Forensic Science International: Genetics 28 (2017) 52–70

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

Research paper

Developmental validation of the MiSeq FGx Forensic Genomics System for Targeted Next Generation Sequencing in Forensic DNA Casework and Database Laboratories

Anne C. Jäger, Michelle L. Alvarez, Carey P. Davis, Ernesto Guzmán, Yonmee Han, Lisa Way, Paulina Walichiewicz, David Silva, Nguyen Pham, Glorianna Caves, Jocelyne Bruand, Felix Schlesinger, Stephanie J.K. Pond, Joe Varlaro, Kathryn M. Stephens , Cydne L. Holt*

Illumina, Inc., 5200 Illumina Way, San Diego, CA 92122, USA

ARTICLE INFO

Article history:
Received 1 February 2016
Received in revised form 8 December 2016
Accepted 23 January 2017
Available online 27 January 2017

Keywords:
Forensic genomics
DNA typing
Forensic DNA
SWGDAM validation
Next-Generation Sequencing (NGS)
Massively parallel sequencing (MPS)
MiSeq FGx™ System
ForenSeq™ DNA Signature Prep Kit
ForenSeq™ Universal Analysis Software (UAS)
Short tandem repeat (STR)
Single nucleotide polymorphism (SNP)

ABSTRACT

Human DNA profiling using PCR at polymorphic short tandem repeat (STR) loci followed by capillary electrophoresis (CE) size separation and length-based allele typing has been the standard in the forensic community for over 20 years. Over the last decade, Next-Generation Sequencing (NGS) matured rapidly, bringing modern advantages to forensic DNA analysis. The MiSeq FGx™ Forensic Genomics System, comprised of the ForenSeq™ DNA Signature Prep Kit, MiSeq FGx™ Reagent Kit, MiSeq FGx™ instrument and ForenSeq™ Universal Analysis Software, uses PCR to simultaneously amplify up to 231 forensic loci in a single multiplex reaction. Targeted loci include Amelogenin, 27 common, forensic autosomal STRs, 24 Y-STRs, 7 X-STRs and three classes of single nucleotide polymorphisms (SNPs). The ForenSeq™ kit includes two primer sets: Amelogenin, 58 STRs and 94 identity informative SNPs (iiSNPs) are amplified using DNA Primer Set A (DPMA; 153 loci); if a laboratory chooses to generate investigative leads using DNA Primer Set B, amplification is targeted to the 153 loci in DPMA plus 22 phenotypic informative (piSNPs) and 56 biogeographical ancestry SNPs (aiSNPs). High-resolution genotypes, including detection of intra-STR sequence variants, are semi-automatically generated with the ForenSeq™ software. This system was subjected to developmental validation studies according to the 2012 Revised SWGDAM Validation Guidelines.

A two-step PCR first amplifies the target forensic STR and SNP loci (PCR1); unique, sample-specific indexed adapters or "barcodes" are attached in PCR2. Approximately 1736 ForenSeq™ reactions were analyzed. Studies include DNA substrate testing (cotton swabs, FTA cards, filter paper), species studies from a range of nonhuman organisms, DNA input sensitivity studies from 1 ng down to 7.8 pg, two-person human DNA mixture testing with three genotype combinations, stability analysis of partially degraded DNA, and effects of five commonly encountered PCR inhibitors. Calculations from ForenSeq™ STR and SNP repeatability and reproducibility studies (1 ng template) indicate 100.0% accuracy of the MiSeq FGx™ System in allele calling relative to CE for STRs (1260 samples), and >99.1% accuracy relative to bead array typing for SNPs (1260 samples for iiSNPs, 310 samples for aiSNPs and piSNPs), with >99.0% and >97.8% precision, respectively. Call rates of >99.0% were observed for all STRs and SNPs amplified with both ForenSeq™ primer mixes. Limitations of the MiSeq FGx™ System are discussed. Results described here demonstrate that the MiSeq FGx™ System meets forensic DNA quality assurance guidelines with robust, reliable, and reproducible performance on samples of various quantities and qualities.

* Corresponding author.
E-mail addresses: ajager@illumina.com (A.C. Jäger), malvarez1@illumina.com (M.L. Alvarez), cdavis1@illumina.com (C.P. Davis), eguzman@illumina.com (E. Guzmán), yhan@illumina.com (Y. Han), lway@illumina.com (L. Way), pwalichiewicz@illumina.com (P. Walichiewicz), dsilva@illumina.com (D. Silva), mpham@illumina.com (N. Pham), gcaves@illumina.com (G. Caves), jbruand@illumina.com (J. Bruand), fschlesinger@illumina.com (F. Schlesinger), spond@illumina.com (S.J.K. Pond), jvarlaro@illumina.com (J. Varlaro), kmstephens@illumina.com (K.M. Stephens), cholt@illumina.com, cydneholt@live.com (C.L. Holt).

http://dx.doi.org/10.1016/j.fsigen.2017.01.011

1. Introduction

Genotyping of human polymorphic, tetra- and pentanucleotide short tandem repeat (STR) loci, based on fragment sizing, has been the mainstream method in forensic biology since the 1990s, having replaced restriction fragment length polymorphism (RFLP) analysis and multiple PCR-based methods [1–3]. Capillary electrophoresis (CE)-based allele calling is achieved by targeting STR loci with fluorescently-tagged PCR primers, followed by amplicon length detection through size separation, and allele calling relative to a physical allelic ladder [4–7]. One of the limitations of fragment-length STR genotyping using CE is that the maximum number of forensically relevant loci that can be multiplexed and simultaneously detected is currently limited to <30 loci [8–12]. Therefore most Y-STRs and any X-STRs must be analyzed in additional PCRs, instrument runs and data analyses beyond an autosomal workflow. Such iterative testing requires additional input template from sometimes limited DNA extracts [9]. This shortcoming can force forensic analysts to limit their amplification targets (*e.g.*, autosomal STRs and Y-STRs or X-STRs or mtDNA) before DNA quality and DNA profiling information have been generated for a particular sample [13,14]. Furthermore, in cases where DNA quality and/or quantity is low, when a sample contains DNA from more than one individual, or when analysis of more than one class of polymorphism is desired or required (*e.g.*, law enforcement databases), the need to choose one subset of markers over another may be challenging and inefficient [13]. A second limitation of CE-based genotyping is that partial STR profiles or inconclusive results from degraded or PCR inhibited DNA samples, commonly encountered in criminal casework, missing person cases, or mass disaster investigations, are exacerbated due to the STR amplicon lengths required across each size range of each fluorescently labeled dye [15–18]. Finally, length-based STR allele typing cannot provide the genetic discrimination power of sequenced-based typing where alleles of same length are identified not only by the number of STR repeats but by the actual nucleotide-by-nucleotide STR sequence itself [19–23]. These limitations can require accredited forensic labs to maintain multiple quality assurance (QA) programs across multiple kit-based marker systems and the associated analysis software module(s), and to perform more than one workflow per sample or consider sending samples to a third-party service provider to attempt additional typing [24–26]. Efforts to compensate for some of these CE-based STR limitations for human identification include improved DNA extraction methods [27,28], dual quantitation assays [29–31], and mini-STRs [32].

Since transitioning from RFLP analysis, the forensic genomics discipline has remained relatively stationary in terms of applying and incorporating established technologies [20,33]. Numerous scientific breakthroughs and advancements in molecular and cellular biology, genetic, and functional dynamics of the human genome have been made that can improve forensic genomics. For example, in 2001 the first working draft of the human genome was announced [34,35], leading directly to the birth of a new era of biology, genetics, and genomics and an innovation explosion in DNA sequencing, analysis, and interpretation using massively parallel sequencing (MPS), also known as next-generation sequencing (NGS) [36]. NGS 'reads' millions of targeted PCR amplicons, base-by-base. One NGS technology in particular, sequencing-by-synthesis (SBS) [36], has been most widely adopted worldwide in multiple areas of research focus and medical diagnostics including oncology [37,38], microbial genomics [39–41], agrigenomics [42,43], and complex disease genomics [44–46]. SBS chemistry uses a high-resolution camera to image the sequential incorporation of a fluorescently labeled dNTP, followed by cleavage of the terminator to allow incorporation of the next complementary base in an amplicon's DNA sequence. A second sequencing technology is available that uses emulsion PCR, and semiconductor sequencing to detect release of hydrogen ions, as indication of nucleotide incorporation, on a complementary metal-oxide semiconductor CMOS chip, by sequentially exposing an amplicon's DNA to individual nucleotides, one at a time [47]. A clear advantage of targeted SBS, relative to emulsion PCR and semiconductor sequencing, is that since all four reversible terminator-bound dNTPs (dATP, dGTP, dTTP, dCTP) are present during each sequencing cycle, natural competition minimizes incorporation bias, nearly eliminating errors and missed base calls associated with homopolymeric regions and repetitive DNA elements [48–51].

Massively parallel pyrosequencing, semiconductor sequencing, and sequencing by synthesis (SBS) were previously evaluated for forensic purposes on both human mtDNA and gDNA (including forensic STRs). Studies have included reproducibility, heteroplasmy analysis, and mixed samples with more than one DNA contributor, with a strategy intended to facilitate NGS integration into standard casework laboratories, including for criminal, missing persons, and disaster victim identification purposes. The future of this "2nd generation sequencing" was envisioned to expand the number and types of loci being analyzed simultaneously, and predicted that comfort with NGS data would meet that of CE-based typing by sizing and Sanger sequencing [52,53]. Analysis of mtDNA with NGS continued to mature with optimized approaches to sequencing the mtDNA genome, as well as considerations related to targeted NGS data analysis for forensic purposes [54,55]. Evaluation of massively parallel pyrosequencing specifically for forensic STR analyses provided early feasibility for deep sequencing to improve casework and databasing analyses [56,57]. Further proof of concept was demonstrated using (1) SBS for high throughput analysis of CODIS STR loci, using an *in silico* reference genome, and open source aligners and custom scripts to locate variant alleles, and (2) pyrosequencing and a different open source software pipeline [58,59]. Open source software tools available for analysis of forensic NGS data also include MyFLq and STRait Razor [60,61]. Other NGS technologies have been evaluated including a semiconductor sequencing approach on the Ion Torrent PGM, using a subset of the expanded core CODIS STR loci, and other STR loci of interest to the forensic community [62,63]. Separate "panels" have also been reported that interrogate only Y-STRs or SNP subsets [64,65]. The history of DNA sequencing in forensic analyses, basics of NGS and solutions for forensic genetics have recently been reviewed, including the beta version of the MiSeq Forensic Genomics System [66].

The MiSeq Forensic Genomics (FGx™) System (Fig. 1) was developed specifically for human identification and generation of investigative leads. The system consists of four components: the ForenSeq™ DNA Signature Prep Kit, the MiSeq FGx™ Reagent Kit, the MiSeq FGx™ sequencing instrument, and the ForenSeq™ Universal Analysis Software. The ForenSeq™ DNA Signature Prep Kit relies on PCR for target amplification and library construction. This assay contains two primer mixes, one of which may be selected as desired per analysis: (1) DNA Primer Mix A (DPMA) that targets Amelogenin, 27 autosomal STRs, 24 Y-STRs, 7 X-STRs, 94 identity informative single nucleotide polymorphisms (iiSNPs) [67–70], and (2) DNA Primer Mix B (DPMB) that targets each of the loci in DPMA, as well as biogeographic ancestry informative SNPs (aiSNPs) [71] and phenotypic informative SNPs (piSNPs) for hair and eye color estimation [72,73], by incorporating an additional 78 SNPs.

The beta version of the MiSeq FGx™ System was evaluated for forensic purposes and reviewed [66,74,75]. Subsequently, the MiSeq FGx™ System was subjected to SWGDAM developmental validation guidelines [76]. Here we report on the following SWGDAM validation studies: species specificity, sensitivity, mixed

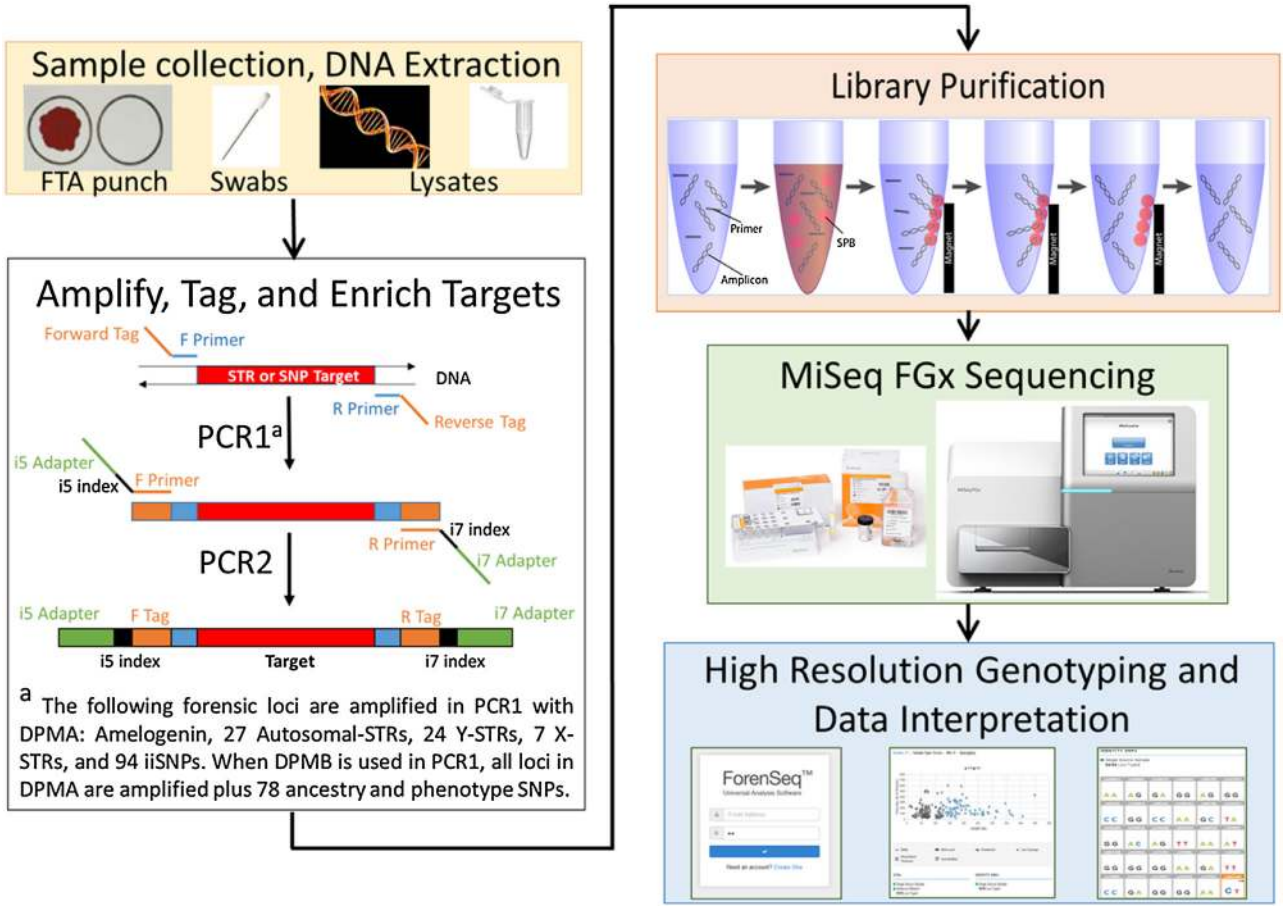

Fig. 1. Workflow diagram of the MiSeq FGx™ Forensic Genomics System. Genomic DNA from a variety of forensic samples, including those collected on FTA cards and swabs, are amenable to analysis using 1 ng PCR template, or less (see Fig. 5). Two rounds of PCR are conducted to first amplify the target forensic STR and SNP loci while attaching universal forward and reverse primer sequence tags (PCR1), and to attach unique, sample-specific indexed adapters or "barcodes" (PCR2). Each amplified, multi-target uniquely tagged library is then purified and normalized using standard molecular biology techniques. These indexed libraries are pooled into a single tube, pipetted into a MiSeq FGx™ reagent cartridge, and inserted into the MiSeq FGx™ sequencer. After targeted sequencing of forensic STR and SNP amplicons [36], individual reads are 'demultiplexed' or separated, using the index sequences. Data are analyzed using the ForenSeq™ Universal Analysis Software for high-resolution genotype and haplotype results, and tertiary data analysis.

samples, stability (inhibitor, degradation), accuracy and precision. In studies presented here, 96 samples were run simultaneously for DPMA samples, and 32 samples were multiplexed when DPMB was employed, to allow for deeper coverage for samples potentially containing a mixture or requiring increased sensitivity. Data indicate that the MiSeq FGx™ Forensic Genomics System meets established forensic guidelines as a reliable method for human DNA profiling with robust and reproducible performance on samples of various quantities and quality.

2. Materials and methods

2.1. Human DNA samples and cell lysates

Human male genomic DNA (gDNA) 2800 M (Promega® Corporation, Madison, WI) was used as the positive amplification control and library preparation control throughout these studies. Four additional human gDNA samples NA12877 (male, Caucasian (CEU)), NA12878 (female, Caucasian (CEU)), NA18507 (male, African (YRI)), and NA19238 (female, African (YRI)) (Coriell Institute for Medical Research, Camden, NJ) were used in studies as described in Results.

