Recent Developments in Y-Short Tandem Repeat and Y-Single Nucleotide Polymorphism Analysis

J. M. B
National Institute of Standards and Technology
Gaithersburg, Maryland
United States of America

TABLE OF CONTENTS

1042-7201/15-02/July 2003/91–111/$10.50 •

I. Y-SHORT TANDEM REPEAT MARKERS

A. Marker Discovery

In 1992, Lutz Roewer and colleagues described the first polymorphic Y-chromosome marker Y-27H39 — now better known as the STR locus DYS19 [73]. For the next ten years, discovery of polymorphic tandem repeat markers on the Y-chromosome progressed much more slowly than for their autosomal counterparts. The year 2002 began with only about 30 markers available to researchers (**Table 2**). In the last year or so, the Y-chromosome has been combed to uncover new STR markers and as of February 2003, information on more than 200 markers has been deposited in the Genome Database (GDB; http://www.gdb.org). The rapid growth in the discovery of new Y-STR markers is a direct result of the availability of DNA sequence information from the Human Genome Project and improved bioinformatics tools for searching DNA sequence databases [3]. Previously, extensive laboratory work was required to uncover new polymorphic Y-chromosome markers such as that described in White et al. [96]. However, much lab work remains to be done with these newly identified markers to determine their relative utilities.

In 1997, the European forensic community settled on a core set of Y-STR markers or "minimal haplotype" that includes DYS19, DYS389I/II, DYS390, DYS391, DYS392, DYS393, and DYS385 a/b with YCAII a/b as an optional marker to create an "extended haplotype" [19,52,75]. Most Y-chromosome data to date has been generated W039 TwrDYStoS385 o5Auary 200491 Working Group on DNA Analysis Methods (SWGDAM) selected a core set of markers that includes in the minimal haplotype plus DYS438 and DYS439. TwrDYStoS3 are available in commercial Y-STR kits (see below). Although new markers will be added to databases as their value is demonstrated and they become part of commercially available kits, TwrDYS11 established markers are likely to continue to be important in future Y-STR research.

B. Chromosomal Locations of Markers

The efforts of the Human Genome Project have generated a publicly available human Y-chromosome sequence that is approximately 51 megabases (Mb) in size. However, a "heterochromatin" region around 20 Mb in size toward the end of the long arm of the Y-chromosome may never be completely deciphered [46,90]. The

Table 2. History of Y-STR marker discoveries over the last decade. Most commonly used markers include DYS19, DYS389I/II, DYS390, DYS391, DYS392, DYS393, and DYS385 a/b. Multi-copy markers are listed W039 "a/b" designations if they are duplicated. The total numbers of markers available are considered by bo39 primer pair used to generate them and by products produced

Year	No. available (W039 multicopy)	Markers	Ref.
1992	1	DYS19	[73]
1994	5 (8)	YCAI a/b, YCAII a/b, YCAIII a/b (DYS413), DXYS156	[61]
1996	11 (14)	DYS389I/II, DYS390, DYS391, DYS392, DYS393	[74]
1996	14 (17)	DYF371, DYS425, DYS426	[44]
1997	16 (19)	DYS288, DYS388	[52]
1998	17 (21)	DYS385 a/b	[81]
1999	22 (26)	A7.1 (DYS460), A7.2 (DYS461), A10, C4, H4	[96]
2000	28 (32)	DYS434, DYS435, DYS436, DYS437, DYS438, DYS439	[3]
2001	30 (34)	DYS441, DYS442	[40]
2002	33 (37)	DYS443, DYS444, DYS445	[39]
2002	34 (38)	DYS462	[8]
2002	48 (56)	DYS446, DYS447, DYS448, DYS449, DYS450, DYS452, DYS453, DYS454, DYS455, DYS456, DYS458, DYS459 a/b, DYS463, DYS464 a/b/c/d	[72]
2002	177	DYS468-DYS596 (+129)	GDB[a]
2003	227	DYS597-DYS645 (+50)	GDB[a]

[a] GDB: Genome Database (*see* http://www.gdb.org).

assembled human Y-chromosome sequence may be downloaded from the University of California-Santa Cruz Genome Bioinformatics website (http://genome.ucsc.edu) or the National Center for Biotechnology Information site (http://www.ncbi.nlm.nih.gov).

