Forensic Science International: Genetics 53 (2021) 102527

Contents lists available at ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsigen

Research paper

Second GHEP-ISFG exercise for DVI: "DNA-led" victims' identification in a simulated air crash

Carlos M. Vullo [a,*], Laura Catelli [a], Adriana A. Ibarra Rodriguez [b], Aikaterini Papaioannou [c], J. Carlos Álvarez Merino [d], AM Lopez-Parra [e], Aníbal Gaviria [f], Carlos Baeza-Richer [e], Carola Romanini [a], Esperanza González-Moya [g], Ferran Casals [h], Francesc Calafell [h], Gabriela Berardi [i], Gian Carlo Iannacone [j], Gloria C. Vicuña Giraldo [k], Gulbanu K. Zorba [c], Ilaria Boschi [l], Jane Valdivia Olarte [j], Juan E. Ruiz Gomez [k], Juan Pablo Acierno [a], Manuel López Soto [g], Manuel Velázquez Miranda [m], Marco D. García King [n], Maria Alessandra Marrucci [l], Maria J. Porto [o], Mariana Herrera Piñero [p], Mercedes Aler [m], Mishel M. Stephenson Ojea [n], Santiago Cobos Navarrete [f], Ulises Toscanini [i], Victor G. Saragoni [q], Walter Bozzo [p], Yeny C. Posada Posada [b], Zlatan Bajunovic [t], Lourdes Prieto Solla [r,s], Thomas Parsons [t]

[a] *Argentine Forensic Anthropology Team (EAAF), Córdoba, Argentina*
[b] *Laboratorio IdentiGEN - Universidad de Antioquia, Medellín, Colombia*
[c] *Committee on Missing Persons Cyprus (CMP) Anthropological Laboratory, Nicosia, Cyprus*
[d] *Laboratorio de Identificación Genética, Facultad de Medicina, Universidad de Granada, Granada, Spain*
[e] *Laboratory of Forensic and Population Genetics, Dept of Toxicology and Health Legislation, Madrid, Spain*
[f] *Lab de Genética Molecular Cruz Vital - Cruz Roja Ecuatoriana, Quito, Ecuador*
[g] *Instituto Nacional de Toxicología y Ciencias Forenses (INTCF), Madrid, Spain*
[h] *Institut de Biologia Evolutiva (CSIC-UPF), Barcelona, Spain*
[i] *PRICAI - Fundación Favaloro, CABA, Argentina*
[j] *Laboratorio De Biología Molecular y Genética Del Instituto De Medicina Legal, Lima, Peru*
[k] *Grupo Genética Forense Dirección Regional Bogotá, Bogotá, Colombia*
[l] *Istituto di Sanità Pubblica-Medicina Legale Policlinico Gemelli, Roma, Italy*
[m] *Instituto de Medicina Legal y Ciencias Forenses, Valencia, Spain*
[n] *Fundación de Antropología Forense de Guatemala (FAFG), Guatemala*
[o] *Instituto Nacional de Medicina Legal e Ciências Forenses, Coimbra, Portugal*
[p] *Banco Nacional De Datos Genéticos, CABA, Argentina*
[q] *Unidad de Genética Forense, Servicio Médico Legal, Santiago de Chile, Chile*
[r] *Grupo de Medicina Xenómica, Instituto de Ciencias Forenses, USC, Santiago de Compostela, Spain*
[s] *Comisaría General de Policía Científica. Madrid*
[t] *International Commission for Missing Persons (ICMP), USA*

ARTICLE INFO

ABSTRACT

* Corresponding author.
E-mail addresses: cvullo@yahoo.com.ar (C.M. Vullo), marialauracatelli@gmail.com (L. Catelli), adriana.ibarra@udea.edu.co (A.A. Ibarra Rodriguez), kate.papaioannou@gmail.com (A. Papaioannou), juanca@ugr.es (J.C.Á. Merino), amlopezparra@med.ucm.es (A. Lopez-Parra), anibalgaviria@hotmail.com (A. Gaviria), cbaezaricher@med.ucm.es (C. Baeza-Richer), carolacorreo@hotmail.com (C. Romanini), esperanza.gonzalez@justicia.es (E. González-Moya), ferran.casals@upf.edu (F. Casals), francesc.calafell@upf.edu (F. Calafell), gberardi@ffavaloro.org (G. Berardi), ggiannacone@yahoo.com (G.C. Iannacone), gloria.vicuna@medicinalegal.gov.co (G.C. Vicuña Giraldo), gulbanu83@yahoo.com (G.K. Zorba), ilaria.boschi@policlinicogemelli.it (I. Boschi), janevaldiviaolarte@gmail.com (J.V. Olarte), jruiz@medicinalegal.gov.co (J.E. Ruiz Gomez), juanpabloacierno@gmail.com (J.P. Acierno), manuel.lopez@justicia.es (M.L. Soto), correotmvm58@gmail.com (M.V. Miranda), marcodgk@gmail.com (M.D. García King), mariaalessandra.marrucci@unicatt.it (M.A. Marrucci), m.joao.porto@inmlcf.mj.pt (M.J. Porto), mherrerapinero@gmail.com (M.H. Piñero), mercealer@gmail.com (M. Aler), mishelmstephenson@gmail.com (M.M. Stephenson Ojea), saint.jamesc@gmail.com (S.C. Navarrete), utoscanini@pricai.com.ar (U. Toscanini), victor.saragoni@gmail.com (V.G. Saragoni), wbozzo@mincyt.gob.ar (W. Bozzo), yeny.posada@udea.edu.co (Y.C. Posada Posada), Zlatan.Bajunovic@icmp.int (Z. Bajunovic), lourditasmt@gmail.com (L.P. Solla), Thomas.Parsons@icmp.int (T. Parsons).

https://doi.org/10.1016/j.fsigen.2021.102527
Received 4 January 2021; Received in revised form 22 April 2021; Accepted 30 April 2021
Available online 19 May 2021

Keywords:
Disaster victim identification
MPI
Database comparison
DVI
Missing persons identification

The Spanish and Portuguese-Speaking Working Group of the International Society for Forensic Genetics (GHEP-ISFG) has organized a second collaborative exercise on a simulated case of Disaster Victim Identification (DVI), with the participation of eighteen laboratories. The exercise focused on the analysis of a simulated plane crash case of medium-size resulting in 66 victims with varying degrees of fragmentation of the bodies (with commingled remains). As an additional difficulty, this second exercise included 21 related victims belonging to 6 families among the 66 missings to be identified. A total number of 228 post-mortem samples were represented with aSTR and mtDNA profiles, with a proportion of partial aSTR profiles simulating charred remains. To perform the exercise, participants were provided with aSTR and mtDNA data of 51 reference pedigrees —some of which deficient—including 128 donors for identification purposes. The exercise consisted firstly in the comparison of the post-mortem genetic profiles in order to re-associate fragmented remains to the same individual and secondly in the identification of the re-associated remains by comparing aSTR and mtDNA profiles with reference pedigrees using pre-established thresholds to report a positive identification. Regarding the results of the post-mortem samples re-associations, only a small number of discrepancies among participants were detected, all of which were from just a few labs. However, in the identification process by kinship analysis with family references, there were more discrepancies in comparison to the correct results. The identification results of single victims yielded fewer problems than the identification of multiple related victims within the same family groups. Several reasons for the discrepant results were detected: a) the identity/non-identity hypotheses were sometimes wrongly expressed in the likelihood ratio calculations, b) some laboratories failed to use all family references to report the DNA match, c) In families with several related victims, some laboratories firstly identified some victims and then unnecessarily used their genetic information to identify the remaining victims within the family, d) some laboratories did not correctly use "prior odds" values for the Bayesian treatment of the episode for both post-mortem/post-mortem re-associations as well as the ante-mortem/post-mortem comparisons to evaluate the probability of identity. For some of the above reasons, certain laboratories failed to identify some victims. This simulated "DNA-led" identification exercise may help forensic genetic laboratories to gain experience and expertize for DVI or MPI in using genetic data and comparing their own results with the ones in this collaborative exercise.

1. Introduction

The identification of missing persons in large-scale events such as disasters (DVI) or, for example, mass graves from past armed conflicts, is a challenge for forensic services because of the complexity that the context may present, which can profoundly influence the difficulty for the correct identification of the victims [1]. Several best practice and procedure recommendations have been published regarding DVI [1,2] and missing persons identification (MPI) [3,4] investigations. The ISFG has also issued recommendations for forensic genetic laboratories in order to help them to deal with the identification process in the context of large number of victims [5].

Several factors can influence the complexity of the identification process in DVI or MPI scenarios [1]. One of them is the number of missing persons —a large number of victims has deep influence on the Bayesian framework of the context and, therefore, on the capability to correctly identify the victims. In addition, the degree of the disarticulation may influence the number of DNA tests required for DNA-based re-association of post-mortem samples. DNA degradation can influence the quality of the retrieved genetic information from the human remains. Autosomal STR markers (aSTRs) are powerful systems to build genetic profile databases for DVI or MPI due to their high discrimination and individualization power [1–4], however partial genetic profiles with extensive allele or locus dropout due to DNA degradation can considerably reduce the power of discrimination of these DNA profiles, which may be even more problematic if multiple close relatives are not available to profile reference samples.

The existence of several related victims among the missing is another factor that can considerably complicate the DNA identification process [6,7]. Regarding reference samples for genetic comparisons, the best DNA sources are victim's ante-mortem biological specimens due to the high power of identification through direct genetic comparisons [6]. Nevertheless, sometimes there is some uncertainty about the real origin of profiles recovered from personal belongings of missing persons. In addition, it may not be possible to obtain ante-mortem biological material from victims as, for example, in cases of mass graves from human rights violations that are investigated long after the event. In such cases, samples from victims' relatives become the appropriate source of genetic in order to carry out the identification process. Hence, the quality of family pedigrees may deeply influence the success of the identifications: deficient pedigrees made up of few first-degree or only second/third-degree relatives may diminish to a great extent the power of identification [7] and may even prevent victim identification by producing weak evidence that is difficult to distinguish from adventitious matches to unrelated individuals [6,7]. In the latter case, the analysis and comparison of lineage markers such as Y-STRs or mtDNA may prove useful to guide the identification of victims having only distant relatives. However, the informativeness of lineage markers is limited when there are related victims belonging to the same lineage. In kinship analyses, inconsistencies in reported family relationships (for example, incidental findings of non-paternity) may hinder the identification of the victim, imposing the need to re-analyze the family pedigree under different hypotheses of relatedness.

