

PAPER

Criminalistics

Validation and beyond: Next generation sequencing of forensic casework samples including challenging tissue samples from altered human corpses using the MiSeq FGx system

Alina Senst MSc¹ | Amke Caliebe PhD² | Eva Scheurer MD¹ | Iris Schulz PhD¹ 

¹Institute of Forensic Medicine, Basel, Switzerland

²Institute of Medical Informatics and Statistics, Kiel University and University-Hospital Schleswig-Holstein, Kiel, Germany

Correspondence

Iris Schulz, Institute of Forensic Medicine, University Basel, Health Department Basel-Stadt, Pestalozzistrasse 22, 4056 Basel, Switzerland.
Email: iris.schulz@bs.ch

Abstract

The proceeding developments in next generation sequencing (NGS) technologies enable increasing discrimination power for short tandem repeat (STR) analyses and provide new possibilities for human identification. Therefore, the growing relevance and demand in forensic casework display the need for reliable validation studies and experiences with challenging DNA samples. The presented validation of the MiSeq FGx system and the ForenSeq™ DNA Signature Prep Kit (1) investigated sensitivity, repeatability, reproducibility, concordance, pooling variations, DNA extraction method variances, DNA mixtures, degraded, and casework samples and (2) optimized the sequencing workflow for challenging samples from human corpses by testing additional PCR purification, pooling adjustments, and adapter volume reductions. Overall results indicate the system's reliability in concordance to traditional capillary electrophoresis (CE)-based genotyping and reproducibility of sequencing data. Genotyping success rates of 100% were obtained down to 62.5 pg DNA input concentrations. Autosomal STR (aSTR) profiles of artificially degraded samples revealed significantly lower numbers of locus and allelic dropouts than CE. However, it was observed that the system still exposed drawbacks when sequencing highly degraded and inhibited samples from human remains. Due to the lack of studies evaluating the sequencing success of samples from decomposed or skeletonised corpses, the presented optimisation studies provide valuable recommendations such as an additional PCR purification, an increase in library pooling volumes, and a reduction of adapter volumes for samples with concentrations ≥ 31.2 pg. Thus, this research highlights the importance of all-encompassing validation studies for implementing novel technologies in forensic casework and presents recommendations for challenging samples.

KEYWORDS

altered human corpses, DNA degradation, ForenSeq™ DNA Signature Prep Kit, Illumina MiSeq FGx system, next generation sequencing, NGS, validation

Presented in part at SGRM Sommertagung 2021, September 4, 2021, in Arlesheim, Switzerland.

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2022 The Authors. *Journal of Forensic Sciences* published by Wiley Periodicals LLC on behalf of American Academy of Forensic Sciences.

Highlights

- Due to a growing demand in forensics, validation of next generation sequencing methods is needed.
- The results show repeatable sequencing data and prove the MiSeq FGx to be robust and reliable.
- The MiSeq FGx shows drawbacks in performance when sequencing degraded and inhibited samples.
- The study presents valuable recommendations and experiences in sequencing challenging samples.
- Additional PCR purification and pooling adjustments are recommended for challenging samples.

1 | INTRODUCTION

Forensic DNA analysts are often confronted with low DNA quantities, mixtures of multiple contributors, and DNA degradation. Especially tissue samples from highly altered human corpses can be challenging in terms of low DNA quantity and quality [1]. For DNA profiling, current capillary electrophoresis (CE)-based short tandem repeat (STR) genotyping has been the gold standard for several years [2–5]. Despite its widespread use, CE displays limitations regarding the required amplicon size and the inability to detect sequence variations in PCR fragments [4,6–10]. Additionally, the number of multiplexed loci is restricted due to the required labelling of similar-sized DNA fragments with different fluorescent dyes [7,8,11,12], leading to a maximal marker capacity of 25–30 loci [3,6,9]. DNA samples from traces or biological material often undergo fragmentation induced by environmental influences like pH value, humidity, temperature, acidic soil, or enzymatic activity [13,14]. Due to the resulting DNA disintegration, larger loci are less likely to be amplified than shorter ones [15–17]. With a common polymerase chain reaction (PCR) fragment range of 80–500 base pairs (bp) and a current spectral overlap of six fluorescent dyes, allele typing based on amplicon size displays a distinct limitation for degraded DNA samples [1,8].

The development of high-throughput DNA sequencing technologies offers promising approaches to advance the resolution of forensic casework samples [18]. Over the past years, massive parallel sequencing (MPS) methods, also known as next generation sequencing (NGS), have expanded the spectrum of DNA analyses providing new opportunities for sequencing the entire human genome or sequences of interest [6,18]. As demonstrated by numerous studies, STR and single-nucleotide polymorphism (SNP) genotyping with NGS reveal a high potential and growing relevance in forensic casework [1–7,11]. In contrast to CE, base-by-base sequencing detects variants in the repeat and flanking region, enhancing the discrimination power [8]. Furthermore, the possibility to multiplex autosomal, X- and Y-STRs, identity informative SNPs (iiSNPs), ancestry informative SNPs (aiSNPs), and phenotype informative SNPs (piSNPs) in a single assay is a major advantage compared with CE [6,7,19,20]. Especially the

potential to predict a person's phenotype and ancestry can aid investigative authorities within a given legal framework [11,21]. Moreover, since the DNA fragments do not have to be separated by size, the amplicon length can be reduced, which benefits the analysis of degraded samples [1,19]. Additionally, the sequencing of SNPs can provide valuable information if CE-based STR typing fails [4]. With the ForenSeq™ DNA Signature Prep Kit (Verogen) and the MiSeq FGx system (Verogen), target PCR amplification and parallel sequencing of up to 231 STR and SNP markers with most amplicon sizes of less than 200 bp were introduced [1,6,22]. The assay contains two PCR primer sets: (1) DNA Primer Mix A (DPMA) aiming for noncoding regions and (2) DNA Primer Mix B (DPMB) predicting, in addition, an individual's phenotype and biogeographic ancestry [6]. However, NGS still provides drawbacks regarding the labour-intensive workflow and costs per sample. Despite many studies concerning the applicability of sequencing technologies in forensic casework, NGS is still not implemented in many forensic laboratories [23]. Commuting the current CE-dominated technology to NGS requires time, qualified staff, and internal validation [23,24]. As stated by the Scientific Working Group on DNA Analysis Methods (SWGDM), NGS-specific studies should address the limits of detection and the quantity and quality of libraries pooled in the sequencing runs [24]. The limited sample input volume (5 μ L) [25] is a considerable restriction for the analyses, in particular of low-concentrated samples. An insufficient amount of DNA provokes an increase in the formation of adapter dimers, which can negatively impact the sequencing quality [26,27]. Especially DNA from postmortem tissue samples of altered corpses can be highly degraded and inhibited due to the decomposition process [18,28]. Identifying human remains with high postmortem intervals is a common task in forensic medicine, yet NGS technologies for STR and SNP analysis of altered tissue samples has not been widely evaluated [4]. Thus, further validation and optimisation of the workflow is necessary, especially for challenging samples. In this study, the capability of the ForenSeq™ DNA Signature Prep Kit (Verogen) on the MiSeq FGx system (Verogen) was thoroughly investigated for forensic samples following the Revised SWGDAM Validation Guidelines [24].

For investigating sensitivity, repeatability, reproducibility, and mixtures, several validation studies used human samples instead of artificial positive controls [2,5,7,9,11,21–23,29–31]. The latter builds a more solid basis for statistical analysis. However, it does not reflect the crime-scene traces of poor quality or quantity, often found in an associated criminal delict. Thus, the presented internal validation adds valuable information using human blood, saliva, and casework samples. Optimisations of the library preparation were evaluated to improve the results of challenging samples from altered corpses. Due to the higher possibility of adapter dimer formation, additional PCR purifications and, for the first time, a reduction of adapter volumes were tested to minimize the occurrence of PCR artifacts. Furthermore, varying pooling volumes were explored to increase the DNA input of low-concentrated samples and improve genotyping success. The study aimed to verify the reliability of the MiSeq FGx system with human materials, identify its limits, and beyond validation to optimize the sequencing workflow, in particular for degraded and inhibited DNA samples from altered human corpses.

2 | MATERIAL AND METHODS

2.1 | Sample collection, DNA extraction, quantification, and capillary electrophoresis

Buccal swabs were taken from four volunteers (male $n = 2$, female $n = 2$) with informed consent, with one of each gender also providing whole blood samples. Five GEDNAP (German DNA Profiling Group) proficiency test samples simulated case-type samples with known DNA results. During autopsies, samples were taken from the *musculus (M.) rectus femoris* of unaltered corpses (postmortem interval [PMI] < 24 h, $n = 2$). From altered remains ($n = 9$), *M. rectus femoris*, *M. pectoralis major*, heart, aorta, liver and lung, buccal swabs, rib fragments, *pars petrosa*, vertebra, femur, and toenails were sampled. The PMI ranged from a few days to several weeks, showing varying degrees of decomposition and skeletonisation. The remains' sampling was approved by the regional Ethical Review Board (No. 2019–02211).

Before extracting DNA from bones with the Bone DNA Extraction Kit (Promega), osseous matters were processed with a modified protocol adapted from Pajnic [32]. Genomic DNA was extracted from blood and tissue samples (100 mg each) and buccal swabs using the Maxwell® FSC DNA IQ™ Casework Kit (Promega) on the Maxwell® RSC instrument (Promega) according to the manufacturer's protocols for solid and liquid samples. For the variance study, DNA from buccal swabs, blood, and tissue samples from unaltered human remains was additionally extracted using the SwabSolution™ Kit (Promega). DNA quantification of all samples was performed on the Applied Biosystems 7500 Real-Time PCR System (Thermo Fisher) using either the PowerQuant® System (Promega) [33] or the Plexor® HY System (Promega) according to the manufacturer's protocol. If stated, the samples were

additionally amplified using the Investigator 24Plex Kit (Qiagen), followed by a fragment length analysis on the ABI Prism3500 xL Genetic Analyzer (Applied Biosystems) according to the manufacturer's protocol. All analyses included required positive and negative controls.

2.2 | Library preparation and sequencing

DNA libraries were prepared using the ForenSeq™ DNA Signature Prep Kit (Verogen) according to the ForenSeq™ DNA Signature Prep Reference Guide [25] unless otherwise noted. For target amplification, samples were amplified either in duplicates or triplicates using the DPMA (27 autosomal STRs, 24 Y-STRs, 7 X-STRs, 94 identity SNPs) and the DPMB (22 phenotypic SNPs, 56 biogeographical ancestry SNPs, and the DPMA loci), each in a reaction volume of 15 μ L. Since the DPMB primer mix also contains the DPMA primers and additionally provides a prediction of the phenotype and ancestry, only DPMB was used for the reproducibility, repeatability, mixed samples, degradation, and optimisation tests. If not otherwise specified, 1-ng DNA input concentrations were used as a template. Target enrichment, library purification, normalization, pooling, and denaturation of libraries were performed as stated in the manufacturer's protocol or with specified adjustments for the optimisation studies. Quality control of purified DNA libraries prior to sequencing was ascertained with the BioAnalyzer 2100 (Agilent) and the High Sensitivity DNA kit (Agilent). Normalized libraries were sequenced on the MiSeq FGx system using MiSeq FGx™ micro flow cells. Unless otherwise noted, the recommended maximal number of 12 pooled samples for DPMB and 36 samples for DPMA was not exceeded. Every sequencing run comprised fully loaded flow cells, including 2800 M Control DNA as positive and nuclease-free water as negative amplification controls.