Human buccal cell samples were collected from volunteers who each signed an informed consent form authorizing the use of de-identified samples for research use. Volunteer buccal cell samples were collected on three types of substrates: sterile cotton swabs (Puritan Medical Products, Guilford, ME), FTA™ cards (GE Healthcare Bio-Sciences Corp., Piscataway, NJ), and Bode Buccal DNA Collectors™ (Bode Technology, Lorton, VA), and allowed to dry overnight at room temperature. Cell lysates were prepared from buccal swabs by incubation in 500 μl QuickExtract™ DNA Extraction Solution (Epicentre®, Madison, WI) for 1 min at 65 °C, inverted five times, incubated for two minutes at 98 °C, and stored at −20 °C. Two microliters of crude lysate were used in each ForenSeq™ reaction. Cell lysates from 1.2 mm FTA™ card punches were prepared in 96-well plate format. Punches were washed with 100 μl of 1 x TBE and shaken for 2 min at 1800 rpm, in a BioShake XP (Quantifoil Instruments, Jena, Germany). TBE was removed from each well and discarded, leaving the punch in the well in preparation for ForenSeq™ amplification. Cell lysates from 1.2 mm Bode Buccal DNA Collector filter paper punches were prepared in 96-well plate format by adding 2 μl Bode PunchPrep™ solution, heating at 70 °C for 20 min to dry the punch, followed by directly proceeding to the ForenSeq™ amplification.

Human DNA samples were quantified prior to amplification using either the Quantifiler® Human DNA Quantification Kit (Life Technologies, Carlsbad, CA) on the Stratagene Mx3000P qPCR System (Agilent, Santa Clara, CA), or using the Qubit® dsDNA HS or BR Assay Kit on a Qubit® 2.0 Fluorometer (Life Technologies), according to the respective manufacturers' instructions.

2.2. Nonhuman DNA samples

Genomic DNA from nonhuman organisms were tested in species specificity studies: two Old World primates (female baboon, Zyagen Laboratories, San Diego, CA; male rhesus monkey, BioChain®, Eureka, CA), nine non-primate mammals (male cat, male dog, mixed male and female ferret, female horse (Zyagen Laboratories, San Diego, CA)), and male cow, mixed male and female hamster, male mouse, male pig, male rat (BioChain®, Eureka, CA), one avian species (male domesticated chicken; Zyagen Laboratories, San Diego, CA), two fungal samples (*Candida albicans*, BioChain®, Eureka, CA; *Saccharomyces cerevisiae*, White Labs, San Diego, CA), and a pooled bacterial sample of six microorganisms that was treated as a single sample (*Rhodobacter sphaeroides, Escherichia coli, Bacillus cereus;* ATCC, Manassas, VA), *Staphylococcus aureus, Neisseria gonorrhoeae, Bacillus subtillis* (Cambridge University Hospital, Cambridge, United Kingdom). Baboon and monkey (1 ng) and all other nonhuman samples (10 ng) were amplified with DPMB and processed according to the ForenSeq™ DNA Signature Prep Guide (Illumina part #15049528) [77].

2.3. Sample preparation for sensitivity, mixture and partially degraded DNA studies

Serial dilutions of the following amounts of template DNA (2800 M and NA12878) were prepared in molecular grade water for sensitivity studies: 1 ng, 500 pg, 250 pg, 125 pg, 62.5 pg, 31.25 pg, 15.625 pg, and 7.82 pg, and amplified with DPMA and DPMB according to the ForenSeq™ DNA Signature Prep Guide [77].

Human genomic DNA mixtures were prepared from the following four purified gDNA samples (Coriell Institute for Medical Research, Camden, NJ): NA12877 (male, Caucasian), NA18507 (male, African), NA12878 (female, Caucasian), and NA19238 (female, African). gDNA mixtures were prepared from two males (MM; NA12877:NA18507) and from two females (FF; NA12878: NA19238) at nine ratios (99.9:0.1, 99:1, 95:5, 93.75:6.25, 90.9:9.1, 90:10, 87.5:12.5, 75:25, 50:50). gDNA from a female and from a male were also prepared as mixtures (FM; NA12878:NA18507) at four ratios (95:5, 90:10, 75:25, 50:50). Each of the four DNA samples that served as "contributors" was also run as single source samples. Each sample was amplified with DPMB and prepared for sequencing according to the ForenSeq™ DNA Signature Prep Guide [77].

Partially degraded samples were prepared using two methods: (1) mechanical shearing alone (500 ng gDNA; 2800 M, NA12878) using an S2 focused-ultrasonicator (Covaris®, Woburn, MA), and (2) mechanical shearing with subsequent DNase I digestion (New England Biolabs Inc., Ipswich, MA) according to manufacturer's instructions. Briefly, 1U of DNase I (RNase-free) was added to the sheared gDNA and incubated at 37 °C for 1 min, after which 0.5 mM EDTA was added and the entire reaction was heat inactivated at 75 °C for 10 min. These "sheared only" and "sheared + DNase I" samples, and the original high molecular weight control DNAs, were amplified with DPMB in triplicate and prepared for sequencing according to the ForenSeq™ DNA Signature Prep Guide [77]. Two template input amounts of partially degraded DNA were tested in PCR1: 1 ng as measured post-shearing/enzymatic treatments, and additionally the maximum volume of sample recommended (5 μl), containing approximately 45 ng based on pre-treatment quantification.

2.4. PCR inhibition study

A variety of known PCR inhibitors were independently "spiked" into the PCR1 reaction before amplification of 2800 M DNA samples (in triplicate) with DPMB. Specifically, 16.66 μM, 33.3 μM, 66.66 μM, and 133.3 μM of hematin, humic acid, indigo dye, tannic acid (Sigma-Aldrich®, St. Louis, MO), or urban dust (National Institute of Standards and Technology, Gaithersburg, MD) were added directly to the reaction. Samples were processed according to the ForenSeq™ DNA Signature Prep Guide [77].

2.5. Repeatability and reproducibility studies

Repeatability and reproducibility studies were performed independently for DPMA (153 loci) and DPMB (231 loci). For repeatability studies, one analyst processed five identical 96-well plates using DPMA and another five plates for DPMB. For reproducibility studies, five different analysts each processed a single plate using DPMA and another single plate using DPMB. The plate layout for samples amplified with DPMA consisted of 96 samples: 95 replicates of 2800 M DNA and one no template control (NTC). The plate layout for samples amplified with DPMB consisted of 32 samples: 31 replicates of 2800 M DNA and one NTC sample. All indexed libraries prepared on a plate were pooled together and sequenced on one MiSeq FGx™ benchtop sequencer. The same MiSeq FGx™ instrument was used within each set of repeatability and reproducibility runs and three instruments were tested in total (the same instrument was used for the repeatability DPMB and reproducibility DPMA studies).

2.6. ForenSeq DNA library preparation

NGS libraries of targeted forensic loci from each sample described here were prepared using the ForenSeq™ DNA Signature Prep Kit (Illumina Inc., San Diego CA) (Fig. 1) with total reaction volumes of 15 μl for PCR1 (amplify and tag forensic targets), and 50 μl for PCR2 (enrich forensic targets), using either DPMA or DPMB, as described for each specific validation study. ForenSeq™ libraries were prepared using the GeneAmp® PCR System 9700 with a gold-plated block (Applied Biosystems, Foster City, CA). Adhesive microseals were applied to 96-well plates, and sealed using a rubber roller, before following steps in the ForenSeq™ protocol for shaking, vortexing, centrifugation and thermalcycling. Microseal 'B' adhesive seals (Bio-Rad, part # MSB-1001) were used for shaking, centrifuging, and long-term storage (*i.e.*, steps conducted between −40 °C to 110 °C), with suitable skirted or semi-skirted PCR plates; Microseal 'A' adhesive seals (Bio-Rad, part number MSA-5001) were used for thermalcycling. PCR1 preparation was conducted in a discrete pre-PCR area, and then transferred to a designated post-PCR room for thermalycling to amplify target loci, as described in ForenSeq™ DNA Signature Prep Guide, and as follows: a 98° C initial incubation (3 min), 8 cycles of [96° C (45 s), 80° C (30 s), 54° C (2 min) (ramp at 0.2 °C per second), 68° C (2 min) (ramp at 0.2 °C per second)], 10 cycles of [96° C (30 s) and 68° C (3 min) (ramp at 0.2 °C per second)], followed by a final extension at 68° C (10 min) and an infinite hold at 10° C. PCR2 set up and thermalcycling, for index addition (i7 and i5), were performed in the post-PCR room, as described in ForenSeq™ DNA Signature Prep Guide, and as follows: a 98° C initial incubation (30 s), 15 cycles of [98° C (20 s), 66° C (30 s), 68° C (90 s)], followed by a final extension at 68° C (10 min) and an infinite hold at 10° C. ForenSeq™ targeted amplicon libraries were purified using Sample Purification Beads (SPB) (Beckman Coulter, Brea, CA), followed by bead based normalization using Library Normalization Beads 1 (LNB1) (Illumina Inc., San Diego CA) according to the ForenSeq™ DNA Signature Prep Guide [77].

Normalized ForenSeq™ libraries were quantified and their quality assessed using the automated capillary electrophoresis DNA Fragment Analyzer™ (Advanced Analytical Technologies, Inc., Ames, IA) according to the manufacturer's instructions. This optional step provides visualization of the general quality of a DNA

library (*e.g.*, pristine, partially degraded) as well as ForenSeq™ targeted amplicon size distribution (longer STRs and shorter SNPs), and is not required for generating high quality data in a routine operational setting. See Supplementary Fig. 2 for an example.

2.7. MiSeq FGx sequencing

The indexing step described above (PCR2) tags each amplicon generated from one individual DNA sample with a unique combination of molecular index sequences on their 5′ and 3′ ends. This allows a MiSeq FGx™ instrument to sequence and separate data from pooled DNA libraries in a single sequencing run. Normalized and indexed ForenSeq™ DNA libraries (5 μl) were pooled into a single microcentrifuge tube and diluted in Hybridization Buffer (HT1). Human Sequencing Control (HSC) (2 μl) from the ForenSeq™ DNA Signature Prep kit was denatured with NaOH (HP3) by incubation at room temperature for 5 min, then added to the pooled libraries, followed by an additional heat denaturation at 96 °C (2 min) to denature the entire library pool. The HSC is a DNA library pool of 23 ForenSeq™ STRs serving as a positive sequencing control for the MiSeq FGx™ instrument. The denatured library pool was immediately pipetted into the MiSeq FGx™ Reagent Cartridge for sequencing on a MiSeq FGx™ instrument (Illumina Inc., San Diego, CA), according to manufacturer's instructions. Sequencing runs were organized as follows: 96 samples were pooled when amplified with DPMA, and 32 samples were multiplexed when DPMB was employed. The sequencing by synthesis (SBS) [36] run consists of 398 total sequencing cycles over approximately 30 h. The first read (Read 1) is 351 SBS cycles, where the first 351 nucleotides (one nucleotide base per cycle) are sequenced in each of the targeted DNA amplicons (or less, depending upon amplicon length). The first index read (Index 1) is 8 cycles that determine the i7 index; the second index read (Index 2) is 8 cycles that determine the i5 index. Finally, the second read (Read 2) is 31 SBS cycles that determine the last 31 nucleotides of each forensic STR and SNP amplicon, in the reverse direction relative to Read 1.

2.8. Orthogonal genotyping for Concordance studies

Orthogonal, lower resolution STR genotyping data (relative to deep sequencing) for DNA samples 2800 M, NA12877, and NA12878 were generated by amplification with the following five multiplexes: AmpFℓSTR® Identifiler® PCR Amplification kit, AmpFℓSTR® NGM Select PCR Amplification kit (Applied Biosystems, Foster City, CA), PowerPlex® Fusion system, PowerPlex® Y23 System (Promega Corp., Madison, WI), and the Investigator Argus X-12 kit (Qiagen, Valencia, CA), according to manufacturers' instructions. Additionally, DNA samples NA18507 and NA19238, used in the mixture studies, were genotyped with the above multiplexes excluding the Argus X-12 kit. PCR products were separated and detected on an AB 3130×l Genetic Analyzer (Life Technologies, Carlsbad, CA) following the manufacturer's recommendations. Samples were injected for 10 s at 3 kV into performance optimized polymer (POP-4™; Life Technologies, Carlsbad, CA) using the HIDFragmentAnalysis36_POP4 Module (Life Technologies, Carlsbad, CA) and a 1500 s run time. Capillary electrophoresis (CE) data were collected using the AB 3130xl Genetic Analyzer Data Collection Software 3.0, and analyzed with GeneMapper® ID software v3.2.1 (Life Technologies, Carlsbad, CA) with peak amplitude thresholds for detection and interpretation (stochastic) each set at 50 relative fluorescence units (RFU).

Orthogonal SNP genotyping data for samples 2800 M, NA12877, NA12878, NA18507, and NA19238 were obtained from the Platinum Genomes pedigree data set (vcf files available at FTP site http://www.illumina.com/platinumgenomes/), and from publicly available whole genome sequencing data from the 1000 Genomes Project (www.1000genomes.org).

2.9. Secondary and tertiary data analysis

MiSeq FGx™ sequencing data were analyzed using the ForenSeq™ Universal Analysis Software (version 1.0) for allele and genotype calling, and tertiary analyses including automated sample comparisons, generation of population statistics such as random match probabilities, estimation of eye and hair color as well as biogeographical ancestry (Illumina, Inc., San Diego, CA). Based on empirical studies, default analysis parameters used throughout were a 1.5% analytical threshold (AT) and 4.5% interpretation threshold (IT), for all loci except for DYS389II (>5.0% AT, >15% IT), DYS448 (>3.3% AT, >10% IT) and DYS635 (>3.3% AT, >10% IT), where noise warranted separate values [77] (Supplemental Table 1). AT and IT values were determined for a locus by multiplying the analysis parameter percentage value by the sum of read counts at that locus. In cases of low coverage, a minimum read number of 650 reads was used for the locus in determination of the threshold values. Default stutter filter percentages for autosomal STR, Y-STR, and X-STR markers are documented in Supplemental Table 1 and range from 7.5% (D2S441, D4S2408, PentaD) to 50% (DYS481). MiSeq FGx™ run quality metrics and target ranges of each were as follows: cluster density (400–1650 K/mm^2) on the MiSeq FGx™ flow cell, clusters passing quality filter (≥80%), as well as phasing (≤0.25%) and pre-phasing (≤0.15%) for Read 1 and Read 2. If a run falls outside of the target ranges occasionally, it can still produce sufficient data for analysis and interpretation; values that deviate substantially from the target range can negatively impact other quality metrics, and decrease the quantity of data produced from the run. Approximately 1736 PCR 1 reactions were analyzed; samples were analyzed and included whether read counts were more or less than the general guideline of 85,000 reads/sample (details in Accuracy, Precision, and Call Rate section), viewable on the Sample Representation tab of the ForenSeq™ Universal Analysis Software's Quality Metric page (Q page). The HSC positive controls passed all default QC metrics, including >250 reads per locus. Negative amplification controls (NTC) where 0–4 loci produced read counts above the default analytical threshold (AT) were considered blank; one NTC yielded 4 loci with read counts >AT (see Results). Semi-automated allele and genotype calls were accomplished using the ForenSeq™ Universal Analysis Software, including application of quality indicators that assisted in manual data review of loci (*e.g.*, to evaluate samples with low coverage at a locus or to designate STR stutter products from parent alleles) (Fig. 2). For the studies presented in this paper, genotype calls were made as follows for STRs and SNPs: reads were called as alleles when greater than the analytical threshold (AT), and not identified as stutter (examples in Fig. 2A were called as 6, 9.3 (Fig. 2-A1); 14,15 (Fig. 2-A2) and 11,12 (Fig. 2-A3)). If a single autosomal allele was greater than the interpretation threshold (IT), it was called as a homozygote (*e.g.*, (12,12) in Fig. 2-A4), whereas if reads for a single allele were detected between the AT and IT, then was designated as an "Ambiguous Genotype" (*e.g.*, (13,*) Fig. 2-A5), to account for possible non-detection of a sister allele. In cases where the highest signal (read counts) was less than the AT (*e.g.*, Fig. 2-A6) an allele was not called. Additionally, for SNP loci, reads greater than 10% of the total reads per locus were designated as alleles (*e.g.*, in Fig. 2-B1 the genotype is A, A; Fig. 2-B2 the genotype is C, T); Fig. 2-B3 the genotype is T, * (where * = reads less than 10% of total reads per locus); and Fig. 2-B4 a genotype was not called.

