Forensic Science International: Genetics 23 (2016) 19–24

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

Short communication

D5S2500 is an ambiguously characterized STR: Identification and description of forensic microsatellites in the genomics age

C. Phillips[a,*], W. Parson[b,c], J. Amigo[d], J.L. King[e], M.D. Coble[f], C.R. Steffen[f], P.M. Vallone[f], K.B. Gettings[f], J.M. Butler[f,g], B. Budowle[e,h]

[a] Forensic Genetics Unit, Institute of Forensic Sciences, University of Santiago de Compostela, Santiago de Compostela, Spain
[b] Institute of Legal Medicine, Medical University of Innsbruck, Innsbruck, Austria
[c] Forensic Science Program, The Pennsylvania State University, PA, USA
[d] Galician Public Foundation in Genomics Medicine (FPGMX), Santiago de Compostela, Spain
[e] Institute of Applied Genetics, Department of Molecular and Medical Genetics, University of North Texas Health Science Center, 3500Camp Bowie Blvd., Fort Worth, TX 76107, USA
[f] U.S. National Institute of Standards and Technology, Applied Genetics Group, Biomolecular Measurement Division, 100 Bureau Drive, Gaithersburg, MD 20899, USA
[g] U.S. National Institute of Standards and Technology, Special Programs Office, 100 Bureau Drive, Mail Stop 4701, Gaithersburg, MD 20899, USA
[h] Center of Excellence in Genomic Medicine Research (CEGMR), King Abdulaziz University, Jeddah, Saudi Arabia

ARTICLE INFO

Article history:
Received 30 January 2016
Received in revised form 4 March 2016
Accepted 6 March 2016
Available online 9 March 2016

Keywords:
D5S2500
Short tandem repeats
STRs
miniSTRs
Qiagen HDplex
AGCU 21plex
Non CODIS STRs

ABSTRACT

In the process of establishing short tandem repeat (STR) sequence variant nomenclature guidelines in anticipation of expanded forensic multiplexes for massively parallel sequencing (MPS), it was discovered that the STR D5S2500 has multiple positions and genomic characteristics reported. This ambiguity is because the marker named D5S2500 consists of two different microsatellites forming separate components in the capillary electrophoresis multiplexes of Qiagen's HDplex (Hilden, Germany) and AGCU ScienTech's non-CODIS STR 21plex (Wuxi, Jiangsu, China). This study outlines the genomic details used to identify each microsatellite and reveals the D5S2500 marker in HDplex has the correctly assigned STR name, while the D5S2500 marker in the AGCU 21plex, closely positioned a further 1643 nucleotides in the human reference sequence, is an unnamed microsatellite. The fact that the D5S2500 marker has existed as two distinct STR loci undetected for almost ten years, even with reported discordant genotypes for the standard control DNA, underlines the need for careful scrutiny of the genomic properties of forensic STRs, as they become adapted for sequence analysis with MPS systems. We make the recommendation that precise chromosome location data must be reported for any forensic marker under development but not in common use, so that the genomic characteristics of the locus are validated to the same level of accuracy as its allelic variation and forensic performance. To clearly differentiate each microsatellite, we propose the name D5S2800 be used to identify the Chromosome-5 STR in the AGCU 21plex.

1. Introduction

The D5S2500 microsatellite is a GATA tetra-nucleotide short tandem repeat (STR) with an average level of polymorphism by modern forensic DNA profiling standards. D5S2500 was an integral component of large-scale STR marker sets developed by Marshfield laboratories in the late 1990s for gene mapping [1]. The same marker sets have also been applied to key studies of population variation, so extensive allele frequency and genomic data exists for the D5S2500 locus and over 600 other STRs used for this purpose [2,3]. The D5S2500 locus first appeared as a forensic marker in 2004 as one of six STRs characterized by Huang et al. [4], using PCR primers: 5′-TTAAAGGAGTGATCTCCCCC-3′ and 5′-GTTACAGTACC-TATGGTCATGCC-3′. These sequences closely match the primers listed for the D5S2500 microsatellite in the NCBI database of sequence tagged sites: NCBI Probe (formally UniSTS). In the same year, the D5S2500 locus was included among 27 STRs assessed for their suitability to monitor donor-recipient chimerism in marrow engraftment therapies [5]. The D5S2500 STR subsequently became part of the Mentype® Chimera® 12-STR multiplex (Biotype Diagnostic GmbH, Dresden, Germany), designed to monitor chimerism but with enough sensitivity and novel STRs to be well

* Corresponding author.
E-mail address: c.phillips@mac.com (C. Phillips).

http://dx.doi.org/10.1016/j.fsigen.2016.03.002
1872-4973/

suited for forensic use. The suitability of the Chimera kit as a supplementary forensic DNA test led to several validation and allele frequency studies [6–11]. The 12 STRs of the Chimera kit were then adapted specifically for forensic analysis by Qiagen as the Investigator HDplex kit (Qiagen, Hilden, Germany) [12,13].

