Evaluation of the Ion Torrent™ HID SNP 169-plex

A SNP typing assay developed for human identification by second generation sequencing

Børsting, Claus; Fordyce, Sarah L; Olofsson, Jill Katharina; Mogensen, Helle Smidt; Morling, Niels

Published in:
Forensic science international. Genetics

DOI:
10.1016/j.fsigen.2014.06.004

Publication date:
2014

Forensic Science International: Genetics 12 (2014) 144–154

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

Evaluation of the Ion Torrent™ HID SNP 169-plex: A SNP typing assay developed for human identification by second generation sequencing

Claus Børsting [1], Sarah L. Fordyce [1,*], Jill Olofsson, Helle Smidt Mogensen, Niels Morling

Section of Forensic Genetics, Department of Forensic Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, DK-2100 Copenhagen, Denmark

ARTICLE INFO

Article history:
Received 22 January 2014
Received in revised form 4 June 2014
Accepted 6 June 2014

Keywords:
Forensic genetics
Human identification
SNP*for*ID
Second generation sequencing
SNP typing
Ion PGM™

ABSTRACT

The Ion Torrent™ HID SNP assay amplified 136 autosomal SNPs and 33 Y-chromosome markers in one PCR and the markers were subsequently typed using the Ion PGM™ second generation sequencing platform. A total of 51 of the autosomal SNPs were selected from the SNP*for*ID panel that is routinely used in our ISO 17025 accredited laboratory. Concordance between the Ion Torrent™ HID SNP assay and the SNP*for*ID assay was tested by typing 44 Iraqis twice with the Ion Torrent™ HID SNP assay. The same samples were previously typed with the SNP*for*ID assay and the Y-chromosome haplogroups of the individuals were previously identified by typing 45 Y-chromosome SNPs. Full concordance between the assays were obtained except for the SNP genotypes of two SNPs. These SNPs were among the eight SNPs (rs2399332, rs1029047, rs10776839, rs4530059, rs8037429, rs430046, rs1031825 and rs1523537) with inconsistent allele balance among samples. These SNPs should be excluded from the panel.

The optimal amount of DNA in the PCR seemed to be ≥0.5 ng. Allele drop-outs were rare and only seen in experiments with <0.5 ng input DNA and with a coverage of <50 reads. No allele drop-in was observed. The great majority of the heterozygote allele balances were between 0.6 and 1.6, which is comparable to the heterozygote balances of STRs typed with PCR–CE. The number of reads with base calls that differed from the genotype call was typically less than five. This allowed detection of 1:100 mixtures with a high degree of certainty in experiments with a high total depth of coverage.

In conclusion, the Ion PGM™ is a very promising platform for forensic genetics. However, the secondary sequence analysis software made wrong genotype calls from correctly sequenced alleles. These types of errors must be corrected before the platform can be used in case work. Furthermore, the sequence analysis software should be further developed and include quality settings for each SNP based on validation studies.

1. Introduction

Second generation sequencing (SGS) technologies are highly interesting for the forensic genetic community. They provide the possibility to (1) obtain detailed sequence information on the traditional forensic genetic markers, (2) type combinations of markers that cannot be typed in the same assay with the currently used methods, and (3) collect massive amounts of data from either one individual or large amount of data from many individuals simultaneously. SGS of STRs [1–6] and mtDNA [7,8] demonstrated the feasibility of sequencing traditional forensic genetic markers with SGS platforms and two of the dominating companies in the SGS market, Illumina® and Life Technologies™, are currently developing sequencing assays for use in forensic genetics.

Life Technologies™ developed a SNP typing assay for human identification named the Ion Torrent™ HID SNP panel. An early version of this panel (v0.1) was recently tested by Budowle and co-workers [9]. In the current version (v2.2), the panel consisted of 169 SNPs; 51 of the 52 SNP*for*ID SNPs [10], 89 of the 92 individual identification SNPs (IISNPs) [11] (four SNPs overlap between these two panels: rs1490413, rs891700, rs2046361 and rs901398) and 33 Y-chromosome markers, that designated the major haplogroups in the Y-chromosome parsimony tree [12]. The assay involved a 169-plex PCR followed by ligation of barcoded adapters, emulsion PCR (emPCR) and sequencing on the Ion Personal Genome Machine (PGM™).

* Corresponding author. Tel.: +45 353226308.
E-mail addresses: sarah.fordyce@sund.ku.dk, slfordyce@gmail.com (S.L. Fordyce).
[1] These authors contributed equally.

http://dx.doi.org/10.1016/j.fsigen.2014.06.004
1872-4973/

In the last decade, SNP typing assays developed for forensic genetic purposes typically involved a multiplex PCR, a multiplex single base extension (SBE) reaction and detection of the SBE products by capillary electrophoresis (CE). PCR–SBE–CE assays are sensitive, relatively cheap and may be performed in just 1 day. The PCR–SBE–CE assays were validated for case work in some laboratories [13–16]. However, PCR–SBE–CE assays are not widely used, most likely because there are no commercial kits available and no commercial software solutions designed for analysis of SBE products. Analyses of the SNP electropherograms were challenging [13,17] and small sized peaks from PCR products or PCR primers extended with ddNTPs in the SBE reaction are sometimes detected and misinterpreted as true alleles [18,19].

SGS of SNPs is an attractive alternative to the PCR–SBE–CE assays and in this work, we evaluated the Ion PGM™ and the Ion Torrent™ HID SNP panel for forensic genetic testing by comparing the SGS results with the results from our ISO 17025 accredited PCR–SBE–CE protocol [13].

2. Materials and methods

2.1. Samples, DNA purification and quantification

Two Danish control samples (one male and one female) and samples from 44 unrelated Iraqi male individuals were selected. DNA was purified from 200 μL of blood using the DNA Blood Mini Kit (Qiagen) as recommended by the manufacturer. DNA was eluted in 50 μL of AE Buffer (Qiagen).

The DNA from the control samples were quantified using the Qubit2.0 (Invitrogen). Serial dilutions were performed to generate DNA concentrations of 10, 5, 2, 1, 0.5, 0.2 and 0.1 ng/μL. The DNA concentrations were verified by quantification on the Qubit®2.0 (Invitrogen).

The study was approved by the Danish Ethical Committee (H-3-2012-023).

2.2. Construction of mixtures

Mixtures of DNA from the two control samples were generated to give mixture ratios of 1000:1, 100:1, 10:1, 5:1, 2:1, 1:1, 1:2, 1:5, 1:10, 1:100 and 1:1000. A total of 5 ng female DNA was mixed with 5 pg, 50 pg, 500 pg and 5 ng male DNA to give female to male mixture ratios of 1000:1, 100:1, 10:1 and 1:1. A total of 1 ng female DNA was mixed with 500 pg male DNA to give a female to male mixture ratio of 2:1. A total of 5 ng male DNA was mixed with 5 pg, 50 pg, 500 pg and 5 ng female DNA to give male to female mixture ratios of 1000:1, 100:1, 10:1 and 1:1. A total of 2.5 ng male DNA was mixed with 500 pg female DNA to give a male to female mixture ratio of 5:1. For the mixture experiments, the DNA input in the PCR ranged from 1.5 to 10 ng DNA.

2.3. Library preparation

DNA libraries were constructed using the Ion AmpliSeq™Library Kit 2.0 (Life Technologies) reagents in conjunction with the HID SNP primer panel v2.2 (Life Technologies). The primer mix was kindly provided by Life Technologies™. The PCR contained 4 μL of 5X Ion AmpliSeqHiFi Master Mix and 10 μL of 2X Ion AmpliSeq Primer Pool (HID SNP panel v2.2). Except for the mixture and sensitivity experiments, the DNA input was 2.5–10 ng. The PCR programme was 2 min at 99 °C, 18 cycles of 15 s at 99 °C and 4 min at 60 °C followed by a 10 °C hold. A total of 2 μL FuPa reagent (Life Technologies) was added to digest excess PCR primers and the ends of the PCR products. The reaction was incubated at 50 °C for 10 min, 55 °C for 10 min and 60 °C for 20 min. All libraries were barcoded using the Ion Xpress™ Barcode Adapters (Life Technologies). A total of 4 μL Switch Solution, 2 μL barcode and adapter mix, and 2 μL DNA Ligase were added to the libraries and incubated at 22 °C for 30 min and72 °C for 10 min. Libraries were purified using AMPure XP Reagents (Agencourt) following the manufacturer's protocol.

Libraries were either quantified using the Ion Library Quantitation kit (Life Technologies) and pooled in equal amounts or 'equalized' to a concentration of 100pM using the Ion Library Equalizer™ Kit (Life Technologies) as recommended by the manufacturer.

2.4. Emulsion PCR and sequencing

Emulsion PCR (emPCR) was performed on the OneTouch™ 2 (OT2) instrument (Life Technologies). The emPCR products were quantified using the Qubit® 2.0 (Invitrogen) as recommended in the protocol in order to quantify the DNA (with the AF647 probes that hybridise to the X adapter sequence) and the number of ISPs (with the AF488 probes that hybridise to the bead-anchor sequence). The amount of DNA should be less than the number of ISPs to reduce polyclonality. The ratio of templated to untemplated ISPs was calculated using the Qubit® 2.0 Easy Calculator (Life Technologies). Samples were only submitted for sequencing if the template ISPs represented 10–30%. The emPCR products were enriched on the Ion OneTouch™Enrichment System (Life Technologies) using the Ion PGM™ Template OT2 200 Kit (Life Technologies) following the manufacturer's protocol.

Sequencing was performed on the Ion PGM™ using the Ion PGM™ Sequencing 200 Kit v2 reagents. Ion Chip sizes varied depending on the experiment.

2.5. Data analysis

Raw data from the sequencing reactions were collected as DAT files by the Ion PGM™. In the primary sequence analysis, the DAT files were processed on the Torrent Suite Server (v3.6.2). Default settings were used for the primary sequence analysis. First, the background signal was estimated from empty wells and subtracted from the raw signal in wells with ISPs. Second, the raw signals were normalised according to the key sequence (TCAG) in the adapter. Third, base calling and quality assessment of each base call (Phred score) were performed. Fourth, sequences of low quality, primer–dimer sequences and sequences from ISPs with more than one template (poly-clonal sequences) were removed. Fifth, the sequences were trimmed according to Phred scores (if the average Phred score for 30 bp sliding windows <15), according to signal unbalance (>3% of the base calls in a sequences were 0.5–0.59 or 1.4–1.49 bp for one base and 1.5–1.59 or 2.4–2.49 bp for two bases) and by removing some base calls in the 3'adapter. Sixth, the sequences were aligned to a reference sequence (human genome build 19) and a binary alignment map (BAM) file was generated.

