



ELSEVIER

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

Research paper

Development and validation of the VISAGE AmpliSeq basic tool to predict appearance and ancestry from DNA

Catarina Xavier^{a,*}, Maria de la Puente^{a,b}, Ana Mosquera-Miguel^b, Ana Freire-Aradas^b, Vivian Kalamara^c, Athina Vidaki^c, Theresa E. Gross^d, Andrew Revoir^e, Ewelina Pośpiech^f, Ewa Kartasińska^g, Magdalena Spólnicka^g, Wojciech Branicki^{f,g}, Carole E. Ames^e, Peter M. Schneider^d, Carsten Hohoff^h, Manfred Kayser^c, Christopher Phillips^b, Walther Parson^{a,i,*}, on behalf of the VISAGE Consortium

^a Institute of Legal Medicine, Medical University of Innsbruck, Innsbruck, Austria

^b Forensic Genetics Unit, Institute of Forensic Sciences, University of Santiago de Compostela, Spain

^c Department of Genetic Identification, Erasmus MC University Medical Center Rotterdam, Rotterdam, the Netherlands

^d Institute of Legal Medicine, Faculty of Medicine and University Hospital Cologne, University of Cologne, Cologne, Germany

^e Metropolitan Police Service London, United Kingdom

^f Malopolska Centre of Biotechnology, Jagiellonian University, Kraków, Poland

^g Central Forensic Laboratory of the Police, Warsaw, Poland

^h Institut für Forensische Genetik GmbH, Münster, Germany

ⁱ Forensic Science Program, The Pennsylvania State University, University Park, PA, USA

ARTICLE INFO

Keywords:

Forensic DNA phenotyping

Appearance and bio-geographical ancestry prediction

MPS Ion S5

AmpliSeq

SNP multiplex

ABSTRACT

Forensic DNA phenotyping is gaining interest as the number of applications increases within the forensic genetics community. The possibility of providing investigative leads in addition to conventional DNA profiling for human identification provides new insights into otherwise “cold” police investigations. The ability of reporting on the bio-geographical ancestry (BGA), appearance characteristics and age based on DNA obtained from a crime scene sample of an unknown donor makes the exploration of such markers and the development of new methods meaningful for criminal investigations. The VISible Attributes through GENomics (VISAGE) Consortium aims to disseminate and broaden the use of predictive markers and develop fully optimized and validated prototypes for forensic casework implementation. Here, the first VISAGE appearance and ancestry tool development, performance and validation is reported. A total of 153 SNPs (96.84 % assay conversion rate) were successfully incorporated into a single multiplex reaction using the AmpliSeq™ design pipeline, and applied for massively parallel sequencing with the Ion S5 platform. A collaborative effort involving six VISAGE laboratory partners was devised to perform all validation tests. An extensive validation plan was carefully organized to explore the assay’s overall performance with optimum and low-input samples, as well as with challenging and casework mock samples. In addition, forensic validation studies such as concordance and mixture tests recurring to the Coriell sample set with known genotypes were performed. Finally, inhibitor tolerance and specificity were also evaluated. Results showed a robust, highly sensitive assay with good overall concordance between laboratories.

1. Introduction

The primary focus of forensic genetics research is the development of new markers and techniques that allow for human individual identification. However, the applications of comparative approaches such as standard forensic STR profiling are limited in cases without suspects and/or national DNA database matches. This has prompted the forensic

genetics community to start the development of tools to infer information about the donor of biological traces found at the crime-scene to be used in police investigations to help find unknown perpetrators of crime. These investigative DNA analyses have been termed Forensic DNA Phenotyping (FDP), which includes three components: the inference of bio-geographical ancestry (BGA), the prediction of externally visible characteristics (EVC), and the estimation of chronological age.

* Corresponding authors at: Institute of Legal Medicine, Medical University of Innsbruck, Müllerstraße 44, 6020 Innsbruck, Austria.

E-mail addresses: catarina.gomes@i-med.ac.at (C. Xavier), walther.parson@i-med.ac.at (W. Parson).

<https://doi.org/10.1016/j.fsigen.2020.102336>

Received 13 February 2020; Received in revised form 28 May 2020; Accepted 8 June 2020

Available online 20 June 2020

1872-4973/ © 2020 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

FDP can serve as biological witness in cases without human eyewitnesses, and also for the corroboration of eyewitness testimonies [1]. FDP is suitable for current and previous unsolved cases (cold cases) without STR profile matches [2–4]. Other applications of FDP are missing person and disaster victim identification (DVI) cases to provide information in order to locate relatives or ante mortem samples. In all these applications, FDP can help the police by narrowing down the typically long list of putative suspects, relatives of victims, or victims.

Earlier FDP tools relied on conventional genotyping technologies deemed suitable for forensic DNA analysis, e.g. SNaPshot-based minisequencing, which, however, are limited in the number of DNA markers that can be simultaneously analyzed [5–19]. More recent studies have started to apply targeted massively parallel sequencing (MPS) technologies for FDP tool developments, typically separating EVCs from BGA [20–27]. Among those, three commercial kits are available i.e. i) the Precision ID Ancestry Panel (Thermo Fisher Scientific, [23]) for BGA, ii) the Ion AmpliSeq™ DNA Phenotyping Panel (Thermo Fisher Scientific) for hair and skin color prediction based on [14] and iii) the ForenSeq DNA Signature Prep Kit (Verogen, [22]) for EVCs and BGA (plus DNA markers for other purposes). The Precision ID Ancestry Panel [23] comprises 165 BGA markers described in previous studies [28,29]. The ForenSeq™ DNA Signature Prep Kit [22], particularly its DNA Primer Mix B, tests 56 BGA markers from a previous study [29] and 24 SNPs for eye and hair color (of which 2 markers overlap) from previous studies [11,12]. Recently, a non-commercial MPS application of the HIRISplex-S SNP panel for simultaneous prediction of eye, hair and skin color was published [25] for both the AmpliSeq-based Ion S5 System (Thermo Fisher Scientific) as well as the MiSeq FGx (Illumina) MPS platforms. Furthermore, other non-commercial SNP panels have been recently developed for BGA inference using targeted MPS, such as the EUROFORGEN Global AIMs SNP set with 128 BGA markers [21] and more recently, the MAPlex panel with 144 BGA SNPs and 20 BGA microhaplotypes [24].

In this context, the VISible Attributes through GENomics (VISAGE) Consortium was created in 2017 to develop and validate new MPS-based prototype tools for predicting BGA, appearance and age of an unknown crime scene sample donor from evidence DNA, among other project goals (<http://www.visage-h2020.eu/>). Here, the development and the validation of the VISAGE Basic Tool for Appearance and Ancestry prediction from DNA (herein BT A&A) is described for its AmpliSeq-based version (herein BT A&A (Amp)), representing the first FDP lab tool that combines DNA analysis of eye, hair, and skin color with continental bio-geographic ancestry in a single assay.

2. Materials and methods

2.1. DNA markers

Selection of DNA markers for the BT A&A focused on previously established knowledge. For appearance, this concerned eye, hair and skin color as previously established appearance traits predictable from DNA, and we employed the 41 DNA markers included in the previously described IrisPlex [11], HIRISplex [14] and HIRISplex-S [30] test systems. For BGA, the most ancestry-informative markers (AIM-SNPs) were taken from the previously developed EUROFORGEN Global AIMs MPS ancestry panel [21,31], supplemented with two additional AIM-SNPs from the 55 AIMs panel of Kiddlab (rs10497191, rs6990312, [29]). Furthermore, eleven additional AIM-SNPs included in the Thermo Fisher Precision ID ancestry panel (rs12130799, rs12629908, rs1834619, rs2269793, rs3737576, rs459920, rs4781011, rs4918664, rs705308, rs7226659, rs870347) were added. At the time of AIM-SNP selection these had just become established for forensic ancestry analysis with the Ion S5 (please note that TFS Precision ID ancestry panel also includes all 55 SNPs of the Kiddlab panel [29]). In addition to continental population comparisons, the differentiation of the sub-continental population groups in Eurasia of Middle East and South Asia