Independently, the D5S2500 locus was part of an initiative at the Applied Genetics Group, National Institute of Standards and Technology (NIST, Gaithersburg, MD, USA) begun in 2004, aimed at developing miniaturized STR markers to improve the typing of degraded DNA with shorter amplicons [14,15]. However, the published primer designs for the NIST-developed D5S2500 locus [14,15] target a different microsatellite positioned 1643 nucleotides from the 'true' D5S2500 marker described in NCBI. Therefore, the NIST D5S2500 locus has been incorrectly identified but given the same name. To compound this ambiguity, the NIST D5S2500 locus is also part of a commercial forensic multiplex of 21 non-CODIS STRs developed by AGCU ScienTech (Wuxi, Jiangsu, China) that retains the incorrect D5S2500 name. The AGCU 21plex kit was recently validated as a potentially informative multiplex of supplementary STRs by Zhu et al. [16]. Although the repeat allele numbers and their allele frequencies are quite distinct between the HDplex D5S2500 and the NIST/AGCU D5S2500, the discrepancies between both markers did not draw the attention of the forensic community. Recently, the evaluation of established forensic STRs for massively parallel sequencing (MPS) analysis has highlighted the misidentification of the D5S2500 STR originally developed by NIST and included in the AGCU 21plex [17,18]. The discrepancy was detected by the observation of discordant genotypes for the standard 9947A control DNA.

This report details the genomic characteristics of both STRs and suggests a new distinct name for the NIST/AGCU locus to differentiate the two loci in all future analyses, whether by capillary electrophoresis or MPS. Each marker is distinguished here by coding the HDplex STR with its NCBI accession number: D5S2500.G08468, and similarly the NIST/AGCU STR as: D5S2500.AC008791. We conclude by outlining a recommended genomic validation framework for any forensic STR of interest but not in common use. Given the capacity of MPS to expand forensic multiplexes to include many novel STRs, the genomic details of new markers must be reported to the same level of detail and accuracy as the current publication guidelines dictate for an STR's population variation and forensic properties.

2. Materials and methods

The following four websites were accessed in January 2016 to compile reference sequence data from GRCh37/GRCh38 human genome assemblies, in order to locate and confirm the identities of HDplex D5S2500.G08468 and NIST/AGCU D5S2500.AC008791 STRs.

(i) 1000 Genomes [http://browser.1000genomes.org/Homo_sapiens/Info/Index]. This portal provides access to the human reference sequence curated by Ensembl as well as the 1000 Genomes Phase III genetic variant database with locus coordinates listed for the GRCh37 human genome assembly

(ii) NCBI dbSNP [http://www.ncbi.nlm.nih.gov/SNP/]. This site provides a catalog of validated short human variants that also includes most of the STRs used in forensic analyses. All dbSNP variants are assigned rs-numbers with D5S2500 = rs111362704 (no rs-number exists for D5S2500.AC008791).

(iii) NCBI Probe (formerly NCBI UniSTS) [http://www.ncbi.nlm.nih.gov/probe/]. This site provides a catalog of microsatellites and other sequence tagged sites in the human genome. When a locus is part of the Marshfield linkage mapping marker sets, the primers used for its amplification are detailed. The NCBI Probe database of sequence tagged sites is directly linked to the NCBI GenBank sequence repository that assigns unique accession numbers to sequence segments used to construct assemblies of the human genome.

Table 1
Defining characteristics of the two forensic STRs named D5S2500. Upper sequence tracts in light gray blocks show NCBI UniSTS primers for D5S2500.G08468, the outermost tract in italic dark gray shows the alternative forward PCR primer used in the initial Chimera multiplex designs [5]. Lower sequence tracts in gray blocks are the two published NIST primer designs for D5S2500.AC008791 listed in Hill et al. ([14] italic dark gray) and Hill et al. 2009 ([15] light gray, common sequence in white).