A proper Bayesian approach to large-scale identifications is based on likelihood ratios coming from DNA comparisons involving paired hypotheses (typically, but not always, the hypothesis of related versus unrelated), multiplied by the prior odds for an identification (typically the inverse of the number of missing persons in an event). A DVI episode may be classified as a "closed" event, for example an air crash in which the number and identity of the missing persons are known, making it a simple matter to define the prior odds as the inverse of the known number of missing persons. In closed events, prior odds can be refined further by considering other contextual or non-DNA evidence, such as age, sex, location, etc. Within other contexts of MPI, for example post-conflict mass graves or enforced disappearances, disappearances may accumulate over time and in different places, and there may be less ante-mortem non-genetic information available, and the event may be "open", without a well-defined number of missing persons. In such open events, defining appropriate prior odds can be more complex, requiring some form of reasoned and operational prior probability to be established considering the context. Furthermore, forensic teams can establish minimum statistical thresholds to consider an identification as reliable, depending on the context of the event [5].

The GHEP-ISFG has previously carried out a simulated MPI collaborative exercise requiring participants to perform bone re-associations and identification of missing persons in a secondary common grave with commingled remains; 11 laboratories participated and there were several lessons learned [8]. In keeping with its interest in collaborative

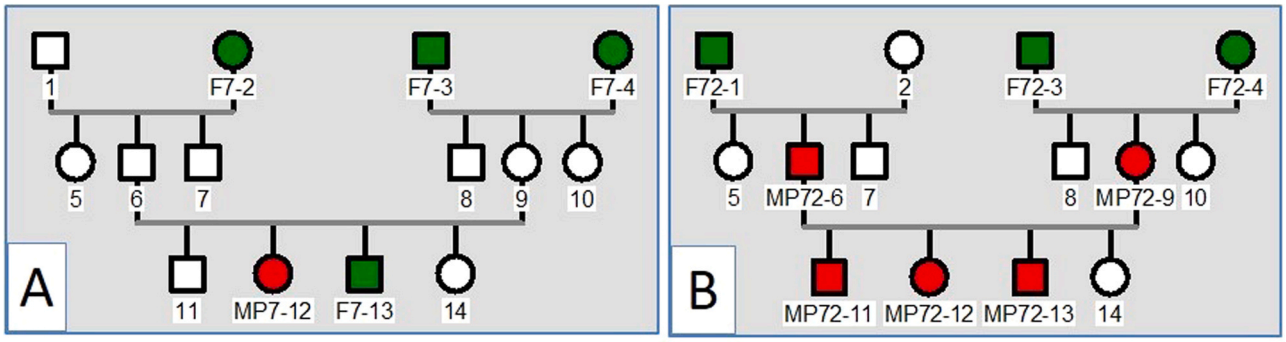

Fig. 1. Genogram representing Families7 and 72. (A) Family7, the missing person shown in red is described as a combination between the family number (F7) and the location of the victim within the family genogram (12), being defined as MP7–12. Similar criterion is used for the references colored in green: F7 is used to indicate Family7 and separated by a hyphen the reference location in the genogram, i.e.: F7–2 (victiḿs paternal grandmother), F7–3 (victim's maternal grandfather), F7–4 (victim's maternal grandmother) and F7–13 (victiḿs full sibling). (B) For Familiy72, a mother (MP72–9), a father (MP72–6) and their three children (MP72–11, MP72–12 and MP72–13) are the victims (red). The reference samples for this case (green) are: F72–3 and F72–4 who are parents of MP72–9 and maternal grandparents of MP72–11, MP72–12 and MP72–13. Reference sample F72–1 is the father of MP72–6 and, at the same time, paternal grandfather of MP72–11, MP72–12 and MP72–13, who are full siblings. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.).

exercises for human identification under DVI or MPI contexts, the GHEP-ISFG organized and documented a second simulated DNA-led exercise within the context of a medium-scale disaster, the results of which are reported here. This simulated scenario was more complex than in the first exercise, including complexity factors such as: fragmented remains requiring re-association through direct post-mortem/post mortem (pm-pm) comparison, partial aSTR profiles due to degraded DNA, ante-mortem/post-mortem (am-pm) comparisons using family references with diverse pedigrees, related victims belonging to the same families, family inconsistencies attributable to mutations, DNA match values below the established statistical threshold, and the requirement to consider mtDNA information to solve matches below threshold.

The goal of this second simulated "DNA-led" collaborative GHEP-ISFG DVI exercise is to continue learning and gaining experience in the comparison of genetic profiles in DVI / MPI contexts. In addition, this exercise may contribute to laboratories interested in preparing themselves in DVI or MPI, permitting a comparison of their results with the consensus obtained by the participating laboratories and highlighting the possible errors that can be made.

2. Methods

2.1. Simulation of family pedigrees, post-mortem (pm) and ante-mortem (am) genetic profiles

2.1.1. Autosomal STR (aSTR) profiles

Genetic profiles for 18 aSTR markers were simulated according to allele frequencies used by GHEP-ISFG in its annual inter-laboratory comparison exercises [9], using the Familias software [10]. To this end, a large family pedigree of 14 individuals including three generations was designed (Fig. 1A and B) and 1000 simulations were performed to obtain the aSTR genotypes for each individual within the pedigree. Once the aSTR profiles in each complete pedigree chart were simulated, a custom Python script was created to process the sample data set. The pseudo-random selection of victims profiles, number of samples, replicates and genetic dropout were achieved using the Random Python library (https://docs.python.org/3/library/random.html). Every family group member was assigned a family code consisting in a number of the family, and separated by a hyphen, a number indicating the exact location into the large pedigree. The location of the missing persons (MP) and references in each simulated pedigree is described further in the *exercise design* (below). A first round of missing person selection was performed per family group: in this first step only one MP was selected per family group. In a subsequent dataset processing, several MPs were added to ensure representation of different types of cases. These MPs were represented by pm profiles. The creation of pm profile replicates to represent fragmented remains and randomization of dropouts was generated as follow: two to 5 replicates were created for each victim profile (pm profiles). Subsequently, each replicate was processed to randomly introduce locus dropouts: firstly, a cut-off of 35% was set, meaning that only 35 out of 100 pm profiles would present dropouts. Then, a number of dropouts ranging from 1 to 8 were randomly generated without considering the STR amplicon size. Finally, family groups with extensive reference family pedigrees as well as groups with deficient pedigrees that is, few individual references within the family, were selected.

2.1.2. mtDNA haplotypes

The mtDNA Control Region (CR) haplotypes used were taken from anonymous data presented in several publications [11–14] to be used as additional maternal lineage genetic information. MtDNA haplotypes were replicated in the pm samples and family references taking into account the maternal relationship in each case.

2. Exercise design

2.1. Scenario description for participants

The present exercise simulates an air crash with victims whose remains are fragmented. The flight list describes the presence of 66 victims including passengers and crew and defines the disaster as a “closed” case. As post-mortem samples (pm) presented various degrees of preservation (some of them were burnt), many of them yielded partial profiles. The agency in charge of coordinating forensic tasks and contacting the victims' relatives gathered biological samples of 128 family references. Family references consist of diverse genealogies which may include first- and second-generation relatives to the victims.

2.2. Post-mortem (pm) samples and profiles

The 228 pm samples were genotyped for different STR autosomal markers as well as for the mtDNA Control Region (CR). The 228 pm aSTR profiles are shown in Table S1 – postmortem aSTR profiles. It was assumed that the aSTR profiles obtained from the pm samples were certain; the participants should not consider the possibility of allele dropout in those loci showing a homozygous genotype, even in pm samples with partial STR profiles. The coordinators of this exercise

asked the participants to use the same criteria for Amelogenin profiles considering an XX result a true female and an XY result a true male without the possibility of dropout. Amelogenin is useful to individualize siblings of different sex. Table S2 – postmortem mtDNA haplotypes shows the mtDNA CR haplotypes from the 228 pm samples.

2.3. Reference (am) samples and profiles

Reference samples from 128 family donors belonging to 51 different nuclear families were collected to identify the 66 victims, as there are some families with several MPs involved.

The aSTR and mtDNA profiles obtained for each of the 128 reference samples are shown in Table S3 – am reference aSTR profiles and Table S4 – am reference mtDNA haplotypes, respectively. The degree of kinship between each relative and the MP is specified in a pedigree chart. The MPs have a mixed code which identifies the family and, separated by a hyphen, a number that represents its location in the pedigree chart. Fig. 1A shows the pedigree chart of Family 7. The missing person is represented as MP7-12 (in red) and family references as F7-2, F7-3, F7-4 and F7-13 (in green).

Those cases presenting families looking for several MPs are described as in the case of Family F72 (Fig. 1B). According to the figure, the codes for the missing persons are MP72-6, MP72-9, MP72-11, MP72-12 and MP72-13 (in red) and the ones for references are F72-1, F72-3 and F72-4 (in green). Family pedigree charts represent the pedigree informed by the victim's relatives.

2.4. aSTR and mtDNA frequencies for statistical calculations

The participants were asked to perform statistical calculations by using the aSTR frequencies in Table S5 - GHEP aSTR allele frequencies, considering a mutation rate of 0.001 for all aSTR markers.

In order to make statistical evaluation of mtDNA matches, the participants were asked to conduct a search in EMPOP[15] in order to estimate the haplotype frequencies. Specific search criteria were established to simplify reporting of statistical values in mtDNA matches. (Detailed in Suppl 1-GHEP-ISFG DVI#2_description).

3. Exercise aims

3.1. Aim 1 - pm-pm comparisons: re-association of profiles

A direct comparison of genetic profiles from pm samples was requested in order to re-associate pm samples to an individual.

A likelihood ratio (LR) value equal to or greater than 1.0E + 07 (LR $\geq$ 10,000,000) was defined as a reliable re-association threshold among pm profiles but laboratories were instructed to not report LR values for pm-pm re-associations but just to group them.

Once the pm samples/profiles were re-associated, participants were requested to group pm samples indicating the code of each fragment in the corresponding columns according to specific instructions and examples given in the Excel file Suppl 2-GHEP-ISFG DVI#2_Results_for participants.

Prior odds values were not specified for re-association, as this criterion was included in a special questionnaire designed to evaluate the statistical treatment of the results within the Bayesian framework of the disaster.

The section explaining how the results were to be reported is described in detail in Suppl 1-GHEP-ISFG DVI#2 description_english, and the datasheet to report the results is shown in Suppl 2_GHEP-ISFG DVI#2 Results_for participants.

