DOI:
10.1016/j.fsigen.2017.10.012

Document Version
Peer reviewed version

Title: Concordance of the ForenSeq™ system and characterisation of sequence-specific autosomal STR alleles across two major population groups

Authors: Laurence Devesse, David Ballard, Lucinda Davenport, Immy Riethorst, Gabriella Mason-Buck, Denise Syndercombe Court

PII: S1872-4973(17)30224-7
DOI: https://doi.org/10.1016/j.fsigen.2017.10.012
Reference: FSIGEN 1800

To appear in: *Forensic Science International: Genetics*

Received date: 24-7-2017
Revised date: 23-10-2017
Accepted date: 31-10-2017

Please cite this article as: Laurence Devesse, David Ballard, Lucinda Davenport, Immy Riethorst, Gabriella Mason-Buck, Denise Syndercombe Court, Concordance of the ForenSeq™ system and characterisation of sequence-specific autosomal STR alleles across two major population groups, Forensic Science International: Genetics https://doi.org/10.1016/j.fsigen.2017.10.012

Concordance of the ForenSeq™ system and characterisation of sequence-specific autosomal STR alleles across two major population groups

Laurence Devesse, David Ballard, Lucinda Davenport, Immy Riethorst, Gabriella Mason-Buck and Denise Syndercombe Court

King's Forensics, Faculty of Life Sciences and Medicine, King's College London, 150 Stamford Street, London SE1 9NH

David.ballard@kcl.ac.uk
King's Forensics
Faculty of Life Sciences and Medicine
King's College London
150 Stamford Street
London
SE1 9NH
United Kingdom

Abstract

By using sequencing technology to genotype loci of forensic interest it is possible to simultaneously target autosomal, X and Y STRs as well as identity, ancestry and phenotypic informative SNPs, resulting in a breadth of data obtained from a single run that is considerable when compared to that generated with standard technologies. It is important however that this information aligns with the genotype data currently obtained using commercially available kits for CE-based investigations such that results are compatible with existing databases and hence can be of use to the forensic community. In this work, 400 samples were typed using commercially available STR kits and CE, as well as using the Ilumina ForenSeq™ DNA Signature Prep Kit and MiSeq® FGx to assess concordance of autosomal STRs and population variability. Results show a concordance rate between the two technologies exceeding 99.98% while numerous novel sequence based alleles are described. In order to make use of the sequence variation observed, sequence specific allele frequencies were generated for White British and British Chinese populations.

Key words: Massively Parallel Sequencing, Concordance, STRs

1. Introduction

Since the 1990's, human identification in forensic genetics has relied almost exclusively on the use of multiplex autosomal short tandem repeats (STRs) and capillary electrophoresis (CE) [1]. In recent years, massively parallel sequencing (MPS, also referred to as next generation sequencing or NGS) has emerged as the technique of choice to overcome some of the limitations of CE-based approaches. For example,

with CE, the number of markers that can be analysed simultaneously is limited by size-based separation and fluorescent dye detection restrictions, and therefore PCR amplicons must be designed in a way that ensure they can be distinguished from one another. As a result of this, most commercial assays target amplicons that range between 80 and 500 bp [2-4].

The ability to simultaneously target many shorter amplicons is of benefit when analysing degraded samples, as it increases both the chance of obtaining information from fragmented material and the power of discrimination when attempting to identify a sample. This power of discrimination is also improved through the increased number of alleles observed using MPS, where sequence information can be used to differentiate alleles of the same size but differing in sequence [5-14].

Before MPS can be implemented into routine casework, the back compatibility of the method against currently used CE-based techniques must be checked in order to ensure the ongoing utility of currently available offender databases. New frequency databases must also be generated to make use of the sequence-variant alleles observed.

In this study, genotypes obtained using a traditional autosomal STR kit and CE were compared to MPS results for 400 samples. The Ilumina ForenSeq™ DNA Signature Prep Kit (Illumina, San Diego, CA) targets 27 autosomal STRs, 7 X-STRs, 24 Y-STRs and 94 Single Nucleotide Polymorphisms (SNPs) simultaneously, and its performance has been evaluated in several publications [15-19]. The MiSeq® FGx was used to generate autosomal STR sequence frequencies for two UK-relevant population groups (White British, n=200 and British Chinese, n=200), which were subsequently checked against genotypes obtained using size-based methods alone. All sequence-based alleles, once characterized, were compared to previous publications and online databases [11-14, 19-25] to identify novel variants.

2. Materials and methods

2.1. Samples

Buccal swab extracts from 400 unrelated individuals that had previously been analysed using commercially available STR kits and CE were chosen for the study. Individuals gave informed consent for their DNA to be used for research purposes and ethical approval for this work was granted by the King's BDM research ethics subcommittee (HR-16/17-2594). DNA had been previously extracted using either Chelex (Sigma-Aldrich, St Louis, USA) or the DNA Investigator Kit (Qiagen, Heidelberg, Germany).

2.2. Library preparation and Sequencing

The Illumina ForenSeq™ DNA Signature Prep Kit [26] was used to prepare the samples for sequencing. An initial two-step PCR reaction was performed to amplify and tag the regions of interest, followed by purification and normalization of the libraries. Primer mix A was used for the first PCR reaction, which contains primers for identity markers including 27 autosomal STRs. A negative and positive amplification control were processed with each run, resulting in a total of 94 samples that could be prepared in one 96-well plate simultaneously. The libraries were prepared according to manufacturer's instruction [26].

The libraries were pooled in batches of 96 and denatured before being loaded into a MiSeq® FGx Reagent Cartridge. The only protocol modification implemented was to increase the volume of pooled libraries from 7 µl to 12 µl. This volume was decided based on internal laboratory validation data, and has been shown to yield better results than the recommended input amount. Denatured human

sequencing control (HSC) was added to the pooled libraries, before dilution and loading into the appropriate well in the cartridge.

The cartridge containing the libraries and sequencing reagents was loaded onto the MiSeq® FGx instrument alongside a flowcell and incorporation buffer according to standard protocol [27].

2.3. Concordance testing

Investigations were performed on the same samples using the STR markers contained within the GlobalFiler® Express kit (Applied Biosystems, Foster City, USA) as per manufacturer's guidelines [2]. 1ng input DNA was used for amplification, and injection was performed at 1.2 kV for 23 seconds on the AB Prism 3130xl Genetic Analyzer (Applied Biosystems) for separation and detection of autosomal STR loci. A detection threshold of 50 relative fluorescence units (RFU) was imposed during data analysis using GeneMapper®ID v3.2 software (Applied Biosystems).

2.4. Additional sequencing

In order to investigate the cause of a null allele at the D5S818 locus, new primers were designed outside the range of the ForenSeq™ amplicon (D5S818-1: tcccatctggatagtggacct, D5S818-2: gcttctaattaaagtggtgtccca). PCR was carried out using the KAPA Multiplex PCR Mix (Kapa Biosystems, Wilmington, USA) with a final primer concentration of 0.3 µM, the addition of 1 ng of genomic template DNA and cycling conditions consisting of 94°C for 3 minutes followed by 28 cycles of 94°C for 20 seconds/ 58°C for 30 seconds/ 72°C for 60 seconds, with a final extension of 72°C for 3 minutes. PCR products were sequenced on the MiSeq® FGx in RUO mode, using a MiSeq v2 300 cycle (Illumina) cartridge following library preparation with the KAPA Hyper Prep Kit for Illumina Platforms (Roche, Basel, Switzerland) according to manufacturer's guidelines, using TruSeq indexes (Illumina).

2.5. Universal Analysis Software

The ForenSeq™ Universal Analysis Software (UAS) was used to create a sample sheet containing the necessary information for demultiplexing the samples based on their index combination, and for subsequent analyses [28]. The software calls alleles based on read counts, and makes use of an analytical and interpretation threshold, which are determined as a percentage of the total number of reads per locus. The thresholds used were 1.5% and 4.5% for the analytical and interpretation thresholds, respectively, taken from the developmental validation conducted by Illumina [29]. The software also contains lower limits for both thresholds of 10 and 30 reads. For the purpose of this study, and as samples were all known to be single-source, all alleles above the analytical threshold were manually called and also used for concordance and allele frequencies. Once all results were analysed in UAS, sample details including allele calls and sequences were exported in Microsoft Office Excel format.

2.6. Data analysis

Length based allele calls from UAS for 20 autosomal loci and Amelogenin were compared to CE results using in-house Microsoft Office Excel workbooks, and any discordant results were recorded and taken for further analysis. Excel was also used to calculate allele frequencies for both population groups. Where necessary, raw data was re-analysed by exporting the FASTQ files, aligning to a bespoke reference

genome containing the STR reference sequence of interest using the mem algorithm within BWA (http://bio-bwa.sourceforge.net/), and visualised using the Integrative Genomics Viewer (IGV) [30]. Discordance was defined as any instance where an allele observed using one technique was not observed with the other. Exceptions were made in the case of heterozygote alleles, where one allele was seen below threshold and the other had dropped out. Here, the discrepancy was attributed to drop out rather than discordance between kits.

2.6.1. IDENTIFICATION OF SEQUENCE VARIANTS

Allelic sequence variants for autosomal STRs were detected using Excel workbooks, and characterised using nomenclature described in the literature [13, 19], and following ISFG recommendations where possible [31]. Each allele was designated using the default output from UAS and therefore limited reported flanking region information was used, as well as the strand direction utilised by the software. For the following markers, a small amount of flanking region sequence directly adjacent to the repeat region is reported: D13S317, D18S51, D19S433, D1S1656, D5S818, D7S820, vWA.

