Forensic Science International: Genetics 10 (2014) 55–63

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

Comparing six commercial autosomal STR kits in a large Dutch population sample☆

Antoinette A. Westen [a], Thirsa Kraaijenbrink [b], Elizaveta A. Robles de Medina [a,b], Joyce Harteveld [a], Patricia Willemse [b], Sofia B. Zuniga [b], Kristiaan J. van der Gaag [b], Natalie E.C. Weiler [a], Jeroen Warnaar [a], Manfred Kayser [c], Titia Sijen [a], Peter de Knijff [b,*]

[a] Department of Human Biological Traces, Netherlands Forensic Institute, P.O. Box 24044, 2490 AA The Hague, The Netherlands
[b] Forensic Laboratory for DNA Research, Department of Human Genetics, Leiden University Medical Centre, Postzone S-05-P, P.O. Box 9600, 2300 RC Leiden, The Netherlands
[c] Department of Forensic Molecular Biology, Erasmus MC University Medical Centre Rotterdam, P.O. Box 2040, 3000 CA Rotterdam, The Netherlands

ARTICLE INFO

Article history:
Received 26 November 2013
Received in revised form 21 January 2014
Accepted 24 January 2014

Keywords:
Forensic science
Short tandem repeat (STR)
Concordancy testing
Allele frequencies
Null alleles
(Next generation) sequencing

ABSTRACT

Regularly, STR results obtained with different PCR amplification kits are compared, for instance with old cases, when revisiting cold cases or when addressing cross-border crimes. It is known that differences in primer design for the same loci in different kits may give rise to null alleles or shifted alleles. In this study, the genotyping results of 2085 Dutch male samples were compared for six autosomal STR kits (Promega's PowerPlex® 16, ESX-16 and ESI-17 Systems, Qiagen's Investigator® ESSplex Kit and Applied Biosystems' AmpFlSTR® Identifiler and NGM PCR Amplification Kits). A total of 19 discordant autosomal genotyping results were obtained that were examined by sequence analysis using Roche-454 next generation sequencing and/or Sanger sequencing. A further 25 discordances were found and sequenced for the Amelogenin locus. The 24 samples showing the same primer binding site mutation at the Amelogenin locus were subjected to X-STR analysis in order to assess whether they could share a common origin, which appeared not to be the case. Based on the sequencing results, we set the final genotypes and determined the allele frequencies of 23 autosomal STRs for the Dutch reference database.

1. Introduction

After expanding the European Standard Set (ESS) of markers used for autosomal forensic short tandem repeat (STR) genotyping [1,2], a number of companies developed commercially available kits incorporating these loci [3]. Examples of these kits are the PowerPlex ESX and ESI Systems (Promega Corporation (Promega), Madison, WI, USA), the Investigator ESSplex (Plus) Kit (Qiagen Benelux B.V. (Qiagen), Venlo, The Netherlands) and the AmpFlSTR NGM™ (SElect) PCR Amplification Kit (Applied Biosystems (AB), Foster City, CA, USA). When comparing genotyping results of the same donor obtained with different PCR amplification kits, differences such as null (a.k.a. silent) alleles and shifted alleles have been observed, primarily due to differences in primer design (e.g. [4–6]). With an increasing number of forensically available kits and a growing number of (international) DNA profile comparisons worldwide, for instance under the European Prüm Treaty, it is informative to know the extent of discordances at specific loci. When certain discordances occur regularly, this information may stimulate companies to include primer adjustments, such as degenerated primers, in newly developed forensic kits.

In this study, 2085 Dutch male samples were typed for six autosomal STR kits: the PowerPlex® 16 (PP16), ESX-16 and ESI-17 Systems (Promega), the Investigator® ESSplex Kit (ESS, Qiagen) and the AmpFlSTR® Identifiler and NGM PCR Amplification Kits (AB). We evaluated the concordancy of the genotyping results obtained with these kits. Discordant allele calls were examined using Roche-454 next generation sequencing (NGS) and/or Sanger sequencing in order to identify the causal mutations. Additional

* Corresponding author. Tel.: +31 71 526 9540; fax: +31 71 526 8278.
E-mail addresses: t.westen@nfi.minvenj.nl (A.A. Westen), t.kraaijenbrink@lumc.nl (T. Kraaijenbrink), l.robles.de.medina@nfi.minvenj.nl (E.A. Robles de Medina), j.harteveld@nfi.minvenj.nl (J. Harteveld), p.willemse@lumc.nl (P. Willemse), s.b.zuniga@lumc.nl (S.B. Zuniga), k.j.van_der_gaag@lumc.nl (K.J. van der Gaag), n.weiler@nfi.minvenj.nl (Natalie E.C. Weiler), j.warnaar@nfi.minvenj.nl (J. Warnaar), m.kayser@erasmusmc.nl (M. Kayser), t.sijen@nfi.minvenj.nl (T. Sijen), p.de_knijff@lumc.nl, knijff@lumc.nl (P. de Knijff).

1872-4973/$ – see front matter © 2014 The Authors. Published by Elsevier Ireland Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.fsigen.2014.01.008

analyses with X-STRs were performed to analyse whether a set of 24 donors having the same Amelogenin primer binding site mutation could share a common origin. Based on the sequencing results, the final genotypes were set and used to create a new Dutch allele frequency database.

2. Material and methods

2.1. DNA samples, extraction and quantification

A total of 2085 male blood donors with self-declared Dutch ancestry were sampled from 99 locations across The Netherlands, while excluding major cities to avoid very recent admixture effects. All volunteers had given their informed consent, and a detailed description of the samples is given in [7]. After anonymising the samples, DNA was robotically extracted either by the Autopure LS® system using the Gentra Puregene Blood Kit (Gentra Systems, Minneapolis, MN) or by the QIAcube using the QIAamp DNA Mini Kit (Qiagen Benelux B.V. (Qiagen), Venlo, The Netherlands). The samples were quantified with the Quantifiler® Duo DNA Quantification Kit on a 7500 Real-Time PCR System (Applied Biosystems (AB), Foster City, CA, USA).