Genotype concordance was assessed by comparing allele calls from forensic loci in the MiSeq FGx™ System (autosomal STRs, X-STRs, Y-STRs, iiSNPs, piSNPs, aiSNPs) to orthogonal STR and SNP

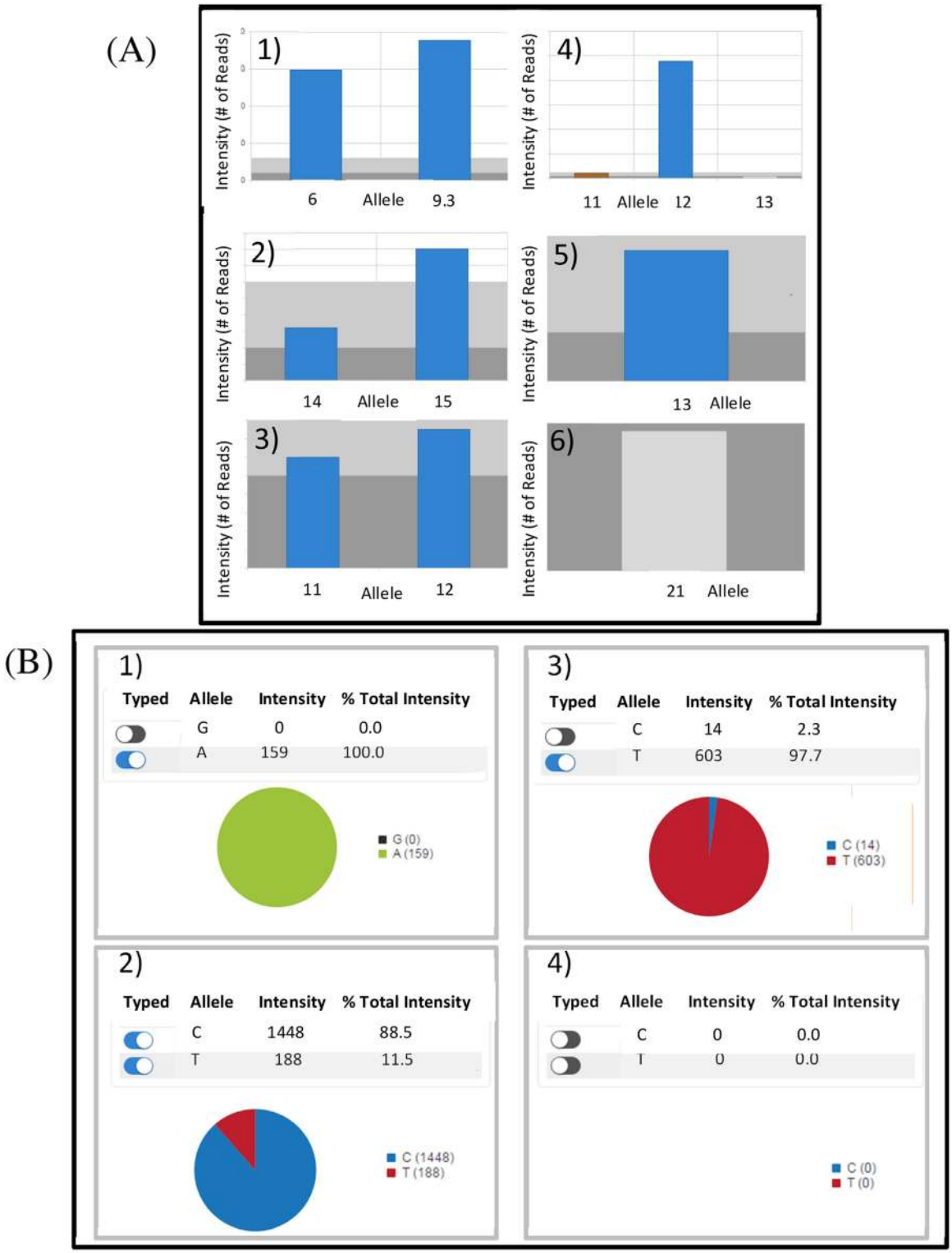

Fig. 2. STR and SNP Allele Calling – ForenSeq™ Universal Analysis Software™.
Top panel: (A) STR locus details for six example data scenarios displayed relative to the number of reads (y-axis), after manual review. Color-coding indicates the following: dark grey and light grey horizontal shading represent the analytical threshold (AT) and interpretation threshold (IT), respectively. Blue vertical bars are typed STR alleles (A1–A5); brown bar indicates STR stutter (A4), and light grey bar (A6) indicates reads below the AT. In examples A2 and A3, quality indicators were displayed for "interpretation threshold" and "imbalanced", for "interpretation threshold" in example A5, and for "low coverage" in example A6. Data were manually reviewed; alleles between the AT and IT were typed. For the purposes of genotype concordance, example A5 is classified as "Ambiguous Genotype", and example A6 classified as "Below AT".
Lower panel: (B) SNP locus details for four example data scenarios. Quantitative data for each of the four example loci are displayed in a table, above a circle plot where green indicates a homozygote (B1) and blue/red indicate proportionally the alleles in putative heterozygotes (B2). The "imbalanced" indicator was displayed for examples B2 and B3 and the "low coverage" indicator was displayed for example B4. For the purposes of genotype concordance, example B4 is classified as "Below AT".

typing methods. The four possible classifications from these comparisons were "Concordant", "Ambiguous Genotype", "Below AT" or "Discordant Allele(s)", as defined by the following criteria. A "Concordant" result was indicated when the STR (autosomal STR, X-STR, Y-STR) or SNP (iiSNP, aiSNP, piSNP) genotype was the same as those generated by the orthogonal methods described previously. An "Ambiguous Genotype" was obtained when reads for one allele were detected with read counts between the analytical and interpretation threshold levels (*e.g.*, Fig. 2, A5 where a (13,*) was called). The "Below AT" (BAT) classification describes examples of sequencing coverage where read counts were detected at less than the default AT (*e.g.*, Fig. 2, A6 and B4). The last classification, "Discordant Allele(s)", contains the following six outcomes where genotype assignment between the ForenSeq result and that of the orthogonal method differs: 1) an STR locus where at least one allele call differs from a CE-generated allele call and the STR genotype is not ambiguous, 2) a Y-STR allele call in a female sample, 3) an X- or Y-STR with two allele calls in a male DNA sample (excluding the multi-copy Y-STR loci DYS385a-b and DYF387S1), 4) SNP loci: orthogonal data indicated a homozygote "A" and deep sequencing detected a homozygote "B" allele, 5) SNP loci: orthogonal data indicated a homozygote and deep sequencing detected a heterozygote, and 6) SNP loci: orthogonal data indicated a heterozygote, deep sequencing detected a homozygote, and the SNP genotype is not ambiguous. For the purposes of statistical analysis for genotype concordance (see Results), only genotypes classified as concordant are considered in accuracy estimates The other three categories enable a more discriminating examination of differences.

3. Results

3.1. Characterization of genetic markers

Mode of inheritance, genomic mapping, and polymorphism type were reconfirmed for the autosomal STRs X-STR, Y-STR, identity informative SNP (iiSNP), phenotypic informative SNP (piSNP), and biogeographical ancestry informative SNP (aiSNP loci amplified with the ForenSeq™ DNA Signature Prep Kit. Briefly, data indicate that these forensic, autosomal STRs and SNPs obey Mendelian inheritance, and that the ForenSeq™ X-STRs and Y-STRs follow the expected segregation of sex-linked loci and uniparental inheritance. Genomic locations of the ForenSeq™ loci (chr:pos) are listed in the ForenSeq™ DNA Signature Prep Guide [77]. The NGS detection methodology of the MiSeq FGx™ instrument is described in Materials and Methods and instrument manual [78,79].

3.2. Collection substrate and cell lysate testing

Performance of the MiSeq FGx™ System (kit, instrument and software) was evaluated on human body fluid samples and three types of sample collection substrates. Human sample lysates were prepared from buccal cells collected on sterile cotton swabs (n = 78), Bode Buccal DNA Collector™ filter paper punches (n = 60), and FTA™ card punches (n = 34). Human blood samples were collected onto FTA™ cards (n = 10). These 182 samples were amplified with DPMA. Additionally, 41 human DNA lysates prepared from human buccal cells collected on sterile cotton swabs (n = 22) and on Bode Buccal DNA Collector™ filter paper punches (n = 19) were amplified with DPMB. Genetic profile completeness for the 58 autosomal STRs, Amelogenin, the 94 iiSNPs and 78 ai,piSNPs (DPMB only) are shown in Table 1 and Supplemental Table 2. Partial data, with most loci not detected (<10% "completeness "), were observed for seven samples, potentially due to variability in sample collection and/or possible PCR inhibition. Full autosomal STR and ii,ai,piSNP profiles were produced from >89% and >87% of the samples tested, respectively.

3.3. Species specificity

Performance of the MiSeq FGx™ System was tested by preparing DNA libraries, from a variety of nonhuman genomic DNA samples, using the ForenSeq™ DNA Signature Prep Kit. The following samples from eukaryotic and prokaryotic species were prepared, in duplicate, along with 2800 M positive control DNA and one NTC sample: two nonhuman primates (1 ng input each), nine non-primate mammalian (10 ng input each), one avian (10 ng input), two fungal (10 ng input each), and a pooled bacterial sample (10 ng input total), sequenced on a MiSeq FGx™ instrument, and analyzed with ForenSeq™ Universal Analysis Software.

The number of sequencing reads (intensity) generated by each of the fifteen species samples, the NTC, and the 2800 M positive controls are shown in Fig. 3. In comparison to the two 2800 M positive control samples which averaged ~270,000 sequencing reads, the majority of the species samples yielded sequencing reads in a range closer to the NTC sample (*e.g.*, 251 reads). The ferret, *S. cerevisiae*, and the bacterial pool produced no sequencing reads. The averaged chicken, dog, horse, hamster, *C. albicans*, pig and cat samples each produced ≤1000 reads and the cow duplicate

Table 1
STR and SNP Profile Percentage Completeness for Substrate and Cell Lysate Testing.

Substrate	n=	Primer Mix	Number of Samples with% aSTR Profile				
			100%	99–90%	89–50%	49–10%	<10%
Cotton Swabs – Buccal	78	DPMA	65	8	1	0	4
Bode Collectors – Buccal	60	DPMA	57	0	1	1	1
FTA Card – Buccal	34	DPMA	33	1	0	0	0
FTA Card – Blood	10	DPMA	10	0	0	0	0
Cotton Swabs – Buccal	22	DPMB	16	5	0	1	0
Bode Collectors – Buccal	19	DPMB	19	0	0	0	0
Substrate	n=	Primer Mix	Number of Samples with% ii,ai,piSNP Profile				
			100%	99–90%	89–50%	49–10%	<10%
Cotton Swabs – Buccal	78	DPMA	66	7	1	0	4
Bode Collectors – Buccal	60	DPMA	57	0	1	0	2
FTA Card – Buccal	34	DPMA	33	1	0	0	0
FTA Card – Blood	10	DPMA	10	0	0	0	0
Cotton Swabs – Buccal	22	DPMB	10	9	1	1	1
Bode Collectors – Buccal	19	DPMB	19	0	0	0	0

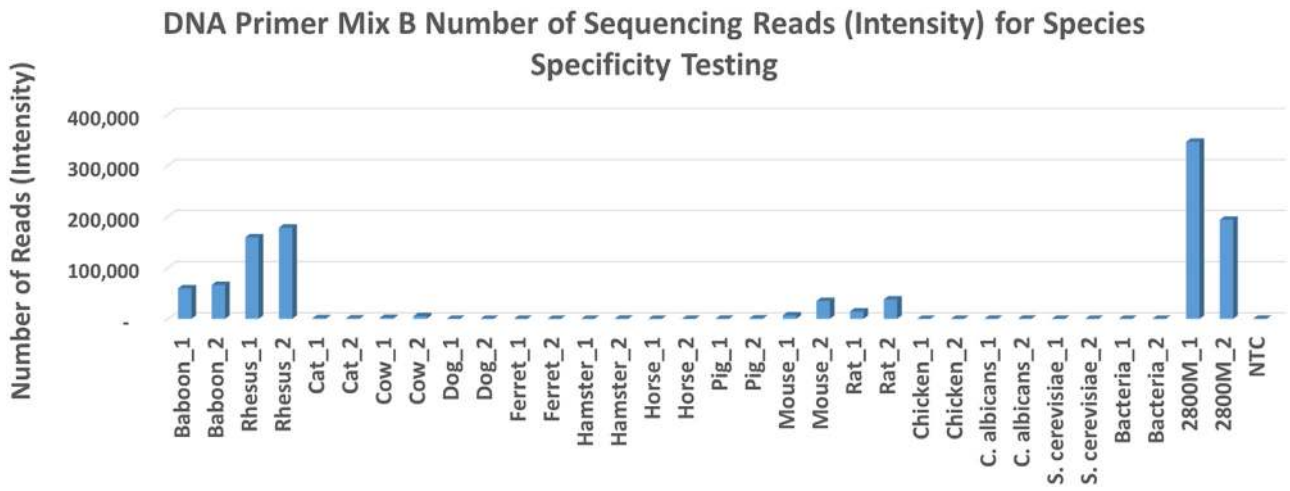

Fig. 3. Number of Sequencing Reads Generated in Species Specificity Testing. The total numbers of sequencing reads (Intensity) are plotted for fifteen species amplified with DPMB in duplicate.

samples produced ~3000 reads. Not surprisingly, the two nonhuman primates, the rhesus monkey and baboon, produced the highest numbers of sequencing reads, ~169 K and ~63 K, respectively. Rodent species (mouse, rat) yielded ~21 K and ~26 K reads, respectively (Supplemental Table 3).

STR and SNP allele detection potential in nonhuman species was evaluated using the ForenSeq™ Universal Analysis Software. Fig. 4 illustrates allele calls generated for the fifteen nonhuman samples and 2800 M, for the 58 STRs and 172 SNPs amplified with DPMB (Supplemental Table 3). Comparable to the sequencing read count data shown in Fig. 4, the species samples with the higher number of sequencing reads also produced more genotyped STR and SNP alleles. Alleles were detected in rhesus monkey and baboon at ~11% of the STR loci and ~25% of the SNP loci, and in bovine sample for ~20% of the STRs and ~15% of the SNPs. Alleles were detected in mouse and rat samples for ~51% and ~60% of the STR and SNP loci, respectively. These data may be compared to the 2800 M positive control sample, where alleles were detected for ~85% of STRs and 100% of SNPs. Overall, the nonhuman species amplifications with the ForenSeq™ DNA Signature Prep Kit are reminiscent of current forensic DNA typing by capillary electrophoresis.

Data from the two rodent species were further analyzed. Differences were noted between the rodent species and human samples, among replicates of each rodent species, and between the rat and mouse samples. The rodent species (10 ng input) each produced partial profiles with elevated numbers of QC Indicators in the ForenSeq™ software, relative to human samples; multiple loci were not detected. The following were observed in the rat samples: For STRs, depending upon the replicate, allele calls were made at 38 of 60 loci, with 20 and 22 loci having no reads, as well as 34 and 36 loci with 1–4 QC Indicators triggered. For iiSNPs, depending upon the replicate, allele calls were made at 56 of 94 loci, with 30 and 38 loci having no reads, as well as 48 and 61 loci with 1–4 QC Indicators triggered. The following were observed in the mouse samples: For STRs, depending upon the replicate, allele calls were seen at 30 and 32 of 60 loci, with 28 and 30 loci having no reads (ND), as well as 43 and 49 loci with 1–4 QC Indicators triggered. For iiSNPs, depending upon the replicate, allele calls were seen at 52 and 65 of 94 loci, with 29 and 42 loci having no reads (ND), as well as 50 and 73 loci with 1–4 QC Indicators triggered. The typed loci and the undetected loci differed between the mouse and rat samples. Different patterns of partial profiles were seen for the piSNPs and aiSNPs between the mouse and rat samples. Eye and hair color estimations were not viable for rodent samples as all piSNPs are required. Biogeographical ancestry estimation placed the rat samples within the East Asian population group on the principle component analysis (PCA) plot in the ForenSeq™ software, and the mouse sample among the Admixed American samples.

More interestingly, multiple sequences were detected within allele-like products of the same length in rat and mouse samples, and not observed in human samples. In rat samples multiple sequences were detected at D13S1358, FGA, DXS10148, DXS10074, DYS505, DYS635, DYS437, DYS612 and DYS460. The numbers of individual sequences were inconsistent across rat replicates. For example, in rat at D3S1358, one replicate produced two 15 alleles (19 and 103 reads) and four 16 alleles (11,12,15 and 120 reads) while in the other replicate there were six 15 alleles detected (11–197 reads) and four 16 alleles (11–133 reads). The sequences of the products resembling human allele 15 differed at one SNP and at four SNPs between replicates; in the other replicate there were three SNPs within the different sequences of the 16 allelic products. Additional visual prompts were apparent in the X-STR data where five sequences (12–105 reads) were detected at DXS10074, and eight sequences (13–405 reads) at DXS10148, differing at five SNP positions. Similar results were detected at the Y-STR loci listed above. Multiple allele-like products detected in mouse samples displayed the hallmarks of the rat data but manifested at different loci as follows: D6S1043, D9S1122, D16S539, D17S1301, D19S433 (and DYS635, as also observed in rat samples). Five human-like sequences were detected in one mouse sample at D19S1122. These observations may serve to assist an analyst in recognizing a rodent specific pattern, and distinguishing from a human sample, or human/rodent mixed sample. Additional analyses to investigate reproducible characteristics across additional rat and mouse sample genotypes could add more information in this regard. These data may also assist analysts who seek to monitor rat and mouse cell lines, using polymorphic STRs.