2.6.2. ALLELE FREQUENCIES

Allele frequencies are given in Table 1. Arlequin software was used to check for non-random association of alleles by testing whether genotypes for each STR marker conformed to Hardy-Weinberg equilibrium (HWE) [32], applying a Bonferroni correction for multiple comparisons (see supplementary materials).

3. Results and discussion

3.1. ALLELES OBTAINED BY SEQUENCE

The number of individual alleles obtained by length was compared to those obtained by sequence, as shown in Figure 1. Samples were split according to population groups (White British, n= 200 and British Chinese, n=200) as variation in marker discrimination is expected between groups. All sequence-based alleles were characterized and are reported in Table 1. Where an allele has not previously been observed [11-14, 19-25], it is highlighted in red and in bold. D12S391 is the most highly polymorphic autosomal STR within the group of markers studied in this work, as demonstrated in Figure 1 and Table 1; 48 additional alleles were seen using sequencing compared to size-based separation (across both population groups).

Eight out of the twenty-seven autosomal loci targeted showed no gain in the number of alleles seen using sequence information (Penta D, Penta E, D22S1045, D16S539, TPOX, TH01, D10S1248, and D19S433). These results differ slightly from the results published by Gettings *et. al*. [12], where increased variation was seen at D10S1248, Penta E, and D19S433, although no sequence variation was observed in their study at D7S820 and D13S317. Novroski *et. al*. [13] reported sequence variation at all autosomal loci except TPOX. This is likely due to the number of samples and different populations investigated. Initial characterisation of alleles in other population groups has revealed variation at Penta E and D19S433 for example. Three loci showed sequence variation in the British Chinese population which was not observed in the White British samples (D7S820, D17S1301 and D20S482), and one locus showed variation only in the White British population (D18S51). This data identifies specific mutations causing sequence variation restricted within a population group.

The limited flanking region information reported by the software increased the diversity of alleles observed at two markers. Alleles characterized for D13S317 include both repeat region variants and flanking region variants, whereas all variation seen at D5S818 was found within the flanking region (Table 1).

3.2. Autosomal STR frequencies

The frequencies for each sequence-based allele across the two population groups studied are given in Table 1. No loci deviated significantly from HWE expectations.

STR	Allele	Bracket sequence	Frequencies	
			White British	British Chinese
CSF1PO*	7	[AGAT]7		0.0025
	8	[AGAT]8	0.0024	
	9	[AGAT]9	0.0192	0.0371
	10	[AGAT]10	0.2644	0.2426
	11	[AGAT]11	0.2957	0.2401
	12	[AGAT]12	0.3293	0.3886
	12	[AGAT]6[AGAC][AGAT]5		0.0050
	13	[AGAT]13	0.0721	0.0767
	14	[AGAT]14	0.0168	0.0050
	15	[AGAT]15		0.0025
D10S1248	8	[GGAA]8	0.0025	
	9	[GGAA]9		0.0024
	11	[GGAA]11	0.0048	0.0025
	12	[GGAA]12	0.0192	0.0767
	13	[GGAA]13	0.2788	0.3490
	14	[GGAA]14	0.3149	0.2203
	15	[GGAA]15	0.1899	0.2178
	16	[GGAA]16	0.1442	0.1114
	17	[GGAA]17	0.0457	0.0149
	18	[GGAA]18		0.0050
D12S391	15	[AGAT]8[AGAC]6AGAT	0.0291	0.0101
	16	[AGAT]8[AGAC]7AGAT		0.0025
	16	[AGAT]9[AGAC]6AGAT	0.0339	
	17	[AGAT]9[AGAC]7AGAT	0.0024	
	17	[AGAT]10[AGAC]6AGAT	0.1138	0.1086
	17	[AGAT]11[AGAC]5AGAT	0.0048	0.0025
	17.3	[AGAT]GAT[AGAT]8[AGAC]7AGAT	0.0169	
	18	[AGAT]10[AGAC]8	0.0024	
	18	[AGAT]10[AGAC]7AGAT	0.0024	0.0152
	18	[AGAT]11[AGAC]7		0.0101
	18	[AGAT]11[AGAC]6AGAT	0.1719	0.2121
	18	[AGAT]12[AGAC]5AGAT	0.0073	
	18	[AGAT]13[AGAC]4AGAT		0.0025
	18.3	[AGAT]GAT[AGAT]9[AGAC]7AGAT	0.0145	
	19	**[AGAC][AGAT]11[AGAC]6AGAT**		0.0025
	19	[AGAT]10[AGAC]8AGAT		0.0025
	19	[AGAT]11[AGAC]8		0.0101

	19	[AGAT]11[AGAC]7AGAT	0.0242	0.0253
	19	[AGAT]12[AGAC]6AGAT	0.0993	0.1364
	19	[AGAT]13[AGAC]5AGAT		0.0076
	19.3	[AGAT]5GAT[AGAT]7[AGAC]6AGAT	0.0048	
	19.3	[AGAT]GAT[AGAT]10[AGAC]7AGAT	0.0073	
	20	[AGAT]11[AGAC]9	0.0218	0.0051
	20	[AGAT]11[AGAC]8AGAT	0.0024	0.0429
	20	[AGAT]12[AGAC]8	0.0073	0.0051
	20	[AGAT]12[AGAC]7AGAT	0.0194	0.0404
	20	[AGAT]13[AGAC]7	0.0024	
	20	[AGAT]13[AGAC]6AGAT	0.0436	0.0808
	20	[AGAT]14[AGAC]5AGAT	0.0024	0.0025
	20.3	**[AGAT]3GAT[AGAT]10[AGAC]6AGAT**	0.0024	
	21	[AGAT]11[AGAC]10	0.0121	0.0025
	21	[AGAT]11[AGAC]9AGAT	0.0024	0.0025
	21	[AGAT]12[AGAC]9	0.0508	0.0303
	21	[AGAT]12[AGAC]8AGAT		0.0556
	21	[AGAT]13[AGAC]8	0.0194	0.0025
	21	[AGAT]13[AGAC]7AGAT	0.0121	0.0152
	21	[AGAT]14[AGAC]6AGAT		0.0227
	21	**[AGAT]14[AGAC]5[AGAT]2**	0.0024	
	21	[AGGT][AGAT]11[AGAC]9	0.0024	
	22	[AGAT]11[AGAC]10AGAT	0.0024	
	22	[AGAT]12[AGAC]10	0.0145	0.0101
	22	[AGAT]12[AGAC]9AGAT	0.0024	0.0025
	22	[AGAT]13[AGAC]9	0.0726	0.0379
	22	[AGAT]13[AGAC]8AGAT	0.0145	0.0227
	22	[AGAT]14[AGAC]8	0.0169	
	22	[AGAT]14[AGAC]7AGAT	0.0048	0.0126
	22	**[AGGT][AGAT]13[AGAC]7AGAT**	0.0024	
	23	[AGAT]12[AGAC]11	0.0024	
	23	[AGAT]12[AGAC]10AGAT	0.0024	
	23	[AGAT]13[AGAC]10	0.0145	0.0051
	23	[AGAT]13[AGAC]9AGAT	0.0048	0.0051
	23	[AGAT]14[AGAC]9	0.0339	0.0177
	23	[AGAT]14[AGAC]8AGAT	0.0218	0.0076
	23	[AGAT]15[AGAC]8	0.0048	
	24	[AGAT]14[AGAC]10	0.0024	0.0025
	24	[AGAT]14[AGAC]9AGAT	0.0024	0.0025
	24	[AGAT]15[AGAC]9	0.0121	0.0076
	24	[AGAT]15[AGAC]8AGAT	0.0024	0.0051

	25	[AGAT]15[AGAC]10		0.0025
	25	[AGAT]15[AGAC]9AGAT	0.0024	
	25	[AGAT]16[AGAC]9	0.0024	
	25	[AGAT]16[AGAC]8AGAT	0.0073	0.0025
	26	[AGAT]16[AGAC]10	0.0048	
	26	[AGAT]17[AGAC]9	0.0073	
	28	**[AGGT][AGAT]18[AGAC]8AGAT**	0.0024	

D13S317	7	[TATC]7 [AATC]2		0.0025
	7	[TATC]7 [TATC][AATC]	0.0024	
	8	[TATC]8 [AATC]2	0.1202	0.3441
	9	[TATC]9 [AATC]2	0.0649	0.1213
	9	[TATC]9 [TATC][AATC]		0.0198
	10	[TATC]10 [AATC]2	0.0433	0.0347
	10	[TATC]10 [TATC][AATC]	0.0048	0.0842
	10	[TATC]10 [TATC]2		0.0124
	11	[TATC]11 [AATC]2	0.1490	0.0248
	11	[TATC]11 [TATC][AATC]	0.1851	0.1559
	11	[TATC]11 [TATC]2		0.0173
	11	[TATC]8[TGTC][TATC]2 [TATC][AATC]		0.0124
	12	**[TATC]5[TAAC][TATC]6 [TATC][AATC]**	0.0024	
	12	[TATC]12 [AATC]2	0.1731	0.0198
	12	[TATC]12 [TATC][AATC]	0.1178	0.1114
	12	[TATC]12 [TATC]2		0.0074
	13	[TATC]13 [AATC]2	0.0649	
	13	[TATC]13 [TATC][AATC]	0.0264	0.0223
	14	[TATC]14 [AATC]2	0.0337	
	14	[TATC]14 [TATC][AATC]	0.0120	0.0099

