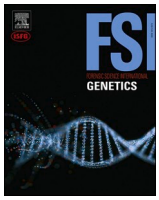
Forensic Science International: Genetics 77 (2025) 103233

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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



Inter-platform evaluation of the MPSplex large-scale tri-allelic SNP panel for forensic identification

J. Ruiz-Ramírez [a,*], F. Bittner [b], T.J. Parsons [b], A. Tillmar [c,d], L. Vangeel [b], I. Grandell [c], M. Eduardoff [b], M.A. Peck [b], A. Ambroa-Conde [a], A. Mosquera-Miguel [a], A. Freire-Aradas [a], M.V. Lareu [a], C. Phillips [a,e], M. de la Puente [a,*]

[a] Forensic Genetics Unit, Institute of Forensic Sciences, Universidade de Santiago de Compostela, Santiago de Compostela, Spain
[b] International Commission on Missing Persons, Koninginnegracht 12, The Hague, Netherlands
[c] Department of Forensic Genetics and Forensic Toxicology, National Board of Forensic Medicine, Linköping, Sweden
[d] Department of Biomedical and Clinical Sciences, Faculty of Medicine and Health Sciences, Linköping University, Linköping, Sweden
[e] King's Forensics, Faculty of Life Sciences and Medicine, King's College, London, UK

ARTICLE INFO

Keywords:
Forensic genetics
Tri-allelic SNPs
Unique Molecular Indices (UMIs)
Missing persons identification
Massively parallel sequencing

ABSTRACT

MPSplex is a large-scale forensic massively parallel sequencing (MPS) panel with 1,270 tri-allelic SNPs, 44 microhaplotypes (MH) and 55 ancestry-informative bi-allelic SNPs (aiSNPs) designed for missing persons identification. We have evaluated MPSplex with the most widely used MPS platforms in the forensic field: the Illumina MiSeq, the Thermo Fisher Scientific Ion S5 and the Qiagen GeneReader. The tri-allelic SNPs of MPSplex were previously identified from the most polymorphic loci with three common alleles in 1000 Genomes Phase III data and combined with the 44 MH and 55 aiSNPs, then implemented into a QIAseq Targeted DNA Custom Panel (Qiagen), a marker panel which uses Unique Molecular Indices or UMIs. The UMI random-sequence DNA molecules are incorporated onto DNA fragments before the Target Enrichment PCR, allowing the identification of reads that originated from the same template and consequently they can be used to correct the errors that may arise within the PCR or the sequencing process. In this study, we present the results of an inter-platform evaluation of the MPSplex panel, characterizing its performance in different forensic scenarios, which assessed aspects that include sensitivity, genotyping accuracy and mixture analysis. MPSplex aims to provide a tool designed for kinship analysis that can be applied beyond the resolution of first- or second-degree relationships, avoiding the need for much bigger forensic panels designed for genealogy purposes, which usually require significantly more sequencing resources. This study provides evaluation of MPSplex using the MPS systems in routine use for forensic genotyping of large-scale panels of SNPs.

1. Introduction

The implementation of massively parallel sequencing (MPS) technologies in the forensic field has led to the creation of many different large-scale panels, including expanded STR panels (e.g., ForenSeq MainstAY from Verogen, PowerSeq® 46GY System from Promega or Precision ID GlobalFiler™ NGS STR Panel v2 from Thermo Fisher Scientific (TFS)), panels of single nucleotide polymorphisms (SNPs), and panels comprising microhaplotypes (MHs) [1–3], a type of marker which cannot be analysed through traditional capillary electrophoresis (CE) procedures. Many such panels are solely composed of SNPs, which can be used not only for identification purposes, but also to predict bio-geographic ancestry (BGA) and externally visible characteristics (EVCs) [4–7].

While SNPs have lower polymorphism levels than STRs so are less informative per locus, the capability of MPS to simultaneously analyse many more markers per run than conventional multiplexed assays based on capillary electrophoresis, and the abundance of SNPs in the human genome, has led to the development of panels targeting up to a thousand or more SNPs. Panels of SNPs at this scale have the most utility in kinship studies where the putative relationship analysed is distant and STRs cannot provide enough power to resolve the degree of relatedness - commonly beyond second-degree relatives. Currently, a variety of MPS SNP panels developed by the community or commercially available,

* Corresponding authors.
E-mail addresses: jruiz.ramirez@usc.es (J. Ruiz-Ramírez), mdelcarmendela.puente@usc.es (M. de la Puente).

https://doi.org/10.1016/j.fsigen.2025.103233
Received 21 June 2024; Received in revised form 25 January 2025; Accepted 28 January 2025
Available online 28 January 2025

range in scale from ~100 SNPs [8,9] to the bigger multiplexes specifically designed for genealogical purposes: Verogen's ForenSeq Kintelligence Kit comprising 10,230 SNPs; and the FORCE panel, containing 3,929 kinship SNPs [10].

We previously developed and optimised the MPSplex panel, designed to enhance the identification of missing persons when STRs could not resolve the putative relationship of human remains to reference samples of relatives [11]. MPSplex consists of 1,270 carefully selected tri-allelic SNPs compiled from 1000 Genomes Phase III data supplemented with 44 MHs and 55 established ancestry informative SNPs (aiSNPs) [12], implemented in a QIAseq custom panel. The main characteristic of the QIAseq MPS panel design is the use of Unique Molecular Indices (UMIs), random nucleotide sequences that, when incorporated into DNA fragments prior to PCR amplification, can be used to identify all sequence reads that originated from the same template molecule. During posterior bioinformatic analysis, all these sequence reads are unified into a single read (the UMI read), allowing bioinformatic removal of errors introduced during PCR and the sequencing process. In addition, the panel utilizes a single region-specific primer for each locus, while the other primer is complementary to the adapter ligated (Supplementary Fig. S1). Although this technology is mainly used outside of forensic applications, such as somatic variant detection or transcript counting in RNA sequence analysis, UMIs have begun to be successfully implemented in several forensic panels [13–16].

In the present study, we evaluated the forensic performance of the MPSplex panel sequenced in the three MPS platforms which have been applied to forensic DNA analysis: Thermo Fisher Scientific's Ion S5, Verogen's MiSeq FGx and Qiagen's GeneReader.

2. Materials and methods

2.1. The MPSplex panel

The version of MPSplex used for this evaluation was developed as a QIAseq Targeted DNA Custom Panel (Catalogue number: CDHS-15861Z-2897) targeting a total of 1,445 SNPs. In addition to the 1,270 tri-allelic SNPs originally selected in [11] (1,241 autosomal plus 29 X-SNPs) and 120 key targeted SNPs within 44 microhaplotypes (MHs) selected from [1] and [17], it includes 55 aiSNPs for ancestry prediction from [12]. Chromosomal locations and SNP frequencies for the tri-allelics were previously compiled in Supplementary Table S1C from [11] (note that SNP allele frequencies for EUR and AFR populations were inverted, so incorrectly described in this data) and information on the 55 aiSNPs was compiled in the original publication [12]. Several MHs were shortened to reduce the targeted sequence amplicon size while designing their primers, using an approach similar that followed in [18]. A list of microhaplotype loci with their respective composite SNPs is in Supplementary Table S1.

For all data analysis in this study, the MH composite SNPs were treated as individual SNPs rather than as components of discrete sets forming the targeted microhaplotypes. Additionally, it was noted that aiSNPs rs870347, rs192655 and rs917115 had incorrect coordinates (e.g., rs870347 was incorrectly positioned on chromosome 6, not 5). Thus, primers for these three SNPs did not target the correct location and were excluded from further analysis.

2.2. Inter-laboratory validation

Evaluation of the performance of MPSplex across the three platforms involved MPS systems sited in three different locations. These were: the Ion S5 System (TFS) at Universidade de Santiago de Compostela, the MiSeq FGx (Verogen) at the National Board of Forensic Medicine from Sweden and the GeneReader (Qiagen) at the International Commission on Missing Persons. A set of shared DNA control samples were selected to be run independently with each platform.