2.2. PCR, capillary electrophoresis and DNA profile analysis

DNA amplifications were performed using six different autosomal STR kits: PP16, ESX-16 and ESI-17 (Promega), ESS (Qiagen) and Identifiler and NGM (AB). PCR amplification programs were performed according to the manufacturer's protocols, except for the reaction volumes (Table 1). The PCR input and reaction volume, the capillary electrophoresis instrument, setup and settings, and the DNA profile analysis software and detection threshold are shown in Table 1 for the six different STR kits.

2.3. Roche-454 next generation sequencing, Sanger sequencing and X-STR analysis

Supplementary Text 1 describes the material and methods for Roche-454 NGS, analysis of NGS data [8–10], Sanger sequencing and X-STR analysis. Primers used during Sanger sequencing and X-STR analysis are presented in Supplementary Table 1.

2.4. Concordancy testing, allele frequencies and statistical analyses

The complete data set was tested for the presence of population substructure as described in [7], and no substructure was detected (data not shown). Therefore, the complete dataset was interpreted as one group. The 2085 genotypes from all six kits were compared using Excel and discordant allele calls were identified. After sequencing analysis of discordant alleles, one final database was created comprising 2085 genotypes for 23 autosomal STRs and Amelogenin. With this database, the allele frequencies and descriptive statistics for this Dutch population sample were determined using the Excel Microsatellite Toolkit [11].

3. Results and discussion

3.1. Discordances between six autosomal STR kits

The 2085 DNA samples were analysed with six autosomal STR kits (PP16, ESX-16, ESI-17, ESS, Identifiler and NGM), comprising a total of 23 different autosomal STR loci and Amelogenin (Fig. 1). The genotyping results for the 20 STRs and Amelogenin that are present in more than one kit (Penta D and E only reside in PP16, SE33 only resides in ESI-17; see Fig. 1) were compared and 44 discordances were detected in 38 DNA samples (Table 2). These discrepancies comprise 43 null alleles on 12 loci and one shifted allele with a size difference of one nucleotide (Table 2). Just when all 2085 samples had been analysed with the NGM kit, Applied Biosystems adjusted the kit (without changing the name; kit lot-numbers 1103010 M and up) by adding extra primers for Amelogenin, D2S441 and D22S1045 that cover reported primer binding site mutations on these loci [12,13]. All discordant samples were reanalysed with this new version of NGM (here denoted NGM2), and the 24 missing X alleles on Amelogenin and the absent allele 14.1 on D2S441 were now detected and concordant with the results from other kits. As a result, the final number of discordant results decreased to 19, and these discordances occurred in 13 of the 2085 samples (for six samples, the same discordant result was seen in two different kits; Table 2). Seven discordant alleles were found for Identifiler and NGM/NGM2, three for ESI-17, and one for both PP16 and ESS (Table 2); ESX-16 showed no discordant allele calls (Table 2).

3.2. Roche-454 NGS and/or Sanger sequencing of samples with discordant results

In order to determine the genetic variations causing the above-mentioned 44 discordant results, the loci involved were sequenced for the affected samples and control samples. First, Roche-454 NGS was applied using primers designed by Kline et al. [5], and when no (clear) results were obtained, Sanger sequencing was used with custom-designed primers both further away from the repeat than those by Kline et al. [5] (Supplementary Table 1). For 42 of the 44 discordances a nucleotide change or nucleotide insertion or deletion was observed that probably caused the discordant genotyping result (Table 2 and Supplementary Data 1). For two samples, with a null allele at D12S391 for the ESI-17 and NGM/NGM2 kits, the discordant allele could not be amplified and no

Table 1
Characteristics of the PCR amplification, capillary electrophoresis and DNA profile analysis for the six autosomal PCR kits analysed in this study.

PCR			Capillary electrophoresis						DNA profile analysis	
Kit	Input (ng)	Reaction volume (μL)	Instrument ABI Prism®	ddH_2O (μL)	Hi-Di formamide (μL)	Size standard	Sample	Settings	Software	Detection threshold (rfu)
PP16	1.5	12.5	3100	11.4	–	0.6 μL ILS 600	1.0 μL	3 kV/11 s	GeneMarker v. 1.75	30
ESX-16	1.5	7.5	3100	11.5	–	0.5 μL ILS 500	1.0 μL[a]	3 kV/5 s	GeneMarker v. 1.75	30
ESI-17	0.5	12.5	3130xl	–	10.0	1.0 μL CC5 ILS 500	1.0 μL	1.5 kV/10 s	GeneMapper ID-X v.1.1.1	50
ESS	1.5	7.5	3100	11.5	–	0.5 μL DSS 550	1.0 μL	3 kV/10 s	GeneMarker v. 1.75	30
Identifiler	4.0	12.5	3100	11.6	–	0.4 μL LIZ 600	1.0 μL	1 kV/22 s	GeneMarker v. 1.75	30
NGM	0.5	12.5	3130xl	–	8.7	0.3 μL LIZ 500	1.0 μL	3 kV/15 s	GeneMapper ID-X v.1.1.1	50

[a] 30x diluted PCR product.

