufficio del Commissario Straordinario per le Persone Scomparse, Roma, Italy

[5] Dipartimento Clinico di Scienze Mediche, Chirurgiche e della Salute, Università di Trieste, Trieste, Italy

[6] Dipartimento di Ricerca Traslazionale e Nuove Tecnologie in Medicina e Chirurgia, Università di Pisa, Pisa, Italy

Corresponding author: Carlo Previderè, Dipartimento di Sanità Pubblica, Medicina Sperimentale e Forense, Università di Pavia, via Forlanini12, 27100 Pavia, Italy, email: carlo.previdere@unipv.it

Highlights

- A boat full of more than 500 migrants sank in the Mediterranean in 2013
- Genetic profiles were extracted both from victims and putative living relatives
- The Identity by State (IBS) and the Identity by Descent (IBD) statistics were calculated
- Genetic identifications were confirmed by pedigree analysis using FAMILIAS

1. INTRODUCTION

Among the few main paths used by migrants in their attempt to reach Europe, the "Central Mediterranean route" (basically from Libya to Italy) has been the most heavily trafficked and the deadliest maritime route in the world [1]. Human traffickers load rubber dinghies and rusty fishing boats with tens to hundreds of people, in the hope that they are rescued by humanitarian or official vessels before they sink. The total number of victims is not known with certainty, but the death toll reported by the Missing Migrants Project of the International Organization for Migration is of the order of several thousands/yr in the past few years [2]. Sadly, not only the number of victims is huge, but also only a small fraction of the bodies may be recovered to allow for identification, which is universally recognized as a fundamental human right [3,4].

A turning point in the general awareness of the dramatic dimension of this ongoing tragedy has been a shipwreck which occurred on October 3rd, 2013. A boat packed with more than 500 migrants capsized and sank in about 40 m of water, half a mile off the shores of Lampedusa, a tiny Italian island south of Sicily [5]. Only 155 survived. It has been reported that most migrants were from Eritrea, Somalia and Ethiopia. One hundred ninety-four bodies were recovered soon after the disaster, and a further 108 were retrieved from the inside of the boat's hull some days later; other corpses were found during the following days. The final count of the dead was 366. Even if the exact number of victims remains unknown, it can be reasonably rounded up to 400. The tragedy shocked the Italian as well as the international public opinion; the media highlighted the news for many days. This triggered the attention of the Italian Authorities towards the need for a specific program dealing with this type of mass disasters. The "Lampedusa's October 3rd shipwreck" was the first major migrant disaster off the Italian coasts for which data on all victims was collected in the same fashion (complete external examination, photographs and DNA sampling) and pooled in the same dataset [6]. External examination, apparent aging and sexing and pictures of all bodies were collected by the Polizia Scientifica before burial, and samples were collected for DNA analysis. Later, the Office of the Commissioner for Missing Persons of the Italian Government approved a protocol aimed at identifying the victims through the search of antemortem material that could be provided by putative living relatives [6, 7].

An article presenting and discussing the general aim of a pilot study conducted on the victims of this disaster and an overview of the results of the anthropological and genetic examinations has been published elsewhere [6]. Here, we describe the methodological and statistical approach leading

to the genetic identification of the victims by kinship analysis, in this open disaster victim identification (DVI) scenario that is a tragedy without an exact number of victims.

In order to address the problem of the genetic and familial composition of the victim sample, we have used a "blind search" approach using the software *FAMILIAS* [8, 9] to spot putative parent-child (PC) pairs based on autosomal STRs genetic profiles. Then, the same approach was applied to the reference persons (RP) paired to the victims to identify putative PC and full-sibling (FS) pairs.

For the statistical evaluation of these large-scale genetic data comparisons, we have defined a posterior probability value for a positive identification, according to the Bayesian approach.

This is the first paper that tries to systematically deal with the genetic identification of African migrants who died in the Mediterranean Sea. The methodological and statistical approach used in this study to achieve genetic identifications will be evaluated for a future application to other similar mass disasters.

2. MATERIALS AND METHODS

2.1. Data collection

2.1.1 STR analysis (European Standard Set of markers, 16 STRs)

Victims - Biological samples from 364 bodies recovered after the shipwreck (saliva, blood, or muscle tissue) were collected depending on the state of preservation of the corpses. DNA profiling of 16 autosomal STRs was performed in duplicate using two commercial kits (Powerplex 17 ESX, Promega or NGMSElect, AppliedBiosystems) by the ISO/IEC 17025 accredited lab of Palermo's Polizia Scientifica as already described elsewhere [6].

Living relatives - Following an international call from 2013 to 2017, 52 alleged relatives of 47 victims belonging to 42 familial groups presumably on this boat were recruited. Biological samples (buccal swabs, hair, saliva, or nails) of these reference persons (RP), after having been collected at the Ministry's quarters in Rome or at the University of Milano, were sent to the Forensic Genetics Laboratory of the University of Pavia. DNA profiling for the same set of 16 autosomal STRs was carried out as described [6].

The DNA profiles of the RP were anonymised, according to [6]. As some reference subjects were relatives to each other and/or were missing more than one relative, we resolved such many-to-many relationships by assigning a unique identifier (Fam x) to each family (Supplementary Table 1). Nine families included more than one RP (Fam 02, 04, 14, 16, 33, 34, 35, 37, 41), whereas two families were missing more than one victim (Fam 30 and 33). The final RP database included 43 subjects from 36 different familial groups typed for 16 loci (Fam01-Fam36 in Supplementary Table 1).

2.1.2 Extended genetic typing

Additional autosomal STRs and lineage markers (Y-STRs or mtDNA) were typed in selected cases, both in the victims and the RP, separately by the two forensic genetics laboratories involved.

Autosomal markers – Five additional autosomal STRs (D7S820, CSF1PO, D13S317, TPOX, and D5S818), included in the AmpFLSTR Identifiler Plus PCR Amplification kit (ThermoFisher Scientific), were amplified following the manufacturer's recommendations, using 0.5-1 ng input DNA amount. The amplified products were run either on the ABI 310 or 3500 genetic analyzers (AppliedBiosystems) in the Pavia or Palermo labs. The electropherograms were analyzed and the alleles called using GeneMapper ver 3.2.1 or Genemapper Id- X software.

Y-chromosome STRs - DNA amounts varying from 0.25 to 1 ng were used to amplify, according to the manufacturers' recommendations, 23-26 Y-STRs contained in the commercial kits *PowerPlex Y23 System* (Promega Corporation) and *Yfiler plus* (Thermofisher). The amplified products were separated through capillary electrophoresis and the alleles called as for the previous autosomal markers. In most of the victim-reference person pairs both subjects were typed with the *PowerPlex Y23* kit while, in two cases, the victim was typed with *PowerPlex Y23* and the corresponding alleged relative with *Yfiler plus*, thus generating a common shared 21 Y-STRs haplotype.