3.4. Sensitivity study

The maximum number of libraries of targeted forensic loci simultaneously sequenced on a MiSeq FGx™ flow cell depends upon the total number of reads (depth of coverage) desired per locus. For example, more samples may be run when single source samples (*e.g.*, with DPMA) are being analyzed as compared to a run where some or all samples may consist of a DNA mixture with one

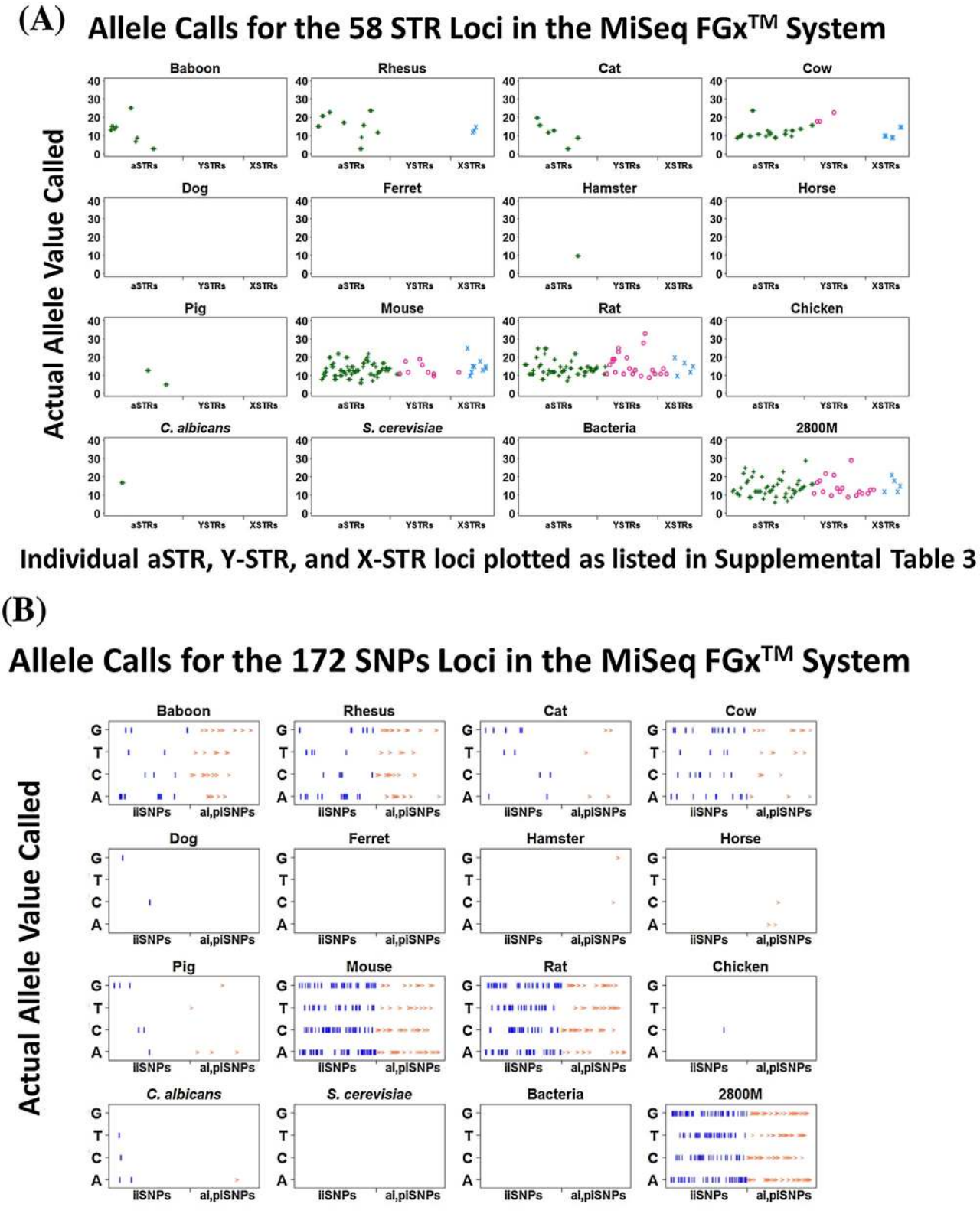

Fig. 4. STR and SNP "alleles " generated in Species Specificity study using ForenSeq™ Universal Analysis Software. Allele calls are shown for fifteen nonhuman species and human DNA 2800 M (prepared in duplicate), amplified with the ForenSeq™ DNA Signature Prep Kit (DNA Primer Mix B; DPMB), sequenced and analyzed with the ForenSeq™ Universal Analysis Software. Calls are shown for each replicate for each individual species or the pooled bacterial sample. The 27 autosomal STRs (green '+'), 24 Y-STRs (pink 'o'), and 7 X-STRs (blue 'x') are shown in (A). The 94 iiSNPs (blue '|') and 78 ai,piSNPs (orange '>') are shown in (B). The x-axis for both plots is listed in Supplemental Table 3. Plots were generated using R version 3.2.0 software (https://www.r-project.org/).

or more minor components of interest (*e.g.*, with DPMB). Sensitivity studies were conducted using DPMA (152 loci) on 96 samples, and using DPMB (230 loci) on 32 samples, to evaluate the MiSeq FGx™ System's ability to generate reliable genotypes and haplotypes at various gDNA template amounts. D22S1045 was not included in these two studies. Two gDNA samples, 2800 M and NA12878, were serially diluted; the input DNA amounts of 1 ng, 500 pg, 250 pg, 125 pg, 62.5 pg, 31.25 pg, 15.625 pg, and 7.82 pg were amplified in quadruplicate. ForenSeq™ DNA Signature Prep Kit STR and SNP genotypes and haplotypes, as analyzed with ForenSeq™ Universal Analysis Software, were compared to orthogonal genotyping data from CE fragment length detection for STRs, and from bead array data for SNPs, to determine allele call rate and comparative accuracy of the sequencing-based genotype data, in relation to the amount of input gDNA.

Sensitivity results for the 2800 M and NA12878 gDNA samples amplified with DPMA (Fig. 5A), and DPMB (Fig. 5B) yielded similar results. Genomic DNA inputs from 1 ng down to 62.5 pg produced 100% call rates (no allele loss) and yielded 100% genotype concordance for all tested autosomal loci with both primer mixes, when compared to conventional genotyping methods and kits (CE for STRs) and SNP arrays. For Y-STR and X-STR loci, gDNA inputs of 1 ng, 500 pg, 250 pg, 125 pg, and 62.5 pg yielded 100% call rates for all 31 of the loci in DPMA and DPMB, with the exception of one 2800 M amplification where DXS10103 was not detected (read count = 0) at 125 pg. When amplified with DPMA these five gDNA input amounts produced the following haplotype concordance percentages, at Y-STRs and X-STRs, respectively: 1 ng (100%/100%), 500 pg (99.5%/100%), 250 pg (100%/100%), 125 pg (99.0%/100%), and 62.5 pg (99.0%/96.4%). One or neither allele(s) at autosomal, X- and Y-STR loci may be detected from less than 62.5 pg gDNA input (*i.e.*, 31.25 pg, 15.625 pg, 7.82 pg), in both DPMA and DPMB (Supplemental Table 4).

Sensitivity results for the 94 iiSNPs in both DPMA (Fig. 5A) and DPMB and the aiSNPs and piSNPs in DPMB (Fig. 5B) were similar. Genomic DNA inputs of 1 ng, 500 pg, and 250 pg yielded >99.9%, >99.6%, and >99.6% iiSNP concordance, respectively, from >99% of called iiSNP loci from both primer mixes with no locus loss. At aiSNPs and piSNPs, these three DNA input amounts yielded 100% allele call rates, and 100% genotype concordance. SNP allele calls from gDNA inputs $\leq$125 pg similarly produced high accuracy values, however, loss of some alleles and some loci, or low coverage (reads below the default analytical threshold), were noted. Specifically, genotype concordance for iiSNPs in DPMA, iiSNPs in DPMB, and for ai,piSNPs in DPMB, were as follows: 99.6%, 98.7%, 99.7% (125 pg), 97.9%, 96.1%, 98.9% (62.5 pg), 90.6%, 86.7%, 92.6%

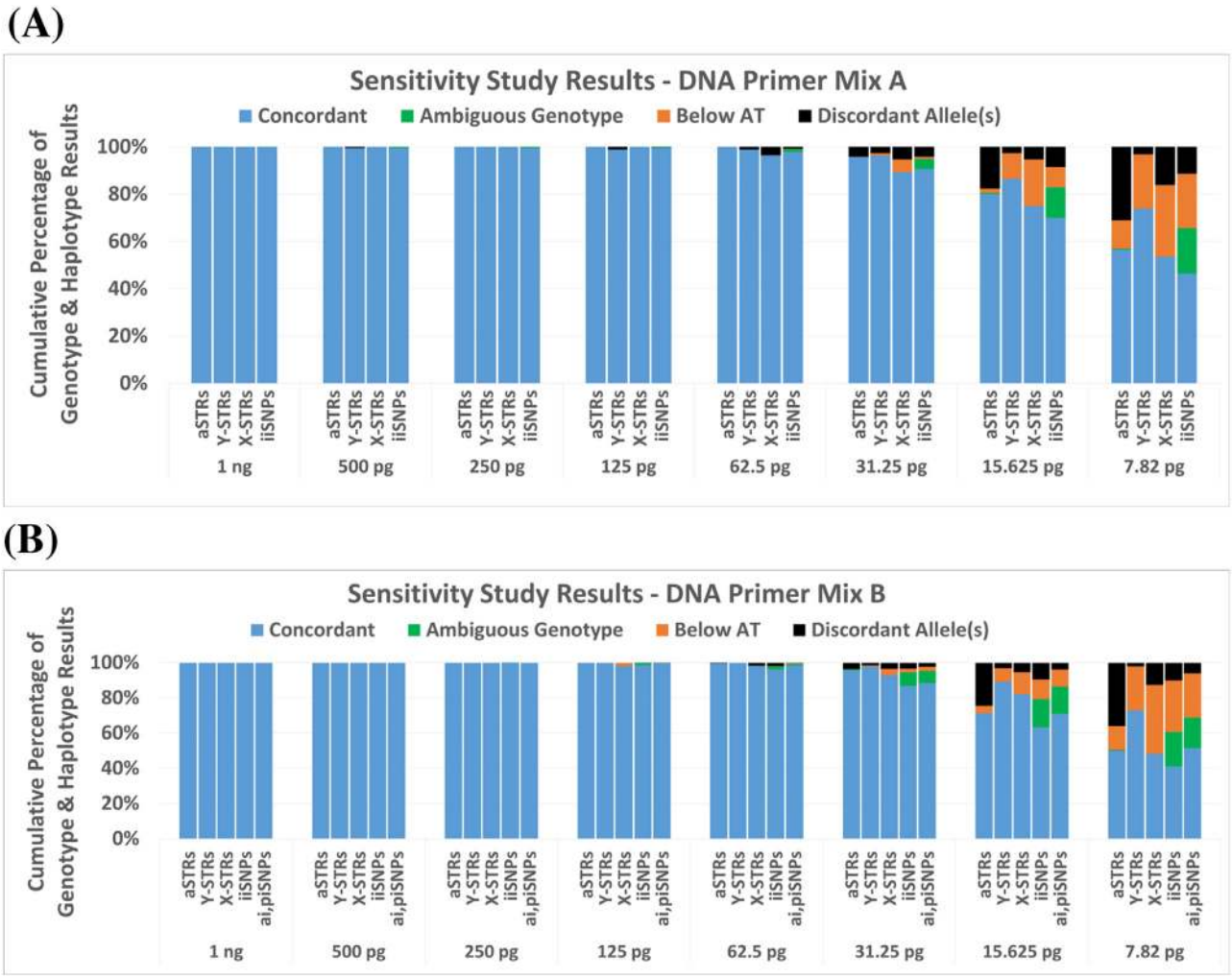

Fig. 5. Sensitivity Study. Two gDNA samples (2800 M and NA12878) were serially diluted; each input DNA amount was amplified in quadruplicate using the ForenSeq™ DNA Signature Prep Kit with DNA primer mix A (A) or DNA primer mix B (B). Genotype and haplotype outcomes, as compared to orthogonal typing methods, are designated as either, "Concordant" (blue), "Ambiguous Genotype" (green), "Below AT" (orange), or "Discordant Allele(s)" (black) (see Materials and methods) and plotted as cumulative percentage of the total number of outcomes for each of the gDNA inputs (1 ng, 500 pg, 250 pg, 125 pg, 62.5 pg, 31.25 pg, 15.625 pg, 7.82 pg). Shown are the average percentages of the four reactions of both gDNA samples.

(31.25 pg), 70.0%, 63.3%, 82.5% (15.625 pg), and 46.4%, 41.0%, 66.4% (7.82 pg) (Supplemental Table 4).

3.5. Genomic DNA mixture studies

Performance of the MiSeq FGx™ Forensic Genomics System was evaluated with mixed source samples from two DNA contributors. Three sets of DNA mixtures were prepared using four commercially available purified DNA samples (male:male (MM), female:male (FM), female:female (FF)) in varied ratios from 99.9:0.1% to 50:50% (FM samples were not tested at 0.1% or 1%). Mixed DNA samples were amplified in triplicate with the ForenSeq™ DNA Signature Prep Kit using DPMB, and processed according to the Illumina MiSeq FGx™ User Guide [77,78]. Sequencing reads were analyzed with ForenSeq™ Universal Analysis Software; genotyping results are provided in Supplemental Table 5.

The ForenSeq™ Universal Analysis Software provides a visual "single source" indicator that displays as green when no mixture is detected and as orange when a mixture may be present. Quality Control Indicators in the software that can indicate the possible presence of a DNA mixture include: "allele count" displayed when more alleles than expected for a single source sample are detected, "imbalanced" displayed when intralocus balance is less than expected and thus contributors to a mixture could share an allele, and "interpretation threshold". Ratios in mixture studies 100:0, 99.9:0.1, 99:1 (Male:Male) and 95:5 (Female:Male) were displayed as single source, using the default AT and IT. All other mixture ratios triggered at least one, or a combination of the QC Indicators for "allele count", "imbalanced" and "interpretation threshold".

ForenSeq™ data analysis of mixed samples consisted of quantification of the number of alleles detected above the default analytical threshold for each STR (Fig. 6) and SNP locus (Fig. 7). Mixture ratios were selected to push the limits of the system, in the context of default AT and IT. When the default AT is applied, no calls are expected for the 0.1% minor DNA contributor, and detection of some minor alleles in the 1% samples are possible. Fig. 6 illustrates that, as the proportion of minor contributor DNA increased, so did the number of alleles detected for all three STR categories. As the percentage of the minor contributor increased from 0% (single source) to 1%, 5%, 10%, 25%, and finally 50%, the number of called autosomal STRs increased from 54 alleles (27 loci x 2 alleles, single source) to an average among replicates of 58, 73, 80, 82, and 83 allele calls, respectively (Fig. 6A). Similarly, for the Y-STRs and X-STRs, a gender-specific increase in allele number was observed, depending on the amount of minor contributor present in each sample. For Y-STRs, male:male (MM) mixtures showed an increase in allele count relative to the single source (26 alleles). As the percentage of the minor male contributor increased to 1, 5, 10, and 25–50% of the total DNA sample, the number of Y-STR alleles identified increased to an average of 27, 41, 43, and 47 alleles, respectively (Fig. 6B). For the female:male (FM) mixtures, Y-STRs were not detected in the single source female sample, yet in the presence of a 5% minor male contributor, 20 Y-STRs were called, and a full 24 locus Y-STR profile was produced for the 10% minor contributor (100 pg of the total DNA input) (Fig. 6B). Y-STR alleles were not detected in the female:female (FF) mixtures. X-STR allele counts showed a similar increase based on the gender, and on the minor contributor percentage. For the FM and FF mixtures, the female single source samples produced full 14 X-STR multilocus profiles, while in the MM mixtures the single source samples produced a full 7 allele multilocus profile of the X chromosome. Upon addition of a female minor contributor at 1% and 5% to the major female sample (FF mixture) the number of alleles counted increased from 14 to 18 and 22, respectively; at each proportion >5%, the total number of X-STR alleles plateaued at 24 (Fig. 6C).