was an important target for the Basic Tool. The most informative Middle East AIM-SNPs were chosen from the EUROFORGEN NAME panel [27] and the most informative South Asian AIM-SNPs from the original candidate SNPs compiled for the *Eurasiaplex* forensic ancestry panel [32–34]. Lastly, 20 candidate tri-allelic SNPs with ancestry-informative allele frequency distributions were targeted to provide scope for mixed DNA detection – identified by the potential presence of three alleles in the sequence data of these SNPs from multiple contributors. The final AIM-SNP proportions in the developed BT A&A tool comprised: 56 AIM-SNPs from the Global AIMs panel with well-balanced and powerful five group differentiation; 12 AIM-SNPs selected for Middle East; 19 AIM-SNPs selected for South Asia; 2 + 11 additional AIM-SNPs from the Thermo Fisher Precision ID ancestry panel; and 15 tri-allelic AIM-SNPs. Three pigmentation predictive SNPs: rs16891982, rs1426654, rs12913832 are also standard BGA markers and were used in the BT A&A for both purposes; leading to a total of 115 BGA markers, of which 3 were already part of the EVC panel and 15 were tri-allelic loci. Considering BGA and appearance DNA markers together, 153 DNA markers (Supplementary Table S1, Supplementary Figure S1) were used for the design of the BT A&A. Even though no formal linkage evaluation was performed for the AIM markers, a rule of 1Mb minimum separation between syntenic SNP pairs was applied for marker selection (except for one transgression between SNPs rs12629908 and rs12498138 of 0.94 Mb).

2.2. Assay development, protocol and data analysis

For the currently described AmpliSeq-based version of the BT A&A, the Ion AmpliSeq Designer algorithm (<https://ampliseq.com/>, Thermo Fisher Scientific, Waltham, Massachusetts, USA, herein TFS) was used to design and manufacture a single primer pool containing all DNA markers for appearance and ancestry inference. The Ion AmpliSeq designer algorithm has proven effective with other similar-sized panels for human identification, ancestry prediction and even for full mitochondrial DNA sequencing [20,21,35]. Furthermore, the size of the amplicons can be selected considering the application in downstream analysis.

All libraries were prepared automatically using the Precision ID DL8 Kit (TFS) and Ion Code barcodes on the Ion Chef System (TFS) following the manufacturer's protocols [36]. All produced pools (each batch of eight libraries produces one pool) were quantified using the Ion Library Quantitation Kit (TFS), and then two pools (16 libraries in total) were combined once again equimolarly at 30 pM, when possible (or undiluted when not reaching a concentration of 30 pM). All final library pools were loaded onto the Ion Chef System for template preparation using the Ion S5 Precision ID Chef & Sequencing Kit (TFS) and loaded automatically onto Ion 530 Chips (TFS). Finally, sequencing was performed using the Ion S5 System (TFS, herein Ion S5), 16 libraries (2 DL8 batches) per Ion 530 Chip.

Raw data were processed using the S5 Torrent Server applying Torrent Suite software V.5.6.0 (TFS). A TMAP alignment against the hg19 genome assembly was performed. All BAM and BAI files produced from the sequencing were manually inspected using Integrative Genome Viewer (IGV) [51]. Genotypes were called with the plugin HID_SNP_Genotyper V.5.2.2 (herein, SNP Genotyper) using default parameters. All statistical analyses were performed using Microsoft Excel and R (<https://www.r-project.org/>, [37,38]).

2.3. Validation

A comprehensive validation plan was designed for assessing the overall performance of the BT A&A (Amp) assay. The plan comprised of common developmental tests, including reproducibility, sensitivity, mixtures and specificity, as well as challenging samples, such as mock casework samples and artificially degraded DNA controls, and inhibitor tolerance tests. The validation testing was shared among several

VISAGE Consortium partners to perform all tests needed for the validation of the BT A&A (Amp) assay. In total, 18 of the 530 chips in nine Ion S5 initializations were run in six VISAGE partner laboratories.

2.3.1. Reproducibility, sensitivity and altered PCR protocol

Reproducibility of the BT A&A (Amp) assay was assessed by preparing three replicates of 2800 M control DNA (Promega, Madison, WI, USA) at three different optimum inputs of 1, 2 and 10 ng in five different VISAGE laboratories (a total of 15 replicates per input were used for analysis). The sequencing results allowed not only to analyze the accuracy and precision of the method, but also to explore read depth distribution, strand bias and rates of misincorporation. Furthermore, typing DNA replicates under optimal conditions allowed the identification and description of poorly performing SNPs.

A sensitivity dilution series from 1 ng to 0.01 ng input (1, 0.5, 0.25, 0.1, 0.05, 0.025, 0.01 ng) of 2800 M control DNA (Promega) was prepared in duplicate and analyzed to identify the sensitivity threshold of the assay and describe the genotype variations or inconsistencies occurring at low DNA inputs. Five different VISAGE laboratories performed the same sequencing run, thus in total, ten replicates of each dilution step were included in the analysis. In addition, an increased cycle number PCR protocol (normal protocol - 22 cycles; modified protocol - 27 cycles) was tested using the same dilution series. The manufacturer (TFS) suggests such increased cycle protocols, particularly when preparing challenging or low-level DNA samples (< 1 ng). This experiment was repeated by three different laboratories; therefore, three replicates per dilution step were used for the final analysis.

2.3.2. Concordance and mixtures

Six Coriell samples (NA07029, NA07000, NA06994, NA11200, NA10540 and NA18498) were analyzed with the BT A&A (Amp) assay for concordance analysis. The concordance tests were analyzed taking into account: a) the known and accessible genotypes of the Coriell samples in the 1000 Genomes Phase III data [39] and Simons Foundation Genome Diversity Project (SGDP) [40] databases; and b) the concordance between genotypes obtained from three distinct VISAGE Consortium laboratories. Discordances with the 1000 Genomes Phase III data were further inspected with the 1000 Genomes Project New York Genome Center high coverage dataset.

Mixture identification and deconvolution is usually not described as a goal for EVC and BGA typing methods, as these typically follow STR analysis in a routine workflow. Hence, the STR results would indicate the presence of a single source sample, which is preferred for EVC and BGA analysis. Only under rare scenarios is it possible to derive EVC and BGA information reliably from mixed DNA samples. However, for the sake of completeness and because mixture detection in SNP panels, and particularly in EVC and BGA panels, has been previously described and published [20,21,25,41], this study was included in the validation. It is based on the fluctuation of allele frequencies and increase of heterozygosity values that indicate the presence of additional contributors. Two Coriell samples with different biogeographical ancestries were chosen to produce mixed profiles, in order to increase the probability of detecting different alleles. Coriell samples NA18498 (African) and NA07000 (European) were mixed in ratios 1:1, 1:3 and 1:9 at 1 ng final input, prepared and sequenced in duplicate.

2.3.3. Challenging samples

Seven GEDNAP (German DNA Profiling Group, <https://www.gednap.org>) proficiency testing samples, prepared from different biological tissues (two blood samples, two saliva samples, two semen samples and one sample of unidentified origin) were chosen to test the assay's performance with mock casework samples. A batch of seven traces was sent to each of five participating laboratories, who applied their in-house extraction and quantification methods before analyzing them with the BT A&A (Amp) assay. Furthermore, in order to evaluate the assay's ability to process degraded/fragmented samples, artificially

degraded samples were prepared and sequenced in two different VISAGE laboratories. A sonication time series (0–360 min) of 007 control DNA was prepared using an ultra-sound cleaner at 40 kHz. All samples were checked for increasing degradation by routine STR typing using the PowerPlex ESI 17 typing kit (Promega). STRs were analyzed on an ABI 3500 XL capillary electrophoresis instrument (Supplementary Fig. S2). All challenging samples were prepared at 1 ng input when possible, or at the highest DNA input possible.

2.3.4. Inhibitor tolerance

MPS tolerance to inhibitors has previously been shown to be lower than common STR/capillary electrophoresis methods [42], with the initial PCR being described as the critical point for failed genotype sequencing. In order to further explore and describe the tolerance thresholds of the Precision ID DL8 library preparation, three known PCR inhibitors at varying concentrations (hematin, humic acid and indigo carmine) were spiked into a 2800 M control DNA at 1 ng. As the library preparation protocol used was fully automated, the final concentration of the inhibitors could not be exactly determined, therefore, inhibitor levels are described considering the total amount used. Hematin was tested at 8×10^{-4} – 2.5×10^{-5} μmol , humic acid at 1600–50 ng and indigo carmine dye at 0.16–0.005 μmol . All inhibitor-spiked samples were prepared in two VISAGE laboratories, so two replicates of each inhibitor concentration were included for analysis.