2.2.1. Reference DNA sample set

The reference DNA set comprised: (i) the 2800 M Control DNA (Promega); (ii) the NIST 2391d Standard Reference Material (SRM), components A and C; and (iii) 7 Coriell cell-line control DNAs of NA12878 and two trios comprising NA24143 (mother) - NA24149 (father) - NA24385 (son) (a CEU sample trio) and, NA06994 (father) - NA07000 (mother) - NA07029 (son) (an Ashkenazim sample trio). These samples were utilized for the assessment of sequencing quality metrics including sequence coverage (both total coverage and number of UMI reads), base misincorporation, strand bias and allele read frequency balance. The non-template negative control was used to measure the sequencing baseline.

2.2.2. Concordance analysis

The Coriell samples were used for both confirmation of mendelian inheritance across all markers and for concordance studies. The reference genotypes for these samples were retrieved from the 1000 Genomes project's high coverage sequencing data [19] for NA06994, NA07000 and NA12878; and the Genome-in-a-Bottle (GIAB) benchmark variant calls v4.2.1 [20] for the Ashkenazim Trio.

2.2.3. Sensitivity

A dilution series of 2800 M was used to evaluate the sensitivity of MPSplex to low-level DNA, with DNA input amounts in the range: 1000 pg; 500 pg; 250 pg; 125 pg; 62 pg; 31 pg; and 7 pg. The consensus profile of the 1000 pg replicate from each platform was used to harmonize the inter-platform analysis of the percentage of concordant loci – i.e., for each SNP, a genotype was assigned when coinciding for at least two of the MPS platforms.

2.2.4. Mixture detection

To evaluate the ability of MPSplex to detect and analyse mixed DNA, SRM 2391d component D (NIST), a precisely constructed 3:1 mixture of components A and C, was analysed together with components A and C from the SRM 2391d reference set.

2.3. Proof-of-concept kinship analysis

A set of 3 samples from an indirect kinship case (details in Section 3.5) showing low LR values when analysed with two CE-STRs kits of GlobalFiler™ (TFS) and Investigator HDplex (Qiagen) were run on the Ion S5 instrument to test the applicability of MPSplex for typical relationship testing casework. Samples had been previously extracted from buccal swabs using the AutoMate *Express*™ Forensic DNA Extraction System and the PrepFiler *Express*™ Forensic DNA Extraction kit (Applied Biosystems) and quantified with Quantifiler™ Trio DNA Quantification Kit (Applied Biosystems).

2.4. Library preparation

Libraries were prepared using the QIAseq Targeted DNA Panel and QIAseq Indexes (Qiagen) for the Ion S5 and MiSeq platforms, following the manufacturer's protocol for standard DNA conditions for each specific platform. The GeneRead QIAact Library Preparation Target Enrichment Reagents and GeneRead QIAact Lung DNA UMI Panel and Adaptor Kit (Qiagen) were used for GeneReader library preparation, using a 1.4x QIAseq beads-DNA ratio for adapter-ligated DNA cleanup and 1.2x for target enrichment and Universal PCR cleanups instead of the 1x ratio recommended in the protocol. Although the QIAseq assay specifications stated that a minimum input of 10 ng input DNA was required for library preparation, it also noted that lower inputs were possible, resulting in a reduction of the number of sequenced UMIs and variant detection sensitivity. Given the forensic applications that the panel was designed for (i.e., samples with low concentration and/or highly degraded), and that the identification of low-frequency variants was not intended, the Ion S5 and GeneReader laboratories used 1 ng for

most of the reference samples. In contrast, the MiSeq laboratory opted to follow the manufacturer's recommendations using a higher DNA input. The specific input DNAs for each sample and MPS platform are outlined in Supplementary Table S2.

2.5. Sequencing

2.5.1. Ion S5

Ion S5 libraries were quantified with the Ion Library TaqMan™ Quantitation Kit on a 7500 Real-Time PCR System (TFS) following the manufacturer's protocol and subsequently pooled at equimolar concentrations at a final library pool of 50 pM. Templating was carried out in an Ion Chef Instrument (TFS) with the Ion 510 & Ion 520 & Ion 530 Kit-Chef using a 400 bp templating protocol. Sequencing was performed on an Ion 530™ Chip (with a capacity of 15–20 million reads per chip) using the Ion S5™ Sequencing Reagents (TFS) on the Ion S5™ Instrument (TFS).

2.5.2. MiSeq

MiSeq libraries were quantified using the Qubit 2.0 fluorometer (TFS) and DNA integrity was evaluated using the High Sensitivity DNA kit on the 2100 Bioanalyzer (Agilent Technologies). Samples were further diluted to 4 nM based on the quantification and fragment size distribution. Libraries were then pooled, denatured, and further diluted to 5 pM. The sequencing process was carried out on a MiSeq FGx instrument with the MiSeq FGx Reagent Kit, providing a flow cell capacity of 12.5 million reads (Verogen) using the paired-end 2×151 bp sequencing protocol. In addition, a QIAseq A Read 1 Custom Primer I was loaded according to the manufacturer's protocol.

2.5.3. GeneReader

Libraries for the GeneReader were quantified with the Qubit 3.0 fluorometer (TFS) and DNA integrity was verified with TapeStation 4200 and D5000 ScreenTape and D5000 reagents (Agilent Technologies). These were further normalised to 100 pg/μL, pooled in equimolar concentrations and diluted to 1 pg/μL. Clonal amplification was performed on the GeneRead QIAcube and sequencing was carried out on the GeneReader platform with the GeneRead UMI Advanced Sequencing Q kit on a GeneRead UMI Advanced Sequencing Q Flow Cell (Qiagen).

2.6. Sequence analysis and genotype calling

2.6.1. UMI analysis

CLC Genomics Workbench v.20.0.1 bioinformatics suite (herein CLC) was used for analysis of the sequencing data. Certain settings were adapted for each platform – e.g., use of paired-end reads for the MiSeq. FASTQ files were imported to CLC and processed with *Remove and Annotate with Unique Molecular Index* and *Universal adapter trimming* tools. Next, reads were aligned to the reference genome (hg19) using the *Map Reads to Reference*; and subsequently the *Calculate Unique Molecular Index Groups* and *Create UMI Reads from Grouped Reads* tools were used to generate UMI reads. Alignments were processed with the *Remove Ligation Artifacts* tool and then *Indel and Structural Variants* and *Prepare Guidance Variant Track* tools were used to identify structural variants. The *Local Realignment* tool was used to improve alignment quality and then primer sequences were removed from reads with *Trim Primers of Mapped Reads*. Finally, the *Identify Variants on Known Positions* utility was used to extract the data for each locus.

Genotype calling was performed on R software 4.3.1 (https://www.r-project.org/, [21]) using a 10 UMI threshold and, for reference samples, a 90 % minimum allele frequency was applied for homozygous calls and a 20 % minimum allele frequency for heterozygous calls. If a marker showed three alleles with allele frequencies above 20 %, the genotype was not called. For mixture analysis, allele calling was made using a 10 UMI threshold and a minimum allele frequency of 10 % for allele calling. This threshold was established considering prior knowledge of the mixture ratio, that represents a minimum allele frequency for the minor component of 12.5 % for the 3:1 ratio (e.g., the genotype from the major contributor is AA and the minor contributor AG).

2.6.2. Independent sequence analyses without using UMI data

Raw FASTQ files were processed independently for the three kinship samples, using a pipeline based on open-source software and disregarding the UMI information. First, UMI tags, along with the common sequence, were trimmed from reads. Then, FASTQ files were aligned to hg19 reference genome using the bwa mem algorithm and later analysed with bam-readcount [22], applying a minimum base quality of 20 and mapping quality of 30. The output was subsequently imported into R and used for genotype calling. Parameters for genotype calling comprised an analytical threshold of 10 reads and an allele frequency of 90 % and 20 % for calling homozygous and heterozygous genotypes, respectively.

2.7. Statistical analyses

R scripts from ILIR and FamLink (v. 1.1) [23,24] were used to generate simulations to obtain kinship probability distributions using the 1000 Genomes European haplotype frequency estimates and map distance estimates from the HapMap project provided in [11]. A posterior version of FamLink (v. 2.3) was used for LR calculation. Familias R package [25] was used to generate the pedigree plots.