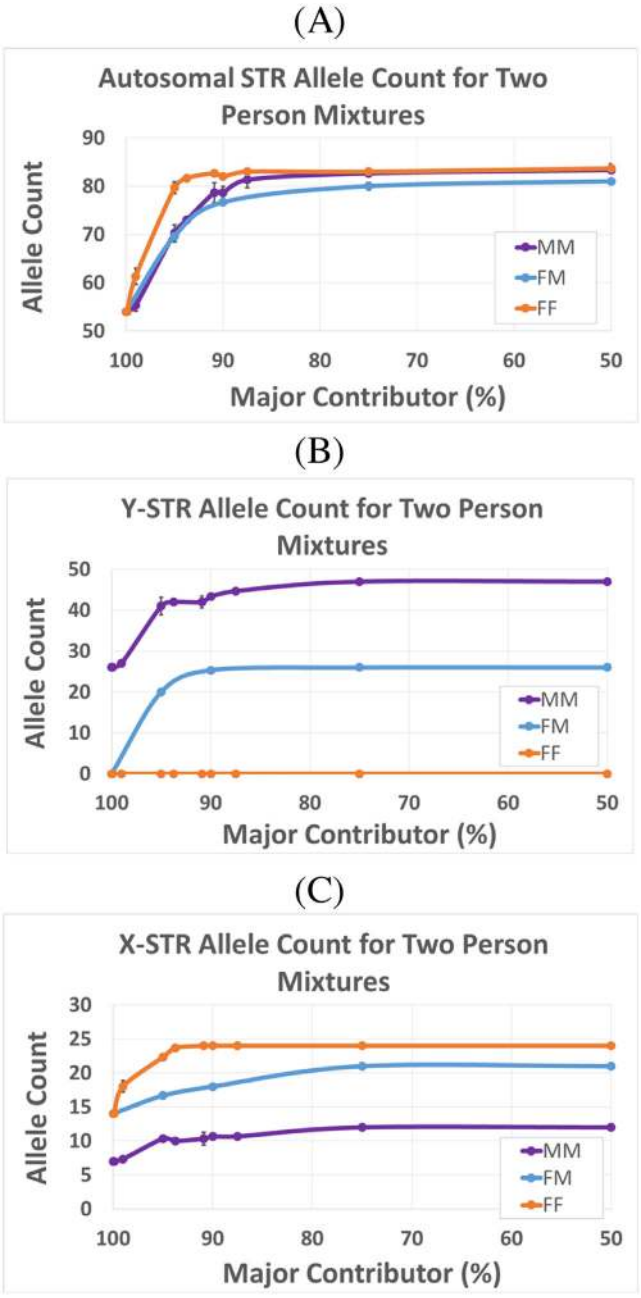

Fig. 6. DNA Mixture Study. Allele detection at 58 forensic STR loci. The percentage of the major contributor in each of three sets of DNA mixtures was plotted relative to the number of STR alleles detected above the default analytical threshold of the ForenSeq™ Universal Analysis Software. Single source samples (i.e., major contributor = 100%; minor contributor = 0%) are shown at the (x,y) origin of each plot where there are 27 total autosomal STR loci (A), 24 Y-STR total loci (B), and 7 X-STR loci (C). Male:male (MM), female:male (FM), and female:female (FF) mixtures exhibited similar increasing trends in the number of alleles called relative to increasing percentage of minor contributors. Data are plotted as average allele counts and standard deviations for each mixture sample ratio, tested in triplicate.

Addition of a minor, male contributor, in either the MM or FM mixtures, also led to an increase in X-STR allele detection such that at 5%, 10%, and 25% contributions 10, 11, 12 and 17, 18, 21 X-STR alleles were called in the MM and FM mixtures, respectively (Fig. 6C).

ForenSeq™ iiSNPs were selected for inclusion in the multiplex based on high heterozygosities and low inbreeding (Fst) coefficients, as required to approach similar random match probabilities as compared to core forensic STR loci [69,70]. Biogeographical

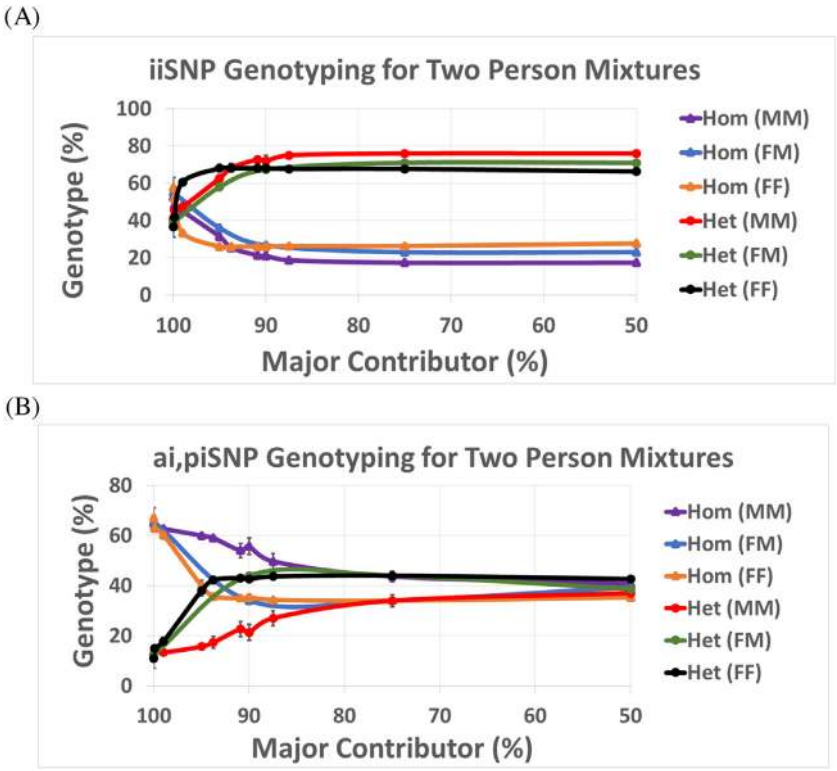

Fig. 7. DNA Mixture Study: Allele detection at 172 forensic SNP loci (homozygous and heterozygous) in Two Person DNA Mixtures. The percentage of the major contributor in each of three sets of DNA mixtures was plotted relative to the total percentage of homozygote (hom, AA or BB) and heterozygote (het, AB) genotypes for the 94 iiSNPs (plot A), and for the 78 combined aiSNPs and piSNPs (plot B). Single source samples (i.e., major contributor = 100%; minor contributor = 0%) are shown at the (x,y) origin of each plot, and illustrate that the ratio of homozygous to heterozygous iiSNPs is approximately equal (58.6%/41.4%) in the male–male (MM), female–male (FM), and female–female (FF) mixtures, and diverge with increasing amounts of minor iiSNP allele contribution. Inversely, the relationship between homozygous and heterozygous aiSNPs and piSNPs is approximately 84.6% and 15.4% in single source samples and converge with increasing minor contributor allele presence. Data are plotted as the average homozygote and heterozygote genotype percentages and standard deviations for each mixture sample ratio, tested in triplicate.

ancestry SNPs (aiSNPs), conversely, possess low heterozygosities and high Fst values [71,80]. Fig. 7 illustrates the proportions of biallelic SNP homozygotes (hom) (*i.e.*, AA or BB) and heterozygotes (het) (*i.e.*, AB), and how those proportions change in the MM, FM, and FF mixtures. Aforementioned properties of the SNP classes are evident in Fig. 7 where for single source samples (100% major contributor) 41.4% of the 94 iiSNP loci were heterozygotic (Fig. 7A), and only 15.4% of the 78 aiSNPs and piSNPs were heterozygotic (Fig. 7B). The homozygote and heterozygote SNP genotype percentages varied for each SNP category as minor contributor DNA was added and the mixture proportion increased. When the minor contributor was present at 5% (50 pg DNA of the total input) the iiSNPs with heterozygous genotypes increased to 66.9% (Fig. 6A). When the minor contributor proportion was $\geq$10%, iiSNP homozygote and heterozygote genotypes plateaued at 25% and 75%, respectively. When the minor contributor was present at 5%, the percentage of aiSNP and piSNP homozygotes and heterozygotes were 65.3% and 34.7%, respectively. When the minor contributor was present at $\geq$10%, then the homozygote and heterozygote genotypes were 52.0% and 48.0%, respectively (Fig. 7B).

The numbers of unique, minor contributor STR and SNP alleles detected in the ForenSeq™ data, above the default analytical threshold of the ForenSeq™ Universal Analysis Software for STRs and >2% of reads for SNPs, were determined for each of the mixed DNA samples. For loci in which an intra-STR sequence variant(s) was present, this information was used to designate unique, minor contributor alleles. 115 total unshared alleles were detected in the MM minor contributor, as follows, by locus type: autosomal STR (31 unique alleles), Y-STR (21 unique alleles), X-STR (4 unique alleles), iiSNP (31 unique alleles), ai,piSNP combined (28 unique alleles). 29, 26, 6, 29 and 29 unique minor contributor alleles were unshared with the major contributor in the FM mixtures for each locus type, respectively (119 total unique); 29, 0, 9, 26 and 29 unshared minor contributor alleles were present in the FF mixtures (93 total unique alleles), respectively. The numbers of unique minor contributor alleles were quantified, at the five categories of forensic loci, for mixture ratios ranging from 0.l% to 50% (Fig. 8). At the 1% minor contributor input amount, 4 and 35 unique alleles were detected in the MM and FF minor contributors, respectively. At the 5% minor contributor input 55, 68, and 86 unique alleles were detected from the minor component in the MM, FM, and FF mixtures, respectively. Similar data were generated in the other tested minor contributor percentages tested (Fig. 8). At 0.1%, $<$0.5% of the minor contributor alleles are detected above the default analytical threshold. At 1%, $\sim$4% (MM) to $\sim$40% (FF) of the minor alleles were detected. In the FF samples at 1%, almost all of the iiSNP differences between the two DNA contributors are detected and resolvable to each individual. At 5%, at least 48% and up 90% of the minor alleles were detected, and 97–100% were detected at 10% minor contributor.

Thirty-two mixture samples were run in each of these DPMB studies. Deeper coverage, and increased allele detection at lower DNA input amounts, can be achieved with NGS by including fewer samples per run, if desired (minimum of 8 samples/library preparation recommended to account for pipetting accuracy in preparing master mix).

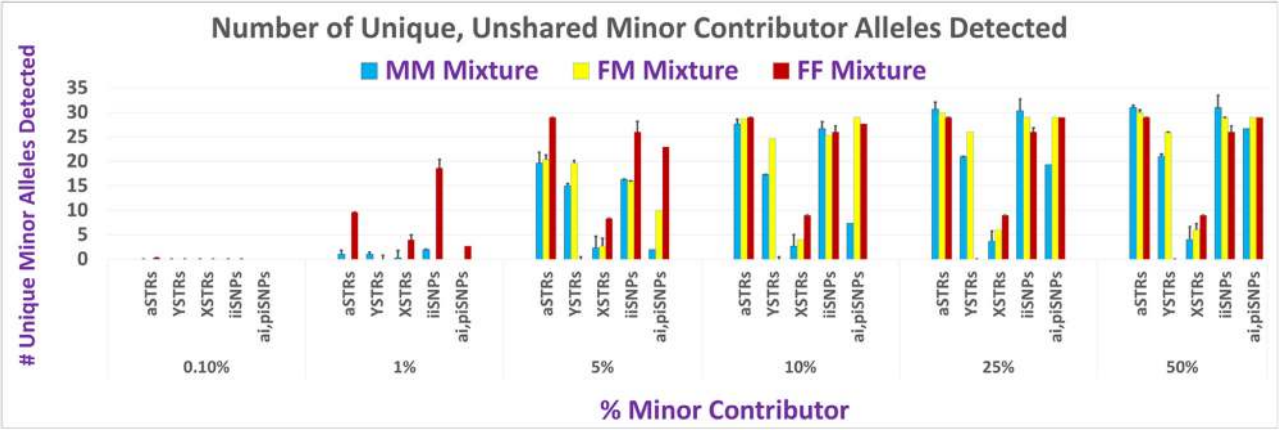

Fig. 8. DNA Mixture Study. Detection of unique, unshared minor contributor alleles in two person DNA mixtures. The ability of the MiSeq FGx™ Forensic Genomics System to detect minor contributor alleles that are foreign to the major contributor in male:male (MM), female:male (FM), and female:female (FF) mixtures was assessed using the ForenSeq™ DNA Signature Prep Kit and Universal Analysis Software. For FM samples, 0.1% and 1% minor ratios were not tested. The average number of unshared minor contributor alleles, above the default analytical threshold, and standard deviations, were plotted for autosomal, Y-, and X-STRs, iiSNPs, and ai-pi-SNPs at each of the mixture ratios tested, in triplicate.

3.6. Degraded DNA studies

Performance of the MiSeq FGx™ Forensic Genomics System was assessed with partially degraded gDNA that mimic forensic samples exposed to environmental and chemical stresses. Two gDNA samples (2800 M, NA12878) were partially fragmented using mechanical shearing. Aliquots of the sheared samples were also digested with DNase I endonuclease to nonspecifically cleave the DNA into smaller fragments. Degraded DNA samples were amplified with DPMB (231 loci) of the ForenSeq™ DNA Signature Prep Kit and processed according to the Illumina MiSeq FGx™ User Guide [77,78] in two phases. For one sample set (phase I), the 1 ng recommended DNA input of both "sheared only" and "sheared + digested" samples was added to PCR1. For the second sample set (phase II), the maximum input template volume (5 μl) of degraded DNA ("sheared only" and "sheared + digested") was added to PCR1. The physical impact of mechanical degradation and DNase I digestion for the two gDNA samples was assessed on an Agilent

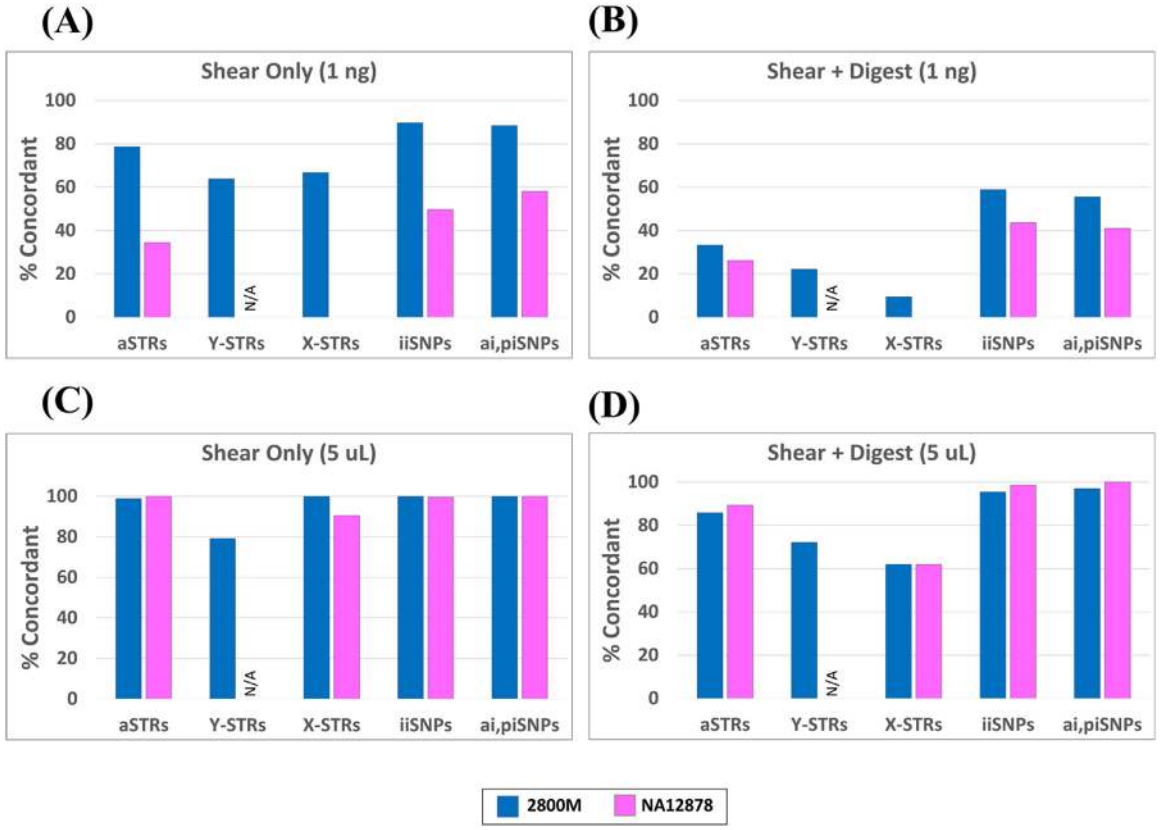

Fig. 9. DNA Degradation Study. STR and SNP alleles detected. Genomic DNA (2800 M (male), NA12878 (female)) was partially degraded and used as input for library preparation with the ForenSeq™ DNA Signature Prep Kit with 1 ng DNA input (A and B) and with the maximum input volume input of 5 μl (C and D). DNAs were partially degraded by mechanical shearing only ("phase I"; A and C) or in combination with 1U DNase I treatment ("phase II"; B and D). The percentage of typed alleles with read counts exceeding default interpretation thresholds were plotted for autosomal, Y-, and X-STRs, iiSNPs, and ai,piSNPs for the two DNA samples, tested in triplicate.

2100 BioAnalyzer. Phase I gDNA samples yielded "sheared only" and "sheared + digested" DNA ranging from 40 to 80 nt and 40–50 nt, respectively. Phase II fragments ranged from 30 to 300 nt and 30–200 nt for "sheared only" and "sheared + digested", respectively. As shown in Supplemental Fig. 1, the size distribution for the 1 ng and 5uL samples under the same treatment (shear only or shear + DNase) is similar. The size distribution of the peaks is quite broad (particularly for the sheared condition), and extends into the larger fragment sizes. For the 5uL inputs, ~40 x more DNA was added than the 1 ng samples, greatly increasing the number of molecules present in the reaction and increasing the probability that the longer fragments were amplified and detected. Thus, when analyzing partially degraded DNA, adding a larger input quantity may generate additional genotypes. ForenSeq™ DNA libraries of these partially degraded DNA samples were visualized after automated capillary electrophoresis on the Fragment Analyzer™ to assess quality and quantity prior to sequencing (Supplemental Fig. 2).

Data indicate, for the phase I input samples (1 ng) created by mechanical shearing alone (Fig. 9A), and those sheared and digested with DNase I (Fig. 9B), loss of amplifiable STRs (autosomal, Y- and X-) and SNPs (ii, ai and pi) for both 2800 M and NA12878 relative to amplicon length. In this study, sample NA12878 was more sensitive to DNA degradation than 2800 M. SNP loci were particularly robust after partial DNA degradation. Specifically, in the "sheared only" samples from 2800 M and NA12878, respectively, >10% more alleles were typed at iiSNP (90%, 50%) and at ai,piSNP (88%, 58%) loci, than at the autosomal STRs (79%, 35%). For the most degraded samples (from shearing and digestion) at 1 ng, >15% more SNP loci were typed as compared to autosomal STR loci.