3.2. Aim 2 - comparison of pm-am reference samples

Once the pm profiles were re-associated, participants were asked to compare the most complete pm profile of the re-associated group with the family reference profiles by kinship analysis. The selection of the most complete pm profile to be used in the kinship analysis was described in detail in Suppl 1-GHEP-ISFG DVI#2 description_english, where examples were given on how to group pm profiles in Excel file Suppl 2-GHEP-ISFG results_for participants under "report example". In pm-am comparisons, participants were asked to do the following:

1. To report the kinship LR value using the scientific notation, i.e: 1.2E + 09. An LR value equal to or greater than 1.0E + 03 (LR $\geq$ 1000) was defined as a reliable identification threshold. This is an artificially low threshold to be used in real cases and has only been set for this exercise to further investigate matches below threshold.
2. LR values below but close to the threshold could be reported as "probable" identification.
3. If necessary, the mtDNA information of the family group should be used for the identification of the victims, and the laboratory should decide, at its own discretion, whether to combine aSTR with mtDNA LRs. Note that mtDNA data might prove particularly useful to individualize victims within the same family (i.e. father/son).

3.3. Aim 3 - Bayesian treatment of results

In order to evaluate the Bayesian statistical treatment of the episode, a questionnaire was given to the participants on the use of prior odds values for pm-pm re-association and for am-pm comparisons to identify victims (Suppl 2- GHEP-ISFG_DVI#2 Results_for participants). In this document they had to report: (i) if prior odds were used in both pm-pm and pm-am comparisons and if so, what value they had used and (ii) the value of the posterior probability of identification, specifying whether they had combined aSTR and mtDNA results or not.

3.4. Additional difficulties of the exercise

The exercise design included some special difficulties that were not revealed to the participants and that may be encountered in an air crash:

a) A description was made of the incident resulting in 66 victims; however, the 228 pm profiles represented only 65 different profiles.
b) A unique profile represented by two pm samples (pm1842, pm1959) yielded no match with any reference family. Therefore, these remains should be reported as unidentified.
c) As a result of a) and b), two reference families (F82 and F88) did not match any pm sample so that, 64 sets of the re-associated pm samples are expected to match 64 families.

3.5. Results evaluation

This scenario was independently analyzed in advance by three experienced laboratories, Forensic Science Institute of the University of Santiago de Compostela (INCIFOR-USC), Argentine Forensic Anthropology Team (EAAF) and International Commission on Missing Persons (ICMP), to establish a known set of correct results, to be used as a benchmark for comparison to results of participating laboratories. For the results of the pm-pm comparisons, a discrepancy was considered when a laboratory failed to re-associate correctly at least one pm sample into the correct re-association group. In the case of am-pm comparisons by kinship analysis, results with values near to the previously established correct value, the mean and SD of the $\log_{(10)}$ of the reported LRs were calculated. Any reported value of LR with a deviation greater than + /- 2 SD from the mean of $\log_{(10)}$LRs reached by consensus was considered a discrepant result. This criterion was selected as participants used different software to solve the exercise which may yield small differences in results depending on the software and settings used. Discrepant results were analyzed individually to determine the reason for the discrepancy.

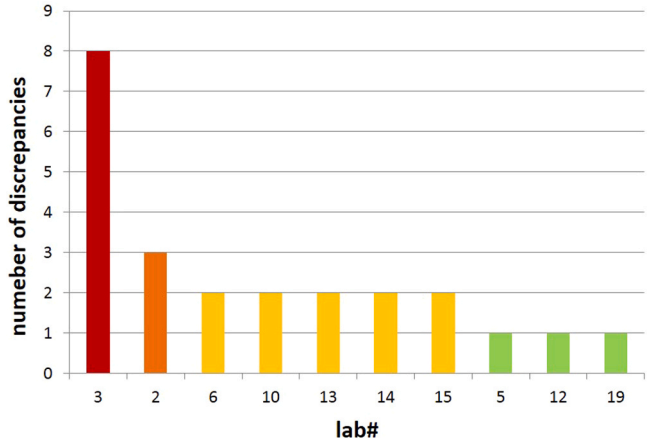

Fig. 2. Discrepancies in pm-pm profiles re-associations according to participating laboratories.

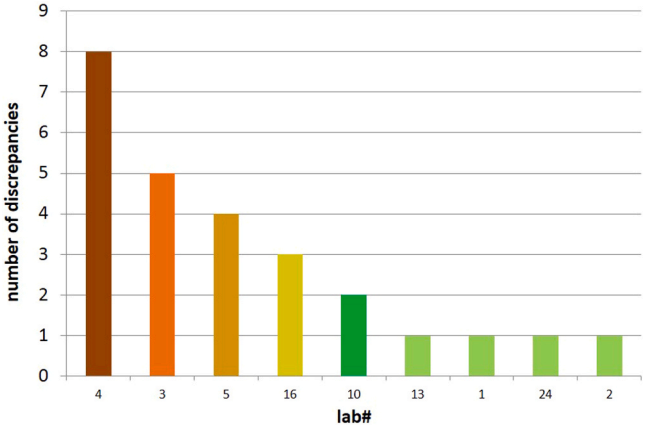

Fig. 3. Discrepancies observed in am-pm comparisons for 44 single victims. Laboratory 14 was not included in this analysis.

5. Results

5.1. Pm-pm comparisons for re-association of "fragmented remains"

The first objective of this exercise was to compare the 228 pm profiles for the re-association of fragmented remains, provided that the LR threshold to report a pm-pm DNA match was LR $\geq$ 1.0E + 07. No pm-pm DNA matches yielding a LR value lower than 1.0E + 07 were included in the exercise, therefore all the pm profiles, in theory, could be re-associated into 65 single profiles.

Regarding the pm-pm re-associations, a divergent result was considered when at least one of the re-associated fragments into the re-association group reported by a laboratory varied from the correct result.

The pm-pm comparisons of the 18 participating laboratories to re-associate 65 unique profiles yielded 1170 observations ($18 \times 65 = 1170$). Twenty four discrepancies (failed or wrong re-associations) were observed out of 1170 comparisons representing a 2% error (24/1170 = 0.02), although the discrepancies were concentrated in a few participants. Fig. 2 shows the distribution of discrepant results among the participating laboratories: ten laboratories reported discrepant pm-pm re-associations, however two laboratories (lab#2 and#3) accumulated 11/24 discrepancies (46% of the total).

5.2. Pm-am comparisons for "victim identification"

Once the pm profiles were re-associated, it was indicated that the pm profile with the highest number of reported markers should be used for pm-am comparisons in kinship analyses. This criterion rather than the profile with the highest discrimination power was chosen in order to simplify the exercise and the results analysis.

The participants were asked to report as positive identifications those cases in which the LR value was LR $\geq$ 1.0E + 03. Forty-four of the 65 victims of the disaster were not related to one another and 21 were related within six families: F30, F57, F67, F72, F86 y F89. In general, the variation of LR values among labs was lower in cases of single individual victims than in cases of related victims within families.

5.2.1. Single victims

Four out of 44 single victim's identifications (9%) were not reported by three laboratories for different reasons that are analyzed further on. However, some participants reported positive identifications but showing discrepant LR values beyond ± 2 SD of $\log_{(10)}$ consensus LR. In order to obtain the mean and SD of $\log_{(10)}$ of the reported LRs, laboratory #14 was excluded from the analysis in specific cases as it made a systematic error: in all the results with consensus LR $\geq$ 1.0E + 05, lab#14 artificially reported LR values five degrees of magnitude higher (i.e: if consensus LR was 1.0E + 07, lab#14 reported 1.0E + 12). This difference is attributable to a problem of this laboratory in transferring the LR values from.txt files generated from Familias3 software to the Excel file provided to report the DNA match results.

There were 26 discrepancies from the consensus LR. Again, these discrepancies were observed mostly in few laboratories. Fig. 3 shows the distribution of the consensus LR discrepancies for the 44 single victims. It can be seen that 3 laboratories accumulated 17/26 discrepancies (65%) while 4 laboratories accumulated 20 of 26 (77%).

Different reasons were detected for these discrepant results: a) partial use of the genetic information of the family references instead of using the whole pedigree for statistical kinship calculation, b) wrong definition of the identity hypothesis (Hi) and/or of non-identity hypothesis (Hni) for the LR calculation. c) failure to identify because of not considering mutations among victims and references.

5.2.1.1. Partial use of references for the calculation of DNA match probability. One of the discrepancies detected in this exercise was the partial use of family references to calculate the DNA match probability. Some labs only used the genetic information of some references disregarding the whole family pedigree. As a result, these labs reported lower DNA match values than correct consensus. Fig. 4A shows the pedigree of Family 38 (top) and the results of the DNA match for MP38-1 (bottom). Family references are F38-11, F38-12 (victim's grandchildren who are full siblings) and F38-9 (victim's daughter in law). Lab#3 did not use information of F38-9 but only of F38-11 and F38-12; consequently, the match value reported by this lab was LR ~ 1.0E + 07 versus the correct consensus LR ~ 1.0E + 10. Similarly, Fig. 4B shows the pedigree of family 52 (top) and the matching results for MP52-13 (bottom). Family references are F52-9 (victim's mother) and F52-12 (victim's full sibling). The consensus result of the DNA match is ~ 4.0E + 10. Lab#4 only used information of F52-12 and disregarded reference F52-9, thus reporting LR ~ 4.0E + 05. This kind of error is also observed for MP3-9, MP7-12, MP19-13, MP20-11, MP37-12, MP52-13 and MP87-11 among others (Individual victim's results are shown in the Excel file Suppl 3- GHEP-ISFG Compiled overall Results).

5.2.1.2. Wrong definition of the identity (Hi) and non-identity (Hni) hypothesis for LR calculation. Another factor that accounts for the discrepant values obtained is that the identity hypothesis (Hi), the non-identity hypothesis (Hni) or both were wrongly defined to calculate the likelihood ratio. Different combinations of incorrect scenarios under Hi and Hni definitions may lead to artificially higher or lower LR values, depending on the error introduced.

