

# Characterization of 26 New miniSTR Loci

Carolyn R. Hill, Michael D. Coble,\* and John M. Butler

National Institute of Standards and Technology, Biochemical Science Division, Gaithersburg, MD 20899-8311

\*Current address: Armed Forces DNA Identification Laboratory, Research Section, Rockville, MD 20850

Email: becky.hill@nist.gov

Phone: 301-975-4275

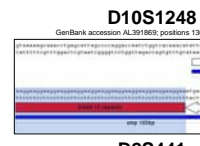
Copy of poster available:

http://www.cstl.nist.gov/biotech/strbase/pub\_pres/Promega2006\_Hill.pdf

A total of 26 novel mini-short tandem repeat (miniSTR) loci have been developed and characterized to aid in the analysis of degraded DNA samples. These new markers produce short PCR products in the target range of 50 – 150 base pairs (bp) by moving the primer sequences as close as possible, if not directly next to the identified repeat region [1]. More than 900 candidate loci were initially screened to determine optimal miniSTR markers based on the following criteria: small amplicon sizes (<125 bp), narrow allele spreads (<24bp), observed heterozygosities (>0.70), and locations on chromosomes unoccupied by the 13 CODIS STR loci, or at least 50 Mb away from them on the same chromosome [2]. The miniSTR loci selected included D1GATA11E02, D1S1627, D1S1677, D2S441, D2S1776, D3S3053, D3S4529, D4S2364, D4S2408, D5S2500, D6S474, D6S1017, D8S1115, D9S1122, D9S2157, D10S1248, D10S1435, D11S4463, D12ATA63A05, D14S1434, D17S974, D17S1301, D18S853, D20S482, D20S1082, and D22S1045. All of these markers were sequenced and evaluated across more than 600 samples, and their population statistics were determined [3]. The heterozygosities of the new loci were compared to those of the 13 CODIS loci and all were found to be comparable. Only seven of the new loci had lower heterozygosity values than the CODIS loci; however, all of these were much smaller in size [3]. This data suggests that these additional 26 miniSTR loci will serve as useful complements to the CODIS loci to aid in the forensic analysis of degraded DNA. In addition, these new loci will be valuable in a variety of scenarios, particularly for paternity cases, missing persons work, or mass fatality DNA identification testing involving kinship samples [2]. In fact, three of these new markers (D10S1248, D2S441, and D22S1045) from the initial six miniSTR loci previously described [2] have recently been recommended for adoption by the European DNA community as new core loci for forensic testing [4,5].

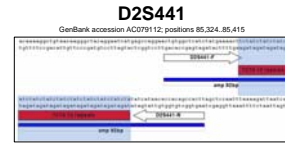
- Butler, J.M., Shen, Y., McCord, B.R. (2003) The development of reduced size STR amplicons as tools for analysis of degraded DNA. *J. Forensic Sci.* 48(6): 1054-1064.
- Coble, M.D., Butler, J.M. (2005) Characterization of new miniSTR loci to aid analysis of degraded DNA. *J. Forensic Sci.* 50(1): 43-53.
- Hill, C.R., Coble, M.D., Butler, J.M. (2006) Development of additional new miniSTR loci for improved analysis of degraded DNA samples. submitted.
- Gill, P., Ferrelly, L., Morling, N., Schneider, P.M. (2006) The evolution of DNA databases—recommendations for new European loci. *Forensic Sci. Int.* 159:242-244.
- Gill, P., Ferrelly, L., Morling, N., Schneider, P.M. (2006) "Letter to the Editor: New multiplexes for Europe – Amendments and clarification of strategic development." *Forensic Sci. Int.* in press.

The repeat unit nomenclature for these STR loci is defined by using the top strand in the GenBank accession reference sequence, selecting the first 5' full tandem repeat, and allowing a single nucleotide change in the repeat structure for a compound repeat (e.g., D22S1045 ATT and ACT). We have altered the nomenclature for D10S1248 and D22S1045 from that previously published in Coble and Butler (2005).