The secondary sequence analysis was performed using the generated BAM file and the HID_SNP_Genotyper (v2.2) plug-in with low-stringency settings as recommended by the manufacturer. HID_SNP_Genotyper (v2.2) generated a variant caller file (VCF) with genotype calls, a quality assessment of the genotype call (*P*-value), the total depth of coverage, the number of sequence reads in forward and reverse directions, the number of base calls for all four bases and the number of no calls (named "Deletions") in the SNP position. The data in the VCF files was further analysed in Excel™ and the statistical software R. Possible deviations from Hardy–Weinberg expectations was calculated using Arlequin 3.5.

Coverage analysis and target uniformity were calculated using the Ion Torrent Coverage Analysis (v3.6.63324) plug-in.

Heterozygote allelic balances for the Iraqi samples and the dilution experiments samples were calculated as the number of

reads for one nucleotide divided by the number of reads for the other nucleotide in the order A, C, G and T.

For the mixture experiment, the theoretical allele balance was calculated as the number of expected base calls from the minor contributor/total number of base calls for SNPs where the two individuals did not share any allele (opposite homozygotes). For SNPs, where the two individuals shared one allele, the theoretical balance was calculated as 0.5 × the expected number of base calls from the minor contributor/total number of base calls (see Supplementary Table S1).

3. Results

3.1. Locus balance and total depth of coverage

The male control sample was used to test the performance of the sequencing assay. In the first series of experiments, the sample was PCR amplified once and one library was constructed. The library was used for two independent emPCRs and each emPCR was used for two sequencing runs on separate Ion 314 Chips. In the second series of experiments, eight independent PCRs and library builds were performed. Two sets of four libraries were pooled prior to emPCR and subsequent sequencing on two Ion 314 Chips.

Fig. 1 and Supplementary Fig. S1 shows the coverage variation for each SNP from the first and second series of experiments, respectively. Some loci were sequenced with up to 10 times more reads than other loci and the locus balance seemed consistent between experiments. The PCR introduced most of the locus imbalance, whereas the emPCR and the sequencing steps generated some, but much less variation (Fig. 1). The lowest coverage was observed for rs917118, rs321198 and rs576261.

Pooling of libraries prior to emPCR and sequencing resulted in coverage depth imbalance between samples. When four samples were pooled on one Ion 314 Chip, the total coverage for each sample ranged from 19% to 32% (99,018–163,896 reads) of the total reads generated during the sequencing run. When 34 samples were pooled on one Ion 318 Chip, the coverage ranged from 1% to 9% (19,883–588,479 reads) of the total reads.

3.2. Analysis of the SNP typing performance

A total of 44 Iraqi individuals were typed twice with the Ion Torrent™ HID SNP panel. The consensus profiles may be found in Supplementary Table S2. When the two profiles for each sample were compared, 178 inconsistencies were observed. Too few reads were detected to call the genotype in 54 of these inconsistencies, which lead to a "no call" rate of 0.36% for the 14,872 genotypes. However, in the remaining 124 inconsistencies, inspection of the sequence data revealed that the inconsistencies were caused by the secondary sequence analysis software, i.e. the HID_SNP_Genotyper (v2.2) plug-in. Analysis of unusual allelic balances in heterozygous individuals and unusually low fractions of reads for homozygous allele calls revealed another 40 errors by the HID_SNP_Genotyper (v2.2) plug-in. These errors were reproduced in both experiments and would have resulted in wrong genotype calls. Overall, the error rate of the HID_SNP_Genotyper (v2.2) plug-in was approximately 1.1%. Errors were detected in many different SNPs, but a large proportion of the errors were detected in four SNPs: rs2399332, rs4606077, rs430046 and rs576261. Ten examples of errors are shown in Supplementary Table S3. The detected errors were manually corrected by changing the genotype call according to the relative numbers of base calls in the VCF file and the genotyping performance of each SNP was analysed in detail (Table 1A). The allelic balance of most SNPs was close to 1:1, and in general, the variation in allelic balance was small. However, the allelic balances of eight autosomal SNPs (rs2399332, rs1029047, rs10776839, rs4530059, rs8037429, rs430046, rs1031825 and rs1523537) were not satisfactory and these SNPs should be excluded from the panel. For most homozygous SNP calls, the percentage of reads with the called nucleotide typically ranged from 97% to 99%, and overall, the number of reads with nucleotide calls that differed from the SNP genotype call was

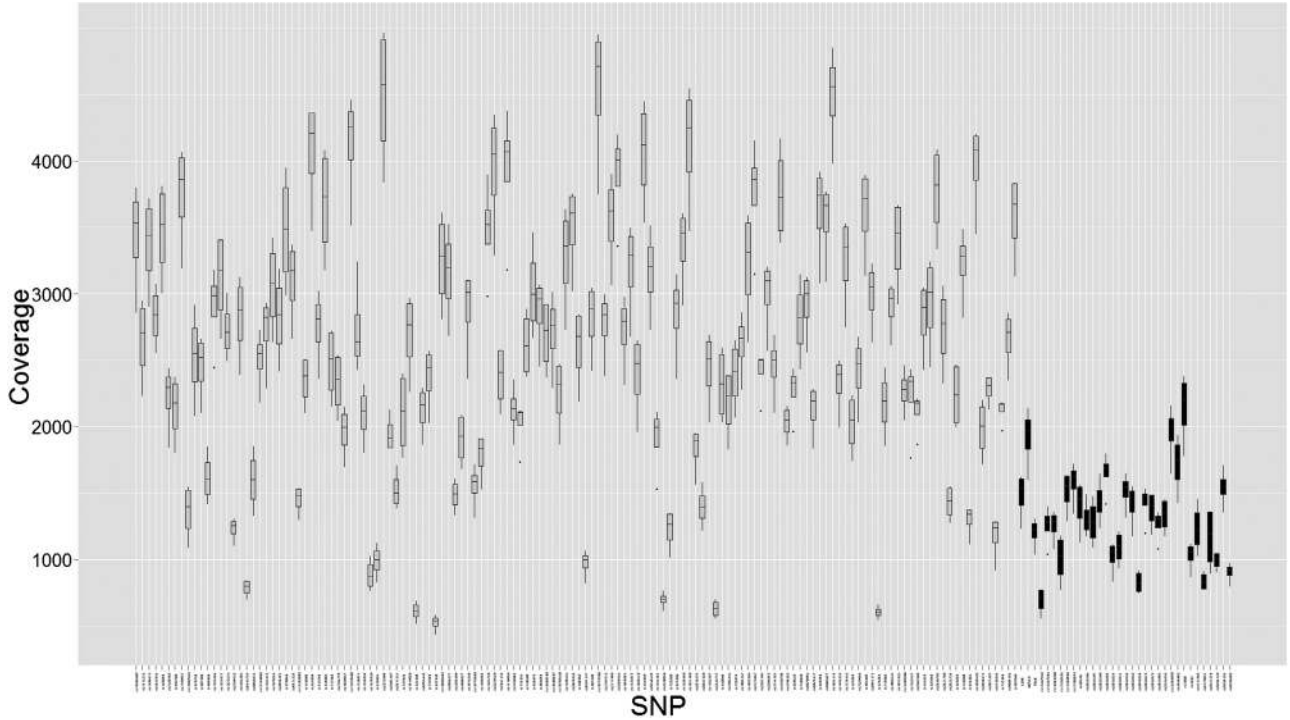

Fig. 1. Locus balance of the Ion Torrent™ HID SNP panel. Four independent emPCRs and sequencing experiments were performed on the same library. The order of the SNPs corresponds with Table 1A. Grey colouring indicates autosomal SNPs, while black represents Y-SNPs.

Table 1A
Analysis of autosomal SNPs.

