Strategy for STR typing of bones from the Second World War combining CE and NGS technology: A pilot study

Irena Zupanič Pajnič, Paolo Fattorini

PII: S1872-4973(20)30173-3

DOI: https://doi.org/10.1016/j.fsigen.2020.102401

Reference: FSIGEN 102401

To appear in: *Forensic Science International: Genetics*

Received Date: 4 May 2020

Revised Date: 21 September 2020

Accepted Date: 26 September 2020

Please cite this article as: { doi: https://doi.org/

Strategy for STR typing of bones from the Second World War combining CE and NGS technology: a pilot study

Irena Zupanič Pajnič[a], Paolo Fattorini[b*]

* Corresponding author

[a]Institute of Forensic Medicine, Faculty of Medicine, University of Ljubljana, Korytkova 2, 1000 Ljubljana, Slovenia

[b]Department of Medicine, Surgery and Health, University of Trieste, Italy

Highlights

- The identification of skeletal remains in mass graves is performed by STR typing.
- Conventional PCR/CE analysis is limited by DNA degradation.
- PCR/MPS of bone samples produces reliable molecular data.
- PCR/MPS of bone samples allows genetic profiles.
- The combined use of PCR/CE and PCR/MPS technologies improves results.

Abstract

The genetic identification of skeletal remains found in Second World War mass graves is complicated because of the poor quality of the samples. The aim of this study was to set up a workflow for STR typing of such samples combining PCR/CE and PCR/NGS technologies. To this end, 57 DNA samples from an equal number of 75-year-old femurs were studied. After a first round of PCR typing using GlobalFiler CE, 42 samples yielded a full profile and were therefore submitted to our standard workflow. The 15 samples that yielded no or a limited number (2–17/21) of autosomal STR markers as well four bone control samples that provided a full profile with the conventional PCR/CE test were typed in duplicate by the GlobalFiler NGS kit. Despite the degradation of the samples, which resulted in lower coverage and a lower % of on-target reads, reliable sequencing data were obtained from 16/19 samples. The use of a threshold of 30× for the locus call led to a consensus profile (cp) of 20–31/31 STR autosomal loci in 10 samples and to a cp of 8–10/31 loci in two samples, whereas the four control samples yielded a cp of 26–31/31

loci. Finally, the data from the NGS typing were combined with those from the CE typing. This last task allowed us to recover (on average) three alleles per sample and to increase the number of the heterozygous patterns in 37 cases. In total, the combined approach proposed here made possible the genetic typing of 65 to 100% of the autosomal STR markers in 10/15 (66.6%) skeletal remains that yielded no or very poor results with the conventional PCR/CE approach. However, because several artefacts (such as allelic drop-out and allelic drop-in) were scored, the risk of mistyping cannot be neglected.

Keywords: bones, STR typing, missing person identification, DNA degradation, Second World War, massive parallel sequencing

1. Introduction

Innumerable hidden mass graves have been unearthed in recent years throughout the world that are a direct result of the Second World War, such as missing Norwegian soldiers buried in Russia [1] and mass graves found in Poland [2,3]. In Slovenia, the discovery of human remains from the Second World War is common. The Slovenian Government Commission on Concealed Mass Graves estimates that over 600 concealed mass graves contain the remains of almost 100,000 persons that were extrajudicially killed during or immediately after the Second World War [4]. The majority of the victims of these crimes remain buried, and their genetic identification started more than a decade ago [5–8].

Skeletal remains, especially aged ones, are among the most challenging biological samples for forensic human identification analyses. Old bones usually contain low amounts of DNA with variable levels of degradation and endogenous (bacterial or fungal) contamination [9]. In addition, the success of DNA typing is limited by substances that act as PCR inhibitors [9–11]. Finally, because the risk of human exogenous contamination is high, both precautions and authentication criteria determined in ancient DNA analysis must be adopted [12].

Autosomal nuclear DNA is usually preferred for the identification of missing persons because it is individual-specific and it provides bi-parental kinship information, with the

polymerase chain reaction / capillary electrophoresis (PCR/CE) analysis of short tandem repeat (STR) markers as the gold standard [1–3,5–8,13–15]. In the last decade, forensic genetic analyses have gained much potential with new technologies, including next-generation sequencing (NGS), which makes possible high throughput and acquisition of a larger amount of genetic information from a single experiment [16,17]. More recently, several kits allowing the NGS of forensically relevant STR markers have been customized and used both in pilot studies and for applied purposes, which have shown the strong discriminatory power of this new technology [18–23]. However, because the cost of the NGS experiment per sample still remains high, it is also true that strategies need to be developed for identifying cases that must be tackled by NGS-based approaches [16]. This could apply to the case of bones from the Second World War, which have shown with a certain frequency (depending on the individual environmental conditions of the mass grave) poor or even inconclusive results with conventional PCR/CE methods [6–8].

Thus, the aim of this study was to develop a strategy for STR typing of bones from the Second World War combining PCR/CE and PCR/NGS technology. In particular, we tried to establish a sample workflow that combines the use of the GlobalFiler CE kit and GlobalFiler NGS kit. The shortness of the loci of the NGS kits (see Table S1) coupled with the high number of STR markers included there should allow more successful typing in cases in which PCR/CE technology has failed. The results are presented for a set of 57 bones from the Second World War.

2. Materials and methods

2.1 Bone samples

Fifty-seven femurs excavated from the Konfin I Mass Grave (in Slovenia) were used in this study. Specifically, 5 to 8 cm of compact cortical diaphysis fragments from each bone sample were cut with a circular diamond saw (Dremel, Racine, WI, USA) and used for DNA extraction. Specific precautions were followed, and the skeletal remains were handled under suggested conditions [12, 24]. To reduce contamination from previous handling and to eliminate potential surface contaminants, the bone samples were cleaned both mechanically and chemically. The exposed bone surface was removed through drilling, and bone material was acquired directly from inside the specimen [25]. To avoid cross-contamination between samples, reusable tools for bone processing and the complete workspace were cleaned after each sample was processed. Cleaning was carried

out by washing with 5% Alconox detergent (Sigma-Aldrich, St. Louis, MO, USA), sterile bi-distilled water (Millipore, Darmstadt, Germany), and 80% ethanol (Fisher Scientific, Loughborough, UK). This was followed by UV irradiation, as described previously [26]. The samples were cleaned and ground in a room designed exclusively for processing old skeletal remains, making use of an MC 3 closed microbiological safety laminar flow cabinet (Iskra Pio, Šentjernej, Slovenia) outfitted with HEPA filters and UV lights.

The bones were ground into powder for 1 to 2 minutes at 30 Hz using a homogenizer (Bead Beater MillMix 20 (Tehtnica, Domel, Železniki, Slovenia). To prevent bone warming during drilling and cutting, a lower speed setting was used for abrasion and cutting, and the bones were frequently cooled in liquid nitrogen. The powder samples were weighed, and 0.5 g from each of them was used for DNA extraction.