The availability of a human reference sequence now permits location of the various Y-STR markers along the Y-chromosome. Chromosomal positions were determined by performing a BLAT search [57] using the reference sequences defined in Table 3. The entire search across the human genome was performed in less than a minute for these 50 Y-STR markers, which include loci with published population data as of early 2003.

Relative positions of the tested markers are shown in **Figure 1**. The minimal haplotype loci, which have been used extensively in population studies, are shown on the left side of the chromosome diagram with all of the other markers on the right side. The sex-determining region SRY occshed

where the "a" allele is equal to "b" allele, the resulting peak is usually twice as high during electrophoretic analysis compared to situations where alleles "a" and "b" are not equal in size and can be individually resolved.

C. Characteristics of New Markers

Perhaps the most interesting polymorphic Y-STR discovered to date is DYS464 [72], which has at least four copies on the Y-chromosome and occurs at around 25 Mb near the DAZ region [46]. Analysis of the directionality of DYS464 sequences along the Y-chromosome indicates that it is really a duplicated duplicate locus rather than an independently quadruplicated one. The alleles within each pair are ~225 kilobasepairs (kbp) apart while the pairs are 1.4 Mb apart (Figure 2).

Examples of several peak patterns produced by amplifying the DYS464 a/b/c/d locus with a single primer pair are illustrated in **Figure 3**. While four peaks may be seen with equivalent heights during genotyping when all four alleles can be separated by size, peak patterns are often a more complex set of two or three imbalanced peaks. Thus, allele calls could be made by either taking the peak heights into account (e.g., 12,13,17) or by only considering the actual alleles seen (e.g., 12,13,17).

Many of the new Y-STRs recently discovered have desirable characteristics for forensic analysis. A high degree of polymorphism and a low degree of stutter product formation are valuable characteristics for STR markers when components of mixtures may need to be resolved from one another. The dinucleotide YCAII [61], which is part of the European "extended haplotype" [75], is very polymorphic and does help resolve some common haplotypes. Unfortunately, YCAII has a high degree of stutter because it is a dinucleotide repeat and prone to strand slippage. Multiple stutter products are produced when amplifying YCAII, with some stutter products as high as 50% of the height of the true allele (**Figure 4**).

Penta- and hexanucleotide repeat loci exhibit a much lower degree of stutter and are therefore desirable in assays used for analysis of forensic evidence [9]. Of the 14 new Y-STR markers described by Redd et al. [72], five are pentanucleotides and one contains a hexanucleotide repeat (*see* **Table 3**). Electropherograms from the pentanucleotide DYS447 and the hexanucleotide DYS448 shown in Figure 4 illustrate that both these markers have less than 2% stutter. DYS447 and DYS448 also rank well in terms of allelic diversity against other markers tested in the same sample set [72, Schoske R, personal communication].

performed with new markers has grown, many of these studies have not evaluated sample sets across all of the

the testing for Ancestry.com and owns a company named

DYS392, DYS393, DYS385 a/b, YCAII a/b, DYS426, DYS437, DYS438, DYS439, DYS460, DYS461, DYS462, GGAAT1B07, Y-GATA-A4 (DYS439), A10, C4, and H4.

II. Y-STR TYPING ASSAYS AND KITS

A. Approaches to Reliable Genotyping

Reliable Y-STR typing results may be obtained in one of three different approaches as illustrated in **Figure 5**. When STR markers are first discovered and are being evaluated in research laboratories, typing of samples is often performed with fixed bin genotyping macros that rely on high run-to-run precision and internal size standards (Figure 5, panel A). This approach easily accommodates new alleles as they are discovered. A sequenced reference sample, containing only one of the alleles, can be used to calibrate repeat number to PCR product size under particular electrophoretic conditions. For example, a sample containing 14 TAGA repeats at DYS19 may size at 246.50 bp; a template with 4-bp increments across the expected allele range could then be used to convert measured size into repeat number.