D16S539	8	[GATA]8	0.0096	0.0124
	9	[GATA]9	0.1154	0.2475
	10	[GATA]10	0.0649	0.1287
	11	[GATA]11	0.2909	0.2723
	12	[GATA]12	0.3221	0.2178
	13	[GATA]13	0.1707	0.1015
	14	[GATA]14	0.0216	0.0124
	15	[GATA]15		0.0050
	16	[GATA]16	0.0048	0.0025

D17S1301	7	[AGAT]7		0.0025
	8	[AGAT]8		0.0074

	9	[AGAT]9	0.0024	0.0248
	10	[AGAT]10	0.0313	0.0545
	11	[AGAT]11	0.3053	0.1881
	12	[AGAT]12	0.4736	0.4356
	12	[AGAT]11[CGAT]		0.0050
	13	[AGAT]13	0.1563	0.2351
	14	[AGAT]14	0.0313	0.0421
	15	[AGAT]15		0.0050

D18S51	10	[AGAA]10 AAAG AGAG AG	0.0024	
	11	[AGAA]11 AAAG AGAG AG	0.0096	
	12	[AGAA]12 AAAG AGAG AG	0.1466	0.0446
	13	[AGAA]13 AAAG AGAG AG	0.1298	0.1584
	14	[AGAA]14 AAAG AGAG AG	0.1538	0.1931
	14	[AGAA][AGCA][AGAA]12 AAAG AGAG AG	0.0048	
	15	[AGAA]15 AAAG AGAG AG	0.1490	0.2005
	16	[AGAA]16 AAAG AGAG AG	0.1298	0.1436
	17	[AGAA]17 AAAG AGAG AG	0.1322	0.0569
	18	[AGAA]18 AAAG AGAG AG	0.0745	0.0619
	18.1	**[AGAA]14[AAAG] AGAG AGGA A[AGAA]** AAAG AGAG AG	0.0024	
	19	[AGAA]19 AAAG AGAG AG	0.0288	0.0371
	20	[AGAA]20 AAAG AGAG AG	0.0120	0.0198
	21	[AGAA]21 AAAG AGAG AG	0.0144	0.0149
	22	[AGAA]22 AAAG AGAG AG	0.0048	0.0371
	23	[AGAA]23 AAAG AGAG AG	0.0024	0.0248
	24	[AGAA]24 AAAG AGAG AG	0.0024	0.0074

D19S433*	4	[AAGG][AAAG][AAGG][TAGG][AAGG]2 AGAG AGGA AGAA AGAG AG		0.0025
	12	[AAGG][AAAG][AAGG][TAGG][AAGG]10 AGAG AGGA AGAA AGAG AG	0.0649	0.0347
	12.1	[AAGG][AAAG][AAGG][TAGG][AAGG]5A[AAGG]5 AGAG AGGA AGAA AGAG AG	0.0024	
	12.2	[AAGG][AA][AAGG][TAGG][AAGG]11 AGAG AGGA AGAA AGAG AG		0.0099
	13	[AAGG][AAAG][AAGG][TAGG][AAGG]11 AGAG AGGA AGAA AGAG AG	0.2524	0.2921
	13.2	[AAGG][AA][AAGG][TAGG][AAGG]12 AGAG AGGA AGAA AGAG AG	0.0120	0.0446
	14	[AAGG][AAAG][AAGG][TAGG][AAGG]12 AGAG AGGA AGAA AGAG AG	0.3702	0.2500
	14.2	[AAGG][AA][AAGG][TAGG][AAGG]13 AGAG AGGA AGAA AGAG AG	0.0433	0.1139
	15	[AAGG][AAAG][AAGG][TAGG][AAGG]13 AGAG AGGA AGAA AGAG AG	0.1683	0.0990
	15.2	[AAGG][AA][AAGG][TAGG][AAGG]14 AGAG AGGA AGAA AGAG AG	0.0313	0.1213
	16	[AAGG][AAAG][AAGG][TAGG][AAGG]14 AGAG AGGA AGAA AGAG AG	0.0385	0.0074
	16.2	[AAGG][AA][AAGG][TAGG][AAGG]15 AGAG AGGA AGAA AGAG AG	0.0096	0.0248
	17	[AAGG][AAAG][AAGG][TAGG][AAGG]15 AGAG AGGA AGAA AGAG AG	0.0048	
	18	[AAGG][AAAG][AAGG][TAGG][AAGG]16 AGAG AGGA AGAA AGAG AG	0.0024	

D1S1656*	10	[TAGA]9 TAGG [TG]5	0.0025	
	11	[TAGA]9 TAGG [TG]5	0.0896	0.0693
	12	[TAGA]12[TG]5	0.0498	0.0269
	12	[TAGA]11 TAGG [TG]5	0.0896	0.0202
	13	[TAGA]13[TG]5	0.0249	0.0480
	13	[TAGA]12 TAGG [TG]5	0.0224	0.0510
	13	[TAGA]11 TAGC TAGA [TG]5	0.0025	
	14	[TAGA]14[TG]5	0.0025	0.0169
	14	[TAGA]13 TAGG [TG]5	0.0746	0.0574
	14.3	[TAGA]2 TGA [TAGA]11 TAGG [TG]5	0.0025	
	15	[TAGA]14 TAAG [TG]5	0.0026	
	15	[TAGA]15[TG]5	0.0104	0.0060
	15	[TAGA]14 TAGG [TG]5	0.1089	0.3306
	15.3	[TAGA]3 TGA [TAGA]11 TAGG [TG]5	0.0269	0.0025
	15.3	[TAGA]4 TGA [TAGA]10 TAGG [TG]5	0.0403	
	16	[TAGA]15 TAAG [TG]5	0.0127	
	16	[TAGA]16 [TG]5	0.0025	
	16	[TAGA]15 TAGG [TG]5	0.0867	0.1955
	16.1	**[TAGA]15 [TAAG] G** [TG]5		0.0025
	16.3	[TAGA]4 TGA [TAGA]11 TAGG [TG]5	0.0697	0.0074
	17	[TAGA]16 TAGG [TG]5	0.0597	0.0693
	17.3	[TAGA]4 TGA [TAGA]12 TAGG [TG]5	0.1517	0.0693
	18	[TAGA]17 TAGG [TG]5	0.0050	0.0050
	18.3	[TAGA]4 TGA [TAGA]13 TAGG [TG]5	0.0498	0.0223
	19.3	[TAGA]4 TGA [TAGA]14 TAGG [TG]5	0.0100	
	20.3	[TAGA]4 TGA [TAGA]15 TAGG [TG]5	0.0025	

D20S482	9	[AGAT]9	0.0168	
	10	[AGAT]10	0.0024	0.0223
	11	[AGAT]11	0.0144	0.0074
	12	[AGAT]12	0.0120	0.0347
	13	[AGAT]13	0.2091	0.2970
	13	[AGAT]12 AGCT		0.0025
	14	[AGAT]14	0.4688	0.4059
	15	[AGAT]15	0.2019	0.1856
	16	[AGAT]16	0.0697	0.0396
	16	**[ACAT] [AGAT]15**		0.0025
	17	[AGAT]17	0.0048	
	19	[AGAT]19		0.0025