Fig. 5 shows examples of different assignments of the victim into the

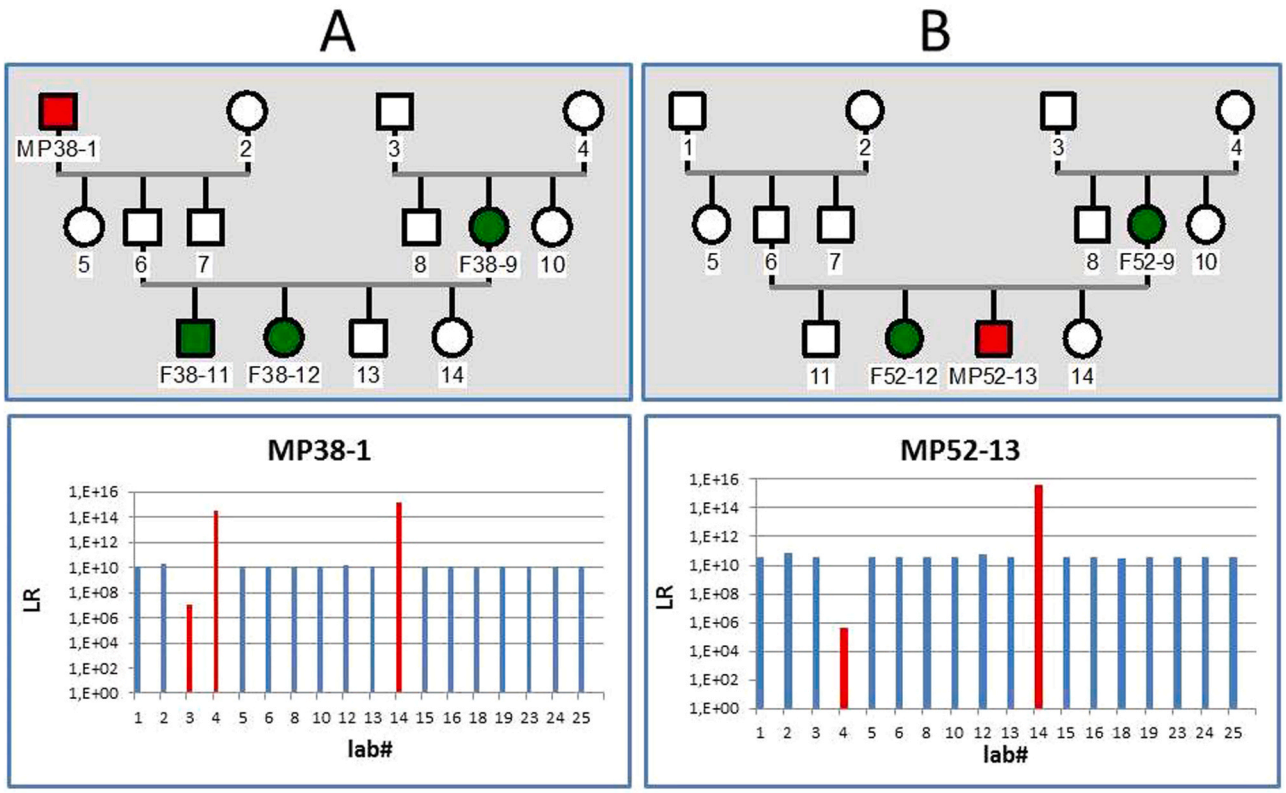

Fig. 4. The figure represents two cases for which discrepancies were due to partial use of references. 4A top: genogram of Family 38 shows victim MP38-1 (red) and references F38-9, F38-11 y F38-12 (green). 4A bottom: LR values for MP38-1. Blue bars show correct consensus values, red bars show out of consensus reported LR values (labs #3, #4 and #14). B top: Genogram of Family 52. MP52-13 (red) and F52-9 and F52-12 (green). B bottom: LR values for MP52-13. Note that lab #14 systematically reports LR values five degrees of magnitude higher than the consensus because in both cases the consensus DNA match results is higher than E + 05. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.).

family pedigree. Fig. 5A shows the pedigree of Family F35 (top left) and the DNA matching results (bottom left) for MP35–6. The consensus LR is approximately 8.7E + 11. However, labs #4 and #5 reported LRs ~ 4.5E + 15. Family F35 has references within a pedigree spanning three generations: F35-1 (victim's father), F35-11 (victim's son) and F35-9 (victim's wife). The family relationship between F35-1 and F35-11 is not questioned (under either Hi or Hni), but these 2 labs wrongly considered the kinship between both reference individuals under Hni. Therefore they reported a four degrees of magnitude higher LR value. On the other hand, lab#14 reported a five degrees of magnitude higher LR value than consensus, but for a different reason (error during data transposition from txt to Excel files, see single victims analysis).

Another similar example is Family F18 (Fig. 5B). In this case, Hi should be defined by designating victim MP18–9 and reference F18-6 as mother and father respectively of F18-12 and 13 (who are full siblings). In Hni, victim MP18-9 should be replaced by a virtual person to keep references F18-6, F18-12 and F18-13 properly linked. Lab# 4 defined Hni by linking only F18-6 as the father of F18-12 and F18-13 but did not replace the victim MP18–9 by any virtual person, thus F18-12 and F18-13 appeared as paternal half-siblings rather than full siblings. Hence, the LR reported is eight degrees of magnitude higher (1.0E+20) than the LR defined by consensus (1.0E + 12). Similar examples can be seen in MP9-12, MP14-9, MP29-9, MP38-1, MP45-11, MP56-9, MP83-6, in which the value of the DNA match reported by some participants is artificially higher than the consensus (Suppl 3- GHEP-ISFG Compiled overall Results).

5.2.1.3. Undetected mutations. The exercise simulated two inconsistencies attributable to one-step mutations in two family groups: Family 21 (D8S1179) and Family 86 (FGA). In Family F21, the mutation involves victim MP21-11 and reference F21-9 (victiḿs mother). A consensus LR ~ 4,5E + 08 was reported by all laboratories except Lab#14, which did not report the match in spite of using the Familias3 software; this discrepancy could have been caused by a problem in setting the mutation rate in the population frequency database, which is critical in this software for the DVI module analysis. Concerning Family F86, as the mutation involves two victims (MP 86-9 and MP 86-12), it has no impact on their identification and few laboratories commented the presence of that inconsistency.

5.2.2. Related victims

The exercise simulated 6 family groups (F30, F57, F67, F72, F86, F89) with several victims within each family representing a total of 21 MP. Most of the family references were sufficient to identify and locate each victim in the family pedigree without the need to elevate any of the victims to the category of a reference sample (DNA match values exceeding LR ≥ 1.0E + 03) except for victims F67-4, MP77-11 and MP78-4 (the last two being single victims). However, the results involving related victims were more discrepant compared to those involving single victims. It was noted that, in addition to the differences resulting from the wrong definition of Hi/Hni or the partial use of references, some laboratories firstly identified one victim into the family group (with LR ≥ 1.0E + 03) and then added the victim firstly identified as a new family reference in order to identify the rest of the victims. Such an approach can be used to make identifications when combined with other non-DNA evidence that confirms the identification, although when considering DNA only, the uncertainty of the first identification is not incorporated into the identification statistics of subsequent identifications and multi-hypothesis approach should be applied (see below *"cases with insufficient references"*). Because of the

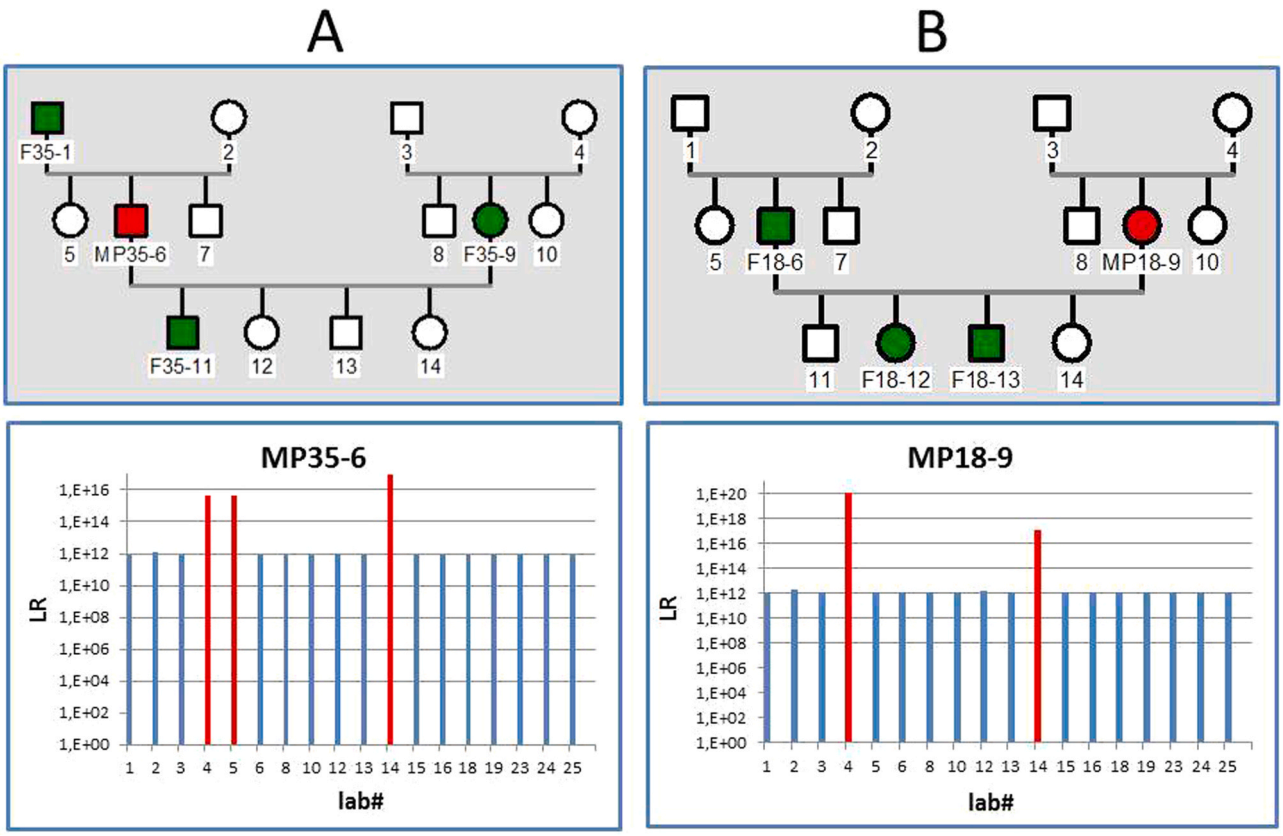

Fig. 5. Representation of two cases with discrepancies due to wrong LR expression. (5A top) Genogram of Family F35 and the DNA match results for MP35–6 (5A bottom). As observed, consensus LR values are LR ~ 8E + 11; labs #4 and #5 reported an LR ~ 4E + 15 (see text for explanation). (5B up): Genogram of Family F18 and reported LR values (5B bottom). Consensus LR ~ E + 12; lab #4 reported a DNA match value eight degrees of magnitude higher than the consensus (see text for explanation).

addition of this new genetic reference information, these laboratories reported LR values much higher than the consensus for the other victims in the same family group. As was commented above, it is worthy of note that references provided did not require the use of this approach to report a reliable match with LR ≥ 1.0E + 03, except for three missing persons: MP67-4, MP77-11 and MP78-4 (MP67-4 and MP77-11 are commented in more detail below under "*cases with insufficient references*").