2.3.5. Species specificity

Fourteen animal DNA samples from different species (Supplementary Table S2) were prepared at 1 ng input and sequenced with the BT A&A (Amp) assay. Samples were selected considering either their relevance in forensic casework analysis, such as domestic animals, or their genetic similarity to human DNA testing primates.

3. Results and discussion

3.1. Assay design and development

Of the 158 EVC and BGA markers initially considered in the design, primers for 153 markers were successfully included into one single, functional multiplex assay using the AmpliSeq design algorithm (TFS). With the purpose of designing an assay for forensic casework, which can involve highly degraded samples, an assay containing the shortest amplicons possible (around 175 bp) was designed. Five tri-allelic BGA candidate markers as well as a single Middle East informative SNP were rejected by the assay design process; representing an MPS assay conversion rate for BGA-SNPs of 95 % (i.e. 116/121 candidates successfully incorporated). All of the appearance informative markers were successfully designed and incorporated into the final assay, thus reaching a 100 % MPS conversion rate for EVC-SNPs.

The BT A&A (Amp) assay was designed and firstly tested with optimum DNA input samples and a sensitivity dilution series to assess its general performance by the Institute of Legal Medicine, Medical University Innsbruck (MUI), before being distributed to five additional VISAGE partner laboratories for further validation testing. No additional optimization of the PCR and library preparation steps was needed by using the fully automated process.

All data processed by the VISAGE partner laboratories were sent to MUI for uniform analysis and validation of the Basic Tool. Overall, 288 DNA samples were sequenced and analyzed during the validation process, successfully completing all the tasks. Here, a compiled analysis of the entire dataset is described.

3.2. Assay characterization and sequence quality

3.2.1. Coverage

In terms of total coverage, all replicates of optimum DNA input reached values above 500,000 reads, with mean coverage values of

1,077,374 \pm 354,052.51, 893,355 \pm 174,788.26 and 1,137,734 \pm 366,024.97 for 1, 2 and 10 ng, respectively. Considering the total capacity of a 530 chip is 15–20 million reads, when dividing by 16 samples the optimum total number of reads per sample should vary between 937,500 to 1,250,000 reads. Both 1 ng and 10 ng mean total coverage values fall within the expected interval, whereas the 2 ng mean value underperforms. A statistical *t*-test was performed considering the total coverage obtained from 1, 2 and 10 ng samples, respectively, resulting in a significant difference only when comparing the 2 ng replicates. However, such results might stem from an artificial bias of the running scheme adopted for the reproducibility samples. In fact, the mean total coverage of the 2 ng replicates is the lowest of the three optimum inputs, but because these replicates were always run simultaneously with the 10 ng replicates, the latter ones might have outcompeted the 2 ng samples. A more detailed analysis of normalized read depth – the read depth per marker divided by the total number of reads (total coverage) – allows exploration of the distribution of reads across the whole assay and pinpoints markers that are under- or over-performing. The comparison between mean normalized read depth distribution of 1 ng replicates against 2 ng and 10 ng, allows an understanding of how the read distribution among markers differed with DNA input. In addition to a formal Kolmogorov-Smirnov statistical test, a consistent trend in marker behavior becomes evident with varying DNA input amounts, as shown in Fig. 1, which clearly indicates that the BT A&A (Amp) assay is not affected by variation in the optimum DNA input. When considering all reads divided equally among the 153 SNPs, normalized read depth per marker should be \sim 0.0065 (represented as dashed lines in Fig. 1A). The analysis of the mean normalized read depth per marker for the 1, 2 and 10 ng replicates presented very similar results, showing 65.4 %, 63.4 % and 61.4 % of the markers above 0.006, respectively. Furthermore, only 25–26 % of component SNPs showed a mean normalized read depth below the first quartile (0.0046 – 0.0049). High mean read depth per marker was obtained for 1 ng replicates (7,041.66 \pm 3,753.98 reads per SNP), all SNPs presenting more than 1000x average coverage, which is above any threshold normally applied to forensic MPS. In fact, the observed mean number of reads per SNP falls within the expected interval of 6,127.45–8,169.93, obtained when dividing the expected optimum total number of reads per sample by the number of markers (considering 16 samples per 530 Ion Chip). In addition, all genotypes compared between the five participating VISAGE laboratories were fully concordant for the 1, 2 and 10 ng 2800 M replicates.

3.2.2. Strand bias

Strand bias was calculated as the ratio between reads covering the forward strand and total reads (forward + reverse) at a specific target site. Previous studies have indicated that bias in strand sequencing might lead to less confident genotype calls, with a tendency for Ion Torrent chemistry to present reverse strand sequencing preference [20]. Analysis of the 1 ng replicate mean results indicated only 20.55 % of the target sites showed a strand bias ratio outside of the optimum interval of $0.45 < sb < 0.55$ (Supplementary Fig. S3A). Furthermore, a more balanced distribution in preference for forward and reverse reads was observed with 9.15 % of markers above 0.55 and 5.23 % below 0.45 on average for 1 ng replicates.

3.2.3. Base misincorporation rates

The base misincorporation rate is commonly defined as the incorporation of an erroneous base at a SNP target site. This variation across different targets is important to describe the precision of the assay as well as to identify specific positions where a sequencing error is observed at a higher than expected frequency. Due to its putative impact on genotype calls, there is the necessity to distinguish between the incorporation of a base that creates an alternative allele in a homozygous genotype (or tri-allelic SNP) and the incorporation of a non-specific base for the site. Supplementary Fig. S3B depicts the variation of misincorporation rates divided into the two variants across the entire assay. Furthermore, minimum allele read frequency thresholds can be adapted to values that benefit the finding of minor contributors without compromising the occurrence of dropins. The observed mean total misincorporation rate was 0.23 %, 31 SNPs showed a total misincorporation rate above the mean and only three SNPs presented a rate above 1% (rs2789823, rs1040934 and rs7084970; all BGA markers, Supplementary Table S3). For these SNPs a manual inspection of the reads should be performed in order to identify misaligned reads and prevent erroneous detection of minor alleles (Supplementary Table S3 and Supplementary File S1 for further details). Particularly, rs2789823 and rs7084970 are located in repetitive regions (GAGGG/ACCC and TTA AAC/TGAAT, respectively), therefore the erroneous incorporation of cytosines in the reverse strands at rs2789823 and adenines in the forward strands at rs7084970 can occur frequently and affect genotype call (Supplementary File S1A-B). Such results are in concordance with previous studies with AmpliSeq chemistry [20,21,43] A *t*-test was applied to examine whether there was a difference in the frequencies of detecting an erroneous non-specific base or an erroneous base creating

