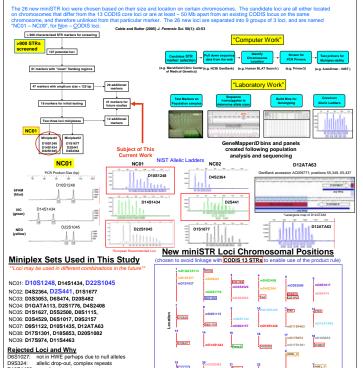


Email: becky.hill@nist.gov Phone: 301-975-4275

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 D1GATA113E02, 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(5): 1054-1064.
- [2] Coble, M.D., Butter, J.M. (2005) Characterization of new miniSTR loci to aid analysis of degraded DNA. J. Forensic Sci. 50(1): 43-53.
 [3] Hill, C.R., Coble, M.D., Butter, J.M. (2006) Development of additional new miniSTR loci for improved analysis of degraded DNA sample
- submitted. [4] Gill, P., Fereday, L., Morting, N., Schneider, P.M. (2006) The evolution of DNA databases--recommendations for new European loci. Forensic Sci.
- 12:244 , Fereday, L., Moring, N., Schneider, P.M. (2006) 'Letter to the Editor: New multiplexes for Europe Amendments and clarification of evelocoment." Forensic Sci. Int. in press.

Selection and Characterization of New miniSTR Loci



AMEL_Y

CODIS PowerPlex 16

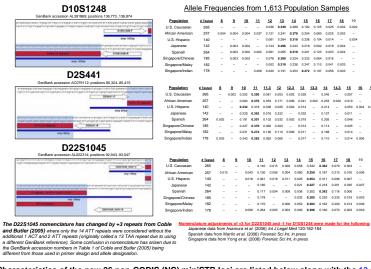
antifiler Sex-Typing

non (MCBI Build

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

> 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).



Poster #44 at Promega Meeting, Nashville, TN, October 10-12, 2006 **REVISED 10/26/06** with final nom

Copy of poster available

http://www.cstl.nist.gov/biotech/strbase/pub pres/Promega2006 Hill.pdf

Issue of Potential Disease Linkage with New and Currently Used STR Loci PubMed searches have noted potential disease linkages with the following STR loci: 1101 – nicolarde padato pasatistica padato padat D21S11 - trisomy-21 (Down's syndrome) D18S51 - trisomy-18 (Edwards' syndrome)



should not prevent the continued use of the STR locus in question." In fact, Colin Kimpton and coworkers from the European DNA Profiling Group recognized early on in the application of STRs for human identity testing that it is likely that many or possibly most STRs will eventually be shown to be useful in following a genetic disease or other genetic trait within a family and therefore this possibility must be recognized at the outset of the use of such systems" [FSI 1995;71:137-152].

The Marshfield genome-scan -400 STR marker set used in searches for disease-causing gene includes TPOX, D7S820, D851179, D13S317, D165S39, and D195433. The majority of the 26 'new' miniSTR loci characterized here are part of the same Marshfield genome-scan STR marker set.

Possibility of Increased Stutter Percentages with Trinucleotide Repeats



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).

