



Forensic Population Genetics - Original Research

Analysis of 36 Y-STR marker units including a concordance study among 2085 Dutch males



Antoinette A. Westen^a, Thirsa Kraaijenbrink^b, Lindy Clarisse^a, Laurens J.W. Grol^a, Patricia Willemse^b, Sofia B. Zuniga^b, Elizaveta A. Robles de Medina^{a,b}, Ron Schouten^b, Kristiaan J. van der Gaag^b, Natalie E.C. Weiler^a, Arnoud J. Kal^a, Manfred Kayser^c, Titia Sijen^{a,1}, Peter de Knijff^{b,1,*}

^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 Center, Postzone S-05-P, P.O. Box 9600, 2300 RC Leiden, The Netherlands

^c Department of Forensic Molecular Biology, Erasmus MC University Medical Center Rotterdam, P.O. Box 2040, 3000 CA Rotterdam, The Netherlands

ARTICLE INFO

Article history:

Received 12 September 2014

Received in revised form 13 October 2014

Accepted 14 October 2014

Keywords:

Forensic science
Rapidly mutating Y-STR marker unit
Concordance study
Haplotype diversity
Allele frequency
Sanger sequencing

ABSTRACT

The genotypes of 36 Y-chromosomal short tandem repeat (Y-STR) marker units were analysed in a Dutch population sample of 2085 males. Profiling results were compared for several partially overlapping kits, i.e. PowerPlex Y, Yfiler, PowerPlex Y23, and two in-house designed multiplexes with rapidly mutating Y-STRs. Nineteen Y-STR marker units, of which two are rapidly mutating, reside in at least two of these multiplexes, and for these markers concordance testing was performed. Two samples showed discordant genotyping results and the probable causative base change was revealed by Sanger sequencing. In addition, we encountered concordant, but aberrant genotyping results including one allele with low peak height and several null alleles. For 12 samples, this involved a null allele in two adjacent loci suggesting a large and recurrent deletion as the samples represent three distinct haplogroups. For each marker unit, the allele counts and frequencies are presented, as are the haplotype counts and haplotype diversities for several combinations of markers.

© 2014 The Authors. Published by Elsevier Ireland Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/3.0/>).

1. Introduction

The number of Y-chromosomal short tandem repeat (Y-STR) markers for routine forensic and population genetic use has grown considerably over the past few years. Initially, a minimal haplotype set of nine Y-STR marker units was recommended for forensic use [1], and expansion of the core set by two additional STRs was recommended by SWGDAM [2]. The subsequently developed and

commercially available multiplexes contain a growing number of Y-STR marker units, such as 12 in the PowerPlex[®] Y System (PPY, Promega, released in 2003), 17 in the AmpFISTR[®] Yfiler[®], (Yfiler, Life Technologies, released in 2004), 23 in the PowerPlex Y23 System (PPY23, Promega, released in 2012) and 27 in the AmpFISTR[®] Yfiler[®] Plus Kit [3] (Life Technologies, to be released in 2014).

Y-STRs can be of great value in stains with small quantities of male DNA and overwhelming amounts of female DNA, for instance in sexual assault cases. They can also be very useful in kinship analyses, due to their strict paternal inheritance pattern. However, as a result of the relatively low mutation rate for the commonly used Y-STRs, it is difficult, if not impossible, to differentiate between closely related males. The introduction of 13 rapidly mutating (RM) Y-STRs with median mutation rates about 6.5 times higher than the Yfiler STRs [4] assists cases where increased discrimination power of Y-STRs is needed [4–6]. Consequently, in a set of 2378 father–son pairs 26.9% could be differentiated using the RM Y-STR set versus 4.5% with Yfiler [6].

* 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.clarisse@nfi.minvenj.nl (L. Clarisse), l.grol@nfi.minvenj.nl (Laurens J.W. Grol), p.willemse@lumc.nl (P. Willemse), s.b.zuniga@lumc.nl (S.B. Zuniga), l.robles.de.medina@nfi.minvenj.nl (E.A. Robles de Medina), r.schouten@lumc.nl (R. Schouten), k.j.van.der.gaag@lumc.nl (K.J. van der Gaag), n.weiler@nfi.minvenj.nl (Natalie E.C. Weiler), a.kal@nfi.minvenj.nl (A.J. Kal), m.kayser@erasmusmc.nl (M. Kayser), t.sijen@nfi.minvenj.nl (T. Sijen), p.de.knijff@lumc.nl (P. de Knijff).

¹ Both authors contributed equally.

In this study, we analysed all 36 Y-STR marker units present in PPY, Yfiler, PPY23 and the RM Y-STR set described in [4]. We use the term “marker unit” for previously defined distinct Y-STR markers, e.g. for DYS385 a separate “a” and “b” part are described and these are counted as two marker units (resulting for instance in 17 marker units for Yfiler in total), while DYF387S1 is counted as one marker unit even though it can show up to three alleles (resulting in 15 RM marker units in total). These 36 marker units were tested in 2085 DNA samples from Dutch male blood donors. For the 19 Y-STR marker units that are present in more than one set, concordance testing was performed and discordant alleles were subsequently analysed with Sanger sequencing. Allele counts and frequencies are reported together with the haplotype counts and haplotype diversities for several marker combinations. All PowerPlex Y23 haplotypes have been submitted to the publicly available Y Chromosome Haplotype Reference Database (YHRD) [7,8].

2. Materials and methods

2.1. DNA samples

A total of 2085 male blood donors with self-defined 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. A detailed

description of the samples is given in [9], and the DNA extraction and quantification are described in [10].