In phase II, performance of the MiSeq FGx™ System on partially degraded DNA was assessed by addition of the maximum template volume (5 μl) accepted by PCR1, in an attempt to recover as much relevant genetic information as possible. Fig. 9C illustrates that from sheared only samples >99%, 79%, and >90% of the autosomal, Y-, and X-STRs were genotyped, respectively, and that 100% of the iiSNPs and ai,piSNPs were recovered. When compared to the 1 ng phase I input results, more actionable data were recovered from the 5 μl input of the most compromised samples (sheared + digested). Correct, automated genotype results were generated for the autosomal, Y-, and X-STRs, iiSNPs and ai,piSNPs at >86%, 72%, 62%, >95% and >97%, respectively, of the total possible loci in a full profile (Fig. 9D). In one NA12878 sample with phase II treatment (sheared only, 5uL (~45 ng) template) at DXS10135, alleles 20 and 32 were detected, while in the other two replicates the genotype was (20,20). Genotype results are provided in Supplemental Table 6.

To examine the degradation results in relation to ForenSeq™ amplicon length, STRs were subdivided into three categories based on their average amplicon length: short STRs, medium STRs, and long STRs (Supplemental Table 7; Supplemental Fig. 3). Genotyping

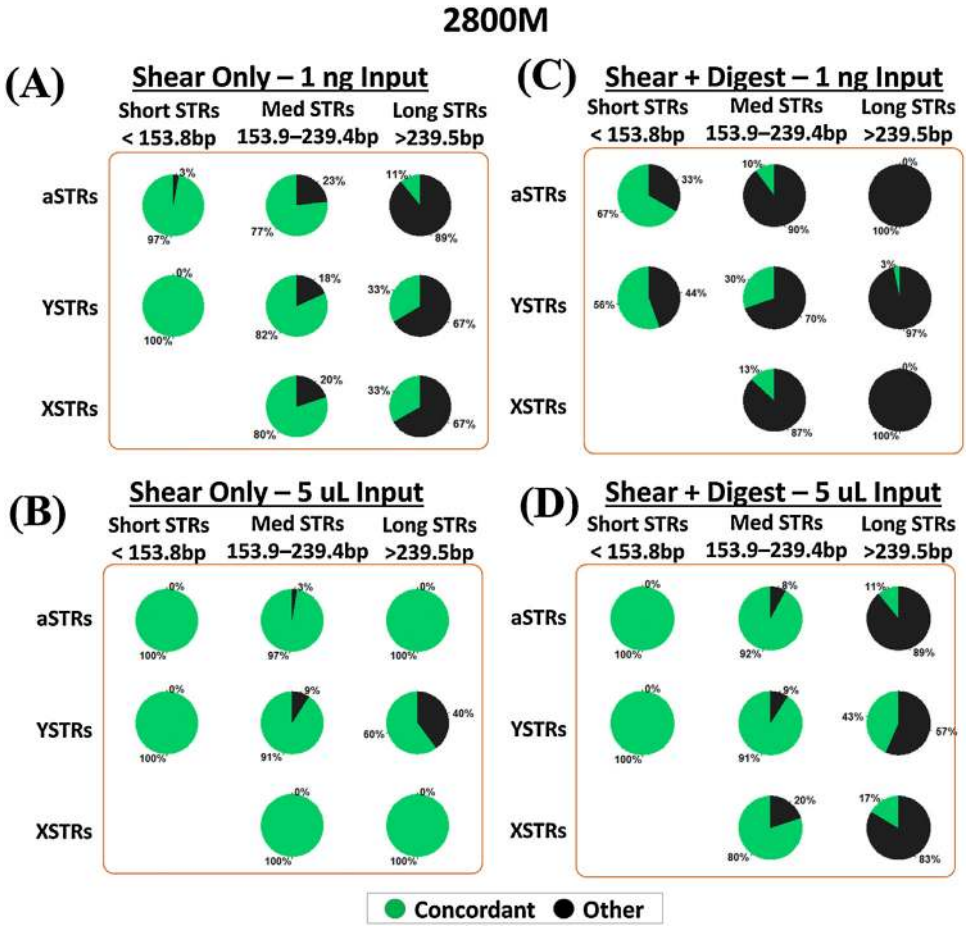

Fig. 10. DNA Degradation Study. Forensic STR genotyping results relative to amplicon length. Genomic DNA (2800 M) was sheared using the Covaris S2 instrument (Shear Only), and sheared plus the addition of a 1U DNase I (Shear + Digest). Shown are results for 1 ng phase I input amount (A and C) and phase II (5 μl) input amounts (B and D), for the sheared, and sheared + digested treatments, respectively. Semi-automatically generated genotypes from the ForenSeq™ Universal Analysis Software were compared to data from orthogonal methods (see Materials and methods), and the results plotted for autosomal STRs, Y-STRs, and X-STRs, relative to the maximum amplicon length of the STR marker type (i.e., short, med, long) as described in Supplemental Table 7. Plot was generated in R v3.2.0.

results for 2800 M and NA12878 were analyzed with respect to each STR size category and locus type (autosomal, Y-, and X-) for each degradation treatment ("Shear", "Shear + Digest") conducted in phases I (1 ng) and II (5 μl). Fig. 10 indicates a pattern where more data from the short and medium STRs was recovered, relative to the longer STRs for 2800 M (for NA12878 see Supplemental Fig. 4). Moreover, the increased, phase II input of partially degraded DNA allowed recovery of actionable data, across all STR sizes.

3.7. PCR inhibition studies

Effects of a set of five known PCR inhibitors on amplification efficiency of SNP and STR amplicons in the ForenSeq™ DNA Signature Prep Kit were characterized. The five inhibitors tested were hematin, humic acid, indigo dye, tannic acid and urban dust, each of which were independently spiked directly into PCR1 (containing DNA Primer Mix B), at the following concentrations: 16.66 μM, 33.3 μM, 66.66 μM and 133.3 μM. 2800 M allele calls were semi-automatically generated using the ForenSeq™ Universal Analysis Software (Supplemental Table 8).

As illustrated in Fig. 11, at the lowest tested concentration (16.66 μM) of each of the five PCR inhibitors, data indicated that a > 99% allele call rate was reached, across the 231 loci targeted by DPMB. Indigo dye, at each of the other three concentrations (33.33, 66.66, 133.3 μM), did not inhibit amplification; 100% genotyping was achieved at each concentration tested. Results were as follows for humic acid and urban dust: complete, 231 locus profiles were generated at 33.33 μM of both inhibitors; >98% call rates were produced at 66.66 μM where some alleles were not detected above the analytical threshold, and at 133.33 μM both of these inhibitors caused nearly complete inhibition of allelic amplification (Fig. 11). In contrast, hematin and tannic acid impacted amplification at lower concentrations: at 33.33 μM, a decrease in call rate was observed where ~25% and ~67% of loci were no longer detected, respectively. At the highest tested concentrations of hematin and tannic acid (66.66 and 133.3 μM), target amplification was completely inhibited.

3.8. Accuracy, precision, and call rate

Data generated on the MiSeq FGx™ instrument in repeatability and reproducibility studies using both DPMA (152 loci) and DPMB (230 loci) of the ForenSeq™ DNA Signature Prep Kit, and analyzed using the ForenSeq™ Universal Analysis Software, were used to calculate genotyping accuracy and precision. The autosomal STR locus D22S1045 was not included, as the primer set was still in development at the time of these studies. The repeatability study was defined as a single analyst preparing five identical, full 96-well plates with 96 samples amplified with DPMA, and five identical 96-well plates with 32 samples amplified with DPMB. Reproducibility studies were defined as five analysts each preparing one full 96-well plate of samples amplified with DPMA, and each preparing one additional 96-well plate of 32 samples amplified with DPMB. Data from these 20 MiSeq FGx™ instrument runs were analyzed with the ForenSeq™ Universal Analysis Software and subjected to manual review of QC indicators. All of the sequencing runs passed the QC metrics for run quality; not all samples were above the 85,000 read general guideline (Supplementary Table 10). All non-NTC samples were included in analysis and interpretation, regardless of read depth. Resulting semi-automated allele calls for 2800 M replicates (DPMA n = 950; DPMB n = 310) were considered as concordant when they were the same as those produced with orthogonal methods (*i.e.*, fragment length based CE of STRs or SNP based bead array typing). Accuracy was determined by concordance analysis of alleles between the orthogonal CE or SNP genotype and haplotype data and the ForenSeq™ DNA Signature Prep Kit, excluding "Below AT" outcomes; those are captured in call rates. Ambiguous genotypes were handled as follows: if one allele at a locus was called, and was concordant with a called allele from an orthogonal method, and one allele was ambiguous, then only one allele was concordant. For example, genotype (12,13) compared to genotype (12,*) was measured as concordant at the 12 allele and discordant at the 13 allele). Precision was calculated by determining the most frequently observed genotype or haplotype at each locus and totaling the percentage that genotype was observed over all replicates, including "Below AT" outcomes. Call Rate was defined as the percentage of loci that resulted in a genotype or haplotype (*i.e.* the percentage of loci that didn't result in a "Below AT" result) (Supplemental Fig. 5).

Accuracy and precision statistics are shown for ForenSeq™ autosomal STRs, Y-STRs, X-STRs, iiSNPs, and for ai,piSNPs, targeted by DNA Primer Mixes A and B, respectively (Table 2 and Supplemental Table 9). The accuracy calculations showed that the 58 autosomal, Y- and X-STR genotypes amplified with DPMA and DPMB PCR1 were 100.00% concordant across all STR loci. The

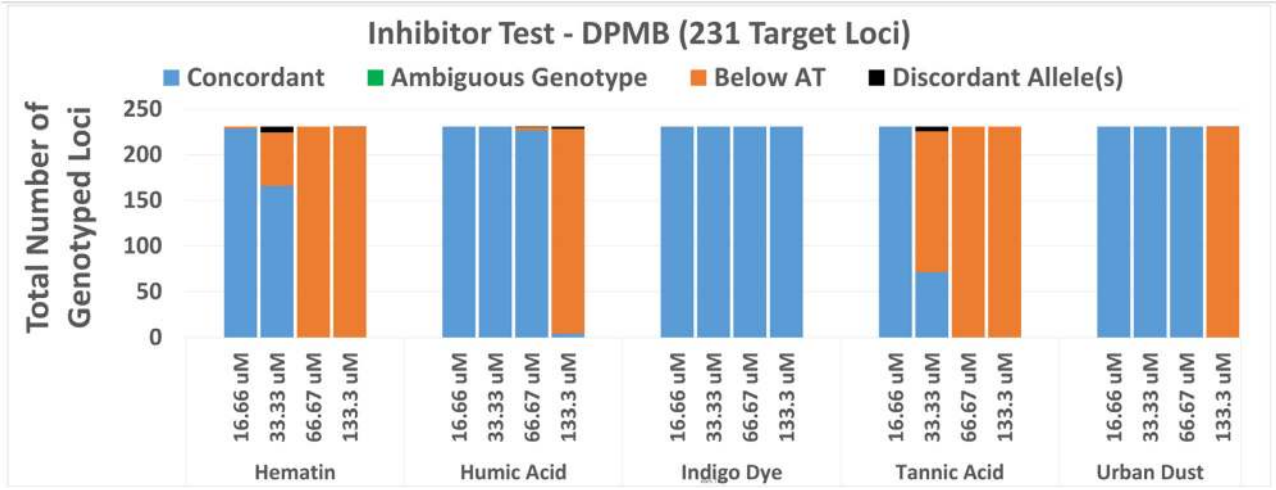

Fig. 11. Genotyping Results from PCR Inhibitor Testing. Five known PCR inhibitors were spiked into PCR1 of the ForenSeq™ Kit, with DNA Primer Mix B. Genotyping results achieved in the presence of varying concentrations of hematin, humic acid, indigo dye, tannic acid and urban dust were plotted vs. triplicate 2800 M control sample genotypes semi-automatically generated using the ForenSeq™ Universal Analysis Software. Genotype results were categorized as "Concordant" (blue), "Ambiguous Genotype" (green), "Below AT" (orange), and "Discordant Allele(s)" (black) (see Materials and methods) at the 231 loci targeted by DPMB.

Table 2
Accuracy and Precision Statistics for the MiSeq FGx™ Forensic Genomics System.

	DPMA			
	aSTRs (%)	Y-STRs (%)	X-STRs (%)	iiSNPs (%)
Accuracy	100.00	100.00	100.00	99.16
Precision	99.89	99.86	99.04	97.84
Call Rate	99.89	99.86	99.04	99.51

	DPMB				
	aSTRs (%)	Y-STRs (%)	X-STRs (%)	iiSNPs (%)	ai,piSNPs (%)
Accuracy	100.00	100.00	100.00	99.73	100.00
Precision	100.00	99.74	100.00	99.45	100.00
Call Rate	100.00	99.74	100.00	100.00	100.00

total 94 iiSNP genotypes targeted by DPMA and DPMB were ≥99.16% concordant and complete. Genotypes at the additional 78 ai,piSNP loci targeted by DPMB were 100.00% concordant and complete.

Further analyses were conducted to investigate impacts of read number per sample, relative to multilocus profile completeness (using default thresholds), as well as to no template controls (NTCs), from the 950 samples amplified using DPMA in the repeatability and reproducibility studies. Numbers of STR loci and iiSNP loci with "Ambiguous Genotype", "Below AT", or "Discordant Allele(s)" were plotted relative to total reads per sample (Fig. 12 and Supplemental Table 10). Data ranged from 100% completeness and concordance to one sample where none of the 58 STRs nor Amelogenin was detected above the default interpretation threshold (Fig. 12A), and to 134 samples where four or more iiSNPs were not called (Fig. 12B). Fig. 11 indicates that as total reads/sample decreases below approximately 85,000 reads (horizontal blue line) there can be an associated reduction in allele calling when default ForenSeq™ Universal Analysis threshold(s) are used. Thus, a general 85 K read guideline may provide a quality control run metric. Samples with <85 K reads may produce complete profiles. If some alleles are not detected (ND), then sufficient data (read counts) may be generated for interpretation of the partial profile.

The twenty NTC samples from repeatability and reproducibility studies (one NTC per sequencing run) were assessed. Ten of the NTCs were amplified with DPMA, and ten with DPMB. Of these, eight NTCs amplified with DPMA and seven NTCs amplified with DPMB were 100% blank (zero reads) for all 152 and 230 loci, respectively. The number of reads above the default AT, ranging from 11 to 160 reads, were detected in the remaining five NTCs at no more than four loci per NTC (Supplemental Table 9).

4. Discussion

Next-Generation Sequencing (NGS), also known as Massively Parallel Sequencing (MPS), allows for targeted forensic NGS and genotyping based on the fundamental nucleotide sequence of STRs and SNPs. The MiSeq FGx™ Forensic Genomics System enables simultaneous PCR amplification and sequencing of STR and SNP loci in a single reaction, with one workflow and analysis tool, for operational efficiency and maximum information potential. Biogeographical ancestry and phenotypic informative SNPs can provide investigative lead information [71–73,80–85]. While sequencing by synthesis (SBS) overcomes some limitations of fragment-length detection by CE, and expands forensic capabilities, there are considerations for MiSeq FGx™ System implementation. Price of the instrument is similar s those of capillary electrophoresis platforms. Price per sample depends on the number of samples/run, and ranges from approximately $50-80 per sample for ForenSeq™ library preparation and MiSeq FGx™ sequencing. ForenSeq™ DNA library preparation, PCR-based and comprised of common molecular biology methods, requires more preparation time than a one-step PCR. Hands-on time for ForenSeq™ PCR1 is estimated at 15 min and 3 h 35 min of thermalcycling, and for PCR2 at 10 min of hands on time and 90 min of thermalcycling. Library purification and normalization using Sample Purification Beads (SPB) require approximately 45 min of hands-on time, and a total of one hour and fifty minutes. Average run time on the MiSeq FGx™ instrument for the protocol described herein is approximately 27 h and 14 min (range 25 h 39 m – 29 h 35m). Data analysis of autosomal, X-, Y-STRs and ii,pi, aiSNPs using the ForenSeq™ Universal Analysis Software typically required 34 min 49 s (range 10 m – 1 h), for a total sequencing and data analysis time of approximately 28 h.