2.3 | Data analysis

ForenSeq Universal Analysis Software (UAS) was used to analyze sequencing data with a default interpretation threshold of 4.5% of sequencing reads (except for DYS635 with a default interpretation threshold of 10%, DYS389II with 15%, and DYS448 with 10%). The default for the analytical threshold was at 1.5% of sequencing reads except for DYS635 (default 3.3%), DYS389II (default 5%), and DYS448 (default 3.3%). Marker coverage below the analytical threshold was considered as locus dropout (LD) or allelic dropout (AD). STR alleles between the analytical and interpretation threshold were manually called. SNPs below the interpretation threshold were called when coverage was ≥ 20 . Biogeographical ancestry prediction was obtained from the principal component analysis provided by the UAS. Statistical analyses of each run's quality metrics and sequencing data were performed using R version 4.1.1 [34] and R studio version 2021.09.0 [35]. Data distribution

was evaluated with the Shapiro–Wilk normality test, density, and Q-Q plots using the *dplyr* [36] and *ggpubr* [37] packages. For normally distributed data, linear regression models, and analysis of variances (ANOVA) and the post hoc Tukey's HSD (honestly significant difference) test were performed using the package *IpSolve* [38] and the function *aov* and *TukeyHD*. Significance was defined at $P < 0.05$, and all tests were two-sided. The total number of reads of the sensitivity study were log₂ transformed and used for the regression and ANOVA models. To measure reproducibility and repeatability, the intraclass correlation coefficient (ICC) was calculated with two-way random-effects and absolute agreement. Data visualization was carried out using the *ggplot2* [39] and *BlandAltmanLeh* [40] packages. Regression lines were plotted using the function *stat_smooth* and 95% confidence bands. CE data analyses were carried out using the GeneMapper ID-X v.1.6 Software (Applied Biosystems). A threshold of 50 relative fluorescence units (rfu) was used for allele typing.

2.4 | Sensitivity

The DNA extract of a male blood sample was quantified and serially diluted with nuclease-free water to quantities ranging from 1000 pg, 500 pg, 250 pg, 125 pg, 62 pg, 31 pg, 15 pg to 7.8 pg. Each dilution was quantified with the Plexor® HY System (Promega) for a second time to confirm the desired input concentration. Amplification was performed using DPMA and DPMB according to the manufacturer's protocol for purified lysates. The eight dilutions were sequenced in triplicates. For concordance, diluted samples were in addition CE genotyped.

2.5 | Variances, reproducibility, and repeatability

For evaluating variances between tissue types and extraction methods, DNA was extracted from two muscle samples of unaltered corpses, two buccal swabs, and two blood samples using both extraction methods. Each sample was amplified with DPMA and DPMB in duplicates with input concentrations of 1 ng using the manufacturer's protocol for purified (Maxwell® FSC DNA IQ™ Casework Kit extracts) and crude lysates (SwabSolution™ Kit extracts).

To measure the method's strength in repeatability, the same analyst sequenced DNA libraries from the study of variances for a second time. Both sequencing runs were performed within 1 week and under the same laboratory conditions. A second analyst reprocessed the same samples within a week and under the same conditions to determine reproducibility.

2.6 | Pooling variations, mixed samples, and concordance

To test the impact of varying pooling quantities of DNA libraries, replicate batches of 31, 36, and 41 samples amplified with DPMA were

pooled and sequenced on three separate flow cells. Replicate batches of 7, 12, and 17 samples were amplified with DPMB and sequenced on three flow cells. Each pool consisted of a corresponding number of replicates of libraries from the same library preparation, including DNA from two buccal swabs and two whole blood samples (Table S1).

Mixtures of quantified DNA samples from buccal swabs were prepared as follows: Trial 1 (male and female sample ratios of 1:1, 5:1, 10:1, 1:5, and 1:10), Trial 2 (male ratios of 1:1, 5:1, and 10:1), and Trial 3 (female ratios of 1:1, 5:1, and 10:1). DPMB was used for amplification, and every mixture was sequenced in duplicates.

Concordance was assessed by comparing sequencing with CE-based genotyping results. Here, DNA was extracted from two whole blood samples, two buccal swabs, and two fresh muscle samples. Each sample was amplified with DPMA and sequenced according to the reference guide [25].

2.7 | Degraded samples

For assessing the system's stability, artificially degraded samples were prepared to simulate DNA damage. Degradation was induced by exposing extracted DNA from a whole blood sample to UV light. Light exposure was conducted in intervals of 0, 10, 15, 20, and 30 min (T_0 to T_{30}) using a UV bank. After quantifying the samples with the PowerQuant Kit (Promega) and measuring the degradation index, samples were amplified with DPMB and sequenced in duplicates according to the Reference Guide [25]. Additionally, each sample was CE-based genotype. Internal quality sensors of the Investigator 24plex QS Kit were used to assess the DNA degradation.

2.8 | Workflow adjustments for challenging samples

The ForenSeq™ DNA Signature Prep Reference Guide [25] includes a library purification step using sample purification beads (SPB). For evaluating the impact of further PCR purifications, the purification was repeated at different steps of the library preparation, and an additional method from Qiagen was used. DNA extracts from a heart sample, toenail, and *pars petrosa* from decomposed human corpses were sequenced without additional purification (RE), with additional purification after target amplification (adjustment 1), with a repetition of the manufacturer's recommended purification step (adjustment 2) and with a subsequent extrapurification using the MinElute Kit (Qiagen) after the protocol's library purification step (adjustment 3). Each sample was amplified using DPMB and sequenced in duplicates. Additionally, tissue samples (*M. rectus femoris*, *M. pectoralis major*, heart, aorta, liver and lung, buccal swabs, rib fragments, *pars petrosa*, and toenails) from decomposed corpses were tested for varying amounts of pooling volumes and their impact on sequencing coverage. A low-concentrated sample (0.05 ng/ μ L) from the *M. pectoralis major* was pooled in volumes of 5, 10, and 15 μ L, with the volumes of the remaining nine tissue samples, each kept at 5 μ L.

Further, the effect of varying amounts of indexed adapters was assessed by reducing the recommended input amount from each 4 μL of index 1 (i7) and index 2 (i5), to 3 and 2 μL , respectively. DNA extract from a male blood sample was diluted 8-fold from 1000 pg to 7.8 pg. DPMB amplicons were enriched using the three different volumes for both index adapters.

To validate the results for inhibited and degraded tissue samples from altered human corpses, three bone samples with input concentrations lower than the recommended (vertebra: 581.23 pg, femur: 127.78 pg, *pars petrosa*: 754.21 pg) were each sequenced with 4, 3, and 2 μL of index 1 and index 2 adapters.

3 | RESULTS

Genotype data of 353 samples from varying tissue types were generated. Each run passed the required quality metrics and showed a mean cluster density of 1165 K/mm² (462–1501 K/mm²). From these runs, on average 92.15% (80.33%–98.59%) passed the chastity filter. Phasing and prephasing rates were below the recommended threshold ($\leq 0.25\%$ and $\leq 0.15\%$) and showed mean values of 0.15% (0.11%–0.25%) and 0.05% (0.01%–0.09%), respectively. In each run, the overall intensity of the human sequencing control (HSC) passed the minimum intensity level and genotype concordance. Unless otherwise noted, predicted phenotype and biogeographic ancestry were consistent with the individuals' descriptions, regardless of potential single allelic dropouts.

3.1 | Sensitivity

Sensitivity samples amplified with both DPMA and DPMB revealed decreasing total read intensities and increasing LDs with declining DNA input concentrations (Figure 1, Figure 2). Mean read intensities

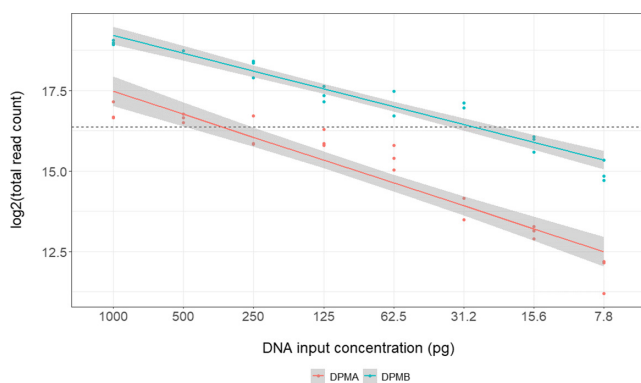


FIGURE 1 Sensitivity study. Total number of reads for DNA input concentrations ranging from 1000 to 7.8 pg amplified with DPMA and DPMB. The total number of reads is log₂-transformed. The dotted line indicates the manufacturer's read count threshold of 85,000 (log₂-transformed) [41]. Regression lines are plotted with 95% confidence bands (gray) [Color figure can be viewed at wileyonlinelibrary.com]

of samples amplified with DPMB ranged from 522,419 (1000 pg input DNA) to 32,330 (7.8 pg). Coverage below the recommended sample read count (85,000 [41]) was obtained from input concentrations ≤ 15.6 pg. A linear regression model was created with log₂-transformed total read intensities (Figure 1). Comparison of log₂-transformed read intensities' mean values showed a significant difference by both DNA input concentration ($p < 0.001$) and primer mix ($p < 0.001$). The interaction of both factors was also significant ($p < 0.001$, all P -values from two-way ANOVA). Concordant and 100% complete autosomal STR (aSTR) profiles were obtained with input concentrations down to 62.5 pg (Figure 2). Only one triplicate (62.5 pg) displayed a read count of one allele at CSF1PO below the interpretation threshold of 4.5%. With 31.2 pg, the first AD was observed at D18S51, and at a DNA level of 7.8 pg, Amelogenin, TPOX, and FGA dropped out. Complete Y-chromosomal STR profiles were obtained with input concentrations down to 62.5 pg. For X-chromosomal STRs, the first LDs occurred at 15.6 pg (DXS10135 and DXS10103).

Compared to STR loci, sensitivity results for iSNPs displayed an LD of rs2920816 at 125 pg. Even though the number of alleles below the interpretation threshold (ABITs) and imbalanced alleles decreased below 250 pg input concentrations, genotyping success rates of $\geq 93\%$ were still observed down to 15.6 pg. Except for one amplification with 31.2 pg and the loss of rs310644, piSNPs and aiSNPs exhibited initial LDs and ADs at 15.6 pg. To the latter DNA input level, the subject's phenotype was assessable, except for one amplification, in which no phenotype estimation was possible due to the LD of rs683. Ancestry was predictable for all dilutions, with a distance to nearest centroids for ancestry estimation ranging from 1.34 to 2.88 provided by the UAS.