Locus details		Human genome reference sequence (5' to 3'): repeat region (centre bold) +/- 200 nucleotides of flanking sequence
NCBI GenBank accession number	G08468 (UniSTS ID: 76230)	AGACCAGCCTGGGCAACATAGAGATATCCTGTCTCTACA AAAATTTTAAAAATTAGCTGGACATGGTGGTGCACACCTG TAGACCTGCACA*CCTGTAGATCGCTGGAGCC*CAAACGTT CAAGGTTACAGTGACCTATGGTCATGCCACTGCACTCCA GCCTGGGCAACACAGACTCTGTTTCTAATACATATATTAG AGA**CTATCTATCTATCTATCTATCTATCTATCTATCT ATCTAT**TTTGATCCCATGGGGGAGATCACTCCTTTAATAA TGCAAGAGAGTGAATAAGATGGGGGAGGGTGGCGTGCA ATTGAGAGATGAGCTTTATACTGGAATACCAAGTTTATGG TGTGGAACTCTGGGGTATTACACTGAAATGTTAGGGGTT CTCAGTGGGTTATTTGT
Temporary Name	D5S2500.G08468	
Synonyms	Marshfield: GATA67D03; Whitehead-YAC: CHLC.GATA67D03: rs111362704	
Forensic Multiplexes	Qiagen HDplex / Mentype® Chimera® kits	
9947A control DNA genotype	15,16	
Repeat motif in reference sequence	$[CTAT]_{11}$	
GRCh38 coordinates of sequence shown	5:59401244-59401655	
GRCh37 coordinates of sequence shown	5:58697070-58697481	
NCBI GenBank accession number	AC008791*	TTGT*ATCATCCCTGCAAAGTAAC*GTTTACTGATAAACCAAA TGATGTGCCATAATTATGTTTTATTTATGGAACAACTTTTTG TTTTTTCTGGAGTTATATATTACCTTCTTTATTTGATTATGT GACATTATCACCAATTTTTCTAGACGTCTCCAAAACATAAT TTTTTAAATTTTAAATTTCTGTTGGTACATAATA**GGTAGGTA GGTAGACAGACAGACAGACAGACAGACAGACAGACAGA TAGATAGATAGATTGATTGATT**TATGGGGCCCACGAGATA TTTTCATACAGGCAAGCAATGCATAATAATATCAGGGTAAA CAGGGAATCTATTACTGCAAACATTTACCCTTTAAGTTA*CA AACAATTCAA*TTATATTCTCCAAACAATTCAATTATATTCTC TTAGTTATTTTGAGATGTATGATAAACTACTGTTGACTGTA GTCACCCTGTTGAGCTATCAAATACCAGATCTTATTTGT
Temporary Name	D5S2500.AC008791	
Synonyms	None identified	
Forensic Multiplexes	NIST miniSTR 26plex / AGCU ScienTech 21-plex	
9947A control DNA genotype	14,23	
Repeat motifs in reference sequence	$[GGTA]_3$ $[GACA]_8$ $[GATA]_3$ $[GATT]_3$	
GRCh38 coordinates of sequence shown	5:59402932-59403399	
GRCh37 coordinates of sequence shown	5:58698758-58699225	

* Cloned genomic DNA of 128,182 nucleotides (CITB-H1_2040J22) containing this microsatellite

(iv) University of California, Santa Cruz (UCSC) Genome Browser In-Silico PCR [http://genome.ucsc.edu/cgi-bin/hgPcr]. This site is a virtual PCR tool based on the input of short sequences (e.g. nucleotides closely flanking a region of interest). In-Silico PCR identifies the primer binding sites in the human genome and shows the predicted amplicon as a sequence string. GRCh37 or GRCh38 human genome assemblies can be searched and the coordinates of the predicted amplified sequence are listed for the selected assembly.

3. Results and discussion

3.1. Genomic characteristics of D5S2500.G08468 and D5S2500.AC008791

The genomic characteristics obtained from online data used to identify each STR are summarized in Table 1. The sequence and nucleotide coordinates of the D5S2500 marker in HDplex match those listed in NCBI probe for D5S2500, therefore this STR has the correctly assigned name. In contrast, the D5S2500 marker in the NIST/AGCU multiplexes comprises sequence positioned 1643 nucleotides further (5′–3′, repeat region start-point nucleotides), but the published primers [14,15] generate sequence different to any microsatellite listed in NCBI Probe. Both original studies by NIST [14,15] assign the correct GenBank accession number for the sequence containing this STR.

The repeat structure of HDplex D5S2500.G08468 is $[CTAT]_a$ as reported in [9] with the reference sequence reverse strand description of $[AGAT]_a$ in [5]. D5S2500.AC008791 has the more complex $[GGTA]_a$ $[GACA]_b$ $[GATA]_c$ $[GATT]_d$ motif structure and was originally described as $[GATA]_a$ $[GATT]_b$ in [14], although early prototype primer designs from NIST consisting of CTGTTGGTACA-TAATAGGTAGGTAGGT and gTCGTGGGCCCCATAAATC, were described as overlapping with the first and fourth repeat motifs (underlined sequence).

3.2. Compilation of published forensic protocols and allele frequency studies

A forensic validation of D5S2500.G08468 in the Qiagen HDplex kit was made by Westen et al. in 2012 [19]. This STR's genomic position was mapped by Phillips et al. in 2011, and showed a large genetic distance from the same-arm STRs of D5S818 and CSF1PO [12]. The most up-to-date publications with optimized protocols for the forensic genotyping of D5S2500.AC008791 describe the use of the NIST 26plex of miniSTRs by Hill [20] and the AGCU 21plex kit by Zhu et al. [16]. Scheible et al. evaluated the D5S2500.AC008791 locus for MPS in 2014 [21] along with 46 other STRs, including the entire set of NIST miniSTRs. The PCR primers used by Scheible were the initial NIST designs for D5S2500.AC008791 listed in Section 3.1 (i.e. overlapping the first/fourth repeat motif) and the study reported some allele dropout for this STR [21].