D2 1S1	26	[TCTA]4[TCTG]6[TCTA]3TA[TCTA]3TCA[TCTA]2TCCATA[TCTA]8	0.0024	

	27	[TCTA]6[TCTG]5[TCTA]3TA[TCTA]3TCA[TCTA]2TCCATA[TCTA]8	0.0072	0.0025
	27	[TCTA]4[TCTG]6[TCTA]3TA[TCTA]3TCA[TCTA]2TCCATA[TCTA]9	0.0242	
	28	[TCTA]6[TCTG]5[TCTA]3TA[TCTA]3TCA[TCTA]2TCCATA[TCTA]9	0.0024	0.0226
	28	**[TCTA]5[TCTG]5[TCTA]3TA[TCTA]2TCA[TCTA]2TCCATA[TCTA]11**		0.0075
	28	[TCTA]5[TCTG]6[TCTA]3TA[TCTA]3TCA[TCTA]2TCCATA[TCTA]9		0.0075
	28	[TCTA]4[TCTG]6[TCTA]3TA[TCTA]3TCA[TCTA]2TCCATA[TCTA]10	0.1594	0.0201
	28.2	[TCTA]5[TCTG]6[TCTA]3TA[TCTA]3TCA[TCTA]2TCCATA[TCTA]8TATCTA		0.0025
	29	[TCTA]6[TCTG]5[TCTA]3TA[TCTA]3TCA[TCTA]2TCCATA[TCTA]10	0.0725	0.1729
	29	[TCTA]5[TCTG]5[TCTA]3TA[TCTA]2TCA[TCTA]2TCCATA[TCTA]12		0.0050
	29	[TCTA]5[TCTG]6[TCTA]3TA[TCTA]2TCA[TCTA]2TCCATA[TCTA]11		0.0025
	29	[TCTA]5[TCTG]6[TCTA]3TA[TCTA]3TCA[TCTA]2TCCATA[TCTA]10	0.0072	0.0075
	29	[TCTA]4[TCTG]6[TCTA]3TA[TCTA]3TCA[TCTA]2TCCATA[TCTA]11	0.1401	0.0952
	29.2	[TCTA]5[TCTG]6[TCTA]3TA[TCTA]3TCA[TCTA]2TCCATA[TCTA]9TATCTA		0.0025
	30	[TCTA]7[TCTG]5[TCTA]3TA[TCTA]3TCA[TCTA]2TCCATA[TCTA]10	0.0048	0.0226
	30	[TCTA]6[TCTG]5[TCTA]3TA[TCTA]3TCA[TCTA]2TCCATA[TCTA]11	0.1353	0.1278
	30	**[TCTA]5[TCTG]6[TCTA]2TA[TCTA]3TCA[TCTA]2TCCATA[TCTA]12**		0.0025
	30	[TCTA]5[TCTG]6[TCTA]3TA[TCTA]3TCA[TCTA]2TCCATA[TCTA]11	0.0314	0.0627
	30	[TCTA]4[TCTG]6[TCTA]3TA[TCTA]3TCA[TCTA]2TCCATA[TCTA]12	0.0773	0.0526
	30	[TCTA]4[TCTG]7[TCTA]3TA[TCTA]3TCA[TCTA]2TCCATA[TCTA]11	0.0024	
	30.2	[TCTA]5[TCTG]5[TCTA]3TA[TCTA]3TCA[TCTA]2TCCATA[TCTA]11TATCTA	0.0217	
	30.2	[TCTA]5[TCTG]6[TCTA]3TA[TCTA]3TCA[TCTA]2TCCATA[TCTA]10TATCTA	0.0048	0.0075
	30.3	[TCTA]6[TCTG]5[TCTA]3TA[TCTA]3TCA[TCTA]2TCCATA[TCTA]5TCA[TCTA]6		0.0025
	31	[TCTA]8[TCTG]5[TCTA]3TA[TCTA]3TCA[TCTA]2TCCATA[TCTA]10	0.0024	0.0050
	31	[TCTA]7[TCTG]5[TCTA]3TA[TCTA]3TCA[TCTA]2TCCATA[TCTA]11	0.0048	
	31	[TCTA]7[TCTG]5[TCTA]3TA[TCTA]3TCA[TCTA]2TCCATA[TCTA]11		0.0201
	31	[TCTA]6[TCTG]5[TCTA]3TA[TCTA]3TCA[TCTA]2TCCATA[TCTA]12	0.0217	0.0075
	31	[TCTA]5[TCTG]6[TCTA]3TA[TCTA]3TCA[TCTA]2TCCATA[TCTA]12	0.0362	0.0501
	31	[TCTA]4[TCTG]6[TCTA]3TA[TCTA]3TCA[TCTA]2TCCATA[TCTA]13	0.0072	
	31.2	[TCTA]5[TCTG]5[TCTA]3TA[TCTA]3TCA[TCTA]2TCCATA[TCTA]12TATCTA		0.0050
	31.2	[TCTA]5[TCTG]6[TCTA]3TA[TCTA]3TCA[TCTA]2TCCATA[TCTA]11TATCTA	0.0966	0.0501
	32	[TCTA]8[TCTG]5[TCTA]3TA[TCTA]3TCA[TCTA]2TCCATA[TCTA]11		0.0125
	32	[TCTA]7[TCTG]5[TCTA]3TA[TCTA]3TCA[TCTA]2TCCATA[TCTA]12		0.0100
	32	[TCTA]6[TCTG]5[TCTA]3TA[TCTA]3TCA[TCTA]2TCCATA[TCTA]13	0.0097	0.0025
	32	[TCTA]5[TCTG]6[TCTA]3TA[TCTA]3TCA[TCTA]2TCCATA[TCTA]13		0.0276
	32	[TCTA]4[TCTG]6[TCTA]3TA[TCTA]3TCA[TCTA]2TCCATA[TCTA]14	0.0024	
	32.2	[TCTA]6[TCTG]6[TCTA]3TA[TCTA]3TCA[TCTA]2TCCATA[TCTA]11TATCTA	0.0024	
	32.2	[TCTA]5[TCTG]5[TCTA]3TA[TCTA]3TCA[TCTA]2TCCATA[TCTA]13TATCTA		0.0050
	32.2	[TCTA]5[TCTG]6[TCTA]3TA[TCTA]3TCA[TCTA]2TCCATA[TCTA]12TATCTA	0.0870	0.1253
	32.2	[TCTA]5[TCTG]6[TCTA]3TA[TCTA]4TCA[TCTA]2TCCATA[TCTA]11TATCTA		0.0050
	32.2	**[TCTA]5[TCTG]7[TCTA]3TA[TCTA]3TCA[TCTA]2TCCATA[TCTA]11TATCTA**		0.0025
	33	[TCTA]6[TCTG]5[TCTA]3TA[TCTA]3TCA[TCTA]2TCCATA[TCTA]14	0.0024	

	33.2	[TCTA]5[TCTG]6[TCTA]3TA[TCTA]3TCA[TCTA]2TCCATA[TCTA]13TATCTA	0.0266	0.0351
	33.2	[TCTA]5[TCTG]7[TCTA]3TA[TCTA]3TCA[TCTA]2TCCATA[TCTA]12TATCTA		0.0025
	34	**[TCTA]9[TCTG]5[TCTA]3TA[TCTA]3TCA[TCTA]2TCCATA[TCTA]12**		0.0025
	34	**[TCTA]8[TCTG]5[TCTA]3TA[TCTA]3TCA[TCTA]2TCCATA[TCTA]13**	0.0024	
	34.2	[TCTA]5[TCTG]6[TCTA]3TA[TCTA]3TCA[TCTA]2TCCATA[TCTA]14TATCTA	0.0048	0.0050

D22S1045	11	[ATT]8 ACT [ATT]2	0.109	0.195
	12	[ATT]9 ACT [ATT]2	0.009	
	13	[ATT]10 ACT [ATT]2		0.006
	14	[ATT]11 ACT [ATT]2	0.059	0.018
	15	[ATT]12 ACT [ATT]2	0.359	0.268
	16	[ATT]13 ACT [ATT]2	0.382	0.299
	17	[ATT]14 ACT [ATT]2	0.077	0.195
	18	[ATT]15 ACT [ATT]2	0.005	0.018

D2S1338*	14	[TGCC]6[TTCC]8	0.0024	
	16	[TGCC]7[TTCC]9		0.0074
	16	[TGCC]6[TTCC]10	0.0409	
	16	[TGCC]5[TTCC]11	0.0024	
	16	[TGCC]4[TTCC]12	0.0024	
	17	[TGCC]6[TTCC]11	0.2043	0.0743
	18	[TGCC]7[TTCC]11	0.0409	0.0668
	18	[TGCC]6[TTCC]12	0.0601	0.0371
	19	[TGCC]8[TTCC]11	0.0024	0.0074
	19	[TGCC]7[TTCC]12	0.0913	0.1881
	19	[TGCC]6[TTCC]13	0.0216	0.0099
	20	[TGCC]7[TCCC][TTCC]12	0.0072	
	20	[TGCC]8[TTCC]12		0.0050
	20	[TGCC]7[TTCC]10[GTCC][TTCC]2	0.0216	0.0025
	20	[TGCC]7[TTCC]13	0.1082	0.0842
	20	[TGCC]7[TTCC]2[TTTC][TTCC]10		0.0099
	20	[TGCC]6[TTCC]14	0.0024	0.0025
	21	**[TGCC]7[TCCC][TTCC]13**	0.0024	
	21	[TGCC]8[TTCC]13		0.0025
	21	[TGCC]7[TTCC]11[GTCC][TTCC]2	0.0120	0.0025
	21	[TGCC]7[TTCC]14	0.0216	0.0248
	21	[TGCC]7[TTCC]2[TTTC][TTCC]11		0.0025
	21	[TGCC]6[TTCC]15		0.0050
	22	[TGCC]7[TTCC]12[GTCC][TTCC]2	0.0385	0.0149
	22	[TGCC]7[TTCC]15	0.0024	
	22	[TGCC]6[TTCC]13[GTCC][TTCC]2	0.0024	0.0198

	22	[TGCC]5[TTCC]14[GTCC][TTCC]2		0.0025
	22	**[TGCC]4[TTCC]15[GTCC][TTCC]2**		0.0025
	23	[TGCC]8[TTCC]12[GTCC][TTCC]2	0.0024	0.0025
	23	[TGCC]7[TTCC]13[GTCC][TTCC]2	0.0865	0.1609
	23	[TGCC]6[TTCC]14[GTCC][TTCC]2	0.0024	0.0347
	23	[TGCC]5[TTCC]15[GTCC][TTCC]2		0.0074
	23	**[TGCC]4[TTCC]16[GTCC][TTCC]2**		0.0025
	24	[TGCC]8[TTCC]13[GTCC][TTCC]2	0.0096	0.0025
	24	[TGCC]7[TTCC]14[GTCC][TTCC]2	0.0938	0.1213
	24	[TGCC]6[TTCC]15[GTCC][TTCC]2	0.0048	0.0223
	24	**[TGCC]5[TTCC]16[GTCC][TTCC]2**		0.0025
	25	[TGCC]8[TTCC]14[GTCC][TTCC]2	0.0048	
	25	[TGCC]7[TTCC]15[GTCC][TTCC]2	0.0889	0.0347
	25	[TGCC]6[TTCC]16[GTCC][TTCC]2		0.0223
	26	[TGCC]8[TTCC]15[GTCC][TTCC]2	0.0024	0.0025
	26	[TGCC]7[TTCC]16[GTCC][TTCC]2	0.0144	0.0099
	26	[TGCC]6[TTCC]17[GTCC][TTCC]2		0.0025
	27	[TGCC]7[TTCC]17[GTCC][TTCC]2	0.0024	