The diversity of results is presented in Fig. 6. Fig. 6A shows the pedigree chart of Family F30. This family has three victims (MP30-9, MP30-12 and MP30-13), of which MP30-9 is the mother of victims MP30-12 and MP30-13, who are full siblings. Family references to identify the victims are F30-6 (father of MP30-12 and MP30-13 and husband of victim MP30-9), and F30-3 and F30-4, parents of MP30-9. The references provided were sufficient to individualize the three victims of Family F30 with high LR values.

Fig. 6B shows the results of the LR values reported by the participants for MP30-9. As both parents of MP30-9 were available as reference samples, a strong DNA match is expected if complete or almost complete DNA profiles are obtained in PM samples. The consensus LR value for MP30-9 was LR ~ 7.0E + 12 (blue bars). However, as can be seen, lab#12 did not report the individual DNA match value for MP30–9 by just using reverse parentage comparison. This laboratory re-associated the pm profiles of each victim correctly, located each missing properly within the family pedigree chart, but did not report individual LR values for each victim. Instead, it reported an LR value for the whole family pedigree applying a multi-hypothesis approach (LR = 3.5E + 39, data not shown in the figure).

Fig. 6C shows the reported LR values for MP30-12; the consensus LR was LR ~ 1.5E + 09 (blue bars). For the identification of MP30-12, the genetic references provided were F30-6 (victim's father), F30-3 and F30-4 (victim's maternal grandparents), which are sufficient to identify the victim with a high DNA match value. It can be seen however, that the reported LR values are very discrepant due to several factors. Lab#3 and #4 positively reported the identification of MP30-12 (LR ≥ 1.0E + 03), although the reported LR values are far below the correct consensus LR due to the partial use of the genetic information provided by the reference family pedigree. These labs compared only the information of F30-6 (victim's father) reporting an LR ~ 7.1E + 04 (yellow bars). Lab#12 did not report the DNA match for victim MP30-12 for the same reason as that commented for MP30-9 above: it reported the match value of the whole family pedigree by a multi-hypothesis approach. Laboratories #10, #15 and #19 incorporated the genetic information of the previously identified MP30-9 (being a mother of MP30-12) to calculate the LR for MP30-12; as a consequence, they reported a LRs several degrees of magnitude higher than the consensus (LR ~ 9.0E + 12), indicated in green bars. Fig. 6D shows a similar picture for MP30-13: values below the correct consensus due to partial use of references and higher values due to having incorporated previously identified victim (MP30-9) as a new reference. The individualization of MP30-12 (female) and MP30-13 (male), was possible considering the Amelogenin information. Fig. 6B, C and D show the systematic deviation of Lab#14 generated when transferring the LR values from.txt files generated by Familias software to the Excel file provided to report the results (difference of five degrees of magnitude, red bar). Other examples from labs that unnecessarily added firstly identified victims as new references in the family could be observed in the families having several relatives among the victims (F57, F72, F86 and F89 in Excel file Suppl 3- GHEP-ISFG Compiled overall Results).

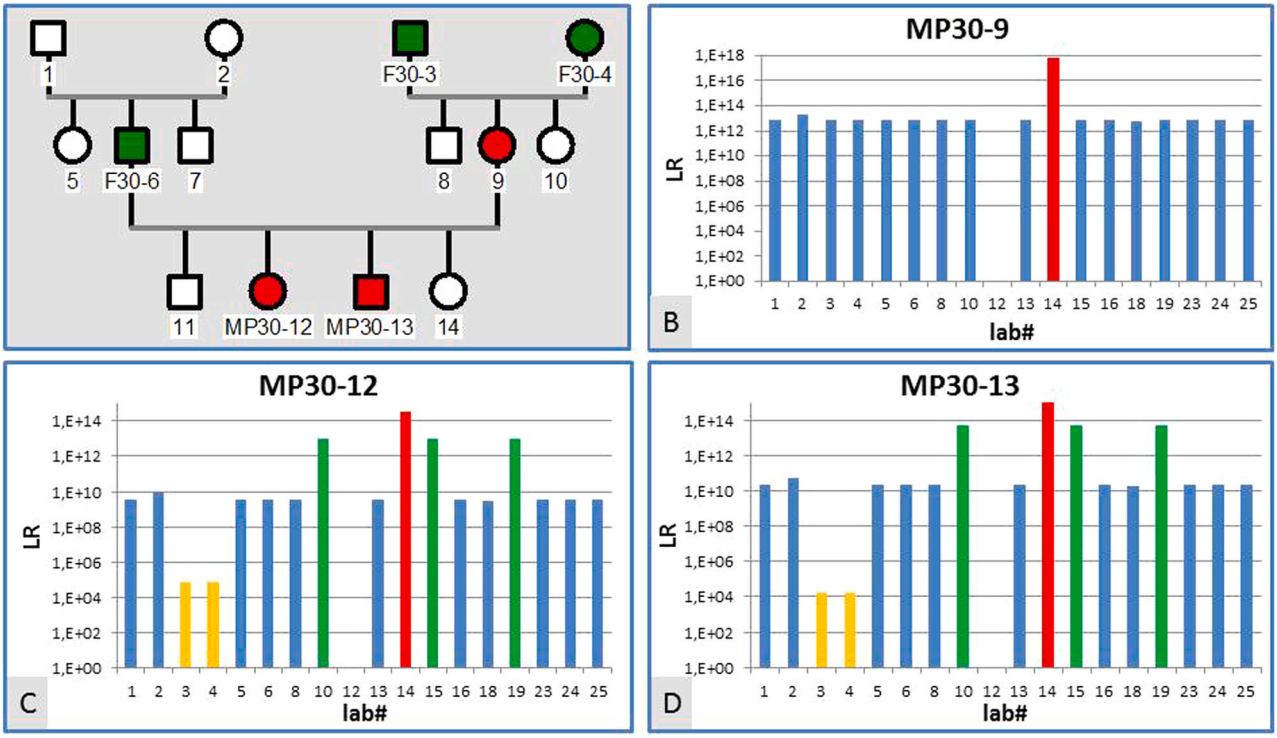

Fig. 6. (A) Genogram representing Family F30: references are F30-6, F30-3 and F30-4 (green) to identify the victims MP30-9, MP30-12 y MP30-13 (red). (B) Blue bars = Consensus aSTRs match values for MP30-9, lab#12 did not reported individual LR values for each victim but reported only the LR value by pedigree multi-hypothesis calculation approach (LR = 3,53E + 39; value not shown in the Fig. 6B, C and D). (C) Blue bars = LRs for MP30–12 showing consensus values; Yellow bars = reported LRs values lower than consensus LR due to partial use of references; Green bars = reported LRs higher than consensus calculated by incorporating genetic information on MP30-9. (D) LR values for MP30-13: colors and reasons for discrepant LR values are similar to the explanation for MP30-12. It is observed that lab#14 (red) shows a systematic error five degrees of magnitude higher than consensus in the three Fig. 6B, C and D. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.).

5.2.3. Cases with insufficient references

As mentioned before, the DVI scenario was designed so that most victims could be identified above the established threshold of LR $\geq$ 1.0E + 03 comparing with the references provided. However, the exercise included some special cases as well: MP67-4, MP77-11 and MP78-4 with deficient family pedigrees, yielding LRs below the reliable threshold.

5.2.4. MP 77–11: a case with LR below threshold 1.0E + 03 and adventitious match

In this air crash, 2 groups of re-associated pm profiles matching family F77 were simulated. The reference samples for MP77-11 are F77-3 and F77-4 (victim's maternal grandparents) as shown in the pedigree chart of Fig. 7A. A re-association group made up of pm001271, pm001571 and pm001720 (group pm1) produced a weak match below threshold (consensus LR ~ 8.1E + 02) with Family F77 (Fig. 7B, blue bars). Furthermore, mtDNA haplotypes of group pm1 and F77-4 produced a match with LR ~ 4.3E + 03 according to EMPOP (brown bars). The LR values of aSTR and mtDNA combined yield a consensus LR ~ 5.0E + 05, exceeding the identification threshold, a result which is also consistent with this as a closed event and a unique mtDNA haplotype; therefore, group pm1 can be identified as belonging to MP77-11.

On the other hand, another group of re-associated pm samples (made up of pm001061, pm001346, pm001824) identified as group pm2 also produced a weak aSTR match with Family F77 but yielding a LR higher than the reliable threshold (LR = 7.8E+03). However, the mtDNA haplotypes from group pm2 and Family F77 (F77-4) mismatched showing clearly different haplotypes, excluding group pm2 as belonging to victim MP77-11. Furthermore, group pm2 yielded a strong aSTR match with Family F38 (MP38-1 could be identified with a consensus of LR ~ 1.3E+10 (see Fig. 4A). It is worth mentioning that victim MP38-1 and family references F38-9, F38-11 y F38-12 do not share the maternal lineage so, mtDNA is not useful for this case. Table 1 summarizes these findings for cases MP77-11 and MP38-1 showing that group pm2 produced a weak adventitious aSTR match with F77. Three laboratories (lab#2, #3 and #14) wrongly associated group pm2 with Family F77 (for MP77-11). Results of the match reported for MP77-11 are shown in Fig. 7B: aSTR LR (blue), mtDNA (brown) and combined STR/mtDNA LRs (green). It can be observed that the three labs obtained an LR value (aSTRs) higher than the threshold of 1.0E + 03 (yellow bars) because they report the "adventitious" match of the group pm2 with F77 instead of the "true" match of the group pm1. Interestingly, lab#2 reported the positive identification of group pm2 as MP77-11 and combined aSTR with mtDNA LRs, despite the mtDNA mismatch between group pm2 and the family reference F77-4 (see Table 1 for mtDNA information). Although lab # 3 reported the group pm2 associated with the family F77 with aSTR LR = 7.0E + 03, this participant indicated that MP77-11 could not be individualized since the mtDNA of the group pm2 matched with family F38. Notably, the three laboratories had also reported a strong DNA match between group pm2 and Family F38 (LR ~ 1.3E+10), positively identifying group pm2 as belonging to MP38-1. A similar case with deficient pedigree and the need to use aSTR and mtDNA genetic information is the case of Family F78: a single reference F78-12 (grandchild) was only available in the case of MP78-4. In short, the inclusion of mtDNA database was useful for many laboratories to distinguish between a "true" and "adventitious" match of pm1 and pm2 MP77-11. Six laboratories used the mtDNA information to solve those cases with insufficient references. Several laboratories (8/18) calculated mtDNA LRs and combined with the aSTR LRs values for all samples, even though it was not necessary for many victim identifications. Finally 4 laboratories did not use mtDNA information at all.