2.2. Marker units, DNA amplification, capillary electrophoresis (CE) and DNA profile analysis

All 2085 DNA samples were amplified with five Y-STR multiplex PCRs, targeting 36 marker units (present in 32 different Y-STRs of which one has a “I” and “II” part (i.e. DYS389) and three have an “a” and “b” part (i.e. DYF403S1, DYS385 and DYS526)). Three of these multiplexes are commercially available: PPY and PPY23 from Promega Corporation (Promega, Madison, WI, USA) and Yfiler from Life Technologies (Life Tech, Foster City, CA, USA). All 12 PPY marker units reside in Yfiler, and all 17 Yfiler marker units are represented in PPY23 (Table 1). The other two multiplexes (RMY1 and RMY2) were redesigned in-house based on the three RM Y-STR multiplexes published in [4,5]. They analyse 15 rapidly mutating Y-STR marker units (that reside in 13 Y-STRs). RMY1 holds six and RMY2 nine marker units, and RMY2 contains two marker units overlapping with PPY23 (Table 1).

DNA amplification with the Yfiler, PPY, and PPY23 multiplexes was performed according to the manufacturer’s protocols, but with half of the reaction volume. PCR products were detected by CE on an ABI Prism 3100 Genetic Analyzer (Life Tech), using a 36 cm array, POP-4 and dye set G5 (for Yfiler and PPY23) or C (for PPY). 1 µL sample or allelic ladder was mixed with 11.6 µL ddH₂O and

Table 1
Marker units present in five Y-STR multiplexes together with the number and percentage discordant results.

Marker	PPY	Yfiler	PPY23	RMY1	RMY2	<i>n</i> alleles/marker unit (avg)	<i>n</i> genotypes analysed	<i>n</i> genotypes discordant	% Genotypes discordant
DYF387S1				x		1–3 (1.79)	2085	n.a. ^b	n.a.
DYF399S1					x	1–5 (2.80)	2085	n.a.	n.a.
DYF403S1a					x	1–4 (2.84)	2085	n.a.	n.a.
DYF403S1b					x	1 ^a	2085	n.a.	n.a.
DYF404S1					x	1–3 (1.77)	2085	n.a.	n.a.
DYS19	x	x	x			1	6255	0	0
DYS385a	x	x	x			1	6255	0	0
DYS385b	x	x	x			1 ^a	6255	0	0
DYS389I	x	x	x			1	6255	0	0
DYS389II	x	x	x			1	6255	0	0
DYS390	x	x	x			1	6255	0	0
DYS391	x	x	x			1	6255	0	0
DYS392	x	x	x			1	6255	0	0
DYS393	x	x	x			1	6255	0	0
DYS437	x	x	x			1	6255	0	0
DYS438	x	x	x			1	6255	0	0
DYS439	x	x	x			1 ^a	6255	0	0
DYS448		x	x			1	4170	1	0.024
DYS449					x	1 ^a	2085	n.a.	n.a.
DYS456		x	x			1	4170	0	0
DYS458		x	x			1	4170	0	0
DYS481			x			1 ^a	2085	n.a.	n.a.
DYS518				x		1 ^a	2085	n.a.	n.a.
DYS526a				x		1	2085	n.a.	n.a.
DYS526b				x		1 ^a	2085	n.a.	n.a.
DYS533			x			1 ^a	2085	n.a.	n.a.
DYS547					x	1 ^a	2085	n.a.	n.a.
DYS549			x			1	2085	n.a.	n.a.
DYS570			x		x	1 ^a	4170	0	0
DYS576			x		x	1 ^a	4170	0	0
DYS612					x	1 ^a	2085	n.a.	n.a.
DYS626				x		1 ^a	2085	n.a.	n.a.
DYS627				x		1 ^a	2085	n.a.	n.a.
DYS635		x	x			1	4170	1	0.024
DYS643			x			1	2085	n.a.	n.a.
Y-GATA-H4		x	x			1	4170	0	0
Total							139,695	2	0.002 ^c

^a A second allele is present in one to six of the 2085 samples.

^b Not applicable.

^c Total % genotypes discordant = total *n* genotypes discordant / total *n* genotypes analysed for two or three kits (i.e. 104,250).

0.4 μ L GeneScanTM LIZ 600 Size Standard (Life Tech) for Yfiler, with 11.5 μ L ddH₂O and 0.5 μ L ILS600 (Promega) for PPY, or with 11 μ L ddH₂O and 1 μ L CC5 ILS500 Y23 (Promega) for PPY23, and analysed after 3 min of denaturation and 3 min on ice. CE injection settings were 1 kV for 22 s for Yfiler and PPY, and 3 kV for 5 s for PPY23. The Y-STR profiles were analysed using GeneMapper v. 3.0 (Life Tech) for PPY or GeneMarker v. 1.75 (Softgenetics, LLC., State College, PA, USA) for Yfiler and PPY23 with a detection threshold of 30 rfu.

RMY1 and RMY2 PCRs were performed in a 10 μ L reaction volume using 1 \times QIAGEN Multiplex PCR Buffer (Qiagen, Venlo, the Netherlands), primers as described in Supplementary Table S1 and 1.0 ng DNA. The PCR protocol starts with a pre-denaturation step for 10 min at 94 °C, followed by a step-down PCR of 10 cycles at 94 °C for 30 s, 65 °C (1 °C/cycle) for 30 s and 72 °C for 1 min, and 23 cycles (for RMY1) or 25 cycles (for RMY2) of 94 °C for 30 s, 50 °C for 30 s and 72 °C for 1 min, with a final extension at 60 °C for 45 min. PCR products were detected by CE on an ABI Prism 3130xl Genetic Analyzer (Life Tech), using a 36 cm array, POP-7 and dye set G5. 1 μ L sample was mixed with 8.7 μ L Hi-DiTM Formamide (Life Tech) and 0.3 μ L GeneScanTM LIZ 600 Size Standard (Life Tech), and analysed after 4 min of denaturation and 5 min on ice. CE injection settings were 3 kV for 10 s. The RM Y-STR profiles were analysed using GeneMapper[®] ID-X v. 1.1.1 (Life Tech) with a detection threshold of 50 rfu. For most markers a stutter filter of 20% was applied, except for DYS518 and DYS526b (both 25%), DYS570 (30%) and DYS612 (35%).

Supplementary Table S1 related to this article can be found, in the online version, at doi:10.1016/j.fsigen.2014.10.012.