D2S441	7	[TCTA]7		0.0025
	9	[TCTA]9	0.0024	
	9.1	A[TCTA]9		0.0347
	10	[TCTA]10	0.0577	0.1188
	10	[TCTA]8[TCTG][TCTA]	0.1370	0.1337
	11	[TCTA]11	0.3173	0.2946
	11	[TCTA]9[TCTG][TCTA]	0.0120	0.0198
	11.3	[TCTA]4[TCA][TCTA]7	0.0601	0.0792
	12	[TCTA]12	0.0144	0.1708
	12	[TCTA]10[TCTG][TCTA]	0.0024	0.0050
	12	[TCTA]9[TTTA][TCTA]2	0.0024	
	12.3	[TCTA]4[TCA][TCTA]8	0.0048	
	13	[TCTA]13	0.0024	0.0248
	13	[TCTA]10[TTTA][TCTA]2	0.0457	
	14	[TCTA]11[TTTA][TCTA]2	0.2909	0.1015
	15	[TCTA]12[TTTA][TCTA]2	0.0505	0.0149

D3S1358	11	TCTA [TCTG]2 [TCTA]8	0.0048	
	12	TCTA [TCTG]2 [TCTA]9		0.0050
	13	TCTA TCTG [TCTA]11	0.0024	
	14	TCTA TCTG [TCTA]12	0.0024	
	14	TCTA [TCTG]2 [TCTA]11	0.1202	0.0272

	15	TCTA TCTG [TCTA]13	0.0288	0.0050
	15	TCTA [TCTG]2 [TCTA]12	0.2163	0.3416
	15	TCTA [TCTG]3 [TCTA]11	0.0072	
	16	TCTA TCTG [TCTA]14	0.0072	0.0025
	16	TCTA [TCTG]2 [TCTA]13	0.1851	0.2228
	16	TCTA [TCTG]3 [TCTA]12	0.0553	0.0718
	16	**TCTA [TCTG]3 TCTA TCTG [TCTA]10**	0.0024	
	17	**[TCTA]2 [TCTG]3 [TCTA]12**		0.0025
	17	TCTA TCTG [TCTA]15	0.0024	
	17	TCTA [TCTG]2 [TCTA]14	0.1274	0.1658
	17	**TCTA [TCTG]2 [TCTC] [TCTA]13**	0.0024	
	17	TCTA [TCTG]3 [TCTA]13	0.0913	0.0990
	18	TCTA [TCTG]2 [TCTA]15	0.0024	0.0198
	18	TCTA [TCTG]3 [TCTA]14	0.1346	0.0347
	19	TCTA [TCTG]2 [TCTA]16	0.0024	
	19	TCTA [TCTG]3 [TCTA]15	0.0048	0.0025

D4S2408	7	[ATCT]7		0.0025
	8	[ATCT]8	0.2308	0.2129
	9	[ATCT]9	0.2837	0.0866
	9	[ATCT] GTCT [ATCT]7	0.0457	0.2426
	10	[ATCT]10	0.2692	0.3292
	10	[ATCT] GTCT [ATCT]8		0.0050
	11	[ATCT]11	0.1514	0.0941
	12	[ATCT]12	0.0192	0.0248
	13	[ATCT]13		0.0025

D5S818 *	7	[AGAT]7[AGAG]		0.0274
	9	[AGAT]9[AGAG]	0.0024	
	9	[AGAT]9[AGAT]	0.0457	0.0748
	10	[AGAT]10[AGAG]	0.0361	0.1920
	10	[AGAT]10[AGAT]	0.0168	0.0050
	11	[AGAT]11[AGAG]	0.3462	0.2918
	11	[AGAT]11[AGAT]	0.0385	0.0399
	12	[AGAT]12[AGAG]	0.2476	0.1796
	12	[AGAT]12[AGAT]	0.0793	0.0249
	13	[AGAT]13[AGAG]	0.1346	0.1521
	13	[AGAT]13[AGAT]	0.0385	0.0100
	14	[AGAT]14[AGAG]	0.0120	0.0025
	14	[AGAT]14[AGAT]	0.0024	

D6S1043*	10	[AGAT]10	0.0096	0.0223
	11	[AGAT]11	0.2933	0.1312
	12	[AGAT]12	0.3125	0.1287
	13	[AGAT]13	0.0625	0.1064
	14	[AGAT]12 ACAT [AGAT]	0.0024	
	14	[AGAT]14	0.0601	0.1832
	15	[AGAT]15		0.0272
	16	[AGAT]10[ACAT][AGAT]5	0.0024	0.0050
	16	**[AGAT]11[ACAT][AGAT]4**		0.0025
	16	[AGAT]14[ACAT][AGAT]	0.0024	
	17	[AGAT]11[ACAT][AGAT]5	0.0625	0.0347
	18	[AGAT]12[ACAT][AGAT]5	0.0721	0.1559
	19	[AGAT]13[ACAT][AGAT]5	0.0817	0.1436
	20	[AGAT]14[ACAT][AGAT]5	0.0337	0.0446
	21	[AGAT]14[ACAT][AGAT]6	0.0024	
	21	[AGAT]15[ACAT][AGAT]5	0.0024	0.0099
	22	[AGAT]16[ACAT][AGAT]5		0.0050

D7S820*	7	[GATA]7 GACA GATT GATA GTTT	0.0217	
	8	[GATA]8 GACA GATT GATA GTTT	0.1594	0.1650
	9	[GATA]9 GACA GATT GATA GTTT	0.1498	0.0375
	9.2	**[GATA]10 GACA GATT GA GTTT**		**0.0025**
	10	[GATA]10 GACA GATT GATA GTTT	0.2657	0.1525
	10.1	A[GATA]10 GACA GATT GATA GTTT		0.0025
	11	[GATA]11 GACA GATT GATA GTTT	0.2005	0.3550
	11	[GATA]3[GGTA][GATA]7 GACA GATT GATA GTTT		0.0175
	12	[GATA]12 GACA GATT GATA GTTT	0.1836	0.2425
	13	[GATA]13 GACA GATT GATA GTTT	0.0193	0.0250

D8S1179	8	[TCTA]8	0.0120	
	9	[TCTA]9	0.0144	
	10	[TCTA]10	0.1106	0.1238
	11	[TCTA]11	0.0529	0.0916
	11	TCTA TCTG [TCTA]9		0.0025
	12	[TCTA]12	0.1538	0.0842
	12	TCTA TCTG [TCTA]10	0.0048	0.0198
	13	[TCTA]13	0.0673	0.0619
	13	[TCTA]2 TCTG [TCTA]10	0.0072	
	13	[TCTA]1 TCTG [TCTA]11	0.2692	0.1262
	14	[TCTA]14	0.0288	0.0272
	14	[TCTA]2 TCTG [TCTA]11	0.0288	0.0792

	14	TCTA TCTG [TCTA]12	0.1106	0.0990
	14	TCTA TCTG TGTA [TCTA]11	0.0024	
	15	[TCTA]15	0.0144	
	15	[TCTA]2 TCTG [TCTA]12	0.0529	0.1287
	15	**TCTA TCTG [TCTA]2 CCTA [TCTA]10**	0.0024	
	15	TCTA TCTG [TCTA]13	0.0361	0.0223
	15	**TCTA [TCTG]2[TCTA]12**		0.0025
	16	[TCTA]2 TCTG [TCTA]13	0.0216	0.1064
	16	TCTA TCTG [TCTA]14	0.0072	
	17	[TCTA]2 TCTG [TCTA]14		0.0223
	17	[TCTA]2[TCTG]2[TCTA]13	0.0024	
	18	[TCTA]2 TCTG [TCTA]15		0.0025

D9S1122	9	[TAGA]9	0.0024	
	9	TAGA TCGA [TAGA]7	0.0024	
	10	[TAGA]10	0.0264	0.0495
	10	TAGA TCGA [TAGA]8	0.0048	0.0074
	11	[TAGA]11	0.1995	0.0965
	11	TAGA TCGA [TAGA]9	0.0337	0.0668
	12	[TAGA]12	0.1490	0.0545
	12	TAGA TCGA [TAGA]10	0.2067	0.2351
	13	[TAGA]13	0.0769	0.0297
	13	TAGA TCGA [TAGA]11	0.2476	0.3762
	13	**TAGA TCGA [TAGA]6 TAGG [TAGA]4**		0.0025
	14	[TAGA]14	0.0072	0.0074
	14	TAGA TCGA [TAGA]12	0.0337	0.0644
	15	TAGA TCGA [TAGA]13	0.0096	0.0050
	15	**TAGA [TCGA]2 [TAGA]12**		0.0025
	16	TAGA TCGA [TAGA]14		0.0025