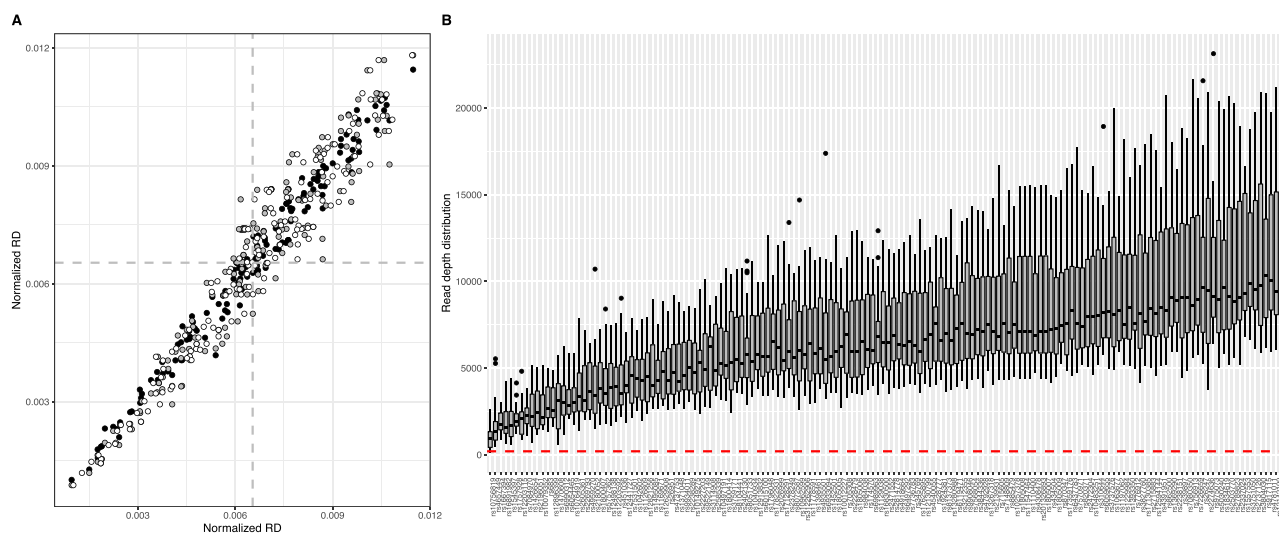


Fig. 1. A) Mean normalized sequence read depth distribution of 1 ng replicates against 2 ng replicates (black dots), 1 ng versus 10 ng (gray dots) and 2 ng versus 10 ng (white dots). Gray dashed lines represent ideal normalized coverage per marker (1/153). B) Sequence read depth distribution per SNP considering all 1 ng replicates of 2800 M control DNA ($n = 25$). The red dashed line represents 200 reads.

the alternative allele; however, no significant difference in rates was observed.

3.2.4. Sequencing baseline

The sequencing baseline was estimated by analyzing the number of target reads at specific positions across the assay in a negative control sample. Ten negative controls (H₂O) were run across the Ion S5 runs of the participating laboratories. In Supplementary Fig. S3C a more detailed graphical representation of the mean distribution of reads across all target sites of the BT assay is shown. As with estimation of the base misincorporation rate, a distinction is made between reads with specific (allelic) and non-specific nucleotide calls. Most observed reads showed a specific allele and therefore, can be assumed as a result from cross contamination between prepared samples and not the presence of extraneous DNA. The mean number of reads obtained per target considering all negative controls was 34; however, one negative control had a higher number of target reads, which possibly indicates contamination during the automated library preparation process (Supplementary Fig. S4). A thorough examination of the generated profile and cross check with the other prepared samples led to the identification of the sample that originated the NTC contamination to be one of the GEDNAP samples prepared in the same library DL8 batch. When excluding this negative control from analysis the mean number of reads per target decreased to 29.

3.2.5. Allele read frequency balance

The SNP Genotyper output measures the MAF (major allele frequency) parameter to help identify imbalanced markers. MAF-flagged markers indicate that the major allele frequency falls outside of the expected intervals for this parameter – set for this exercise to be 95–100 % for homozygotes and 35–65 % for heterozygotes (SNP Genotyper default). To explore the overall assay performance in terms of allele balance, particularly in heterozygotes, all 1 ng samples from 2800 M and Coriell samples were considered for this analysis. The Coriell samples provided additional genotype diversity due to their different bio-geographical backgrounds. Supplementary Fig. S5A shows an overview of this analysis, outlining MAF values for all control samples. SNPs with outlying MAF values were categorized according to the number of times values were outside the expected intervals (Supplementary Fig. S5B). The majority of SNPs had only one or two incidences (in 43 samples). However, SNPs rs1398461 and rs2605361 had 3–5 outlying MAF values and SNPs rs2789823 and rs7084970 had more than five (highlighted in Supplementary Fig. S5A). Furthermore, aiming to distinguish between the influence of the SNP on allele balance vs. the DNA sample in any one sequencing analysis; all 43 samples used for this analysis were categorized based on the total number of MAF flagged SNPs (Supplementary Fig. S5C). Most samples showed one or two SNPs with a MAF flag, four samples had 3–5 MAF flagged SNPs and four had more than 5 SNPs. The latter four samples comprised of two 2800 M replicates and two NA07029 replicates.

Finally, only the SNPs rs2789823 and rs7084970 (both BGA markers) gave MAF values that consistently fell outside the expected intervals (Supplementary Fig. S5A), indicating a balanced and consistent sequencing performance for the assay as a whole. Nevertheless, special care should be taken in routine genotype calling for rs2789823 and rs7084970, and more detail on underperforming SNPs is given in section 3.2.6.

3.2.6. Underperforming SNPs and IGV inspection

All samples were manually inspected using IGV to screen for markers with recurrent alignment issues, including polynucleotide tracts and the possible presence of insertion/deletion polymorphisms (indels) that could impair successful sequencing of the full amplicon. Seventeen SNPs were identified to have possible alignment issues that could impair sequencing of the full amplicon, including potential influence on strand bias or SNPs showing reads with low mapping quality, which

could lead to higher misincorporation rates. Mapping quality represents the probability of read misplacement and is normally reported using the Phred scale [44]. For example, reads hitting a highly repetitive region commonly have low mapping quality, as these have higher probability of false alignments. All reproducibility replicates showed full genotype concordance and no dropouts, but particular attention is needed when considering low level/degraded DNA samples. The SNPs rs2789823 and rs7084970 previously discussed in regard to allele read frequency imbalance also had high total misincorporation rates (4.39 and 6.94 %, respectively). Despite the latter being predominantly non-specific base misincorporation (3.82 and 6.85 %, respectively), these high rates could have led to increased occurrence of MAF flags. As previously described, alignment visualization with IGV of rs7084970 showed a poly-A tract immediately upstream of the SNP site, which is likely to have caused the erroneous call of an “A” instead of a “T” base. Similarly, SNP rs2789823 had a homopolymeric tract on the reverse strand, likely to have impacted the base calling. SNPs rs2196051 and rs2605361, showed strand bias towards the reverse strand (0.19 and 0.30, respectively). SNP rs2196051 showed several indels upstream of the SNP site, which impaired the sequence of the forward strand, and rs2605361 presented a poly-A tract upstream of the SNP site, causing the same effect. SNPs rs2180052 and rs12498138 had strand bias towards the forward strand. Examples of alignment visualizations are outlined in Supplementary File S1. The full list of markers, alignment comments and manual inspection advice is provided in Supplementary Table S3.

3.3. Sensitivity

The performance of the BT A&A (Amp) assay at low DNA input levels was assessed with a dilution series of a known control DNA (Fig. 2). As expected, a reduction in read depth was observed with the decreasing DNA input, except for a better performance of 0.1 ng replicates (Fig. 2A). However, this can be explained as an artificial result since 0.1 ng replicates were joined with lower input samples in the run configuration, therefore, probably being overrepresented in the final library pool produced by the Ion Chef. Automatic library preparation with the Ion Chef produces a final library pool from an initial 8 sample batch, which are combined after a bead normalization and regardless of individual library quantity. Percentages of correct genotypes (CGT) per mean dilution step are represented in black in Fig. 2B. Genotype inconsistency, such as allele dropouts and dropins, incorrect genotype calls (IGT), no detection (NN) and no reads are depicted in Fig. 2B. On average, 0.1 ng input level replicates led to full SNP profiles and at 0.05 ng still 90.1 % of the genotypes were correct, while at 0.025 ng and 0.01 ng, only 79.3 % and 67.8 % of the full profile was successfully genotyped, respectively. Such results agree with a previous study which have also found the lower DNA input limit for full profiles around 0.1 ng [21], however it must be considered that sensitivity limits will vary according to chip throughput, number of samples and assay size. At 0.05 ng and below, genotype discordances (dropouts, dropins and incorrect calls) and no calls were observed. Among possible genotype inconsistencies, allele dropouts occurred at the highest frequency in all mean replicate dilutions. At 0.01 ng the mean amount of no calls (both no reads and lack of sufficient reads for software detection) was higher than the number of allele dropouts, however no SNP was identified with a consistent dropout across the dilution series.

In SNP assays for individual identification and for BGA, a single dropout and/or no call does not generally lead to strong consequences for likelihood ratio calculations and ancestry estimation. In contrast, the inference of EVCs can be much more sensitive to SNP loss and can lead to erroneous predictions from biased likelihood values, depending on the particular EVC-SNP missing [45]. Therefore, establishing a lower DNA input limit for the BT A&A (Amp) tool is important and further testing with data of consistently underperforming SNPs should be carried out.