2.3. Sanger sequencing

Twenty-five microliters singleplex PCR reactions were performed using PCR buffer I (Life Tech) with 1.5 mM MgCl₂, 0.2 mM dNTP mix (Life Tech), 2 units AmpliTaq Gold (Life Tech) and 2 pmol of each HPLC-purified primer (Supplementary Table S1). The amplification, purification, sequencing, detection and sequence analysis was performed as described in [10].

2.4. Allele counts, allele frequencies, haplotypes and haplotype diversities

Based on the Y-STR data, haplotypes were constructed and compared using Excel (Microsoft, Redmond, WA, USA) for all 2085 donors. For each allele in each marker unit, the number of occurrences was counted. Allele frequencies were calculated by dividing the allele count for a specific allele through the total number of counted alleles for that marker unit (which was not always 2085, due to null alleles or additional alleles in multi copy marker units). Haplotype diversities were calculated using Arlequin v3.5.1.3 [11] and an adjusted version of an Excel worksheet kindly provided by Ballantyne (personal communication) to be able to calculate numbers with more than four digits after the decimal point (our version of the worksheet is available on request). In case of multi copy marker units, the “empty cells” were filled with a dummy variable for donors that showed less than the maximum number of alleles.

2.5. Haplogroups and familial relationships

For 12 donors, Y-SNP analysis was performed to determine their haplogroup using the methods described in [12]. For another 22 persons, the autosomal STR data that were determined in [10] were used to infer the most likely familial relationships with Bonaparte [13] (<http://www.bonaparte-dvi.com>). To this end, fictive family trees were produced in which one of the donors of a pair was fixed

(grey square in Fig. 1) and the other donor was tested for all the other possible male relationships (eight white squares in Fig. 1). Additional relationship testing was performed with a version of RelPair [14,15] that was adjusted to enable the analysis of a dataset containing 2085 individuals (details are available on request).

3. Results and discussion

3.1. Discordances and null alleles

DNA samples of 2085 male donors were analysed with five Y-STR multiplexes: PPY, Yfiler, PPY23, RMY1 and RMY2 (both in-house designed, based on the markers published in [4,5]). Of the 36 Y-marker units analysed by these multiplexes, 19 reside in two or three systems (Table 1) and enable concordance testing. Two discordances were found (Table 2): for one person DYS448 showed an allele 19 for PPY23 and no allele with Yfiler, while for another person Yfiler resulted in an allele call 23 for DYS635 with no result for PPY23. Using Sanger sequencing, for both discordances single base changes were disclosed: an A > G transition 49 nucleotides prior to the DYS448 repeat motif, and a T > A transversion 7 nucleotides before the DYS635 repeat structure. As the primer positions for these markers are not publicly available, we cannot check whether these nucleotide changes are located at the primer binding sites for the kits showing the null allele. Both Davis et al. [16] and Larmuseau et al. [17] did not find any discordance in the 17 overlapping loci between Yfiler and PPY23 in their sample sets of 951 American and 535 Belgian donors, respectively. This befits the low percentage of 0.002% discordance that we observe in our larger Dutch dataset (Table 1).

Beside the above-described two discordances, 32 other null alleles were observed. For seven donors, a null allele was found on DYF403S1b, which is only present in RMY2 (Table 2). For one person DYS439 showed no results in all three commercial kits (PPY, Yfiler and PPY23; Table 2). In 12 different samples both DYS448 (present in Yfiler and PPY23) and DYS626 (present in RMY1) showed no results (Table 2). These marker units are located 52.2 kbp from each other with none of the other markers situated between them [18]. We gather that this “double null allele” is due to a large deletion. Several papers describe null alleles at DYS448 (e.g. [19–21]), but since DYS626 is less commonly typed it is unclear whether these have such a double null allele as well. In order to test whether the 12 persons with this double null allele in our sample set are related in the male lineage, their haplogroups were determined using Y-SNPs. Six of them demonstrated haplogroup I, five had haplogroup R1a and one showed haplogroup R1b. Therefore, we deduce that the 12 persons with this double null allele do not originate from one male lineage, and that this double null allele is recurrent and identical by state among the different haplogroups. When examining the Y-STR haplotypes for the persons belonging to the same haplogroup (I or R1a), it was noticed that two donors in haplogroup I showed haplotypes with only one difference between them, while all others displayed at least five differences (data not shown). This one difference was detected in the rapidly mutating DYF403S1b marker and we infer that these two donors may be (closely) related in the male lineage, which would mean that, in this case, the double null allele is identical by descent. However, relationship testing based on 23 autosomal STRs did not suggest a first or second degree relationship between these donors (data not shown).

A noteworthy observation occurred for single copy marker DYS576, as in one person an additional allele 14 of low peak height was found next to a much higher allele 18 in both PPY23 and RMY2 profiles. The peak height ratio between both alleles varied between 0.12 and 0.31 in four independent amplifications with both

multiplexes. The presence of the two alleles was confirmed with Sanger sequencing, although the signals for allele 14 were again very low and did not allow detecting a possible primer binding site mutation. As the PCR primers for Sanger sequencing were positioned at least 100 nucleotides further up- and downstream than those used in RMY2 (and the primer positions for PPY23 are unknown), we infer either the presence of multiple primer binding site mutations, or a chimeric situation that is specific for this Y-STR marker as none of the other Y-STR or autosomal markers showed additional weak alleles. More detailed sequence information may be obtained from next generation sequencing [10], but for now it remains unclear what causes the presence of the second lower allele on DYS576 in this sample.

3.2. Allele counts, allele frequencies and haplotypes

Four RM Y-STR marker units (DYF387S1, DYF399S1, DYF403S1a and DYF404S1) most often show multiple alleles per marker (between one and five alleles, Table 1), and are therefore categorised as multi copy markers. The other 32 marker units are considered single copy markers, although 14 of these, including the previously described DYS576, show a second allele in one to six of the 2085 samples (Table 1). In 26 of the 32 cases, the second allele differs only one repeat length in size from the first allele, but size differences up to six repeat lengths have been found. Except for the previously described sample showing two unbalanced alleles in DYS576, both alleles are balanced in the other cases. Therefore, one needs to realise that finding a second allele on a

marker unit that is believed to be a single copy marker, not always reflects a mixture of two donors.