FGA*	17	[TTTC]3[TTTT][TTCT][CTTT]9[CTCC][TTCC]2		0.0050
	18	[TTTC]3[TTTT][TTCT][CTTT]10[CTCC][TTCC]2	0.0168	0.0249
	19	[TTTC]3[TTTT][TTCT][CTTT]11[CTCC][TTCC]2	0.0769	0.0448
	20	[TTTC]3[TTTT][TTCT][CTTT]12[CTCC][TTCC]2	0.1538	0.0423
	20.2	[TTTC]3[TTTT][TT][CTTT]13[CTCC][TTCC]2	0.0024	
	21	[TTTC]3[TTTT][TTCT][CTTT]13[CTCC][TTCC]2	0.1683	0.1045
	21.2	[TTTC]3[TTTT][TT][CTTT]14[CTCC][TTCC]2	0.0072	0.0025
	22	[TTTC]3[TTTT][TTCT][CTTT]14[CTCC][TTCC]2	0.1899	0.1990
	22.2	[TTTC]3[TTTT][TT][CTTT]15[CTCC][TTCC]2	0.0120	0.0025
	23	[TTTC]3[TTTT][TTCT][CTTT]15[CTCC][TTCC]2	0.1659	0.2239
	23.2	[TTTC]3[TTTT][TT][CTTT]16[CTCC][TTCC]2	0.0048	0.0050

	24	[TTTC]3[TTTT][TTCT][CTTT]16[CTCC][TTCC]2	0.1058	0.1692
	24.2	[TTTC]3[TTTT][TT][CTTT]17[CTCC][TTCC]2		0.0075
	24.3	**[TTTC]3[TTTT][TTCT][CTTT]15[CTT][CTTT][CTCC][TTCC]2**	0.0024	
	25	[TTTC]3[TTTT][TTCT][CTTT]17[CTCC][TTCC]2	0.0697	0.1045
	25.2	[TTTC]3[TTTT][TT][CTTT]18[CTCC][TTCC]2		0.0025
	26	[TTTC]3[TTTT][TTCT][CTTT]18[CTCC][TTCC]2	0.0168	0.0498
	26	[TTTC]3[TTTT][TTCT][CTTT]16[GTTT][CTTT][CTCC][TTCC]2	0.0072	
	26.2	[TTTC]3[TTTT][TT][CTTT]19[CTCC][TTCC]2		0.0050
	27	[TTTC]3[TTTT][TTCT][CTTT]13[CCTT][CTTT]5[CTCC][TTCC]2		0.0025
	27	[TTTC]3[TTTT][TTCT][CTTT]19[CTCC][TTCC]2		0.0025
	28	**[TTTC]3[TTTT][TTCT][CTTT][CTCT][CTTT]18[CTCC][TTCC]2**		0.0025

Penta D	7	[AAAGA]7	0.0102	0.0065
	8	[AAAGA]8	0.0127	0.0654
	9	[AAAGA]9	0.1929	0.4183
	10	[AAAGA]10	0.1497	0.1176
	11	[AAAGA]11	0.1447	0.1275
	12	[AAAGA]12	0.2259	0.1340
	13	[AAAGA]13	0.1726	0.0882
	14	[AAAGA]14	0.0635	0.0359
	15	[AAAGA]15	0.0203	0.0065
	16	[AAAGA]16	0.0025	
	17	[AAAGA]17	0.0051	

Penta E*	5	[AAAGA]5	0.0651	0.0526
	7	[AAAGA]7	0.2101	
	8	[AAAGA]8	0.0148	
	9	[AAAGA]9	0.0089	0.0351
	10	[AAAGA]10	0.0976	0.0702
	11	[AAAGA]11	0.1272	0.2018
	12	[AAAGA]12	0.1923	0.1842
	13	[AAAGA]13	0.0799	0.0965
	14	[AAAGA]14	0.0355	0.1228
	15	[AAAGA]15	0.0414	0.0351
	16	[AAAGA]16	0.0473	0.0526
	17	[AAAGA]17	0.0414	0.0526
	18	[AAAGA]18	0.0148	0.0263
	19	[AAAGA]19	0.0089	
	20	[AAAGA]20	0.0089	0.0175
	21	[AAAGA]21	0.0059	0.0351
	22	[AAAGA]22		0.0175

TH01	6	[AATG]6	0.1971	0.1139
	7	[AATG]7	0.1731	0.2822
	8	[AATG]8	0.1130	0.0545
	9	[AATG]9	0.1466	0.4554
	9.3	[AATG]6[ATG][AATG]3	0.3582	0.0322
	10	[AATG]10	0.0120	0.0619

TPOX	8	[AATG]8	0.5433	0.5941
	9	[AATG]9	0.0769	0.1163
	10	[AATG]10	0.0481	0.0272
	11	[AATG]11	0.2813	0.2450
	12	[AATG]12	0.0505	0.0173

vWA*	14	[TCTA] [TCTG] [TCTA] [TCTG]4[TCTA]3[TCCA][TCTA]3 TCCA TCCA	0.0801	0.2600
	14	[TCTA] [TCTG]3[TCTA]10 TCCA TCTA	0.0267	
	14	[TCTA] [TCTG]4[TCTA]9 TCCA TCTA	0.0024	0.0025
	15	[TCTA] [TCTG] [TCTA] [TCTG]4[TCTA]3[TCCA] [TCTA]3[TCCA] TCCA TCCA	0.0049	
	15	[TCTA] [TCTG]3[TCTA]11 TCCA TCTA	0.0680	0.0050
	15	[TCTA] [TCTG]4[TCTA]10 TCCA TCTA	0.0121	0.0275
	15	[TCTA] [TCTG]4[TCTA]10 TCTA TCTA	0.0024	
	16	[TCTA] [TCTG]3[TCTA]12 TCCA TCTA	0.0534	0.0075
	16	[TCTA] [TCTG]4[TCTA]11 TCCA TCTA	0.1845	0.1525
	17	[TCTA] [TCTG]3[TCTA]13 TCCA TCTA	0.0243	0.0025
	17	[TCTA] [TCTG]4[TCTA]12 TCCA TCTA	0.2282	0.2650
	18	[TCTA] [TCTG]3[TCTA]14 TCCA TCTA		0.0075
	18	[TCTA] [TCTG]4[TCTA]13 TCCA TCTA		0.0025
	18	[TCTA] [TCTG]4[TCTA]13 TCCA TCTA	0.2039	0.1675
	18	[TCTA] [TCTG]5[TCTA]12 TCCA TCTA	0.0049	
	18	**[TCTA] [TCTG]4[TTTA][TCTA]12** TCCA TCTA	0.0024	
	18	**[TCTG]5[TCTA]13** TCCA TCTA		0.0050
	19	[TCTA] [TCTG]3[TCTA]15 TCCA TCTA	0.0024	
	19	[TCTA] [TCTG]4[TCTA]14 TCCA TCTA	0.0825	0.0750
	19	[TCTA] [TCTG]5[TCTA]13 TCCA TCTA	0.0049	
	20	[TCTA] [TCTG]4[TCTA]15 TCCA TCTA	0.0097	0.0125
	21	[TCTA] [TCTG]4[TCTA]16 TCCA TCTA	0.0024	0.0075

Table 1: List of individual sequence-based alleles observed in the two population groups studied. Novel variants are highlighted in red and in bold, whilst allele designations that do not follow ISFG recommendations are marked with an asterisk. Flanking region sequences reported by the software are highlighted in light grey to avoid confusion with repeat region sequences.