Sensitivity results for DPMA amplicons revealed lower total read intensities than DPMB for the investigated DNA range (Figure 1). Mean read intensity for DNA input concentration of 1000 pg was 117,557, and 3863 for 7.8 pg, respectively, resulting in coverages already dropping below the recommended sample read count at 250 pg. The DPMA genotype success rate of 100% for aSTRs was obtained down to 62.5 pg, except for no read counts for CSF1PO in one amplification (data not shown). First ADs occurred with 31.2 pg input, and ABITs were already observed at 250-pg input. Y- and X-chromosomal STRs genotyping success rates were $\geq 88\%$ and $\geq 71\%$, respectively, down to 15.6 pg. For iSNPs, allele loss was already detected for input concentrations of 125 pg. Concordant CE-based analyses yielded similar results compared with NGS-based genotyping, with success rates of 100% down to 62.5 pg. The kit-specific quality sensors showed expected peak heights and confirmed successful amplification.

3.2 | Variances

For evaluating possible variances, the influence of different tissue types and extraction methods on the total read intensities were calculated. The outcome was significantly different between tissue types ($p < 0.001$, for DPMA and DPMB), whereas the extraction

FIGURE 2 Sensitivity study for DPMB. Mean amount of locus dropouts, imbalanced alleles, dropins, allelic dropouts, and alleles under the interpretation threshold for DNA input concentrations ranging from 1000 pg to 7.8 pg. For each DNA concentration, profile quality is separated in aiSNPs, aSTRs, iiSNPs, piSNPs, X-STRs, and Y-STRs

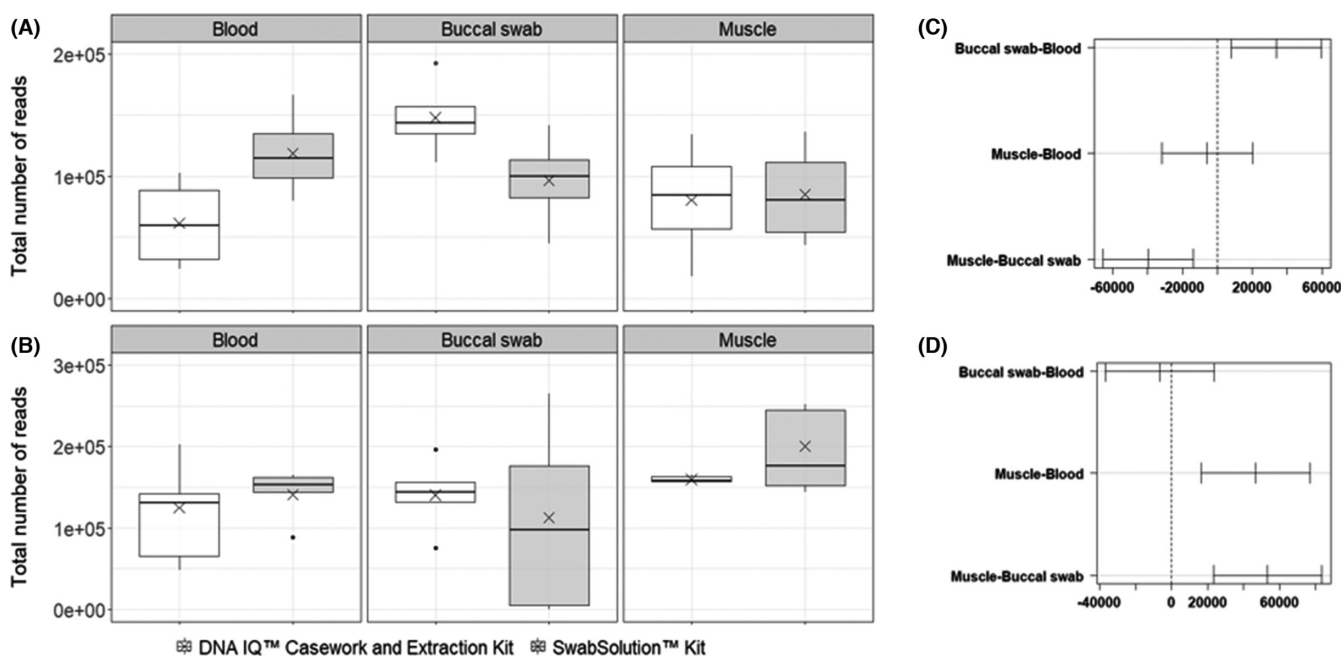
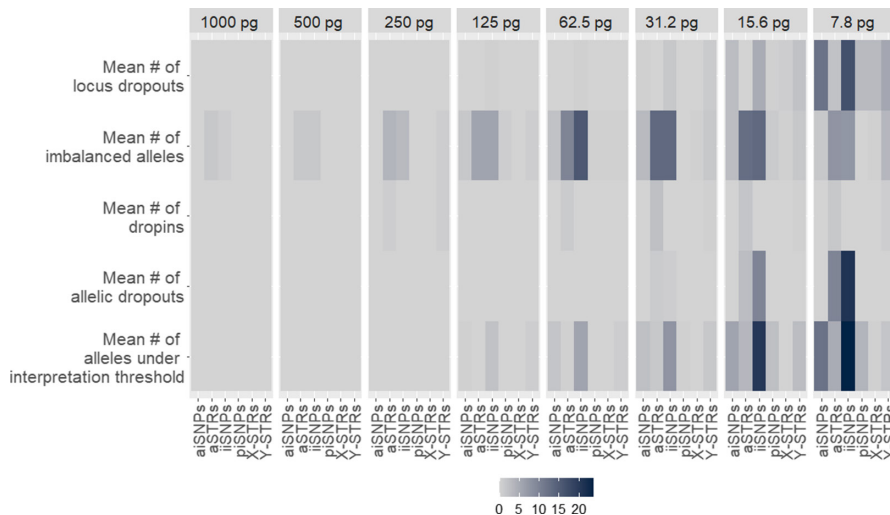


FIGURE 3 Variance study. Total read intensities for DNA extracted with the DNA IQ Casework and Extraction Kit and the SwabSolution Kit from buccal swabs, blood, and muscle samples amplified with DPMA (A) and DPMB (B). Displayed in each plot is the mean (x) and median (-). For both extraction methods, multiple pairwise comparisons of the mean difference (Tukey's honestly significant difference test) were plotted for DPMA (C) and DPMB (D). Confidence intervals for the mean value between the groups crossing the zero line indicate significant differences between groups

method showed no significant influence (DPMA: $p = 0.753$; DPMB: $p = 0.364$). For both primer mixes, a significant interaction between the two factors was found (DPMA: $p < 0.001$; DPMB: $p = 0.023$, all p -values two-way ANOVA). Therefore, each DNA extraction method and sequencing protocol for target amplification demonstrated effective removal of PCR inhibitors. DNA extracted with the SwabSolution™ Kit and amplified with DPMB had the greatest interquartile range (IQR) for DNA extracted from buccal swabs (Figure 3B). Statistically significant differences between each tissue obtained with multiple pairwise comparisons of the mean difference (Tukey's range test, data from both extraction methods were

included) are shown in Figure 3C, D. For DPMA, the confidence intervals for the mean value between the tissue groups do not cross the zero line when comparing buccal swabs with blood and muscle with buccal swabs, showing significant differences between the tissues ($p < 0.001$, $p < 0.001$, respectively). For DPMB, a comparison of muscle and blood, and muscle and buccal swabs, shows significant differences ($p < 0.001$, $p < 0.001$, respectively). Genotyping success rate and concordance of autosomal, X- and Y-chromosomal STRs amplified with DPMA was 100% for all sequenced tissue types. Profile completeness of iiSNPs obtained from blood and buccal swabs extracted with the Maxwell® FSC DNA IQ™ Casework Kit reached

98% and 97%, respectively. For the same tissue types extracted with the SwabSolution™ Kit, 99% of iiSNP loci were typed successfully (LD of rs1736442 and rs1031825, respectively). No significant differences were observed when comparing the number of ABITs. Samples amplified with DPMB showed a 100% genotyping success rate of STRs and SNPs only for blood and muscle samples extracted with the Maxwell® FSC DNA IQ™ Casework Kit. DNA samples from buccal swabs extracted with the SwabSolution™ Kit generated no or partial profiles. Due to the high number of piSNP LDs, no phenotype estimation was possible from two buccal swabs. Otherwise, phenotype and biogeographic predictions showed no differences between extraction methods.

3.3 | Repeatability and reproducibility

Agreement between both repeatability runs was measured by plotting the mean and the difference between both runs' total read intensities in a Bland–Altman plot (Figure 4A). The 95% limits of agreement were 109,330.10 reads and –328,235.10 reads, indicating low agreement. In most cases, the difference was negative, with a mean of –109,452.51 reads, showing that read intensities were higher in the repeated run. Additionally, the ICC was calculated with a two-way random-effects model to estimate the strength of agreement. The repeatability results revealed an ICC of –0.24, representing poor agreement. With differences of up to 300,000 reads, no repeatability of read intensities is given. For both runs, genotyping success rate was 100% for compared aSTRs, X- and Y-chromosomal STRs, iiSNPs, piSNPs, and aiSNPs. All targeted loci were concordant and yielded profile completeness. With no significant difference in the number of ABITs, repeatability in the profile completeness is given.

The agreement was also assessed for the reproducibility analysis by plotting mean total read intensities and the difference between the two runs (Figure 4B). The 95% limits of agreement were –63,016.89 and –425,581.31, also indicating low agreement. In all cases, the difference was negative with a mean of –244,299.12, showing that read intensities were higher in the reproduced run.

The measured ICC of 0.03 also indicates no agreement. Differences of up to 400,000 reads revealed no reproducibility in read intensities. Furthermore, genotyping, phenotype, and ancestry predictions were successful for every sample.

3.4 | Pooling variations

The manufacturer recommends a maximal number of 12 pooled samples for DPMB and 36 samples for DPMA on micro flow cells [42]. With an increased number of pooled samples, both primer mixes show decreasing total read intensities per sample (Figure 5). Significant differences were observed between the batches for DPMA ($p = 0.132$, one-way ANOVA) and DPMB ($p < 0.001$, one-way ANOVA). The Tukey's HSD p -values for significant differences between the runs' mean values are displayed in Figure 5. Differences between the batches were also observed when comparing the average read intensities of STRs and SNPs separately (Figure 6). There was no distinct decline in read intensities from the lowest to the highest number of pooled samples, except for iiSNPs. Instead, pooling 36 samples amplified with DPMA showed the highest average numbers. The most remarkable differences between the recommended number of pooled samples and the variations were observed for X-STRs. In DPMB amplified samples, the highest average numbers were obtained by pooling seven samples, with distinct differences detected for iiSNPs. However, even though the total number of reads decreased with increasing numbers of pooled samples, the genotyping success was not affected. For $n = 31$ (DPMA), only one LD (DYS389II) was observed in a sample obtained from a buccal swab. The batch of $n = 36$ revealed no LD, and $n = 41$ showed two LDs (rs1736442 and rs1031825). No ADs were detected, and with an increasing number of pooled samples, the total number of ABITs increased only slightly. All samples in batches amplified with DPMB revealed genotyping success rates of 100% with no LD, AD, and ABITs. Therefore, no impact on the phenotype and biogeographic ancestry estimation was observed.