3. Results

3.1. Inter-platform comparisons and overall performance assessments of MPSplex

3.1.1. Total sequence coverage, UMI coverage and sequence read/UMI ratio

Fig. 1A represents the distribution of UMI coverage per marker for the reference samples that were amplified with 1–2 ng DNA input – 9, 3 and 10 samples for the Ion S5, MiSeq and GeneReader, respectively. These samples had an average of 159, 137 and 123 UMI reads per locus for the Ion S5, MiSeq and GeneReader, respectively. The proportions of the sequenced target loci of 99.8 %, 99.8 % and 99.0 %, respectively, had a UMI coverage value above the calling threshold. Fig. 1B shows the coverage per marker for the full reference DNA sample set ordered by mean UMI coverage (excluding the high input samples). The Fig. 1B plot shows the differential amplification among markers typically seen in targeted sequencing MPS panels (e.g. [26,27]), in which some markers display a coverage above the average while a few other markers fail to reach the 10 UMI threshold. Supplementary Fig. S2 contains individual boxplots for total coverage, UMI coverage and reads per UMI distribution for each of the reference samples plus the no template control.

It is noteworthy that most reference samples typed on the MiSeq and one on the Ion S5 (NA12878) were amplified using $\geq$ 10 ng of input DNA (Supplementary Table S2). With the greater initial number of DNA copies as input, the UMI coverage for these samples was significantly higher despite all samples having similar depth of coverage, given all samples were added at equimolar concentrations, as indicated in Fig. 1B and Supplementary Fig. S2 for the aforementioned samples. Consequently, most samples with $\geq$ 10 ng input had an average of 1–2 reads per UMI, whereas for most of the samples with lower DNA input each UMI family contained $>$ 3 reads. While the variation in DNA input for reference samples was not premeditated and resulted from the different criteria used between laboratories, these findings highlight the importance of DNA input in UMI panels and how sequencing resources should be allocated accordingly. Therefore, samples with higher DNA input should be sequenced at a higher depth of coverage to maintain the reads per UMI proportion, as multiple reads for each UMI are required for error correction.

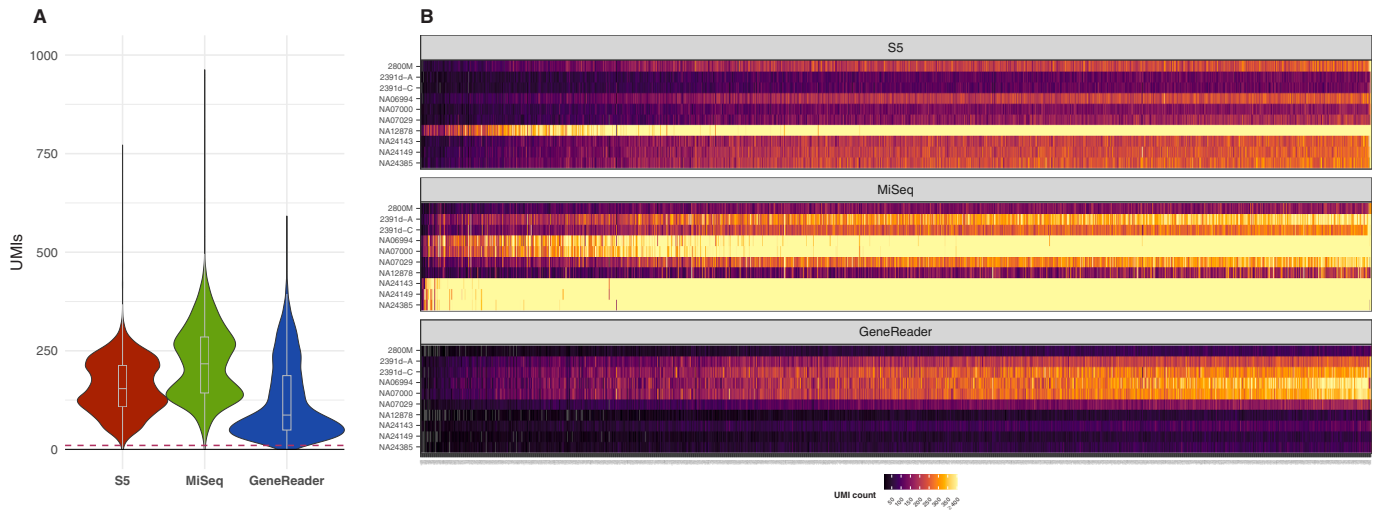


Fig. 1. A. Violin plot of the distribution of UMI counts for reference samples amplified using 1–2 ng DNA input. Maroon dashed line represents the 10 UMI calling threshold. N = 9, 3 and 10 for the Ion S5, MiSeq and GeneReader, respectively. Fig. 1**B.** Raster plot with UMI coverage for each marker for the reference sample set. Markers in grey represent loci with a UMI count lower than the calling threshold of 10 UMIs.

3.1.2. Base misincorporation

The base misincorporation rate was estimated from the percentage of reads of non-called alleles divided by total coverage for the marker. Mean misincorporation values were relatively low, with mean values below 1 %; i.e., levels similar to previous studies evaluating MPS panels for forensic use [26–28]. We obtained 0.564 ± 1.350, 0.253 ± 0.619 and 0.318 ± 0.920 percentage values for the Ion S5, MiSeq and the GeneReader, respectively; while a median value of zero was recorded across the whole panel due to the large proportion of markers showing no base misincorporation.

Overall, the Ion S5 exhibited the highest number of markers with high misincorporation values, with 73 loci presenting a mean misincorporation rate above 2 %, 17 of which were above 5 %. In contrast, the MiSeq had 15 loci above 2 % and 1 above 5 %, and the GeneReader had 18 above 2 % and 2 above 5 %. Three markers, rs4991456, rs60216576 and rs6764984, were observed to have high misincorporation rates for the three platforms. Another five showed high rates in two platforms: rs10795925, rs58318997 (MiSeq and GeneReader); rs13172903, rs370892 (S5 and MiSeq); and rs1762533 (S5 and GeneReader). Supplementary Fig. S3 contains misincorporation boxplots for each marker and Supplementary Table S3 lists all the SNPs with mean values above 2 % and their respective mean, standard deviation and median values.

Although no SNP surpassed a 20 % misincorporation rate (the genotype calling threshold used for heterozygotes), rs4991456 in the MiSeq and GeneReader and rs9933604, rs2184289 and rs2843150 in the Ion S5 gave noticeably high base misincorporation rates of ~18–19 % for particular samples, so these five SNPs have a substantially higher potential to produce erroneous genotype calls, particularly in mixed DNA samples. The three markers with a high misincorporation rate on the Ion S5 are all located close to homopolymeric tracts (poly-T tract near rs2184289, poly-G tracts near rs9933604 and rs2843150), so manual curation of these genotypes may be required. Inspection of rs4991456 did not reveal any alignment issues.

3.1.3. Strand bias

Supplementary Fig. S4 outlines the percentage of sequence reads in the forward strand vs reverse strand. Overall, markers showed a good balance between forward and reverse strand coverage. The Ion S5 showed the most balanced strand ratios, with 55.7 % of markers having median values in the 45–55 % range. Similarly, the MiSeq had 53.6 % of markers in this range, while the GeneReader had a noticeably less balanced output with 35.4 % of loci in the 45–55 % strand ratio range, likely to be caused by lower UMI coverage.

It should be noted that SNPs that tend to be read only in one strand on the Ion S5 and the GeneReader, show a tendency to be read on the other strand on the MiSeq – i.e., markers that are mostly read in the forward strand on MiSeq are read on the reverse strand on the other platforms and vice versa. Although the reason for this phenomenon remains unknown, we suggest this could be caused by the way paired-end reads in the MiSeq are interpreted by the software, rather than related to the chemistry.