2.2 DNA extraction

Decalcification and purification of DNA were performed following the procedure described previously [27]. Briefly, total demineralization was performed by adding 10 ml of 0.5 M Na_2EDTA (Promega, Madison, WI, USA) to 0.5 g of bone powder. The samples were incubated at 37 °C overnight in a Thermomixer comfort (Eppendorf, Hamburg, Germany) with shaking at 950 rpm. After centrifugation, the supernatant was discharged and the precipitate was washed with bi-distilled water (Millipore), centrifuged, and G2 extraction buffer, proteinase K (both Qiagen, Hilden, Germany), and 1M DTT (Sigma-Aldrich) were added to the precipitate and it was incubated 2 to 3 hours at 56 °C and 750 rpm. Extracted DNA was purified in a Biorobot EZ1 device (Qiagen) using the EZ1 DNA Investigator Kit (Qiagen) and EZ1 DNA Investigator Card (Qiagen). Following the manufacturer's instructions [28], the "trace" protocol was selected for purification and 50 µl of DNA was obtained from each sample. Extraction-negative controls (ENC) were carried out [12,24] in parallel. For the elimination database, buccal swabs were obtained from everyone engaged in any part of handling the samples.

2.3 DNA quantification

A qPCR-based assay was performed. PowerQuant kit (Promega) was used at the suggested conditions for each sample in duplicate. Instrument calibration, setup, and programming were all carried out in line with the technical manual [29]. Raw data were

obtained using the ABI 7500 Real-Time PCR System (Applied Biosystems, AB, Foster City, CA, USA), and the results were then converted from raw data by using PowerQuant Analysis Tool software (www.promega.com/resources/tools/powerquantanalysis-tool). The imported data were then used to evaluate standard curves (acceptable r^2 and slope values), to determine quantities of Auto (Autosomal), Deg (Degradation), and Y targets in the samples, and to calculate the [Auto]/[Deg] ratio together with the IPC (Internal Positive Control) shift value. The minimum value for the IPC shift was set at 0.30. Negative template controls and ENC were analyzed in order to verify the cleanliness of laboratory plastics and reagents.

2.4 STR typing using PCR/CE technology

The GlobalFiler CE (Thermo Fisher Scientific, TFS, Waltham, MA, USA) kit was used at 29 cycles of PCR according to the manufacturer's instructions [30] on a Nexus Master Cycler (Eppendorf). The maximum volume of extracts (15 µl) was used for amplification of all samples, corresponding to 0.485±0.466 ng (median value= 0.365; min= 0.039; max= 2.718). Simultaneously with bone samples, negative PCR controls and the ENC were amplified, and the maximum volume of these extracts was used for amplifications.
The amplicons were separated on an automatic ABI PRISM™ 3130 Genetic Analyzer (Applied Biosystems, AB) using the 3130 Performance Optimized Polymer 4 (AB). The genetic profiles were determined using Data Collection v 4.0 and GeneMapper ID-x v 1.5 (AB) software with a peak amplitude threshold of 50 relative fluorescence units (RFU) for all dyes. Stutter band threshold values were set according to the manufacturer's instructions [30] and the stutter file was imported to GeneMapper ID-x software.
After visual evaluation of the electropherograms, 38 out of 42 samples that showed full or almost full profiles (23/24 loci) underwent a duplicate PCR test under the same analytical conditions. The data from the duplicate tests were used to build both composite and consensus autosomal genetic profiles. For the first, all the alleles scored in the two replicates were included [31]. For the second, instead, only the alleles observed > $n/2+1$ replicates (e.g., both duplicates) were reported, as previously described [32].
Genetic profiles from the elimination database were compared with those obtained from bones to monitor possible contamination of aged DNA with contemporary DNA.

2.5 STR typing using PCR/NGS technology

The GlobalFiler NGS STR panel version 2 (TFS) was used. The DNA libraries and template preparations were run automatically on the Ion Chef System (TFS), and an Ion S5 System (TFS) was used for sequencing. This method was used for duplicate analyses of 15 samples that yielded no or partial profiles using PCR/CE technology as well four bone samples that yielded full profiles and were therefore used as controls (see Table 1 for the full list of these samples). The ENC was also analyzed in duplicate for a total of 40 tests.

PCR was run at a standard number of 24 cycles by using 15 µl of each DNA sample, containing between 0.039 and 0.670 ng of template. Fully automated library preparation was performed using the Precision ID DL8 Kit for Chef, and eight libraries (for Ion 520 chip) and 32 libraries (for Ion 530 chip) were combined into one tube for each chip in line with the manufacturer's user guide [33]. The concentration of the combined library pool was determined in duplicate together with standards and negative control by qPCR with the Ion Library TaqMan Quantification Kit (TFS). The E. coli DH10B Ion Control Library was diluted to a final concentration of 6.8 pM, 0.68 pM, and 0.068 pM in sterile bi-distilled water, and quantification of the diluted library sample (1:100 dilution) was performed using 7500 Real-Time PCR with HID Real-Time PCR Analysis Software version 1.2 in line with the manufacturer's user guide [33]. Two library pools -32 pM combining eight samples and 18 pM combining 32 samples- were used for fully automated DNA template preparation on the Ion Chef System. The templates were prepared using the Ion S5 Precision ID Chef Reagents and loaded using the single chip loading workflow. Sequencing made use of the Ion S5 Precision ID Sequencing Reagents and Ion S5 Precision ID Sequencing Solutions [33].

Ion Torrent Suit Software 5.6 (TFS) and Converge software version 2.0 (TFS) were used for NGS analysis of STR markers by applying the manufacturer's default relative settings of 0.05 for both the analytical and stochastic thresholds, with the exceptions reported in Table S1 [34]. The default stutter ratios (ranging from 0.08 for the Penta_D marker to 0.41 for the D12ATA63 marker) were also applied (see Table S1). Coverage analysis was carried out with the Coverage Analysis v 5.6.0.1 plugin, which provides statistics and graphs describing the level of sequence coverage produced for targeted regions. Information about mapped reads, on-target percentage, mean depth, and uniformity of coverage were downloaded for each sample library (Barcode Summary Report file). The resulting Excel files were then used for data analysis.

2.5.1. Data analysis and genotyping

The relative depth of coverage (rDoC) of markers was calculated in each sample as the ratio between the mapped reads for a specific marker and the total mapped reads of the sample [18,35]. The rDoC values were used for comparison between samples with an r^2 test. The data from eight high-molecular-weight DNAs, run in an Ion 520 Chip during a training test performed before this study, were used for comparison (in this case, 1 ng of each DNA was amplified through 24 cycles of PCR). The average molecular weight (mw) of each of the autosomal STR markers was computed as follows: (mw of the shortest amplicon + mw of the longest amplicon) / 2.