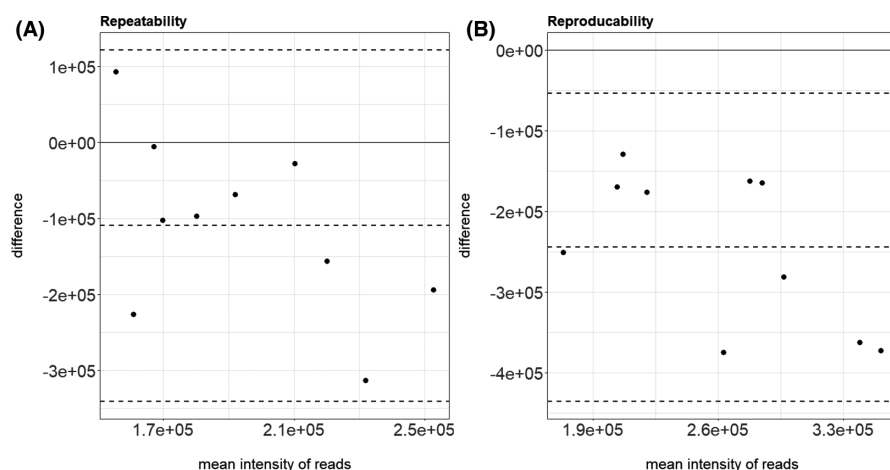


FIGURE 4 Repeatability and reproducibility study for DPMB. Bland–Altman plots for assessing repeatability and reproducibility. Shown are the mean of the total intensity of reads of both runs (x-axis) and the difference between both values (y-axis). The middle dotted line indicates the bias, and the upper and lower dotted lines indicate 95% limits of agreement

FIGURE 5 Pooling variation study for DPMB. Total number of reads for batches of 31, 36, and 41 pooled libraries (DPMA, A) and 7, 12, and 17 pooled libraries (DPMB, B). Shown are the p-values of the ANOVA analysis and pairwise p-values obtained with the Tukey's post hoc test. Displayed in each plot are the mean (x) and median (-) and

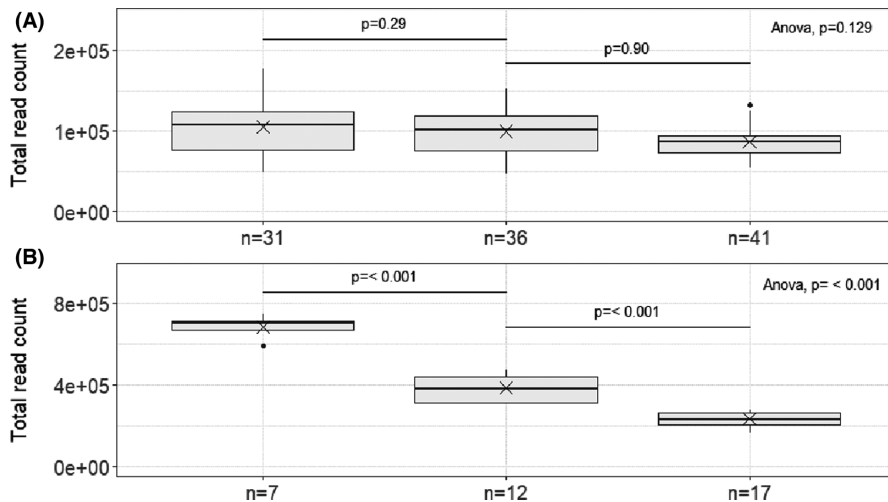


FIGURE 6 Pooling variation study. Mean number of reads for batches of 31, 36, and 41 pooled libraries (DPMA, A) and 7, 12, and 17 pooled libraries (DPMB, B) separated in Y-STRs, X-STRs, iiSNPs, aSTRs, and piSNPs/aiSNPs. Error bars indicate the standard error [Color figure can be viewed at wileyonlinelibrary.com]

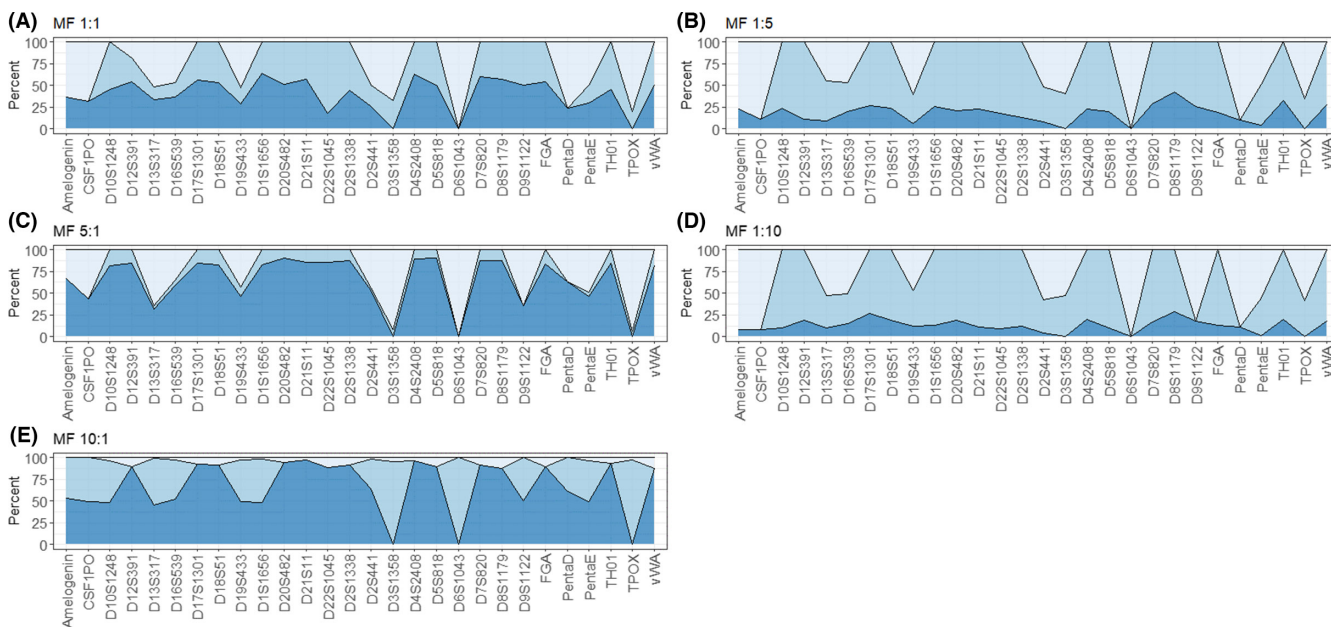
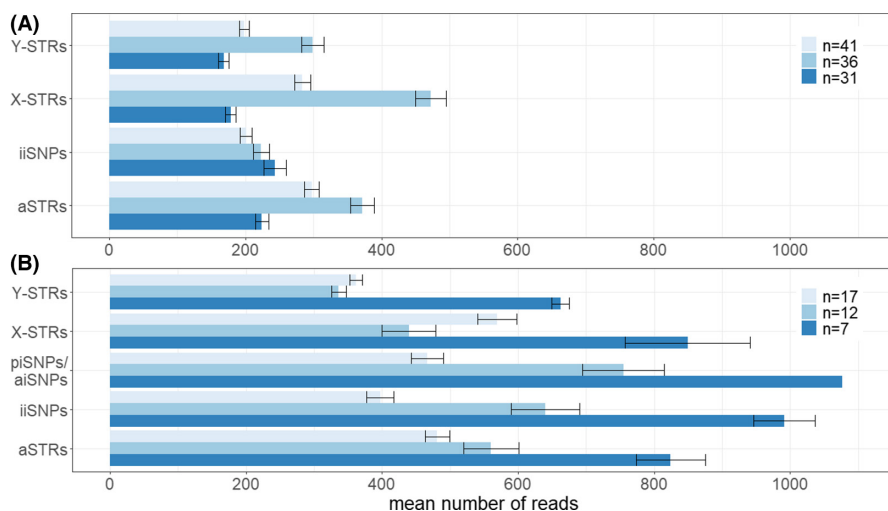


FIGURE 7 Mixture study for DPMB. Percentage of female (F) and male (M) read intensities per marker for different ratios of a male and female (MF) sample (A-E). Shared alleles were summarized as female/male [Color figure can be viewed at wileyonlinelibrary.com]

TABLE 1 Concordance study

Sample	Gender	Extraction method	CE		NGS (DPMA)	
			Profile completeness (%) CE (aSTRs)	Quality sensors (QS)	Profile completeness (%) NGS (aSTRs)	Avg. no. of reads (aSTRs)
Buccal swab	Male	Maxwell® FSC DNA IQ™	100	Present	100	506
		SwabSolution™ Kit	100	Present	100	304
	Female	Maxwell® FSC DNA IQ™	100	Present	100	513
		SwabSolution™ Kit	100	Present	100	304
Blood	Male	Maxwell® FSC DNA IQ™	100	Present	100	334
		SwabSolution™ Kit	100	present	100	402
	Female	Maxwell® FSC DNA IQ™	100	Present	100	188
		SwabSolution™ Kit	100	Present	100	403
M. rect. Femoris	Male	Maxwell® FSC DNA IQ™	100	Present	100	336
		SwabSolution™ Kit	43	QS2 absent	100	403
	Female	Maxwell® FSC DNA IQ™	100	Present	100	483
		SwabSolution™ Kit	0	QS1 and QS2 absent	100	259

Note: Genotyping success rates separated for NGS and CE showed in profile completeness (%) for buccal swabs, blood samples, and samples from the *M. rectus femoris*.

3.5 | Mixed samples

For male–female mixtures (MF), read intensity was above 85,000 for each mixture ratio, and concordant results were obtained from duplicates. Compared to known single-source reference profiles, only one AD at D1S1656 was observed at a ratio of 10:1. However, aSTRs of the minor contributor could be differentiated in every ratio and marker (Figure 7). As expected, the read intensities of the minor contributor decreased with a reduction of the input volume. Even in ratios of 10:1, the male minor contributor showed a total read count of 9881 (mean = 253), compared to the female major contributor (read count = 46,683, mean = 1556). Compared to the results of the female–male mixture, the male–male mixture (MM) revealed a higher number of ADs. Each read intensity was above 85,000. However, the ratio of 5:1 revealed ADs of the minor contributor at D2S441 and vWA, and a complete LD of the Y-STR DYS481. Consequently, the ratio of 10:1 also showed ADs of the minor contributor at CSF1PO, vWA, PentaE, D21S11, and PentaD. In contrast to a total read count of 83,186 for the major contributor (mean = 1698), the minor's was only 5451 (mean = 130). For the female–female mixture (FF), no ADs of aSTRs and X-STRs were observed for a ratio of 1:1, and two ADs of D1S1656 and D21S11 for a ratio of 5:1. With a mean read count of 226, the read intensity of the minor contributor was slightly lower than for the major contributor (mean = 347). The ratio of 10:1 revealed an AD of D5S818 and mean read counts of 585 and 160, respectively.