Mitochondrial DNA - The hypervariable regions HVR-I and HVR-II of the mtDNA were amplified separately using primer sets and PCR conditions described by Ginther *et al* [10] in a total volume of 25 µl, through 34 PCR cycles, using 0.25 ng input DNA amount. The amplified products were checked on 2% agarose gel and then purified with NucleoSpin® Gel and PCR Clean-up (Macherey-Nagel GmbH & Co KG, Germany) following the manufacturer's instructions. Sequencing was carried out using the BigDye® Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) and the same primers were employed in the sequencing reactions. The sequenced products were further purified to remove unincorporated dye terminators using Performa® DTR Gel Filtration Cartridges (Edge BioSystems, Gaithersburg, MD, USA), according to the manufacturer's protocol. Sequences were resuspended in 20 µl of formamide and separated by capillary electrophoresis on ABI PRISM® 310 Genetic Analyzer (Applied Biosystems). Data was analyzed using the Software Sequencing Analysis v.5.2 and then compared to the rCRS sequence [11].

2.2 Data analysis

2.2.1 Pairwise sample comparison

In blind search analysis, all individuals of a sample are paired to each other, or, otherwise, all individuals of a sample are paired to all individuals of a second sample. In our settings, the first case

was applied to the victim sample and the second case was applied to the RP paired to the victims. Two statistics were calculated for each pair, based on the Identity By State (IBS) and the Identity By Descent (IBD) approaches, respectively.

The IBS approach is a non-parametric (or "model free") method, where no specific relationship is hypothesized between any two subjects, and pairs of individuals are ranked by the number of loci with 0, 1 or 2 shared alleles (z_0, z_1, and z_2), or by the total number of shared alleles (z_t). Calculations were carried out by a modified version of the spreadsheet AlleleSharingSheet [12]. The IBD approach is the usual LR calculation, where the likelihood of a specified relationship is contrasted with the hypothesis of unrelatedness for any pair of subjects (or pedigree); calculations were performed by the DVI module of FAMILIAS3 [8, 9].

The evidential weight of the observed $z_{(.)}$ and LR values against those expected for unrelated individuals was evaluated by reshuffling the database of the victims. In this procedure, each of the 32 allele arrays of the victim database was randomly permuted across the genetic profiles. This method leaves the allele frequencies unchanged, whereas the individual's genotypes are reshuffled at all loci. Ten such random databases were obtained. Each of these sets can be interpreted as a sample of unrelated subjects obtained from the population represented by the victims. The $z_{(.)}$ and LR calculations were repeated for each set, both for the random subjects paired to each other and for the reference persons paired to the random subjects.

2.2.2 Bayesian approach

In a Bayesian setting, the posterior odds that a given victim (among N victims, a fraction of which may be untyped) is the missing individual claimed by a reference family is computed as the prior odds times the likelihood ratio. We hypothesized N = 400 total victims and set the posterior probability for a positive identification to 0.999 [13-15], assuming a uniform prior. Calculations were carried out in FAMILIAS3 [8, 9]. In this implementation, the probability of identifying a given missing individual of a given family is partitioned among all unidentified persons; specifically, calling PPi the posterior probability and O'i the posterior odds that subject i is the true missing subject of a given family, PPi is calculated with a formula equivalent to PPi = O'i /Σj O'j, where the sum is conducted over all unidentified persons 1, ..., j, …, N. In this way, the sum of the PPj is constrained to be 1.

2.2.3 Population genetic analysis

For LR and posterior probability calculations, the allele frequencies of a large population sample from the Horn of Africa typed for 35 autosomal STRs, including the 16-21 markers characterized

for the genetic screening of the victims, was used, which included estimates of mutation rates [16]. We refer to this sample as the "Dupuy frequencies".

Allele frequencies were also estimated from the victim sample, after removing some first-degree relatives (parent-child and full siblings) that were spotted in the blind search. The sample was analyzed by the exact test of Hardy-Weinberg equilibrium using ARLEQUIN 3 (v. 3.5.1.3) [17], and for population sub-structuring using STRUCTURE [18]; in addition, the Wright's fixation index F (= 1 – H_{obs}/H_{exp}, where H_{obs} and H_{exp} are the observed and expected heterozygosities, respectively) was calculated for each locus. This parameter corresponds to the definition of F_{IT} in subdivided populations [19]. The resulting database was compared with that of Dupuy, and also with other East African population samples typed for a smaller set of markers [20, 21]. The analyses were performed by computing pairwise and total F_{ST} and testing for population differentiation using ARLEQUIN 3.

The Y-chromosome haplotypes were searched in the YHRD database [22, 23], considering different sets of markers (the "minimal" 8-markers and the 21 or 23 markers haplotypes) and different sets of African population samples (metapopulations).

The mtDNA haplotypes typed were searched in the EMPOP database [24, 25] and in its African metapopulations.

3 RESULTS AND DISCUSSION

3.1 Population genetic analyses

Victims sample - The final database included 335 full genetic profiles. ARLEQUIN's exact test of Hardy-Weinberg equilibrium showed two loci with nominal p-values <0.05, but significance was not maintained after the Bonferroni correction. Wright's fixation index (equivalent to F_{IT}) was included between –0.06 (D10S1248, with an excess of heterozygotes) and +0.05 (SE33, with an excess of homozygotes), and the mean overall value was near zero. In addition, the program STRUCTURE did not show any evidence of population admixture; forcing k = 2, all individuals had nearly equal posterior probability of being assigned to either group, and the same occurred with k = 3 and k = 4. Even though no population substructure has been identified in the sample, we cannot exclude the presence of a population admixture so limited that it could not be detected, making this a complicated issue in such context.

The allele frequencies were compared with those reported for other African population samples, including the one studied by Dupuy [16, 20-21]. As shown in Table 1, the F_{ST} values were generally small, both by locus and overall (maximum value: TH01 = 0.0086, average value = 0.0026 ± 0.002), indicating a remarkable similarity of allele frequencies among samples. ARLEQUIN's exact

test of differentiation was significant for some of the pairwise locus comparisons at the nominal P-value of 0.01, but the false-discovery rate was high due to the high number of comparisons (n = 78) and these observations were considered irrelevant. Of interest is the number of alleles that have been observed only once in the four samples (the so-called "private alleles", in brackets in Table 1). Most were observed in single copy; it is worth noting that their number was disproportionately higher for the two markers SE33 and FGA (16 and 14, respectively, almost the number of all other loci combined); the Dupuy population sample showed the largest number of private alleles (36), undoubtedly because of its large sample size.