3.1.4. Allele read frequency balance

Major allele read frequency (MAF) data, defined in this study as UMI reads for the allele with the highest sequence coverage divided by total UMI coverage, is represented in Supplementary Fig. S5. Ideally, homozygous genotype calls should be close to 100 % and heterozygous calls around 50 % per allele, meaning that the sequencing assay is perfectly balanced for each allele. The MiSeq had the most balanced output out of the three platforms, with 96.0 % of SNP calls falling in the homozygote coverage > 90 % and heterozygote coverage of 50–60 %. The Ion S5 had 91.6 % of loci in these ranges, and the GeneReader 88.6 %. The proportion of genotype calls with more acute imbalance, that is, with major allele frequencies between 50 % and 70 %, was 1.4 %, 2.8 % and 3.2 % for the MiSeq, Ion S5 and GeneReader, respectively. Despite most of the MiSeq samples having a higher DNA input, which may have led to better performance for this platform, the comparison of samples with ~1 ng input versus samples with higher input did not reveal any differences (data not shown).

Although observing a small proportion of markers with imbalanced number of reads is expected, especially in a multiplex of this size, several markers consistently gave imbalanced genotypes across the different platforms (i.e., within the 70–90 % MAF range). SNPs rs66525445, rs7623855, rs9617584, rs1097454, rs58318997, rs370892, rs10151146, rs4906213, rs8029886, rs10795925 and rs11103107 produced the highest number of imbalanced genotypes. SNPs rs370892 and rs10795925 were observed to also have high levels of base misincorporation, whilst rs58318997 had additional issues (detailed later). Although no other issues were observed in the eight remaining SNPs, their genotypes should also be interpreted with caution, as intermittent allele dropouts might be caused by undocumented neighbouring sequence features such as segmental duplications or copy number variation.

3.1.5. Baseline noise

Estimates of baseline sequencing noise from UMI coverage rates in the non-template negative control are outlined in Supplementary Fig. S6. Mean UMI coverage for the negative control was lower than 1 UMI per marker in the three platforms, with 29, 147 and 4 loci showing reads (representing ~2, 10 and 0.3 % of the markers in the multiplex) for the Ion S5, the MiSeq and the GeneReader, respectively. The Ion S5 and GeneReader had a maximum number of 7 and 1 UMI reads, respectively; both below the 10 UMI calling threshold. In contrast, the MiSeq negative control had several markers above this calling threshold (rs7657799, rs784685, rs11531789 and rs10147017), detecting up to 23 UMI reads for rs10147017. Although these measures are based on a single non-template control and an internal validation is required when implementing this tool in the analysis of casework DNA, these SNPs should have manual inspection.

3.2. Genotype concordance analysis

Genotype concordance assessed using reference samples with online genotype data from the independent sequencing analyses of the 1000 Genomes Project and GIAB Consortium, produced 7435 comparisons for each platform (the reference data did not contain full profiles in all cases). A concordance rate of 99.5 % was obtained for the Ion S5, 99.7 % for the MiSeq and 99.6 % for the GeneReader, discounting 63, 31 and 129 no-calls; respectively. We found 33 discordances for the Ion S5 (consisting in 26 drop-ins, 1 drop-out and 6 incorrect genotypes), 19 for the MiSeq (11 drop-ins, 5 drop-outs and 3 incorrect genotypes) and 26 for the GeneReader (9 drop-ins, 15 drop-outs and 2 incorrect genotypes). Despite the overall high genotyping concordance rates, several markers were found to produce consistent genotype discordancies, some platform-specific, and others consistently observed across all three MPS systems. Alignments for SNPs that were suspected to present alignment artifacts were manually inspected with IGV and are outlined in Supplementary File S1.

On the Ion S5, SNPs rs319305 and rs6764984 produced erroneous calls due to the presence of homopolymeric tracts in the neighbouring region. In rs319305, a 3' 7-T tract next to the target site, combined with an InDel, coincides with an A drop-in, and in rs6764984, a 5' InDel next to the target site and the 3' GGGAAAAAAAA tract next to the target site are likely to affect reliable sequence alignment. The presence of other high-frequency variants in the vicinity of the target SNPs also hindered correct alignment and manual revision was challenging due to the complexity of the alignment (see Supplementary File S1). Another SNP, rs7623855, which was observed previously to have acute imbalance in several samples, had allelic dropouts in several samples typed on the GeneReader due to the differences in coverage between both alleles. Additionally, various allele dropouts were found for rs1809429 on the MiSeq and GeneReader, while Ion S5 genotype calls in this SNP were concordant. Reads containing the missing allele were found to align to an off-target region (coinciding with the SNP rs183110408), indicating that the drop-outs observed could be caused by the incorrect alignment of sequences due to this homology.

The results obtained in SNPs rs58318997 (previously observed to have high misincorporation rates and to be highly imbalanced) and rs80266372 gave rise to incorrect genotypes or drop-ins in several samples, independent of the platform. IGV inspection did not reveal any features hindering alignment despite some alleles with sequence reads for more than two different alleles. A BLAT query for the reads showed similarities to other regions, corresponding to SNPs rs201552694 and rs1253935158 respectively, where reads aligning in these regions were found (Supplementary File 1). These findings suggested off-target amplification causing incorrect genotyping, therefore these SNPs should be removed from any analysis. Finally, the genotypes obtained for rs7629097 and rs4461915 were discordant with the 1000 Genomes data for NA06994 and NA07000. However, the results were concordant amongst the sequencing platforms of our study, so this discordancy remains without explanation.

The two Coriell control DNA trios produced some departures from mendelian inheritance patterns. In the CEU trio, non-mendelian incompatibilities were observed in rs1809429 due to the dropout of markers (e.g., rs1809429 on the MiSeq caused by imbalanced reads for both alleles, rs6478720 on the Ion S5 due to the high misincorporation values due to the target SNP being close to a poly-C tract). Nonetheless, two persistent incompatibilities were found in the CEU trio for SNPs rs2009433 and rs399737. Both parents' genotypes were concordant with the genotypes from the 1000 Genomes Project and all examinations made of the sequencing output failed to provide an explanation (no apparent dropin/dropout, no hinderance to alignments). SNP rs399737 is reported to map to multiple locations in the 1000 Genomes Ensembl browser, so should be discounted as a component marker from all future analyses using MPSplex.

3.3. Sensitivity analysis

Fig. 2A shows the UMI coverage distribution for the 2800 M dilution series and Fig. 2B plots the number of reads in each UMI family. Additional information about the total read depth observed per UMI is outlined in Supplementary Fig. S7. As would be expected, there was a decrease of UMI counts when the input DNA was reduced due to the correlation of UMIs detected with the original DNA copies. Total read depth on the other hand also shows a tendency to decrease as the DNA input is reduced, although this effect is not so perceptible due to the normalisation step, where all samples are quantified and pooled at equimolar concentrations. It is noteworthy that the 125 pg sample on the Ion S5 has a higher UMI coverage than the 250 pg, which could be explained by the higher depth of coverage from the sample that, in turn, could derive from an underestimation in the quantification.

Fig. 2C illustrates the proportion of loci that were concordant with the consensus 1000 pg profile from the three platforms. Both the Ion S5 and MiSeq presented high levels of sensitivity up to 125 pg, with concordance levels of ~95 % or higher. At lower DNA inputs levels, the Ion S5 exhibited a higher sensitivity, with ~50 % correctly called genotypes at 15 pg input and ~30 % with 7 pg of DNA. In comparison, the MiSeq the reduction in sequence data with DNA input observed is more acute, with ~20 % correctly called genotypes at 31 pg input and less than 10 % concordant calls with 15 pg. The GeneReader performed the least well with lower levels of DNA, displaying lower concordance rates than the other two platforms, with > 90 % concordance at 500 pg input DNA, but less than 50 % correctly called genotypes at 125 pg.

3.4. Mixture detection

Although tri-allelic SNPs are not as proficient as STRs or microhaplotypes for mixture detection, when compared to bi-allelic SNPs, they enhance the detection of mixtures, as not only are their allele read frequency ratios imbalanced, but more importantly, in the MPSplex panel which compiled tri-allelic SNPs with the maximum heterozygosity levels, a significant proportion of loci in mixed DNA will show sequence data consisting of three distinct alleles.