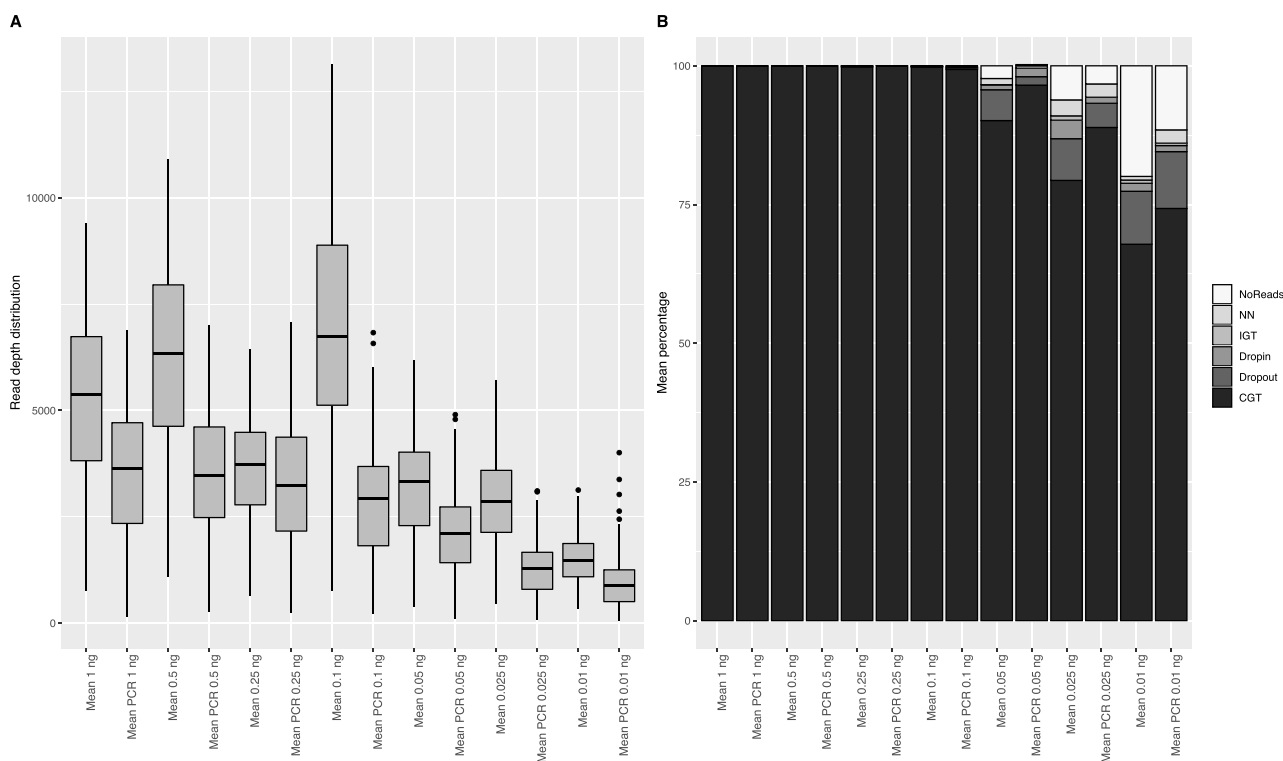


Fig. 2. A) Mean replicate sequence read depth distribution per dilution step of the sensitivity study and per method (normal protocol and increased PCR cycles protocol – PCR). B) Mean percentage of correct genotype (CGT) recovery and discordances (Dropout – allele dropout, Dropin – allele dropin, IGT – incorrect genotype, NN – SNP Genotyper call failed and No Reads – locus dropout) per dilution step of the sensitivity study and by method (normal protocol and increased PCR cycles protocol – PCR).

An increased PCR cycle protocol (+ 5 cycles) resulted in higher percentages of recovered genotypes for the low-input samples, 0.05 ng (90.1 % improving to 96.5 %), 0.025 ng (79.3 %–88.9 %) and 0.01 ng (67.8 %–74.3 %); corroborating the manufacturer’s suggestion for protocol adjustment when typing low level or degraded DNA. Even though there was an increase in successful genotype detection, there was no increase in read depth distribution for the adjusted protocol replicates (Fig. 2), which could also be the result of the artificial overrepresentation of the 0.1 ng replicates in the sensitivity dilution series. In fact, the increased PCR protocol samples were all prepared within one batch (in duplicates, one batch in each laboratory) of library DL8 chemistry and clearly showed a gradual decrease in read depth distribution. On the contrary, the sensitivity series, which were prepared in two batches, resulted in a more inconsistent read depth distribution between different inputs.

3.4. Concordance

Genotype concordance was assessed in two stages: i) between laboratories running the same set of samples (inter-laboratory concordance) and ii) between the genotypes obtained with the BT A&A (Amp) assay for six Coriell samples and the publicly available genotypes in 1000 Genomes and SGDP. The inter-laboratory concordance study compared genotypes from six samples run independently in three VISAGE Consortium laboratories. Results showed full concordance for all samples tested. There was a single discordant genotype between BT A&A (Amp) sequencing results and genotype data stored in online databases in 609 genotypes: a concordance rate of 99.83 %. SNP rs2789823 had a GG genotype for sample NA18498 using the BT A&A (Amp) assay, which was an AG genotype in 1000 Genomes. An IGV image of one replicate of NA18498 for the rs2789823 SNP position is shown in Supplementary File S1, with no apparent alignment problems observed at this site. Further investigation into the discordant position

was performed by comparing the BT A&A (Amp) genotype with the data from the 1000 Genomes New York Genome Center high coverage dataset, which was found fully concordant, reaching 100 % concordance.

3.5. Mixtures

Mixture detection using binary markers is challenging, but has been successfully achieved in previous forensic MPS studies [20,21,25,41,43]. The basis for mixture detection with such markers is the ratio of allele read frequencies observed when more than one contributor is present. Allele read frequencies in single source samples are expected to fall within the 90–100 % and 40–60 % ranges (with up to 10 % baseline noise, although rarely this high, for the uncalled allele). The allele read frequencies observed for single source and mixed profiles are depicted in Fig. 3A and Supplementary Fig. S6. In routine use, the SNP Genotyper output helps to identify a mixed profile by flagging the major allele frequency parameter when outside the ranges established for the assay (95–100 % and 35–65 % default ranges were set for BT A&A). However, due to the SNP Genotyper limitation to call binary SNPs, the extra information contained in the tri-allelic SNPs is not automatically analyzed. Furthermore, the standard genotype calling thresholds of 10 % minor allele read frequency also limits the calling of mixture contributors present at very low levels.

An increased heterozygosity is also expected in mixed profile samples, a representation of the percentage of markers showing homozygous and heterozygous genotypes for the single source samples and the mixed replicates at different ratios is shown in Fig. 3B. The expected mixed profile was also represented as an *in silico* combination of both single source samples’ genotypes. The percentage of obtained heterozygous SNPs found in replicates for 1:1 and 1:3 ratios match the expected *in silico* values, however 1:9 replicates showed smaller heterozygosity percentages. This could be an artifact of the 10 % minor allele