The similarity of the allele frequencies between the victim sample and other population-based surveys is important, as it represents an independent confirmation that most victims were truly from East Africa, meaning that the Dupuy frequencies are appropriate for kinship analysis.

3.2 Familial groups within the victims

To examine the possibility that the victims sample included some familial groups, all individuals with full genetic profiles (n = 347) were paired to each other and for each pair the number of loci with 0, 1 and 2 identical by state (IBS) alleles (z_0, z_1, and z_2) and the total number of shared alleles (z_t) were computed. Fourteen pairs showed no “exclusions” (at least one allele shared at all loci, z_0 = 0), thus qualifying for being putative parent-child (PC) pairs. The corresponding LR PC/NR (NR stands for non-relatives) were 10^3 to 10^8 and 10^4 to 10^7, using the allele frequencies by Dupuy [16] and those calculated from the victim group, respectively.

The expected number of pairs with z_0 = 0 in a database of unrelated pairs of 347 individuals was estimated by ten random permutations of the victim database (see Data Analysis, section 2.2). We found 7 “PC” matches, leading to a point estimate of the probability of a false PC pair of 1.17×10^{-5} and to an expected number of false PC pairs in the real database of 0.8. Although this number is not negligible, the 14 pairs were investigated in more detail. Four of them were present in more than one pair, namely, one person was connected with three subjects, and three other persons were connected each with two subjects. By considering their estimated age (from medico-legal records), it emerged that these four groups composed by more than one family member were a mother with three children, two mothers with two children and a pair of parents with a child. The other five putative single PC pairs did not show evidence of other first-degree relatives among the victims. In conclusion, nine independent putative familial groups were tentatively identified among the victims through a parent-child relationship. Other putative pairs of first-degree relatives (full siblings) could be recognized by looking at the total distribution of allele sharing and LRs. However, we deferred any further analysis of the victim database until possible living relatives were identified.

3.3 Reference persons and families

As outlined in Parsons [26], we first performed a blind search between victims and reference persons (each RP was contrasted against all victims), disregarding the relationships reported with their missing relatives. This approach can mitigate issues deriving from unexpected pedigree relationships. Only after a victim was recognized as a putative relative of a given reference person was the relationship claimed by the living relative used to confirm the identification in a pedigree analysis.

3.3.1 Blind search analysis

The 43 reference persons were paired to the 347 victims with full genetic profile, and the values of z_0, z_1, and z_2 and z_t were determined for each pair. In parallel, a pairwise matching was performed by the DVI module of FAMILIAS3 (also including the victims with partial profiles), checking for parent-child and full sibling relations.

Eleven pairs with $z_0 = 0$, i.e. with no exclusions, were observed. FAMILIAS recognized them as putative parent-child with LR PC/NR from 10^4 to 10^{10}. These eleven pairs included 8 different victims.

In addition, a reference person (Fam01) showed a single exclusionary locus with a victim (D21S11; reference person 31.2/32.2, victim 29/33.2), for which a germline mutation was inferred. Using the mutation model included in the Dupuy database available in FAMILIAS website (www.familias.no), the resulting LR PC/NR was 5×10^5. In conclusion, nine victims showed parent-child relationships with 12 reference persons and these pairs were checked in pedigree analysis.

For full sibling pairs, no obvious cut-off values of any statistics can be considered, so we examined the distributions of the total number of shared alleles (z_t) and of the $\log_{10}$ of the LR FS/NR in the reference persons paired to the ten replicas of the randomized victim databases. As shown in Table 2, 44 total pairs showed LR FS/NR > 100, two of which with LR > 1000, resulting in 4.4 and 0.2 expected numbers in the real data, respectively. Thus, the chance of false FS with LR > 10,000 was negligible, provided that none of the reference persons were first-degree relatives of any of the victims. Considering the z_t distribution (not shown), sixty-five total pairs (out of 149,210) showed $z_t = 17$, eighteen showed $z_t = 18$ and none showed $z_t > 18$, so that the corresponding expectations were 6.5 ($z_t = 17$), 1.8 ($z_t = 18$), and undetermined, but close to 0 ($z_t > 18$), respectively.

The identified FS pairs that had already been considered as putative parent-child relationships were removed from the results (14,921 pairs – 12 PC = 14,909 pairs). Among these 14,909 pairs, 19 showed LR FS/NR > 10^4 (all with $z_t > 18$) and were selected for further analysis. These pairs

involved 17 reference persons and 17 victims. Eight pairs included three victims that had already been detected by the PC analysis (Fam02, Fam04 and Fam16) and three victims that had been related to two reference persons previously involved in parent-child relationships (Fam33). These victims were the missing relatives of families with multiple reference persons. The other 11 pairs included a single reference person missing a single full sibling and a single case of a missing half-brother (Fam21). These pairs were considered in the subsequent kinship analysis.

Other eight pairs showed LR FS/NR 10^2 to 10^4 (Fam05, Fam06, Fam07, Fam08, Fam23, Fam24, Fam26, Fam30) (see Figure 1). As the chance of false FS pairs is not negligible in this range (Table 2), we decided to extend the genetic typing, including five additional autosomal STRs, and either Y-chromosome STRs or mtDNA variants.

The remaining eight families (Fam03, Fam11, Fam14, Fam17, Fam19, Fam20, Fam27 and Fam36) showed LR values lower than 10^2. In three cases (Fam14, Fam19, Fam27), the reference persons were looking for second degree relatives and are considered in the next paragraph (see 3.3.3). The remaining families involved putative relatives seeking a father (Fam17) and a brother (Fam3, Fam11, Fam20, fam36). In the first case, no matches were identified using both IBS and IBD approaches and the search was considered negative; for the other four, the maximum LR observed in the blind search was considered too low to be investigated with the extended typing approach, although definite exclusions could not be attained.

All the results of PC and FS relations were plotted in Figure 1.

3.3.2 Pedigree analysis

Family pedigrees were built for the 27 reference persons associated to one or more victims in the blind search test, and were paired to the entire victim group using *FAMILIAS*. Pedigree LRs were calculated using the allele frequencies both from the victims database and the East African population studied by Dupuy [16], while posterior probabilities were calculated assuming the number of victims rounded up to 400 and by only using the Dupuy database. The pedigree LRs were calculated, for each reference family, by contrasting the hypotheses that the missing individual was any of the victims, rather than some unknown person, and provided values between 10^4 and 10^{16}. The corresponding posterior probabilities were greater than 99.9% in all cases except for Fam18, whose members were submitted to the extended genetic typing.