All haplotypes are presented in Supplementary Table S2. For all 36 marker units, the alleles present in the 2085 DNA samples were counted and their frequencies were calculated (Table 3). DYS393 and DYS437 show the smallest allelic range with only five different alleles in our Dutch population sample; DYF399S1 has the largest range with 36 different alleles.

Supplementary Table S2 related to this article can be found, in the online version, at doi:10.1016/j.fsigen.2014.10.012.

Next, we examined the haplotypes resulting from different combinations of Y-STR marker units: the minimum YHRD marker set, the various commercial kits (PPY, Yfiler and PPY23), the rapidly mutating Y-STRs (RMY1 + RMY2), and all 36 marker units together (PPY23 + RMY1 + RMY2). Table 4 shows the level of uniqueness of haplotypes (the number of times a haplotype was observed) and how many haplotypes have that level of uniqueness (the number of occurrences in our 2085 samples). In general, with more Y-STR markers, more unique haplotypes are found. The PPY23 markers resulted in 92.5% unique haplotypes (1929 haplotypes occurred only once (Table 4), haplotype diversity = 0.999959494976 (Table 5)), which is in the same range as the 93.5% described for the European group analysed with PPY23 by Purps et al. [21]. For the RM Y-STRs (RMY1 + RMY2), 98.4% unique haplotypes were observed (2052 haplotype singletons (Table 4), haplotype diversity = 0.999991714881 (Table 5)), which is somewhat lower than the 100% reported by Ballantyne et al. [6] for the 112 Dutch samples in their set. When combining all 36 Y-STR marker units, 2065

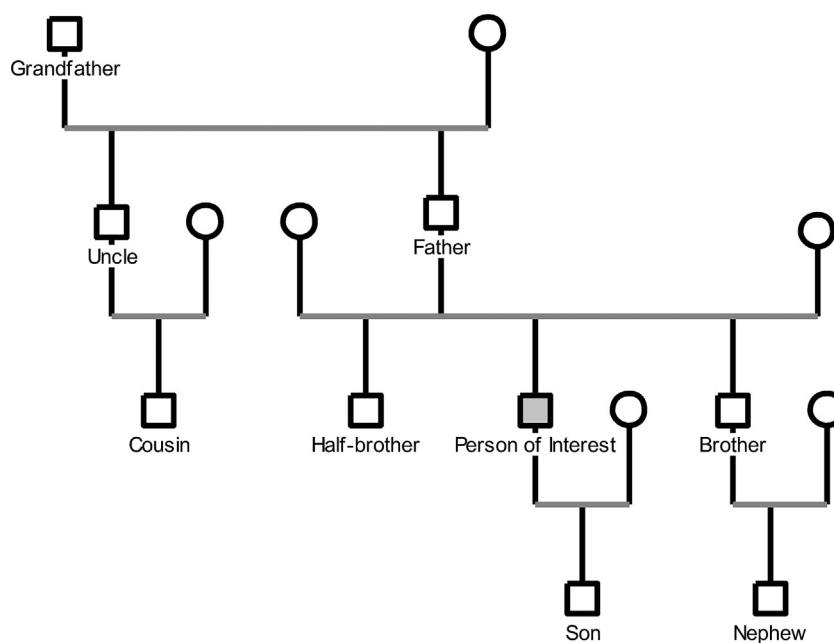


Fig. 1. Fictive family tree used to deduce the most likely family relationship between two donors. Based on genotyping information at 23 autosomal STRs, Bonaparte software is used to calculate the $\log_{10}(LR)$ for the different relationships by fixing one of the donors of a pair (grey square) and testing the second donor for the other possible male relationships (eight white squares).

Table 2
Y-STR discordances and null alleles in a population sample of 2085 Dutch males.

Marker unit	Observation	PPY	Yfiler	PPY23	RMY1	RMY2	<i>n</i>	Remark
DYF403S1b	Null allele	–	–	–	–	Null	7	–
DYS439	Null allele	Null	Null	Null	–	–	1	–
DYS448	19/null	–	Null	19	–	–	1	Sanger: A > G 49 nt before repeat structure
DYS448 + DYS626	Null allele	–	Null ^{DYS448}	Null ^{DYS448}	Null ^{DYS626}	–	12	Haplogroup: 6 × I, 5 × R1a, 1 × R1b
DYS635	23/null	–	23	Null	–	–	1	Sanger: T > A 7 nt before repeat structure

Table 3
Allele counts and frequencies per marker unit.