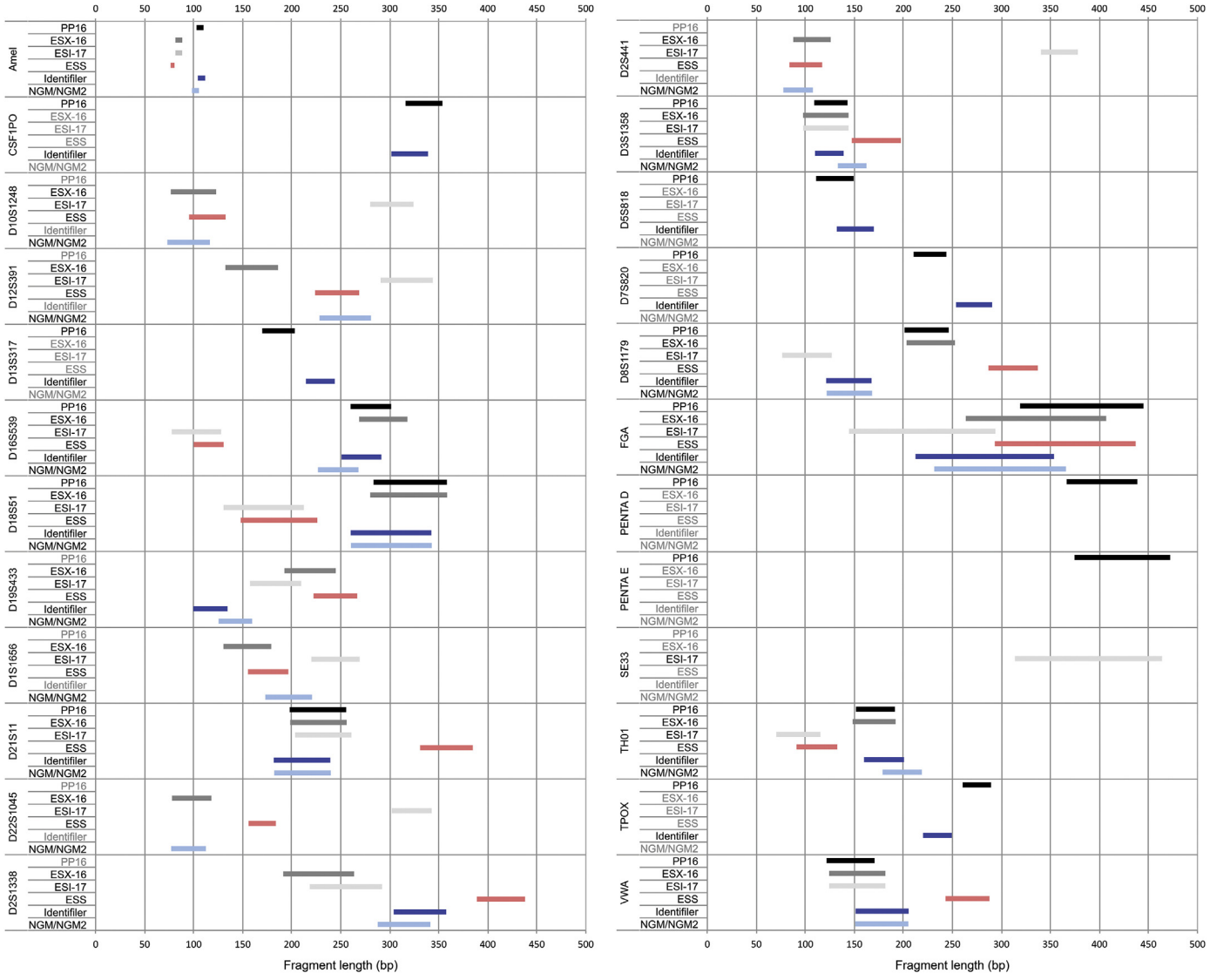

Fig. 1. Fragment lengths for 23 autosomal STR markers and Amelogenin and their presence in six forensic STR kits: PP16, ESX-16 and ESI-17 (Promega, grey shades), ESS (Qiagen, red shade), Identifiler and NGM (Applied Biosystems, blue shades). Amplicon sizes are based on the shortest and longest fragment length in the allelic ladder for each marker. When a kit name is presented in grey, the marker is not present in the kit.

sequence change was revealed. There is a clear relation, however, between the occurrence of the null allele and the location of the various primers: kits that do detect the allele have reverse primers close to the repeat structure (for ESX the reverse primer starts 30 nt and for ESS 43 nt after the repeat; Supplementary Data 1), while the kits and custom-designed primers that do not reveal the allele have reverse primers further on (primer start for NGM is at position 125 (Supplementary Data 1) and for the various sequencing primers at nucleotide 213 [5], 271 or 345). We infer that both null alleles are caused by a deletion of at least a few hundred nucleotides, residing after the position of the ESS reverse primer and covering the binding sites of the sequencing primers and the reverse primers in NGM/NGM2 and ESI-17 (for this kit, the exact primer start is unknown).

Many studies have been conducted to determine the molecular basis of discordant results (e.g. [5,14–17]). When null alleles are observed relatively often (for instance in specific populations [16,18,19]) and are due to a primer binding site mutation, manufacturers may add degenerate primers to the commercially available STR kits in order to recover the specific allele [13,18,19]. The majority of discordant results, however, are rare, like most of the discordances found in our study. Despite their rarity, at least three of the autosomal discordances that we detected have been described before: 1) a G>A substitution 172 bp after the D18S51 repeat motif, described as position 75615 in GenBank sequence AC021803 by Delamoye et al. [15], and Vanderheyden et al. [17], 2) a C>T substitution 81 bp before the repeat structure of TH01, which was described by Li et al. [20], and 3) an A>T substitution 52 bp before the repeat motif of vWA, described as position 1631 in GenBank sequence M25858 by Alves et al. [21], Delamoye et al. [15], Vanderheyden et al. [17], and Kline et al. [5].