Because the minimum depth of coverage to assign a genotype depends on NGS technology and on the end-point of the study [17,36], a fixed threshold value of 30× coverage was used for the locus call in this study. As done for the PCR/CE method, the genotyping data from the duplicate PCR/NGS tests were used to build both composite [31] and consensus [32] profiles.

2.6 Calculations and graphs

Excel was used for calculation and graphs. Significance for the *t*-test was assumed with p-values < 0.05.

3. Results and discussion

DNA was extracted from 57 femurs from the Konfin I Mass Grave using the Biorobot EZ1 device (Qiagen). The main molecular features of these samples are described as follows.

3.1. DNA quantification

The PowerQuant kit (Promega) was used for quantification of the DNA obtained from 57 samples of bones from the Second World War. Each sample was assessed in duplicate, and the average results are reported in Table S2, which shows the results of DNA quantification of Auto, Deg, and Y target, IPC shift value, and degradation index ([Auto]/[Deg] ratio).

Skeletal remains found in soil often contain PCR inhibitors such as humic acids and calcium chloride [37], and therefore PCR failure can unfortunately occur. The average

IPC shift value was −0.298±0.299 (median value = −0.33), thus indicating that PCR inhibitors were removed efficiently by the magnetic particles [38]. The only exception was sample D with an IPC shift value of 0.43, slightly above the threshold fixed at 0.3.

In this study, the lowest LOD (Limit of Quantification) used was 3.2 pg/µl for both the Auto and the Y targets [29], whereas it was 0.5 pg/µl for the Deg target [39]. In total, 56/57 samples showed the Auto target within the LOQ (mean value = 32.3±31.1 pg/µl; median value = 4.3; min = 2.6; max = 181.0). For each sample, the quantity of DNA obtained from 1 g of bone sample was calculated from the Auto target and expressed in nanograms of DNA/g of bone. The average recovery was 1.62±1.55 ng DNA/g of bone (median value = 1.22; min = 0.13; max = 9.06), thus highlighting the enormous variability (up to 70 times) that exists from sample to sample.

The Deg target was < the lowest LOQ in eleven samples, whereas it was not amplified at all in five samples. Only for calculations, in these five cases, the [Auto]/[Deg] ratio was arbitrarily fixed at 98 (e.g., mean + 3 standard deviations of the ratios of the remaining 52 samples).

With the amplification of the Y target, the presence of male DNA was shown in all bone samples, with an [Auto]/[Y target] ratio of 9.1±8.1, a discrepancy resulted most likely due to the higher molecular weight of the Y target [40,41]. When analyzing ENC and PCR-negative controls, no Auto, Deg, or Y targets were amplified.

3.2 PCR/CE typing

A first round of PCR was performed on all 57 samples. Out of these, 40 full profiles and two almost full (23/24 loci) profiles were obtained. Out of the remaining 15 samples, no STR allele was scored in three cases, whereas partial profiles (with 2–17 autosomal STR markers; median value = 8) were scored in 12 cases (see Table S3). The comparison of the 42 samples with full or almost full profiles with the 15 samples with no or partial profiles showed that the first had a degradation index slightly lower (30.8 vs. 41.8; *p*-value = 0.188), and overall that much more template was used for amplification (0.595 ng vs. 0.179 ng; *p*-value = 6.1×10^{-6}). This last finding provides the most likely explanation for the different outcomes of the two sets of samples [9].

Following the experiment design, replicate STR/CE tests were carried out on 38 samples (37 full profiles as well as sample E, which showed the D2S1338 locus out). For each of the 38 samples, both the composite [31] and the consensus [32] profiles are shown in

Table S3, and Figure S1 summarizes the results of the typing. The composite method showed heterozygous typing in 17.5 loci per sample (median value; max = 21; min = 11), whereas the consensus method showed heterozygous typing in 14.0 loci per sample (median value; max = 18; min = 3), corresponding to a mean heterozygosity per sample of 0.807 and 0.654, respectively. On the other hand, the different values provided by the two methods (p-value = 2.6 x 10^{-6}) is due to the drop-out phenomena that were not computed in the composite one. Only ten out of 38 samples showed the same typing using the two different methods. Of interest is that these ten samples had a lower degradation index (13.9 vs. 30.9; p-value = 0.010) and that higher amounts of such templates were used for the amplifications (0.98 ng vs. 0.51 ng; p-value = 0.037).

Figure S2 shows the behavior of the 21 autosomal STR markers when typed by the composite and consensus methods, respectively. The typing of low-molecular-weight markers such as D8S1179, D3S1338, D10S1248, D1S1656, D22S1045, and D2S441 was not or was scarcely affected by the method used, whereas the typing of high-molecular-weight loci such as CSFP01, D7S80, D2S1338, TOPX, D16S539, and SE33 was influenced by more stringent condition of the consensus method. For example, out of a total of 38 samples, the number of heterozygous typing of the CSFP01 locus passed from 23 (with the composite method) to 11 (with the consensus method). Only in two cases were more than two alleles per locus scored. In total, no consensus profile was achieved for 12 loci only.

For disaster victim identification (DVI) it has been estimated [42,43] that at least 16 STR markers need to be reliably typed. In addition, because PCR artefacts usually lead to loss of genetic information (due to allelic drop-outs), loci typed as heterozygous are considered more reliable. The employment of duplicate PCR/CE tests allowed heterozygous typing in $\geq$ 16 loci in 12/38 (31.5%) samples by the consensus method and in 31/38 (81.5%) samples by the composite method. It seems obvious, however, that genotypes showing a lack of consistency between the two profiling methods must be evaluated with caution and that potential matches (with victims' relatives) must be reviewed carefully by experts even considering the possibility of performing the last replicate from the remaining available aliquot of 15 µl. Finally, even if it is time-consuming and expensive, one more chance is to re-extract the bone sample, as already suggested [5–8].

3.3. PCR/NGS analysis

The Precision ID GlobalFiler NGS STR Panel (TFS) was used to analyze 19 bone samples in duplicate at the suggested conditions of 24 cycles of PCR [33]. Out of these 19 samples, 15 samples yielded no or partial profiles after the first PCR/CE round (challenging samples). The remaining four samples (namely, samples 5, 12, 34, and 50) yielded full profiles, and they were therefore used as controls. The results are treated as follows.

3.3.1 Sequencing data

The principal parameters of the two chips run in this study are reported in Table S4 (which also contains data on the Ion 520 Chip run with eight un-degraded samples). Out of the addressable wells, only 31.4% showed ISPs (Ion Sphere Particles), with more than 98.2% of them represented by the libraries. About 23.5% of the ISPs showed polyclonality. Adaptamer dimers represented on average 20.5% of the sequences. The final library ISP % values were 18.3 and 24.7, respectively. In summary, all these data represent normal findings when dealing with suboptimal samples [16,33,44].