DYF387S1	Count	Frequency	13	1195	57.31	DYS526a	Count	Frequency
32	3	0.08	14	311	14.92	10	1	0.05
33	40	1.07	15	13	0.62	11	7	0.34
34	100	2.68	DYS389II	Count	Frequency	12	151	7.24
34.2	2	0.05	25	1	0.05	13	311	14.92
35	956	25.66	26	3	0.14	14	928	44.51
35.1	3	0.08	27	20	0.96	15	297	14.24
35.2	4	0.11	28	487	23.36	16	330	15.83
36	916	24.58	29	937	44.94	17	59	2.83
36.1	1	0.03	30	418	20.05	18	1	0.05
36.2	1	0.03	31	148	7.10	DYS526b	Count	Frequency
37	797	21.39	32	62	2.97	27	1	0.05
37.1	1	0.03	33	9	0.43	30	3	0.14
37.2	2	0.05	DYS390	Count	Frequency	31	19	0.91
38	584	15.67	20	1	0.05	32	82	3.92
39	220	5.90	21	18	0.86	33	64	3.06
39.3	3	0.08	22	414	19.86	34	90	4.30
40	72	1.93	23	729	34.96	35	181	8.66
41	17	0.46	24	714	34.24	36	495	23.67
42	3	0.08	24.3	1	0.05	37	616	29.46
43	1	0.03	25	189	9.06	38	370	17.69
DYF399S1	Count	Frequency	26	18	0.86	39	141	6.74
5	1	0.02	27	1	0.05	40	24	1.15
14	2	0.03	DYS391	Count	Frequency	41	5	0.24
16	2	0.03	9	25	1.20	DYS533	Count	Frequency
16.2	1	0.02	10	1077	51.65	7	1	0.05
17	10	0.17	11	922	44.22	9	39	1.87
17.1	3	0.05	12	53	2.54	10	39	1.87
17.2	30	0.51	13	7	0.34	11	550	26.37
18	23	0.39	14	1	0.05	11.3	1	0.05
18.1	3	0.05	DYS392	Count	Frequency	12	1290	61.84
18.2	18	0.31	9	1	0.05	13	152	7.29
19	85	1.46	10	8	0.38	14	14	0.67
19.1	10	0.17	10.2	3	0.14	DYS547	Count	Frequency
19.2	5	0.09	11	701	33.62	41	1	0.05
20	273	4.68	12	146	7.00	42	14	0.67
20.1	27	0.46	13	1156	55.44	43	19	0.91
21	746	12.78	14	59	2.83	44	60	2.87
21.1	32	0.55	15	9	0.43	45	154	7.38
21.2	1	0.02	16	2	0.10	46	392	18.78
22	828	14.19	DYS393	Count	Frequency	47	560	26.83
22.1	68	1.17	11	1	0.05	48	487	23.33
23	871	14.92	12	146	7.00	48.2	3	0.14
23.1	132	2.26	13	1648	79.04	49	240	11.50
23.2	1	0.02	14	234	11.22	50	121	5.80
24	784	13.43	15	56	2.69	51	31	1.49
24.1	250	4.28	DYS437	Count	Frequency	52	5	0.24
25	399	6.84	13	3	0.14	DYS549	Count	Frequency
25.1	331	5.67	14	458	21.97	9	1	0.05
26	167	2.86	15	1123	53.86	10	16	0.77
26.1	368	6.31	16	491	23.55	11	136	6.52
27	64	1.10	17	10	0.48	12	993	47.63
27.1	191	3.27	DYS438	Count	Frequency	13	773	37.07
28	15	0.26	8	5	0.24	14	150	7.19
28.1	69	1.18	9	75	3.60	15	16	0.77
29	1	0.02	10	675	32.37	DYS570	Count	Frequency
29.1	22	0.38	11	150	7.19	14	4	0.19
30.1	3	0.05	12	1127	54.05	15	18	0.86
DYF403S1a	Count	Frequency	13	51	2.45	16	151	7.24
5	3	0.05	14	1	0.05	17	759	36.39
7	1	0.02	16	1	0.05	18	501	24.02
8	10	0.17	DYS439	Count	Frequency	19	356	17.07
9	105	1.77	9	2	0.10	20	197	9.44
9.1	1	0.02	10	95	4.56	21	71	3.40
10	624	10.53	11	802	38.47	22	21	1.01
10.3	1	0.02	12	908	43.55	23	5	0.24
11	597	10.07	13	242	11.61	24	2	0.10
12	545	9.20	14	32	1.53	26	1	0.05
12.1	1	0.02	15	4	0.19	DYS576	Count	Frequency
12.2	1	0.02	DYS448	Count	Frequency	12	1	0.05
13	813	13.72	16	1	0.05	14	18	0.86
13.3	1	0.02	17	12	0.58	15	71	3.40
14	429	7.24	18	121	5.84	16	446	21.37
14.2	1	0.02	19	1120	54.03	17	722	34.60
15	458	7.73	19.2	2	0.10	18	522	25.01
15.2	1	0.02	20	689	33.24	19	223	10.69
16	884	14.91	20.2	1	0.05	20	71	3.40

Table 3 (Continued)

17	707	11.93	21	113	5.45	21	11	0.53
18	419	7.07	22	12	0.58	22	1	0.05
19	218	3.68	23	1	0.05	23	1	0.05
20	77	1.30	24	1	0.05	DYS612	Count	Frequency
21	20	0.34	DYS449	Count	Frequency	30	1	0.05
22	9	0.15	25	1	0.05	31	5	0.24
29	1	0.02	26	13	0.62	32	14	0.67
DYF403S1b	Count	Frequency	27	61	2.92	33	48	2.30
42	4	0.19	28	162	7.77	34	135	6.46
43	4	0.19	29	397	19.03	35	237	11.33
43.2	1	0.05	29.2	1	0.05	36	572	27.36
44	14	0.67	30	689	33.03	37	514	24.58
45	38	1.83	31	386	18.50	38	353	16.88
46	220	10.58	31.2	1	0.05	39	162	7.75
47	671	32.26	32	196	9.40	40	41	1.96
48	379	18.22	33	113	5.42	41	7	0.33
49	209	10.05	34	47	2.25	42	1	0.05
49.2	1	0.05	35	12	0.58	43	1	0.05
50	240	11.54	36	6	0.29	DYS626	Count	Frequency
51	148	7.12	37	1	0.05	24	7	0.34
52	89	4.28	DYS456	Count	Frequency	25	40	1.93
53	41	1.97	12	7	0.34	26	14	0.68
54	15	0.72	13	39	1.87	27	24	1.16
55	5	0.24	14	377	18.08	28	111	5.35
55.2	1	0.05	15	717	34.39	29	620	29.89
DYF404S1 1	Count	Frequency	16	603	28.92	30	513	24.73
9	1	0.03	17	283	13.57	31	343	16.54
11	18	0.49	18	56	2.69	32	264	12.73
12	358	9.72	19	3	0.14	33	105	5.06
12.2	7	0.19	DYS458	Count	Frequency	34	23	1.11
13	388	10.54	13	11	0.53	35	8	0.39
13.2	38	1.03	14	84	4.03	36	2	0.10
14	1273	34.57	15	444	21.29	DYS627	Count	Frequency
14.2	6	0.16	16	444	21.29	15	5	0.24
15	1067	28.98	16.2	5	0.24	16	47	2.25
15.2	5	0.14	17	684	32.81	17	47	2.25
16	412	11.19	17.2	10	0.48	18	100	4.79
17	95	2.58	18	281	13.48	18.2	2	0.10
18	14	0.38	18.2	4	0.19	19	331	15.86
DYS19	Count	Frequency	19	92	4.41	19.2	1	0.05
11	1	0.05	19.2	1	0.05	20	322	15.43
12	3	0.14	20	17	0.82	21	459	21.99
13	103	4.94	21	5	0.24	21.1	1	0.05
14	1460	70.02	22	3	0.14	22	478	22.90
15	394	18.90	DYS481	Count	Frequency	23	205	9.82
16	92	4.41	16	2	0.10	24	77	3.69
17	28	1.34	17	1	0.05	25	11	0.53
18	4	0.19	18	1	0.05	26	1	0.05
DYS385a	Count	Frequency	19	7	0.34	DYS635	Count	Frequency
9	5	0.24	20	22	1.05	19	1	0.05
10	41	1.97	21	138	6.61	20	69	3.31
11	1129	54.15	22	876	41.97	21	361	17.31
12	170	8.15	23	339	16.24	22	233	11.18
13	403	19.33	24	143	6.85	23	1152	55.25
14	188	9.02	25	245	11.74	24	212	10.17
15	93	4.46	26	173	8.29	25	55	2.64
16	44	2.11	27	111	5.32	26	2	0.10
17	11	0.53	28	18	0.86	DYS643	Count	Frequency
18	1	0.05	29	7	0.34	7	1	0.05
DYS385b	Count	Frequency	30	2	0.10	8	15	0.72
11	17	0.81	31	2	0.10	9	90	4.32
12	12	0.57	DYS518	Count	Frequency	10	1125	53.96
13	116	5.56	33.1	1	0.05	11	241	11.56
13.2	2	0.10	34	8	0.38	12	522	25.04
14	1258	60.28	35	28	1.34	13	75	3.60
15	431	20.65	36	141	6.75	14	15	0.72
16	135	6.47	37	376	18.01	16	1	0.05
17	71	3.40	38	622	29.79	Y-GATA-H4	Count	Frequency
18	30	1.44	39	449	21.50	9	5	0.24
19	12	0.57	40	278	13.31	10	148	7.10
20	3	0.14	41	116	5.56	11	849	40.72
DYS389I	Count	Frequency	42	53	2.54	12	953	45.71
9	1	0.05	43	12	0.57	13	123	5.90
11	12	0.58	44	3	0.14	14	7	0.34
12	553	26.52	45	1	0.05			