Autosomal SNP	Hardy–Weinberg test			Homozygous calls		Heterozygous calls		
	Observed number of heterozygotes	Expected number of heterozygotes	*P*-value	Fraction of reads (Allele 1)[a]	Fraction of reads (Allele 2)[a]	Allele balance[b]	Number of base calls differing from the genotype call[a]	Number of deletions[a]
rs1490413	0.431	0.492	0.536	0.98 (0.97–0.99)	0.99 (0.99–1)	0.95 ± 0.08	0.87 (0–8)	7.82 (0–43)
rs7520386	0.409	0.501	0.238	0.99 (0.99–1)	0.99 (0.99–1)	1.11 ± 0.09	1.02 (0–9)	0.12 (0–3)
rs4847034	0.363	0.379	1	0.99 (0.99–0.99)	0.98 (0.95–0.99)	0.96 ± 0.09	15.2 (0–75)	0.02 (0–1)
rs560681	0.477	0.446	0.740	0.99 (0.99–1)	0.99 (0.98–0.99)	0.97 ± 0.08	1.54 (0–11)	3 (0–18)
rs10495407	0.454	0.438	1	0.96 (0.93–0.98)	0.99 (0.99–1)	0.92 ± 0.09	1.03 (0–9)	16.1 (0–137)
rs891700	0.522	0.505	1	0.99 (0.99–1)	0.99 (0.99–1)	0.99 ± 0.10	1.69 (0–11)	0.01 (0–1)
rs1413212	0.454	0.479	0.757	0.99 (0.98–1)	0.99 (0.99–1)	0.92 ± 0.07	2.34 (0–17)	5.25 (0–26)
rs876724	0.272	0.329	0.345	0.99 (0.99–1)	0.99 (0.99–1)	1.21 ± 0.15	1.22 (0–19)	0.12 (0–2)
rs1109037	0.659	0.505	0.067	0.99 (0.99–0.99)	0.99 (0.99–1)	1.01 ± 0.11	2.17 (0–29)	0.34 (0–2)
rs993934	0.431	0.505	0.374	0.99 (0.93–1)	0.99 (0.98–1)	0.93 ± 0.19[c]	0.75 (0–6)	0.06 (0–2)
rs12997453	0.340	0.367	0.682	0.99 (0.99–1)	0.99 (0.97–1)	0.91 ± 0.15[c]	0.20 (0–2)	2.45 (0–16)
rs907100	0.5	0.489	1	0.99 (0.98–1)	0.99 (0.99–1)	1.01 ± 0.14	1.76 (0–9)	0.95 (0–8)
rs1357617	0.545	0.489	0.537	0.99 (0.99–1)	0.98 (0.96–1)	1.00 ± 0.07	4.46 (0–18)	4.66 (0–38)
rs4364205	0.431	0.454	0.821	0.99 (0.99–1)	0.99 (0.99–1)	1.03 ± 0.07	5.30 (0–31)	0.34 (0–4)
rs9866013[e,f]	0.227	0.379	0.013	0.98 (0.96–1)	0.99 (0.96–1)	0.97 ± 0.11	3.76 (0–105)	1.30 (0–22)
rs2399332[f,g]	0.142	0.495	0.000	0.97 (0.86–1)	0.97 (0.85–1)	0.24 ± 0.16	4.20 (0–29)	0.25 (0–4)
rs1872575	0.454	0.501	0.558	0.99 (0.98–1)	0.99 (0.99–1)	0.90 ± 0.10	0.63 (0–6)	3.37 (0–26)
rs1355366	0.522	0.474	0.535	0.99 (0.98–0.99)	0.99 (0.98–1)	1.00 ± 0.10	1.12 (0–6)	3.53 (0–35)
rs6444724	0.454	0.479	0.757	0.99 (0.99–1)	0.99 (0.99–1)	1.00 ± 0.11	0.47 (0–3)	0.10 (0–2)
rs2046361	0.522	0.474	0.535	0.98 (0.94–1)	0.99 (0.97–1)	0.98 ± 0.13	2.18 (0–71)	0.06 (0–1)
rs279844	0.454	0.401	0.464	0.99 (0.98–1)	0.99 (0.98–1)	1.00 ± 0.10	1.87 (0–24)	0.03 (0–1)
rs13134862	0.318	0.329	1	0.99 (0.98–0.99)	0.99 (0.98–1)	0.97 ± 0.07	0.67 (0–6)	0 (0–0)
rs1554472	0.454	0.496	0.758	0.98 (0.96–1)	0.99 (0.99–1)	0.92 ± 0.07	3.82 (0–30)	0.43 (0–4)
rs6811238	0.5	0.501	1	0.99 (0.99–1)	0.99 (0.99–1)	0.96 ± 0.08	1.59 (0–8)	1.37 (0–7)
rs1979255	0.522	0.430	0.178	1 (1–1)	0.99 (0.99–1)	1.03 ± 0.14	0.52 (0–6)	0.04 (0–2)
rs717302	0.363	0.479	0.122	0.99 (0.98–1)	0.99 (0.99–1)	0.95 ± 0.07	3.89 (0–29)	0.39 (0–5)
rs159606	0.386	0.367	1	0.99 (0.99–1)	0.99 (0.99–1)	0.95 ± 0.11	1.31 (0–10)	0.31 (0–4)
rs13182883[e]	0.454	0.468	1	0.99 (0.99–1)	0.99 (0.99–1)	0.99 ± 0.12	0.43 (0–9)	0.11 (0–2)
rs7704770	0.568	0.474	0.213	0.99 (0.98–1)	0.99 (0.99–1)	1.01 ± 0.13	1.15 (0–16)	0.80 (0–4)
rs315791	0.386	0.503	0.138	0.99 (0.99–1)	0.99 (0.98–1)	1.02 ± 0.10	2.62 (0–17)	0.83 (0–6)
rs251934	0.431	0.446	1	0.99 (0.99–1)	0.99 (0.99–1)	1.05 ± 0.09	2.05 (0–12)	0.33 (0–3)
rs338882	0.431	0.484	0.534	0.99 (0.99–1)	0.99 (0.99–1)	1.03 ± 0.06	1.30 (0–12)	0.02 (0–1)
rs1029047[g]	0.386	0.315	0.320	–	0.98 (0.69–1)	1.76 ± 2.02	2.82 (0–21)	2.75 (0–11)
rs13218440	0.613	0.492	0.127	0.99 (0.99–0.99)	0.99 (0.99–1)	0.99 ± 0.07	1.73 (0–29)	0.09 (0–2)
rs2811231	0.386	0.390	1	0.99 (0.98–1)	0.99 (0.99–1)	0.81 ± 0.10	2.36 (0–11)	2.39 (0–10)
rs1336071	0.5	0.496	1	0.99 (0.97–1)	0.99 (0.99–1)	1.07 ± 0.27[c]	0.38 (0–6)	0.11 (0–2)
rs1478829	0.409	0.438	0.731	0.99 (0.9–1)	0.99 (0.99–1)	1.03 ± 0.42[c]	0.60 (0–5)	0.01 (0–1)
rs1358856[f]	0.340	0.505	0.037	0.99 (0.99–1)	0.99 (0.99–1)	0.72 ± 0.16[c]	0.80 (0–6)	0 (0–0)
rs2503107	0.545	0.505	0.762	0.99 (0.99–1)	0.99 (0.99–1)	0.79 ± 0.13	0.84 (0–5)	0.15 (0–3)
rs2272998[f]	0.318	0.468	0.048	0.99 (0.99–1)	0.99 (0.99–1)	0.98 ± 0.08	1.40 (0–9)	0.33 (0–5)
rs214955	0.5	0.496	1	0.99 (0.99–1)	0.99 (0.98–1)	0.96 ± 0.18[c]	0.39 (0–3)	0.01 (0–1)
rs727811	0.568	0.505	0.545	0.99 (0.99–1)	0.97 (0.95–0.99)	1.04 ± 0.16	0.20 (0–2)	7.29 (0–54)
rs6955448	0.295	0.411	0.072	0.99 (0.99–1)	0.99 (0.99–1)	0.94 ± 0.11	2.11 (0–16)	0.40 (0–8)
rs917118	0.5	0.421	0.283	0.99 (0.99–1)	0.99 (0.99–1)	1.00 ± 0.10	0.32 (0–2)	0.17 (0–1)
rs1019029	0.363	0.421	0.469	0.99 (0.99–1)	0.99 (0.98–1)	1.03 ± 0.10	2.74 (0–17)	0.52 (0–6)
rs321198	0.5	0.496	1	0.99 (0.98–1)	0.99 (0.99–1)	1.00 ± 0.21[c]	0.41 (0–6)	0.06 (0–2)
rs737681	0.477	0.474	1	0.99 (0.98–1)	0.99 (0.99–1)	0.98 ± 0.08	2.37 (0–23)	3.55 (0–12)
rs10092491[f]	0.295	0.446	0.037	0.99 (0.98–1)	0.99 (0.98–1)	1.02 ± 0.10	6.02 (0–33)	3.68 (0–22)
rs4288409	0.295	0.315	0.643	0.99 (0.99–0.99)	0.99 (0.97–1)	1.10 ± 0.13	0.96 (0–6)	5.58 (0–33)
rs2056277	0.295	0.390	0.126	0.99 (0.98–1)	0.99 (0.99–0.99)	1.00 ± 0.14	2.18 (0–8)	5.22 (0–26)
rs4606077[e,f]	0.523	0.386	0.011	0.99 (0.99–1)	–	0.77 ± 0.22	0.60 (0–5)	0.32 (0–4)
rs1015250	0.409	0.379	0.705	0.99 (0.99–1)	0.99 (0.99–1)	1.05 ± 0.09	1 (0–11)	0.13 (0–2)
rs2270529	0.386	0.342	0.657	0.99 (0.99–0.99)	0.99 (0.99–1)	1.06 ± 0.10	1.10 (0–7)	0.02 (0–1)
rs7041158	0.5	0.489	1	0.99 (0.99–1)	0.99 (0.98–1)	1.02 ± 0.09	0.36 (0–5)	0.86 (0–6)
rs1463729	0.545	0.496	0.552	0.99 (0.98–1)	0.99 (0.99–1)	0.90 ± 0.08	1.60 (0–10)	0.5 (0–3)
rs1360288	0.477	0.492	1	0.99 (0.99–1)	0.99 (0.98–1)	0.99 ± 0.10	1.19 (0–24)	0.02 (0–1)
rs10776839[f,g]	0.659	0.492	0.030	0.99 (0.96–1)	0.97 (0.95–1)	0.69 ± 0.49	1.69 (0–18)	0 (0–0)
rs826472	0.477	0.446	0.740	0.99 (0.99–1)	0.99 (0.99–1)	1.07 ± 0.19[c]	0.41 (0–8)	0.17 (0–3)
rs735155	0.477	0.503	0.768	0.99 (0.99–1)	0.99 (0.99–1)	0.91 ± 0.09	0.84 (0–7)	0.11 (0–1)
rs3780962	0.363	0.468	0.191	0.99 (0.99–1)	0.99 (0.98–1)	0.99 ± 0.08	1.95 (0–12)	0.67 (0–5)
rs1410059	0.340	0.461	0.103	0.99 (0.99–1)	0.99 (0.99–1)	1.02 ± 0.12	2 (0–12)	0.01 (0–1)
rs740598	0.5	0.504	1	0.99 (0.98–0.99)	0.99 (0.99–1)	1.00 ± 0.10	1.90 (0–10)	0 (0–0)
rs964681	0.454	0.505	0.553	0.99 (0.98–0.99)	0.99 (0.99–1)	0.93 ± 0.09	1.86 (0–13)	2.24 (0–24)
rs10768550	0.295	0.254	0.568	0.99 (0.99–1)	–	1.01 ± 0.10	0.44 (0–4)	0.24 (0–3)
rs10500617	0.295	0.254	0.568	–	0.99 (0.98–1)	1.01 ± 0.10	1.58 (0–6)	0.55 (0–3)
rs1498553[f]	0.318	0.509	0.008	0.99 (0.99–1)	0.99 (0.98–1)	0.99 ± 0.09	2.10 (0–10)	1.38 (0–10)
rs901398	0.5	0.427	0.614	0.99 (0.99–1)	0.99 (0.99–1)	0.99 ± 0.09	2.88 (0–13)	0.09 (0–2)
rs6591147	0.477	0.474	1	0.99 (0.97–1)	0.99 (0.99–1)	1.04 ± 0.18	0.82 (0–6)	0.39 (0–4)
rs10488710	0.545	0.489	0.536	0.99 (0.99–1)	0.99 (0.99–1)	1.00 ± 0.06	1.02 (0–12)	0.11 (0–1)
rs590162	0.522	0.505	1	0.99 (0.99–1)	0.99 (0.96–1)	1.19 ± 0.24	0.63 (0–9)	0.31 (0–2)
rs2076848	0.568	0.499	0.379	0.99 (0.99–1)	0.99 (0.99–1)	0.97 ± 0.08	1.93 (0–12)	0.02 (0–1)
rs2107612	0.340	0.411	0.279	0.99 (0.97–1)	0.99 (0.97–1)	0.68 ± 0.16[c]	0.5 (0–4)	0.15 (0–2)