As reported in Table S5, eight bone samples were loaded in the Ion Chip 520, whereas the Ion Chip 530 contained the duplicates of these eight samples, both duplicates of the remaining eleven bone samples as well as the NECs. On average, no difference was found in the eight samples run in the two chips regarding the number of mapped reads, the percentage on target, the mean depth of coverage, and the uniformity of coverage (*p*-values ≥ 0.237). When compared to the un-degraded DNA samples, the 19 bone samples yielded, on average, fewer mapped reads (72,894 vs. 367,468), a lower percentage of on-target reads (66.8% vs. 87.5%), a lower mean depth of coverage (1,368 vs. 8,842), and lower uniformity of coverage (79.3% vs. 97.2%; *p*-values $\leq 4.9 \times 10^{-5}$; see Table S6). In addition, slight differences could be found even between the challenging samples and the four controls. The first, in fact, showed fewer mapped reads and lower mean depth of coverage and uniformity (*p*-values from 0.027 to 0.039). All bone samples, however, showed low replicability, as indicated by the r^2 values computed from the 19 duplicate analyses. The r^2 values ranged from −0.010 to 0.588 (median value −0.010), whereas this value was 0.884 in the unique duplicate performed from the un-degraded controls. These last data are in agreement with the performance of PCR/NGS technologies from degraded samples [35,44].

The rDoC of each of the 31 autosomal STR markers (sorted by molecular weight of the amplicons) is reported in Figure 1A, which shows the results for both the bone samples and the un-degraded controls. In agreement with previous findings [20-23], the coverage of the markers is not linked to the molecular weight of the amplicons, but instead it reflects the peculiar amplification efficiency of each marker. It is of interest that the rDoC of each locus in the bone samples follows the same trends of the controls (see Figure 1B), thus indicating that the kit design fits the aim of the forensic laboratory well in the analysis of degraded samples. In addition, it has to be noted that, whereas the rDoC of markers showed low CV % values in the un-degraded controls (median value = 7.8%; min = 2.1% (in D2S1776); max = 23.8% (in Penta E)), the CV % values were always high in the bone samples (median value = 71.6%; min = 42.3% (in D2S441); max = 120.5% (in Penta D)). However, these results are believed to be in agreement with the sample-to-sample variability that is common in degraded samples when amplified *via* PCR [9,35,44].

3.3.2 Genotyping

The sequencing results of all the samples are reported in Table S7. In total, because each sample should provide data on 31 autosomal STR markers, data on 1,178 markers were expected. Nevertheless, because 62 markers yielded no reads and 12 markers showed a coverage < 30× (i.e., the threshold value for the locus call), the statistics reported here are based on data from 1,104 markers (the sex-specific markers amelogenin, SRY, rs2032678 (InDel) and DYS391 are treated separately at the end of this section).

For each of the 19 samples, the results of both the composite and consensus profiles are reported in Table S8, and Figure 2 shows the results of the typing. As shown in Tables 2A and B, differences were found between the 15 challenging samples and the controls. The first showed a higher number of loci that were not typed at all (p-value $\leq$ 0.026). Specifically, one and nine loci per sample were not typed by the composite and the consensus method, respectively (vs. zero and two loci per sample of the controls). In addition, a higher frequency of homozygous typing was found in the challenging samples when the composite method was used (p-value = 0.027). Of interest is that the two sample sets did not differ regarding the amount of template used for the PCR amplification (p-value = 0.388) and the degradation ratio (p-value = 0.192).

Importantly, when comparing the two profiling methods, differences also arose. In particular, as reported in Table S9, the consensus method showed higher frequencies of

homozygous typing both in the 15 challenging samples (p-value = 1.3×10^{-4}) and in the controls (p-value = 0.043). Even if part of the STR markers of the kit (in particular the Non-CODIS ones) show low degrees of heterozygosity, our results clearly show that the use of the consensus method led to an excess of homozygosity due to drop-out phenomena. It is therefore obvious that this potential lack of genetic information needs to be taken into account when using such profiles in kinship comparisons with the victims' relatives. As reported in Tables 2A and B, the finding of more than two alleles per locus was quite common because up to eight markers per sample could exhibit this undesirable pattern when the composite method was used in both set of samples. However, the number of these multi-allelic patterns (due to drop-in phenomena) decreased dramatically by using the consensus method (p-values $\leq 1.2 \times 10^{-5}$). As reported in Table 1, in fact, a tri-allelic pattern was gathered in only four cases (two at the D16S539 locus, one at the FGA locus, and one at the TH01 locus, respectively). Regarding the drop-ins found here, they represent stutters above the thresholds as well as other kinds of length artefacts [45] that cannot be distinguished from true alleles because of their high coverage. For example, the 12.2 allele at the D10S1248 marker reached up to 0.35 of the coverage of the locus, which is far higher than the analytical threshold of 0.12 [46] applied here (see Figure S3). On the other hand, however, the spurious origin of these PCR products is suggested by the data from the PCR/CE technology, which never found allele 12.2 in D10S1248 markers successfully typed in the course of this study (n= 85). In about 25 to 30% of the cases, the drop-ins were instead represented by isometric artefacts [22,47], which could even exhibit different sequences between the duplicates (see Table S10).

The behavior of the 31 autosomal STR markers typed by the composite and the consensus method, respectively, is reported in Figure 3, which shows that they responded to the two methods differently. In particular, certain markers (such as D2S1338, D1S1656, Penta E and D22S1045, for example) were found to show a higher frequency of no typing when using the consensus method. It is of interest that these markers usually showed low coverage and high sample-to-sample variability (see Figure 1A), thus providing a molecular explanation for the corresponding genetic outcome.

As reported in Table S8, the sex-determining markers Amel. X, Amel. Y, SRY, and rs2032678 (InDel) showed no reads in 2/38, 6/38, 4/38, and 14/38 cases, respectively. These locus-out phenomena mainly occurred in samples G and H, two of the three samples that did not yield results with PCR/CE analysis. Interesting drop-in phenomena

were found in one of the duplicates of sample 5, which showed three amplicons. As shown in Figure S4, in addition to the wild sequence Amel.-[G]1 (of 2,004×), the artefactual Amel.-A[G]1 (of 255×) and Amel.-[G]1 TA (of 250×) were also scored. Locus DYS391 dropped out in 10/38 tests, and consensus typing was achieved in twelve cases (2/4 in the controls).