Table 4
Uniqueness of Y-STR haplotypes in 2085 Dutch samples.

Uniqueness: # times a haplotype was observed	# occurrences in 2085 samples					
	Min YHRD ^a	PPY ^b	Yfiler ^b	PPY23 ^b	RMY1 + 2 ^b	PPY23 + RMY1 + 2
1	608	938	1650	1929	2052	2065
2	110	145	130	70	15	10
3	46	60	30	4	1	
4	28	24	7	1		
5	16	13	5			
6	16	11	2			
7	6	4	1			
8	5	4				
9	2	3				
10	4	1				
11	2					
12	2	1				
13	1	3	1			
14	1					
15		1				
17	3	2				
18	1	1				
19	1					
22	1					
23	1					
24	1	1				
25	1					
28		1				
30	1					
31		1				
43		1				
52		1				
53	1					
57		1				
63	1					
84	1					
99	1					
107	1					
# different haplotypes	862	1217	1826	2004	2068	2075

^a The minimum YHRD marker set consists of DYS19, DYS385a, DYS385b, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393.

^b See Table 1 for the marker units present in these Y-STR sets.

Table 5
Y-STR haplotype diversity.

Y-STR marker set ^a	<i>n</i>	Haplotype diversity	Variance	Standard deviation
Min YHRD	2085	0.990190879926	0.000000505263	0.000710818237
PPY	2085	0.996830021587	0.000000084913	0.000291397648
Yfiler	2085	0.999797014596	0.00000001017	0.000031889077
PPY23	2085	0.999959494976	0.000000000284	0.000016837934
RMY1 + 2	2085	0.999991714881	0.000000000233	0.000015271408
PPY23 + RMY1 + 2	2085	0.999995397156	0.000000000227	0.000015065196

^a See Table 1 for the marker units present in these Y-STR sets.

haplotypes were seen just once (99.0% unique haplotypes (Table 4), haplotype diversity = 0.999995397156 (Table 5)) and ten were each seen twice (representing ten haplotype pairs), resulting in 2075 different haplotypes for the complete set of 2085 samples.

For these ten haplotype pairs we performed additional analyses using the information of 23 autosomal STR markers [10]. Bonaparte software was used to deduce the most likely family relationship between the two donors residing in one haplotype pair, based on fictive family trees in which one of the donors of a pair was fixed (grey square in Fig. 1) and the other donor was tested for all the other possible male relationships (eight white squares in Fig. 1). When the donors were switched, slightly different $\log_{10}(LR)$ scores were obtained, due to the differences in genotypes and their corresponding allele frequencies in the formulae, but all results were comparable, as expected (results not shown). Based on the $\log_{10}(LR)$ results, we infer that two of the haplotype pairs have a

father/son relationship ($\log_{10}(LR)$ of 8.1 or 10.5), two have a brother/brother relationship ($\log_{10}(LR)$ of 6.3 or 12.2) and the other six are likely to have a more distant relationship than the eight relationships tested in Fig. 1 ($\log_{10}(LR)$ between -28.3 and 1.6). These results were confirmed by RelPair analyses (results not shown). Although the study was designed to sample unrelated individuals, it appears that few family relationships are present in such a large population sample.

4. Concluding remarks

All the commercial multiplexes (PPY, PPY23 and Yfiler) and the redesigned RM Y-STR multiplexes (RMY1 and RMY2) tested in this study functioned well and efficiently generated genotyping data for all 2085 Dutch donors. Very little discordance (0.002%) was detected in our data set, which contained 19 Y-STR marker units

that were present in multiple (two or three) kits. This might be due to little nucleotide variation in the areas around the targeted markers, or companies using similar primers. The percentage of unique haplotypes was 92.5% for the 23 marker units in PPY23, 98.4% for the 15 RM Y-STR marker units, and it was even raised to 99.0% when all 36 marker units were combined, resulting in a very high discriminating power for Y-STR standards.