Except for Amelogenin, the markers that are present in both Identifiler and NGM are said to have the same primer binding positions [12]. Notwithstanding, null alleles were detected for D2S1338 and D16S539 for Identifiler that were not observed for NGM (Table 2). We think the prolonged annealing time in the NGM protocol (3 min versus 1 min) made it possible to detect these

Table 2
Discordancy between six autosomal STR kits: PP16, ESX-16, ESI-17, ESS, Identifiler and NGM, based on 2085 Dutch male donor samples. Discordant results are shown in grey.

Marker	Discordancy	Promega			Qiagen	AB			*n* observations		Sequencing method	Putative cause	
		PP16	ESX-16	ESI-17	ESS	Identifiler	NGM	NGM2[a]	5 kits + NGM	5 kits + NGM2		Mutation position before/ after repeat[c]	Location in primer[d]
Amel.	null allele X	XY	XY	XY	XY	XY	Y	XY	24	0	454	A: 44 C>T	21 nt after start NGM R-primer
Amel.	null allele Y	XY	XY	XY	XY	XY	X	as NGM	1	1	454/Sanger	A: 51 G>A	14 nt after start NGM R-primer
D2S1338	null allele 17	–	17–20	17–20	17–20	20	17–20	as NGM	1	1	Sanger	A: 174 G>A	23 nt after start Identifiler R-primer
D2S441	null allele 14.1	–	11–14.1	11–14.1	11–14.1	–	11	11–14.1	1	0	454	B: 1 insA	23 nt after start NGM F-primer
D7S820	null allele 12	12–13	–	–	–	13	–	–	1	1	454	A: 125 C>A	start Identifiler R-primer unknown
D8S1179	null allele 12.1	12.1–13	12.1–13	13	12.1–13	12.1–13	12.1–13	as NGM	1	1	454	insT final repeat [TCTTA]	start ESI-17 R-primer unknown
D12S391	null allele 21	–	18–21	18	18–21	–	18	as NGM	2[b]	2[b]	454/Sanger	large del involving NGM, ESI-17, Kline [5] and 2 custom-made[e] R-primers	
D12S391	null allele 21	–	21–22	22	21–22	–	22	as NGM	2[b]	2[b]	454/Sanger	large del involving NGM, ESI-17, Kline [5] and 2 custom-made[e] R-primers	
D13S317	null allele 8	11	–	–	–	8–11	–	–	1	1	454	A: 27 C>A	18 nt after start PP16 R-primer
D16S539	null allele 13	12–13	12–13	12–13	12–13	12	12–13	as NGM	1	1	454	B: 140 G>A	17 nt after start Identifiler F-primer
D18S51	null allele 16	15–16	15–16	15–16	15–16	15	15	as NGM	2[b]	2[b]	454	A: 172 G>A	8 nt after start Identifiler + NGM R-primer
D19S433	allele shift 14 to 13.3	–	14–15	14–15	13.3–15	14–15	14–15	as NGM	1	1	454	B: 50 delA	in between ESS and other (known) F-primers
TH01	null allele 6	6–9	6–9	6–9	6–9	9	9	as NGM	2[b]	2[b]	454	B: 81 C>T	14 nt after start Identifiler + NGM F-primer
vWA	null allele 17	17–19	17–19	17–19	17–19	19	19	as NGM	2[b]	2[b]	454	B: 52 A>T	19 nt after start Identifiler + NGM F-primer
vWA	null allele 17	16–17	16–17	16–17	16–17	16	16	as NGM	2[b]	2[b]	454	B: 52 A>T	19 nt after start Identifiler + NGM F-primer
	Total	1	0	3	1	7	32	7	44	19			

[a] NGM2 is only tested on the discordant samples.
[b] Both observations involve the same donor and the discordancy was seen in two different kits.
[c] Mutation position before (B) or after (A) the repeat, and in the direction as given in STRbase (accessed October 2013); see also Supplementary Data 1.
[d] For most primers, the (start) position is given in Supplementary Data 1.
[e] Up to 345 nt from repeat.

alleles, albeit with relatively low peak heights compared to the other allele in the heterozygous allele set (peak height ratios were 0.06 and 0.13, respectively).

All 24 samples that showed a null allele for X on Amelogenin when typed with NGM have the same G>A substitution (Table 2 and Supplementary Data 1). Green et al. [13] described that the discovery of rare population-specific variant alleles for AMEL prompted the adjustment of the AMEL primers, which was put into effect in the summer of 2011. The use of this updated NGM kit (here denoted NGM2), resolved all AMELX null alleles in our samples, indicating that this mutation is not very rare in the Dutch dataset as it occurs with a frequency of 1.15%. Interestingly, none of the other reported AMELX null alleles [22–26], with estimated frequencies up to 2% in specific populations, were found in our dataset. In order to analyse whether the Dutch male donors of these samples could share a common origin, we tested X-STRs in all these Amelogenin mutation carriers. A homemade multiplex was used that contains the SRY male gender marker and 13 X-STRs of all four linkage groups [27,28]. Three of the markers in this multiplex (DXS8378, DXS9902 and DXS6807) reside in the same area of the X-chromosome as Amelogenin, but no apparent similarities at these three loci were seen among the 24 samples (results not shown). Thus, no apparent association between the X-STR results and the G>A substitution in the Amelogenin gene could be made. This suggests there is no recent common origin for this mutation in our study population.