3.6 | Casework samples

All five GEDNAP samples revealed concordant aSTR results of mixtures or single-source samples with no LDs or ADs. The intensity of

reads was above 85,000 for both mixtures (with a mean of 91,751 and 106,893, respectively). For three samples, one per duplicate fell below the threshold of 85,000 (83,529, 60,726, and 80,895, respectively). However, no decline in data quality was observed.

3.7 | Concordance

For the buccal swabs and blood samples, all aSTR genotypes were concordant between NGS- and CE-based genotyping and complete with regard to each method-specific marker set (Table 1). For STRs analyzed with CE, both low-concentrated DNA samples from muscle tissues (0.02 ng and 0.01 ng) extracted with the SwabSolution™ Kit showed a lower genotyping success rate for male (43%) and female profiles (0%), respectively. For these samples, one or both PCR quality sensors were missing. Unexpectedly, the same NGS-based STR typing revealed complete aSTR profiles. On average, four loci with imbalanced alleles were observed.

3.8 | Degradation

Quantification of artificially degraded DNA samples showed a decrease in autosomal DNA concentration and an increased ratio of the autosomal target relative to the degradation target ($[\text{Auto}]/[\text{Deg}]$). Each sample exposed to UV light displayed no internal positive control quantification threshold (IPC Cq) shift but values exceeding the manufacture's threshold of two, indicating the presence of degraded DNA [43] (Table 2). The comparison of read intensities and genotyping success revealed a significant decrease with increasing minutes of UV light exposure ($p < 0.001$, one-way ANOVA) (Figure 8). First

TABLE 2 Degradation study

Sample	UV light exposure (Tmin)	ng [Auto]	ng [Deg]	ng [Y]	[Auto]/[Deg]	[Auto]/[Deg] threshold
Blood	T0	14.11	17.93	14.96	0.79	Below
Blood	T10	7.90	0.26	4.21	30.25	Above
Blood	T15	5.92	0.08	2.77	72.05	Above
Blood	T20	4.53	0.03	1.97	155.09	Above
Blood	T30	3.33	0.01	1.27	302.95	Above

Note: Quantification results were obtained with the PowerQuant Kit (Promega) for a blood sample exposed to UV light. [Auto]/[Deg] ratio greater than the threshold of two is marked in bold.

LD of rs354439 (iiSNP) and alleles below the interpretation threshold were shown for UV light exposures of 10 min. For longer UV light exposure times, a frequent LD and AD of PentaE were observed. Genotyping success rate for iiSNPs ranged from 100% (0 and 10 min) to 99% (15 and 20 min) and 93% (30 min). Phenotype and ancestry prediction were possible for all samples, except for one duplicate exposed to UV light for 30 min with a loss of rs683. Each DNA profile of degraded samples obtained with CE exhibited a distinct “ski-slope effect” [44] with a loss of the larger loci D2S1338, D21S11 D5S818, D7S820, D8S1179, FGA, and vWA. Evaluation of the influence of UV light exposure and analysis system on the number of dropouts revealed significantly higher numbers in samples analyzed with CE ($p = 0.019$, Figure 9). The factor UV light exposure showed no significant influence ($p = 0.255$). No significant interaction between the two factors was found ($p = 0.633$, all p -values from two-way ANOVA). UV light exposure times of 10 min led to first dropouts in CE-based genotyping (D2S1338, D7S820), compared with one AD at UV T₁₅ for NGS-based genotyping (PentaE). For UV T₃₀, seven times more LD were observed for CE-STRs.

3.9 | Additional sample purification for challenging samples

Comparison of total read intensities revealed a distinct decrease after purification adjustment 1 (Figure 10). Consequently, the NGS genotyping success was relatively low (Table 3). The purification adjustment 2 showed the highest intensities and genotyping success rates for each tissue type compared with the reference. Samples from the *pars petrosa* showed comparable low read numbers due to low DNA input concentration (0.015 ng/μL). The predicted phenotypes from purification adjustments 2 and 3 corresponded to the reference and the person’s visual phenotype. With regard to ancestry prediction, samples clustered between European and admixed American ancestry in a principal component analysis. No estimations were possible for samples with purification adjustment 1.

3.10 | Pooling adjustments for challenging samples

Increasing the input volume of a low-concentrated tissue sample from a human corpse within the pool led to an increase in its total

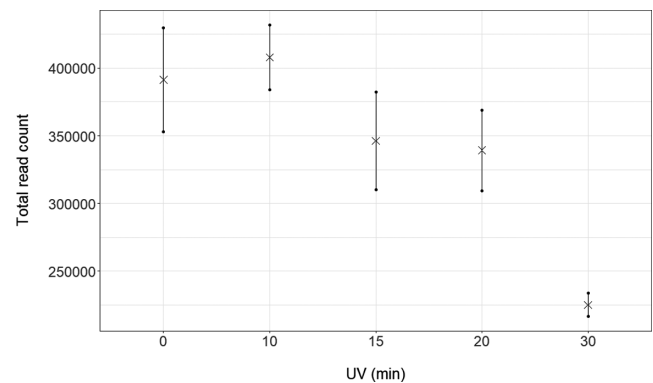


FIGURE 8 Degradation study for DPMB. Total number of reads from a DNA extract exposed to UV light for 0, 10, 15, 20, and 30 min. Displayed in each plot is the mean (x) of both samples

read intensities (Figure 11). However, even with an input volume of 15 μL, the recommended maximal threshold of 85,000 reads [41] was not reached. However, the NGS genotyping success increased, with an associated LDs and ADs decrease (Figure 12). Profile NGS completeness for autosomal and gonosomal STRs as well as SNPs was 79% for 5 μL, 84% for 10 μL, and 96% for 15 μL. Increasing the low-concentrated sample volume showed no considerable influence on the read count and genotyping success of the other samples within the pool. However, even though the read numbers of the low-concentrated sample increased, no phenotype prediction was possible due to an LD of rs1805009 in each run. Despite LDs of aiSNPs, accurate European ancestry was predicted for every sample.

3.11 | Reduction of adapters for challenging samples

A regression model was used to determine the influence of adapter reduction on the resulting read intensities of various input concentrations (Figure 13). With reduced adapter amounts, read intensities dropped below the recommendation of 85,000 [41] from samples with input concentrations ≤15.6 pg. The factor DNA input concentration showed significant influence ($p < 0.001$), whereas the factor volume of adapters displayed no significant influence on the read count ($p = 0.399$). Additionally, a significant interaction

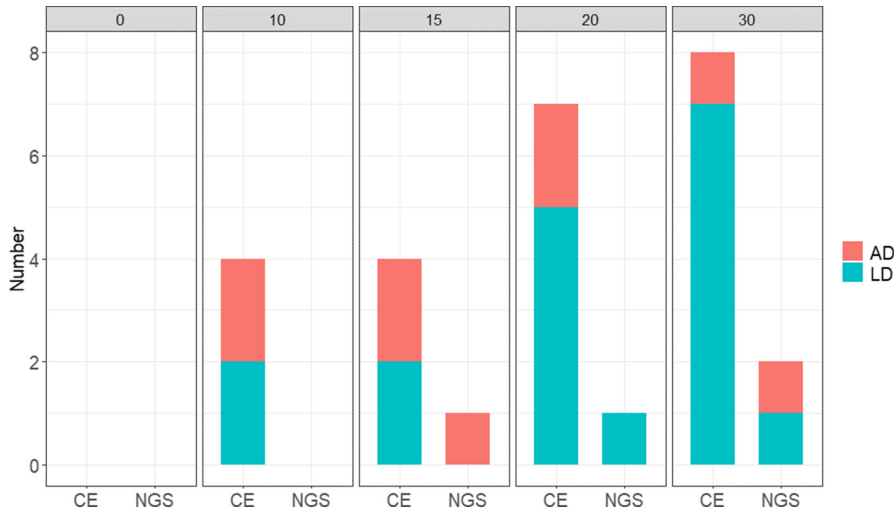


FIGURE 9 Degradation study for DPMB. Profile quality comparison of NGS and CE. Number of locus dropouts (LD) and allelic dropouts (AD) obtained from a DNA extract exposed to UV light for 0, 10, 20, and 30 min [Color figure can be viewed at wileyonlinelibrary.com]

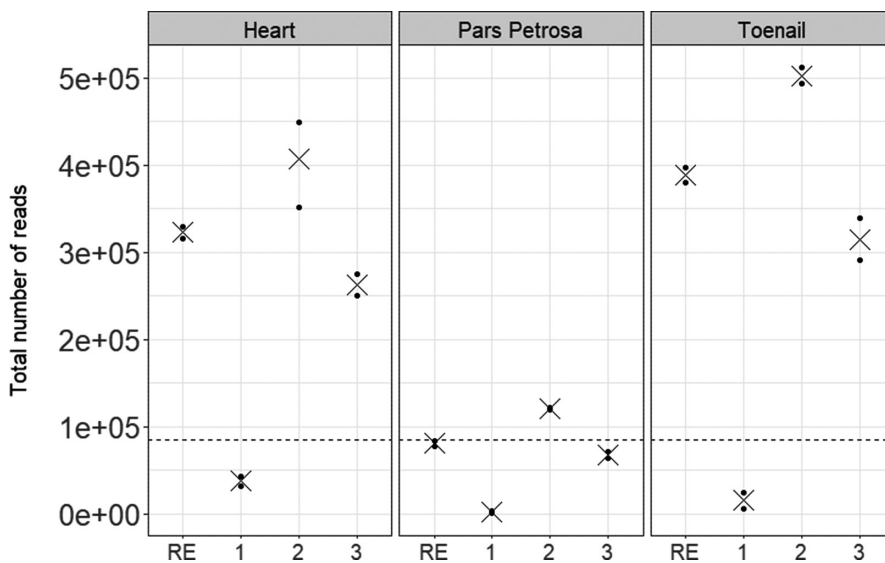


FIGURE 10 Optimisation study: Additional PCR purification for DPMB. Total number of reads obtained for samples from heart, toenail, and *pars petrosa* without additional purification (RE), with additional purification after target amplification (adjustment 1), with a repetition of the manufacturer's recommended purification step (adjustment 2) and with an extra purification using the MinElute kit (Qiagen) after the protocols library purification step (adjustment 3). Displayed in each plot is the mean (x), and the dotted line marks the threshold of 85,000 reads

between the two factors was found ($p < 0.001$, all p -values two-way ANOVA).