Among the found matches, a putative second-degree pair (Fam21, half-brothers sharing a common father) provided a posterior probability greater than 99.9%. In order to confirm the paternal lineage, Y- STR markers were analyzed both in the victim's and the corresponding living relative's DNA (see section 3.3.3.).

In conclusion, pedigree analysis based on 16 autosomal STRs provided posterior probabilities greater than 99.9% for 22 missing relatives of 19 reference families by assigning equal priors to each of an estimated number of 400 victims (see Table 3, section "Pedigree Analysis").

3.3.2.1 Extended genetic typing.

Nine victim-RP pairs showed FS/NR LR lower than 10^4 but greater than 10^2 in the blind search. This is a "gray zone", where both the chances of false-positives and false-negatives are high. Therefore, we extended the genetic typing, including five autosomal STR markers and lineage markers (Y-chromosome STRs or mtDNA). As shown in the bottom part of Table 3 (section "Extended Genetic Typing"), the 21-loci LRs increased by two to five orders of magnitude for six of the nine pairs, leading to posterior probabilities ≥ 99.9% in all cases except three (Fam07, Fam08, and Fam26), which are further described below. The lineage markers confirmed the former six putative relationships.

Each haplotype was searched in the forensic Y-STR or mtDNA databases YHRD [22, 23] and EMPOP [24, 25], respectively. All the haplotypes were unique in the entire databases when the complete set of markers was considered. The number of shared Y-STR and mtDNA haplotypes and the corresponding frequencies in the general database and in specific African metapopulations are reported in Supplementary Tables 2 and 3, respectively.

We did not calculate a combined LR of the autosomal STRs and the lineage markers as suggested in [14-15, 27]. In most cases, in fact, the combined LRs would have been boosted by at least three orders of magnitude, since all the complete observed Y-chromosome and mtDNA haplotypes were unique, even in the general population databases comprising the African metapopulations. Unfortunately, there is still limited high-quality forensic data about lineage markers in African populations, and this may lead to under-estimating the haplotype frequencies. However, we did not ignore the high discriminatory power of extended Y-chromosome and mtDNA haplotypes in the final process of identification of a victim. In fact, the finding of shared haplotype in relative-victim pairs was considered as a strong support to the positive identification of the six victims.

3.3.3 Second-degree relationships

As the autosomal loci do not distinguish second-degree relatives (such as half-sib, avuncular or grandparent-grandchild) in kinship analysis, the persons missing half siblings or nieces/nephews (six missing relatives in total) were grouped. Two reference persons were missing a half-brother (Fam19 and Fam21), two were missing a nephew and a niece, respectively (Fam14 and Fam27), and one (Fam30) was missing a nephew and a niece, his sister's children. The last case was

resolved by the extended genetic typing, which confirmed a full-sibling relationship with a female victim (PP = 0.9996, Family 30, Table 3), who had already been identified as the mother of two other victims; the possibility that the putative niece could have been the actual sibling of the reference person was ruled out, as the PP of this relationship was $< 10^{-10}$. This case perfectly fits the situation described in [26] where the DNA profile of an identified victim was used as a reference sample for other corresponding missing relatives.

In addition, the half-brother paternal relationship previously described for Fam21 was also confirmed by Y-STR markers.

Three reference persons, from Fam14, Fam19 and Fam27 respectively, were missing a single second-degree relative each, and the highest value of the posterior probability obtained by searching all the victims was lower than 50% in all three cases. This ruled out the option to perform additional analyses in these cases.

3.3.4 Multidisciplinary approach to identification

Two familial groups (Fam26 and Fam07) showed posterior probability values very close to 99.9% and were submitted to lineage marker analysis.

Fam26 included a brother looking for his sister. Kinship analysis resulted in LR FS/NR equal to 4.0×10^4, corresponding to a posterior probability of 99.86%. Since mtDNA analysis confirmed the maternal origin, and a biological match was also detected by anthropological and medico-legal investigations, all these findings supported the identification of the victim.

Fam07 was a man looking for his brother. In this case, LR and posterior probability were lower than the selected threshold, corresponding to 3.0×10^3 and 98.54%, respectively. The analysis of Y-chromosome STRs highlighted the same paternal origin, and anthropological and medico-legal investigations revealed a biological match. Therefore, as in the previous case, all these findings supported the identification of the victim.

These results highlight the importance of the support of other primary and secondary identifiers in DVI. As it is well known, a multidisciplinary approach involving anthropology, odontology and genetics can increase the number of the identified victims. This was clearly pointed out in the 2007 ISFG recommendations regarding the role of forensic genetics in disaster victim identification [14] and in the paper reporting the ICMP experience [26].

Finally, the last pair included in the analysis, Fam08, was excluded to be a true full sibling relationship. In fact, in this case, the analysis of the additional autosomal markers led to a decrease of LR FS/NR and posterior probability (from 4.0×10^2 to 6.6×10^1 and from 87.0% to 52.48%, respectively). The results were confirmed by lineage markers which did not highlight a common paternal origin.

4. CONCLUSIONS

In the end, the present work identified 29 first-degree relatives out of 35 (83%) missing by the reference families, and 3 second-degree relatives out of 6 (50%). With reference to the paper by Olivieri et al. [6], eight additional identifications are reported here, which are supported by kinship analysis only; in another case, the genetic match confirmed a previous identification obtained by anthropological investigations.

The main reason for this increase was the typing of additional autosomal markers and haploid systems in the pairs that previously did not reach a posterior probability of 99.9%. This suggests that a 21 autosomal STRs profiling should be considered as a standard screening approach in future DVI analysis of the same kind and that lineage markers could support the identification process. However, more population data are needed, especially for the Sub-Saharan countries where most migrants come from. This is especially true for lineage markers, though even for autosomal STRs it is very difficult to find large published population surveys for the complete 21 marker set.

Another reason of the overall high rate of victim identification is undeniably the great care taken by the interviewers of the putative victim's relatives (a team composed by a trained psychologist, a forensic anthropologist/odontologist, a forensic pathologist and a cultural/linguistic mediator), in such a way that the final recorded information was highly reliable.