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. We thank Kaye Ballantyne for her assistance in the haplotype diversity calculations.

References

- [1] P. de Knijff, M. Kayser, A. Caglia, D. Corach, N. Fretwell, C. Gehrig, G. Graziosi, F. Heidorn, S. Herrmann, B. Herzog, M. Hidding, K. Honda, M. Jobling, M. Krawczak, K. Leim, S. Meuser, E. Meyer, W. Oesterreich, A. Pandya, W. Parson, G. Penacino, A. Perez-Lezaun, A. Piccinini, M. Prinz, L. Roewer, Chromosome Y microsatellites: population genetic and evolutionary aspects, *Int. J. Leg. Med.* 110 (1997) 134–149.
- [2] SWGDAM, Report on the current activities of the scientific working group on DNA analysis methods Y-STR subcommittee, *Forensic Sci. Commun.* 6 (2004).
- [3] K.B. Gettings, Yfiler Plus Kit, Improved Haplotype Discrimination Using “Rapidly Mutating” Y-STR Markers in a Large Multiplex Kit, 2014 http://www.cstl.nist.gov/strbase/pub_pres/YfilerPlus_HotTopics%20Workshop_MAAFS_2014.pdf.
- [4] K.N. Ballantyne, M. Goedbloed, R. Fang, O. Schaap, O. Lao, A. Wollstein, Y. Choi, K. van Duijn, M. Vermeulen, S. Brauer, R. Decorte, M. Poetsch, N. von Wurmb-Schwark, P. de Knijff, D. Labuda, H. Vezina, H. Knoblauch, R. Lessig, L. Roewer, R. Ploski, T. Dobosz, L. Henke, J. Henke, M.R. Furtado, M. Kayser, Mutability of Y-chromosomal microsatellites: rates, characteristics, molecular bases, and forensic implications, *Am. J. Hum. Genet.* 87 (2010) 341–353.
- [5] K.N. Ballantyne, V. Keerl, A. Wollstein, Y. Choi, S.B. Zuniga, A. Ralf, M. Vermeulen, P. de Knijff, M. Kayser, A new future of forensic Y-chromosome analysis: rapidly mutating Y-STRs for differentiating male relatives and paternal lineages, *Forensic Sci. Int. Genet.* 6 (2012) 208–218.
- [6] K.N. Ballantyne, A. Ralf, R. Aboukhalid, N.M. Achakzai, M.J. Anjos, Q. Ayub, J. Balazic, J. Ballantyne, D.J. Ballard, B. Berger, C. Bobillo, M. Bouabdellah, H. Burri, T. Capal, S. Caratti, J. Cardenas, F. Cartault, E.F. Carvalho, M. Carvalho, B. Cheng, M.D. Coble, D. Comas, D. Corach, M.E. D’Amato, S. Davison, P. de Knijff, M.C. De Ungria, R. Decorte, T. Dobosz, B.M. Dupuy, S. Elmrghni, M. Gliwinski, S.C. Gomes, L. Grol, C. Haas, E. Hanson, J. Henke, L. Henke, F. Herrera-Rodriguez, C.R. Hill, G. Holmlund, K. Honda, U.D. Immel, S. Inokuchi, M.A. Jobling, M. Kaddura, J.S. Kim, S.H. Kim, W. Kim, T.E. King, E. Klausriegler, D. Kling, L. Kovacevic, L. Kovatsi, P. Krajewski, S. Kravchenko, M.H. Larmuseau, E.Y. Lee, R. Lessig, L.A. Livshits, D. Marjanovic, M. Minarik, N. Mizuno, H. Moreira, N. Morling, M. Mukherjee, P. Munier, J. Nagaraju, F. Neuhuber, S. Nie, P. Nilasitsatoporn, T. Nishi, H.H. Oh, J. Olofsson, V. Onofri, J.U. Palo, H. Pamjav, W. Parson, M. Petlach, C. Phillips, R. Ploski, S.P. Prasad, D. Primorac, G.A. Purnomo, J. Purps, H. Rangel-Villalobos, K. Rebal, B. Rerkamnuaychoke, D.R. Gonzalez, C. Robino, L. Roewer, A. Rosa, A. Sajantila, A. Sala, J.M. Salvador, P. Sanz, C. Schmitt, A.K. Sharma, D.A. Silva, K.J. Shin, T. Sijen, M. Sirker, D. Sivakova, V. Skaro, C. Solano-Matamoros, L. Souto, V. Stenzl, H. Sudoyo, D.S. Court, A. Tagliabracci, D. Taylor, A. Tillmar, I.S. Tsybovsky, C. Tyler-Smith, K.J. van der Gaag, D. Vanek, A. Volgyi, D. Ward, P. Willemse, E.P. Yap, R.Y. Yong, I.Z. Pajnic, M. Kayser, Towards male individualization with rapidly mutating Y-chromosomal STRs, *Hum. Mutat.* 35 (2014) 1021–1032.
- [7] S. Willuweit, L. Roewer, Y chromosome haplotype reference database (YHRD): update, *Forensic Sci. Int. Genet.* 1 (2007) 83–87.
- [8] Y Chromosome Haplotype Reference Database (YHRD), 2013 <http://www.yhrd.org/>.
- [9] O. Lao, E. Altena, C. Becker, S. Brauer, T. Kraaijenbrink, M. van Oven, P. Nurnberg, P. de Knijff, M. Kayser, Clinal distribution of human genomic diversity across the Netherlands despite archaeological evidence for genetic discontinuities in Dutch population history, *Investig. Genet.* 4 (2013) 9.
- [10] A.A. Westen, T. Kraaijenbrink, E.A. Robles de Medina, J. Hartevelde, P. Willemse, S.B. Zuniga, K.J. van der Gaag, N.E. Weiler, J. Warnaar, M. Kayser, T. Sijen, P. de Knijff, Comparing six commercial autosomal STR kits in a large Dutch population sample, *Forensic Sci. Int. Genet.* 10 (2014) 55–63.