The most commonly used method in forensic laboratories involves allelic ladders where samples are compared to a set of common alleles run under the same electrophoretic conditions [9,24,88] (Figure 5, panel B). The ladder is run with each batch of samples and contains the same internal size standard as the individual samples being tested. Allele sizes in the ladder sample are then compared to sequentially run samples. Each allele in the allelic ladder should be sequenced and the alleles should span the expected range of common alleles [24]. A company supplying the allelic ladder as part of a kit typically performs sequencing of the alleles in the ladder. The major advantage of using an allelic ladder is that results can ou, IL)5 Tw (labolocus-s25 ifthfTc ckedde(LSBs(Figure 5, pammodates)

The first multiplex developed in our laboratory involved 10 loci: DYS19, DYS391, DYS392, DYS435, DYS436, DYS437, DYS438, DYS439, Y-GATA-A7.1 (DYS460), and Y-GATA-H4. The primers and PCR conditions were described for this multiplex at the International Symposium on Human Identification in October 2000 [77] and made available on our STRBase website (http://www.cstl.nist.gov/biotech/strbase/y_strs.htm). A complete description of the primers and the multiplex design process was published more recently [83]. Laboratories in Finland, Japan, and the United States have performed population studies with this multiplex [34,49,91]. These loci were selected to examine newly discovered markers [3,96] for evaluation and possible use in additional assays. The Y-STRs DYS435 and DYS436 showed little variation in the samples tested and were therefore dropped from consideration.

The Y-STR 20plex assay developed in the summer of 2001 includes the 11 markers of the European extended haplotype, the trinucleotide loci DYS388 and DYS426, the tetranucleotide loci DYS437, DYS439, GATA A7.1 (DYS460) and H4, the pentanucleotide loci DYS438 and DYS447, and the hexanucleotide marker DYS448 [12]. Efforts were made to avoid X-chromosome homology in the primer design, particularly in the case of DYS391 [15,27]. PCR product size ranges were packed together through careful examination of known allele ranges in order to keep all alleles less than 350 bp. Allelic ladders were not created with our original multiplex assays because

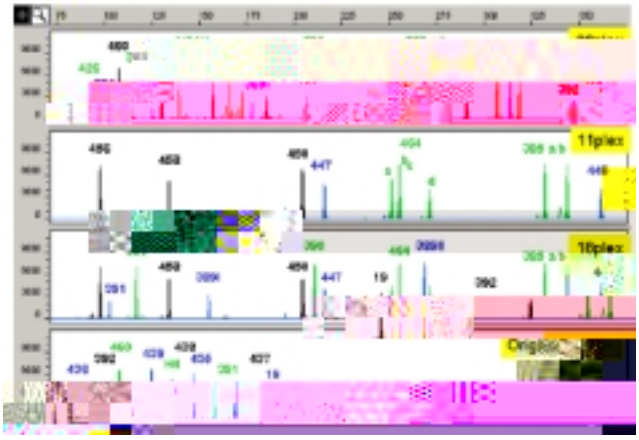


Figure 6. NIST Y-STR multiplexes. The same sample was amplified with four different multiple assays. The DYS marker names are listed above the corresponding PCR product peak.

in many cases we did not know the full allele range or have available alleles to create one. Instead, population data has been collected with a high degree of intralaboratory precision along with sequenced reference materials to correlate sizing results to allele calls (see section on reference materials below).

More recently an 11plex assay has been developed that generates Y-STR amplicons using the markers DYS447, DYS448, DYS450, DYS456, DYS458, DYS385 a/b, and DYS464 a/b/c/d (Schoske, in preparation). The PCR product sizes for these new markers were designed to allow incorporation of the minimal haplotype loci around

Table 5. Comparison of Y-STR markers present in commercial kits and NIST multiplex assays indicated by dye label color

Marker	Y-Plex™ 6	Y-Plex™ 5	PowerPlex® Y	NIST 20plex	NIST 11plex	NIST 10plex
DYS19	Blue		Green	Yellow		Blue
DYS385 a/b	Yellow		Yellow	Green	Green	
DYS389 I		Blue	Blue	Blue		
DYS389 II	Blue					
DYS390	Yellow		Yellow	Green		
DYS391	Yellow		Blue	Blue		Green
DYS392		Yellow	Green	Yellow		Yellow
DYS393	Blue		Yellow	Green		
DYS438		Yellow	Green	Blue		Yellow
DYS439		Green	Blue	Blue		Blue
DYS437			Green	Blue		Yellow
YCAII a/b				Green		
DYS388				Yellow		
DYS426				Green		
DYS435						Blue
DYS436						Blue
DYS447				Red	Blue	
DYS448				Red	Blue	
DYS450					Yellow	
DYS456					Yellow	
DYS458					Yellow	
DYS460 (A7.1)				Yellow		Green
DYS464 a/b/c/d					Green	