Results of literature searches compiling studies of each STR are summarized in Supplementary Table S1. Excluding publications in Chinese, 20 papers report studies of D5S2500.G08468 (8 as a component of Qiagen HDplex) and 24 papers report studies of D5S2500.AC008791 (17 as a component of AGCU 21plex).

Table 2 and Fig. 1A summarize the allele frequency data obtained from 14 published population studies of D5S2500.G08468, indicating a repeat allele range from 9 to 18 with much rarer 7- and 8-repeat plus 16.1/17.1 intermediate repeat alleles. Table 3 and Fig. 1B summarize the allele frequency data from 10 population studies of D5S2500.AC008791. The only surveys of D5S2500.AC008791 in multiple populations were made by Hill et al. [14] and Abrahams et al. [22]. As a large number of population

Table 2
Allele frequency compilations for D5S2500.G08468 from five East Asian, one African and nine European population surveys. HGDP-CEPH CEPH: Human Genome Diversity Panel, Het: Heterozygosity, Pop: Population, PMID: PubMed ID. Allele frequencies in bold are plotted in Fig. 1A. A full list of PubMed IDs and publication details for all D5S2500.G08468 population studies are given in Supplementary Table S1. One other small E ASN population survey is listed in Supplementary Table S2A.

	East Asian (E ASN)					African	European (EUR)												Average values	
Study	Xie, 2015	Huang, 2004	Zhang, 2013	Ozeki, 2013	Phillips, 2012	Phillips, 2012	Phillips, 2012	Schmid, 2005	Kuzniar, 2006	Henke, 2007	Becker, 2007	Pepinski, 2011	Westen, 2012	Tilmar, 2013	Turrina, 2015				E ASN	EUR
PMID	26264960	15027574	23192632	24112991	24315587			15939167	16280219	17821954	19083767	20457029	22752809	22921958	25205546					
Pop	Han, N China	Han, Wuhan	Han, Shanghai	Japanese	17HGDP-CEPH*	7HGDP-CEPH*	8HGDP-CEPH*	German	Polish	NW European	German	Lithuania	Dutch	Swedish	Italian					
N	1027	284	484	175	227	103	157	123	353	203	524	200	335	217	303	**N**	1289	2415		
7	0	0	0	0	0	**0**	0	0	0	0	0	0	0	0	0.001†	**7**	**0**	< **0.001**		
8	< 0.001†	0	0	0	0	**0**	0	0	0	0	0	0	0	0	0	**8**	< **0.001**	**0**		
9	0.003	0	0.004	0.003	0.009	**0.024**	0	0.004	0.006	0.003	0.005	0.007	0.003	0	0.009	**9**	**0.004**	**0.004**		
10	0.015	0.014	0.010	0.034	0.024	**0.155**	0.073	0.106	0.072	0.106	0.077	0.084	0.093	0.081	0.117	**10**	**0.020**	**0.090**		
11	0.296	0.276	0.273	0.357	0.299	**0.248**	0.269	0.281	0.302	0.258	0.304	0.328	0.261	0.320	0.288	**11**	**0.300**	**0.290**		
12	0.201	0.171	0.157	0.163	0.175	**0.112**	0.168	0.163	0.194	0.186	0.179	0.172	0.184	0.166	0.159	**12**	**0.173**	**0.174**		
13	0.039	0.046	0.058	0.023	0.035	**0.078**	0.070	0.065	0.037	0.044	0.057	0.052	0.070	0.067	0.035	**13**	**0.040**	**0.055**		
14	0.074	0.086	0.082	0.054	0.077	**0.083**	0.054	0.028	0.054	0.064	0.059	0.045	0.082	0.044	0.053	**14**	**0.075**	**0.054**		
15	0.277	0.294	0.314	0.309	0.288	**0.180**	0.253	0.252	0.197	0.232	0.218	0.218	0.219	0.210	0.228	**15**	**0.296**	**0.225**		
16	0.080	0.099	0.086	0.046	0.064	**0.078**	0.089	0.065	0.092	0.077	0.080	0.084	0.064	0.085	0.079	**16**	**0.075**	**0.079**		
16.1	0	0	0	0	0	**0**	0	0.004†	0	0	0	0	0	0	0	**16.1**	**0**	< **0.001**		
17	0.014	0.014	0.016	0.011	0.024	**0.024**	0.025	0.024	0.038	0.024	0.018	0.005	0.019	0.018	0.028	**17**	**0.016**	**0.022**		
17.1	0	0	0	0	0	**0**	0	0.004†	0	0	0	0	0	0	0	**17.1**	**0**	< **0.001**		
18	0.001	0	0.001	0	0.004	**0.019**	0	0.004	0.008	0.005	0.003	0.005	0.004	0.009	0.003	**18**	**0.001**	**0.005**		
Het	78.15%	78.84%	78.47%	74.39%	78.49%	84.95%	81.39%	80.99%	81.29%	82.12%	80.86%	79.64%	82.52%	80.53%	81.50%	**Het**	77.67%	81.20%		