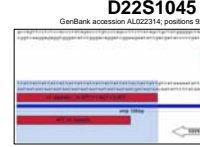


### Allele Frequencies from 1,613 Population Samples

Population	# tested	n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
U.S. Caucasian	265	-	-	-	-	0.06	0.30	0.29	0.19	0.25	0.05	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
African American	257	0.04	0.04	0.04	0.02	0.11	0.24	0.27	0.24	0.10	0.05	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.04
U.S. Hispanic	140	-	-	-	-	-	0.01	0.24	0.28	0.28	0.14	0.04	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Japanese	142	-	-	-	-	0.04	0.04	-	0.13	0.26	0.24	0.10	0.02	0.01	0.04	-	-	-	-	-	-
Spanish	264	-	-	-	-	0.02	0.02	0.02	0.01	0.25	0.31	0.23	0.03	0.04	-	-	-	-	-	-	-
Singapore/Chinese	185	-	-	-	-	0.03	0.03	-	0.09	0.30	0.24	0.10	0.04	0.01	-	-	-	-	-	-	-
Singapore/Malay	182	-	-	-	-	-	-	-	0.02	0.19	0.26	0.21	0.11	0.01	0.03	-	-	-	-	-	-
Singapore/Indian	178	-	-	-	-	-	-	-	0.03	0.20	0.19	0.23	0.17	0.06	0.03	-	-	-	-	-	-



Population	# tested	n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
U.S. Caucasian	265	-	0.10	0.32	0.30	0.01	0.03	0.02	0.09	-	0.46	-	-	-	-	-	-	-	-	-	-
African American	257	-	0.04	0.37	0.03	0.11	0.06	0.04	0.02	0.23	0.02	0.01	0.04	-	-	-	-	-	-	-	-
U.S. Hispanic	140	-	-	0.32	0.18	0.36	0.02	0.04	0.14	-	0.24	-	-	-	-	-	-	-	-	-	-
Japanese	142	-	-	0.29	0.30	0.18	0.22	-	0.03	-	0.17	-	-	-	-	-	-	-	-	-	-
Spanish	264	0.02	-	0.19	0.30	0.12	0.02	0.02	0.19	-	0.36	-	-	-	-	-	-	-	-	-	-
Singapore/Chinese	185	-	-	0.27	0.37	0.02	0.23	-	0.14	-	0.14	-	-	-	-	-	-	-	-	-	-
Singapore/Malay	182	-	-	0.21	0.27	0.10	0.10	0.06	0.11	-	0.18	-	-	-	-	-	-	-	-	-	-
Singapore/Indian	178	0.03	-	0.34	0.32	0.02	0.05	-	0.17	-	0.10	-	-	-	-	-	-	-	-	-	-



Population	# tested	n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
U.S. Caucasian	265	-	-	-	0.14	0.51	0.00	0.08	0.32	0.32	0.07	0.04	-	-	-	-	-	-	-	-	-
African American	257	0.10	-	0.43	0.10	0.06	0.04	0.00	0.29	0.17	0.21	0.16	0.06	0.06	-	-	-	-	-	-	-
U.S. Hispanic	140	-	-	0.18	0.61	0.19	0.01	0.05	0.48	0.11	0.06	0.07	-	-	-	-	-	-	-	-	-
Japanese	142	-	-	-	0.10	-	-	-	0.21	0.37	0.19	0.05	0.07	0.07	-	-	-	-	-	-	-
Spanish	264	-	-	-	0.11	0.04	0.08	0.08	0.42	0.16	0.10	0.04	-	-	-	-	-	-	-	-	-
Singapore/Chinese	185	-	-	-	0.17	-	-	-	0.32	0.36	0.20	0.23	0.19	0.03	-	-	-	-	-	-	-
Singapore/Malay	182	-	-	-	0.17	-	-	-	0.08	0.52	0.32	0.16	0.04	0.06	-	-	-	-	-	-	-
Singapore/Indian	178	-	-	-	0.06	0.24	0.03	0.03	0.36	0.16	0.07	0.03	0.03	0.03	-	-	-	-	-	-	-