The success rate achieved suggested the reliability of the strategy and its applicability in cases of mass disasters. Moreover, even in this challenging context, the results obtained highlighted the beneficial role of DNA analysis in victim identification as already described in other DVI scenario [28]. However, the lack of suitable frequency databases concerning sub Saharan populations, both for autosomal and lineage markers, highlighted the need to fill this gap in order to perform more accurate statistical analyses. In addition, while agreement was reached among recommendations on the role of forensic genetics in DVI events concerning sampling and genetic characterization of victims and reference samples [14, 29, 30], different approaches have been adopted about statistical evaluation, especially concerning the association of lineage markers with autosomal STRs in a combined likelihood ratio value [14, 30-35]. This lack of a shared statistical approach should be filled in view of future identification cases from most catastrophic humanitarian tragedies, such as that which occurred on April 18th 2015 [36], in which about 1,000 migrants coming from different African countries drowned in the Mediterranean Sea.

ACKNOWLEDGMENTS

The authors wish to thank Prof. Thore Egeland and Dr. Daniel Kling (University of Oslo, Norway) for their suggestions, Terre des Hommes and

Isacchi Samaja for their assistance in the project and the staff of the Commissioner for Missing Persons for the support.

REFERENCES

[1] T.G. Hammond, The Mediterranean Migration Crisis, Foreign Policy Journal. 2015.

[2] International Organization for Migration (IOM), World Migration Report, 2018. https://publications.iom.int/system/files/pdf/wmr_2018_en.pdf, 2019 (access 29 April 2019).

[3] European Convention on Human Rights, Article 2, 2007.

[4] UN General Assembly, Human Rights Council, Progress report of the Human Rights Council Advisory Committee on best practices on the issue of missing persons (A/HRC/14/42). 22 March 2010, p. 4.

[5] The new York Times. Migrants Die as Burning Boat Capsizes Off Italy. http://www.nytimes.com/2013/10/04/world/europe/scores-die-in-shipwreck-off-sicily.html?_r=0, 2013 (access 23 March 2019).

[6] L. Olivieri, D. Mazzarelli, B. Bertoglio, D. De Angelis, C. Previderè, P. Grignani, A. Cappella, S. Presciuttini, C. Bertuglia, P. Di Simone, N. Polizzi, A. Iadicicco, V. Piscitelli, C. Cattaneo, Challenges in the identification of dead migrants in the Mediterranean: The case study of the Lampedusa shipwreck of October 3rd 2013, Forensic Sci Int. 285 (2018) 121-128. doi: 10.1016/j.forsciint.2018.01.029.

[7] Protocollo d'Intesa, 2014. Memorandum of Understanding 2014. http://www.interno.gov.it/sites/default/files/allegati/protocollo_persone_scomparse_naufragi_lampedusa_3-11_ottobre_2013.pdf, 2019 (access 22 March 2019).

[8] D. Kling, A.O. Tillmar, T. Egeland, Familias 3 - Extensions and new functionality, Forensic Sci Int Genet. 13 (2014) 121-127. doi: 10.1016/j.fsigen.2014.07.004.

[9] T. Egeland, P. Mostad, B. Mevåg, M. Stenersen, Beyond traditional paternity and identification cases. Selecting the most probable pedigree, Forensic Sci Int. 110 (2000) 47-59.

[10] C. Ginther, L. Issel-Tarver, M.C. King, Identifying individuals by sequencing mitochondrial DNA from teeth, Nat Genet. 2 (1992) 135-138.

[11] R.M. Andrews, I. Kubacka, P.F. Chinnery, R.N. Lightowlers, D.M. Turnbull, N. Howell, Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA, Nat Genet. 23 (1999) 147.

[12] S. Presciuttini, N. Cerri, S. Turrina, B. Pennato, M. Alù, A. Asmundo, A. Barbaro, I. Boschi, L. Buscemi, L. Caenazzo, E. Carnevali, D. De Leo, C. Di Nunno, R. Domenici, M. Maniscalco, G. Peloso, S. Pelotti, A. Piccinini, D. Podini, U. Ricci, C. Robino, L. Saravo, A. Verzeletti, M. Venturi,

A. Tagliabracci, Validation of a large Italian Database of 15 STR loci, Forensic Sci Int. 156 (2006) 266-268.

[13] C.H. Brenner, B.S. Weir, Issues and strategies in the DNA identification of World Trade Center victims, Theor Popul Biol. 63 (2003) 173-178.

[14] M. Prinz, A. Carracedo, W.R. Mayr, N. Morling, T.J. Parsons, A. Sajantila, R. Scheithauer, H. Schmitter, P.M. Schneider, International Society for Forensic Genetics, 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. 1 (2007) 3-12. doi:10.1016/j.fsigen.2006.10.003.

[15] I. Zupanic Pajnic, B. Gornjak Pogorelc, J. Balazic, Molecular genetic identification of skeletal remains from the Second World War Konfin I mass grave in Slovenia, Int J Legal Med. 124 (2010) 307-17. doi:10.1007/s00414-010-0431-y.

[16] B.M. Dupuy, M. Stenersen, D. Kling, Frequency data for 35 autosomal STR markers in a Norwegian, an East African, an East Asian and Middle Asian population and simulation of adequate database size, Forensic Sci.Int Genet Supplement Series. 4 (2013) e378-e379.

[17] L. Excoffier, G. Laval, S. Schneider, Arlequin (version 3.0): an integrated software package for population genetics data analysis, Evol Bioinform Online. 23 (2007) 47-50.

[18] L. Porras-Hurtado, Y. Ruiz, C. Santos, C. Phillips, A. Carracedo, M.V. Lareu, An overview of STRUCTURE: applications, parameter settings, and supporting software, Front Genet. 4 (2013) 98. doi: 10.3389/fgene.2013.00098.

[19] S. Wright, The genetical structure of populations, Ann Eugen. 15 (1951) 323-354.

[20] A.O. Tillmar, G. Bäckström, K. Montelius, Genetic variation of 15 autosomal STR loci in a Somali population, Forensic Sci Int Genet. 4 (2009) e19-e20. doi:10.1016/j.fsigen.2009.01.004.

[21] C. Tomas, H.S. Mogensen, S.L. Friis, C. Hallenberg, M.C. Stene, N. Morling, Concordance study and population frequencies for 16 autosomal STRs analyzed with PowerPlex® ESI 17 and AmpFℓSTR® NGM SElect™ in Somalis, Danes and Greenlanders, Forensic Sci Int Genet. 11 (2014) e18-e21. doi: 10.1016/j.fsigen.2014.04.004.

[22] S. Willuweit, L. Roewer, The new Y Chromosome Haplotype Reference Database, Forensic Sci Int Genet. 15 (2015) 43-48. doi: 10.1016/j.fsigen.2014.11.024.

[23] Y-Chromosome STR Haplotype Reference Database.
http://yhrd.org, 2019 (access 05 April 2019).

[24] W. Parson, A. Dür, EMPOP--a forensic mtDNA database, Forensic Sci Int Genet. 1 (2007) 88-92. doi: 10.1016/j.fsigen.2007.01.018.