Table 1A (*Continued*)

Autosomal SNP	Hardy–Weinberg test			Homozygous calls		Heterozygous calls		
	Observed number of heterozygotes	Expected number of heterozygotes	*P*-value	Fraction of reads (Allele 1)[a]	Fraction of reads (Allele 2)[a]	Allele balance[b]	Number of base calls differing from the genotype call[a]	Number of deletions[a]
rs2269355	0.431	0.492	0.536	0.99 (0.96–1)	0.99 (0.99–1)	1.39 ± 1.60[d]	2.26 (0–16)	0.17 (0–1)
rs2111980	0.431	0.492	0.535	0.99 (0.99–1)	0.99 (0.99–1)	1.02 ± 0.10	1.25 (0–9)	0.01 (0–1)
rs10773760	0.522	0.484	0.753	0.99 (0.99–1)	0.99 (0.99–1)	1.00 ± 0.05	2.87 (0–23)	0.09 (0–1)
rs1335873	0.409	0.421	1	0.99 (0.98–1)	0.99 (0.98–1)	1.00 ± 0.11	5.54 (0–23)	0.02 (0–1)
rs1886510	0.431	0.474	0.746	0.99 (0.97–1)	0.99 (0.98–1)	0.91 ± 0.08	0.60 (0–5)	0.76 (0–5)
rs9546538	0.409	0.468	0.515	0.99 (0.99–1)	0.99 (0.99–1)	1.06 ± 0.16	0.43 (0–8)	0.10 (0–3)
rs1058083	0.5	0.438	0.489	0.99 (0.99–0.99)	0.99 (0.99–1)	1.03 ± 0.08	1.38 (0–10)	0.33 (0–2)
rs354439	0.5	0.504	1	0.99 (0.98–1)	0.99 (0.99–0.99)	0.95 ± 0.09	2.37 (0–26)	0 (0–0)
rs1454361	0.340	0.492	0.062	0.99 (0.99–1)	0.99 (0.99–1)	0.94 ± 0.06	3.26 (0–13)	1.26 (0–12)
rs722290	0.545	0.501	0.759	0.99 (0.99–1)	0.99 (0.99–1)	1.01 ± 0.17	0.39 (0–4)	0.11 (0–2)
rs873196	0.318	0.379	0.414	0.99 (0.99–1)	0.99 (0.98–1)	1.05 ± 0.10	5.56 (0–28)	0.03 (0–1)
rs4530059[g]	0.477	0.461	1	0.99 (0.99–0.99)	0.99 (0.99–1)	1.45 ± 0.94	0.72 (0–10)	0.15 (0–2)
rs2016276	0.25	0.286	0.584	0.99 (0.98–0.99)	0.98 (0.89–1)	0.97 ± 0.11	6.77 (0–35)	0.31 (0–5)
rs1821380	0.5	0.479	1	0.99 (0.99–1)	0.99 (0.99–0.99)	1.04 ± 0.08	2.46 (0–13)	1.24 (0–7)
rs8037429[g]	0.340	0.499	0.064	0.99 (0.97–1)	0.99 (0.88–1)	0.55 ± 0.49	0.74 (0–8)	0.10 (0–2)
rs1528460	0.431	0.430	1	0.99 (0.99–1)	0.99 (0.99–1)	1.03 ± 0.13	0.77 (0–6)	0.26 (0–2)
rs729172	0.431	0.461	0.745	0.99 (0.99–1)	0.99 (0.99–1)	0.98 ± 0.13	1.39 (0–9)	0.29 (0–4)
rs2342747	0.454	0.379	0.245	0.99 (0.99–1)	0.99 (0.99–1)	0.96 ± 0.10	0.29 (0–3)	0.40 (0–4)
rs7205345	0.386	0.492	0.213	0.99 (0.99–1)	0.99 (0.99–1)	0.99 ± 0.06	3.02 (0–18)	0.04 (0–2)
rs430046[e,f,g]	0.25	0.461	0.003	0.98 (0.63–1)	0.85 (0.75–0.89)	1.88 ± 0.23	0.31 (0–4)	52.8 (0–582)
rs1382387[f]	0.5	0.379	0.041	0.99 (0.98–1)	–	1.00 ± 0.09	2.95 (0–15)	1.48 (0–6)
rs9905977	0.431	0.461	0.744	0.99 (0.99–1)	0.99 (0.99–1)	0.99 ± 0.11	1.27 (0–11)	0.25 (0–2)
rs740910	0.409	0.379	0.704	0.99 (0.99–1)	0.99 (0.98–1)	0.96 ± 0.09	1.84 (0–14)	0.55 (0–5)
rs4796362	0.5	0.505	1	0.99 (0.98–1)	0.99 (0.99–1)	0.99 ± 0.24[d]	1.17 (0–9)	1.37 (0–8)
rs2175957	0.590	0.504	0.361	0.99 (0.99–1)	0.99 (0.99–1)	0.96 ± 0.07	3.06 (0–13)	0.18 (0–2)
rs8070085	0.522	0.503	1	0.99 (0.99–1)	0.99 (0.99–1)	0.97 ± 0.08	1.11 (0–15)	0.06 (0–2)
rs1004357	0.386	0.390	1	0.99 (0.98–1)	0.99 (0.95–1)	1.04 ± 0.14	1.63 (0–12)	3.90 (0–18)
rs1027895	0.5	0.421	0.284	0.99 (0.99–1)	0.99 (0.98–0.99)	0.96 ± 0.08	1.89 (0–13)	0.04 (0–1)
rs938283	0.340	0.315	1	0.99 (0.99–0.99)	0.99 (0.99–1)	1.00 ± 0.09	1.15 (0–8)	0.13 (0–1)
rs8078417	0.477	0.474	1	0.99 (0.99–0.99)	0.98 (0.97–0.99)	0.92 ± 0.08	5.94 (0–40)	0.31 (0–5)
rs2291395	0.454	0.479	0.758	0.99 (0.97–1)	0.99 (0.99–1)	1.05 ± 0.08	1.11 (0–8)	0 (0–0)
rs4789798	0.454	0.479	0.758	0.99 (0.99–1)	0.99 (0.98–1)	1.04 ± 0.11	3.91 (0–22)	0.10 (0–2)
rs689512	0.522	0.499	0.769	0.99 (0.99–1)	0.99 (0.99–1)	0.95 ± 0.10	1.39 (0–10)	0.11 (0–2)
rs3744163	0.522	0.499	0.769	0.99 (0.98–1)	0.99 (0.98–1)	0.97 ± 0.11	1.93 (0–17)	2.62 (0–13)
rs2292972	0.5	0.496	1	0.99 (0.98–0.99)	0.99 (0.99–1)	0.99 ± 0.08	3.81 (0–22)	0 (0–0)
rs1493232	0.340	0.411	0.279	0.99 (0.98–0.99)	0.99 (0.99–1)	1.02 ± 0.07	6.38 (0–55)	0 (0–0)
rs9951171	0.5	0.504	1	0.99 (0.98–0.99)	0.99 (0.99–1)	0.96 ± 0.09	3.66 (0–39)	0.11 (0–2)
rs7229946	0.409	0.504	0.237	0.99 (0.99–1)	0.99 (0.99–1)	0.97 ± 0.07	1.81 (0–11)	0.10 (0–1)
rs985492	0.431	0.503	0.376	0.99 (0.98–0.99)	0.99 (0.99–1)	0.94 ± 0.07	2.66 (0–23)	1.17 (0–6)
rs521861	0.636	0.501	0.125	0.99 (0.99–1)	0.99 (0.98–1)	1.01 ± 0.09	2.73 (0–24)	0.47 (0–3)
rs1736442	0.386	0.499	0.218	0.99 (0.98–1)	0.99 (0.99–1)	0.96 ± 0.08	2.04 (0–20)	1.43 (0–19)
rs1024116	0.386	0.503	0.138	0.99 (0.98–1)	0.99 (0.99–1)	1.02 ± 0.13	1.83 (0–10)	1.52 (0–11)
rs719366	0.454	0.468	1	0.99 (0.99–1)	0.99 (0.99–1)	0.99 ± 0.09	1 (0–8)	0.39 (0–4)
rs576261[e]	0.5	0.468	0.749	0.99 (0.99–1)	0.98 (0.97–0.99)	1.05 ± 0.18	1.16 (0–19)	0.04 (0–1)
rs1031825[f,g]	0.767	0.496	0.000	0.99 (0.99–0.99)	0.97 (0.96–0.98)	0.79 ± 0.44	5.80 (0–100)	1.66 (0–10)
rs445251	0.477	0.503	0.768	0.99 (0.98–1)	0.99 (0.98–1)	0.80 ± 0.16	1.23 (0–9)	0.58 (0–4)
rs12480506	0.363	0.454	0.197	0.99 (0.99–1)	0.99 (0.99–1)	0.98 ± 0.05	1.25 (0–7)	0.03 (0–2)
rs2567608	0.477	0.446	0.741	0.99 (0.98–1)	0.99 (0.99–0.99)	0.96 ± 0.18[d]	3.77 (0–35)	0.02 (0–1)
rs1005533[f]	0.318	0.489	0.029	0.99 (0.99–1)	0.99 (0.99–1)	0.99 ± 0.08	0.79 (0–8)	1.87 (0–14)
rs1523537[f,g]	0.567	0.467	0.004	0.67 (0.57–0.76)	0.91 (0.58–1)	2.57 ± 3.91	62.8 (0–676)	28.6 (0–301)
rs722098	0.295	0.315	0.644	0.99 (0.99–1)	0.99 (0.99–1)	0.99 ± 0.07	2.12 (0–12)	0.04 (0–2)
rs464663	0.5	0.501	1	0.99 (0.99–1)	0.99 (0.99–1)	1.01 ± 0.08	2.22 (0–15)	0.27 (0–7)
rs2830795	0.386	0.390	1	0.99 (0.99–1)	0.99 (0.99–1)	0.99 ± 0.10	2.32 (0–14)	0.83 (0–9)
rs2831700	0.454	0.501	0.557	0.99 (0.98–1)	0.99 (0.98–1)	1.01 ± 0.11	1 (0–8)	0.37 (0–3)
rs2833736	0.5	0.496	1	0.93 (0.88–0.99)	0.99 (0.99–1)	0.90 ± 0.06	0.48 (0–5)	28.6 (0–250)
rs914165	0.522	0.492	0.761	0.99 (0.97–0.99)	0.99 (0.99–1)	1.01 ± 0.13	1.60 (0–26)	0.26 (0–2)
rs221956	0.409	0.438	0.732	0.99 (0.99–1)	0.99 (0.99–1)	1.02 ± 0.07	1.29 (0–6)	0.39 (0–7)
rs9606186	0.522	0.492	0.761	0.99 (0.99–1)	0.99 (0.99–1)	1.07 ± 0.08	2.09 (0–11)	0.17 (0–2)
rs5746846[e]	0.545	0.504	0.762	0.99 (0.99–1)	0.98 (0.90–1)	1.01 ± 0.14	0.66 (0–6)	4.91 (0–114)
rs2073383	0.409	0.489	0.351	0.99 (0.99–1)	0.99 (0.99–1)	0.93 ± 0.08	1.84 (0–13)	0.01 (0–1)
rs733164	0.363	0.401	0.701	0.92 (0.90–0.93)	0.99 (0.99–1)	0.93 ± 0.08	1.19 (0–8)	23.7 (0–223)
rs987640	0.5	0.504	1	0.99 (0.99–1)	0.99 (0.99–1)	0.99 ± 0.08	3.69 (0–16)	0.94 (0–8)
rs2040411	0.386	0.484	0.213	0.99 (0.98–1)	0.99 (0.99–1)	0.95 ± 0.10	2.95 (0–21)	0.41 (0–3)
rs1028528	0.5	0.421	0.284	0.99 (0.99–1)	0.99 (0.99–0.99)	1.02 ± 0.07	2.06 (0–14)	0.13 (0–2)