As expected, the amount of LDs, ADs, ABITs, dropins, and imbalanced alleles increased with decreasing DNA input concentrations (Figure 14). Concentrations down to 31.2 pg showed no distinct differences between adapter volumes. For an input concentration of 15.6 pg, the ADs increased from 31 (4 μ L) to 119 (2 μ L), with most dropouts observed for iiSNPs and aiSNPs. The most apparent differences were observed when comparing ADs of input concentrations of 7.8 pg. Samples with 4 μ L of each index showed mostly dropouts of iiSNPs, while 2 μ L in addition led to dropouts of aiSNPs. No phenotype prediction was possible for 7.8 pg and each tested adapter volume. Despite dropouts of aiSNPs, estimation of European ancestry was predicted for all adapter volume variations. Concordant genotype successes were obtained for each adapter volume and the vertebra and *pars petrosa* sample. No decrease in NGS-STR and iiSNP genotyping success rates was observed. For the femur sample, typing success of STRs slightly

decreased from 98% (4 μ L) to 97% (2 μ L) and profile completeness of iiSNP decreased from 98% (4 μ L) to 95% (2 μ L). Quality control of the purified libraries conducted with the BioAnalyzer revealed large peaks at about 170 bp, which is the length for the adapter dimers, indicating their presence in each sample. The range of the ForenSeq target fragments was between the expected 200–600 bp [45]. When comparing adapter input volumes, decreases were observed for adapter dimer concentrations. For the femur samples, peak height was reduced from 693 fluorescence units (FU) (4 μ L) to 400 FU (2 μ L), for the vertebra sample from 249 FU (4 μ L) to 208 FU (2 μ L) and the *pars petrosa* sample from 971 FU (4 μ L) to 423 FU (2 μ L).

4 | DISCUSSION

Over the past years, NGS moved more and more in the focus of forensic genetics, providing new opportunities for forensic DNA

TABLE 3 Optimization study: Additional PCR purification for DPMB

	Genotype success rate (aSTRs) (%)	Genotype success rate (Y-STRs) (%)	Genotype success rate (X-STRs) (%)	Genotype success rate (iiSNPs) (%)	Genotype success rate (ai SNPs) (%)	Genotype success rate (piSNPs) (%)	Phenotype concordance with reference (%)	Ancestry concordance with reference (%)
Reference	100	100	100	100	100	100	100	100
Purification after target amplification (1)	≥27	≥71	≥43	≥1	≥1	≥3	NA	NA
Repetition of purification step (2)	100	100	100	≥93	100	100	100	100
MinElute kit (Qiagen) (3)	≥95	100	100	≥92	≥99	100	100	100

Note: Genotyping success rates (%) obtained for aSTRs, Y-STRs, X-STRs, iiSNPs, aiSNPs, and piSNPs. Samples from heart, toenail, and pars petrosa with additional purification after target amplification (1), with a repetition of the manufacturer's recommended purification step (2), and with an extra purification using the MinElute kit (Qiagen) after the protocols' library purification step (3).

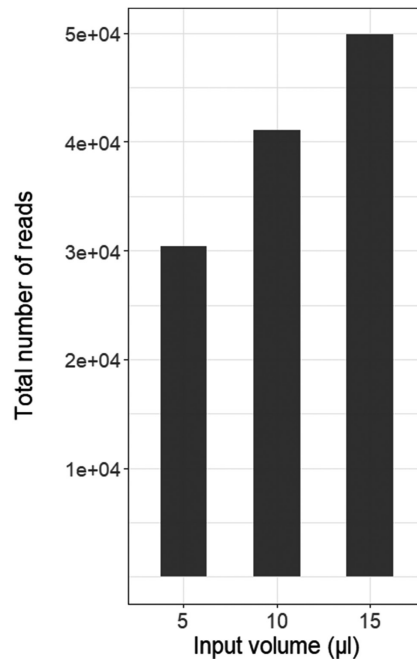


FIGURE 11 Optimisation study: Pooling adjustment for DPMB. Intensity of reads obtained from a low-concentrated DNA sample with input volumes of 5, 10, and 15 µL and same input volumes (5 µL) for the other samples within each pool

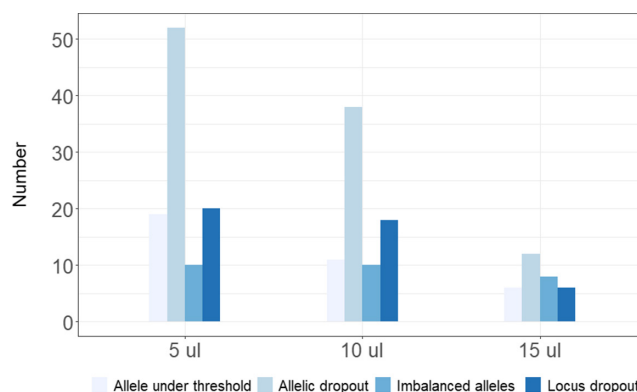


FIGURE 12 Optimization study: Pooling adjustment for DPMB. Number of locus dropouts, allelic dropouts, imbalanced alleles, and alleles under threshold for a sample pooled in volumes of 5, 10, and 15 µL [Color figure can be viewed at wileyonlinelibrary.com]

analyses. Its advantages of parallel sequencing of many autosomal and gonosomal STRs as well as SNPs, and detecting intravariations of STRs result in an increase in discrimination power compared with CE-based genotyping. However, prior to using novel systems or methods, internal validation studies must evaluate their potential power and limits within the forensic environment [24], especially for challenging samples like DNA mixtures of multiple persons and low-concentrated, degraded, or inhibited samples. Here, the extensive internal validation and optimisation study demonstrated the limitations and reliability of the MiSeq FGx system and the ForenSeq™ DNA Signature Prep Kit.

4.1 | Sensitivity

In agreement with published work, the sensitivity results demonstrated the possibility to generate complete DNA profiles with less than the recommended input concentration of 1 ng [6,7,11,31]. As shown in Jäger et al., it was possible to obtain 100% genotype success rates for concentrations down to 62.5 pg [6]. The total read numbers of samples amplified with DPMA dropped below the recommended threshold much earlier than those amplified with DPMB, which could be explained by the higher number of markers included in the primer mix B. Compared with STRs, piSNPs, and aiSNPs, the mean percent of LDs (6.3%), ADs (17.0%), and ABITs (16.0%) was the highest in iiSNPs, probably because of its highest number of markers in the ForenSeq™ DNA Signature Prep Kit. Additionally, the marker DXS10103 underperformed at lower concentrations as previously reported by Hollard et al. and Köcher et al., among others [2,6,11,22,23,31,46]. However, in contrast to these studies, no artificial DNA sample, but a real blood sample was used for dilution.

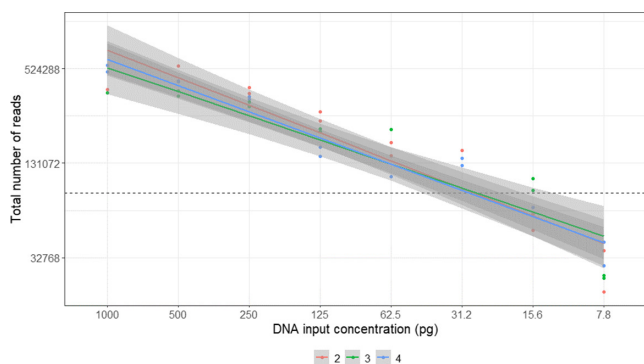


FIGURE 13 Optimisation study: Reduction of adapter for DPMB. Total number of reads for a serial dilution from 1000 to 7.8 pg DNA and adapter volumes of 2, 3, and 4 µL. Regression line for each adapter volume is plotted with 95% confidence bands. The dotted line indicates the manufacturer's read count threshold of 85,000 [Color figure can be viewed at wileyonlinelibrary.com]

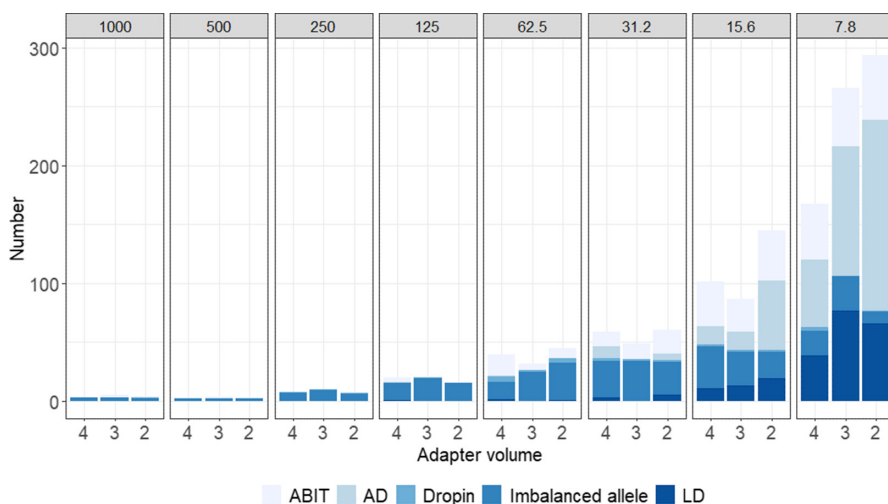


FIGURE 14 Optimisation study: Reduction of adapter for DPMB. Number of alleles below the interpretation threshold (ABITs), allelic dropouts (ADs), dropins, imbalanced alleles, and locus dropouts (LDs) for a serial dilution from 1000 to 7.8 pg DNA and adapter volumes of 2, 3, and 4 µL [Color figure can be viewed at wileyonlinelibrary.com]

4.2 | Variances

The variance study analyzed the impact of purified and crude lysates from different body fluids and tissue types. No significant differences were assessed between DNA extraction methods, indicating the protocols' reliability of target amplification regardless of the extraction method. Hence, even not purified DNA samples revealed robust sequencing results when using the protocol for crude lysates. Despite equal input material and fresh samples, tissue-specific differences were observed, potentially due to deviations during the complex and manual library preparation. In particular, buccal swabs extracted with the SwabSolution™ Kit and amplified with DPMB showed the most unpredictable number of reads and the lowest genotyping success. Such a deviation was also seen in a previous work [28], but the underlying samples were from altered and degraded human materials, for which divergent results are expected. Potentially different sequencing efficiencies could cause a general poor agreement in read intensities despite equal DNA input from fresh samples, which were found in the repetition and reproduction studies. However, due to the small sample size, further validation is necessary.