*Frequency estimates from combined populations † Singletons

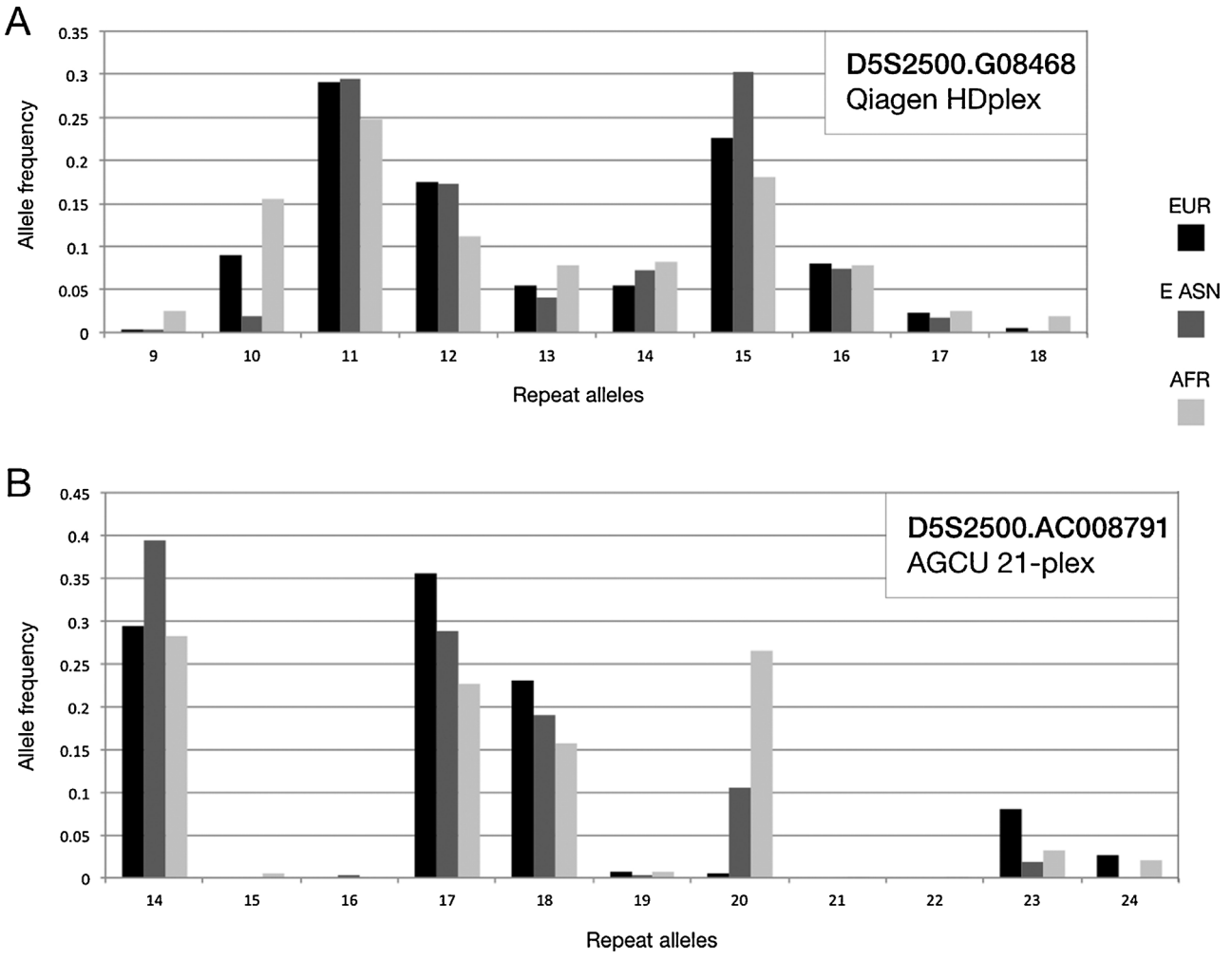

Fig. 1. (A) Compilation of allele frequencies from published population studies of HDplex STR D5S2500.G08468 (shown as bold values in Table 2). EUR = European (average of nine studies), E ASN = East Asian (average of five studies), AFR = African (average of seven populations). (B) Compilation of allele frequencies from published population studies of NIST or AGCU 21plex STR D5S2500.AC008791 (shown as bold values in Table 3). EUR = European (US Caucasians, NIST primers), E ASN = East Asian (Han Chinese, AGCU 21plex), AFR = African (African Americans, NIST primers).

surveys report data in East Asian ethnic populations using the AGCU kit, allele frequency estimates are listed in Supplementary Table S2 from 17 of these studies. A repeat allele range is observed from 14 to 24 with rarer 12- and 13-repeat alleles, 18.2 intermediate repeat alleles, plus 15-, 16-, 22-repeat alleles found in only 1–3 individuals. Observed allele frequencies translate into much higher discrimination power for the HDplex D5S2500. G08468 STR. Estimated Heterozygosity values for D5S2500. G08468 are 81.2% for Europeans; 85% for Africans; 77.7% for East Asians, compared to 72.7%; 77.3%; 71.5%, respectively, for D5S2500. AC008791. However, the complex repeat structure in D5S2500. AC008791 indicates it is likely to be a highly informative STR for forensic MPS analysis, with a confirmed advantage as a short amplicon locus [14–16]. It is noteworthy that the study of Scheible observed a marked increase in informativeness from MPS analysis of sequence variation in D5S2500.AC008791 [21]. This STR was one of three giving double the number of alleles using sequence information instead of fragment size alone [21].