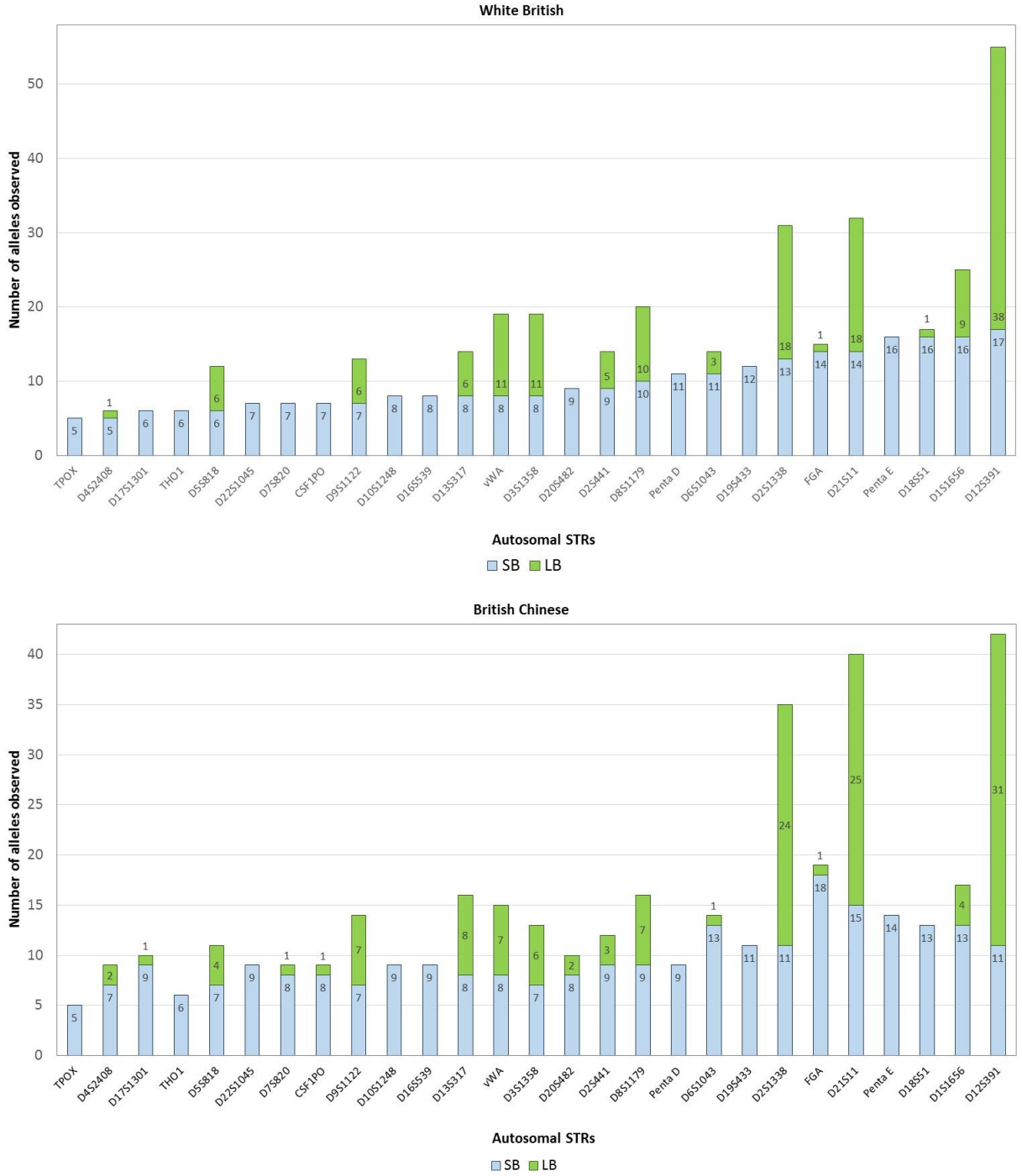

Figure 1: Increase in variation seen across the 27 autosomal STR markers targeted by the ForenSeq™ kit when comparing sequence-based (SB) alleles to length-based (LB) ones. Markers are arranged in order of increasing discrimination based on length in the White British population.

3.3. LOCI OF PARTICULAR INTEREST

During the characterisation of unique alleles observed in this study, several interesting sequence variants were found. These alleles are described below, and the corresponding sequences given in Table 2.

1. CSF1PO is one of the markers in the ForenSeq™ set that shows very limited sequence diversity. Across the 800 CSF1PO alleles characterised in this study, there was only one instance where the sequence diverged from the $[AGAT]_n$ motif. This sequence variation at allele 12 differs to any reported by Gettings *et. al*., who also observed limited variation at this locus [12]. The only other reported instance of this sequence motif was published by Novroski *et. al*. [13], who also only recorded one count at an allele 12, within the same population group as that observed in this study.

2. One novel variant was observed at D18S51, a locus which exhibits limited sequence variation. Although a size-based 18.1 allele has been seen before, no sequence information is available in the literature [22]. The common repeat motif for D18S51 is $[AGAA]_n$. The divergence at this allele would suggest either a duplication into an allele 14, or an insertion into an allele 15. This theory is supported by the flanking region sequences, where the same sequence is observed. Here, the availability of flanking region sequences in the software output enables an interpretation to be made regarding the allele designation.

3. The utility of flanking region information is also demonstrated at D7S820, where a 9.2 allele was observed which is composed of 10 tetra-nucleotide repeats within its repeat region. The two bases underlined in the flanking region show the deletion at this allele. Looking at the repeat region sequence alone in this case would not match to the size-based allele and would therefore be discordant with CE result.

	STR locus	Allele	Sequence
1	CSF1PO	12	$[AGAT]_6[AGAC][AGAT]_5$
2	D18S51	18.1	$[AGAA]_{14}$ **AAAG AGAG AG** GAA [AGAA] AAAG AGAG AG
3	D7S820	9.2	$[GATA]_{10}$ GACA GATT GA-- GTTT

Table 2: Interesting alleles described in section 3.3. The proposed insertion/duplication at D18S51 is highlighted in bold, whilst the deletion at D7S820 is shown as GA--.

3.4. CONCORDANCE OF AUTOSOMAL STRS WITH CE

Due to instances of dropout, 800 alleles were not always typed for each marker targeted (Table 3). Allelic dropout rates of 14% and 28% were observed at D1S1656 and D22S1045, respectively. In the case of D22S1045, heterozygote imbalance has been seen by the manufacturer, and a note about interpretation of homozygous genotypes is included in the protocol [28]. For all other markers exhibiting drop out, either both alleles were not typed, or a single allele was seen under the interpretation threshold. The majority of D1S1656 allelic drop out occurred for samples analysed on one run where sequence

clustering was sub-optimal (although still within the manufacturer's recommended specifications), and this marker was found to underperform in these circumstances. For any marker where 800 alleles were not typed (D12S391, D13S317, D18S51, D19S433, D1S1656, D22S1045 and vWA), the sequence-based allele frequencies were adjusted as described in the supplementary materials.

	Alleles typed (/800)	Concordant alleles	% Concordance
AM	800	800	100
CSF1PO	800	800	100
D10S1248	800	800	100
D12S391	795	795	100
D13S317	799	799	100
D16S539	800	800	100
D18S51	797	797	100
D19S433	799	799	100
D1S1656	692	692	100
D21S11	800	800	100
D22S1045	576	576	100
D2S1338	800	800	100
D2S441	800	800	100
D3S1358	800	800	100
D5S818	800	799	99.88
D7S820	800	799	99.88
D8S1179	800	800	100
FGA	800	800	100
TH01	800	800	100
TPOX	800	800	100
vWA	795	795	100

Table 3: Percentage concordance across all autosomal STRs overlapping with GlobalFiler® and Amelogenin.

Discrepancies were found in 2 samples, where the genotypes obtained using the MiSeq FGx were not concordant with results obtained using the CE-based kit. In one case, a sample presented with a 7 allele at D7S820, despite it being genotyped as a 6.3 with several CE-based methods [2-4]. Here, raw data was re-analysed in order to look at the flanking regions for this allele. The discrepancy was determined to be due to a rare deletion (rs540346880) found in the flanking region of the 7 allele (Figure 2) [21, 33]. This variant has been described in the literature, with a frequency of less than 0.01% in a Caucasian population [34]. In the case of the CE-based method, the deletion would have caused the amplicon to be 1 base pair shorter than expected, hence the resulting 6.3 genotype. This issue raises a question regarding whether flanking region information should be reported for all markers, to facilitate a nomenclature that offers full back-compatibility with CE-based methods.

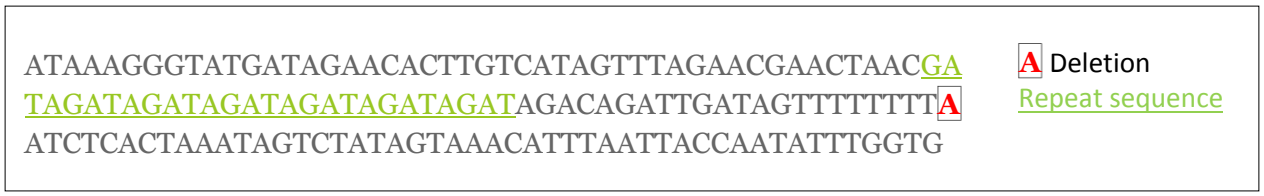

Figure 2: rs540346880 deletion in a 7 allele at D7S820

The second discrepancy was found in a sample reported by UAS as a homozygous 11 genotype at D5S818, whereas a heterozygous 9, 11 genotype was observed using the CE-based kit. Custom primers described in section 2.4. were used to sequence a larger amplicon containing the primer binding regions, and the null allele was found to be caused by a SNP in the reverse primer binding site. The sequence of this null 9 allele is shown in Figure 3, with the base change generating the null allele highlighted in red. This mutation does not have an assigned RS number, and is therefore assumed to be rare. Due to the presence of this null allele, the sample was removed from frequency calculations for this marker.

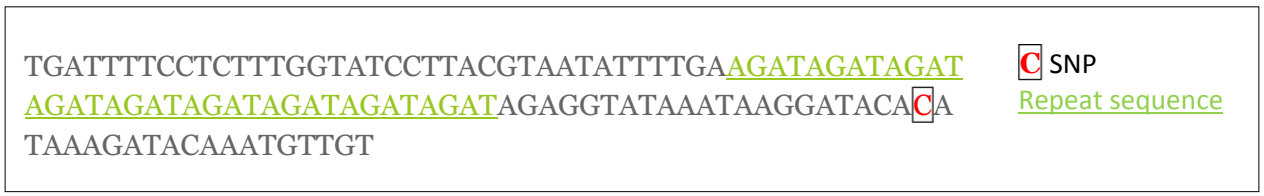

Figure 3: G>C mutation in a 9 null allele at D5S818

Across a total of 16453 allele comparisons, 2 discrepancies were observed, producing a concordance rate exceeding 99.98%. This value is comparable to that observed when comparing CE-based kits [35], and suggests that the ForenSeq™ system is compatible with current technologies.

4. Conclusions and future work

This study demonstrates that data generated using massively parallel sequencing is concordant with current size-based separation techniques to the same degree that CE kits are concordant with each other. It is vital that the forensic community is able to continue using databases generated using CE, and these results confirm that it is possible to compare profiles generated using the two technologies with confidence. Significant sequence variation was seen within a number of autosomal STR markers, with only a minority showing no increased discrimination when comparing sequenced based alleles with length based alleles. This sequence variation will be particularly useful when approaching mixture deconvolution, due to the increased heterozygosity seen at some of these loci. In this work, flanking region sequences were only considered in the context of a discordance event, but it is also likely that additional variation is present within these flanking regions, and future work will focus on exploring this. In order to make use of the additional alleles observed through sequencing, new population frequency data is necessary. This study describes several novel sequence variants, and population data for White British and British Chinese populations are now available as a result of this work. Population-based databases for these alleles are essential, and further work should continue to investigate sequence variation across different populations.