[a] Average and range.
[b] Average ± standard deviation.
[c] Relatively large variation caused by low coverage.
[d] Relatively large variation caused by a reproducible but unusual allele balance in one sample.
[e] Biased sequencing (>5 times) of either forward or reverse strand.
[f] Not in Hardy–Weinberg equilibrium ($P < 0.05$).
[g] Poorly performing SNP.

Table 1B
Analysis of Y-chromosome SNPs.

Y-chromosome SNP	Homozygous calls		Number of base calls differing from the genotype call[a]	Number of deletions[a]
	Fraction of reads (Allele 1)[a]	Fraction of reads (Allele 2)[a]		
rs2534636	0.99 (0.98–1)	0.99 (0.98–1)	0.5 (0–6)	0.02 (0–1)
rs9786608	0.99 (0.99–1)	0.99 (0.99–1)	0.48 (0–4)	0.08 (0–1)
rs35284970	0.99 (0.99–1)	–	0.41 (0–5)	0 (0–0)
rs9786184	0.99 (0.99–1)	0.99 (0.99–1)	1.53 (0–7)	0.01 (0–1)
rs9786139	0.98 (0.96–1)	0.99 (0.98–1)	0.53 (0–8)	0.97 (0–13)
rs16981290	–	0.99 (0.99–1)	0.19 (0–2)	0.04 (0–1)
L298	0.99 (0.99–1)	0.99 (0.98–1)	1.06 (0–5)	0.03 (0–1)
P256	–	0.99 (0.99–1)	0.20 (0–3)	0.20 (0–2)
rs17306671	0.99 (0.99–1)	0.99 (0.98–1)	0.89 (0–10)	0.17 (0–1)
rs4141886	0.99 (0.98–1)	0.99 (0.99–1)	0.91 (0–7)	0.02 (0–1)
rs2032595	–	0.99 (0.95–1)	0.82 (0–5)	0.01 (0–1)
rs2032597	0.99 (0.98–1)	–	0.44 (0–3)	0.06 (0–2)
rs2032599	–	0.99 (0.96–1)	0.34 (0–3)	0.19 (0–2)
rs20320	0.99 (0.99–1)	0.99 (0.99–1)	0.59 (0–4)	0.03 (0–2)
rs2032602	–	0.99 (0.97–1)	1.19 (0–7)	0.03 (0–1)
MSY2.2	n.d.	n.d.	n.d.	n.d.
rs8179021	0.99 (0.99–1)	0.99 (0.99–1)	0.30 (0–4)	0.13 (0–7)
rs2032624	0.99 (0.98–1)	0.99 (0.99–1)	0.32 (0–4)	0.06 (0–1)
rs2032636	0.97 (0.95–1)	0.97 (0.97–0.97)	1.59 (0–8)	7.47 (0–27)
rs9341278	–	0.99 (0.98–1)	0.12 (0–2)	0.04 (0–1)
rs2032674	–	0.99 (0.99–1)	1.29 (0–8)	0 (0–0)
rs2032658	0.99 (0.97–1)	0.99 (0.99–1)	0.36 (0–4)	2.74 (0–12)
rs17269816	0.99 (0.99–1)	–	0.32 (0–4)	0.05 (–1)
rs3848982	0.99 (0.98–1)	0.99 (0.98–1)	0.82 (0–8)	0.13 (0–3)
rs3900	0.99 (0.99–1)	0.99 (0.99–1)	0.58 (0–4)	0.10 (0–2)
rs3911	0.99 (0.99–1)	0.99 (0.98–1)	0.30 (0–7)	0 (0–0)
rs2032631	0.99 (0.97–1)	0.98 (0.88–1)	4.31 (0–28)	0.02 (0–1)
rs2032673	–	0.99 (0.98–1)	0.86 (0–6)	0.22 (0–2)
rs2032652	0.99 (0.98–1)	0.99 (0.98–1)	1.16 (0–5)	0.01 (0–1)
rs13447443	0.99 (0.99–1)	–	0.20 (0–2)	0 (0–0)
rs13447352	0.99 (0.99–1)	0.99 (0.97–1)	0.18 (0–3)	0 (0–0)
rs17250535	0.99 (0.99–1)	0.99 (0.95–1)	0.34 (0–5)	0.48 (0–4)
rs2033003	0.99 (0.98–1)	0.99 (0.99–1)	1.93 (0–8)	0 (0–0)

[a] Average and range.
n.d., not determined. MSY2.2 is a length polymorphism.

typically less than five. The number of reads with no calls (named "Deletions" in the VCF file) was typically less than one. However, in five autosomal SNPs (rs10495407, rs430046, rs1523537, rs2833736 and rs733164), the average number of reads with no calls was more than 10. A large number of reads with no calls may be a sign of misalignment of the reads in the primary sequence analysis. rs10495407, rs430046and rs733164 are positioned in homopolymer regions with four or five identical nucleotides which may explain the high number of no calls. However, rs1523537 and rs2833736 are not located in homopolymer stretches.

A total of 13 SNPs failed the X^2-test for Hardy–Weinberg equilibrium (Table 1B). Five of these were among the SNPs with skewed allelic balance that were already identified as poorly performing SNPs. A strong sequence bias was observed for rs4606077 and rs9866013 in the forward and reverse direction, respectively. This increases the risk of wrong genotype calls and may explain the Hardy–Weinberg disequilibrium. In five SNPs, only one of the two homozygous genotypes was observed among the 44 Iraqis and two of these SNPs, rs1382387 and rs4606077, failed the Hardy–Weinberg test. Another five SNPs (rs1005533, rs1498553, rs10092491, rs2272998 and rs1358856) deviated from Hardy–Weinberg expectations without any apparent reason. This may be due to random effects caused by the relatively small sample size.

In Fig. 2, the allele balance of all heterozygous genotype calls, except for the genotype calls from the eight poorly perming SNPs, is shown as a function of the number of reads. As expected, the allele balance varied more and the risk of allele drop-out was higher when the coverage was low. With 2.5–10 ng DNA in the PCR and a minimum coverage of 200 reads, the allele balance ranged from approximately 0.6–1.66, which is the commonly used boundaries for the heterozygote balance of STRs typed with PCR–CE [20].

3.3. Concordance study

The 44 Iraqi samples were previously typed with the SNP*for*ID assay [21] and for 45 Y-chromosome SNPs [22].

Except for two SNPs, rs1029047 and rs1031825, there was complete concordance between the results from the Ion Torrent™ HID SNP panel and the ISO17025 accreditated SNP*for*ID assay. The SNPs rs1029047 and rs1031825 were two of the eight SNPs identified as poorly performing SNPs. The sequence context surrounding rs1029047 and rs1031825 are ATTT[A/T]AAAAAAAAAC and CTAA[A/C]CCCCG, respectively. Both SNPs are located between homopolymer stretches and the two possible alleles are identical to the stretches on either side of the SNP. Therefore, misalignment of reads and wrong call of alleles may be expected. Another of the poorly performing SNPs, rs8037429, was also typed with the SNP*for*ID assay. However, all the genotype calls based on the Ion Torrent™ HID SNP panel were confirmed by the SNP*for*ID assay. Two SNPs that failed the Hardy–Weinberg test, rs1005533 and rs1382387, were typed with the SNP*for*ID assay. No discrepancies were observed between the assays, even though only one of the two homozygous genotypes for rs1382387 was observed among the Iraqis. In the previous study [21], where 101 unrelated Iraqis were typed, rs1005533 and rs1382387 were not in Hardy–Weinberg disequilibrium, which indicates that it was a random effect.