[25] EMPOP.
http://empop.online, 2019 (access 29 April 2019).

[26] T.J. Parsons, R.M.L. Huel, Z. Bajunović, A. Rizvić, Large scale DNA identification: The ICMP experience, Forensic Sci Int Genet. 38 (2019) 236-244. doi:10.1016/j.fsigen.2018.11.008.

[27] L. Bradford, J. Heal, J. Anderson, N. Faragher, K. Duval, S. Lalonde, Disaster victim investigation recommendations from two simulated mass disaster scenarios utilized for user acceptance testing CODIS 6.0, Forensic Sci Int Genet. 5 (2011) 291–296.

[28] L.G.Biesecker, J.E. Bailey-Wilson, J. Ballantyne, H. Baum, F.R. Bieber, C. Brenner, B. Budowle, J.M. Butler, G. Carmody, P.M. Conneally, B. Duceman, A. Eisenberg, L. Forman, K.K. Kidd, B. Leclair, S. Niezgoda, T.J. Parsons, E. Pugh, R. Shaler, S.T. Sherry, A. Sozer, A. Walsh, Epidemiology. DNA identifications after the 9/11 World Trade Center attack, Science. 310 (2005) 1122-1123.

[29] B. Budowle, F.R. Bieber, A.J. Eisenberg, Forensic aspects of mass disasters: strategic considerations for DNA-based human identification, Leg Med (Tokyo). 7 (2005) 230-243.

[30] C.M. Vullo, M. Romero, L. Catelli, M. Šakić, V.G. Saragoni, M.J. Jimenez Pleguezuelos, C. Romanini, M.J. Anjos Porto, J. Puente Prieto, A. Bofarull Castro, A. Hernandez, M.J. Farfán, V. Prieto, D. Alvarez, G. Penacino, S. Zabalza, A. Hernández Bolaños, I. Miguel Manterola, L. Prieto, T. Parsons, GHEP-ISFG collaborative simulated exercise for DVI/MPI: Lessons learned about large-scale profile database comparisons, Forensic Sci Int Genet. 21 (2016) 45-53. doi: 10.1016/j.fsigen.2015.11.004.

[31] A. Amorim, A cautionary note on the evaluation of genetic evidence from uniparentally transmitted markers, Forensic Sci Int Genet. 2 (2008) 376-378. doi: 10.1016/j.fsigen.2008.04.001.

[32] L. Bradford, J. Heal, J. Anderson, N. Faragher, K. Duval, S. Lalonde, Disaster victim investigation recommendations from two simulated mass disaster scenarios utilized for user acceptance testing CODIS 6.0, Forensic Sci Int Genet. 5 (2011) 291-296. doi: 10.1016/j.fsigen.2010.05.005.

[33] J. Buckleton, S. Myers, Combining autosomal and Y chromosome match probabilities using coalescent theory, Forensic Sci Int Genet. 11 (2014) 52-55. doi: 10.1016/j.fsigen.2014.02.009.

[34] Scientific Working Group on DNA Analysis Methods (SWGDAM), Interpretation Guidelines for Mitochondrial DNA Analysis by Forensic DNA Testing Laboratories, 2013, pp. 1–23.

http://media.wix.com/ugd/4344b0_c5e20877c02f403c9ba16770e8d41937.pdf

[35] J.S. Buckleton, M. Krawczak, B.S. Weir, The interpretation of lineage markers in forensic DNA testing, Forensic Sci Int Genet. 5 (2011) 78-83. doi: 10.1016/j.fsigen.2011.01.010.

[36] The Guardian. UN says 800 migrants dead in boat disaster as Italy launches rescue of two more vessels.
https://www.theguardian.com/world/2015/apr/20/italy-pm-matteo-renzi-migrant-shipwreck-crisis-srebrenica-massacre, 2015 (access 05 August 2019).