The sterility of tools and plastic equipment and the purity of extraction and purification reagents were examined with the help of ENC and PCR-negative controls. When analyzing these controls with qPCR and PCR/CE technology, no amplification was detected. In the same manner, ENC tested by PCR/NGS proved to be uncontaminated. In spite of the strict precautions to prevent contamination, however, the results of samples B and G are believed to be due to contamination that occurred during the NGS procedure, as suggested by the very poor replicability of the duplicates and the overall high number of alleles per locus (see Figure S5).

In conclusion, out of the 15 challenging bones samples tested using NGS technology, two samples (namely B and G) yielded inconclusive results because of a contamination issue, the analysis of sample H was judged inconclusive because no more than three loci provided a consensus profile, two samples (namely I and 35) provided partial genetic information on eight to 12 loci, and the remaining 10 samples provided genetic information on 20 to 31 loci (see Table 1). The four control samples yielded consensus profiles on 26 to 31 loci

3.4 Combining PCR/CE and PCR/NGS data

As shown in Table S11, the data from the PCR/NGS analysis were compared with those of the PCR/CE analysis for all but samples B, G, and H. Out of the 277 alleles scored in the CE chromatograms, 257 (92.8 %) were found in at least one of the two corresponding PCR/NGS duplicates, with a significant difference (p= 0.012) between the four control samples (97.7 %) and the challenging ones (83.1 %). With regard to the 20 alleles scored only by PCR/CE technology, in 16 cases they were found in loci for which PCR/NGS yielded no amplicons (n = 3) and only a unique allele/locus (n = 13), respectively. Therefore, these 16 amplicons could reflect the original genetic features of the samples. In the remaining four cases, instead, the PCR/CE amplicons were found in loci for which PCR/NGS yielded two or three alleles/locus, thus suggesting that an artefactual origin is more likely for these four amplicons.

Table 1 summarizes the results of the consensus profile gathered from the duplicate PCR/NGS analysis plus the single PCR/CE analysis (see Table S11 for details). In total, this approach allowed the recovery of 49 alleles (on average three alleles per sample; min = 0; max = 10) that were not reported after PCR/NGS data analysis because they were scored only once. In all these cases, in fact, the additional information from the PCR/CE analysis made it possible to confirm the presence of the alleles questioned, which dropped out once in the PCR/NGS duplicates. Thus, with the help of PCR/CE data, 32 loci passed from homozygous typing to heterozygous, and 12 loci that yielded no result after PCR/NGS analysis were typed as homozygous (n = 7) and heterozygous (n = 5), respectively.

In conclusion, after this task, the power of the genetic information was improved in 15/16 samples, and 14 samples provided a consensus profile on more than 20 loci (median value = 29).

4. Concluding remarks

The aim of this study was to set up a workflow for STR typing of Second World War skeletal remains combining PCR/CE and PCR/NGS technologies. Mainly because of the high cost per sample of NGS technologies, only selected samples should be tackled using these means [16]. Thus, out of 57 degraded DNA samples extracted from an equal number of 75-year-old femurs, 19 samples were chosen. Of these, 15 had yielded no or a limited number of amplicons after a first PCR/CE typing with GlobalFiler kit, and four samples that had yielded full or almost full profiles were used as controls. For the NGS analysis, the GlobalFiler NGS kit was used in duplicate by amplifying - in average - 185 pg of template through 24 cycles of PCR; that is, using less DNA than recommended (1 ng) [22,23,33], but not as low to justify additional PCR cycles. All the other procedures were performed at standard conditions using the Ion Chef System, the Ion S5 sequencer, and Coverage Analysis v5.6.0.1 software [33,34,46], with a threshold of 30× for the locus call.

In general, the PCR/NGS of the 19 bone samples showed less coverage, lower percentage of on-target reads, lower uniformity of coverage, and lower replicability than those usually found in un-degraded samples. All of this, however, was expected because of the poor quality of the samples [16,35,36,44], and it is also known that DNA degradation

promotes PCR artefacts such as drop-out and drop-in phenomena [9,45]. In spite of the severe degradation of the samples, the sequencing data were sufficient to build a consensus profile for 20 to 31 loci in 14/19 samples, whereas two samples (samples I and 35) yielded very partial consensus profiles (8–12 loci) in which only homozygous patterns could be scored. Confirmation that at least part of the homozygous patterns found in the samples were artefactual comes from combining the genotyping data from the duplicate PCR/NGS tests with those of the PCR/CE tests. By using the genotyping data from the initial PCR/CE test, in fact, 32 loci passed from a homozygous pattern to a heterozygous one (see Table S11).

The excess of homozygous typing was found to be an issue here when using the consensus profile method [32]. Conversely, the employment of the less stringent composite profile method [31] included alleles scored only once and therefore potentially spurious. Therefore, when dealing with such results to perform kinship comparisons to family reference profiles, it must be borne in mind that the use of the consensus method increases the risk of a false exclusion whereas the use of the composite method could lead to a false inclusion. In agreement with ref. [43], we believe that all loci should be evaluated for consistency between the composite profile of the bone sample and the victim's reference profiles, but statistical evaluation should include only the alleles shared with the consensus profile.

Further potential sources of mistyping were the drop-ins due to stutters above the threshold values used here [46], other length artefacts [45], and isometric artefacts [22,47]. These undesirable phenomena occurred in loci whose coverage showed a broad range from 94 to 4,497 reads, thus showing that higher thresholds for "locus call" (e.g., 100 reads of coverage) could cut off only a minimal percentage (less than 5 %) of the loci affected by these artefacts. It should be considered, instead, as already done for PCR/CE technology [48], that the threshold values [46] employed for non-degraded samples can be inadequate for highly degraded samples such as those employed in this study. Finally, the performance of replicates seems to remain the one strategy for displaying these as well other PCR/NGS artefacts.

It was beyond the goal of this pilot study to assess the weight of the additional information that arises from the STR/NGS analysis of these 75-year-old bones (i.e., sequence variability within the STR motif and SNPs of the flanking regions) [16,18-24]. Rather, the aim of this study was to investigate whether the high number of markers included in

the Globalfiler NGS panel coupled with the high sensitivity of the NGS technique was able to provide genetic length data from samples for which the conventional STR/CE approach failed. In addition, because the use of this additional information requires that even the victim's reference samples have to be typed by NGS technology, this approach will be evaluated in future studies.

In conclusion, employing the combined approach described here allowed the genetic typing of 20–31/31 autosomal STR loci in 10/15 (66.6%) bone samples that yielded no or very poor results with the conventional PCR/CE approach. Therefore, even if limited to the skeletal remains found in a unique mass grave, these data should be considered very promising. In addition, the results of this study show that the kit design meets forensic laboratory requirements in dealing with degraded samples, although expert opinion plays an indispensable role in evaluating artefacts that arise from PCR/NGS technology.

Ethical standards

This research project was approved by the Medical Ethics Committee of the Republic of Slovenia (KME 0120-246/2020/3), and informed consent was obtained from individuals for inclusion in the elimination database.