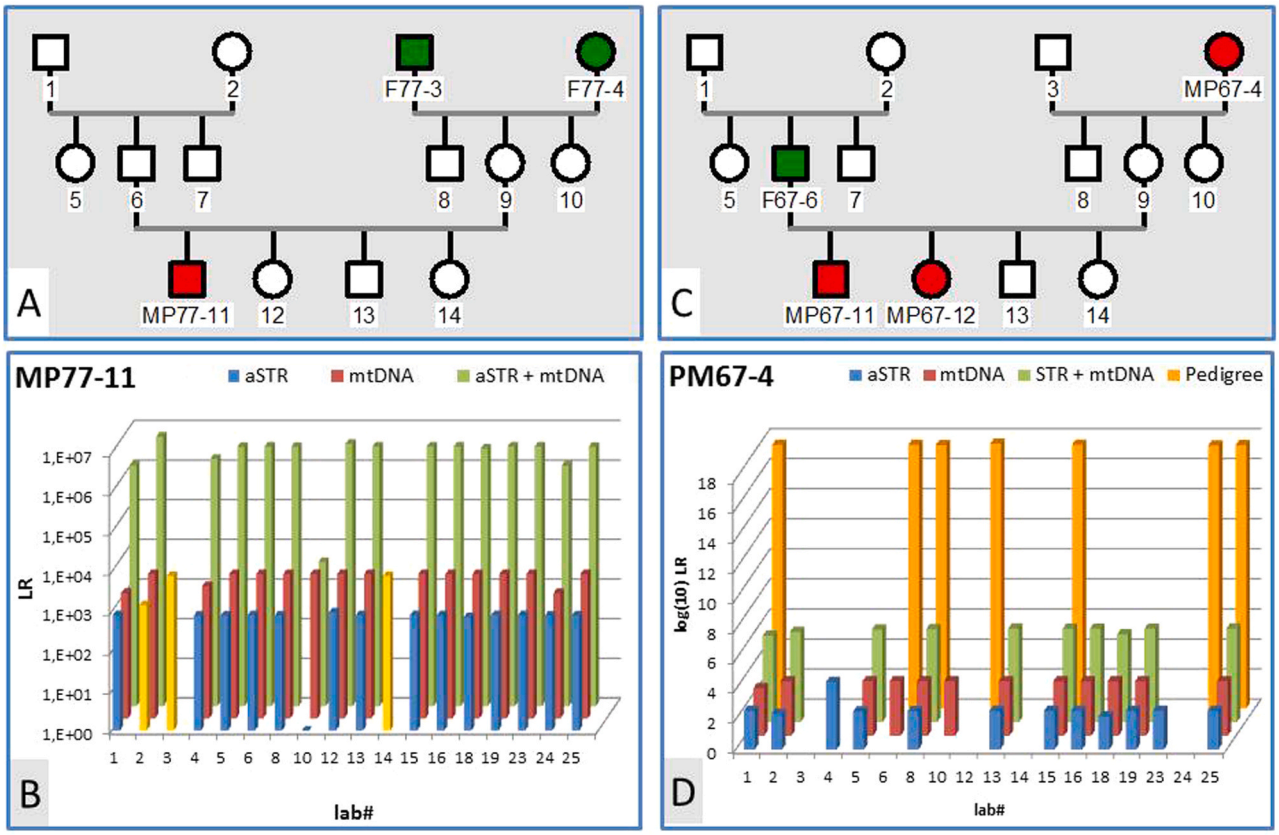

Fig. 7. (A) Genogram of Family F77. The reference F77–4 to identify MP77–11 gives little aSTR genetic information, but enables mtDNA comparison. (B) aSTR LR values (blue bars), mtDNA LRs (brown bars) and combined aSTR/mtDNA LRs (green bars) representing MP77-11 match with F77-4. Labs#2, #3 and #14 reported higher aSTR LR values than consensus LR using group pm2 genetic information (adventitious match) instead of group pm1 (yellow bars). Lab#10 only reported the mtDNA match. (C) Genogram of F67 showing three victims (red) and only one reference (green) that gives aSTR information rendering the LR not enough to identify MP67-4. (D) LR values for MP67-4: aSTR (blue); mtDNA (brown); combined aSTR/mtDNA LRs (green) and LR calculated by pedigree multi-hypothesis approach (orange). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.).

5.2.5. Family F67 involving MP67-4, MP67-11 and MP67-12: a) use of a victim profile as a reference for other missing persons, b) multi-hypothesis approach

As shown in Fig. 7C, two of these related victims (MP67-11 and MP67-12) have F67-6 as a single common reference (father of both). MP67-4 has no genetic references since F67-6 is the son-in-law of this missing person. Therefore, in this case, two approaches may be possible to solve the case:

a) The use of MP67-11 and MP67-12 as new reference samples would be useful for identifying MP67-4. Nevertheless, this can only be done under certain circumstances (e.g., if MP67-11 and MP67-12 were also identified by non-genetic methods). However, in this specific case, the mitochondrial DNA information can be used to link the three victims MP67-4, MP67-11 and MP67-12 as belonging to the same maternal lineage, allowing the exclusion of the other victims, since the shared haplotype is unique for the episode.
b) The correct way to proceed with this case involves a multi-hypothesis simultaneous identification approach. This approach is more complicated since, in general, a single LR value cannot simply be reported, but rather a range of relative likelihoods that need to be evaluated with different priors for each hypothesis [21,28]. Details about this approach are given in Suppl 4 –F67 multi-hypothesis.

Fig. 7D shows a wide variety of results reported for MP67-4. Some laboratories (lab#3 and lab#14) failed to report identification of MP67-4. Most laboratories firstly identified MP67-11 and MP67-12 with high match value (LR ~ 1.0E + 06 for each MP), elevate them as new references and then re-analyzed the aSTR data (blue bars), reporting a weak LR below threshold for MP67-4 (LR ~ 3.8E + 02). Lab#4 reported an aSTR above the reliability threshold (LR = 4.4E + 04) due to an error in expressing Hi/Hni (MP67-11 and MP67-12 were not considered as full siblings). Some laboratories also reported the mtDNA match (LR ~ 4.3E + 03) shown in brown bars and combined aSTR and mtDNA LRs (green bars). Only lab#18 indicated that the mtDNA haplotype of MP67-4, MP67-11 and MP67-12 is unique, and did not match any other family reference for the episode. Other laboratories reported the match and calculated the pedigree match probability using the multi-hypotheses approach (corresponding to the hypothesis H1 in Suppl 4 –F67 multi-hypothesis with LR ~ 3.0E + 17) and represented by orange bars. Lastly, other laboratories (Lab#1,#8, #15 and #25) reported the four LR values (aSTR, mtDNA, combined LRs and H1 from multi-hypotheses), as shown in Fig. 7D.

5.2.6. Other observations

5.2.6.1. Pm samples matching no family references. Two post-mortem samples: pm1842 and pm1959 re-associate with identical genetic profiles (aSTR and mtDNA) but do not match with any family, as described in Methods. All labs except one (lab#14) reported both re-associated profiles without matching any family references. Likewise, most of the labs reported that families F82 and F88 do not produce a match with any sample.

5.2.6.2. Use of "prior odds" values for the incident. The simulated air crash is classified as a "closed" disaster with 66 victims (in fact 65 as

Table 1
Description of the 2 pm sample groups that match Family F77. Group pm1: yields a weak aSTR match (below the reliability threshold) but matched mtDNA haplotypes. Group pm2: yields a weak random match (aSTR) with F77 (above reliability threshold) but the mtDNA mismatch excludes pm remains as belonging to MP77–11. Note that the mtDNA information on Family F38 is not useful to identify MP38-1 (see family pedigree chart in Fig. 4A).

Group pm	Samples	Matched family (aSTR)	LR aSTR	LR mtDNA	Comments
1	1271, 1571, 1720	F77	8.1E + 02	4.3E + 03	Reference F77–4 and victim MP77–11 share maternal lineage: mtDNA is informative
2	1061, 1346, 1824	F77	7.8E + 03	Mismatch	aSTR random match
		F38	1.3E + 10	not informative	References F38–9, F38–11 and F38–12 and victim MP38–1 does not share maternal lineage because mtDNA is not useful.
pm and reference mtDNA haplotypes					
Group pm1		16126C, 16298C, 72C, 263G, 309.1C, 315.1C			
F77-4		16126C, 16298C, 72C, 263G, 309.1C, 315.1C			
Group pm2		16519C, 263G, 309.1C, 315.1C			
F38-9, F38-11 and F38-12		16519C, 152C, 263G, 309.1C, 315.1C, 573.1C			

described in Methods) with varying degrees of fragmentation. In order to analyze the Bayesian evaluation of the episode by the participants, laboratories answered a questionnaire as to whether or not they would use a "prior odds" value to evaluate the statistics in this incident.

5.2.6.3. Use of Prior odds for pm-pm re-associations (fragmented and commingled remains). As described under Methods above, participants were instructed that for pm-pm profiles re-associations the reliable threshold is LR $\geq$ 1.0E + 07 without considering "prior odds" values. Nevertheless, in the questionnaire, labs were asked to indicate whether they would use "prior odds" values for pm profile re-associations in the case of fragmented remains: only 7/18 labs (39%) answered they would use a prior odds value of 1/66 (some indicated 1/65) to be combined with the LR to estimate the posterior odds of re-association.

5.2.6.4. Use of Prior odds for pm-am comparisons (victims' identification). Regarding the use of "prior odds" for pm comparisons with family references to report the posterior odds of identification, 11/18 (61%) laboratories used a value of 1/66 (or 1/65) as "prior odds".

5.2.6.5. Software used to solve the exercise. All participating laboratories used at least one software with facilities for massive comparisons in DVI or MPI; 15 labs used Familias v3 [10], 4 used DNA-VIEW [16], 3 labs used Codis7 [17] and 1 lab used M-FISys [18]. Six labs additionally used other software or a combination of some of the four mentioned above.

4. Discussion and conclusions

This simulated DVI exercise considered different complexities that forensic genetic laboratories may face when massively comparing genetic profiles in order to identify victims in a disaster of these characteristics. The exercise was designed as a "DNA-led" DVI project allowing the identification of all the victims with the exception of the group pm1842/pm1959. These pm profiles did not match any reference family, while two families, F82 and F88, did not match any remains' profile. In addition, victims MP72-11 and MP72-13 can be located within the F72 family but cannot be identified as they are two male siblings, and Amelogenin cannot distinguish them.