the 11plex amplicons. **Figure 6** illustrates the NIST Y-STR multiplex assays completed as of Fall 2002. The original 10plex, published 20plex, new 11plex, and an 18plex that combines the minimal haplotype loci and the 11plex markers are shown using the same male DNA sample. These multiplex assays, particularly the 20plex and 11plex, have allowed our laboratory to rapidly generate population data on hundreds of samples and directly evaluate which markers are most polymorphic in the same sample set (Schoske, in preparation). However, most forensic laboratories are more comfortable with using commercial kits due to primer quality control issues and

Table 6. Characteristics of 246 Y-SNP markers [98]. (*See* also http://ycc,biosci.arizona.edu/nomenclature_system/data.html.)

Marker	Ancest/Der	YCC Hg	Marker	Ancest/Der	YCC Hg	Marker	Ancest/Der	YCC Hg	Marker	Ancest/Der	YCC Hg
M2	A->G	E3a	M70	A->C	K2	M138	C->T	H1c	M207	A->G	R
M3	C->T	Q3	M71	C->T	A2	M139	5G->4G	B-R	M208	C->T	
M4	A->G	M	M72	A->G	I1a3	M141	T->A	A2	M209	A->G	
M5	C->T	M	M73	2-bp DE[a]	R1b4	M143	G->T	Q2	M210	A->T	
M6	T->C	A2	M74	G->A	P-R	M144	T->C	A3b	M211	C->T	B2b4b
M7	C->G	O3d	M75	G->A	E2	M145	G->A	D-E	M212	C->A	
M8	G->T	C1	M76	T->G	L1	M146	A->C	B1	M213	T->C	F-R
M9	C->G	K-R	M77	C->T	C3c	M147	1-bp IN[a] (T)	K3	M214	T->C	O
M10	T->C	E3a6	M78	C->T	E3b1	M148	A->G	Eb1a	M215	A->G	
M11	A->G	L	M81	C->T	E3b2	M149	G->A	E3a3	M216	C->T	C
M12	G->T	J2e	M82	-2bp	H1	M150	C->T	B2a	M217	A->C	C3
M13	G->C	A3b2	M82	-2bp	H1b	M151	G->A	D2b2	M218	C->T	
M14	T->C	A2	M85	C->A	E2b	M152	C->T	B2a1	M219	T->C	
M15	9-bp IN[a]	D1	M86	T->G	C3c	M153	T->A	R1b6	M220	A->G	A3b
M16	C->A	M2a	M87	T->C	R1a1c	M154	T->C	E3a4	M221	G->A	
M17	4G->3G	R1a1	M88	A->G	O2a1	M155	G->A		M223	C->T	
M18	2-bp IN[a]	R1b1	M89	C->T	F-R	M156	A->G	E3a6	M224	T->C	
M19	T->A	Q3a	M90	C->G	E2b	M157	A->C	R1a1b	YAP	Alu- ->Alu+	D-E
M20	A->G	L	M91	9T->8 T	A	M158	G->A	J2d	P1	C->T	E3a
M21	A->T	I1a2	M92	T->C	J2f1	M159	A->C	O3c	P2	C->T	E3
M22	A->G	L	M93	C->T	C3a	M160	A->C	R1b7	P3	G->A	A2
M23	A->G	A2	M94	C->A	B-R	M161	C->A	I1b2a	P4	C->T	A2
M25	G->C	Q2	M95	C->T	O2a	M163	A->C	J2f2	P5	C->T	A2
M26	G->A	I1b2	M96	G->C	E	M164	T->C	O3b	P6	G->C	B2b1
M27	C->G	L1	M97	T->G	H1b	M165	A->G	"E3a5, E3b2b"	P7	T->C	B2b4
M28	T->G	A3a	M98	G->C	E2b	M166	G->A	J2f2	P8	G->A	B2b4a
M30	G->A	B2b3	M99	1-bp DE[a]	J2e1a	M168	C->T	C-R	P9	C->A	C-R
M31	G->C	A1	M101	C->T	O1a	M169	T->C	B2b2	P14	C->T	F-R
M32	T->C	A3	M102	G->C	J2e1	M170	A->C	I	P15	C->T	G2
M33	A->C	E1	M103	C->T	O1b	M171	G->C	A3b2a	P16	A->T	G2a
M34	G->T	E3b3a	M105	C->T	C1	M172	T->G	J2	P18	C->T	G2a1
M35	G->C	E3b	M106	A->G	M	M173	A->C	R1	P19	T->G	I
M36	T->G	H1a	M107	A->G	E3b2a	M174	T->C	D	P20	C DE[a]	G1
M37	C->T	"I1b,R1b2"	M108	T->C	"B2a2, B2b3a"	M175	-5bp	O	P21	C->A	N3a1
M38	T->G	C2	M109	C->T	B2a1	M178	T->C	N3a	P22 (M104)	G/A->A	M2
M39	C DE[a]	H1c	M110	T->C	O1b	M179	C->T		P25	C->A	R1b
M42	A->T	B-R	M111	2-bp (TT) DE[a]	O2a1	M180	T->C		P27	G->A	P-R
M43	A->G	B2a2a	M112	G->A	B2b	M181	T->C	B	P28	C->T	A2b
M44	G->C	E1a	M113	A->G	O3d1	M182	C->T	B2	P29	A->C	E
M45	G->A	P-R	M114	T->C	A2a	M183	A->C		P31	T->C	O2
M47	G->A	J2a	M115	C->T	B2b2	M184	G->A		P33	T->C	C2a
M48	A->G	C3c	M116.2	"A->C, triallelic"	D2b,E3a2	M185	C->T		P36	G->A	Q
M49	T->C	A2	M117	4-bp DE[a]	O3e1	M186	1-bp DE[a]	M	P37	T->C	D2
M50	T->C	O1b	M118	A->T	A3b2b	M188	C->T		P44	G->A	C3
M51	G->A	A3b1	M119	A->C	O1	M189	G->T	M	SRY_{4064}	G->A	E
M52	A->C	H	M120	T->C	Q1	M190	A->G	A3b	SRY_{9138}	C->T	K1
M54	G->A	E2b	M121	5 bp DE	*M6 0 0 7.6 486(C->T)-9919.6(P3 [(C5-3277.7(Q)]TJ -26.4316D-1.2316 TD 022001 Tc 0..0002 Tc [(9 Tt-1491.40 00 29.1(J2a)--3896(M118)4>A)H/F8 1 Tf 06.1(SRH0 TD -0.bp*					