3.3. Number and percentage discordant and concordant results

A consolidated database (denoted 2085-database) comprising 2085 genotypes of 23 autosomal STR loci and Amelogenin was created. For the samples that showed a null allele, the detected allele was included in the database. For the sample with the shifted allele in ESS for D19S433, we included the allele call presented by the other four kits, as the deleted nucleotide was found in the flanking regions of the repeat (Table 2 and Supplementary Data 1). Next, we determined the number and percentage of discordant genotypes and alleles for each locus (Table 3). The highest percentage of discordant results was found for locus D12S391 (genotypes: 0.048% and alleles: 0.024%), which was typed by four different kits. Fully concordant results for all kits tested were obtained for 9 loci typed with 2 to 6 kits. For three loci (Penta D, Penta E and SE33) no comparisons could be made, as they were only present in one of the tested kits. Overall, 99.991% of the genotypes and 99.995% of the alleles show fully concordant results for our Dutch reference dataset of 2085 samples.

3.4. Allele frequencies and descriptive statistics

The frequencies in the new Dutch allele frequency database are given in Table 4. Compared to the former Dutch database [29], which was based on 231 samples for the SGM Plus (AB) and 201 for the Profiler (AB) loci, 52 new alleles were detected that had between one and thirteen occurrences in the 2085-dataset. One allele that was detected once in the former database was not revealed in the 2085-dataset. The largest difference between the allele frequencies of the former and the new database is 0.048. In the new allele frequency database 97 of the 361 different alleles have frequencies below 0.001 ($\leq$ four occurrences) of which 48 alleles occur only once.

The summary statistics together with the heterozygosity and PIC values are shown in Supplementary Table 2. SE33 shows the highest heterozygosity and PIC values, which can be explained by the high number of microvariants resulting in 59 different alleles (Table 4) detected for this locus compared to the average of 15.65 different alleles per marker.

Table 3
Number and percentage of discordant results on genotype and allele level obtained with six autosomal STR kits.

	Present in # kits	*n* samples	*n* analysed		*n* discordances		% discordances	
			Genotype	Allele	Genotype	Allele	Genotype	Allele
Amelogenin	6	2085	12,510	25,020	1	1	0.008	0.004
D16S539	6	2085	12,510	25,020	1	1	0.008	0.004
D18S51	6	2085	12,510	25,020	2	2	0.016	0.008
D21S11	6	2085	12,510	25,020	0	0	0.000	0.000
D3S1358	6	2085	12,510	25,020	0	0	0.000	0.000
D8S1179	6	2085	12,510	25,020	1	1	0.008	0.004
FGA	6	2085	12,510	25,020	0	0	0.000	0.000
TH01	6	2085	12,510	25,020	2	2	0.016	0.008
VWA	6	2085	12,510	25,020	4	4	0.032	0.016
D19S433	5	2085	10,425	20,850	1	1	0.010	0.005
D2S1338	5	2085	10,425	20,850	1	1	0.010	0.005
D10S1248	4	2085	8340	16,680	0	0	0.000	0.000
D12S391	4	2085	8340	16,680	4	4	0.048	0.024
D1S1656	4	2085	8340	16,680	0	0	0.000	0.000
D22S1045	4	2085	8340	16,680	0	0	0.000	0.000
D2S441	4	2085	8340	16,680	0	0	0.000	0.000
CSF1PO	2	2085	4170	8340	0	0	0.000	0.000
D13S317	2	2085	4170	8340	1	1	0.024	0.012
D5S818	2	2085	4170	8340	0	0	0.000	0.000
D7S820	2	2085	4170	8340	1	1	0.024	0.012
TPOX	2	2085	4170	8340	0	0	0.000	0.000
PENTA D	1	2085	2085	4170	n.a.	n.a.	n.a.	n.a.
PENTA E	1	2085	2085	4170	n.a.	n.a.	n.a.	n.a.
SE33	1	2085	2085	4170	n.a.	n.a.	n.a.	n.a.
Total	n.a.[a]	50,040	20,2245	40,4490	19	19	0.009	0.005

[a] Not applicable.

Table 4
Overview of the number and frequency of alleles per marker.

Allele	Count	Frequency
D1S1656		
10	7	0.00168
11	312	0.07482
12	517	0.12398
13	239	0.05731
14	365	0.08753
14.3	7	0.00168
15	547	0.13118
15.3	312	0.07482
16	495	0.11871
16.1	1	0.00024
16.3	229	0.05492
17	241	0.05779
17.1	3	0.00072
17.3	592	0.14197
18	16	0.00384
18.3	224	0.05372
19	2	0.00048
19.3	57	0.01367
20.3	4	0.00096
D2S441		
8	2	0.00048
9	1	0.00024
10	787	0.18873
11	1431	0.34317
11.3	177	0.04245
12	198	0.04748
12.3	7	0.00168
13	107	0.02566
13.3	2	0.00048
14	1225	0.29376
14.1	1	0.00024
15	207	0.04964
16	25	0.00600
D2S1338		
12	1	0.00024
14	1	0.00024
15	2	0.00048
16	159	0.03813
17	809	0.19400
18	349	0.08369
19	486	0.11655
20	632	0.15156
21	127	0.03046
22	131	0.03141
23	391	0.09376
24	512	0.12278
25	469	0.11247
26	90	0.02158
27	7	0.00168
28	4	0.00096
D3S1358		
10	1	0.00024
11	8	0.00192
13	25	0.00600
14	512	0.12278
15	1028	0.24652
16	1008	0.24173
17	902	0.21631
18	644	0.15444
19	39	0.00935
20	3	0.00072
D5S818		
7	2	0.00048
8	15	0.00360
9	127	0.03046
10	211	0.05060
11	1468	0.35204
12	1571	0.37674
13	729	0.17482
14	40	0.00959
15	6	0.00144
16	1	0.00024

Table 4 (*Continued*)