Data reported here from SWGDAM developmental validation studies of the MiSeq FGx™ System demonstrates robust, reliable, reproducible and semi-automated allele calling that meets established forensic validation guidelines. Backward compatibility of allele calling with existing law enforcement STR databases is maintained. Robustness was assessed through analysis of a range of environmental and situational circumstances, including species testing, DNA sensitivity studies, two-person DNA mixture testing, stability analysis of partially degraded DNA samples and of

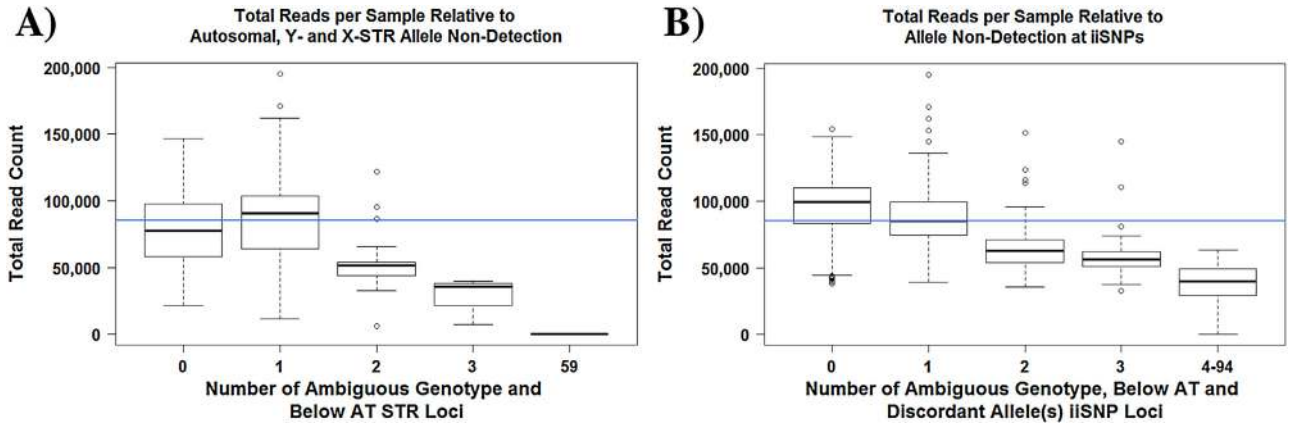

Fig. 12. Total Reads per Sample Relative to STR and iiSNP Genotype Concordance. The total numbers of ForenSeq™ reads per sample in DPMA (y-axis) are shown relative to binned outcomes of the number of "Ambiguous Genotype", "Below AT" or "Discordant Allele(s)" categories (see Materials and methods) for 2800 M (n = 950). 85,000 total reads (horizontal blue line) provides a general guideline for minimum read count needed to approach a full profile. Boxplots were generated using R v3.1.

reactions that included known PCR inhibitors. Across all studies, quality indicators in the ForenSeq™ Universal Analysis Software assisted analysts in evaluating information content in each run to verify semi-automated allele calls relative to default thresholds and filters. Individual laboratories may choose to modify these values and analysis parameters in enacting specific internal laboratory policies or in relation to particular questions and scenarios at hand.

Data used to assess reliability and reproducibility of the MiSeq FGx™ System allowed for measurements of accuracy and of precision in allele typing. The gDNA sample 2800 M was sequenced hundreds of times with >99% accuracy and precision in both STR and SNP genotype calling. A call rate of >99% was observed for the 152 and 230 loci targeted by DNA Primer Mix A (DPMA) and DNA Primer Mix B (DPMB), respectively. Conventional amplicon sizing by capillary electrophoresis can be inaccurate regarding actual length determination, but precise when length-based allele calls are made relative to ±0.5 bp bins around alleles of a physical allelic ladder [5]. This report offers a look at both the accuracy and the precision of higher resolution allele typing that is not based upon amplicon sizing.

The desired number of reads (depth of coverage) per locus determines the maximum number of ForenSeq™ DNA libraries processed simultaneously, in one run. The MiSeq FGx™ System was assessed with various numbers of libraries; data here are reported on 96 samples for high-quality, single source DNA samples, using DPMA, and 32 samples with DPMB. Under these conditions, the success rate for profiling lysates from various blood and buccal collection substrates was observed as follows: for DPMA >88% of samples yielded full profiles (153 loci), while 68% of samples with DPMB produced full profiles (231 loci). >92% of DPMB samples produced genotypes at >90% of loci (208 loci). Depending upon the depth of read coverage and level of profile completeness required to address the question at hand, individual laboratories may determine that more or less than these numbers of samples per run are appropriate. For example, if a lower limit of detection of minor component(s) in mixed DNA samples is desired, then <32 samples per run can provide deeper coverage per locus than studies presented here (and vice versa for higher LOD). The autosomal STR locus D22S1045 should be interpreted with caution as more imbalance in read counts may be seen between alleles of a heterogyzote than observed at other loci. Three SNP loci (rs10776839, rs7041158, rs6955448) displayed higher relative levels of intralocus imbalance than the majority of loci. The rapidly mutating Y loci demonstrate higher stutter percentages relative to the other Y loci, consistent with CE methods.

Results from the species specificity studies are not unexpected, as species cross reactivity has been reported for the GlobalFiler® Express kit, PowerPlex® ESI 16/17 Fast, PowerPlex® ESX 16/17 Fast Systems, the AmpFℓSTR® Yfiler® PCR Amplification kit, and several other forensic PCR multiplexes [10,86,87]. The MiSeq FGx™ System generated complete, accurate and reproducible genotypes from input DNA ranging from 62.5 pg to 1 ng in sensitivity studies (serial dilutions).

Some mixed samples may not be detected with CE systems due to limit of detection differences, equi-length alleles that differ by sequence, or minor sequence variant alleles that are masked at a stutter position of a parent allele. In some mixtures, CE data indicate a single source sample or a two-person mixture, where NGS detects alleles from three and upwards of five or more DNA contributors (David Ballard, personal communication). Since NGS detects and identifies intra-STR sequence variants [88–90], the total complement of alleles are provided, at the nucleotide level. Mixture studies described here demonstrated the ability of the MiSeq FGx™ System to detect shared, and unshared (unique, obligate) minor contributor alleles at less than 5% of the major donor. Initial considerations regarding an approach to nomenclature of refined STR alleles are available from the DNA Commission of the International Society of Forensic Genetics (ISFG) [91].

Recovery of actionable genetic information from partially degraded DNA samples is possible using the MiSeq FGx™ System because >50% of the 230 targeted amplicons are less than 205nt in length. Partial profiles from template input as low as 7 pg produced random match probabilities as rare, or rarer, than as those from the 13 CODIS core STR loci.

5. Conflict of interest

All authors are current employees of Illumina, Inc., and declare no other competing interests.

Acknowledgements

The authors thank White Labs for providing the *Saccharomyces cerevisiae* sample and Claudio Koser from the University of Cambridge, based at Addenbrooke's hospital, for the *Staphylococcus aureus*, *Neisseria gonorrhoeae* and *Bacillus subtillis* DNA samples used in the species specificity studies. A special thanks to Jon-Matthew Belton for his contributions to the development of the workflow diagram of the MiSeq FGx™ Forensic Genomics System, and to Meghan Didier and John Walsh for careful manuscript review.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.fsigen.2017.01.011.

References

[1] C.J. Frégeau, R.M. Fourney, DNA typing with fluorescently tagged short tandem repeats: a sensitive and accurate approach to human identification, Biotechniques 15 (1993) 100–119. (accessed 17.08.2015) http://www.ncbi.nlm.nih.gov/pubmed/8103347.

[2] H.A. Hammond, L. Jin, Y. Zhong, C.T. Caskey, R. Chakraborty, Evaluation of 13 short tandem repeat loci for use in personal identification applications, Am. J. Hum. Genet. 55 (1994) 175–189. (accessed 20.08.2015) http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1918216&tool=pmcentrez&rendertype=abstract.

[3] W. Wang, M. Fukuda, T. Kishida, Y. Tamaki, Quintuplex PCR-amplification of microsatellites, Nihon Hoigaku Zasshi. 50 (1996) 231–236. (accessed August 20, 2015) http://www.ncbi.nlm.nih.gov/pubmed/8810743.

[4] J.M. Wallin, C.L. Holt, K.D. Lazaruk, T.H. Nguyen, P.S. Walsh, Constructing universal multiplex PCR systems for comparative genotyping, J. Forensic Sci. 47 (2002) 52–65.

[5] K. Lazaruk, P.S. Walsh, F. Oaks, D. Gilbert, B.B. Rosenblum, S. Menchen, et al., Genotyping of forensic short tandem repeat (STR) systems based on sizing precision in a capillary electrophoresis instrument, Electrophoresis 19 (1998) 86–93, doi:http://dx.doi.org/10.1002/elps.1150190116.

[6] D.L. Deforce, R.E. Millecamps, D. Van Hoofstat, E.G. Van den Eeckhout, Comparison of slab gel electrophoresis and capillary electrophoresis for the detection of the fluorescently labeled polymerase chain reaction products of short tandem repeat fragments, J. Chromatogr. A 806 (1998) 149–155. (accessed August 20.08.2015) http://www.ncbi.nlm.nih.gov/pubmed/9639885.

[7] T.R. Moretti, A.L. Baumstark, D.A. Defenbaugh, K.M. Keys, A.L. Brown, B. Budowle, Validation of STR typing by capillary electrophoresis, J. Forensic Sci. 46 (2001) 661–676. (accessed 20.08.2015) http://www.ncbi.nlm.nih.gov/pubmed/11373005.

[8] J.M. Butler, C.R. Hill, Biology and genetics of new autosomal str loci useful for forensic DNA analysis, Forensic Sci. Rev. 24 (2012) 16–26.

[9] P. Martín, L.F. de Simón, G. Luque, M.J. Farfán, A. Alonso, Improving DNA data exchange: validation studies on a single 6 dye STR kit with 24 loci, Forensic Sci. Int. Genet. 13 (2014) 68–78, doi:http://dx.doi.org/10.1016/j.fsigen.2014.07.002.

[10] L.K. Hennessy, N. Mehendale, K. Chear, S. Jovanovich, S. Williams, C. Park, et al., Developmental validation of the GlobalFiler(®) express kit, a 24-marker STR assay, on the RapidHIT(®) System, Forensic Sci. Int. Genet. 13 (2014) 247–258, doi:http://dx.doi.org/10.1016/j.fsigen.2014.08.011.

[11] D.Y. Wang, S. Gopinath, R.E. Lagacé, W. Norona, L.K. Hennessy, M.L. Short, et al., Developmental validation of the GlobalFiler(®) Express PCR Amplification Kit: a 6-dye multiplex assay for the direct amplification of reference samples,

Forensic Sci. Int. Genet. 19 (2015) 148–155, doi:http://dx.doi.org/10.1016/j.fsigen.2015.07.013.
[12] K. Oostdik, K. Lenz, J. Nye, K. Schelling, D. Yet, S. Bruski, 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, doi:http://dx.doi.org/10.1016/j.fsigen.2014.04.013.
[13] M. Barbisin, J.G. Shewale, Assessment of DNA extracted from forensic samples prior to genotyping, Forensic Sci. Rev. 22 (2010) 199–214. (accessed 20.08.2015) http://www.ncbi.nlm.nih.gov/pubmed/26242597.
[14] M. Turnbough, A. Eisenberg, L.L. Shade, J.G. Shewale, Training of forensic DNA scientists – a commentary, Forensic Sci. Rev. 24 (2012) 143–150. (accessed 20.08.2015) http://www.ncbi.nlm.nih.gov/pubmed/26244269.
[15] R.A. van Oorschot, K.N. Ballantyne, R.J. Mitchell, Forensic trace DNA: a review, Investig. Genet. 1 (2010) 14, doi:http://dx.doi.org/10.1186/2041-2223-1-14.
[16] J.P. Whitaker, T.M. Clayton, A.J. Urquhart, E.S. Millican, T.J. Downes, C.P. Kimpton, et al., Short tandem repeat typing of bodies from a mass disaster: high success rate and characteristic amplification patterns in highly degraded samples, Biotechniques 18 (1995) 670–677. (accessed 20.08.2015) http://www.ncbi.nlm.nih.gov/pubmed/7598902.
[17] B. Budowle, F.R. Bieber, A.J. Eisenberg, Forensic aspects of mass disasters: strategic considerations for DNA-based human identification, Leg. Med. (Tokyo) 7 (2005) 230–243, doi:http://dx.doi.org/10.1016/j.legalmed.2005.01.001.
[18] S.R. Hughes-Stamm, K.J. Ashton, A. van Daal, Assessment of DNA degradation and the genotyping success of highly degraded samples, Int. J. Legal Med. 125 (2011) 341–348, doi:http://dx.doi.org/10.1007/s00414-010-0455-3.
[19] K. Lazaruk, J. Wallin, C. Holt, T. Nguyen, P.S. Walsh, Sequence variation in humans and other primates at six short tandem repeat loci used in forensic identity testing, Forensic Sci. Int. 119 (2001) 1–10, doi:http://dx.doi.org/10.1016/S0379-0738(00)00388-1.
[20] J.M. Butler, Advanced Topics in Forensic DNA Typing: Methodology, Elsevier Inc., 2012. http://www.scopus.com/inward/record.url?eid=2-s2.0-84882202563&partnerID=tZOtx3y1.
[21] C. Phillips, L. Fernandez-Formoso, M. Gelabert-Besada, M. Garcia-Magariños, C. Santos, M. Fondevila, et al., Development of a novel forensic STR multiplex for ancestry analysis and extended identity testing, Electrophoresis 34 (2013) 1151–1162, doi:http://dx.doi.org/10.1002/elps.201200621.
[22] M.V. Lareu, S. Barral, A. Salas, C. Pestoni, A. Carracedo, Sequence variation of a hypervariable short tandem repeat at the D1S1656 locus, Int. J. Legal Med. 111 (1998) 244–247. (accessed 20.08.2015) http://www.ncbi.nlm.nih.gov/pubmed/9728750.
[23] C. Phillips, L. Fernandez-Formoso, M. Garcia-Magariños, L. Porras, T. Tvedebrink, J. Amigo, et al., Analysis of global variability in 15 established and 5 new European Standard Set (ESS) STRs using the CEPH human genome diversity panel, Forensic Sci. Int. Genet. 5 (2011) 155–169, doi:http://dx.doi.org/10.1016/j.fsigen.2010.02.003.
[24] FBI DNA Advisory Board Quality Assurance Standards for Forensic DNA Testing Laboratories, Forensic Sci. Commun. [Online]. (1998). https://www.fbi.gov/about-us/lab/biometric-analysis/codis/qas_testlabs.
[25] FBI DNA Advisory Board Quality Assurance Standards for Convicted Offender DNA Databasing Laboratories, Forensic Sci. Commun. [Online]. (1999).
[26] FBI DNA Advisory Board Quality Assurance Standards for DNA Databasing Laboratories, Forensic Sci. Commun. [Online]. (2011).
[27] M.G. Brevnov, H.S. Pawar, J. Mundt, L.M. Calandro, M.R. Furtado, J.G. Shewale, Developmental validation of the PrepFiler Forensic DNA Extraction Kit for extraction of genomic DNA from biological samples, J. Forensic Sci. 54 (2009) 599–607, doi:http://dx.doi.org/10.1111/j.1556-4029.2009.01013.x.
[28] S.A. Montpetit, I.T. Fitch, P.T. O'Donnell, A simple automated instrument for DNA extraction in forensic casework, J. Forensic Sci. 50 (2005) 555–563. (accessed 24.08.2015) http://www.ncbi.nlm.nih.gov/pubmed/15932086.
[29] W.R. Hudlow, M.D. Chong, K.L. Swango, M.D. Timken, M.R. Buoncristiani, A quadruplex real-time qPCR assay for the simultaneous assessment of total human DNA, human male DNA degradation and the presence of PCR inhibitors in forensic samples: a diagnostic tool for STR typing, Forensic Sci. Int. Genet. 2 (2008) 108–125, doi:http://dx.doi.org/10.1016/j.fsigen.2007.09.001.
[30] B.E. Krenke, N. Nassif, C.J. Sprecher, C. Knox, M. Schwandt, D.R. Storts, Developmental validation of a real-time PCR assay for the simultaneous quantification of total human and male DNA, Forensic Sci. Int. Genet. 3 (2008) 14–21, doi:http://dx.doi.org/10.1016/j.fsigen.2008.07.004.
[31] M. Barbisin, R. Fang, C.E. O'shea, L.M. Calandro, M.R. Furtado, J.G. Shewale, Developmental validation of the Quantifiler Duo DNA Quantification kit for simultaneous quantification of total human and human male DNA and detection of PCR inhibitors in biological samples, J. Forensic Sci. 54 (2009) 305–319, doi:http://dx.doi.org/10.1111/j.1556-4029.2008.00951.x.
[32] J.J. Mulero, C.W. Chang, R.E. Lagacé, D.Y. Wang, J.L. Bas, T.P. McMahon, et al., Development and validation of the AmpFlSTR MiniFiler PCR Amplification Kit: a MiniSTR multiplex for the analysis of degraded and/or PCR inhibited DNA, J. Forensic Sci. 53 (2008) 838–852, doi:http://dx.doi.org/10.1111/j.1556-4029.2008.00760.x.
[33] D.G. Hert, C.P. Fredlake, A.E. Barron, Advantages and limitations of next-generation sequencing technologies: a comparison of electrophoresis and non-electrophoresis methods, Electrophoresis 29 (2008) 4618–4626, doi:http://dx.doi.org/10.1002/elps.200800456.
[34] E.S. Lander, L.M. Linton, B. Birren, C. Nusbaum, M.C. Zody, J. Baldwin, et al., Initial sequencing and analysis of the human genome, Nature 409 (2001) 860–921, doi:http://dx.doi.org/10.1038/35057062.
[35] J.C. Venter, M.D. Adams, E.W. Myers, P.W. Li, R.J. Mural, G.G. Sutton, et al., The sequence of the human genome, Science 291 (2001) 1304–1351, doi:http://dx.doi.org/10.1126/science.1058040.
[36] D.R. Bentley, S. Balasubramanian, H.P. Swerdlow, G.P. Smith, J. Milton, C.G. Brown, et al., Accurate whole human genome sequencing using reversible terminator chemistry, Nature 456 (2008) 53–59, doi:http://dx.doi.org/10.1038/nature07517.
[37] T.L. Meiring, A.T. Salimo, B. Coetzee, H.J. Maree, J. Moodley, I.I. Hitzeroth, et al., Next-generation sequencing of cervical DNA detects human papillomavirus types not detected by commercial kits, Virol. J. 9 (2012) 164, doi:http://dx.doi.org/10.1186/1743-422X-9-164.
[38] N. Kothari, M.J. Schell, J.K. Teer, T. Yeatman, D. Shibata, R. Kim, Comparison of KRAS mutation analysis of colorectal cancer samples by standard testing and next-generation sequencing, J. Clin. Pathol. 67 (2014) 764–767, doi:http://dx.doi.org/10.1136/jclinpath-2014-202405.
[39] J. Qin, R. Li, J. Raes, M. Arumugam, K.S. Burgdorf, C. Manichanh, et al., A human gut microbial gene catalogue established by metagenomic sequencing, Nature 464 (2010) 59–65, doi:http://dx.doi.org/10.1038/nature08821.
[40] F. Broecker, M. Kube, J. Klumpp, M. Schuppler, L. Biedermann, J. Hecht, et al., Analysis of the intestinal microbiome of a recovered Clostridium difficile patient after fecal transplantation, Digestion 88 (2013) 243–251, doi:http://dx.doi.org/10.1159/000355955.
[41] B. Liu, L.L. Faller, N. Klitgord, V. Mazumdar, M. Ghodsi, D.D. Sommer, et al., Deep sequencing of the oral microbiome reveals signatures of periodontal disease, PLoS One 7 (2012) e37919, doi:http://dx.doi.org/10.1371/journal.pone.0037919.
[42] Y. Kawahara, M. de la Bastide, J.P. Hamilton, H. Kanamori, W.R. McCombie, S. Ouyang, et al., Improvement of the Oryza sativa Nipponbare reference genome using next generation sequence and optical map data, Rice (N.Y.) 6 (2013) 4, doi:http://dx.doi.org/10.1186/1939-8433-6-4.
[43] H. Sakai, S.S. Lee, T. Tanaka, H. Numa, J. Kim, Y. Kawahara, et al., Rice Annotation Project Database (RAP-DB): an integrative and interactive database for rice genomics, Plant Cell Physiol. 54 (2013) e6, doi:http://dx.doi.org/10.1093/pcp/pcs183.
[44] L. Pihlstrøm, A. Rengmark, K.A. Bjørnarå, M. Toft, Effective variant detection by targeted deep sequencing of DNA pools: an example from Parkinson's disease, Ann. Hum. Genet. 78 (2014) 243–252, doi:http://dx.doi.org/10.1111/ahg.12060.
[45] Y. Wang, Y. Yang, J. Liu, X.-C. Chen, X. Liu, C.-Z. Wang, et al., Whole dystrophin gene analysis by next-generation sequencing: a comprehensive genetic diagnosis of Duchenne and Becker muscular dystrophy, Mol. Genet. Genomics. 289 (2014) 1013–1021, doi:http://dx.doi.org/10.1007/s00438-014-0847-z.
[46] H.C. Martin, G.E. Kim, A.T. Pagnamenta, Y. Murakami, G.L. Carvill, E. Meyer, et al., Clinical whole-genome sequencing in severe early-onset epilepsy reveals new genes and improves molecular diagnosis, Hum. Mol. Genet. 23 (2014) 3200–3211, doi:http://dx.doi.org/10.1093/hmg/ddu030.
[47] J.M. Rothberg, W. Hinz, T.M. Rearick, J. Schultz, W. Mileski, M. Davey, et al., An integrated semiconductor device enabling non-optical genome sequencing, Nature 475 (2011) 348–352, doi:http://dx.doi.org/10.1038/nature10242.
[48] S.J. Salipante, T. Kawashima, C. Rosenthal, D.R. Hoogestraat, L.A. Cummings, D.J. Sengupta, et al., Performance comparison of Illumina and Ion Torrent next-generation sequencing platforms for 16S rRNA-based bacterial community profiling, Appl. Environ. Microbiol. 80 (2014) 7583–7591, doi:http://dx.doi.org/10.1128/AEM.02206-14.
[49] L.M. Bragg, G. Stone, M.K. Butler, P. Hugenholtz, G.W. Tyson, Shining a light on dark sequencing: characterising errors in ion torrent PGM data, PLoS Comput. Biol. 9 (2013), doi:http://dx.doi.org/10.1371/journal.pcbi.1003031.
[50] M. Quail, M.E. Smith, P. Coupland, T.D. Otto, S.R. Harris, T.R. Connor, et al., A tale of three next generation sequencing platforms: comparison of Ion torrent, pacific biosciences and illumina MiSeq sequencers, BMC Genomics. 13 (2012) 1, doi:http://dx.doi.org/10.1186/1471-2164-13-341.
[51] N.J. Loman, R.V. Misra, T.J. Dallman, C. Constantinidou, S.E. Gharbia, J. Wain, et al., Performance comparison of benchtop high-throughput sequencing platforms, Nat. Biotechnol. 30 (2012) 434–439, doi:http://dx.doi.org/10.1038/nbt.2198.
[52] M.M. Holland, M.R. McQuillan, K.A. O'Hanlon, Second generation sequencing allows for mtDNA mixture deconvolution and high resolution detection of heteroplasmy, Croat. Med. J. 52 (2011) 299–313, doi:http://dx.doi.org/10.3325/cmj.2011.52.299.
[53] J. Irwin, R. Just, M. Scheible, O. Loreille, Assessing the potential of next generation sequencing technologies for missing persons identification efforts, Forensic Sci. Int. Genet. Suppl. Ser. 3 (2011) e447–e448, doi:http://dx.doi.org/10.1016/j.fsigss.2011.09.085.
[54] J.A. McElhoe, M.M. Holland, K.D. Makova, M.S.W. Su, I.M. Paul, C.H. Baker, et al., Development and assessment of an optimized next-generation DNA sequencing approach for the mtgenome using the Illumina MiSeq, Forensic Sci. Int. Genet. 13 (2014) 20–29, doi:http://dx.doi.org/10.1016/j.fsigen.2014.05.007.
[55] R.S. Just, J.A. Irwin, W. Parson, Mitochondrial DNA heteroplasmy in the emerging field of massively parallel sequencing, Forensic Sci. Int. Genet. 18 (2015) 131–139, doi:http://dx.doi.org/10.1016/j.fsigen.2015.05.003.
[56] A.M. Divne, H. Edlund, M. Allen, Forensic analysis of autosomal STR markers using Pyrosequencing, Forensic Sci. Int. Genet. 4 (2010) 122–129, doi:http://dx.doi.org/10.1016/j.fsigen.2009.07.004.