For the assessment of the mixture detection capabilities of MPSplex, the sequence patterns of SRM 2391d component D were analysed in detail. Fig. 3 summarises the percentage of loci where three alleles were detected, the proportion of alleles from the major and minor component detected in the mixture, and the major allele read frequency distribution of the single-source samples and the mixture for each of the MPS platforms. Clear signals of mixed DNA were observed in all platforms, with three alleles detected in ~18–21 % of tri-allelic SNPs, and more than 90 % of the genotypes of the minor contributor were obtained. Additionally, MAF patterns were altered compared to single-source samples, with many SNPs showing MAF ratios beyond their expected ranges observed in unadmixed DNA analyses.

In addition, even though we did not formally assess the capability of MPSplex for mixture deconvolution, it should be possible to retrieve

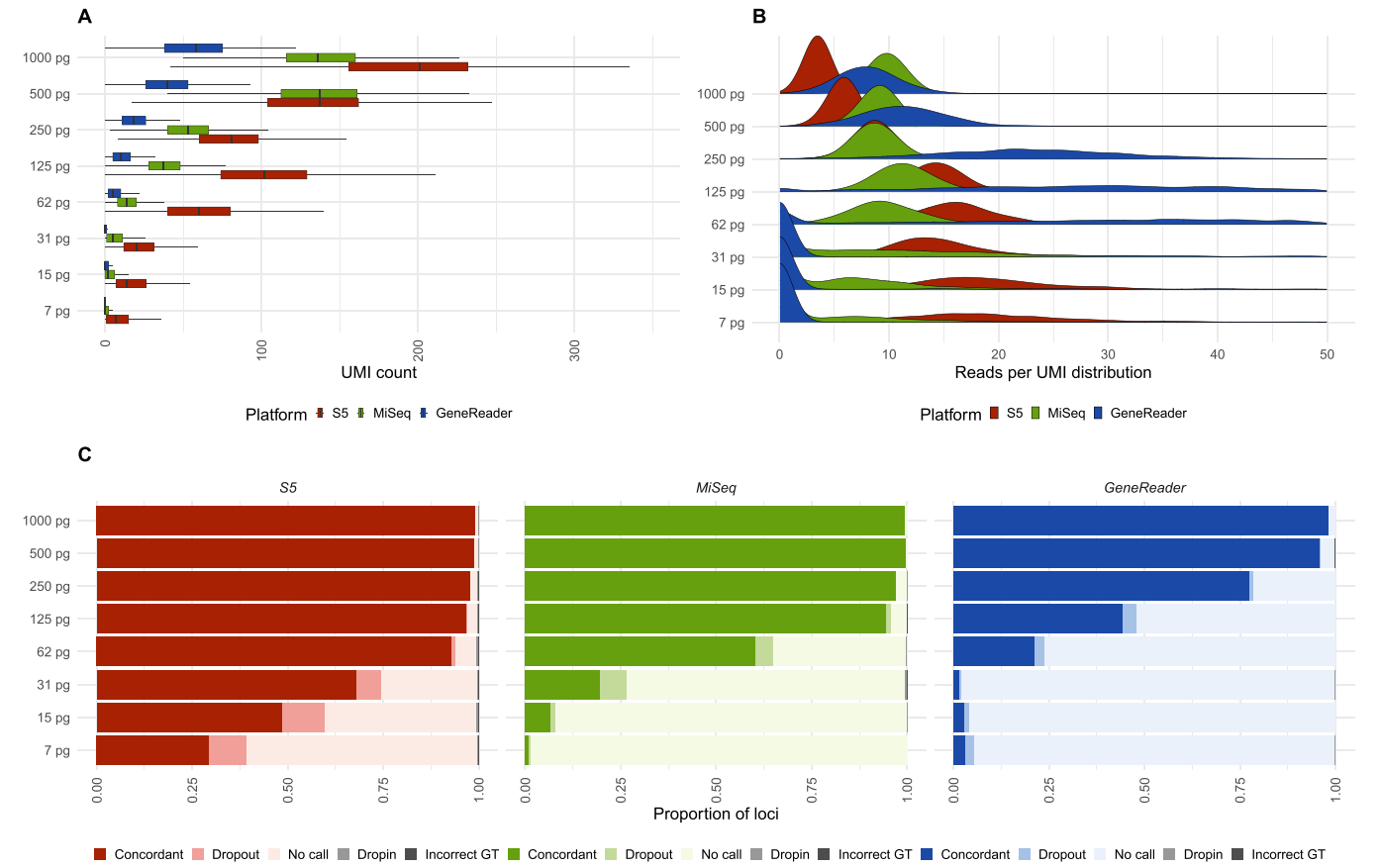


Fig. 2. Overall performance of the 2800 M dilution series in each of the three platforms. Fig. 2**A**. Boxplots of UMI coverage distribution per locus. Fig. 2**B** Ridgeline plot of number of reads per UMI read distribution. Fig. 2**C** Proportion of concordant genotypes obtained for each sample compared to the consensus 1000 pg reference profile. Incorrect GT: incorrect genotype.

partial profiles from the minor contributor based on skewed allele read frequency ratios from loci where three alleles are detected. However, factors such as high misincorporation for some SNPs or the possible effect of using UMI reads must be considered if this approach is applied further.

3.5. *Proof-of-concept kinship analysis*

A kinship case in which routine STR kits (Globalfiler™ and Investigator HDplex) did not provide enough statistical power for its resolution consisted of an indirect paternity test in which the paternity was analysed using two putative half-sisters. The hypotheses tested are detailed in Fig. 4A.

A total of 1061, 1124 and 1153 genotype calls out of 1235 autosomal tri-allelic SNPs were obtained for the CH1 and suggested sibships CH2 and CH3, respectively. It should be noted that aiSNPs, tri-allelic SNPs from the X chromosome and the microhaplotype composite SNPs were not used for the analyses. Additionally, the 6 autosomal tri-allelic SNPs identified as problematic for the Ion S5 (rs2009433, rs399737, rs58318997, rs80266372, rs319305 and rs6764984, as detailed in Section 3.2) were also disregarded.

Fig. 4B shows the probability distributions from the simulations performed with FamLink for both relationship hypotheses, using all the markers in the panel and with profiles containing calls for 70, 80 and 90 % of the SNPs. Due to the large number of markers present in the panel and the nature of the pedigrees assessed, several simulations gave genetic incompatibilities (therefore, failing to give an LR value), when the profiles were generated assuming H2 to be true, ~5,900 for the 10,000 simulations assuming a full profile. In other words, for the profiles simulated for CH1, CH2 and CH3 assuming H2, with CH2-CH3 being full siblings and CH1 as non-related to CH2 and CH3; some markers presented genetic inconsistences that could not be explained if H1 were true (e.g., CH2 and CH3 are homozygotes for alleles 2 and 3, respectively, and CH1 is homozygous for allele 1).

The LR value obtained from FamLink was 1.339e+044, around 42 orders of magnitude greater than the results obtained for STRs (an LR of 89.17) and falling within the expected LR distribution for the suggested relationship. Nonetheless, it was also observed that the LR value was positioned in the lower end of the LR distribution when compared to the expected value from simulation data using the whole set of SNPs, as the case samples did not result in genotypes for all MPSplex SNPs, while the simulations assumed complete SNP profiles. A total of 80.16 % of the single-site autosomal tri-allelic SNPs in MPSplex overlapped between CH1 and CH2. While CH1 and CH3 shared calls for 82.10 % of the markers and the three (CH1–3) shared 76.84 %. Therefore, simulations using partial profiles provided a better adjustment to the LR value obtained.