Conflict of interest

The authors declare that they have no conflict of interest.

Acknowledgements

The authors would like to thank the Slovenian Government Commission on Concealed Mass Graves for its support in the exhumations of Second World War victims. This study was financially supported by the Slovenian Research Agency (project "Determination of the most appropriate skeletal elements for molecular genetic identification of aged human remains" J3-8214). The AAs wish to thank two anonymous reviewers for their helpful comments.

References

[1] I. Morild, S.S. Hamre, R. Huel, T.J. Parsons, Identification of Missing Norwegian World War II Soldiers in Karelia Russia, J. Forensic Sci. 60(4) (2015) 1104-1110.

[2] A. Ossowski, M. Diepenbroek, M. Zwolski, A. Falis, M. Wróbel et al., A case study of an unknown mass grave. Hostages killed 70 years ago by a Nazi firing squad identified thanks to genetics, Forensic Sci. Int. 278 (2017) 173-176.

[3] A. Ossowski, M. Kuś, T. Kupiec, M. Bykowska, G. Zielińska G et al., The Polish Genetic Database of Victims of Totalitarianisms, Forensic Sci. Int. 258 (2016) 258:41-49.

[4] M. Ferenc, Topografija evidentiranih grobišč (Topography of documented mass graves). In: Dežman J (ed) Poročilo Komisije Vlade Republike Slovenije za reševanje vprašanj prikritih grobišč 2005-2008. Družina, Ljubljana, 2008:7-27.

[5] I. Zupanič Pajnič, Molekularno genetska identifikacija domobranskih žrtev (Molecular genetic identification of Slovenian home guard victims), Zdrav Vestn. 77 (2008) 745-750.

[6] I. Zupanič Pajnič, B. Gornjak-Pogorelc, J. Balažic, Molecular genetic identification of skeletal remains from the Second world war Konfin I mass grave in Slovenia, Int. J. Legal Med. 124 (2010) 307-317.

[7] I. Zupanič Pajnič, T. Zupanc, J. Balažic, Ž.M. Geršak, O. Stojković et al., Prediction of autosomal STR typing success in ancient and Second World War bone samples, Forensic Sci. Int. Genet. 27 (2017) 17-26.

[8] I. Zupanič Pajnič, M. Obal, T. Zupanc, Identifying victims of the largest Second World War family massacre in Slovenia, Forensic Sci Int. 306 (2020) 110056, https://doi.org/10.1016/j.forsciint.2019.110056.

[9] R. Alaeddini, S.J. Walsh, A. Abbas, Forensic implications of genetic analyses from degraded DNA-a review, Forensic Sci. Int. Genet. 4 (2010) 148–157.

[10] E. Zietkiewicz, M. Witt, P. Daca, J. Zebracka-Gala, M. Goniewicz et al., Current genetic methodologies in the identification of disaster victims and in forensic analysis, J. Appl. Genet. 53 (2012) 41-60.

[11] J.A. Irwin, R.S. Just, O.M. Loreille, T.J. Parsons, Characterization of a modified amplification approach for improved STR recovery from severely degraded skeletal elements, Forensic Sci. Int. Genet. 6 (2012) 578-587.

[12] S. Pääbo, H. Poinar, D. Serre, V. Jaenicke-Després, J. Hebler et al., Genetic Analyses from Ancient DNA, Annu. Rev. Genet. 38 (2004) 645-679.

[13] 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.

[14] G. Quatrehomme, S. Toupenay, T. Delabarde, B. Padovani, V. Alunni, Forensic answers to the 14th of July 2016 terrorist attack in Nice, Int. J. Legal Med. 133 (2019) 277-287.

[15] E. Pilli, S. Boccone, A. Agostino, A. Virgili, G. D'Errico et al., From unknown to known: Identification of the remains at the mausoleum of fosse Ardeatine, Science & Justice, 58 (2018) 469-478.

[16] C. Børsting, N. Morling, Next generation sequencing and its applications in forensic genetics, Forensic Sci. Int. Genet. 18 (2015) 78–89.

[17] S. Goodwin, J.D. McPherson, W.R. McCombie, Coming of age: Ten years of nextgeneration sequencing technologies, Nat. Rev. Genet. 17 (6) (2016) 333–351.

[18] J.D. Churchill, S.E. Schmedes, J.L. King, B. Budowle, Evaluation of the Illumina(®) Beta Version ForenSeq™ DNA Signature Prep Kit for use in genetic profiling, Forensic Sci. Int. Genet. 20 (2016) 20-29.

[19] A.L. Silvia, N. Shugarts, J. Smith, A preliminary assessment of the ForenSeq™ FGx System: next generation sequencing of an STR and SNP multiplex, Int. J. Legal Med. 131 (2017) 73-86.

[20] Z. Wang, D. Zhou, H. Wang, Z. Jia, J. Liu J et al., Massively parallel sequencing of 32 forensic markers using the Precision ID GlobalFiler™ NGS STR Panel and the Ion PGM™ System, Forensic Sci. Int. Genet. 31 (2017) 126-134.

[21] L. Hui, Z. Xueying, M. Ke, C. Yu, Z. Huaigu et al., Applying massively parallel sequencing to paternity testing on the Ion Torrent Personal Genome Machine, Forensic Sci. Int. Genet. 31 (2017) 155-159.

[22] P. Müller, A. Alonso, P.A., Barrio, B. Berger, M. Bodner et al., Systematic evaluation of the early access applied biosystems precision ID Globalfiler mixture ID and Globalfiler NGS STR panels for the ion S5 system, Forensic Sci. Int. Genet. 36 (2018) 36:95-103.

[23] R. Tao, W. Qi, C. Chen, J. Zhang, Z. Yang et al., Pilot study for forensic evaluations of the PrecisionIDGlobalFiler™ NGSSTR Panel v2 with the Ion S5™ system, Forensic Sci. Int. Genet. 43 (2019) 102147.

[24] W. Parson, L. Gusmão, D.R. Hares, J.A. Irwin, W.R. Mayr et al., DNA Commission of the International Society for Forensic Genetics: Revised and extended guidelines for mitochondrial DNA typing, Forensic Sci. Int. Genet. 13 (2014) 134-142.

[25] S. Amory, R. Huel, A. Bilić, O. Loreille, T.J. Parsons, Automatable full

demineralization DNA extraction procedure from degraded skeletal remains, Forensic Sci. Int. Genet. 6 (2012) 398-406.

[26] N. Rohland, M. Hofreiter, Ancient DNA extraction from bones and teeth, Nat. Protoc. 2 (2007) 1756-1762.

[27] I. Zupanič Pajnič, Extraction of DNA from human skeletal material, in: Goodwin W (Ed), Forensic DNA Typing Protocols, Methods Mol. Biol., 1420 (2016) 89-108.