		U.t.											
Locus Name	GenBank (repeat #)	Chromosomal Position	Location	Observed Size (bp)	Allele Range	Repeat Motif	Heterozygosity Overall Af. Am. Cau. Hisp.		N	Forward Primer (5'dve labels shown)	Reverse Primer (extra G on 5'end> +A)		
D1GATA113	Z97987 (11)	Chr 1 7.377 Mb	1p36.23	81 - 105	7 - 13	GATA	0.668	0.673	0.632	0.727	654	IVICI - TCTTAGCCTAGATAGATACTTGCTTCC	GTCAACCTTTGAGGCTATAGGAA
D1S1627	AC093119 (13)	Chr 1 106 676 Mb	1p21.1	81 - 100	10 - 16	ATT	0.746	0.783	0.737	0.693	660	[VIC] - CATGAGGTTTGCAAATACTATCTTAAC	GTTTTAATTTTCTCCAAATCTCCA
D1S1627 D1S1677 (NC02)	AL513307 (15)	Chr 1 160.747 Mb	1023.3	81 - 117	9 - 18	TTCC	0.746	0.743	0.749	0.743	660	[NED] - TTCTGTTGGTATAGAGCAGTGTTT	GTGACAGGAAGGACGGAATG
D2S441 (NC02)	AC079112 (12)	Chr 2 68.214 Mb	2p14	78 - 110	9-10	TCTA	0.774	0.798	0.749	0.743	660	IVICI - CTGTGGCTCATCTATGAAAACTT	GAAGTGGCTGTGGGTGTTATGAT
D2S1776	AC009112 (12) AC009475 (11)	Chr 2 169 471 Mb	2024.3	127 - 161	9 - 17 7 - 15	AGAT	0.763	0.790	0.801	0.721	654	[FAM] - TGAACACAGATGTTAAGTGTGTATATG	GTCTGAGGTGGACAGTTATGAA
D3S3053	AC069259 (9)	Chr 3 173.234 Mb	2q24.3 3q26.31	84 - 108	7 - 15	TATC	0.783	0.740	0.724	0.734		IVICI - TCTTTGCTCTCATGAATAGATCAGT	GTTTGTGATAATGAACCCACTCAG
D3S4529	AC117452 (13)	Chr 3 85.935 Mb	3p12.1	111 - 139	13 - 20	ATYT	0.761	0.752	0.723	0.829	660	[VIC] - CCCAAAATTACTTGAGCCAAT	<u>G</u> AGACAAAATGAAGAAACAGACAG
D4S2364 (NC02)	AC022317 (9)	Chr 4 93.976 Mb	4q22.3	67 - 83	8 - 12	GRAT	0.511	0.385	0.551	0.664	660	[FAM] - CTAGGAGATCATGTGGGTATGATT	<u>G</u> CAGTGAATAAATGAAC <u>GAATGGA</u>
D4S2408	AC110763 (9)	Chr 4 30.981 Mb	4p15.1	85 - 109	7 - 13	ATCT	0.722	0.752	0.709	0.691	654	[NED] - AAGGTACATAACAGTTCAATAGAAAGC	GTGAAATGACTGAAAAATAGTAACCA
D5S2500	AC008791 (17)	Chr 5 58.735 Mb	5q11.2	85 - 126	14 - 24	GRYW	0.747	0.757	0.747	0.729	664	[NED] - CTGTTGGTACATAATAGGTAGGTAGGT	<u>G</u> TCGTGGGCCCCATA <u>AATC</u>
D6S474	AL357514 (17)	Chr 6 112.986 Mb	6q21	107 - 136	10 - 17	[AGAT][GATA]	0.761	0.765	0.802	0.679	648	[NED] - GGTTTTCCAAGAGATAGACCAATTA	<u>G</u> TCCTCTCATAAATCCCTACTCATATC
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] - CCACCCGTCCATTTAGGC	<u>G</u> TGAAAAAGTAGATATAATGGTTGGTG
D8S1115	AC090739 (9)	Chr 8 42.656 Mb	8p11.21	63 - 96	9 - 20	ATT	0.663	0.629	0.660	0.729	664	[FAM] - TCCACATCCTCACCAACAC	<u>G</u> CCTAGGAAGGCTACTGTCAA
D9S1122	AL161789 (12)	Chr 9 76.918 Mb	9q21.2	93 - 125	9 - 17	TAGA	0.734	0.753	0.742	0.686	659	[VIC] - GGGTATTTCAAGATAACTGTAGATAGG	GCTTCTGAAAGCTTCTAGTTTACC
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] - CAAAGCGAGACTCTGTCTCAA	GAAAATGCTATCCTCTTTGGTATAAAT
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] - TTAATGAATTGAACAAATGAGTGAG	<u>GCAACTCTGGTTGTATTGTCTTCAT</u>
D10S1435	AL354747 (11)	Chr 10 2.233 Mb	10p15.3	82 - 139	5 - 19	TATC	0.766	0.798	0.770	0.700	663	[FAM] - TGTTATAATGCATTGAGTTTTATTCTG	GCCTGTCTCAAAAATAAAGAGATAGACA
D11S4463	AP002806 (14)	Chr 11 130.338 Mb	11q25	88 - 116	10 - 17	TATC	0.730	0.780	0.676	0.743	664	[FAM] - TCTGGATTGATCTGTCTGTCC	GAATTAAATACCATCTGAGCACTGAA
D12ATA63	AC009771 (13)	Chr 12 106.825 Mb	12q23.3	76 - 106	9 - 19	YAA	0.829	0.788	0.842	0.879	659	[FAM] - GAGCGAGACCCTGTCTCAAG	GAAAAGACATAGGATAGCAATTT
D14S1434 (NC01)	AL121612 (13)	Chr 14 93.298 Mb	14q32.13	70 - 98	13 - 20	CTRT	0.696	0.685	0.721	0.650	663	[VIC] - TGTAATAACTCTACGACTGTCTGTCTG	<u>G</u> AATAGGAGGTGGATGGATGG
D17S974	AC034303 (10)	Chr 17 10.459 Mb	17p13.1	95 - 124	5 - 12	CTAT	0.732	0.757	0.702	0.743	664	[VIC] - GCACCCAAAACTGAATGTCATA	GTGAGAGTGAGACCCTGTC
D17S1301	AC016888 (12)	Chr 17 70.193 Mb	17q25.1	114 - 139	9 - 15	AGAT	0.649	0.626	0.717	0.564	664	[FAM] - AAGATGAAATTGCCATGTAAAAATA	GTGTGTATAACAAAATTCCTATGATGG
D18S853	AP005130 (11)	Chr 18 3.981 Mb	18p11.31	82 - 104	9 - 16	ATA	0.711	0.772	0.645	0.721	664	[NED] - GCACATGTACCCTAAAACTTAAAAT	GTCAACCAAAACTCAACAAGTAGTAA
D20S482	AL121781 (14)	Chr 20 4.454 Mb	20p13	85 - 126	9 - 19	AGAT	0.691	0.673	0.689	0.729	648	[FAM] - CAGAGACACCGAACCAATAAGA	GCCACATGAATCAATTCCTATAATAAA
D20S1082	AL158015 (14)	Chr 20 53.299 Mb	20q13.2	73 - 101	8 - 17	ATA	0.696	0.792	0.653	0.600	664	[VIC] - ACATGTATCCCAGAACTTAAAGTAAAC	GCAGAAGGGAAAATTGAAGCTG
D22S1045 (NC01)	AL022314 (17)	Chr 22 35.779 Mb	22q12.3	82 - 115	8 - 19	ATT	0.784	0.817	0.785	0.721	663	[NED] - ATTTTCCCCGATGATAGTAGTCT	<u>GCGAATGTATGATTGGCAATATTTTT</u>