5. Acknowledgements

This work was supported in part by a knowledge transfer partnership project, and the authors would like to thank Innovate UK and Illumina for financial and technical support.

References

1. Jobling, M.A. and P. Gill, *Encoded evidence: DNA in forensic analysis.* Nature Reviews Genetics, 2004. **5**(10): p. 739-751.
2. Life_Technologies, *GlobalFiler™ Express PCR Amplification Kit User Guide.* Publication Part Number 4477672 Rev. A, 2012.
3. Promega, *PowerPlex® ESI 17 Pro System Technical Manual.* 2017.
4. QIAgen, *Investigator® 24plex QS Handbook.* 2016.
5. Dalsgaard, S.R., Eszter; Gelardi, Chiara; Børsting, Claus; Fordyce, Sarah Louise; Morling, Niels, *Characterization of mutations and sequence variations in complex STR loci by second generation sequencing.* Forensic Science International: Genetics. Supplement Series, 2013. **4**(1): p. e218-e219.
6. Fordyce, S.L., M.C. Avila-Arcos, E. Rockenbauer, C. Borsting, R. Frank-Hansen, F.T. Petersen, E. Willerslev, A.J. Hansen, N. Morling, and M.T. Gilbert, *High-throughput sequencing of core STR loci for forensic genetic investigations using the Roche Genome Sequencer FLX platform.* Biotechniques, 2011. **51**(2): p. 127-33.
7. Friis, S.L., A. Buchard, E. Rockenbauer, C. Borsting, and N. Morling, *Introduction of the Python script STRinNGS for analysis of STR regions in FASTQ or BAM files and expansion of the Danish STR sequence database to 11 STRs.* Forensic Sci Int Genet, 2016. **21**: p. 68-75.
8. Rockenbauer, E., S. Hansen, M. Mikkelsen, C. Borsting, and N. Morling, *Characterization of mutations and sequence variants in the D21S11 locus by next generation sequencing.* Forensic Sci Int Genet, 2014. **8**(1): p. 68-72.
9. Van Neste, C., F. Van Nieuwerburgh, D. Van Hoofstat, and D. Deforce, *Forensic STR analysis using massive parallel sequencing.* Forensic Sci Int Genet, 2012. **6**(6): p. 810-8.
10. Borsting, C. and N. Morling, *Next generation sequencing and its applications in forensic genetics.* Forensic Sci Int Genet, 2015. **18**: p. 78-89.
11. Gelardi, C., E. Rockenbauer, S. Dalsgaard, C. Borsting, and N. Morling, *Second generation sequencing of three STRs D3S1358, D12S391 and D21S11 in Danes and a new nomenclature for sequenced STR alleles.* Forensic Sci Int Genet, 2014. **12**: p. 38-41.
12. Gettings, K.B., K.M. Kiesler, S.A. Faith, E. Montano, C.H. Baker, B.A. Young, R.A. Guerrieri, and P.M. Vallone, *Sequence variation of 22 autosomal STR loci detected by next generation sequencing.* Forensic Sci Int Genet, 2016. **21**: p. 15-21.
13. Novroski, N.M., J.L. King, J.D. Churchill, L.H. Seah, and B. Budowle, *Characterization of genetic sequence variation of 58 STR loci in four major population groups.* Forensic Sci Int Genet, 2016. **25**: p. 214-226.
14. Wendt, F.R., J.D. Churchill, N.M. Novroski, J.L. King, J. Ng, R.F. Oldt, K.L. McCulloh, J.A. Weise, D.G. Smith, S. Kanthaswamy, and B. Budowle, *Genetic analysis of the Yavapai Native Americans from West-Central Arizona using the Illumina MiSeq FGx forensic genomics system.* Forensic Sci Int Genet, 2016. **24**: p. 18-23.
15. Silvia, A.L., N. Shugarts, and J. Smith, *A preliminary assessment of the ForenSeq FGx System: next generation sequencing of an STR and SNP multiplex.* Int J Legal Med, 2017. **131**(1): p. 73-86.
16. Churchill, J.D., S.E. Schmedes, J.L. King, and B. Budowle, *Evaluation of the Illumina((R)) Beta Version ForenSeq DNA Signature Prep Kit for use in genetic profiling.* Forensic Sci Int Genet, 2016. **20**: p. 20-9.
17. Xavier, C. and W. Parson, *Evaluation of the Illumina ForenSeq DNA Signature Prep Kit - MPS forensic application for the MiSeq FGx benchtop sequencer.* Forensic Sci Int Genet, 2017. **28**: p. 188-194.

18. Hussing, C., C. Borsting, H.S. Mogensen, and N. Morling, *Testing of the Illumina (R) ForenSeq (TM) kit.* Forensic Science International Genetics Supplement Series, 2015. **5**: p. E449-E450.
19. Just, R.S., L.I. Moreno, J.B. Smerick, and J.A. Irwin, *Performance and concordance of the ForenSeq system for autosomal and Y chromosome short tandem repeat sequencing of reference-type specimens.* Forensic Sci Int Genet, 2017. **28**: p. 1-9.
20. van der Gaag, K.J., R.H. de Leeuw, J. Hoogenboom, J. Patel, D.R. Storts, J.F. Laros, and P. de Knijff, *Massively parallel sequencing of short tandem repeats-Population data and mixture analysis results for the PowerSeq system.* Forensic Sci Int Genet, 2016. **24**: p. 86-96.
21. Kline, M.C., C.R. Hill, A.E. Decker, and J.M. Butler, *STR sequence analysis for characterizing normal, variant, and null alleles.* Forensic Sci Int Genet, 2011. **5**(4): p. 329-32.
22. Ruitberg, C.M., D.J. Reeder, and J.M. Butler, *STRBase: a short tandem repeat DNA database for the human identity testing community.* Nucleic Acids Res, 2001. **29**(1): p. 320-2.
23. Cruz, C., T. Ribeiro, C. Vieira-Silva, I. Lucas, R. Espinheira, and H. Geada, *vWA STR locus structure and variability.* Progress in Forensic Genetics 10, 2004. **1261**: p. 248-250.
24. Dauber, E.M., W. Bar, M. Klintschar, F. Neuhuber, W. Parson, E. Mueller-van der Spruit, and W.R. Mayr, *New sequence data of allelic variants at the STR loci ACTBP2 (SE33), D21S11, FGA, vWA, CSF1PO, D2S1338, D16S539, D18S51 and D19S433 in Caucasoids.* Progress in Forensic Genetics 10, 2004. **1261**: p. 191-193.
25. Grubwieser, P., R. Muhlmann, H. Niederstatter, M. Pavlic, and W. Parson, *Unusual variant alleles in commonly used short tandem repeat loci.* International Journal of Legal Medicine, 2005. **119**(3): p. 164-166.
26. Illumina, *ForenSeq™ DNA Signature Prep Reference Guide.* Document #15049528 v01, 2015.
27. Illumina, *MiSeq FGx™ Instrument Reference Guide.* Part # 15050524 Rev. C, 2015.
28. Illumina, *ForenSeq™ Universal Analysis Software Guide.* Document #15053876 v01 2016.
29. Jager, A.C., M.L. Alvarez, C.P. Davis, E. Guzman, Y. Han, L. Way, P. Walichiewicz, D. Silva, N. Pham, G. Caves, J. Bruand, F. Schlesinger, S.J. Pond, J. Varlaro, K.M. Stephens, and C.L. Holt, *Developmental validation of the MiSeq FGx Forensic Genomics System for Targeted Next Generation Sequencing in Forensic DNA Casework and Database Laboratories.* Forensic Sci Int Genet, 2017. **28**: p. 52-70.
30. Robinson, J.T., H. Thorvaldsdottir, W. Winckler, M. Guttman, E.S. Lander, G. Getz, and J.P. Mesirov, *Integrative genomics viewer.* Nat Biotechnol, 2011. **29**(1): p. 24-6.
31. Parson, W., D. Ballard, B. Budowle, J.M. Butler, K.B. Gettings, P. Gill, L. Gusmao, D.R. Hares, J.A. Irwin, J.L. King, P. Knijff, N. Morling, M. Prinz, P.M. Schneider, C.V. Neste, S. Willuweit, and C. Phillips, *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, 2016. **22**: p. 54-63.
32. Excoffier, L., G. Laval, and S. Schneider, *Arlequin (version 3.0): an integrated software package for population genetics data analysis.* Evol Bioinform Online, 2007. **1**: p. 47-50.
33. Gettings, K.B., R.A. Aponte, P.M. Vallone, and J.M. Butler, *STR allele sequence variation: Current knowledge and future issues.* Forensic Sci Int Genet, 2015. **18**: p. 118-30.
34. Allor, C., D.D. Einum, and M. Scarpetta, *Identification and characterization of variant alleles at CODIS STR loci.* J Forensic Sci, 2005. **50**(5): p. 1128-33.
35. Hill, C.R., M.C. Kline, D.L. Duewer, and J.M. Butler, *Concordance testing comparing STR multiplex kits with a standard data set.* Forensic Science International: Genetics Supplement Series Supplement Series, 2011. **3**(1): p. e188-e189.