4. Discussion

From our results and experience of running multiple MPS chemistries and sequence detection systems we observed that the MPSplex assay developed with QIAseq technology gives robust and reliable performance as a forensic SNP genotyping assay, and at a larger genotyping scale than most such assays. Each of the three MPS systems evaluated had individual beneficial aspects – e.g., the Ion S5 displayed higher sensitivity, the MiSeq had lower base misincorporation values, and the GeneReader had fewer non-specific sequence reads. Despite these differences in sequencing performance, overall, more than 99 % of the SNPs had sequence coverage levels above the calling threshold. High

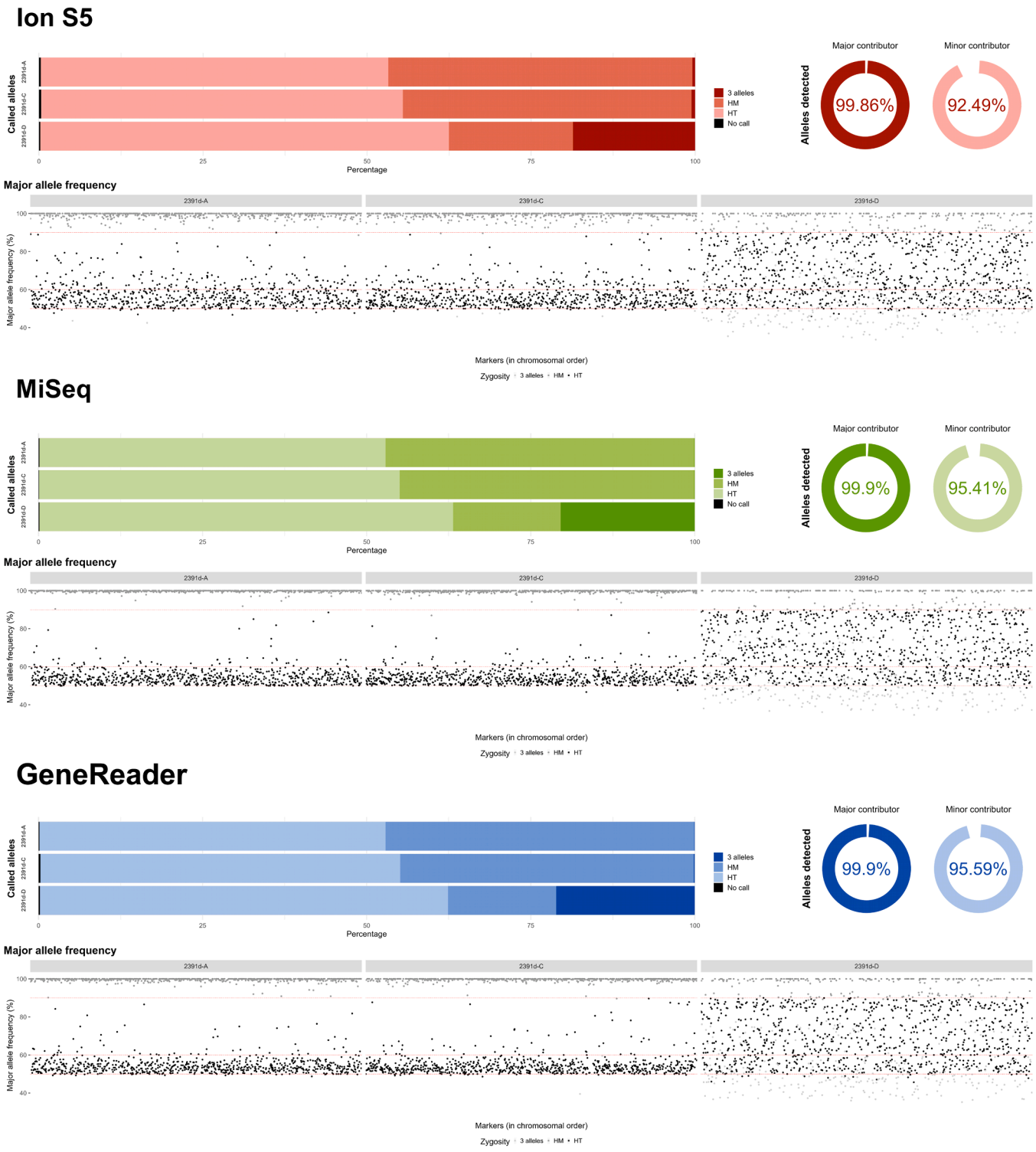


Fig. 3. Mixture analysis with MPSplex single-site autosomal tri-allelic SNPs on each sequencing platform. Upper left bar charts show percentage of heterozygous (HT), homozygous (HM), three-allele and no genotype calls for the two single-source samples and the mixture from the SRM 2391d control DNA set. Upper right circular plots show proportion of alleles called in the mixture for the major (2391d-A) and the minor (2391d-C) components. Bottom scatter plots show the major allele frequency for the SNPs analysed in three sample panes. Red dashed lines represent the optimum 50–60 % and > 90 % intervals for heterozygous and homozygous genotype calls, respectively.

levels of sensitivity were observed with 250 pg of input DNA and as low as 31–62 pg on the MiSeq and Ion S5 platforms. Mixture detection, albeit evaluated with a limited control DNA experiment, was also successful in each platform, although this is, in part, due to the detection of three alleles for a sizeable proportion of the tri-allelic component SNPs of MPSplex.

Despite having high genotype concordance values (≥ 99.5 % concordant results), some SNPs were identified as problematic, in many cases because primers are targeting SNPs previously not identified to be sited in duplicated regions. SNPs rs58318997 and rs80266372 were found to have three alleles in some samples; and rs2009433 and rs399737 were found to depart from mendelian inheritance patterns, so

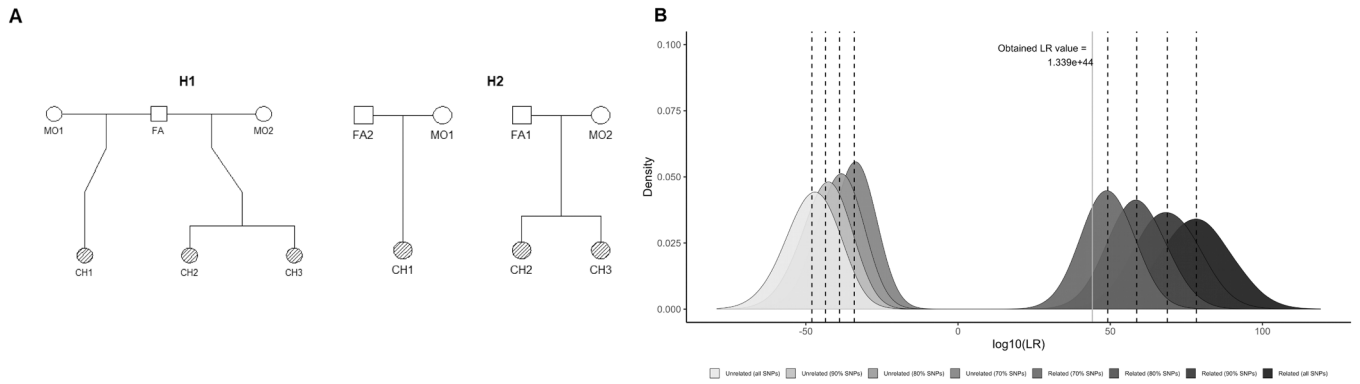


Fig. 4. A. Pedigrees of the two hypotheses assessed in the kinship study, comprising: CH1 is related to CH2 and CH3 (H1) vs CH1 unrelated (H2), with shaded symbols representing individuals that were genotyped. Fig. 4**B**. Log LR distributions for the simulations performed with FamLink. Log LR distributions have graded greyscale shading to represent different degrees of SNP genotype data completeness of 70 %, 80 %, 90 % and 100 % used in the simulations. Black dashed lines represent mean LR values for the simulations and the grey line represents the LR value obtained in support of H1 with MPSplex.

these four SNPs should be excluded from further analyses using the current MPSplex primer sets. In terms of individual marker performance, SNP rs4991456 on the MiSeq and GeneReader; and SNPs rs9933604, rs2184289 and rs2843150 on the Ion S5 were found to have high levels of base misincorporation, which would require manual scrutiny of sequence output using IGV, and thus involving a labour-intensive process. SNPs rs319305 and rs6764984 also produced erroneous calls on the Ion S5 due to the presence of allele drop-ins, while rs7623855 on the MiSeq and rs1809429 on the MiSeq and GeneReader had miscalls due to allele drop-out. Supplementary Table S4 summarises all the underperforming SNPs identified in our study along with details of the issues that were identified. In addition, IGV screenshots for the markers for which alignment issues were observed are included in Supplementary File S1.