[32]. Until 1997 only about a dozen biallelic markers had been described on the Y-chromosome. These Y-SNPs included sY81 (DYS271) [84], DYS199 (M3) [92], 92R7 [61], and SRY -8299, -1532, -2627 [97]. The use of denaturing high performance liquid chromatography (DHPLC) by Peter Underhill's group at Stanford University for discovery of SNPs has added several hundred more Y-SNPs to the available marker set [93,94,98].

Table 6 lists characteristics for 246 Y-SNP markers [98]. The marker names are listed as "M" numbers were discovered and named by the Stanford group. Marker numbers listed in Table 6 are discontinuous because of selected removal of numbered microsatellite and homopolymer polymorphisms. In addition, markers discovered by other groups, such as Tat (M46), were given Stanford marker numbers and then later removed from the list. Some of these duplicates include YAP (M1), sY81 (M2), P3 (M29), SRY 4064 (M40), SRY 9138 (M177), and SRY 2627 (M167). In addition to the marker name, information on "ancestral" and "derived" allele calls for each Y-SNP are listed in Table 6 along with the haplogroup defined by a derived allele when variation is observed at a particular marker.

B. Unified Nomenclature for Y-Single Nucleotide Polymorphism Haplogroups

One of the biggest problems with Y-SNPs has been the different naming schemes for haplogroup designation developed by the various Y-chromosome research groups around the world. Before 2002, if a "G" (derived state) was observed in a sample when typing the M2 (sY81 or DYS271) marker, then the sample could be reported as belonging to haplogroup (Hg) 8 by Jobling's nomenclature [46], Hg III by Underhill'