We designed a medium-scale disaster avoiding the processing and analysis of hundreds or thousands of victims and genetic profiles. This allowed the organizers to compare the results from the labs more easily by detecting and identifying the reasons for the discrepancies noted. With the experience from a previous GHEP-ISFG MPI exercise [8], the present collaborative exercise includes additional difficulties commonly present in real DVI or MPI processes. In large-scale disasters [19,20] or mass graves with commingled remains [21,22], the re-association of fragmented and/or commingled remains is essential and can be carried out through direct comparisons of pm profiles. Likewise, degraded DNA in poorly preserved samples may yield partial profiles with locus dropout. Although this GHEP-ISFG exercise simulated partial post-mortem profiles due to locus dropout, inconsistencies due to allele dropout were not considered in database comparisons to avoid complex statistical calculations [23,24]. Other common difficulties found in DVI/MPI identifications included in this exercise were: presence of related victims, reference families with variable genealogies, some of which give insufficient genetic information to identify the victim, mutations in aSTR markers, DNA match values below the reliability threshold with the need to integrate previously identified victims into the family group to identify the remaining victims or the need of multi-hypotheses approach, weak adventitious matches and incorporation of the uniparental mtDNA marker to individualize victims of the simulated disaster. All these difficulties were instructive as to the mistakes that can be made when comparing large scale of genetic databases for DVI or MPI.

The direct comparison of the 228 pm profiles for the re-association of fragmented victims showed that some laboratories had difficulties to correctly re-associate profiles of the same victim, even though all the pm profiles of the incident had high discrimination power with at least 10 reportable aSTRs. This is difficult to understand considering that all participants used some powerful software for DVI/MPI. However, errors in the remains re-associations were mostly performed by only few laboratories (Fig. 2). The discrepancies in the profiles re-association were not particularly related to partial profiles of the related victims, in which the power of discrimination of the profiles with locus dropout decreases and may impact on an erroneous association of remains. Nine out of 10 labs that incorrectly re-associated non-matching profiles or failed to associate some profiles to the correct group used the Familias3 software, which has an appropriate DVI [25] module for direct profiles comparisons. However, several labs that used this free software indicated in the questionnaire that the laboratory is not continuously involved in the large-scale profiles comparison for DVI / MPI, so, the reason for the discrepancies is probably due to difficulties in the setting and handling the software used, or in data handling errors in compiling the results for reporting.

The simulated event included 44 single victims and 21 related victims belonging to six families, a very common situation in an air crash. The identification of victims by comparing family references showed two different patterns in the results reported: the results of the single 44 victims showed less variation than those of the related victims. However, just few laboratories reported most of the differences from the correct consensus. The results of the 44 single victims of the episode were evaluated against the results of the consensus LR. A total of 26 discrepancies were observed in different laboratories, although 65% (17/26) of such discrepant results were obtained by only three labs (Fig. 3).

One of the reasons for the discrepancies with the consensus LR was the partial use of the reference family genetic information to be compared with the remains. This error usually gives a lower DNA match value than the correct LR, an error that had been observed in the previous GHEP-ISFG collaborative exercise [8].

Another reason why discrepant values were observed was that some laboratories wrongly defined the identity hypothesis (Hi) or the non-identity hypothesis (Hni) or both for the DNA match calculations, producing artificially higher or lower values than the correct LR. In DVI or MPI in which relatives are used as a reference to identify missing persons, complex and variable genealogies can influence the correct assignment of victims to the reference family. Many guidelines on best practices for DVI [1,2] or MPI [3,4] recommend the use of pedigree chart to describe the relationship between family references and victims. In this simulated disaster, in contrast to the previous exercise [8], pedigree charts were used —instead of simply listing individual relationships— to precisely define the relationships of all family members to each other. An additional challenge to the participants of this exercise was to then transfer these relationships into the DVI software using proper format and relationship designations. This can be tricky for laboratories that do not regularly conduct this type of work. In fact, a questionnaire sent to the participants revealed that many of them are not often involved in DVI cases (data not shown).

There were several cases where laboratories reported identifications based on the serial identification of missing persons, where a missing person is categorized as "identified" based on reference profile comparisons, and the profile from this victim is then used as a reference for the identification of other family members. This can be an important approach in real world incidents, where DNA evidence from the initial match can be confirmed by other information, so that the identification can be considered confirmed or official and providing justification for the elevation of the victim profile to that of a reference sample. However, from the standpoint of DNA only, this approach is not strictly correct, as there is no basis to transform a finite LR of the first match to an infinite LR (complete certainty) upon which to base the second identification. Several laboratories used serial identifications to report elevated match statistics on cases where the existing reference samples were sufficient to exceed the match threshold without using missing persons profiles as additional reference profiles.

There was only one case (Family 67, missing person MP67-4, Fig. 7C) in the exercise where existing references were not sufficient to reach the match threshold, and in that case many labs took the serial identification approach which enabled them to report a correct identification. It is important to emphasize that, in order to apply this approach, a process of elimination of the other victims must be carried out by exclusion of the DNA match; in this case mtDNA was useful as the haplotype of the related victims of family F67 was unique for the aircrash. However, a strictly correct approach to solving this case based on DNA alone is to do a simultaneous calculation involving all victims and references, and to consider the evidentiary weight associated with each possible relationship scenario involving hypotheses or relatedness and non-relatedness. An example of this multi-hypothesis approach based on [21] can be found in Suppl 4- F67 Multi-hypothesis in this article.

When comparing large databases of pm profiles with am profiles by kinship analysis, in addition to the presence of weak DNA matches below the established reliability threshold, adventitious or random matches may occur [6]. Both situations: "true" weak and false "adventitious" matches may be due to the presence of partial pm profiles with low discrimination power (ie: DNA degradation), as well as deficient family reference pedigrees to produce a strong match. As a consequence of these two situations, weak true DNA match values may be hidden by adventitious matches, making it difficult to distinguish a true identification from a random match [26]. In these cases, the analysis of other lineage markers as mtDNA or Y-STR may be of help to distinguish true from adventitious matches. Considering this situation, the exercise included a special case (MP77-11) as a challenge, in which the (aSTR) LR value of the "adventitious" match was higher than the "true" match; however, mtDNA data could solve the identification of the victim. This case served as a lesson, as some labs linked the same pm remains with two different families.

In DVI or MPI contexts, probabilities must be considered within the Bayesian framework [26,27]. Large numbers of missing profoundly influence the probability of identity of the victims, since the identifications must necessarily be analyzed in the context of the episode. The value of prior odds is usually represented by 1/MP (MP = number of missing persons), or 1/(MP + 1) if we want to consider that the MP we are looking for is not among the victims (i.e.: open episodes). This exercise simulated a "closed" episode, an air crash with 66 passengers who were found with various degrees of fragmentation and a total of 228 fragments recovered and analyzed; the correct prior odds for the episode is 1/66 (or 1/65 once it is revealed that there were 65 single profiles and assuming that the list of passengers was wrong or the remains of a victim were not recovered). The participants in this exercise answered a questionnaire to indicate the criteria they used to evaluate the Bayesian statistics for the episode. Surprisingly, only 7/18 (39%) laboratories correctly answered that they would use a prior odds value of 1/66 (or 1/65); it is worth emphasizing that this is an event with fragmented and commingled remains, so it is appropriate to consider a prior odds value for fragments re-associations. Although the use of prior odds (1/66) combined with the LR has a negligible impact on the posterior odds for re-associations by direct match (all simulated pm profiles reported 10–18 aSTR markers with high discrimination power) prior odds should still be considered within a proper Bayesian approach to the analysis.

Regarding the use of prior odds for pm-am comparisons to "identify" victims, only 61% (11/18) of the labs correctly considered to use prior odds of 1/66 or 1/65 for the episode. Conversely, 7 out of 18 laboratories answered that they would not use any prior value for the incident. It is worth noting that Bayesian treatment considering prior odds values is extremely important when identifying large numbers of victims because of the impact in the posterior probability of identity. For the purposes of this exercise, an LR of 1000 was set as sufficient for an identification, however this is an unrealistically low value to use in real incidents, as it corresponds to only a 93% posterior probability (surety of match) when 1/66 is used as prior odds. Setting identification thresholds based on posterior probabilities (for example 99.95%) is a far better approach, and can, for example, assist in discarding adventitious matches as in the case of MP77-11.

In processes of identification of large numbers of victims, it is common to incidentally find inconsistencies between social and genetic family pedigrees; likewise, there may be differences between the number of victims reported and the remains recovered. This exercise described 66 victims, although only 65 unique profiles were included, as if victim number 66 was not recovered. These 65 single genetic profiles produced a match with only 64 of the 66 family references provided; the remaining families F82 and F88 showed no match with any pm sample. Conversely, a unique profile represented by two re-associated pm samples did not produce match with any family (pm001842 pm001959), simulating an unidentified victim/unreported missing person: this finding could be interpreted as the presence of an unknown passenger on the flight or an incidental finding where the victim is not biologically related to the family. Only 14 of 18 laboratories (78%) answered that two families (F82 and F88) and pm001842/pm001959 profiles remained unmatched. The remaining participants did not mention this finding in the questionnaire or reported that all the pm samples gave DNA matches with the 66 reference families.

DVI or MPI situations with large numbers of data, necessarily requires the use of software suitable for storage, handling and comparisons. There are several software packages with different facilities for DVI [15–17,24,28]; all participating laboratories used at least one software with appropriate modules for DVI/MPI or a combination of these. However, the 3 labs that accumulated 17/26 discrepancies for the identification of the 44 single victims used Familias3, and 1 lab that reported 3 discrepancies use CODIS7 and Familias3. Similarly, 2 labs reported 11/24 discrepancies in direct pm-pm comparisons using Familias3. While for large numbers of database comparisons the use of appropriate software is essential, it is fundamental to set and manage it correctly.