Allele	Count	Frequency
D7S820		
7	80	0.01918
8	690	0.16547
9	750	0.17986
10	1081	0.25923
10.3	1	0.00024
11	803	0.19257
11.3	1	0.00024
12	533	0.12782
13	190	0.04556
14	38	0.00911
15	3	0.00072
D8S1179		
8	80	0.01918
9	70	0.01679
10	371	0.08897
11	323	0.07746
12	665	0.15947
12.1	1	0.00024
13	1333	0.31966
14	813	0.19496
15	383	0.09185
16	110	0.02638
17	21	0.00504
D10S1248		
7	1	0.00024
10	3	0.00072
11	11	0.00264
12	148	0.03549
13	1318	0.31607
14	1281	0.30719
15	819	0.19640
16	489	0.11727
17	85	0.02038
18	14	0.00336
19	1	0.00024
D12S391		
14	1	0.00024
15	173	0.04149
16	156	0.03741
16.3	1	0.00024
17	419	0.10048
17.3	95	0.02278
18	732	0.17554
18.3	92	0.02206
19	430	0.10312
19.3	41	0.00983
20	503	0.12062
20.2	1	0.00024
20.3	2	0.00048
21	546	0.13094
22	423	0.10144
23	313	0.07506
24	154	0.03693
25	65	0.01559
26	15	0.00360
27	6	0.00144
28	1	0.00024
29	1	0.00024
D13S317		
7	6	0.00144
8	458	0.10983
9	309	0.07410
10	275	0.06595
11	1236	0.29640
12	1225	0.29376
13	468	0.11223
14	190	0.04556
15	3	0.00072
D16S539		
7	1	0.00024
8	70	0.01679
9	540	0.12950
10	272	0.06523

Table 4 (*Continued*)

Allele	Count	Frequency
11	1341	0.32158
12	1102	0.26427
13	748	0.17938
14	94	0.02254
15	1	0.00024
16	1	0.00024
D18S51		
9	4	0.00096
10	33	0.00791
11	58	0.01391
12	655	0.15707
13	575	0.13789
14	687	0.16475
15	596	0.14293
16	514	0.12326
17	448	0.10743
18	299	0.07170
19	153	0.03669
20	82	0.01966
21	38	0.00911
22	18	0.00432
23	4	0.00096
24	6	0.00144
D19S433		
10	2	0.00048
11	21	0.00504
12	279	0.06691
12.1	4	0.00096
12.2	1	0.00024
13	890	0.21343
13.2	76	0.01823
14	1527	0.36619
14.2	97	0.02326
15	763	0.18297
15.2	176	0.04221
16	222	0.05324
16.2	70	0.01679
17	22	0.00528
17.2	14	0.00336
18	1	0.00024
18.2	4	0.00096
19.2	1	0.00024
D21S11		
24.2	1	0.00024
25.2	1	0.00024
26	7	0.00168
27	179	0.04293
28	706	0.16930
28.2	1	0.00024
29	839	0.20120
29.2	3	0.00072
29.3	2	0.00048
30	1079	0.25875
30.2	137	0.03285
31	382	0.09161
31.2	325	0.07794
32	67	0.01607
32.2	306	0.07338
33	12	0.00288
33.2	109	0.02614
34	1	0.00024
34.1	2	0.00048
34.2	8	0.00192
35	2	0.00048
35.2	1	0.00024
D22S1045		
10	6	0.00144
11	586	0.14053
12	56	0.01343
13	25	0.00600
14	212	0.05084
15	1356	0.32518
16	1546	0.37074
17	351	0.08417
18	28	0.00671
19	4	0.00096
CSF1PO		
6	1	0.00024
7	5	0.00120
8	10	0.00240
9	94	0.02254
10	1063	0.25492
10.3	1	0.00024
11	1280	0.30695
12	1389	0.33309
13	267	0.06403
14	52	0.01247
15	7	0.00168
16	1	0.00024
FGA		
17	5	0.00120
18	75	0.01799
19	253	0.06067
19.1	2	0.00048
19.2	4	0.00096
20	559	0.13405
20.2	1	0.00024
21	696	0.16691
21.2	13	0.00312
22	718	0.17218
22.1	1	0.00024
22.2	31	0.00743
23	614	0.14724
23.2	14	0.00336
24	574	0.13765
24.2	5	0.00120
25	428	0.10264
26	156	0.03741
27	20	0.00480
28	1	0.00024
VWA		
10	1	0.00024
12	2	0.00048
13	8	0.00192
14	412	0.09880
15	407	0.09760
16	859	0.20600
17	1143	0.27410
18	833	0.19976
19	442	0.10600
20	60	0.01439
21	3	0.00072
PENTA D		
2.2	1	0.00024
5	3	0.00072
6	2	0.00048
7	28	0.00671
8	85	0.02038
9	870	0.20863
10	413	0.09904
11	568	0.13621
12	968	0.23213
13	885	0.21223
13.4	5	0.00120
14	283	0.06787
14.1	3	0.00072
15	41	0.00983
16	13	0.00312
17	1	0.00024
18	1	0.00024
PENTA E		
5	298	0.07146
6	10	0.00240
7	712	0.17074
8	43	0.01031
9	44	0.01055
10	367	0.08801
11	398	0.09544
12	825	0.19784

Table 4 (*Continued*)