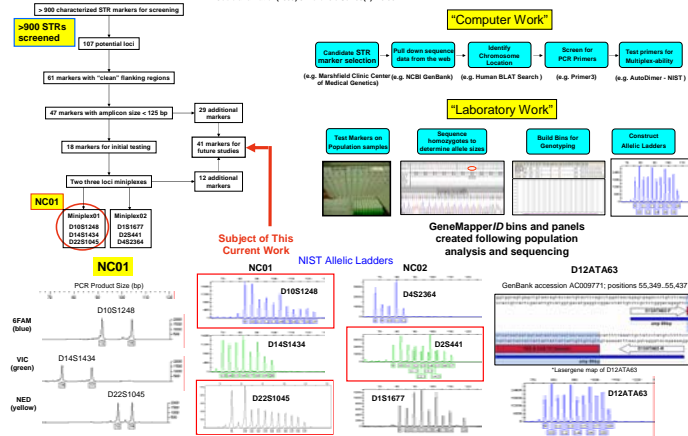
The D22S1045 nomenclature has changed by +3 repeats from Coble and Butler (2005) where only the 14 ATT repeats were considered without the additional 1 ACT and the 13 TAA repeat due to using a different GenBank reference. Some confusion in nomenclature has arisen due to the GenBank accession numbers in Table 1 of Coble and Butler (2005) being different from those used in primer design and allele designation.

Nomenclature adjustments of +3 for D22S1045 and +1 for D10S1248 were made for the following: Japanese data from Asamura et al. (2006) *Int J Legal Med* 120:162-184; Spanish data from Martin et al. (2006) *Forensic Sci Int* in press; Singapore data from Yong et al. (2006) *Forensic Sci Int* in press.

## Selection and Characterization of New miniSTR Loci

The 26 new miniSTR loci were chosen based on their size and location on certain chromosomes. The candidate loci are all either located on chromosomes that differ from the 13 CODIS core loci or are at least ~50 Mb apart from an existing CODIS locus on the same chromosome, and therefore unlinked from that particular marker. The 26 new loci are separated into 9 groups of 3 loci, and are named "NC01" – "NC09", for Non – CODIS loci.

Coble and Butler (2005) *J. Forensic Sci.* 50(1): 43-53



## Miniplex Sets Used in This Study

\*Loci may be used in different combinations in the future\*\*

- NC01: D10S1248, D14S1434, D22S1045
- NC02: D4S2364, D2S441, D1S1677
- NC03: D3S3053, D6S474, D9S482
- NC04: D1GATA11, D2S1776, D4S2408
- NC05: D1S1627, D5S2500, D8S1115,
- NC06: D3S4529, D6S1017, D9S2157
- NC07: D9S1122, D10S1435, D12ATA63,
- NC08: D17S1301, D18S853, D20S1082
- NC09: D17S974, D11S4463

## Rejected Loci and Why

- D6S1027: not in HWE perhaps due to null alleles
- D9S2324: allelic drop-out, complex repeats
- D10S1430: complex repeats
- D10S2327: tri-, quad-allelic profiles
- D14S297: poor heterozygosity
- D15S817: tri-, quad-allelic profiles

\*\*32 loci x 663 samples = 21,216 total data points in this study

For more information, see STRBase:

http://www.cstl.nist.gov/biotech/strbase/newSTRs.htm

http://www.cstl.nist.gov/biotech/strbase/miniSTR.htm

## Characteristics of the new 26 non-CODIS (NC) miniSTR loci are listed below along with the 13 CODIS and 4 additional kit STR loci (D2S1338, D19S433, Penta D, and Penta E).