[57] S.L. Fordyce, M.C. Ávila-Arcos, E. Rockenbauer, C. Børsting, R. Frank-Hansen, F.T. Petersen, et al., High-throughput sequencing of core STR loci for forensic genetic investigations using the Roche Genome Sequencer FLX platform, Biotechniques 51 (2011) 127–133, doi:http://dx.doi.org/10.2144/000113721.
[58] D.M. Bornman, M.E. Hester, J.M. Schuetter, M.D. Kasoji, A. Minard-Smith, C.A. Barden, et al., Short-read, high-throughput sequencing technology for STR genotyping, Biotech. Rapid Dispatches (2012), doi:http://dx.doi.org/10.2144/000113857 Apr: 1–6.
[59] C. Van Neste, F. Van Nieuwerburgh, D. Van Hoofstat, D. Deforce, Forensic STR analysis using massive parallel sequencing, Forensic Sci. Int. Genet. 6 (2012) 810–818, doi:http://dx.doi.org/10.1016/j.fsigen.2012.03.004.
[60] C. Van Neste, Y. Gansemans, D. De Coninck, D. Van Hoofstat, W. Van Criekinge, D. Deforce, et al., Forensic massively parallel sequencing data analysis tool: implementation of MyFLq as a standalone web- and illumina BaseSpace (®)-application, Forensic Sci. Int. Genet. 15 (2015) 2–7, doi:http://dx.doi.org/10.1016/j.fsigen.2014.10.006.
[61] D.H. Warshauer, D. Lin, K. Hari, R. Jain, C. Davis, B. Larue, et al., STRait Razor: a length-based forensic STR allele-calling tool for use with second generation sequencing data, Forensic Sci. Int. Genet. 7 (2013) 409–417, doi:http://dx.doi.org/10.1016/j.fsigen.2013.04.005.
[62] F. Guo, Y. Zhou, F. Liu, J. Yu, H. Song, H. Shen, et al., Evaluation of the early access STR kit v1 on the ion torrent PGM platform, Forensic Sci. Int. Genet. 23 (2016) 111–120, doi:http://dx.doi.org/10.1016/j.fsigen.2016.04.004.
[63] S.L. Fordyce, H.S. Mogensen, C. Børsting, R.E. Lagacé, C.W. Chang, N. Rajagopalan, et al., Second-generation sequencing of forensic STRs using the ion torrent HID STR 10-plex and the ion PGM, Forensic Sci. Int. Genet. 14 (2014) 132–140, doi:http://dx.doi.org/10.1016/j.fsigen.2014.09.020.
[64] J.D. Churchill, J. Chang, J. Ge, N. Rajagopalan, S.C. Wootton, C.-W. Chang, et al., Blind study evaluation illustrates utility of the Ion PGM™ system for use in human identity DNA typing, Croat. Med. J. 56 (2015) 218–229, doi:http://dx.doi.org/10.3325/cmj.2015.56.218.
[65] X. Zhao, K. Ma, H. Li, Y. Cao, W. Liu, H. Zhou, et al., Multiplex Y-STRs analysis using the ion torrent personal genome machine (PGM), Forensic Sci. Int. Genet. 19 (2015) 192–196, doi:http://dx.doi.org/10.1016/j.fsigen.2015.06.012.
[66] C. Børsting, N. Morling, Next generation sequencing and its applications in forensic genetics, Forensic Sci. Int. Genet. 18 (2015) 78–89, doi:http://dx.doi.org/10.1016/j.fsigen.2015.02.002.
[67] J.J. Sanchez, C. Phillips, C. Børsting, K. Balogh, M. Bogus, M. Fondevila, et al., A multiplex assay with 52 single nucleotide polymorphisms for human identification, Electrophoresis 27 (2006) 1713–1724, doi:http://dx.doi.org/10.1002/elps.200500671.
[68] L.A. Dixon, C.M. Murray, E.J. Archer, A.E. Dobbins, P. Koumi, P. Gill, Validation of a 21-locus autosomal SNP multiplex for forensic identification purposes, Forensic Sci. Int. 154 (2005) 62–77, doi:http://dx.doi.org/10.1016/j.forsciint.2004.12.011.
[69] A.J. Pakstis, W.C. Speed, R. Fang, F.C.L. Hyland, M.R. Furtado, J.R. Kidd, et al., SNPs for a universal individual identification panel, Hum. Genet. 127 (2010) 315–324, doi:http://dx.doi.org/10.1007/s00439-009-0771-1.
[70] K.K. Kidd, J.R. Kidd, W.C. Speed, R. Fang, M.R. Furtado, F.C.L. Hyland, et al., Expanding data and resources for forensic use of SNPs in individual identification, Forensic Sci. Int. Genet. 6 (2012) 646–652, doi:http://dx.doi.org/10.1016/j.fsigen.2012.02.012.
[71] K.K. Kidd, W.C. Speed, A.J. Pakstis, M.R. Furtado, R. Fang, A. Madbouly, et al., Progress toward an efficient panel of SNPs for ancestry inference, Forensic Sci. Int. Genet. 10 (2014) 23–32, doi:http://dx.doi.org/10.1016/j.fsigen.2014.01.002.
[72] W. Branicki, F. Liu, K. van Duijn, J. Draus-Barini, E. Pośpiech, S. Walsh, et al., Model-based prediction of human hair color using DNA variants, Hum. Genet. 129 (2011) 443–454, doi:http://dx.doi.org/10.1007/s00439-010-0939-8.
[73] S. Walsh, F. Liu, K.N. Ballantyne, M. van Oven, O. Lao, M. Kayser, 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 (2011) 170–180, doi:http://dx.doi.org/10.1016/j.fsigen.2010.02.004.
[74] J.D. Churchill, S.E. Schmedes, J.L. King, B. Budowle, Evaluation of the Illumina® beta version ForenSeq™ DNA Signature Prep Kit for use in genetic profiling, Forensic Sci. Int. Genet. 20 (2016) 20–29, doi:http://dx.doi.org/10.1016/j.fsigen.2015.09.009.
[75] S. Caratti, S. Turrina, M. Ferrian, E. Cosentino, D. De Leo, MiSeq FGx™ Sequencing System: a new platform for forensic genetics, Forensic Sci. Int. Genet. Suppl. Ser. 1768 (2015), doi:http://dx.doi.org/10.1016/j.fsigss.2015.09.040.
[76] Scientific Working Group on DNA Analysis Methods Validation Guidelines for DNA Analysis Methods, (2004) 1–13.
[77] Illumina Inc., ForenSeq ™ DNA Signature Prep Reference Guide, (2015).
[78] Illumina Inc., MiSeq FGx™ Instrument Reference Guide, (2015).
[79] Illumina Inc., ForenSeq™ Universal Analysis Software User Guide, (2015).
[80] J.R. Kidd, F.R. Friedlaender, W.C. Speed, A.J. Pakstis, F.M. De La Vega, K.K. Kidd, Analyses of a set of 128 ancestry informative single-nucleotide polymorphisms in a global set of 119 population samples, Investig. Genet. 2 (2011) 1, doi:http://dx.doi.org/10.1186/2041-2223-2-1.
[81] M. Kayser, Forensic DNA Phenotyping: predicting human appearance from crime scene material for investigative purposes, Forensic Sci. Int. Genet. 18 (2015) 33–48, doi:http://dx.doi.org/10.1016/j.fsigen.2015.02.003.
[82] C. Phillips, Forensic genetic analysis of bio-geographical ancestry, Forensic Sci. Int. Genet. 18 (2015) 49–65, doi:http://dx.doi.org/10.1016/j.fsigen.2015.05.012.
[83] S. Walsh, L. Chaitanya, L. Clarisse, L. Wirken, J. Draus-Barini, L. Kovatsi, 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, doi:http://dx.doi.org/10.1016/j.fsigen.2013.12.006.
[84] F. Liu, K. van Duijn, J.R. Vingerling, A. Hofman, A.G. Uitterlinden, a.C.J.W. Janssens, et al., Eye color and the prediction of complex phenotypes from genotypes, Curr. Biol. 19 (2009) 192–193, doi:http://dx.doi.org/10.1016/j.cub.2009.01.027.
[85] M. Fondevila, C. Phillips, C. Santos, A. Freire Aradas, P.M. Vallone, J.M. Butler, 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 (2013) 63–74, doi:http://dx.doi.org/10.1016/j.fsigen.2012.06.007.
[86] R.S. McLaren, J. Bourdeau-Heller, J. Patel, J.M. Thompson, J. Pagram, T. Loake, et al., Developmental validation of the PowerPlex(®) ESI 16/17 fast and PowerPlex(®) ESX 16/17 fast systems, Forensic Sci. Int. Genet. 13 (2014) 195–205, doi:http://dx.doi.org/10.1016/j.fsigen.2014.08.004.
[87] J.J. Mulero, C.W. Chang, L.M. Calandro, R.L. Green, Y. Li, C.L. Johnson, et al., Development and validation of the AmpFlSTR Yfiler PCR amplification kit: a male specific, single amplification 17 Y-STR multiplex system, J. Forensic Sci. 51 (2006) 64–75, doi:http://dx.doi.org/10.1111/j.1556-4029.2005.00016.x.
[88] K.B. Gettings, K.M. Kiesler, S.A. Faith, E. Montano, C.H. Baker, B.A. Young, et al., Sequence variation of 22 autosomal STR loci detected by next generation sequencing, Forensic Sci. Int. Genet. 21 (2015) 15–21, doi:http://dx.doi.org/10.1016/j.fsigen.2015.11.005.
[89] K.B. Gettings, R.A. Aponte, P.M. Vallone, J.M. Butler, STR allele sequence variation: current knowledge and future issues, Forensic Sci. Int. Genet. 18 (2015) 118–130, doi:http://dx.doi.org/10.1016/j.fsigen.2015.06.005.
[90] D.H. Warshauer, J.D. Churchill, N. Novroski, J.L. King, B. Budowle, Novel Y-chromosome short tandem repeat variants detected through the use of massively parallel sequencing, Genom. Proteom. Bioinf. 13 (2015) 250–257, doi:http://dx.doi.org/10.1016/j.gpb.2015.08.001.
[91] W. Parson, D. Ballard, B. Budowle, J.M. Butler, K.B. Gettings, P. Gill, et al., Massively parallel sequencing of forensic STRs: considerations of the DNA commission of the International Society for Forensic Genetics (ISFG) on minimal nomenclature requirements, Forensic Sci. Int. Genet. 22 (2016) 54–63, doi:http://dx.doi.org/10.1016/j.fsigen.2016.01.009.