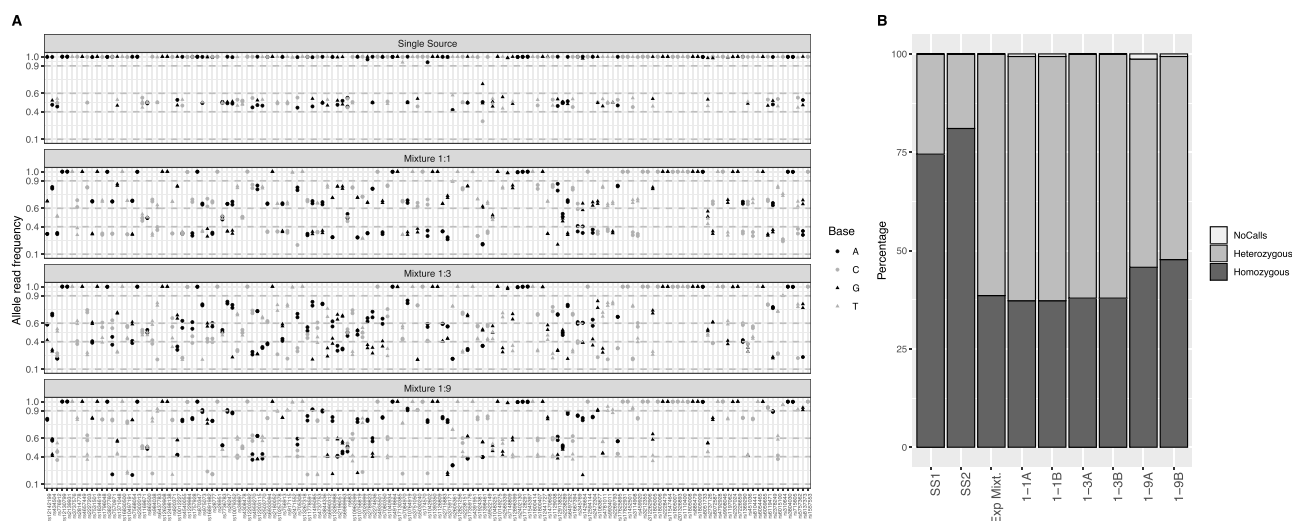


Fig. 3. A) Mean allele read frequency (minimum allele read frequency ≥ 0.2) per marker considering both single source samples and replicates of mixed ratios. Dashed gray lines depict 10, 40, 60 and 90 % of allele read frequency. B) Percentage of observed homozygote and heterozygote genotypes per single source and mixed sample. ‘Exp. Mixt.’ represents the expected percentage of homozygous and heterozygous genotypes when mixing *in silico* both single source genotypes.

read frequency threshold used by SNP Genotyper as the standard threshold. Previous studies show decreasing this threshold helps to identify minor contributors present at lower levels in mixed DNA [21,43].

3.6. Challenging samples

Different types of challenging samples were used to assess the performance of the assay both with samples obtained from different biological tissues and mimicking typical forensic casework DNA; as well as samples with various degrees of degradation/fragmentation.

3.6.1. Casework type samples– GEDNAP

A batch of seven GEDNAP stains (Supplementary Table S4) were sent to five VISAGE Consortium laboratories for extraction and quantification according to their in-house validated protocols and prepared following the BT A&A (Amp) analysis workflow. All genotypes obtained per sample were compared (5,355 comparisons in total) and the observed concordance rate was 99.8 %. In fact, only 11 discordant genotypes were observed, and nine of these occurred with a single sample (one replicate of 49S2). In general, this sample has produced low quantification results from all laboratories (average of 2.45 ng/ μ L) and particularly this replicate has shown the lowest quantification results (0.07 ng/ μ L, Supplementary Table S4). Eight of the affected SNPs showed low normalized read depths (< 0.0038 , ideal nRD = 0.0065) for the mean 1 ng reproducibility replicates. Two of the markers were also listed as problematic SNPs (rs2196051 and rs12498138). The remaining two discordances occurred with replicates of one sample (44S3) run in two different laboratories, one discordance derived from a locus dropout and the second one called a TT genotype whereas the remaining replicates showed a CT genotype, both in the same SNP (rs5757362). Such results underline the robustness and versatility of the BT A&A (Amp) assay in handling samples from different biological tissues and analyzed with a range of differing extraction protocols.

3.6.2. Artificially degraded DNA samples

The read depth distribution of artificially degraded DNA samples showed a gradual decrease with increasing sonication time, but this did not affect genotyping success. Only the longest sonication replicates (360 min) showed a visible decrease in correct genotype calls and occurrence of allele dropouts and dropins plus several no calls (Fig. 4A and Supplementary Fig. S7). In fact, one of the 360 min replicates showed such low read depth values that it was not possible to apply

SNP Genotyper, and this sample was excluded from further analysis. The progressive degradation/fragmentation status of the sonicated samples was verified by both real-time PCR quantification [46] and STR typing (Supplementary Fig. S2), although the degradation did not affect the VISAGE BT A&A (Amp) genotyping performance to the same degree. One reason for this reduced effect is likely to be the assay design considering shorter amplicon sizes for the SNP targets to achieve amplified fragments of reduced size, however the limited number of samples analyzed per chip (16 samples) could be another explanation for the genotyping success.

3.7. Inhibitor tolerance

The 2800 M control DNA replicates separately spiked with PCR inhibitors showed a gradual decrease of read depth distribution with increasing inhibitor concentration (Fig. 4B). However, the percentage of correct genotype calls dropped dramatically above 4×10^{-4} μ mol (total input) of hematin, 200 ng of humic acid and 0.02 μ mol of indigo dye, clearly indicating the maximum PCR tolerance of these inhibitors at such concentrations (Fig. 4C). It should be noted that all inhibitors are described at total amount and not their concentration per reaction volume. In fact, the DL8 automated library preparation solution uses 15 μ L of input sample, but as all reagents (primers and enzyme) were added automatically in the Ion Chef, calculation of the inhibitor concentration per reaction volume was not possible. All results were concordant between replicates, further indicating the upper inhibitor tolerance limits of the assay. MPS inhibitor tolerance has not been thoroughly studied by the forensic community, however, early results showed another MPS library preparation method to be more sensitive to inhibitors than routine STR profiling with capillary electrophoresis methods [42]. Our results confirm this finding; e.g., the BT A&A (Amp) assay is at least 18.7x and 15x more sensitive to the inhibiting effects of hematin and humic acid respectively, as the commercial Promega PowerPlex Fusion STR System (Promega, [47]). Such results increase the necessity of further studying the effects of PCR inhibitors on MPS assays and, more importantly, of improving current chemistry before fully implementing them into forensic casework.

3.8. Species specificity

The results of primer specificity tests showed no evident amplification of non-human DNA. Read depth distributions were very similar to the negative control (Supplementary Fig. S8A), except for the