4.3 | Repeatability and reproducibility

With respect to read intensities, both repeatability and reproducibility studies showed poor agreement between the sequencing runs. Differences of up to 300,000 reads were observed between two runs, repeated by the same analyst. This could be explained by divergences during the complex library preparation and/or deviating efficiencies during cluster generation. According to Hollard et al., the automation of all library preparation steps with the Hamilton ID STARlet robotic platform led to repeatable results in terms of depth of coverage (DoC) [23]. Additionally, in contrast to other studies [2,6], no artificial reference control like the 2800 M DNA was used, but DNA from human buccal swabs, blood, and muscle samples to

represent real casework samples. However, regarding the genotyping success, profile completeness was completely reproducible and repeatable despite significant differences in read intensities. Even though only one run was repeated by the same analyst and reproduced by a second analyst, the genotyping success rates correspond to the results of comparable studies [2,7,31,47]. For SNP genotypes, biogeographic ancestry and phenotype prognosis, accurate, and concordant results were obtained for each evaluated run, regardless of the analyst. The results demonstrate reproducibility and repeatability and were also achieved in the study by Frégeau [21].

4.4 | Pooling variations

Adjusting the manufacturer's recommended number of pooled samples resulted in significant differences between total read intensities. With higher numbers of samples, the decrease was more distinct in samples amplified with DPMB, which could be linked to the greater number of markers included in the primer mix. Furthermore, in forensic casework, not only the sample's total intensity is crucial but also the performance of STRs and SNPs separately, as shown in Figure 6. Average intensities vary the most for SNPs amplified with DPMB, demonstrating that lower numbers of pooled samples should be preferred to achieve higher read numbers. However, for DPMB, no negative impact on the data quality and, for DPMA, only minor decreases in genotyping success rates were determined. Even though standard MiSeq FGx™ flow cells were evaluated, Moreno et al. also observed no decrease in profile quality when pooling 24, 32, and 40 samples with DPMB [47]. Nevertheless, as noted by Just et al., exceeding the number of pooled libraries by almost twice the recommended amount leads to decreasing numbers of recovered loci [30]. Thus, overclustering has a negative impact on sequencing performance, likely due to the difficulty of image analyses, including loss of focus [48].

4.5 | Concordance

The concordance of both methods is essential for implementing NGS into the currently CE-dominated routine work of forensic genetics. The aSTRs amplified with the ForenSeq™ DNA Signature Prep Kit and the Investigator 24plex QS Kit (Qiagen) showed concordant genotypes for buccal swabs and blood samples. Surprisingly, for muscle samples from corpses, aSTR markers enclosed in both kits revealed higher typing rates when sequenced with NGS. DNA extracted with the SwabSolution™ Kit surprisingly only led to 43% partial or zero profiles with CE but to complete profiles with NGS. In our laboratory routine casework, the SwabSolution™ Kit is validated and demonstrates sufficient results even for tissue samples. However, the ForenSeq protocol's purification step obviously removed inhibitors efficiently when compared to the nonpurified SwabSolution extracts.

4.6 | Degradation

As expected, with higher degrees of degradation, the total number of reads and NGS genotyping success decreased. Yet, 87% of iSNPs was still typed with UV exposure times of 30 minutes, and the only aSTR marker dropping out was PentaE. LD of PentaE was probably due to its second-longest amplicon length (392 bp) within the multiplex. The intensity of the longest amplicon (DXS8378, 450 bp) was just above the analytical threshold. In agreement with Jäger et al. [6] and Fattorini et al. [18], SNP markers showed a mean genotyping success of 98% and were about as stable as the STR markers (97%). The slight difference in the typing rate could be associated with the smaller amplicon size of SNPs.[18,27] Especially for degraded samples, sizes of ≤ 125 bp increase the chance to obtain sufficient genotypes [1,25]. Most likely, the reduced amplicon size is also the reason for the highly diverging profile completeness results obtained with the traditional CE method. Equivalent to the results of the concordance study, DNA profiles obtained with NGS revealed significantly higher genotype success rates than CE. This enhanced potential and clear advantage were also observed by Almohammed et al., who demonstrated significant differences in obtaining sufficient profiles from degraded bone samples analyzed with NGS and CE (GlobalFiler™ kit) [49].

4.7 | Additional sample purification for challenging samples

The MiSeq FGx system can prematurely abort a sequencing run when too many low-quality samples are pooled, as experienced for internal sequencing runs with highly degraded and inhibited samples from altered human remains (data not shown). Comprised input material can increase the formation of adapter dimers that might remain in the solution after purification. Due to their short size, such dimers have a higher amplification efficiency than DNA libraries and can interfere with cluster generation. Consequently, excessive cluster formation of dimers can result in underclustering of actual libraries [27,45,50]. Additionally, as experienced in sequencing runs (data not shown) and observed by Guo et al., highly inhibited samples can not only affect their genotyping success but also influence the cluster generation of the initially not inhibited samples on the same flow cell [22]. This potential contamination led to the decision against using artificially inhibited samples in this study.

For the aforementioned aborted runs, pooling several highly inhibited and degraded samples resulted in a total loss of sequencing data. To improve the sequencing of such challenging samples, three purification tests were conducted with highly inhibited and degraded samples from decomposed corpses. By repeating the manufacturer's purification workflow, including the magnetic beads' step, the removal of PCR artifacts was improved, and the sequencing results were enhanced. The total read intensities increased with a decrease in adapter dimers. The purification with additional spin

columns showed only slightly less improved genotyping success rates. Since the *spin column method* deviates from the actual workflow, repeating the manufacturer's purification workflow is more efficient and recommended.

4.8 | Pooling adjustments for challenging samples

By increasing the pooling volume of a low-concentrated muscle sample from a decomposed corpse to 15 μ L, the genotyping success rate increased by 17% to an almost complete profile. In addition, the volume-wise excess of low-concentrated DNA does not affect the remaining samples on the same flow cell, at least if they contain sufficient DNA amounts. Therefore, when sequencing highly degraded and inhibited samples, we recommend an additional PCR purification, an adjusted pooling volume, and adding higher quality samples to the library pool.

4.9 | Reduction of adapters for challenging samples

Furthermore, to minimize the formation of adapter dimers, a reduction of adapter volumes was tested down to 2 μ L for the entire DNA range. Library quality results of tissue samples from altered remains showed a decrease in adapter dimer concentration and concordant genotyping success rates for the vertebra and *pars petrosa* sample and each adapter volume. Although the opposite results were expected, the profile completeness of the femur sample and DNA input concentrations ≤ 15.6 pg decreased with reducing adapter volumes, indicating an insufficient amount of adapter. Therefore, a reduction of adapter is only recommended for samples with concentrations ≥ 31.2 pg and expected high-adapter dimer occurrence.

5 | CONCLUSION

The presented study comprised extensive analysis of the MiSeq FGx system's and ForenSeq™ DNA Signature Prep Kit's sensitivity, repeatability, reproducibility, concordance to CE, evaluated pooling variations, and validated different DNA extraction methods and casework, degraded, and mixed samples. Overall results showed the system to be reliable, robust, and implementable for forensic casework samples. In agreement with previous validation studies, the sequencing data are accurate and reproducible. Compared with CE, NGS revealed clear advantages in terms of marker multiplexing and concordant or even improved genotyping results. Especially for degraded samples, the reduced amplicon sizes lead to enhanced amplification efficiencies and typing rates.

However, the restricted amount of DNA input for target amplification and interference of cluster generation by adapter dimers is still a main drawback for forensic applications. Particularly highly degraded and inhibited tissue samples from altered corpses expose the system's limits. According to the presented results

of the optimisation studies, adjustments of library preparation prior to sequencing are recommended. An additional PCR purification step should be added, and the pooling volumes for low-concentrated DNA samples should be increased to prevent run failures. Therefore, further studies are required to improve the genotyping success of challenging tissue samples. Nevertheless, the MiSeq FGx has been successfully validated internally, and the results can be used as a basis for further implementation of NGS in forensic laboratories.

ACKNOWLEDGMENTS

The authors would like to thank Katja Anslinger for the valuable comments on the paper. Open access funding provided by Universitat Basel. WOA Institution: Universitat Basel Blended DEAL: CSAL.

DATA AVAILABILITY STATEMENT

All data generated or analyzed during this study are included in this published article.

RESEARCH INVOLVING HUMAN PARTICIPANTS AND/OR ANIMALS, HUMAN AND/OR ANIMAL SUBJECTS

This study involved samples from voluntary participants and human corpses obtained for forensic purposes. No formal consent was required for this study, which was approved by the regional Ethical Review Board (No. 2019-02211). All procedures were performed in compliance with the relevant laws and institutional guidelines.

ORCID

Iris Schulz  <https://orcid.org/0000-0002-7933-2563>

REFERENCES

1. Carrasco P, Inostroza C, Didier M, Godoy M, Holt CL, Tabak J, et al. Optimizing DNA recovery and forensic typing of degraded blood and dental remains using a specialized extraction method, comprehensive qPCR sample characterization, and massively parallel sequencing. *Int J Leg Med.* 2020;134(1):79–91. <https://doi.org/10.1007/s00414-019-02124-y>
2. Köcher S, Müller P, Berger B, Bodner M, Parson W, Roewer L, et al. Inter-laboratory validation study of the ForenSeq™ DNA Signature Prep Kit. *Forensic Sci Int Genet.* 2018;36:77–85. <https://doi.org/10.1016/j.fsigen.2018.05.007>
3. Wu J, Li J-L, Wang M-L, Li J-P, Zhao Z-C, Wang Q, et al. Evaluation of the MiSeq FGx system for use in forensic casework. *Int J Leg Med.* 2019;133(3):689–97. <https://doi.org/10.1007/s00414-018-01987-x>
4. Hwa H-L, Wu M-Y, Chung W-C, Ko T-M, Lin C-P, Yin H-I, et al. Massively parallel sequencing analysis of nondegraded and degraded DNA mixtures using the ForenSeq™ system in combination with EuroForMix software. *Int J Leg Med.* 2019;133(1):25–37. <https://doi.org/10.1007/s00414-018-1961-y>
5. Van Neste C, Van Nieuwerburgh F, Van Hoofstat D, Deforce D. Forensic STR analysis using massive parallel sequencing. *Forensic Sci Int Genet.* 2012;6(6):810–8. <https://doi.org/10.1016/j.fsigen.2012.03.004>
6. Jäger AC, Alvarez ML, Davis CP, Guzmán E, Han Y, Way L, et al. Developmental validation of the MiSeq FGx forensic genomics system for targeted next generation sequencing in forensic DNA casework and database laboratories. *Forensic Sci Int*