The kinship study we performed served as a simple proof-of-concept of the potential of MPSplex, and demonstrated its applicability to relationship tests where routine STR and mtDNA analyses cannot provide sufficient power of discrimination. Although in the current field, the Kintelligence and FORCE forensic panels have now been developed, they are focussed on genealogical applications and the information they provide may exceed what would be required in many relationship testing scenarios. These panels also require the allocation of extra capacity in the sequencing run, increasing the cost per sample. Therefore, MPSplex represents a simpler, but highly informative "middle ground" solution for cases in which the application of MPS technologies is required but the degree of relatedness between missing persons and relatives is less distant.

In comparison to other smaller scale MPS forensic panels, MPSplex displayed comparable levels of performance. The error rate was low when considering the whole multiplex, shown by the median misincorporation value of 0 obtained for the three platforms and the high levels of genotype concordance. Some SNPs nonetheless had high levels of base misincorporation, especially for the Ion S5 platform. Although the number of markers showing high levels of misincorporation was higher than other MPS panels, this factor should be considered in the light of the bigger size of the MPSplex multiplex. Previous studies also report a similar behaviour using this platform, on many occasions due to the presence of homopolymeric tracts target SNP site that hinder correct alignment [2,26]. In terms of sensitivity, we observed similar or less sensitive outputs when compared to lower-scale SNP multiplexes on the Ion S5 and MiSeq [9,13,26,27]. Bigger multiplexes on the other hand showed reduced sensitivity with less complete results in terms of percentage of markers genotypes, although the total number of markers genotyped was higher in some instances due to the larger number of SNPs present in those panels [14,29].

Although we did not formally assess the improvement of sequencing results obtained when the UMI information is considered, the studies of Woerner et al., Crysup et al., and Staadig et al., [14–16] indicate that this technology increases the accuracy of the results. Nevertheless, several factors should be taken into consideration when working with UMI-based sequencing technology in a forensic context. First, UMI technology was previously developed for non-forensic applications, where neither DNA concentration nor DNA integrity are a problem. Therefore, the UMI workflow contains several steps that may pose a problem when dealing with forensic samples. A total of four magnetic bead clean-ups are required for library preparation, which is likely to increase the risk of target DNA loss. We note that the recently introduced QIAseq PRO kit could potentially alleviate this issue, as it replaces the bead clean-ups after adapter ligation and target enrichment with enzymatic clean-ups. In addition, a fragmentation step is performed at the beginning of the protocol, and this procedure could present issues for the analysis of degraded or low-level DNA. Although the protocol contains an adaptation for FFPE DNA with reduced fragmentation time and Staadig's study obtained promising results with a UMI panel used to analyse bone samples, the applicability of UMI-based MPS for this type of sample needs further evaluation of the effects of variation in fragmentation time by sequencing a sufficiently broad-based collection of highly degraded samples.

Another issue that is likely to arise with UMI technology in forensic applications is the increased complexity of analysis demanded for the interpretation of sequencing data, as not only coverage needs to be considered, but also the UMI count for each locus and the number of reads for each UMI family. Additionally, prior information on the likely quantity and quality of the DNA is required to correctly pool each library. Otherwise, as evident in the samples amplified with ≥ 10 ng of DNA, each UMI family could be under-represented and consequently that will affect the quality of the sequencing output.

Despite the above reservations, our findings indicate that the use of MPSplex for missing persons analysis based on UMI technology will fulfil the need for a medium-scale marker panel that can provide informative likelihoods in challenging kinship tests where relationships are either too distant for conventional STR data to resolve, or the pedigree of surviving relatives is too incomplete and requires a much higher number of markers. At the time of writing, Qiagen are withdrawing their GeneReader technology, but our study shows the workflows of MiSeq and Ion S5 MPS platforms can be readily adapted to perform UMI-based sequencing with the expectation that a large proportion of complex kinship tests can be resolved when use of STRs alone fails to provide high enough likelihoods.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

J.R. was supported by the "Programa de axudas á etapa predoutoral" funded by the Consellería de Cultura, Educación e Ordenación Universitaria e da Consellería de Economía, Emprego e Industria from Xunta de Galicia, Spain (ED481A-2020/039). MVL is supported by grant PID2019–107876RB-I00 funded by the Ministerio de Educación, Cultura y Ciencia, Spain (MCIN/AEI/10.13039/501100011033) and grant PID2022–141224OB-I00 funded by MCIN/AEI/10.13039/501100011033 and "ERDF A way of making Europe". MdlP is supported by grant IJC2020–042638-I funded by the Gobierno de España MCIN/AEI/10.13039/501100011033 and the European Union "NextGenerationEU/PRTR". Data analysis was partially undertaken in the FinisTerrae III supercomputer from the Centro de Supercomputación de Galicia (CESGA). We would like to thank Keith Elliot and Holger Karas from Qiagen for their support and technical assistance. A special thank you goes to the ISFG board for supporting J.R. with the Short Time Fellowships program allowing this study to be accomplished at ICMP, The Hague, Netherlands.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.fsigen.2025.103233.