Allele	Count	Frequency
13	391	0.09376
14	256	0.06139
15	210	0.05036
16	218	0.05228
17	205	0.04916
18	101	0.02422
19	44	0.01055
20	31	0.00743
21	12	0.00288
22	3	0.00072
23	2	0.00048
TH01		
5	30	0.00719
6	875	0.20983
7	754	0.18082
8	463	0.11103
8.3	1	0.00024
9	564	0.13525
9.3	1445	0.34652
10	38	0.00911
TPOX		
7	2	0.00048
8	2258	0.54149
9	395	0.09472
10	239	0.05731
11	1077	0.25827
12	193	0.04628
13	6	0.00144
SE33		
10.2	1	0.00024
11	2	0.00048
11.2	2	0.00048
12	19	0.00456
12.2	1	0.00024
13	32	0.00767
13.2	7	0.00168
13.3	2	0.00048
14	112	0.02686
14.2	6	0.00144
14.3	1	0.00024
15	147	0.03525
15.2	2	0.00048
15.3	5	0.00120
16	182	0.04365
16.2	1	0.00024
16.3	4	0.00096
17	265	0.06355
17.2	1	0.00024
17.3	4	0.00096
18	280	0.06715
18.2	3	0.00072
18.3	4	0.00096
19	261	0.06259
19.1	2	0.00048
19.2	22	0.00528
20	202	0.04844
20.1	2	0.00048
20.2	65	0.01559
20.3	1	0.00024
21	134	0.03213
21.2	74	0.01775
22	36	0.00863
22.2	119	0.02854
23	11	0.00264
23.2	113	0.02710
24	3	0.00072
24.2	134	0.03213
25.2	149	0.03573
25.3	4	0.00096
26.2	265	0.06355
27	2	0.00048
27.2	355	0.08513
27.3	1	0.00024
28.2	347	0.08321
29.2	301	0.07218

Table 4 (*Continued*)

Allele	Count	Frequency
30.2	236	0.05659
31	1	0.00024
31.2	111	0.02662
32	8	0.00192
32.2	46	0.01103
33	40	0.00959
33.2	19	0.00456
34	10	0.00240
34.2	3	0.00072
35	5	0.00120
35.2	2	0.00048
36	2	0.00048
37	1	0.00024

4. Concluding remarks

Nineteen discordant alleles were detected in 2085 Dutch male donor samples that were analysed for six autosomal STR kits (giving 99.995% concordant alleles overall). The numbers of null or shifted alleles vary for the different commercial kits: seven in Identifiler, seven in NGM2, three in ESI-17, one in PP16, one in ESS and none in ESX-16. Upon Roche-454 and/or Sanger sequencing nucleotide changes, insertions and deletions (single nucleotide or stretches of at least few hundred nucleotides) were found that could account for the discordant genotypes. The use of as much as 2085 samples generated a highly detailed allele frequency database, which can be of assistance in evidentiary value calculations.

Acknowledgements

This study was supported by a grant from The Netherlands Genomics Initiative/Netherlands Organization for Scientific Research (NWO) within the framework of the Forensic Genomics Consortium Netherlands. The authors would like to thank Applied Biosystems, Promega and Qiagen for their unconditional in kind support, and Arnoud Kal for critically reading the manuscript.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.fsigen.2014.01.008.

References

[1] P. Gill, L. Fereday, N. Morling, P.M. Schneider, New multiplexes for Europe-amendments and clarification of strategic development, Forensic Sci. Int. 163 (2006) 155–157.
[2] P. Gill, L. Fereday, N. Morling, P.M. Schneider, The evolution of DNA databases – recommendations for new European STR loci, Forensic Sci. Int. 156 (2006) 242–244.
[3] A.A. Westen, H. Haned, L.J. Grol, J. Harteveld, K.J. van der Gaag, P. de Knijff, T. Sijen, Combining results of forensic STR kits: HDplex validation including allelic association and linkage testing with NGM and identifiler loci, Int. J. Legal Med. 126 (2012) 781–789.
[4] C.R. Hill, D.L. Duewer, M.C. Kline, C.J. Sprecher, R.S. McLaren, D.R. Rabbach, B.E. Krenke, M.G. Ensenberger, P.M. Fulmer, D.R. Storts, J.M. Butler, Concordance and population studies along with stutter and peak height ratio analysis for the PowerPlex (R) ESX 17 and ESI 17 Systems, Forensic Sci. Int. Genet. 5 (2011) 269–275.
[5] M.C. Kline, C.R. Hill, A.E. Decker, J.M. Butler, STR sequence analysis for characterizing normal, variant, and null alleles, Forensic Sci. Int. Genet. 5 (2011) 329–332.
[6] L.A. Welch, P. Gill, C. Phillips, R. Ansell, N. Morling, W. Parson, J.U. Palo, I. Bastisch, European Network of Forensic Science Institutes (ENFSI): evaluation of new commercial STR multiplexes that include the European Standard Set (ESS) of markers, Forensic Sci. Int. Genet. 6 (2012) 819–826.
[7] O. Lao, E. Altena, C. Becker, S. Brauer, T. Kraaijenbrink, O.M. van, P. Nurnberg, K.P. de, M. Kayser, Clinal distribution of human genomic diversity across the Netherlands despite archaeological evidence for genetic discontinuities in Dutch population history, Invest. Genet. 4 (2013) 9.