- [11] L. Excoffier, H.E. Lischer, Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows, *Mol. Ecol. Resour.* 10 (2010) 564–567.
- [12] E. Altena, et al., The Dutch Y-chromosomal landscape, 2014 (in preparation).
- [13] C.J. Bruinjing-van Dongen, K. Slooten, W. Burgers, W. Wiegnerinck, Bayesian networks for victim identification on the basis of DNA profiles, *Forensic Sci. Int. Genet. Suppl. Ser. 2* (2009) 466–468.
- [14] M. Boehnke, N.J. Cox, Accurate inference of relationships in sib-pair linkage studies, *Am. J. Hum. Genet.* 61 (1997) 423–429.
- [15] M.P. Epstein, W.L. Duren, M. Boehnke, Improved inference of relationship for pairs of individuals, *Am. J. Hum. Genet.* 67 (2000) 1219–1231.
- [16] C. Davis, J. Ge, C. Sprecher, A. Chidambaram, J. Thompson, M. Ewing, P. Fulmer, D. Rabbach, D. Storts, B. Budowle, Prototype PowerPlex(R) Y23 System: a concordance study, *Forensic Sci. Int. Genet.* 7 (2013) 204–208.
- [17] M.H. Larmuseau, N. Vanderheyden, A. Van Geystelen, R. Decorte, A substantially lower frequency of uninformative matches between 23 versus 17 Y-STR haplotypes in north Western Europe, *Forensic Sci. Int. Genet.* 11 (2014) 214–219.
- [18] E.K. Hanson, J. Ballantyne, Comprehensive annotated STR physical map of the human Y chromosome: forensic implications, *Leg. Med. (Tokyo)* 8 (2006) 110–120.
- [19] P. Balaresque, G.R. Bowden, E.J. Parkin, G.A. Omran, E. Heyer, L. Quintana-Murci, L. Roewer, M. Stoneking, I. Nasidze, D.R. Carvalho-Silva, C. Tyler-Smith, P. de Knijff, M.A. Jobling, Dynamic nature of the proximal AZFc region of the human Y chromosome: multiple independent deletion and duplication events revealed by microsatellite analysis, *Hum. Mutat.* 29 (2008) 1171–1180.
- [20] B. Budowle, X.G. Aranda, R.E. Lagace, L.K. Hennessy, J.V. Planz, M. Rodriguez, A.J. Eisenberg, Null allele sequence structure at the DYS448 locus and implications for profile interpretation, *Int. J. Leg. Med.* 122 (2008) 421–427.
- [21] J. Purps, S. Siegart, S. Willuweit, M. Nagy, C. Alves, R. Salazar, S.M. Angustia, L.H. Santos, K. Anslinger, B. Bayer, Q. Ayub, W. Wei, Y. Xue, C. Tyler-Smith, M.B. Bafalluy, B. Martinez-Jarreta, B. Egyed, B. Balitzki, S. Tschumi, D. Ballard, D.S. Court, X. Barrantes, G. Bassler, T. Wiest, B. Berger, H. Niederstatter, W. Parson, C. Davis, B. Budowle, H. Burri, U. Borer, C. Koller, E.F. Carvalho, P.M. Domingues, W.T. Chamoun, M.D. Coble, C.R. Hill, D. Corach, M. Caputo, M.E. D’Amato, S. Davison, R. Decorte, M.H. Larmuseau, C. Ottoni, O. Rickards, D. Lu, C. Jiang, T. Dobosz, A. Jonkisz, W.E. Frank, I. Furac, C. Gehrig, V. Castella, B. Grskovic, C. Haas, J. Wobst, G. Hadzic, K. Drobnic, K. Honda, Y. Hou, D. Zhou, Y. Li, S. Hu, S. Chen, U.D. Immel, R. Lessig, Z. Jakovski, T. Ilievski, A.E. Klann, C.C. Garcia, P. de Knijff, T. Kraaijenbrink, A. Kondili, P. Miniati, M. Vouropoulou, L. Kovacevic, D. Marjanovic, I. Lindner, I. Mansour, M. Al-Azem, A.E. Andari, M. Marino, S. Furfuro, L. Locarno, P. Martin, G.M. Luque, A. Alonso, L.S. Miranda, H. Moreira, N. Mizuno, Y. Iwashima, R.S. Neto, T.L. Nogueira, R. Silva, M. Nastainczyk-Wulf, J. Edelman, M. Kohl, S. Nie, X. Wang, B. Cheng, C. Nunez, M.M. Pancorbo, J.K. Olofsson, N. Morling, V. Onofri, A. Tagliabracci, H. Pamjav, A. Volgyi, G. Barany, R. Pawlowski, A. Maciejewska, S. Pelotti, W. Pepinski, M. Abreu-Glowacka, C. Phillips, J. Cardenas, D. Rey-Gonzalez, A. Salas, F. Brisighelli, C. Capelli, U. Toscanini, A. Piccinini, M. Piglionica, S.L. Baldassarra, R. Ploski, M. Konarzewska, E. Jastrzebska, C. Robino, A. Sajantila, J.U. Palo, E. Guevara, J. Salvador, M.C. Ungria, J.J. Rodriguez, U. Schmidt, N. Schlauderer, P. Saukko, P.M. Schneider, M. Sirker, K.J. Shin, Y.N. Oh, I. Skitsa, A. Ampati, T.G. Smith, L.S. Calvit, V. Stenzl, T. Capal, A. Tillmar, H. Nilsson, S. Turrina, D. De Leo, A. Verzeletti, V. Cortellini, J.H. Wetton, G.M. Gwynne, M.A. Jobling, M.R. Whittle, D.R. Sumita, P. Wolanska-Nowak, R.Y. Yong, M. Krawczak, M. Nothnagel, L. Roewer, A global analysis of Y-chromosomal haplotype diversity for 23 STR loci, *Forensic Sci. Int. Genet.* 12 (2014) 12–23.