Õj 04o5 at 0.0088 Tw (Tw2(j l nclunsTj /F8 gew methodIn addition, (Õ)Tjs (Õ)-26l Tat500pugf
a paTffarcultth Y-SNand cenclature
M4iclud0 -1.1943 TD -0.0037 612((M2many waysasampRoa-0.0037 612((M2manywMa)Tjonc

use of multi-color fluorescence gel or capillary electrophoresis equipment readily available in most forensic DNA laboratories. Inagaki and coworkers [41] examined 15 Y-SNPs in two SNaPshot multiplexes. Markers used in these assays included M9, M105, M122, M125, M128, M130, SRY465, and 8 new Y-SNPs from a Japanese SNP database. They observed 13 different haplogroups in 159 Japanese males [41]. Kayser and coworkers [54] also used SNaPshot to examine M95, M104, M173, M210, and M217 as part of a study of New Guinea populations. At the November 2002 Third International Forensic Y-User Workshop held in Porto, Portugal, the ability to multiplex 35 Y-SNPs in a single SNaPshot assay was reported [79].

Our group at NIST has examined medium-size SNaPshot multiplexes in order to evaluate several dozen

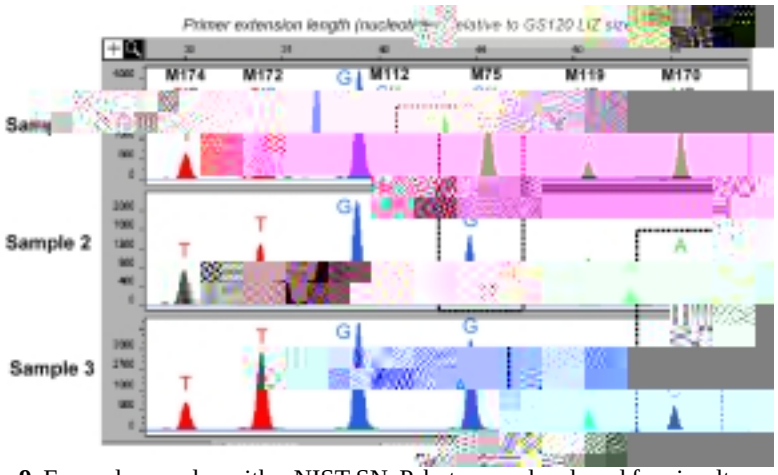


Figure 9. Example samples with a NIST SNaPshot assay developed for simultaneous analysis of 6 Y-SNPs. A 6plex PCR multiplex is the template for the 6plex SnaPshot assay (Vallone and Butler, in preparation). Allele comparisons in boxes are distinguished by size and/or color.

Y-SNPs for relevance to U.S. populations. **Figure 9** demonstrates three samples with different Y-SNP results using a 6plex SNaPshot assay for the markers M75, M112, M119, M170, M172, and M174. We have examined a total of 50 Y-SNPs in appoximately 200 U.S. Caucasian and African American population samples using the SNaPshot and Luminex SNP typing approaches (Vallone and Butler, in preparation).

E. Luminex Assay

Another technology that permits evaluation of Y-SNP markers in a highly multiplexed fashion is based on the Luminex platform with allele-specific hybridization [2]. **Figure 10** illustrates the process in the Luminex assay. PCR is used to amplify the SNP site (e.g., A or G) and to label the PCR product with a fluorescent dye. The labeled

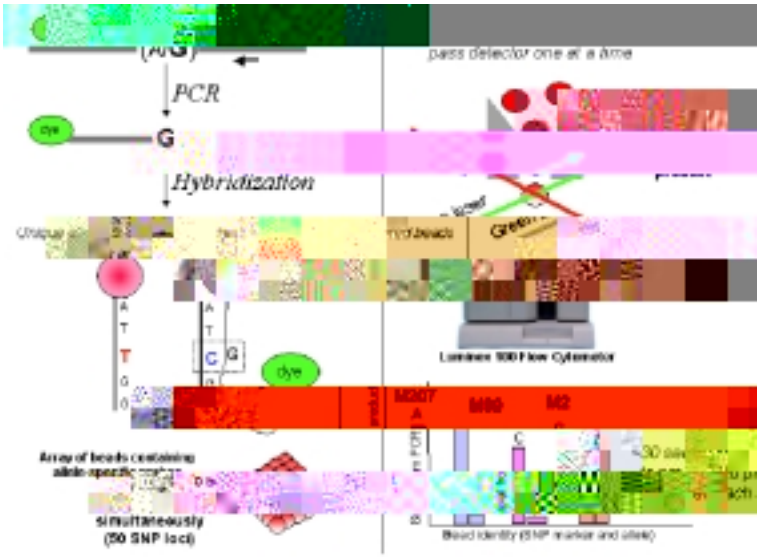


Figure 10. Schematic of Luminex bead hybridization assay for SNP analysis.