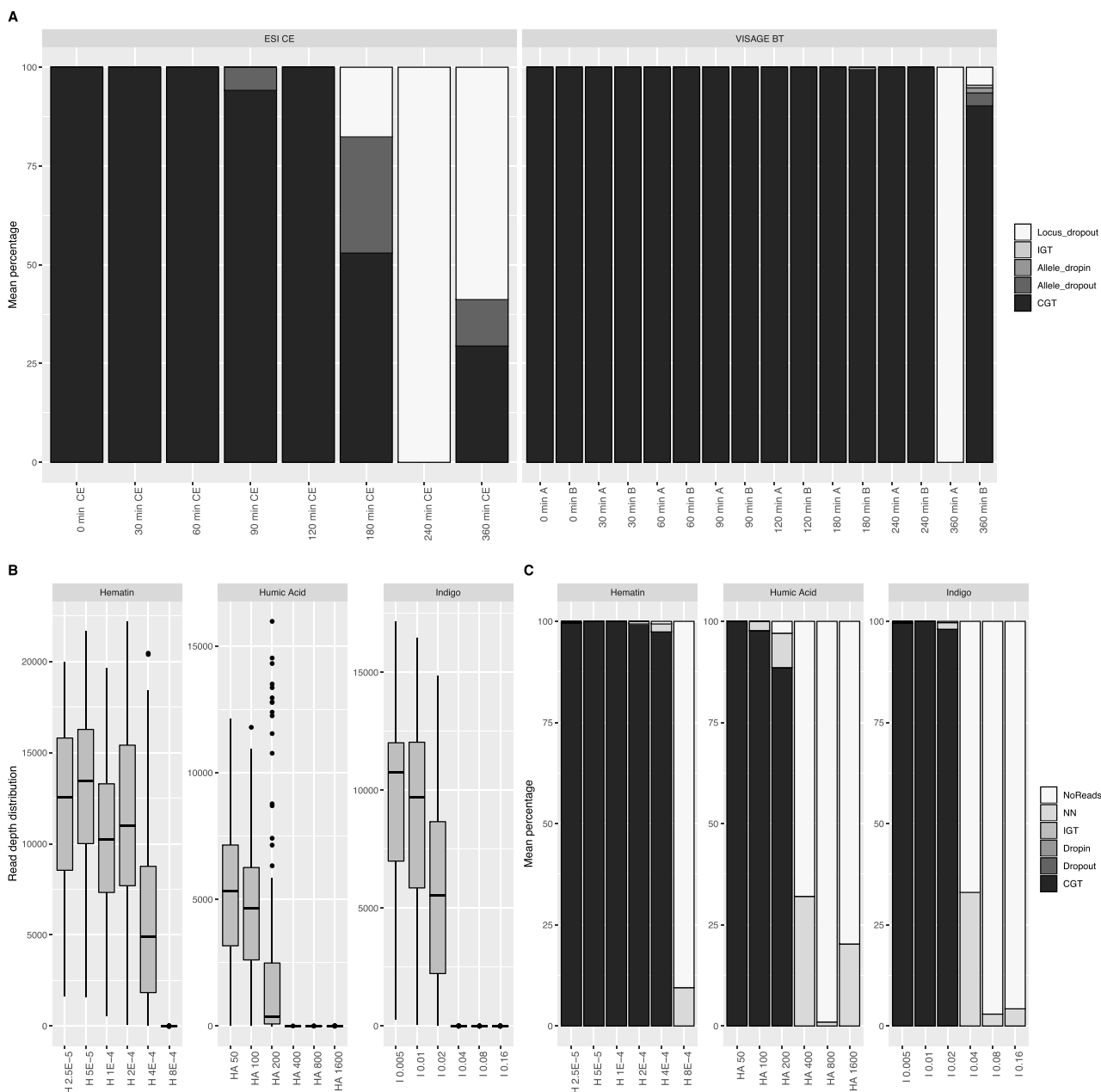


Fig. 4. A) Percentage of correct genotype (CGT) recovery and discordances (Allele dropout, Allele dropin, IGT – incorrect genotype and Locus dropout) per sample used in the sonication time series and subjected to STR amplification (ESI) and capillary electrophoresis, and the BT A&A inference in duplicates. B) Duplicate mean sequence read depth distribution for each series of inhibitor treatment (spiked 2800 M control DNA) – hematin, humic acid and indigo dye. C) Mean percentage of correct genotype (CGT) recovery and discordances (Dropout – allele dropout, Dropin – allele dropin, IGT – incorrect genotype, NN – SNP Genotyper call failed and No Reads – locus dropout) for each series of inhibitor treatment (spiked 2800 M control DNA) – hematin, humic acid and indigo dye.

Bonobo sample, which had a wider range of read depth values. Considering genotype calls, almost all targets failed for most samples, with only a few successful genotype calls. Primate DNA samples, particularly the Bonobo but also others, showed higher percentages of called genotypes (Supplementary Fig. S8B). Amplification with primate DNA is not unusual and has been widely described in the literature [47–49]; although the chances of finding primate DNA in forensic casework are low. Even though the species specificity results were compromised to some extent due to the underperforming positive control sample in this batch (achieved only 77 % of the genotype calls, similar to the Bonobo with 73 %), the higher percentage of genotype calls and higher read depth from 2800 M control samples achieved in the reproducibility set (both 16 sample sets per chip) indicate this to be an anomalous result.

4. Conclusion

The BT A&A tool – a prototype MPS tool for simultaneous analysis of eye, hair and skin color as well as continental biogeographic ancestry - in particular its AmpliSeq version presented here - showed very good overall performance in terms of sequence quality and coverage across the whole range of component SNPs in the panel. The reproducibility study demonstrated the BT A&A (Amp) assay to be sensitive to DNA input for total coverage but not for the read depth distribution across the SNPs of the assay, clearly indicating its robustness to different DNA input levels. In agreement with previously reported AmpliSeq-based methods applied on the Ion S5 sequencing instrument ([43], unpublished data), balanced strand representation in sequence number was also seen for the BT A&A (Amp) assay. Low baseline noise and base

misincorporation rates were observed, which however did not influence genotype calling and interpretation. Allele read frequency values indicate a well-designed assay, with the exception of two SNPs, which consistently presented major allele read frequency values outside the optimum intervals (35–65 %). The panel's ability to process low level or degraded DNA typically found in forensic casework was shown by its high sensitivity with full SNP profiles obtained down to 100 pg of input DNA or 240 min of sonication. In addition, the increased PCR cycles protocol, suggested by the manufacturer for challenging samples, gave a slight increase in the recovery of correct genotypes. The assay achieved 99.8 % genotyping concordance from casework-type samples, demonstrating its success in analyzing DNA from different extraction methods and tissue sources. However, the BT A&A (Amp) assay showed poor inhibitor tolerance, particularly when compared with routine CE-based STR typing kits. Such results agree with other forensic MPS studies [42] and underline the necessity for MPS suppliers to dedicate resources for preparing forensically relevant MPS assay reagents that contain components for the control of inhibitors; as are found in CE-based STR reagent kits.

Mixture detection proved to be relatively straightforward when observing imbalanced allele read frequencies and compiling the MAF flags produced by SNP Genotyper – although the incapacity of the software to call third alleles limits the efficiency of tri-allelic SNPs for mixture genotyping and should be addressed in future versions of this plug-in.

In conclusion, the presented AmpliSeq version of the VISAGE Basic Tool for Appearance and Ancestry prediction provides a well-balanced, specific and robust SNP sequencing assay suitable to be implemented and applied in forensic DNA casework when information on continental biogeographic ancestry as well as eye, hair, and skin color of an unknown crime scene sample donor are suitable to aid police investigations, if legislation in the country of application allows [50]. Statistical prediction tools to convert the genotype outcomes of the presented MPS tool into probabilities of eye, hair, skin color categories and continental BGA in a combined way are currently being developed by the VISAGE Consortium. For the time being, existing separate tools are available for use via the HIRISplex (<https://hirisplex.erasmusmc.nl/>) and the Snipper websites (<http://mathgene.usc.es/snipper/>), respectively. Future work on the implementation of this panel into forensic routine laboratories will bring a better insight into the performance of the BT A&A (Amp) assay with real casework samples in the routine forensic DNA analysis environment. Further developments on BT A&A versions that are not based on AmpliSeq as well as expanding the VISAGE A&A tool towards additional appearance traits and more detailed BGA are currently underway.

Acknowledgements

The authors would like to thank Antonia Heidegger, Harald Niederstätter, Christina Strobl and the Forensic Genetics and Casework section of the Institute of Legal Medicine, Innsbruck, Kerstin Schöbel of the Institute of Legal Medicine, Cologne, Aleksandra Pisarek from Malopolska Centre of Biotechnology, Jagiellonian University in Krakow and Anna Woźniak of the Central Forensic Laboratory of the Police in Warsaw for valuable discussion and technical support. The study received support from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement No. 740580 within the framework of the VISible Attributes through GENomics (VISAGE) Project and Consortium. MdIP is supported by a postdoctoral fellowship awarded by the Consellería de Cultura, Educación e Ordenación Universitaria and the Consellería de Economía, Emprego e Industria from Xunta de Galicia (Modalidade A, ED481B 2017/088). AFA is supported by a post-doctorate grant funded by the Consellería de Cultura, Educación e Ordenación Universitaria e da Consellería de Economía, Emprego e Industria from Xunta de Galicia, Spain (Modalidade B, ED481B 2018/010). The 1000 Genomes high coverage

sequence data were generated at the New York Genome Center with funds provided by NHGRI Grant 3UM1HG008901-03S1.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:<https://doi.org/10.1016/j.fsigen.2020.102336>.