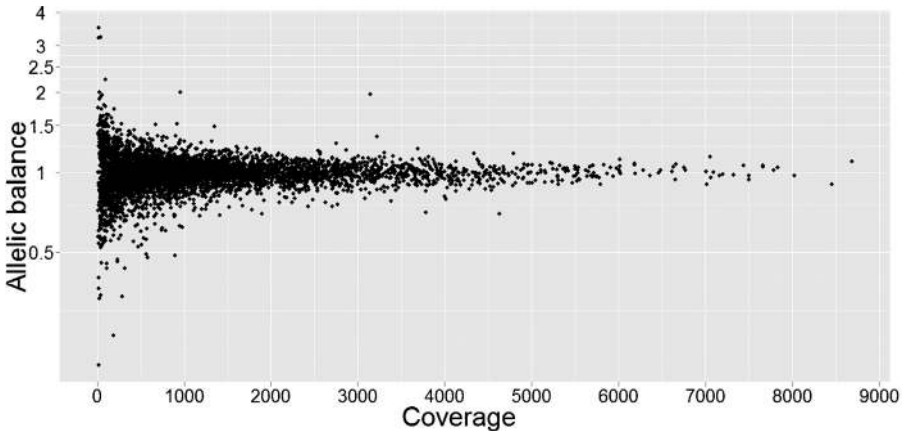

Fig. 2. Allele balances of heterozygous genotype calls from duplicate typing of 44 Iraqis. The results from the eight poorly performing SNPs were excluded.

Most of the Y-chromosome SNPs typed in the previous work [22] were not included in the Ion Torrent™ HID SNP panel. However, for each of the 44 Iraqi individuals, the Y-haplogroup designation was as expected based on the 32 Y-chromosome SNPs in the Ion Torrent™ HID SNP panel (Fig. 3. The Y-chromosome marker MSY2 was not taken into consideration (Table 1A) because MSY2 is a length polymorphism [23] and it was uncertain whether the HID_SNP_Genotyper (v2.2) plug-in was capable of typing MSY2 correctly. The HID_SNP_Genotyper (v2.2) plug-in reported the nucleotide in position 15,015,500 on the human Y-chromosome. This position is 55 bp upstream of the first repeat unit of MSY2 and it is, to our knowledge, not varying in humans.

3.4. Sensitivity study

Serial dilutions in the range 0.1–10 ng DNA of the two control samples were made and each dilution was typed twice with the Ion Torrent™ HID SNP panel using two Ion 316 Chips. The average call rates for each dilution are shown in Table 2 after the results from MSY2 and the eight poorly performing SNPs (see above) were excluded from the analysis and 19 errors by the HID_SNP_Genotyper (v2.2) plug-in were manually corrected. The number of locus drop-outs was high in experiments with less than 0.5 ng of input DNA. Nevertheless, almost 50% of the SNPs were correctly typed from only 0.1 ng of DNA. A total of 16 allele drop-outs were observed, all of them in experiments with less than 0.5 ng DNA. The coverage for these genotype calls was less than 50 reads (the average was 17.3 reads), and in all except one example, the allele that dropped out was detected in a small fraction of the sequences. However, the allele balance was skewed and the software ignored these reads. It was noteworthy that there was no drop-in in any of the samples.

The allele balance for the heterozygote SNPs are shown in Fig. 4. As expected, the allele balance varied more in experiments with low amounts of DNA. The optimal amount of DNA in the PCR seemed to be $\geq$0.5 ng.

3.5. SNP typing of mixtures

Eleven mixtures of the two control samples ranging from 1:1 to 1:1000 were made and each mixture was typed twice with the Ion Torrent™ HID SNP panel using two Ion 318 Chips. Fig. 5 shows the expected and observed allele balances for all SNP loci, where the genotypes of the two individuals differed. The results from the eight poorly performing SNPs and SNPs where $<$10 reads were observed for one of the alleles were omitted from the analysis. There was a clear linear correlation between the expected and observed allele balances ($R^2 = 0.984$) which indicated that the sequencing assay generated a loyal representation of the DNA sample.

The two control samples were homozygous for different alleles at 16 loci. Fig. 6 shows the number of reads from the minor and major contributor in the 1:10, 1:100 and 1:1000 mixtures for these SNPs. The number of reads from the minor contributor was sufficiently high to identify the 1:100 mixtures as mixtures when taking into account that the number of reads with nucleotide calls that differed from the SNP genotype call was typically less than five (Table 1B). It was also noteworthy that 66%, 6% and 0% of the Y-chromosome SNPs were detected in the 1:10, 1:100 and 1:1000 male:female mixtures, respectively.

4. Discussion

In this study, the performance of the Ion PGM™ and the Ion Torrent™ HID SNP panel was tested by studying SNP typing concordance with the SNP*for*ID assay, sensitivity and mixtures. In general, the sequencing performance of the platform was impressive. The allele balance for heterozygotes was close to 1:1 for almost all SNPs and the number of reads with nucleotide calls that differed from the SNP genotype call (base call errors) was typically less than five. In combination, this made it possible to identify mixtures of 1:100 with a high degree of certainty if the total depth of coverage for the sample was high. Consider a 1:100 mixture where the coverage for a given locus was 5000 reads and the major contributor was homozygous aa. Approximately 50 or 25 reads with the second allele, b, would be expected if the minor contributor was homozygous bb or heterozygous ab, respectively. This is more than the typical number of base call errors and since similar allele imbalances will be observed in many loci, it indicates that the sample is a mixture. Detection of mixtures with SNPs is possible by analysing the allele balance [15,24]. However, it is essentially impossible to deconvolute a mixture unless the precise mixture ratio is known, which is unlikely in real case work samples [15].

The optimal amount of input DNA in the Ion Torrent™ HID SNP assay was $>$0.5 ng. It seems likely that the sensitivity may be improved by further optimisation of the PCR or by removing some of the weakly amplified SNPs from the panel. However, the sensitivity study also showed that the allele balance suffered when the amount of input DNA was lower than 0.5 ng. The allele balances of eight (rs2399332, rs1029047, rs10776839, rs4530059, rs8037429, rs430046, rs1031825 and rs1523537) of the 169 markers were not consistent and these SNPs should be excluded

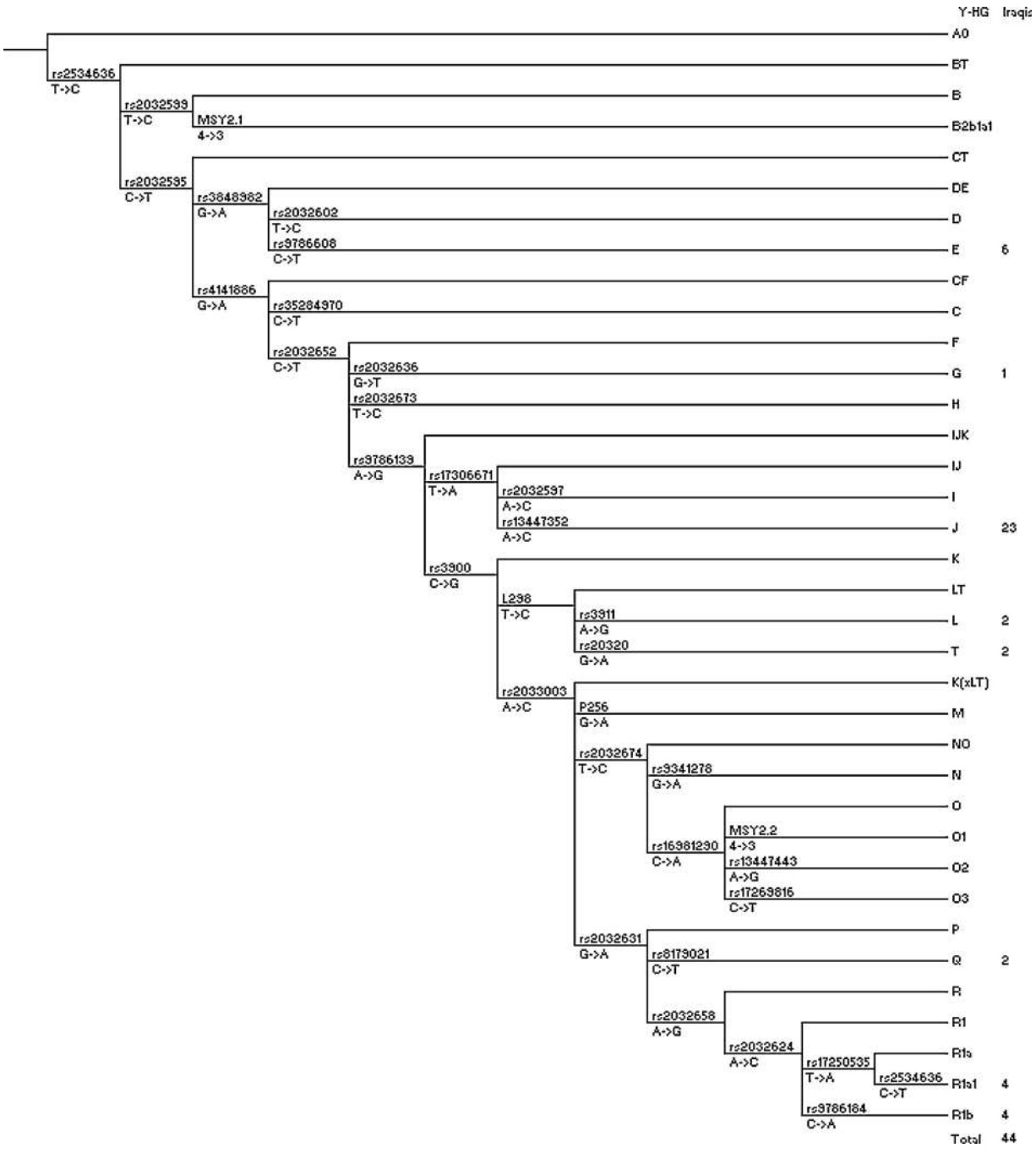

Fig. 3. Phylogenetic distribution of Y-chromosome haplogroups with the Ion Torrent™ HID SNP panel. The distribution of haplogroups in the Iraqi population is shown on the right.

from the panel. Also, the length polymorphism MSY2 seemed an odd choice for this panel and should be reconsidered.