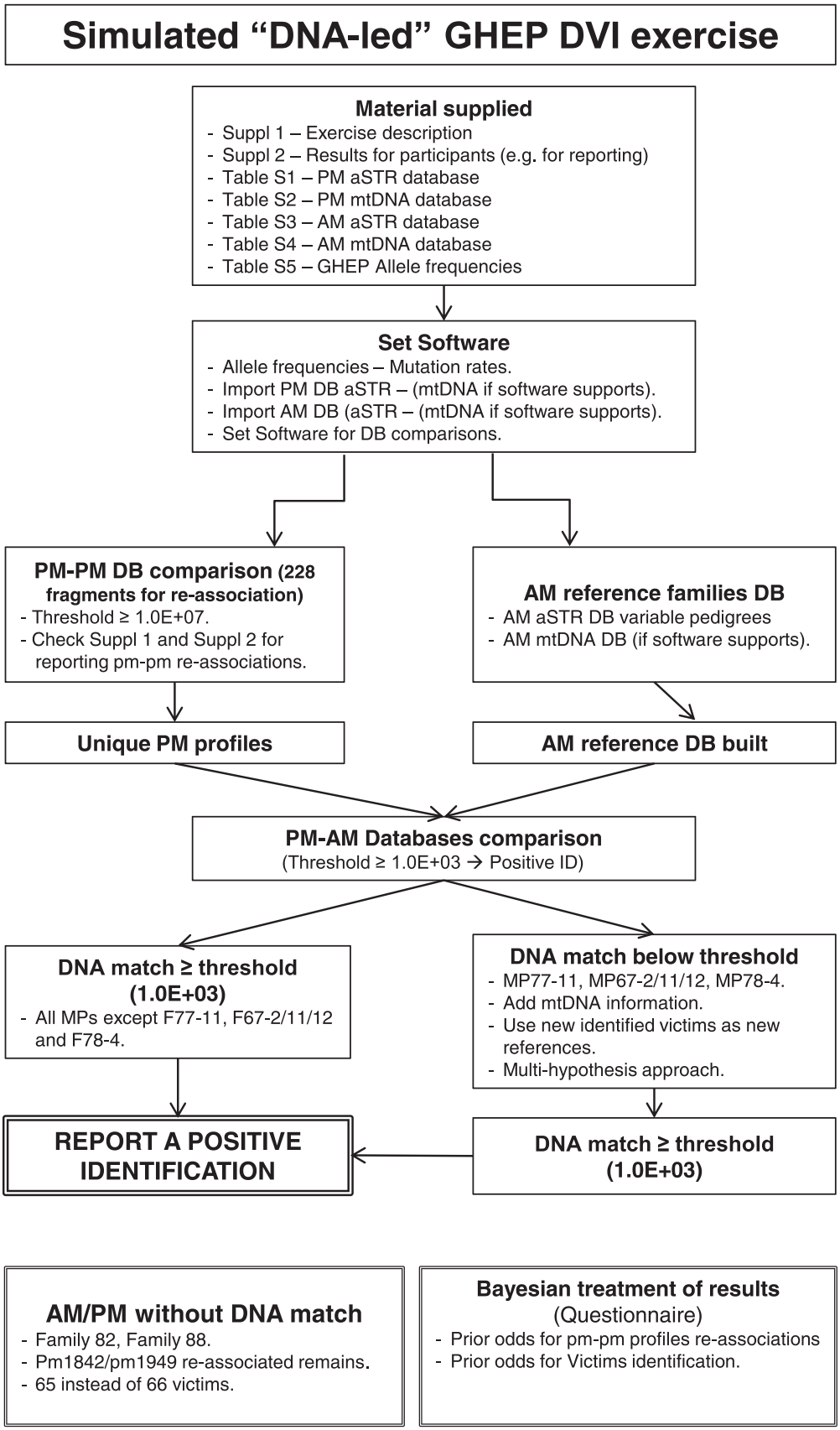

Fig. 8. Flowchart showing the steps to be followed to test the entire DVI exercise. DB = database; PM = post-mortem; AM = ante-mortem; MP = missing person.

Fig. 8 shows a flow chart describing the steps to be followed to perform this DVI exercise; special difficulties and the Bayesian treatment of the episode are shown separately at the bottom.

This new GHEP-ISFG collaborative exercise on "DNA-based" victim identification in a DVI process revealed several sources of error that forensic genetics laboratories may need to confront when searching DVI DNA data. The exercise was designed with several problematic samples and pedigrees to add complexity to the genetic identification of victims. As a result, valuable lessons have been learned from all aspects of this exercise: fragment re-associations, victim identification through kinship analysis, related victims, presence of mutations, insufficient number of family references, Bayesian framework, and correct use of DVI software. The underlying genetic profiles and all results of this exercise have been made available and can be used by other laboratories that wish to evaluate their performance in a "DNA-led" DVI scenario.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Acknowledgments

We thank the 18 laboratories that participated in this exercise.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.fsigen.2021.102527.

References

[1] A. Sozer, M. Baird, M. Beckwith, B. Harmon, D. Lee, G. Riley, S. Schmitt, Guidelines for Mass Fatality DNA Identification Operations, AABB, 2015 ⟨http://www.aabb.org/programs/disasterresponse/Documents/aabbdnamassfatalityguidelines.pdf⟩ (Accessed in November, 2020).
[2] Interpol DVI guide (Version 2018). ⟨https://www.interpol.int/How-we-work/Forensics/Disaster-Victim-Identification-DVI⟩ (Accessed in November, 2020).
[3] Scientific Working Group on DNA Analysis Methods (SWGDAM). Guidelines for Missing Persons Casework. ⟨https://www.swgdam.org/publications⟩ (Accessed in November, 2020).
[4] Missing People, DNA Analysis and Identification of Human Remains, A guide to best practice in armed conflicts and other situations of armed violence. International Committee of the Red Cross (ICRC), ⟨https://www.icrc.org/en/doc/assets/files/other/icrc_002_4010.pdf⟩ (Accessed in November, 2020).
[5] M. Prinz, A. Carracedo, W.R. Mayr, N. Morling, T.J. Parsons, A. Sajantila, R. Scheithauer, H. Schmitter, P.M. Schneider, 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, https://doi.org/10.1016/j.fsigen.2006.10.003.
[6] ENFSI DNA Working Group. DNA database management review and recommendations – April 2017. ⟨https://enfsi.eu/wp-content/uploads/2017/09/DNA-databasemanagement-review-and-recommendatations-april-2017.pdf⟩ (Accessed in November 2020).
[7] C.H. Brenner, Some mathematical problems in the DNA identification of victims in the 2004 tsunami and similar mass fatalities, Forensic Sci. Int. 157 (2006) 172–180, https://doi.org/10.1016/j.forsciint.2005.11.003.
[8] 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, https://doi.org/10.1016/j.fsigen.2015.11.004.
[9] K. Fernández, J. Gómez, J. García-Hirschfeld, E. Cubillo, C. Sánchez de la Torre, G. Vallejo, Accreditation of the GHEP-ISFG proficiency test: One step forward to assure and improve quality, Forensic Sci. Int. Genet SS 5 (2015) e515–e517, https://doi.org/10.1016/j.fsigss.2015.09.204.
[10] 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, https://doi.org/10.1016/s0379-0738(00)00147-x.
[11] F. Alshamali, A. Brandstätter, B. Zimmermann, W. Parson, Mitochondrial DNA control region variation in Dubai, United Arab Emirates, Forensic Sci. Int Genet. 2 (2008) 9–10, https://doi.org/10.1016/j.fsigen.2007.08.005.
[12] Seung Beom Hong, Ki. Cheol Kim, Wook Kim, Population and forensic genetic analyses of mitochondrial DNA control region variation from six major provinces in the Korean population, Forensic Sci. Int. Genet. 17 (2015) 99–103, https://doi.org/10.1016/j.fsigen.2015.03.017.
[13] A. Brandstätter, B. Egyed, B. Zimmermann, A. Tordai, Z. Padar, W. Parson, Mitochondrial DNA control region variation in Ashkenazi Jews from Hungary, Forensic Sci. Int. Genet. 2 (1) (2008) 4–6, https://doi.org/10.1016/j.fsigen.2007.07.006.
[14] R. Jankova-Ajanovska, B. Zimmermann, G. Huber, A.W. Röck, M. Bodner, Z. Jakovski, B. Janeska, A. Duma, W. Parson, Mitochondrial DNA control region analysis of three ethnic groups in the Republic of Macedonia, Forensic Sci. Int. Genet. 13 (2014) 1–2, https://doi.org/10.1016/j.fsigen.2014.06.013.
[15] EMPOP (EDNAP mtDNA Population Database); ⟨https://empop.online/⟩.
[16] C. Brenner, ⟨http://dna-view.com/dnaview.htm⟩ (Accessed in November, 2020).
[17] CODIS; Federal Bureau of Investigation. ⟨https://www.fbi.gov/services/laboratory/biometric-analysis/codis⟩ (Accessed in November, 2020).
[18] H.D. Cash, J.W. Hoyle, A.J. Suttond, Development under extreme conditions: forensic bioinformatics in the wake of the World Trade Center disaster, Pac. Symp. Biocomput (2003) 638–653.
[19] National Institute of Justice, Lessons Learned from 9/11: DNA Identification in Mass Fatality Incidents, US Department of Justice, Washington, DC, 2005 accessed in November, 2020.
[20] S. Donkervoort, S.M. Dolan, M. Beckwith, T.P. Northrup, A. Sozer, Enhancing accurate data collection in mass fatality kinship identifications: lessons learned from Hurricane Katrina, Forensic Sci. Int. Genet. 2 (2008) 354–362, https://doi.org/10.1016/j.fsigen.2008.05.008.
[21] Thomas J. Parsons, Rene M.L. Huel, Zlatan Bajunović, Adnan RizvićFicha, Large scale DNA identification: the ICMP experience, Forensic Sci. Int. Genet. 38 (2019) 236–244, https://doi.org/10.1016/j.fsigen.2018.11.008.
[22] E. Huffine, J. Crews, B. Kennedy, K. Bomberger, A. Zinbo, 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. 42 (June (3)) (2001), 271–27. PMID: 11387637.
[23] Peter Gill, Hinda Haned, Oyvind Bleka, Oskar Hansson, Guro Dørum, Thore Egeland. Genotyping and interpretation of STR-DNA: low-template, mixtures and database matches—Twenty years of research and development, Forensic Sci. Int. Genet 18 (2015) 100–117, https://doi.org/10.1016/j.fsigen.2015.03.014.
[24] John Buckleton, Chris Trigg, Dealing with allelic dropout when reporting the evidential value in DNA relatedness analysis, Forensic Sci. Int. 160 (2006) 134–139, https://doi.org/10.1016/j.forsciint.2005.08.023.
[25] D. Kling, A.O. Tillmar, T. Egeland, Familias 3—extensions and new functionality, Forensic Sci. Int. Genet. 13 (2014) 121–127, https://doi.org/10.1016/j.fsigen.2014.07.004.
[26] C.H. Brenner, B.S. Weir, Issues and strategies in the identification of World Trade Center victims, Theor. Pop. Biol. 63 (2003) 173–178, https://doi.org/10.1016/s0040-5809(03)00008-x.
[27] C.H. Brenner, Some mathematical problems in the DNA identification of victims in the 2004 tsunami and similar mass fatalities, Forensic Sci. Int. 157 (2006) 172–180, https://doi.org/10.1016/j.forsciint.2005.11.003.
[28] C.J. van Dongen, K. Slooten, M. Slagter, W. Burgers, W. Wiegerinck, Bonaparte: application of new software for missing persons program, e119–e12, Forensic Sci. Int. Genet Suppl. Ser. 3 (2011), https://doi.org/10.1016/j.fsigss.2011.08.059.