3.3. D5S2503

To avoid further ambiguity, we highlight one other 'D5S250X' locus referenced in the literature as a forensic STR. This is D5S2503 in a 9-plex of novel non-CODIS STRs published in 2014 [23]. Few genomic details for D5S2503 are provided, but the study reports a GATA tetra-nucleotide repeat STR with allele sizes of 350–390 nucleotides. The NCBI Probe primers for D5S2503 locate the sequence at GRCh37 = 5:23591242–23591611 and GRCh38 = 5:23591133–23591502 (>35 Mb separation from both D5S2500 STRs). The reference sequence generated by the NCBI Probe primers has 13 AGAT repeats with size 370 nucleotides, so it highly likely they were used to amplify D5S2503 in the study's PCR [23]. Genomic and population details for D5S2503 are given in Supplementary Tables S3A and S3B.

3.4. A recommended framework for the genomic characterization of novel forensic markers and a suggested unique identifier for D5S2500. AC008791

The existence for almost ten years of two distinct forensic STRs named D5S2500, one correctly identified and the other incorrectly identified but with the same name, emphasizes an obvious need for a properly constructed framework for the genomic identification of forensic STRs. In the first description of D5S2500. AC008791 by NIST the NCBI GenBank accession number was correctly identified for the sequence segment carrying this microsatellite. However, describing its position as 5q11.2 and Chr 5 58.735 Mb (Table 4 of [14]), gives insufficient detail to find

Table 3
Allele frequency compilations for D5S2500.AC008791 from population surveys of Hill et al. [14], Abrahams et al. [22] and eight East Asian populations. Het: Heterozygosity, Pop: Population, PMID: PubMed ID. Allele frequencies in bold are plotted in Fig. 1B. A full list of PubMed IDs and publication details for all D5S2500.AC008791 population studies is given in Supplementary Table S1. Allele frequencies from nine additional Chinese ethnic population studies using the AGCU 21plex kit are listed in Supplementary Table S2B.

Study	Hill, 2008			Abrahams, 2011			Zhu, 2011	Teng, 2012	Yuan, 2012	Yuan, 2012	Shen, 2013	Jin, 2013	Zha, 2014	Yuan, 2014
PMID	18005005			20457088			21042917	21617945	22245836	23065199	23043955	24237828	24041912	24132724
Pop	US Caucasian	US African American	US Hispanic	Afrikaner Caucasian	Asian Indian	Mixed Ancestry[#]	Tibetan	Salar	Han	Tujia	Bai	Korean	Mongolian	Kazak
N	265	259	140	105	112	115	≃200[a]	120	220	107	106	411	523	114
12	**0**	**0**	0	0	0	0	0	0	**0**	0	0	0.002	0	0
13	**0**	**0**	0	0	0	0	0	0	**0**	0.005[†]	0	0	0.001[†]	0
14	**0.294**	**0.282**	0.225	0.301	0.319	0.287	0.351	0.338	**0.393**	0.425	0.396	0.371	0.364	0.276
15	**0**	**0.006**	0	0	0	0.004[†]	0	0	**0**	0	0	0.002	0	0
16	**0**	**0**	0	0	0.014	0	0	0.004[†]	**0.002**[†]	0	0	0.001[†]	0.003	0
17	**0.355**	**0.226**	0.368	0.296	0.269	0.248	0.322	0.338	**0.289**	0.304	0.307	0.345	0.315	0.412
18	**0.230**	**0.156**	0.229	0.306	0.296	0.230	0.216	0.271	**0.190**	0.201	0.184	0.251	0.218	0.206
19	**0.008**	**0.008**	0.007	0	0.009	0.030	0	0.004[†]	**0.002**[†]	0.005[†]	0	0.001[†]	0.008	0.004[†]
20	**0.006**	**0.265**	0.036	0.015	0.046	0.122	0.053	0.017	**0.104**	0.056	0.080	0.024	0.071	0.070
21	**0**	**0.002**[†]	0	0	0	0	0.010	0	**0**	0	0	0	0.007	0
22	**0**	**0.002**[†]	0.007	0	0	0.004[†]	0	0	**0**	0	0	0	0.001[†]	0
23	**0.081**	**0.033**	0.118	0.058	0.042	0.061	0.043	0.029	**0.019**	0.005	0.033	0.002	0.011	0.022
24	**0.026**	**0.021**	0.011	0.024	0.005[†]	0.013	0.005	0	**0**	0	0	0	0.002	0.009
Het	72.72%	77.35%	74.64%	72.73%	73.58%	78.72%	72.17%	69.69%	71.46%	68.34%	70.76%	67.97%	71.55%	70.56%