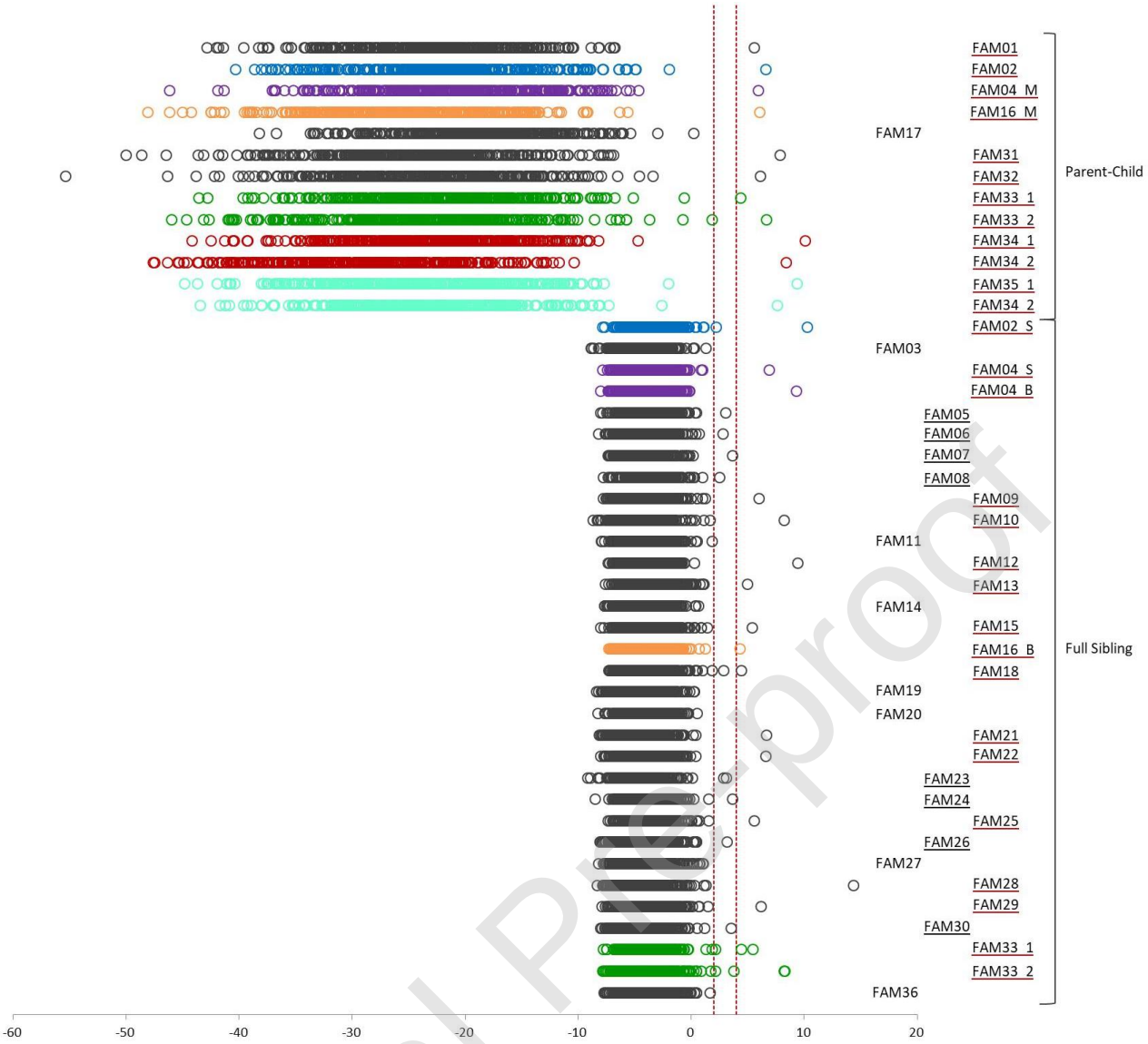


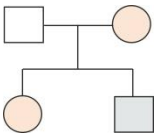
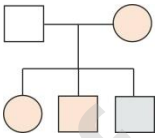
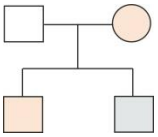
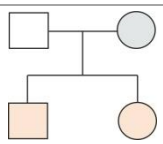
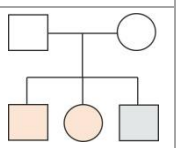
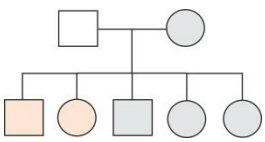
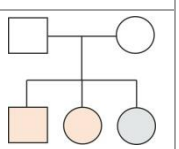
Figure 1. **Plot of the $\log_{10}$(LR PC and FS/NR) representing the blind search analysis performed between all the victims against all the reference persons.** In the upper and lower part of the graph, the results obtained for parent-child and full sibling searching using the Dupuy frequencies are shown. For parent-child relationships, pairs with no exclusions or LR greater than 10^4 are reported, including, moreover, the only RP missing a father, which provided low LR values matches (LR < 2) in the analysis (Fam17). The first family at the top (Fam01) is the PC relation with the germline mutation. Reference persons belonging to the same family are marked with the same color. Families showing LR values greater than 10^4 are reported on the right (underlined in red); families providing LR values between 10^2 and 10^4 are reported in the center (underlined in black); families showing LR values lower than 10^2 are reported on the left (B: brother, M: mother, S: sister). The red lines correspond to the 10^2 and 10^4 LR values.

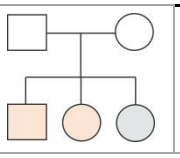
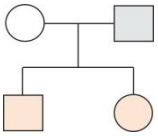
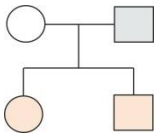
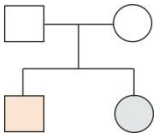
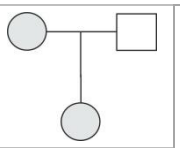
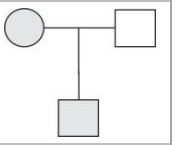
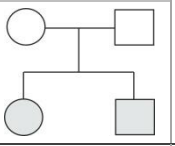
	This study (335)		Dupuy (799)		Tomas (198)		Tilmar (404)		
Marker	Na (Np)	H_{exp}	Na (Np)	H_{exp}	Na (Np)	H_{exp}	Na (Np)	H_{exp}	F_{ST}
D3S1358	9 (0)	0.760	14 (4)	0.753	8 (0)	0.754	6 (0)	0.737	0.0007
VWA	10 (0)	0.804	12 (2)	0.819	9 (0)	0.802	9 (0)	0.800	0.0024
D16S539	7 (0)	0.794	10 (2)	0.801	8 (0)	0.802	8 (0)	0.782	0.0017
D2S1338	13 (1)	0.877	14 (1)	0.872	12 (0)	0.854	12 (0)	0.858	0.0043
D8S1179	11 (0)	0.822	12 (0)	0.806	10 (0)	0.777	11 (0)	0.790	0.0033
D21S11	16 (0)	0.826	24 (4)	0.852	17 (1)	0.836	18 (0)	0.826	0.0045
D18S51	20 (0)	0.907	27 (4)	0.902	20 (1)	0.902	21 (0)	0.902	0.0033
D19S433	14 (1)	0.841	12 (0)	0.833	11 (0)	0.807	13 (0)	0.818	0.0032
TH01	7 (1)	0.770	8 (2)	0.752	6 (0)	0.745	6 (0)	0.716	0.0086
FGA	23 (6)	0.874	31 (8)	0.868	16 (0)	0.853	19 (1)	0.860	0.0004
D12S391	16 (0)	0.865	17 (1)	0.863	15 (0)	0.836			0.0014
D1S1656	16 (2)	0.869	14 (0)	0.872	14 (1)	0.866			0.0022
D2S441	11 (0)	0.787	11 (0)	0.796	8 (0)	0.790			0.0011
D10S1248	10 (0)	0.790	9 (0)	0.750	9 (0)	0.791			0.0010
D22S1045	9 (0)	0.716	10 (1)	0.790	8 (0)	0.724			0.0012
SE33	45 (8)	0.940	46 (7)	0.935	34 (1)	0.930			0.0030
Average	14.8	0.828	16.9	0.829	12.8	0.817	12.3	0.809	0.0026
SD	9.2	0.059	10.2	0.055	6.9	0.056	5.4	0.057	0.0020

Table 1. **Population data of four samples from East Africa.** Numbers right of sample names are the number of typed individuals. Na: Number of alleles; Np: number of private alleles; Hexp: expected heterozygosity.

x	Observed number with log10(LR) > x	Estimate of P(log10 (LR) > x)	Expected number with log10(LR) > x
0	1,363	0.0091	136.3
0.5	630	0.0042	63.0
1	259	0.0017	25.9
1.5	109	0.0007	10.9
2	44	3.0E-04	4.4
2.5	14	9.4E-05	1.4
3	2	1.34E-05	0.2
3.5	0	0	0.0
4	0	0	0.0

Table 2. **Distribution of the $\log_{10}$(LR FS/NR) in the randomized and the real data.** Observed number (2nd column) of pairs with log10 (LR FS/NR) larger than the given x value (1st column) in the 43 reference persons paired to the ten replicas of the randomized database (149,210 total pairs), and the corresponding expected number in the real data (14,921 pairs), in the hypothesis that none of the reference persons were relatives of any of the victims.