PCR product is then hybridized to allele-specific probes attached to latex beads. Oligonucleotide probes for each possible SNP allele are attached to a different color bead. A hundred different bead colors are possible, enabling up to 50 biallelic markers to be examined simultaneously. The beads are then evaluated one at a time through flow cytometry using two different lasers. One laser detects fluorescence from the labeled PCR products and the other evaluates the color of bead passing by the detector. Signal from the PCR product is placed into various bins associated with bead color and hence SNP marker and allele call. The relative amounts of signal from the two possible alleles can be compared to determine the SNP call. Each sample can be processed through the Luminex 100 flow cytometer instrument in approximately 30 seconds. Thus, a 96-well plate can be run in less than an hour.

Marligen Biosciences, Inc. (Ijamsville, MD) has developed a Y-SNP testing kit capable of analyzing 42 Y-SNPs with 5 different multiplex PCR reactions that works on the Luminex platform (see http://www.marligen.com/products/signetysnp.htm). These 42 Y-SNPs define 38 possible haplogroups covering most of the YCC tree (**Figure 11**). Multiplex 1 includes markers that examine the major branch points of the tree, whereas Multiplex 5 markers seek to further differentiate YCC haplogroup R. Note that there is some redundancy in the Marligen kit markers. For example, M42 and M94 (all but Hg A) provide the same information, as do P3 and P4 (Hg A2*). It is also worth noting that not all Y-SNP markers are equally useful in population analysis.

F. Optimal Y-SNP Markers

An analysis of 20 U.S. Caucasian and 20 African American samples with the 42 Marligen Y-SNPs illustrates that most of the markers do not vary in the small sample set shown here (**Table 8**). In fact, only 8 different haplogroups were observed among the 40 samples. However, separation of the population-of-origin (i.e., ethnic discrimination) for the samples is striking. Most of the African American samples are derived at M2 and are thus in the E3a haplogroup while a majority of the U.S. Caucasians are derived at M207 and fall into haplogroup R. A larger study of almost 200 individuals showed similar characteristics (**Figure 12**). While there is a degree of admixture between U.S. populations, Y-SNP markers may be able to play a role in inferring the population-of-origin for a crime-scene stain should that ability be desired in the future [47].

Y-SNP population studies to date have primarily focused on human migration patterns or evolutionary studies [5,7,33,50,51,54,62,63,92,94,95,100,101]. These studies have been conducted with relatively small sample sets from diverse populations. The studies necessary to truly evaluate the forensic relevance of Y-SNPs in larger, more homogeneous population data sets are just getting underway. It is likely that Y-SNPs will be used in a complementary role with Y-STRs rather than as a stand-alone approach for examining male genetic variation in a forensic context.

IV. REFERENCE MATERIALS AND STANDARDIZATION

Reference materials permit calibration of analytical methods as well as monitoring the quality of these methods over time. Need for standardization of information going into DNA databases has stressed the importance of quality reference materials. In addition, allele nomenclatures for typing systems must be consistent so that DNA databases can efficiently exchange information among laboratories.

Table 8. Typing results from 42 Marligen Y-SNPs with 20 African American (AA) and 20 U.S. Caucasian (C) males. Derived alleles are shown with italic. The other 32 Y-SNPs did not vary in the tested samples. Note the redundancy in M207 and M45 and the fact that ethnic discrimination is not 100% with these population samples. YCC haplogroup (Hg) designations (*see* Ref. [98]) and frequencies are on the right side of the table