[a] Estimated value, published study does not give sample size.
[#] Complex admixture of South Asian, European and African ancestries.
[†] Singletons.

the locus. A similar lack of precision is repeated in the recent report of D5S2503 [23], where only the genomic position 5p14 is provided for the STR. Therefore, we propose that published studies of novel forensic markers must provide sufficient genomic data to identify each locus as a unique site with a properly defined position in the human reference sequence (i.e. in a stated genome assembly such as GRCh38). The sequence, chromosome coordinates and the genome assembly used to describe the coordinates given in Table 1 for the two STRs named D5S2500 should be considered an appropriate minimum data framework for microsatellites (most human single nucleotide polymorphisms and insertion-deletion polymorphisms benefit from better positional data compiled in NCBI dbSNP). The details for both STRs were relatively easy to compile from the open-access human genome data resources of 1000 Genomes, Santa Cruz and NCBI. We illustrate the usefulness of such an approach further by applying the same sequence data framework in Supplementary Table S3A to detail D5S2503 [23]: a novel STR of forensic interest that lacks a proper genomic description. Some authors may choose not to publish primer designs for novel loci, but a minimum 200 nucleotides of flanking sequence allows forensic end-users to locate the genomic region and explore sequence characteristics such as repeat structures and flanking variants, that can influence the forensic performance and informativeness of the marker, particularly in MPS. As more markers can be typed with the larger multiplexes MPS offers, likelihood calculations are required to take increasing account of linkage and the precise genomic positioning of forensic loci is now a key step in this process [12].

The incorrectly identified D5S2500.AC008791 locus clearly requires a name that can be applied to its use in a commercially established forensic multiplex and is suitably distinct from D5S2500. During explorations of genomic data in NCBI Probe, microsatellite identifiers from D5S2501 through to D5S2600 were systematically searched and of the few D5S numbers in use, none matched the genomic details of the NIST/AGCU D5S2500.AC008791 STR. To the best of our knowledge, the **D5S2800** locus name has not been applied to any Chromosome-5 STR, therefore we suggest this name be used to identify the NIST/AGCU D5S2500.AC008791 STR from now on.

4. Conclusions

As forensic DNA profiling moves increasingly into the era of genomic analysis, imprecise positional descriptions for new genetic markers such as '5q11.2' or '5p14' are clearly inadequate and, it can be argued, were not sufficiently detailed in the first place. In identifying the correctly named D5S2500 STR for this study, we outline a simple and effective minimum genomic data framework that provides unequivocal identification of any novel locus at a unique position in the human genome.

To address the need to name the incorrectly identified D5S2500 STR that is already part of at least three forensic multiplex designs [16,20,21], we propose D5S2800.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgements

The authors thank Robert Lagacé, Sharon Wootton and Chien-Wei Chang of Life Technologies, Thermo Fisher Scientific, South San Francisco, USA, for very helpful discussions. The authors are also grateful to Chris Tyler Smith, Sanger Institute, Hinxton, UK, for informative advice on the genomic characterization of polymorphic loci of potential forensic use. Points of view in this document are those of the authors and do not necessarily represent the official position or policies of the National Institute of Standards and Technology. Certain commercial equipment, instruments, and materials are identified in order to specify experimental procedures as completely as possible. In no case does such identification imply a recommendation or endorsement by the National Institute of Standards and Technology, nor does it imply that any of the materials, instruments, or equipment identified are necessarily the best available for the purpose.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.fsigen.2016.03.002.

References

[1] K.W. Broman, J.C. Murray, V.C. Sheffield, R.L. White, J.L. Weber, Comprehensive human genetic maps: individual and sex-specific variation in recombination, Am. J. Hum. Genet. 63 (1998) 861–869.

[2] N.A. Rosenberg, J.K. Pritchard, J.L. Weber, H.M. Cann, K.K. Kidd, L.A. Zhivotovsky, M.W. Feldman, Genetic structure of human populations, Science 298 (2002) 2381–2385.

[3] T.J. Pemberton, C.I. Sandefur, M. Jakobsson, N.A. Rosenberg, Sequence determinants of human microsatellite variability, BMC Genomics 10 (2009) 612.

[4] D. Huang, Q. Yang, C. Yu, Allele frequencies of six STR loci (D3S4536, D4S2633, D5S2500, D9S925, D9S1118 and D20S481) in Chinese Han population, J. Forensic Sci. 49 (2004) 413–414.

[5] C. Thiede, M. Bornhäuser, G. Ehninger, Evaluation of STR informativity for chimerism testing—comparative analysis of 27 STR systems in 203 matched related donor recipient pairs, Leukemia 18 (2004) 284–254.