Locus	GenBank (repeat #)	Chromosomal Position	Location	Observed Size (bp)	Allele Range	Repeat Motif	Heterozygosity					Forward Primer (5'dye labels shown)	Reverse Primer (extra G on 5'end ->> AA)
							Overall	Am. Af.	Cau.	Hisp.	N		
D1GATA11	Z97987 (11)	Chr 7 1,377.7 Mb	1p36.23	81-105	7-13	GATA	0.668	0.673	0.632	0.727	654	[VIC] - TCTTAGCATAGATAGACTTCGCTCC	GTCAACCTTGGAGCTATAGGAA
D1S1627	AC093113 (11)	Chr 1 106.676 Mb	1p21.1	81-100	10-16	ATT	0.746	0.783	0.737	0.693	660	[VIC] - CATGAGATCGGAATACCTATCAAC	GTTTAATTTTCCCAACTCCCA
D1S1677 (NC02)	AL513307 (15)	Chr 1 160.747 Mb	1q23.3	81-117	9-18	TTC	0.746	0.743	0.749	0.743	660	[NED] - TCTTGTTGGTATGAAGCACTGTT	GCACAGGAAGAGCGAAGTAA
D2S441 (NC02)	AC079112 (12)	Chr 2 68.214 Mb	2p14	78-110	9-17	TGTA	0.774	0.798	0.780	0.721	660	[VIC] - CTGTGGCTCATCTGAAAGAACT	GAAGTGGCTGGCTTGTATGAT
D2S1776	AC009475 (11)	Chr 2 169.471 Mb	2q24.3	127-161	7-15	AGAT	0.763	0.740	0.801	0.734	654	[FAM] - TGAACACAGATGTTAAAGTGTTATG	GTCTGAGGTGGCACTTATGAAA
D3S3053	AC069259 (9)	Chr 3 173.234 Mb	3q26.31	84-108	7-13	TATC	0.739	0.713	0.724	0.814	648	[VIC] - TCTTTCCTCTCATGAAGCTACAG	GTTTTGTAATAGAACCACTACAG
D3S4529	AC117452 (13)	Chr 3 85.935 Mb	3p12.1	111-139	13-20	ATTT	0.761	0.752	0.723	0.829	660	[VIC] - CCGAAATATGAAGAAGCAAGACAG	GAGCAAAATATGAAGAAGCAAGACAG
D4S2364 (NC02)	AC022317 (8)	Chr 4 93.976 Mb	4q22.3	67-83	8-12	GRAT	0.511	0.385	0.551	0.664	660	[FAM] - CTGGAGTACTGTGGTATGAT	GCATGAAATAATGAACCAAGTGA
D5S2408	AC110763 (9)	Chr 4 30.981 Mb	4p15.1	85-109	7-13	ATCT	0.722	0.752	0.709	0.691	654	[NED] - AAGGATCAACAGCTTCAAGTAAGGC	GTGAATGACTGAAGTAATGAACCA
D5S2500	AC008791 (17)	Chr 5 58.735 Mb	5p11.2	85-126	14-24	GRWY	0.747	0.757	0.747	0.729	664	[NED] - CTGTGTAGCAATAAGTCTAGTAGTGT	GTCTGGGCCCAATAAATCT
D6S474	AL357514 (17)	Chr 6 112.986 Mb	6p21	107-136	10-17	[AGAT]GATA	0.761	0.765	0.802	0.679	648	[NED] - GGTTTCCCAAGATGACCAACATA	GTCTCTTCAATTAACCTTCACTTAT
D6S1017	AL035588 (10)	Chr 6 41.785 Mb	6p21.1	81-110	6-13	ATCC	0.740	0.807	0.698	0.693	664	[VIC] - CCACCGCTCATATTAGC	GTGAAAAGTAGATATACTGGTGGTG
D8S1115	AC090739 (9)	Chr 8 42.656 Mb	8p11.21	63-96	9-20	ATT	0.663	0.629	0.660	0.729	664	[VIC] - TCCACATCCCAACCAAC	GCCTGAGAGGCTACTGTCAA
D9S1122	AL161789 (12)	Chr 9 76.918 Mb	9q12.1	93-125	9-17	TAGT	0.734	0.753	0.742	0.686	659	[VIC] - GGATATTCAGTAAGCTAGTAGG	CTCTCTGGAAGCTTCTGTTTACC
D9S2157	AL162417 (10)	Chr 9 133.065 Mb	9q34.2	71-107	7-19	ATA	0.844	0.884	0.840	0.779	661	[FAM] - CAAGGAGCACTGCTCTCAAC	GAATATGCTCTCTTGTGTAATAA
D10S1248 (NC01)	AL391869 (13)	Chr 10 130.567 Mb	10q26.3	79-123	8-19	GGAA	0.792	0.825	0.785	0.743	663	[FAM] - TTAATGAAATGAAACAAGTAGG	GCACACTGGTGTATGTTCTTCACT
D10S1435	AL354747 (11)	Chr 10 2.233 Mb	10p15.3	82-139	5-19	TATC	0.766	0.798	0.770	0.700	663	[VIC] - TGGTATATGCAATGAGTTTATTC	CGCTCTCAAAAATGAAGGATAGACA
D11S4463	AP002806 (14)	Chr 11 130.338 Mb	11q25	88-116	10-17	TATC	0.730	0.780	0.676	0.743	664	[FAM	