From Butle	r, J.M. (2006) Genetics and	Heterozygosity									
ocus	GenBank (repeat #)	Chromosome Position	Location	Size (bp)	Alleles	Repeat	Overall	Af. Am.	Cau.	Hisp.	N
CSF1PO	X14720 (12)	Chr 5 149.436 Mb	5q33.1	276 - 320	5 - 16	TAGA	0.745	0.759	0.733	0.743	65
GA	M64982 (21)	Chr 4 155.866 Mb	4q31.3	196 - 348	12.2 - 51.2	CTTT	0.886	0.883	0.889	0.886	65
TH01	D00269 (9)	Chr 11 2.149 Mb	11p15.5	160 - 204	3 - 14	TCAT	0.745	0.759	0.721	0.764	65
POX	M68651 (11)	Chr 2 1.472 Mb	2p25.3	209 - 257	4 - 16	GAAT	0.707	0.763	0.668	0.679	65
/WA	M25858 (18)	Chr 12 19.83 Mb	12p13.31	152 - 212	10 - 25	TCTR	0.826	0.802	0.836	0.850	65
03S1358	AC099539 (16)	Chr 3 45.557 Mb	3p21.31	97 - 149	8 - 21	TCTR	0.763	0.767	0.763	0.757	65
5S818	AC008512 (11)	Chr 5 123.139 Mb	5q23.2	134 - 178	7 - 18	AGAT	0.721	0.735	0.702	0.729	65
75820	AC004848 (13)	Chr 7 83.433 Mb	7q21.11	253 - 297	5 - 16	GATA	0.806	0.763	0.817	0.864	6
8S1179	AF216671 (13)	Chr 8 125.976 Mb	8q24.13	123 - 175	7 - 20	TCTR	0.774	0.763	0.779	0.786	6
13\$317	AL353628 (11)	Chr 13 81.620 Mb	13q31.1	193 - 237	5 - 16	TATC	0.747	0.693	0.748	0.843	6
016S539	AC024591 (11)	Chr 16 84.944 Mb	16q24.1	233 - 277	5 - 16	GATA	0.766	0.786	0.733	0.793	6
018S51	AP001534 (18)	Chr 18 59.100 Mb	18q21.33	264 - 394	7 - 40	AGAA	0.876	0.860	0.870	0.914	6
021S11	AP000433 (29)	Chr 21 19.476 Mb	21q21.1	138 - 256	12 - 41.2	TCTR	0.844	0.829	0.844	0.871	6
02S1338	AC010136 (20)	Chr 2 218.705 Mb	2q35	289 - 341	15 - 28	TKCC	0.882	0.903	0.882	0.843	6
019S433	AC008507 (16)	Chr 19 35.109 Mb	19q12	106 - 140	9 - 17.2	AAGG	0.803	0.876	0.752	0.764	6
Penta D	AP001752 (13)	Chr 21 43.880 Mb	21q22.3	376 - 449	2.2 - 17	AAAGA					
Penta E	AC027004 (5)	Chr 15 95.175 Mb	15q26.2	379 - 474	5 - 24	AAAGA					

Summarv

- New miniSTR markers are being characterized and information will be made available on STRBase (http://www.cstl.nist.gov/biotech/strbase/newSTRs.htm).
- Several of these miniSTR loci have been recommended for adoption by the European DNA community as new core loci. (Gill et al. 2006)
- · In addition to increasing the successful typing of degraded materials, these loci can also provide additional discrimination in complex paternity cases or missing persons cases.

Acknowledgments and Disclaimer

This project was funded by the National Institute of Justice through interagency agreement 2003-IJ-R-029 to the NIST Office of Law Enforcement Standards. Points of view are those of the authors and do not necessarily represent the official position or policies of the US Department of Justice. Certain commercial neutosanty represent nue official position or policies or nue do Depariment of obside; Centar Commercial equipment, instruments and materials are identification imply a recommendation or endorsement 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. We thank Margaret Kline, Jan Redman, Richard Schoske, Peter Vallone, and Amy Decker for initial preparation and quantitation of the NIST U.S. population samples.

For more information, see STRBase: http://www.cstl.nist.gov/biotech/strbase/newSTRs.htm

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

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

D10S1430: complex repeats D10S2327: tri-, quad-allelic profiles D14S297: poor heterozygosity

D15S817: tri-,quad-allelic profiles