The secondary sequence analysis software, the HID_SNP_Genotyper (v2.2) plug-in, sometimes interpreted the sequence data wrongly. It was hard to identify why that happened, although, the interpretations of some SNPs were more prone to errors than others. During the preparation of this paper, Life Technologies launched an update of the Torrent Suite Server and the variant caller function used by the HID_SNP_Genotyper (v4.0) plug-in. We re-analyzed the first run of the 44 Iraqi samples with the new software using the default settings. The number of sequence reads and base calls were not changed by the new analysis. However, all the genotype errors were corrected and only two new ones (error rate: 0.02%) were introduced. Both were heterozygote calls based on only 10 reads. According to the selected parameters for analysis, they should have been designated as no calls, because there were less than 6 forward and reverse reads.

Clearly, the new software is a major improvement. It illustrates the necessity of making comprehensive testing and validation of SGS softwares and it underlines the need for close collaboration between the forensic genetic community and the manufacturers. With SGS data, it is not possible to analyse the sequences manually or even to analyse the genotype calls manually. Therefore, the software solution(s) must be completely trustworthy and thoroughly validated before they can be used in real case work. A forensic genetic tool for analysis of SGS data should also have a biological aspect in addition to the bioinformatics approach used for analysis of most SGS data. For each locus, it should be possible

Table 2
Sensitivity study.

SNPs	DNA (ng/μL)	Number of locus dropouts	Number of allele dropouts	Correctly typed SNPs[a]
Autosomal	0.1	256	7	48.2%
Autosomal	0.2	73	8	83.9%
Autosomal	0.5	3	1	99.2%
Autosomal	1	3	0	99.4%
Autosomal	2	0	0	100%
Autosomal	5	0	0	100%
Autosomal	10	0	0	100%
Y	0.1	33	–	48.4%
Y	0.2	30	–	53.1%
Y	0.5	11	–	82.8%
Y	1	5	–	92.2%
Y	2	0	–	100%
Y	5	0	–	100%
Y	10	0	–	100%

[a] MSY2 and 8 poorly performing SNPs were excluded.

to define the expected allele balance, the expected fraction of reads for homozygous genotype calls, minimum thresholds for coverage and reads with no calls, and in the case of a homozygous genotype call, a minimum threshold for reads with base calls of the second known allele. If any of these criteria are not fulfilled, the software should warn the user that the result was unexpected. A forensic genetic tool may also be used to analyse the genotypes and report, e.g. the Y-chromosome haplogroup based on the Y-SNPs in the Ion Torrent™ HID SNP panel or e.g. report the expected eye colour based on the Iris Plex markers.

There are three different size Ion Chips (314, 316 and 318) for the Ion PGM™. We estimated that it was possible to run 4, 15 and 30 samples typed with the Ion Torrent™ HID SNP panel on the 314, 316 and 318 chips, respectively, and obtain sufficiently high coverage. The cost of the reagents will be 108, 68 and 59 Euro per sample, respectively. This is a relatively high price compared to the commercial STR kits or the PCR–SBE–CE SNP typing method (approximately three and seven times higher, respectively). On the other hand, the cost of an Ion PGM™ is only half the price of one CE-instrument.

The sequencing results presented in this work demonstrated that the Ion PGM™ is a very attractive platform for forensic genetics. If the problems with the secondary sequence analysis are solved and future panels perform as well as the Ion Torrent™ HID SNP panel, the Ion PGM™ may become a standard tool in many forensic laboratories. Obviously, the platform should also be able to type STRs and in order to develop an ideal panel for forensic genetics STRs and SNPs should be typed in the same reaction. The Ion Torrent™ HID SNP panel typed 136 autosomal SNPs and 33 Y-chromosome markers. The autosomal SNPs were collected from two independent selection panels [10,11] that each may provide mean matching probabilities in the 10^{-18} range, which is comparable to the mean matching probabilities of the commonly used STR kits with more than 15 STRs [25]. Furthermore, the Y-chromosome SNPs have limited use in forensic case work. Thus, a large number of the SNPs were redundant and may be replaced by other types of markers, e.g. ancestry informative markers, markers for phenotypical traits and, most importantly, the core STRs. A panel with this set of markers would be very attractive to the forensic genetic community and by combining different sets of markers, the higher price and lower sensitivity of the assay may be acceptable since more information is obtained in a single experiment.

In general, sequencing of PCR amplicons will most likely be the preferred SGS method for forensic genetic applications, because PCR ensures high sensitivity and speed, and because the number of

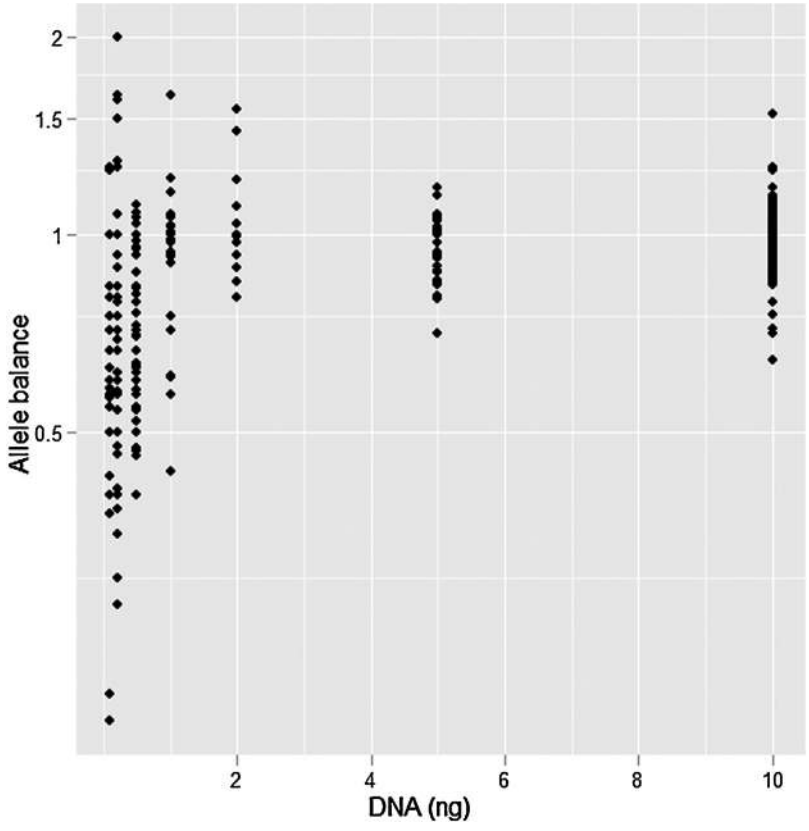

Fig. 4. Allele balances of heterozygous genotype calls from the sensitivity study. The results from the eight poorly performing SNPs were excluded.

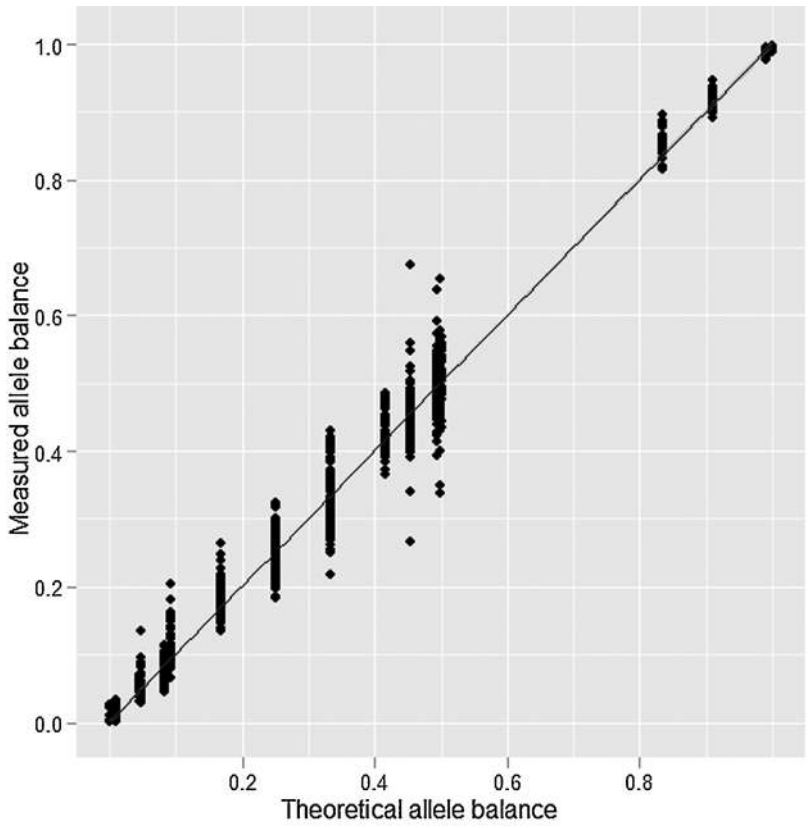

Fig. 5. Expected and observed allele balances for 1782 genotypes from the mixture study. The results from the 2:1 and 1:5 male/female mixtures were excluded in the final analysis because the two tubes with the mixtures were swabbed prior to PCR. The indicated line from the linear regression analysis intercepted on the *Y*-axis at 0.01, the slope was 0.998 and the correlations coefficient, $R^2 = 0.984$. The results from the eight poorly performing SNPs were excluded.