References

[1] N. Hiroaki, F. Koji, K. Tetsushi, S. Kazumasa, N. Hiroaki, S. Kazuyuki, Approaches for identifying multiple-SNP haplotype blocks for use in human identification, Leg. Med. 17 (5) (2015) 415–420.
[2] M. de la Puente, C. Phillips, C. Xavier, J. Amigo, A. Carracedo, W. Parson, M. V. Lareu, Building a custom large-scale panel of novel microhaplotypes for forensic identification using MiSeq and Ion S5 massively parallel sequencing systems, Forensic Sci. Int. Genet. 45 (2020) 102213.
[3] F. Oldoni, D. Bader, C. Fantinato, S.C. Wootton, R. Lagace, K.K. Kidd, D. Podini, A sequence-based 74plex microhaplotype assay for analysis of forensic DNA mixtures, Forensic Sci. Int. Genet. 49 (2020) 102367.
[4] K. Breslin, B. Wills, A. Ralf, M. Ventayol Garcia, M. Kukla-Bartoszek, E. Pospiech, A. Freire-Aradas, C. Xavier, S. Ingold, M. de La Puente, K.J. van der Gaag, N. Herrick, C. Haas, W. Parson, C. Phillips, T. Sijen, W. Branicki, S. Walsh, M. Kayser, HIrisPlex-S system for eye, hair, and skin color prediction from DNA: Massively parallel sequencing solutions for two common forensically used platforms, Forensic Sci. Int. Genet. 43 (2019) 102152.
[5] M. Diepenbroek, B. Bayer, K. Schwender, R. Schiller, J. Lim, R. Lagace, K. Anslinger, Evaluation of the ion AmpliSeq PhenoTrivium panel: MPS-based assay for ancestry and phenotype predictions challenged by casework samples, Genes (Basel) 11 (12) (2020).
[6] J. Ruiz-Ramirez, M. de la Puente, C. Xavier, A. Ambroa-Conde, J. Alvarez-Dios, A. Freire-Aradas, A. Mosquera-Miguel, A. Ralf, C. Amory, M.A. Katsara, T. Khellaf, M. Nothnagel, E.Y.Y. Cheung, T.E. Gross, P.M. Schneider, J. Uacyisrael, S. Oliveira, M.D.N. Klautau-Guimaraes, C. Carvalho-Gontijo, E. Pospiech, W. Branicki, W. Parson, M. Kayser, A. Carracedo, M.V. Lareu, C. Phillips, V. Consortium, Development and evaluations of the ancestry informative markers of the VISAGE enhanced tool for appearance and ancestry, Forensic Sci. Int Genet 64 (2023) 102853.
[7] M. Kayser, W. Branicki, W. Parson, C. Phillips, Recent advances in Forensic DNA Phenotyping of appearance, ancestry and age, Forensic Sci. Int Genet 65 (2023) 102870.
[8] I. Grandell, R. Samara, A.O. Tillmar, A SNP panel for identity and kinship testing using massive parallel sequencing, Int. J. Leg. Med. 130 (4) (2016) 905–914.
[9] M. de la Puente, C. Phillips, C. Santos, M. Fondevila, A. Carracedo, M.V. Lareu, Evaluation of the Qiagen 140-SNP forensic identification multiplex for massively parallel sequencing, Forensic Sci. Int. Genet. 28 (2017) 35–43.
[10] A. Tillmar, K. Sturk-Andreaggi, J. Daniels-Higginbotham, J.T. Thomas, C. Marshall, The FORCE Panel: an all-in-one SNP marker set for confirming investigative genetic genealogy leads and for general forensic applications, Genes (Basel) 12 (12) (2021).
[11] C. Phillips, J. Amigo, A.O. Tillmar, M.A. Peck, M. de la Puente, J. Ruiz-Ramirez, F. Bittner, S. Idrizbegovic, Y. Wang, T.J. Parsons, M.V. Lareu, A compilation of tri-allelic SNPs from 1000 Genomes and use of the most polymorphic loci for a large-scale human identification panel, Forensic Sci. Int. Genet. 46 (2020) 102232.
[12] K.K. Kidd, W.C. Speed, A.J. Pakstis, M.R. Furtado, R. Fang, A. Madbouly, M. Maiers, M. Middha, F.R. Friedlaender, J.R. Kidd, Progress toward an efficient panel of SNPs for ancestry inference, Forensic Sci. Int. Genet. 10 (2014) 23–32.
[13] D. Truelsen, A. Freire-Aradas, M. Nazari, A. Aliferi, D. Ballard, C. Phillips, N. Morling, V. Pereira, C. Borsting, Evaluation of a custom QIAseq targeted DNA panel with 164 ancestry informative markers sequenced with the Illumina MiSeq, Sci. Rep. 11 (1) (2021) 21040.
[14] A. Staadig, J. Hedman, A. Tillmar, Applying unique molecular indices with an extensive all-in-one forensic SNP panel for improved genotype accuracy and sensitivity, Genes (Basel) 14 (4) (2023).
[15] B. Crysup, S. Mandape, J.L. King, M. Muenzler, K.B. Kapema, A.E. Woerner, Using unique molecular identifiers to improve allele calling in low-template mixtures, Forensic Sci. Int. Genet. 63 (2023) 102807.
[16] A.E. Woerner, S. Mandape, J.L. King, M. Muenzler, B. Crysup, B. Budowle, Reducing noise and stutter in short tandem repeat loci with unique molecular identifiers, Forensic Sci. Int. Genet. 51 (2021) 102459.
[17] K.K. Kidd, W.C. Speed, A.J. Pakstis, D.S. Podini, R. Lagace, J. Chang, S. Wootton, E. Haigh, U. Soundararajan, Evaluating 130 microhaplotypes across a global set of 83 populations, Forensic Sci. Int. Genet. 29 (2017) 29–37.
[18] C. Phillips, D. McNevin, K.K. Kidd, R. Lagacé, S. Wootton, M. de la Puente, A. Freire-Aradas, A. Mosquera-Miguel, M. Eduardoff, T. Gross, L. Dagostino, D. Power, S. Olson, M. Hashiyada, C. Oz, W. Parson, P.M. Schneider, M.V. Lareu, R. Daniel, MAPlex - A massively parallel sequencing ancestry analysis multiplex for Asia-Pacific populations, Forensic Sci. Int.: Genet. 42 (2019) 213–226.
[19] M. Byrska-Bishop, U.S. Evani, X. Zhao, A.O. Basile, H.J. Abel, A.A. Regier, A. Corvelo, W.E. Clarke, R. Musunuri, K. Nagulapalli, S. Fairley, A. Runnels, L. Winterkorn, E. Lowy, C. Human Genome Structural Variation, F. Paul, S. Germer, H. Brand, I.M. Hall, M.E. Talkowski, G. Narzisi, M.C. Zody, High-coverage whole-genome sequencing of the expanded 1000 Genomes Project cohort including 602 trios, Cell 185 (18) (2022) 3426–3440, e19.
[20] J. Wagner, N.D. Olson, L. Harris, Z. Khan, J. Farek, M. Mahmoud, A. Stankovic, V. Kovacevic, B. Yoo, N. Miller, J.A. Rosenfeld, B. Ni, S. Zarate, M. Kirsche, S. Aganezov, M.C. Schatz, G. Narzisi, M. Byrska-Bishop, W. Clarke, U.S. Evani, C. Markello, K. Shafin, X. Zhou, A. Sidow, V. Bansal, P. Ebert, T. Marschall, P. Lansdorp, V. Hanlon, C.A. Mattsson, A.M. Barrio, I.T. Fiddes, C. Xiao, A. Fungtammasan, C.S. Chin, A.M. Wenger, W.J. Rowell, F.J. Sedlazeck, A. Carroll, M. Salit, J.M. Zook, Benchmarking challenging small variants with linked and long reads, Cell Genom. 2 (5) (2022).
[21] R Core Team, R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing (2023).
[22] Bam-readcount software at: https://github.com/genome/bam-readcount.
[23] D. Kling, T. Egeland, A.O. Tillmar, FamLink–a user friendly software for linkage calculations in family genetics, Forensic Sci. Int. Genet. 6 (5) (2012) 616–620.
[24] A.O. Tillmar, C. Phillips, Evaluation of the impact of genetic linkage in forensic identity and relationship testing for expanded DNA marker sets, Forensic Sci. Int.: Genet. 26 (2017) 58–65.
[25] T. Egeland, P.F. Mostad, B. Olaisen, A computerised method for calculating the probability of pedigrees from genetic data, Sci. Justice 37 (4) (1997) 269–274.
[26] C. Xavier, M. de la Puente, A. Mosquera-Miguel, A. Freire-Aradas, V. Kalamara, A. Ralf, A. Revoir, T.E. Gross, P.M. Schneider, C. Ames, C. Hohoff, C. Phillips, M. Kayser, W. Parson, V. Consortium, Development and inter-laboratory evaluation of the VISAGE Enhanced Tool for Appearance and Ancestry inference from DNA, Forensic Sci. Int. Genet. 61 (2022) 102779.
[27] C. Xavier, M. de la Puente, M. Sidstedt, K. Junker, A. Minawi, M. Unterlander, Y. Chantrel, F.X. Laurent, A. Delest, C. Hohoff, I. Bastisch, J. Hedman, K.J. van der Gaag, T. Sijen, W. Parson, Evaluation of the VISAGE basic tool for appearance and ancestry inference using ForenSeq(R) chemistry on the MiSeq FGx(R) system, Forensic Sci. Int. Genet. 58 (2022) 102675.
[28] M. Eduardoff, T.E. Gross, C. Santos, M. de la Puente, D. Ballard, C. Strobl, C. Borsting, N. Morling, L. Fusco, C. Hussing, B. Egyed, L. Souto, J. Uacyisrael, D. Syndercombe Court, A. Carracedo, M.V. Lareu, P.M. Schneider, W. Parson, C. Phillips, E.U.-N. Consortium, W. Parson, C. Phillips, Inter-laboratory evaluation of the EUROFORGEN Global ancestry-informative SNP panel by massively parallel sequencing using the Ion PGM, Forensic Sci. Int. Genet. 23 (2016) 178–189.
[29] J. Antunes, P. Walichiewicz, E. Forouzmand, R. Barta, M. Didier, Y. Han, J. C. Perez, J. Snedecor, C. Zlatkov, G. Padmabandu, L. Devesse, S. Radecke, C. L. Holt, S.A. Kumar, B. Budowle, K.M. Stephens, Developmental validation of the ForenSeq(R) Kintelligence kit, MiSeq FGx(R) sequencing system and ForenSeq Universal Analysis Software, Forensic Sci. Int. Genet. 71 (2024) 103055.