[6] D. Schmid, K. Anslinger, B. Rolf, Allele frequencies of the ACTBP2 (=SE33), D18S51 D8S1132, D12S391, D2S1360, D3S1744, D5S2500, D7S1517, D10S2325 and D21S2055 loci in a German population sample, Forensic Sci. Int. 151 (2005) 303–305.

[7] P. Kuzniar, E. Jastrzebska, R. Ploski, Validation of nine non-CODIS STR loci for forensic use in a population from Central Poland, Forensic Sci. Int. 159 (2006) 258–260.

[8] L. Henke, M. Muche, A. Blaauw, P.H. Van Eede, W. Martin, C. Helmken, B. Budowle, J. Henke, Validation of a new short tandem repeat (STR) fluorescent multiplex system and report of population genetic data, Clin. Lab. 53 (2007) 477–482.

[9] D. Becker, K. Bender, J. Edelmann, F. Götz, L. Henke, S. Hering, C. Hohoff, K. Hoppe, M. Klintschar, M. Muche, B. Rolf, R. Szibor, V. Weirich, M. Jung, W. Brabetz, New alleles and mutational events at 14 STR loci from different German populations, Forensic Sci. Int. Genet. 1 (2007) 232–237.

[10] W. Pepinski, A. Niemcunowicz-Janica, M. Skawronska, J. Janica, Polymorphism of 11 non-CODIS STRs in a population sample of Lithuanian minority residing in northeastern Poland, Forensic Sci. Int. Genet. 5 (2011) e37.

[11] W. Pepinski, M. Skawronska, J. Janica, A. Niemcunowicz-Janica, Polymorphism of 11 non-CODIS STRs in a population sample of religious minority of Old Believers residing in northeastern Poland, Adv. Med. Sci. 55 (2010) 328–332.

[12] C. Phillips, D. Ballard, P. Gill, D. Syndercombe Court, Á. Carracedo, M.V. Lareu, The recombination landscape around forensic STRs: Accurate measurement of genetic distances between syntenic STR pairs using HapMap high density SNP data, Forensic Sci. Int. Genet. 6 (2012) 354–365.

[13] C. Phillips, L. Fernandez-Formoso, M. Gelabert-Besada, M. García-Magariños, J. Amigo, Á. Carracedo, M.V. Lareu, Global population variability in Qiagen Investigator HDplex STRs, Forensic Sci. Int. Genet. 8 (2014) 36–43.

[14] C.R. Hill, M.C. Kline, M.D. Coble, J.M. Butler, Characterization of 26 miniSTR loci for improved analysis of degraded DNA samples, J. Forensic Sci. 53 (2008) 73–80.

[15] C.R. Hill, J.M. Butler, P.M. Vallone, A 26plex autosomal STR assay to aid human identity testing, J. Forensic Sci. 54 (2009) 1008–1015.

[16] B.F. Zhu, Y.D. Zhang, C.M. Shen, W.A. Du, W.J. Liu, H.T. Meng, H.D. Wang, G. Yang, R. Jin, C.H. Yang, J.W. Yan, X.H. Bie, Developmental validation of the AGCU 21 + 1 STR kit: a novel multiplex assay for forensic application, Electrophoresis 36 (2015) 271–276.

[17] F. Wendt, X. Zeng, J. Churchill, J.L. King, B. Budowle, Analysis of short tandem repeat (STR) and single nucleotide polymorphism (SNP) loci from single source samples using a custom HaloPlex Target Enrichment System panel, Am. J. Forensic Med. Pathol. (2016) (accepted manuscript).

[18] W. Parson, D. Ballard, B. Budowle, J.M. Butler, K.B. Gettings, P. Gill, L. Gusmão, D. R. Hares, J. Irwin, J.L. King, P. de Knijff, N. Morling, M. Prinz, P.M. Schneider, C. Van Neste, S. Willuweit, 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. 22 (2016) 54–63.

[19] A.A. Westen, H. Haned, L.J.W. 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.

[20] C.R. Hill, Capillary electrophoresis and 5-channel LIF detection of a 26plex autosomal STR assay for human identification, Methods Mol. Biol. 830 (2012) 17–29.

[21] M. Scheible, O. Loreille, R. Just, J. Irwin, Short tandem repeat typing on the 454 platform: strategies and considerations for targeted sequencing of common forensic markers, Forensic Sci. Int. Genet. 12 (2014) 107–119.

[22] Z. Abrahams, M.E. D'Amato, S. Davison, M. Benjeddou, Allele frequencies of six non-CODIS miniSTR loci (D1S1627, D3S4529 D5S2500, D6S1017, D8S1115 and D9S2157) in three South African populations, Forensic Sci. Int. Genet. 5 (2011) 354–355.

[23] L.M. Pinto, C.L. de Oliveira, L.L. Dos Santos, E. Tarazona-Santos, Molecular characterization and population genetics of non-CODIS microsatellites used for forensic applications in Brazilian populations, Forensic Sci. Int. Genet. 9 (2014) e16–17.