[28] Qiagen Companies, EZ1 DNA Investigator Handbook, 2014, Hilden.

[29] Promega Corporation, PowerQuant System. Technical Manual, 2019, Madison, WI.

[30] Thermo Fisher Scientific, GlobalFiler PCR amplification kit. User Guide, 2016, Waltham, MA.

[31] J-A. Bright, P. Gill, J. Buckleton, Composite profiles in DNA analysis, Forensic Sci. Int. Genet. 6(3) (2012) 317-321.

[32] P. Taberlet, S. Griffin, B. Goossens, S. Questiau, S. Manceau et al., Reliable genotyping of samples with very low DNA quantities using PCR, Nucleic Acids Res. 24 (1996) 3189–3194.

[33] Thermo Fisher Scientific, Precision ID GlobalFiler™ NGS STR Panel v.2 with the Ion S5™ System. Application Guide, 2017, Waltham, MA.

[34] Thermo Fisher Scientific, Converge Software. Setup and Reference Guide, November 2017, Waltham, MA.

[35] P. Fattorini, C. Previderè, I. Carboni, G. Marrubini, S. Sorçaburu-Cigliero et al., Performance of the ForenSeqTM DNA Signature Prep kit on highly degraded samples, Electrophoresis 38 (2017) 1163–1174.

[36] M. Eduardoff, C. Santos, M. de la Puente, T.E. Gross, M. Fondevila et al., Inter-laboratory evaluation of SNP-based forensic identification by massively parallel sequencing using the Ion PGMTM, Forensic Sci. Int. Genet. 17 (2015) 110-121.

[37] B.E. Krenke, N. Nassif, C.J. Sprecher, C. Knox, M. Schwandt et al., Developmental validation of a real-time PCR assay for the simultaneous quantification of total human and male DNA, Forensic Sci. Int. Genet. 3 (2008) 14-21.

[38] R. Kishore, W.R. Hardy, V.J. Anderson, N.A. Sanchez, M.R. Buoncristiani, Optimization of DNA extraction from low-yield and degraded samples using the biorobot EZ1 and biorobot M48, J. Forensic Sci. 51 (2006) 1055-1061.

[39] M.M. Ewing, J.M. Thompson, R.S. McLaren, V.M. Purpero, K.J. Thomas et al., Human DNA quantification and sample assessment: Developmental validation of the

PowerQuant system, Forensic Sci. Int. Genet. 23 (2016) 166-177.

[40] M.J. Brisco, S. Latham, P.A. Bartley, A.A. Morley, Incorporation of measurement of DNA integrity into qPCR assays, BioTechniques 49 (2010) 893–897.

[41] P. Fattorini, C. Previderè, S. Sorçaburu-Cigliero, G. Marrubini, M. Alù et al., The molecular characterization of a depurinated trial DNA sample can be a model to understand the reliability of the results in forensic genetics, Electrophoresis 35 (2014) 3134–3145.

[42] M. Prinz, A. Carracedo, W.R. Mayr, N. Morling, T.J., Parsons TJ et al. DNA Commission of the International Society for Forensic Genetics (ISFG): Recommendations regarding the role of forensic genetics for disaster victim identification (DVI), Forensic Sci. Int. Genet. 1 (2007) 3-12.

[43] A. Sozer, M. Baird, M. Beckwith, B. Harmon, D. Lee et al., Guidelines for mass fatality DNA identification operations, AABB, Bethesda, 2010.

[44] E. Salata, A. Agostino, I. Ciuna, S. Wootton, L. Ripani L. et al., Revealing the challenges of low template DNA analysis with the prototype Ion AmpliSeq™ Identity panel v2.3 on the PGM™ Sequencer, Forensic Sci. Int. Genet. 22 (2016) 25-36.

[45] S. Pääbo, D. Irwin, A. Wilson, DNA damage promotes jumping between templates during enzymatic amplification, J. Biol. Chem. 265 (1990) 4718–4721.

[46] Thermo Fisher Scientific, Technical note. Performance of the Precision ID GlobalFiler™ NGS STR Panel v2: Artifacts, Thresholds and Chip Loading, TF, 06 October, 2019, Waltham, MA.

[47] P. Fattorini, B. Gornjak Pogorelc, S. Sorçaburu-Ciliero, T. Zupanc, C. Previderè et al., MPS reveals isometric PCR artefacts in degraded samples, Forensic Sci. Int. Genet. (Suppl. Series) 7 (2019) 368-369.

[48] P. Gill, H. Haned, O. Bleka, O. Hansson,G. Dørum, T. 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.

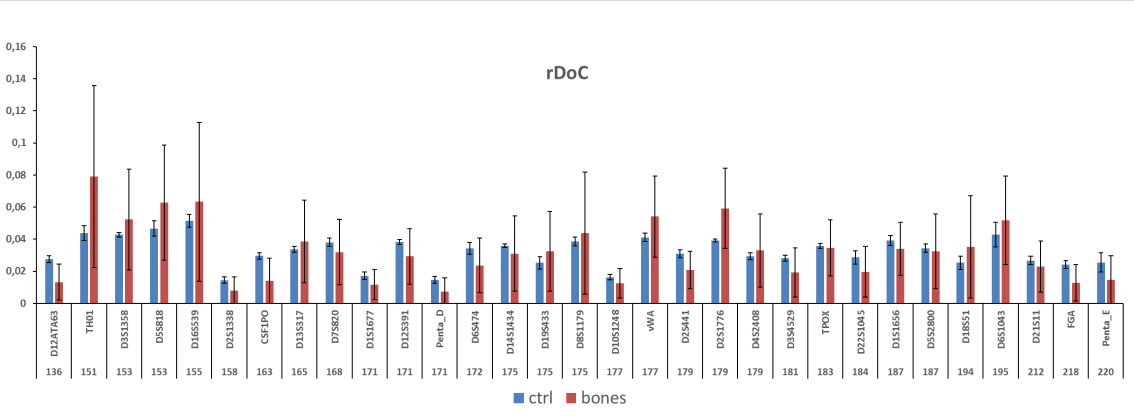

Figure 1A. Relative depth of coverage (rDoC) of the 31 autosomal STR markers computed in the 19 bone samples and in the eight un-degraded controls (control). The error bars represent the standard deviation. The data are sorted by the average molecular weight of the amplicons.