[8] S.Y. Anvar, K.J. van der Gaag, J.W.F. van der Heijden, M.H.A.M. Veltrop, R.H.A.M. Vossen, R.H. de Leeuw, C. Breukel, H.P.J. Buermans, J.S. Verbeek, P. de Knijff, J.T. den Dunnen, J.F.J. Laros, TSSV: A tool for characterization of complex allelic variants in pure and mixed genomes, Bioinformatics (2014), http://dx.doi.org/10.1093/bioinformatics/btu068.
[9] J.D. Thompson, D.G. Higgins, T.J. Gibson, CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice, Nucleic Acids Res. 22 (1994) 4673–4680.
[10] T.A. Hall, BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT, Nucleic Acids Symp. Ser. 41 (1999) 95–98.
[11] S.D.E. Park, Trypanotolerance in West African Cattle and the Population Genetic Effects of Selection, University of Dublin, Ireland, 2001.
[12] N. Oldroyd, R. Green, J. Mulero, L. Hennessy, J. Tabak, Development of the AmpFℓSTR® NGM SElect™ Kit: new sequence discoveries and implications for genotype concordance, Forensic News (2011), http://www3.appliedbiosystems.-com/cms/groups/applied_markets_marketing/documents/generaldocuments/cms_090694.pdf.
[13] R.L. Green, R.E. Lagace, N.J. Oldroyd, L.K. Hennessy, J.J. Mulero, Developmental validation of the AmpFlSTR(R) NGM SElect PCR Amplification Kit: a next-generation STR multiplex with the SE33 locus, Forensic Sci. Int. Genet. 7 (2013) 41–51.
[14] B. Budowle, A. Masibay, S.J. Anderson, C. Barna, L. Biega, S. Brenneke, B.L. Brown, J. Cramer, G.A. DeGroot, D. Douglas, B. Duceman, A. Eastman, R. Giles, J. Hamill, D.J. Haase, D.W. Janssen, T.D. Kupferschmid, T. Lawton, C. Lemire, B. Llewellyn, T. Moretti, J. Neves, C. Palaski, S. Schueler, J. Sgueglia, C. Sprecher, C. Tomsey, D. Yet, STR primer concordance study, Forensic Sci. Int. 124 (2001) 47–54.
[15] M. Delamoye, C. Duverneuil, K. Riva, M. Leterreux, S. Taieb, M.P. De, False homozygosities at various loci revealed by discrepancies between commercial kits: implications for genetic databases, Forensic Sci. Int. 143 (2004) 47–52.
[16] T.M. Clayton, S.M. Hill, L.A. Denton, S.K. Watson, A.J. Urquhart, Primer binding site mutations affecting the typing of STR loci contained within the AMPFlSTR SGM Plus kit, Forensic Sci. Int. 139 (2004) 255–259.
[17] N. Vanderheyden, A. Mai, A. Gilissen, J.J. Cassiman, R. Decorte, Identification and sequence analysis of discordant phenotypes between AmpFlSTR SGM Plus and PowerPlex 16, Int. J. Legal Med. 121 (2007) 297–301.
[18] C. Leibelt, B. Budowle, P. Collins, Y. Daoudi, T. Moretti, G. Nunn, D. Reeder, R. Roby, Identification of a D8S1179 primer binding site mutation and the validation of a primer designed to recover null alleles, Forensic Sci. Int. 133 (2003) 220–227.
[19] M.S. Nelson, E.N. Levedakou, J.R. Matthews, B.E. Early, D.A. Freeman, C.A. Kuhn, C.J. Sprecher, A.S. Amin, K.C. McElfresh, J.W. Schumm, Detection of a primer-binding site polymorphism for the STR locus D16S539 using the Powerplex 1.1 system and validation of a degenerate primer to correct for the polymorphism, J. Forensic Sci. 47 (2002) 345–349.
[20] F. Li, J. Xuan, J. Xing, M. Ding, B. Wang, H. Pang, Identification of new primer binding site mutations at TH01 and D13S317 loci and determination of their corresponding STR alleles by allele-specific PCR, Forensic Sci. Int. Genet. 8 (2014) 143–146.
[21] C. Alves, A. Amorim, L. Gusmao, L. Pereira, VWA STR genotyping: further inconsistencies between Perkin-Elmer and Promega kits, Int. J. Legal Med. 115 (2001) 97–99.
[22] B. Shadrach, M. Commane, C. Hren, I. Warshawsky, A rare mutation in the primer binding region of the amelogenin gene can interfere with gender identification, J. Mol. Diagn. 6 (2004) 401–405.
[23] C. Alves, M. Coelho, J. Rocha, A. Amorim, The amelogenin locus displays a high frequency of X homologue failures in Sao Tome Island (West Africa), Int. Congress Series 1288 (2006) 271–273.
[24] S. Caratti, G. Voglino, V. Cirigliano, A. Ghidini, R. Taulli, C. Torre, C. Robino, Amplification failure of the amelogenin gene (AMELX) caused by a primer binding site mutation, Prenat. Diagn. 29 (2009) 1180–1182.
[25] A. Maciejewska, R. Pawlowski, A rare mutation in the primer binding region of the Amelogenin X homologue gene, Forensic Sci. Int. Genet. 3 (2009) 265–267.
[26] X. Ou, W. Chen, H. Chen, F. Zhao, J. Zheng, D. Tong, Y. Chen, A. Chen, H. Sun, Null alleles of the X and Y chromosomal amelogenin gene in a Chinese population, Int. J. Legal Med. 126 (2012) 513–518.
[27] S. Hering, J. Edelmann, C. Augustin, E. Kuhlisch, R. Szibor, X chromosomal recombination—a family study analysing 39 STR markers in German three-generation pedigrees, Int. J. Legal Med. 124 (2010) 483–491.
[28] M. Nothnagel, R. Szibor, O. Vollrath, C. Augustin, J. Edelmann, M. Geppert, C. Alves, L. Gusmao, M. Vennemann, Y. Hou, U.D. Immel, S. Inturri, H. Luo, S. Lutz-Bonengel, C. Robino, L. Roewer, B. Rolf, J. Sanft, K.J. Shin, J.E. Sim, P. Wiegand, C. Winkler, M. Krawczak, S. Hering, Collaborative genetic mapping of 12 forensic short tandem repeat (STR) loci on the human X chromosome, Forensic Sci. Int. Genet. 6 (2012) 778–784.
[29] A.D. Kloosterman, The development and implementation of forensic DNA typing technologies in the Netherlands, 2002.