Pedigree analysis						
Fam	**Reference samples**	**Missing person**	**LR (Victims)**	**LR (Dupuy)**	**PP (Dupuy)**	**Pedigree**
Fam01	RP 1 (father)	Son	2.2×10^{5}	7.6×10^{5}	99.99%	
Fam02	RP 2 (mother) RP 3 (sister)	Son/Brother	3.8×10^{13}	1.3×10^{14}	> 99.99%	
Fam04	RP 5 (mother) RP 6 (sister) RP 7 (brother)	Son/Brother	5.4×10^{14}	2.7×10^{15}	> 99.99%	
Fam09	RP 12 (brother)	Brother	1.4×10^{6}	1.2×10^{6}	99.99%	
Fam10	RP 13 (sister)	Sister	4.8×10^{7}	2.0×10^{8}	> 99.99%	
Fam12	RP 15 (sister)	Brother	2.3×10^{8}	3.1×10^{9}	> 99.99%	
Fam13	RP 16 (sister)	Brother	1.0×10^{5}	1.2×10^{5}	99.90%	
Fam15	RP 18 (sister)	Sister	1.1×10^{5}	2.8×10^{5}	99.96%	
Fam16	RP 19 (mother) RP 20 (brother)	Son/Brother	1.3×10^{9}	1.3×10^{9}	> 99.99%	
Fam18	RP 22 (sister)	Brother	1.0×10^{4}	3.3×10^{4}	96.90%	
Fam21	RP 25 (half-brother)	Half-brother	3.5×10^{4}	7.7×10^{5}	99.94%	
Fam22	RP 26 (sister)	Brother	7.0×10^{6}	4.5×10^{6}	> 99.99%	
Fam25	RP 29 (brother)	Brother	2.0×10^{5}	4.5×10^{5}	99.97%	
Fam28	RP 32 (brother)	Brother	4.5×10^{13}	2.7×10^{14}	> 99.99%	
Fam29	RP 33 (brother)	Brother	2.9×10^{5}	1.9×10^{6}	99.99%	
Fam31	RP 35 (son)	Mother	7.6×10^{7}	9.3×10^{7}	> 99.99%	
Fam32	RP 36 (son)	Father	9.2×10^{5}	1.5×10^{6}	99.99%	
Fam33	RP 37 (son) RP 38 (daughter)	Mother	4.4×10^{11}	3.5×10^{11}	> 99.99%	
	RP 37 (brother) RP 38 (sister)	Brother	3.0×10^{10}	1.9×10^{10}	> 99.99% *	 
	RP 37 (brother) RP 38 (sister)	Sister	3.7×10^{13}	4.3×10^{14}	> 99.99% *	

	RP 37 (brother) RP 38 (sister)	Sister	1.8 x 10^{12}	1.9 x 10^{12}	> 99.99% *	
Fam34	RP 39 (son) RP 40 (daughter)	Father	2.1 x 10^{15}	5.3 x 10^{16}	> 99.99%	
Fam35	RP 41 (daughter) RP 42 (son)	Father	2.4 x 10^{12}	7.0 x 10^{12}	> 99.99%	
Extended genetic typing						
Fam	**Reference samples**	**Missing person**	**LR (PP) 16 STR**	**LR (PP) 21 STR**	**Lineage**	**Pedigree**
Fam05	RP 8 (brother)	Brother	1.4 x 10^{3} (96.52%)	8.2 x 10^{5} (99.99%)	21/21 Y-STR	
Fam06	RP 9 (brother)	Brother	7.9 x 10^{2} (93.57%)	1.9 x 10^{7} (> 99.99%)	23/23 Y-STR	
Fam07	RP 10 (brother)	Brother	5.6 x 10^{3} (99.20%)	3.0 x 10^{3} (98.54%)	21/21 Y-STR	
Fam18	RP 22 (sister)	Brother	3.3 x 10^{4} (96.90%)	1.2 x 10^{6} (99.91%)	HVR1/2	
Fam23	RP 27 (brother)	Sister	1.5 x 10^{3} (62.50%)	1.4 x 10^{6} (99.93%)	HVR1/2	
Fam24	RP 28 (brother)	Brother	5.0 x 10^{3} (98.29%)	2.3 x 10^{5} (99.96%)	23/23 Y-STR	
Fam26	RP 30 (brother)	Sister	1.8 x 10^{3} (97.02%)	4.0 x 10^{4} (99.86%)	HVR1/2	
Fam30	RP 34 (brother)	Sister	4.1 x 10^{3} (98.51%)	1.9 x 10^{5} (99.96%)	HVR1/2	
	---	Mother Daughter	1.4 x 10^{6} (99.99%) *	9.6 x 10^{5} (99.99%) *	---	
	---	Mother Son	1.6 x 10^{6} (99.99%) *	1.9 x 10^{6} (99.99%) *	---	
	---	Sister Brother	5.1 x 10^{6}	1.0 x 10^{9}	---	
Fam21	RP 25 (half-brother)	Half-brother	7.7 x 10^{5} (99.94%)	---	23/23 Y-STR	

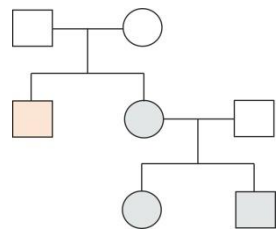


Table 3. **Results of the pedigree analysis and the extended genetic typing.** Top section: LRs and posterior probabilities (PP) for the 20 families whose reference persons were identified as putative parent/child or full sibling by

blind search analysis. Computations were performed using both victim and Dupuy databases. As the results obtained with the two datasets were very similar, only the posterior probabilities calculated with the Dupuy database are reported. The parental relationship is reported for each reference person and missing relative between round brackets. In the last column a graphic representation of the pedigree for the complex families is shown (colors identify the following persons: orange, reference persons; grey, victims; white, relatives not available).

Bottom section: likelihood ratios, posterior probabilities (PP) and Y/mitochondrial results for the ten families submitted to the extended genetic typing. For each victim-reference person pairs, LR and PP values calculated using the Dupuy frequencies for 16 and 21 markers are reported. In the "lineage" column, the number of shared Y-STR alleles between victim and paternal relative is shown. For mitochondrial DNA analysis, HVR1/2 represents the same shared mitochondrial haplotype between pairs. In the last column a graphic representation of the complex pedigree of Fam 30 is shown. An asterisk in PP columns means that the posterior probability is calculated for that specific victim, after removing from the analysis, in turn, the victims that were strong candidates to be the other missing relatives of that family. Reference persons were anonymized.