References

- [1] M. Kayser, Forensic DNA Phenotyping: predicting human appearance from crime scene material for investigative purposes, *Forensic Sci. Int. Genet.* 18 (2015) 33–48.
- [2] J. Pieters, Twenty Years in Prison for 1992 Zaanadam Murder, Available from: NLTimes, 2018, <https://nltimes.nl/2018/12/11/twenty-years-prison-1992-zaadam-murder>.
- [3] L. Jong, A. M'charek, The high-profile case as 'fire object': following the Marianne Vaatstra murder case through the media, *Crime Media Cult.* 14 (3) (2017) 347–380.
- [4] B.-J. Koops, M.H.M. Schellekens, Forensic DNA phenotyping: regulatory issues, *Columbia Sci. Technol. Law Rev.* 9 (1) (2006).
- [5] E. Musgrave-Brown, et al., Forensic validation of the SNPforID 52-plex assay, *Forensic Sci. Int. Genet.* 1 (2) (2007) 186–190.
- [6] C. Phillips, et al., Inferring ancestral origin using a single multiplex assay of ancestry-informative marker SNPs, *Forensic Sci. Int. Genet.* 1 (3–4) (2007) 273–280.
- [7] F. Liu, et al., Eye color and the prediction of complex phenotypes from genotypes, *Curr. Biol.* 19 (5) (2009) R192–R193.
- [8] P. Kersbergen, et al., Developing a set of ancestry-sensitive DNA markers reflecting continental origins of humans, *BMC Genet.* 10 (2009) 69.
- [9] R.K. Valenzuela, et al., Predicting phenotype from genotype: normal pigmentation*, *J. Forensic Sci.* 55 (2) (2010) 315–322.
- [10] S. Walsh, et al., Developmental validation of the IrisPlex system: determination of blue and brown iris colour for forensic intelligence, *Forensic Sci. Int. Genet.* 5 (5) (2011) 464–471.
- [11] S. Walsh, et al., IrisPlex: a sensitive DNA tool for accurate prediction of blue and brown eye colour in the absence of ancestry information, *Forensic Sci. Int. Genet.* 5 (3) (2011) 170–180.
- [12] W. Branicki, et al., Model-based prediction of human hair color using DNA variants, *Hum. Genet.* 129 (4) (2011) 443–454.
- [13] R. Pereira, et al., Straightforward inference of ancestry and admixture proportions through ancestry-informative insertion deletion multiplexing, *PLoS One* 7 (1) (2012) e29684.
- [14] S. Walsh, et al., The HIRISplex system for simultaneous prediction of hair and eye colour from DNA, *Forensic Sci. Int. Genet.* 7 (1) (2013) 98–115.
- [15] M. Fondevila, et al., Revision of the SNPforID 34-plex forensic ancestry test: assay enhancements, standard reference sample genotypes and extended population studies, *Forensic Sci. Int. Genet.* 7 (1) (2013) 63–74.
- [16] Y. Ruiz, et al., Further development of forensic eye color predictive tests, *Forensic Sci. Int. Genet.* 7 (1) (2013) 28–40.
- [17] S. Walsh, et al., Developmental validation of the HIRISplex system: DNA-based eye and hair colour prediction for forensic and anthropological usage, *Forensic Sci. Int. Genet.* 9 (2014) 150–161.
- [18] K.B. Gettings, et al., A 50-SNP assay for biogeographic ancestry and phenotype prediction in the U.S. Population, *Forensic Sci. Int. Genet.* 8 (1) (2014) 101–108.
- [19] O. Maroñas, et al., Development of a forensic skin colour predictive test, *Forensic Sci. Int. Genet.* 13 (2014) 34–44.
- [20] M. Eduardoff, et al., Inter-laboratory evaluation of SNP-based forensic identification by massively parallel sequencing using the Ion PGM™, *Forensic Sci. Int. Genet.* 17 (2015) 110–121.
- [21] M. Eduardoff, et al., 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.
- [22] A.C. Jäger, 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.* 28 (2017) 52–70.
- [23] M. Al-Asfi, et al., Assessment of the precision ID ancestry panel, *Int. J. Legal Med.* 132 (6) (2018) 1581–1594.
- [24] C. Phillips, et al., MAPlex - A massively parallel sequencing ancestry analysis multiplex for Asia-Pacific populations, *Forensic Sci. Int. Genet.* 42 (2019) 213–226.
- [25] K. Breslin, et al., 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.
- [26] E. Pospiech, et al., Towards broadening Forensic DNA Phenotyping beyond pigmentation: improving the prediction of head hair shape from DNA, *Forensic Sci. Int. Genet.* 37 (2018) 241–251.
- [27] V. Pereira, et al., Development and validation of the EUROFORGEN NAME (North African and Middle Eastern) ancestry panel, *Forensic Sci. Int. Genet.* 42 (2019) 260–267.
- [28] R. Kosoy, et al., Ancestry informative marker sets for determining continental origin and admixture proportions in common populations in America, *Hum. Mutat.* 30 (1) (2009) 69–78.
- [29] K.K. Kidd, et al., Progress toward an efficient panel of SNPs for ancestry inference, *Forensic Sci. Int. Genet.* 10 (2014) 23–32.
- [30] L. Chaitanya, et al., The HIRISplex-S system for eye, hair and skin colour prediction from DNA: introduction and forensic developmental validation, *Forensic Sci. Int.*

- Genet. 35 (2018) 123–135.
- [31] C. Phillips, et al., Building a forensic ancestry panel from the ground up: the EUROFORGEN Global AIM-SNP set, *Forensic Sci. Int. Genet.* 11 (2014) 13–25.
- [32] C. Phillips, et al., Eurasiaplex: a forensic SNP assay for differentiating European and South Asian ancestries, *Forensic Sci. Int. Genet.* 7 (3) (2013) 359–366.
- [33] M.F. Seldin, et al., European population substructure: clustering of northern and southern populations, *PLoS Genet.* 2 (9) (2006) e143.
- [34] J.M. Galanter, et al., Development of a panel of genome-wide ancestry informative markers to study admixture throughout the Americas, *PLoS Genet.* 8 (3) (2012) e1002554.
- [35] C. Strobl, et al., Evaluation of the precision ID whole MtDNA genome panel for forensic analyses, *Forensic Sci. Int. Genet.* 35 (2018) 21–25.
- [36] Thermo Fisher Scientific, Precision ID SNP Panels With the HID Ion S5™ /HID Ion GeneStudio™ S5 System - APPLICATION GUIDE, (2019) Publication Number: MAN0017767(Revision C.0).
- [37] R Development Core Team, R: a Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, 2010.
- [38] H. Wickham, ggplot2: Elegant Graphics for Data Analysis, Springer-Verlag, New York, 2016.
- [39] The Genomes Project, C, A global reference for human genetic variation, *Nature* 526 (7571) (2015) 68–74.
- [40] S. Mallick, et al., The Simons Genome Diversity Project: 300 genomes from 142 diverse populations, *Nature* 538 (7624) (2016) 201–206.
- [41] Ø. Bleka, et al., Open source software EuroForMix can be used to analyse complex SNP mixtures, *Forensic Sci. Int. Genet.* 31 (2017) 105–110.
- [42] M. Sidstedt, et al., The impact of common PCR inhibitors on forensic MPS analysis, *Forensic Sci. Int. Genet.* 40 (2019) 182–191.
- [43] M. de la Puente, et al., 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.
- [44] H. Li, J. Ruan, R. Durbin, Mapping short DNA sequencing reads and calling variants using mapping quality scores, *Genome Res.* 18 (11) (2008) 1851–1858.
- [45] S. Walsh, et al., Developmental validation of the HrisPlex system: DNA-based eye and hair colour prediction for forensic and anthropological usage, *Forensic Sci. Int. Genet.* 9 (2014) 150–161.
- [46] C. Xavier, et al., SD quant - Sensitive detection tetraplex-system for nuclear and mitochondrial DNA quantification and degradation inference, *Forensic Sci. Int. Genet.* 42 (2019) 39–44.
- [47] K. Oostdik, et al., Developmental validation of the PowerPlex® Fusion System for analysis of casework and reference samples: a 24-locus multiplex for new database standards, *Forensic Sci. Int. Genet.* 12 (2014) 69–76.
- [48] J.M. Thompson, et al., Developmental validation of the PowerPlex(R) Y23 System: a single multiplex Y-STR analysis system for casework and database samples, *Forensic Sci. Int. Genet.* 7 (2) (2013) 240–250.
- [49] M.G. Ensenberger, et al., Developmental validation of the PowerPlex® fusion 6C system, *Forensic Sci. Int. Genet.* 21 (2016) 134–144.
- [50] G. Samuel, B. Prainsack, The Regulatory Landscape of Forensic DNA Phenotyping in Europe, VISAGE, 2018.
- [51] James T. Robinson, et al., Integrative genomics viewer, *Nat. Biotechnol.* 29 (1) (2011) 24–26, <https://doi.org/10.1038/nbt.1754>.