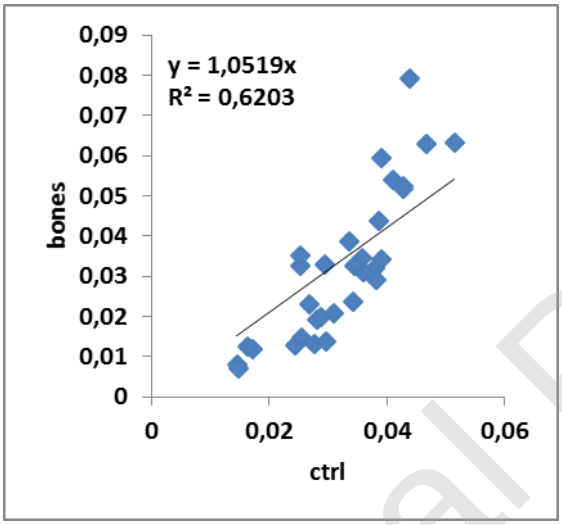

Figure 1B. Comparison of the average rDoC of each of the 31 STR autosomal markers between the un-degraded controls (*x*-axis) and the 19 bone samples (*y*-axis).

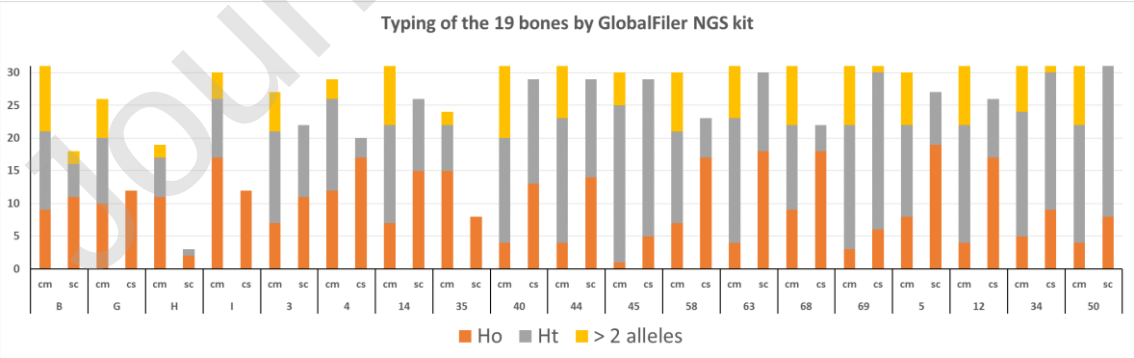

Figure 2. Autosomal STR typing of each of the 19 bones from the Second World War by the composite (cm) and consensus (cs) methods. Ho = homozygous; Ht = heterozygous; *y*-axis = number of markers. The four controls (samples 5, 12, 34, and 50) are on the right.

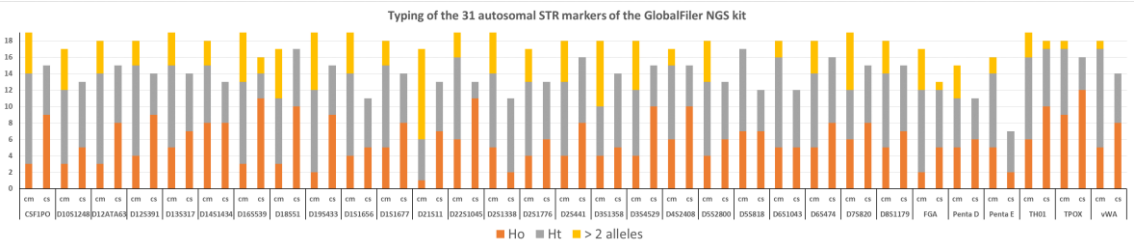

Figure 3. Typing of each of the 31 autosomal STR markers by composite (cm) and consensus (cs) method, respectively. Ho = homozygous; Ht = heterozygous; *y*-axis = number of samples.

Sample	CE	NGS	NGS+CE
B	0/21	11, 5 (2)	n.p.
G	0/21	12, 0	n.p.
H	0/21	2, 1	n.p.
I	3/21	12, 0	13, 1
3	17/21	10, 11	8, 14
4	5/21	17, 3	17, 3
5*	21/21	19, 8	13, 16
12*	21/21	17, 9	14, 15
14	3/21	15, 11	14, 12
34*	20/21	9, 21 (1)	8, 22 (1)
35	6/21	8, 0	10, 1
40	12/21	13, 16	10, 19
44	8/21	14, 15	11, 18
45	14/21	5, 24	3, 26
50*	21/21	8, 23	5, 26
58	8/21	17, 6	15, 8
63	12/21	18, 12	16, 15
68	7/21	18, 4	19, 4
69	13/21	6, 24 (1)	5, 25 (1)

Table 1. Samples used for NGS analysis. CE = number of autosomal STR markers that yielded amplicons after the first PCR/CE test. NGS = results of the NGS analysis (autosomal STR consensus typing). For each sample, the number of loci typed as homozygous, heterozygous, and showing more than two alleles (in brackets) is given. NGS+CE = autosomal STR consensus typing combining NGS and CE technologies. n.p. = not performed (see text and Table S11 for details). * indicates samples used as controls in the course of the study.

	15 challenging samples				ctrl (samples 5, 12, 34 and 50)			
	no typing	**Ho**	**Ht**	**>2 alleles**	**no typing**	**Ho**	**Ht**	**>2 alleles**
median	1.0	7.0	14.0	8.0	0	5.0	18.0	8.5
max	12	17	24	11	1	8	19	9
min	0	2	6	2	0	4	14	7
mean ± s.d.	2.2 ± 3.5 (0.071)*	8.0 ± 4.5 (0.292) §	14.1 ± 4.9 (0.480) §	6.7 ± 2.9 (0.227) §	0.3 ± 0.5 (0.008)*	5.3 ± 1.9 (0.171) §	17.3 ± 2.2 (0.560) §	8.3 ± 1.0 (0.268) §
p-value	0.026	0.027	0.063	0.077	/	/	/	/

Table 2A.

	15 challenging samples				ctrl (samples 5, 12, 34 and 50)			
	no typing	**Ho**	**Ht**	**>2 alleles**	**no typing**	**Ho**	**Ht**	**>2 alleles**
median	9.0	12.0	6	0	2.0	13.0	15.0	0
max	28	19	24	2	5	19	23	1
min	0	2	0	0	0	8	8	0
mean ± s.d.	10.1 ± 8.8 (0.324)*	12.0 ± 5.0 (0.646) §	8.7 ± 8.2 (0.344) §	0.2 ± 0.6 (0.009) §	2.3 ± 2.6 (0.072)*	13.3 ± 5.6 (0.476) §	15.3 ± 7.8 (0.515) §	0.3 ± 0.5 (0.008) §
p-value	0.004	0.133	0.125	0.447	/	/	/	/

Table 2 B.

Tables 2A and B. NGS typing of the samples by the composite (2A) and consensus (2B) method. The values refer to the number of markers per sample. In brackets are the frequencies that were calculated on the total number of markers (*) and on the number of typed loci (§), respectively. Ho = homozygous typing; Ht = heterozygous typing; *p*-value = result of the *t*-test comparison with the control samples (samples 5, 12, 34, and 50).