- Genet. 2017;28:52–70. <https://doi.org/10.1016/j.fsigen.2017.01.011>
7. Xavier C, Parson W. Evaluation of the Illumina ForenSeq™ DNA Signature Prep Kit–MPS forensic application for the MiSeq FGx™ benchtop sequencer. *Forensic Sci Int Genet.* 2017;28:188–94. <https://doi.org/10.1016/j.fsigen.2017.02.018>
 8. van der Gaag KJ, de Leeuw RH, Hoogenboom J, Patel J, Storts DR, Laros JF, et al. Massively parallel sequencing of short tandem repeats—Population data and mixture analysis results for the PowerSeq™ system. *Forensic Sci Int Genet.* 2016;24:86–96. <https://doi.org/10.1016/j.fsigen.2016.05.016>
 9. Zeng X, King JL, Stoljarova M, Warshauer DH, LaRue BL, Sajantila A, et al. High sensitivity multiplex short tandem repeat loci analyses with massively parallel sequencing. *Forensic Sci Int Genet.* 2015;16:38–47. <https://doi.org/10.1016/j.fsigen.2014.11.022>
 10. Børsting C, Morling N. Next generation sequencing and its applications in forensic genetics. *Forensic Sci Int Genet.* 2015;18:78–89. <https://doi.org/10.1016/j.fsigen.2015.02.002>
 11. Churchill JD, Schmedes SE, King JL, Budowle B. Evaluation of the Illumina® beta version ForenSeq™ DNA signature prep kit for use in genetic profiling. *Forensic Sci Int Genet.* 2016;20:20–9. <https://doi.org/10.1016/j.fsigen.2015.09.009>
 12. Butler JM. *Fundamentals of forensic DNA typing.* Burlington, MA: Academic Press; 2009.p. 191–5.
 13. Diegoli TM, Farr M, Cromartie C, Coble MD, Bille TW. An optimized protocol for forensic application of the PreCR™ Repair Mix to multiplex STR amplification of UV-damaged DNA. *Forensic Sci Int Genet.* 2012;6(4):498–503. <https://doi.org/10.1016/j.fsigen.2011.09.003>
 14. Burger J, Hummel S, Herrmann B, Henke W. DNA preservation: A microsatellite-DNA study on ancient skeletal remains. *Electrophoresis.* 1999;20(8):1722–8.
 15. Roeper A, Reichert W, Mattern R. The Achilles tendon as a DNA source for STR typing of highly decayed corpses. *Forensic Sci Int.* 2007;173(2–3):103–6. <https://doi.org/10.1016/j.forsciint.2007.02.004>
 16. Piccinini A, Cucurachi N, Betti F, Capra M, Coco S, D'Avila F, et al. Forensic DNA typing of human nails at various stages of decomposition. *Int Congr Ser.* 2006;1288:586–8. <https://doi.org/10.1016/j.ics.2005.08.029>
 17. McCord B, Opel K, Funes M, Zoppis S, Meadows JL. An investigation of the effect of DNA degradation and inhibition on PCR amplification of single source and mixed forensic samples. *National Criminal Justice References Report 236692* 2011. <https://www.ojp.gov/pdffiles1/nij/grants/236692.pdf>. Accessed March 1, 2022.
 18. Fattorini P, Previderé C, Carboni I, Marrubini G, Sorçaburu-Cigliero S, Grignani P, et al. Performance of the ForenSeq™ DNA Signature Prep kit on highly degraded samples. *Electrophoresis.* 2017;38(8):1163–74. <https://doi.org/10.1002/elps.201600290>
 19. Carboni I, Fattorini P, Previderè C, Cigliero SS, Iozzi S, Nutini A, et al. Evaluation of the reliability of the data generated by Next Generation Sequencing from artificially degraded DNA samples. *Forensic Sci Int Genet Suppl Ser.* 2015;5:e83–e5. <https://doi.org/10.1016/j.fsigss.2015.09.034>
 20. Iozzi S, Carboni I, Contini E, Pescucci C, Frusconi S, Nutini A, et al. Forensic genetics in NGS era: New frontiers for massively parallel typing. *Forensic Sci Int Genet Suppl Ser.* 2015;5:e418–9. <https://doi.org/10.1016/j.fsigss.2015.09.166>
 21. Frégeau CJ. Validation of the Verogen ForenSeq™ DNA Signature Prep kit/Primer Mix B for phenotypic and biogeographical ancestry predictions using the Micro MiSeq® Flow Cells. *Forensic Sci Int Genet.* 2021;53:102533. <https://doi.org/10.1016/j.fsigen.2021.102533>
 22. Guo F, Yu J, Zhang L, Li J. Massively parallel sequencing of forensic STRs and SNPs using the Illumina® ForenSeq™ DNA signature prep kit on the MiSeq FGx™ forensic genomics system. *Forensic Sci Int Genet.* 2017;31:135–48. <https://doi.org/10.1016/j.fsigen.2017.09.003>
 23. Hollard C, Ausset L, Chantrel Y, Jullien S, Clot M, Faivre M, et al. Automation and developmental validation of the ForenSeq™ DNA Signature Preparation kit for high-throughput analysis in forensic laboratories. *Forensic Sci Int Genet.* 2019;40:37–45. <https://doi.org/10.1016/j.fsigen.2019.01.010>
 24. Scientific Working Group on DNA Analysis Methods (SWGDM). *Validation guidelines for DNA analysis methods.* 2016. <https://www.swgdam.org/>.
 25. MiSeq FGx™ Forensic Genomics System. Solve more cases and generate more leads with the power and accuracy of Illumina next-generation sequencing. System Specification Sheet: Forensic Genomics. Illumina. Pub. No. 1470–2014–004. San Diego, CA: illumina; 2016 April.
 26. Bronner IF, Quail MA, Turner DJ, Swerdlow H. Improved protocols for illumina sequencing. *Curr Protoc Hum Genet.* 2013;79(1):18.2.1–18.2.42. <https://doi.org/10.1002/0471142905.hg1802579>
 27. England R, Harbison S. A review of the method and validation of the MiSeq FGx™ Forensic Genomics Solution. *Wiley Interdisciplinary Rev Forensic Sci.* 2020;2(1):e1351. <https://doi.org/10.1002/wfs2.1351>
 28. Senst A, Scheurer E, Gerlach K, Schulz I. Which tissue to take? A retrospective study of the identification success of altered human remains. *J Forensic Leg Med.* 2021;84:102271. <https://doi.org/10.1016/j.jflm.2021.102271>
 29. Hussing C, Huber C, Bytyci R, Mogensen HS, Morling N, Børsting C. Sequencing of 231 forensic genetic markers using the MiSeq FGx™ forensic genomics system—An evaluation of the assay and software. *Forensic Sci Res.* 2018;3(2):111–23. <https://doi.org/10.1080/20961790.2018.1446672>
 30. Just RS, Moreno LI, Smerick JB, Irwin JA. Performance and concordance of the ForenSeq™ system for autosomal and Y chromosome short tandem repeat sequencing of reference-type specimens. *Forensic Sci Int Genet.* 2017;28:1–9. <https://doi.org/10.1016/j.fsigen.2017.01.001>
 31. Almalki N, Chow HY, Sharma V, Hart K, Siegel D, Wurmbach E. Systematic assessment of the performance of illumina's MiSeq FGx™ forensic genomics system. *Electrophoresis.* 2017;38(6):846–54. <https://doi.org/10.1002/elps.201600511>
 32. Pajnič IZ. *Extraction of DNA from human skeletal material. Forensic DNA typing protocols.* Cham, Switzerland: Springer Nature; 2016. p. 89–108.
 33. Promega Corporation. *PowerQuant® System Technical Manual #TMD047, Rev. 01/20.* Madison, WI: Promega Corporation; 2020.
 34. Team R. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2021.
 35. Team R. RStudio: Integrated development for R. Boston, MA: RStudio, PBC; 2020.
 36. Wickham H, François R, Henry L, Müller K. *dplyr: A grammar of data manipulation. R package version 1.0.7.* 2021 2021. <https://CRAN.R-project.org/package=dplyr>.
 37. Kassambara A. *ggpubr: 'ggplot2' based publication ready plots. R package version 0.4.0.* 2020. <https://CRAN.R-project.org/package=ggpubr>.
 38. Berkelaar M. *lpSolve: Interface to 'Lp_solve' v. 5.5 to solve linear/integer programs. R package version 5.6.15.* 2020. <https://CRAN.R-project.org/package=lpSolve>.
 39. Wickham H. *ggplot2: Elegant graphics for data analysis.* New York, NY: Springer-Verlag; 2016.
 40. Lehnert B. *BlandAltmanLeh: Plots (slightly extended). Bland-Altman Plots. R package version 0.3.1.* 2015. <https://CRAN.R-project.org/package=BlandAltmanLeh>.
 41. Verogen. *ForenSeq™ Universal Analysis Software guide. Sample and run results.* Verogen. Document # VD2018007. Rev. A. June 2018. San Diego, CA: Verogen; 2018.

42. Verogen. ForenSeq™ DNA Signature Prep reference guide. Document # VD2018005 Rev. C. August 2020. San Diego, CA: Verogen; 2020.
43. Kraemer M, Prochnow A, Bussmann M, Scherer M, Peist R, Steffen C. Developmental validation of QIAGEN Investigator® 24plex QS Kit and Investigator® 24plex GO! Kit: Two 6-dye multiplex assays for the extended CODIS core loci. *Forensic Sci Int Genet*. 2017;29:9–20. <https://doi.org/10.1016/j.fsigen.2017.03.012>
44. Butler JM. *Advanced topics in forensic DNA typing: Interpretation*. San Diego, CA: Academic Press; 2014.p. 109–10.
45. England R, Nancollis G, Stacey J, Sarman A, Min J, Harbison S. Compatibility of the ForenSeq™ DNA Signature Prep Kit with laser microdissected cells: An exploration of issues that arise with samples containing low cell numbers. *Forensic Sci Int Genet*. 2020;47:102278. <https://doi.org/10.1016/j.fsigen.2020.102278>
46. Silvia AL, Shugarts N, Smith J. A preliminary assessment of the ForenSeq™ FGx System: next generation sequencing of an STR and SNP multiplex. *Int J Leg Med*. 2017;131(1):73–86. <https://doi.org/10.1007/s00414-016-1457-6>
47. Moreno LI, Galusha MB, Just R. A closer look at Verogen's ForenSeq™ DNA Signature Prep kit autosomal and Y-STR data for streamlined analysis of routine reference samples. *Electrophoresis*. 2018;39(21):2685–93. <https://doi.org/10.1002/elps.201800087>
48. Optimizing cluster density on Illumina sequencing systems. Understanding cluster density limitations and strategies for preventing under- and overclustering. Illumina. Pub. No. 770-2014-038. San Diego, CA: Illumina. 2016 April. p. 1-12.
49. Almohammed E, Zgonjanin D, Iyengar A, Ballard D, Devesse L, Sibte H. A study of degraded skeletal samples using ForenSeq™ DNA Signature Kit. *Forensic Sci Int Genet Suppl Ser*. 2017;6:e410–2. <https://doi.org/10.1016/j.fsigs.2017.09.158>
50. Head SR, Komori HK, LaMere SA, Whisenant T, Van Nieuwerburgh F, Salomon DR, et al. Library construction for next-generation sequencing: overviews and challenges. *Biotechniques*. 2014;56(2):61–77. <https://doi.org/10.2144/000114133>

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Senst A, Caliebe A, Scheurer E, Schulz I. Validation and beyond: Next generation sequencing of forensic casework samples including challenging tissue samples from altered human corpses using the MiSeq FGx system. *J Forensic Sci*. 2022;67:1382–1398. <https://doi.org/10.1111/1556-4029.15028>