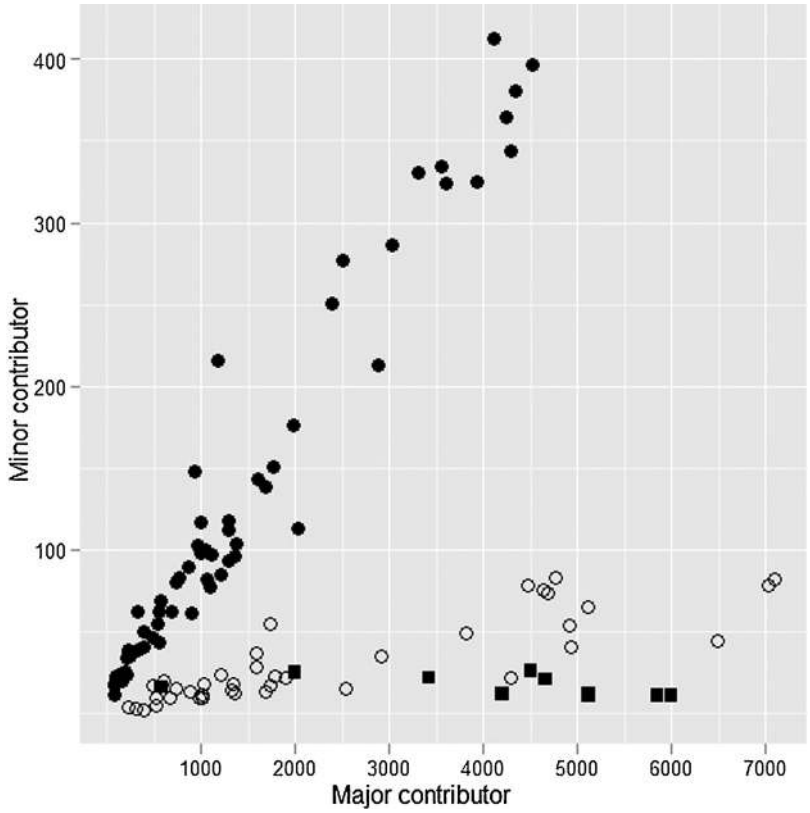

Fig. 6. The number of reads from the minor and major contributor from the 16 loci where the two individuals were homozygous for different alleles. Closed circles: 1:10 mixtures. Open circles: 1:100 mixtures. Closed squares: 1:1000 mixtures.

relevant loci will be limited to a few hundred. The output from the various SGS platforms ranges from 0.1 to 3000 million individual sequences (reads). Sequencing of a few hundred, relatively short (<200 bp) PCR products is a simple task even for the smallest platforms. Sequencing of e.g. 250 loci with an average coverage of 200 reads per locus in 100 individuals requires 5 million reads and will be possible with already available bench top machines. In contrast, shotgun sequencing of the entire genome or sequencing of targeted regions via probe capture consumes large amounts of DNA (10–250 ng), which is not applicable for investigations of the majority of case work samples.

Acknowledgements

We thank Ruune H. Larsen and Sarah N. Vitcetz for technical assistance. We thank Rune Frank-Hansen, Jeppe Dyrberg Andersen and Carmen Tomas for assistance with data analysis.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.fsigen.2014.06.004.

References

[1] S.L. Fordyce, M.C. Avila-Arcos, E. Rockenbauer, C. Børsting, R. Frank-Hansen, F.T. Petersen, E. Willerslev, A.J. Hansen, N. Morling, M.T. Gilbert, High-throughput sequencing of core STR loci for forensic genetic investigations using the Roche Genome Sequencer FLX platform, Biotechniques 51 (2011) 127–133.

[2] D.M. Bornman, M.E. Hester, J.M. Schuetter, M.D. Kasoji, A. Minard-Smith, C.A. Barden, S.C. Nelson, G.D. Godbold, C.H. Baker, B. Yang, J.E. Walther, I.E. Tornes, P.S. Yan, B. Rodriguez, R. Bundschuh, M.L. Dickens, B.A. Young, S.A. Faith, Short-read high-throughput sequencing technology for STR genotyping, Biotechniques (2012) 1–6. , http://dx.doi.org/10.2144/000113857.

[3] D.H. Warshauer, D. Lin, K. Hari, R. Jain, C. Davis, B. LaRue, J.L. King, B. Budowle, 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.

[4] C. Van Neste, M. Vandewoestyne, W. Van Criekinge, D. Deforce, F. Van Nieuwerburgh, My-Forensic-Loci-queries (MyFLq) framework for analysis of forensic STR data generated by massive parallel sequencing, Forensic Sci. Int. Genet. 9 (2014) 1–8.

[5] E. Rockenbauer, S. Hansen, M. Mikkelsen, C. Børsting, N. Morling, Characterization of mutations and sequence variants in the D21S11 locus by next generation sequencing, Forensic Sci. Int. Genet. 8 (2014) 68–72.

[6] S. Dalsgaard, E. Rockenbauer, A. Buchard, H.S. Mogensen, R. Frank-Hansen, C. Børsting, N. Morling, Non-uniform phenotyping of D12S391 resolved by second generation sequencing, Forensic Sci. Int. Genet. 8 (2014) 195–199.

[7] M. Mikkelsen, E. Rockenbauer, A. Wächter, L. Fendt, B. Zimmermann, W. Parson, S.A. Nielsen, T. Gilbert, E. Willerslev, N. Morling, Application of full mitochondrial geneome sequencing using 454 GS FLX pyrosequencing, Forensic Sci. Int. Genet. Suppl. 2 (2009) 518–519.

[8] W. Parson, C. Strobl, G. Huber, B. Zimmermann, S.M. Gomes, L. Souto, L. Fendt, R. Delport, R. Langit, S. Wootton, R. Lagacé, J. Irwin, Evaluation of next generation mtGenome sequencing using the ion Torrent Personal Genome Machine (PGM), Forensic Sci. Int. Genet. 7 (2013) 543–549.

[9] S.B. Seo, J.L. King, D.H. Warshauer, C.P. Davis, J. Ge, B. Budowle, Single nucleotide polymorphism typing with massively parallel sequencing for human identification, Int. J. Legal Med. 127 (2013) 1079–1086.

[10] J.J. Sanchez, C. Phillips, C. Børsting, K. Balogh, M. Bogus, M. Fondevila, C.D. Harrison, E. Musgrave-Brown, A. Salas, D. Syndercombe-Court, P.M. Schneider, A. Carracedo, N. Morling, A multiplex assay with 52 single nucleotide polymorphisms for human identification, Electrophoresis 27 (2006) 1713–1724.

[11] A.J. Pakstis, W.C. Speed, R. Fang, F.C. Hyland, M.R. Furtado, J.R. Kidd, K.K. Kidd, SNPs for a universal individual identification panel, Hum. Genet. 127 (2010) 315–324.

[12] T.M. Karafet, F.L. Mendez, M.B. Meilerman, P.A. Underhill, S.L. Zegura, M.F. Hammer, New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree, Genome Res. 18 (2008) 830–838.

[13] C. Børsting, E. Rockenbauer, N. Morling, Validation of a single nucleotide polymorphism (SNP) typing assay with 49 SNPs for forensic genetic testing in a laboratory accredited according to the ISO 17025 standard, Forensic Sci. Int. Genet. 4 (2009) 34–42.

[14] S. Walsh, A. Lindenbergh, S.B. Zuniga, T. Sijen, P. de Knijff, M. Kayser, K.N. Ballantyne, Developmental validation of the IrisPlex system: determination of blue and brown iris colour for forensic intelligence, Forensic Sci. Int. Genet. 5 (2011) 464–471.

[15] C. Børsting, M. Mikkelsen, N. Morling, Kinship analysis with diallelic SNPs – experiences with the SNPforID multiplex in an ISO17025 accreditated laboratory, Transfus. Med. Hemother. 39 (2012) 195–201.

[16] T. Schwark, P. Meyer, M. Harder, J.H. Modrow, N. von Wurmb-Schwark, The SNPforID assay as a supplementary method in kinship and trace analysis, Transfus. Med. Hemother. 39 (2012) 187–193.

[17] C. Børsting, C. Tomas, N. Morling, Typing of 49 autosomal SNPs by single base extension and capillary electrophoresis for forensic genetic testing, Methods Mol. Biol. 830 (2012) 87–107.

[18] E. Musgrave-Brown, D. Ballard, K. Balogh, K. Bender, B. Berger, M. Bogus, C. Børsting, M. Brion, M. Fondevila, C. Harrison, C. Oguzturun, W. Parson, C. Phillips, C. Proff, E. Ramos-Luis, J.J. Sanchez, D.P. Sanchez, R.B. Sobrino, B. Stradmann-Bellinghausen, C. Thacker, A. Carracedo, N. Morling, R. Scheithauer, P.M. Schneider, D. Syndercombe Court, Forensic validation of the SNPforID 52-plex assay, Forensic Sci. Int. Genet. 1 (2007) 186–190.

[19] J.J. Sanchez, C. Børsting, K. Balogh, B. Berger, M. Bogus, J.M. Butler, A. Carracedo, D.S. Court, L.A. Dixon, B. Filipovic, M. Fondevila, P. Gill, C.D. Harrison, C. Hohoff, R. Huel, B. Ludes, W. Parson, T.J. Parsons, E. Petkovski, C. Phillips, H. Schmitter, P.M. Schneider, P.M. Vallone, N. Morling, Forensic typing of autosomal SNPs with a 29 SNP-multiplex—results of a collaborative EDNAP exercise, Forensic Sci. Int. Genet. 2 (2008) 176–183.

[20] J. Bright, J. Turlington, J. Buckleton, Examination of the variability in mixed DNA profile parameters for the Identifiler™ multiplex, Forensic Sci. Int. Genet. 4 (2010) 111–114.

[21] C. Tomas, I.E. Diez, E. Moncada, C. Børsting, N. Morling, Analysis of 49 autosomal SNPs in an Iraqi population, Forensic Sci. Int. Genet. 1 (2013) 198–199.

[22] J.J. Sanchez, C. Hallenberg, C. Børsting, A. Hernandez, N. Morling, High frequencies of Y chromosome lineages characterized by E3b1, DYS19-11, DYS392-12 in Somali males, Eur. J. Hum. Genet. 13 (2005) 856–866.

[23] W. Bao, S. Zhu, A. Pandya, T. Zerjal, J. Xu, Q. Shu, R. Du, H. Yang, C. Tyler-Smith, MSY2: a slowly evolving minisatellite on the human Y chromosome which provides a useful polymorphic marker in Chinese populations, Gene 244 (2000) 29–33.

[24] C. Børsting, H.S. Mogensen, N. Morling, Forensic genetic SNP typing of low-template DNA and highly degraded DNA from crime case samples, Forensic Sci. Int. Genet. 7 (2013) 345–352.

[25] R.L. Green, R.E. Lagacé, N.J. Oldroyd, L.K. Hennessy, J.J. Mulero, Developmental validation of the AmpFℓSTR® NGM SElect™ PCR Amplification Kit: a next-generation STR multiplex with the SE33 locus, Forensic Sci. Int. Genet. 7 (2013) 41–51.