SWGDM sample	M207 A/G	M45 G/A	M89 C/T	DYS391 C/G	M2 A/G	M170 A/C	M172 T/G	M201 G/T	M153 T/A	SRY10831 A/G	Hg	Fregueny
AA1	A	G	C	*G*	*G*	A	T	G	T	G	E3a	40%
AA2	A	G	C	*G*	*G*	A	T	G	T	G		
AA3	A	G	C	*G*	*G*	A	T	G	T	G		
AA4	A	G	C	*G*	*G*	A	T	G	T	G		
AA6	A	G	C	*G*	*G*	A	T	G	T	G		
AA7	A	G	C	*G*	*G*	A	T	G	T	G		
AA8	A	G	C	*G*	*G*	A	T	G	T	G		
AA10	A	G	C	*G*	*G*	A	T	G	T	G		
AA11	A	G	C	*G*	*G*	A	T	G	T	G		
AA12	A	G	C	*G*	*G*	A	T	G	T	G		
AA15	A	G	C	*G*	*G*	A	T	G	T	G		
AA16	A	G	C	*G*	*G*	A	T	G	T	G		
AA18	A	G	C	*G*	*G*	A	T	G	T	G		
AA19	A	G	C	*G*	*G*	A	T	G	T	G		
AA20	A	G	C	*G*	*G*	A	T	G	T	G		
AA5	A	G	C	*G*	*G*	A	T	G	T	G		
C9	A	G	T	*G*	A	A	T	G	T	G	E3*	3%

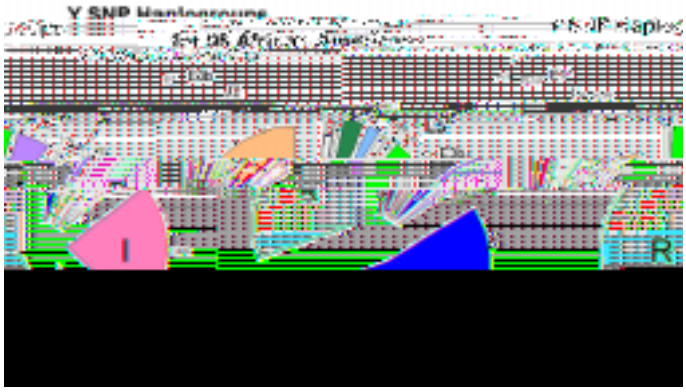


Figure 12. Y-SNP haplogroup frequencies in 95 African American and 94 Caucasian males defined by analysis of 42 Marligen Y-SNPs. Only 15 different groups were observed from 189 individuals.

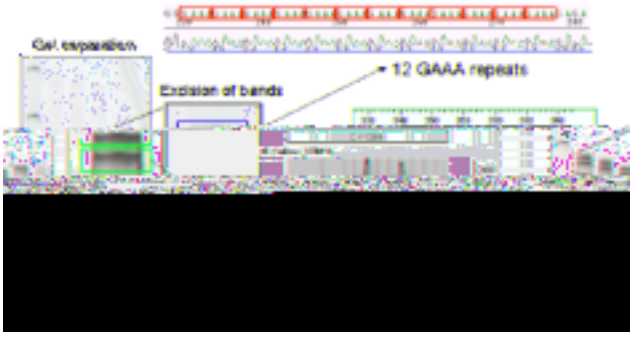


Figure 13. Characterization of DYS385 alleles in SRM 2395 by sequence analysis and Y-Plex™ 6 kit typing.

www.nist.gov/srm). The male samples have been sequenced at more than 20 Y-STR loci and typed at more than 40 Y-SNPs (Butler, in preparation). An example of the sequence information obtained with two DYS385 alleles is shown in **Figure 13**. Laboratories wishing to verify that their assays were run properly with any primer set can use these reference materials. The recent availability of commercial STR kits and their allelic ladders will also promote standardization in allele calls.

B. Allele Nomenclature Issues

One of the major challenges with comparing results from Y-STR markers beyond the well-characterized minimal haplotype loci involves the issue of allele nomenclature. For example, the same DYS439 alleles have been reported three different ways in the literature [3,25,26]. Ayub et al. [3] use only the core variable repeat unit in their allele designations, whereas Griganni and coworkers [26] use seven additional invariant repeat units found upstream of the core variable repeat block. Gonzalez-Neira et al. [25] added two more invariant repeats beyond those used by Griganni in their DYS439 allele nomenclature. Thus, without a common set of rules correlating results between different laboratories can be quite challenging.

The DNA Commission of the International Society of Forensic Genetics (ISFG) published recommendations in July 2001 on Y-STR markers [24]. The guidelines state that Y-STR locus nomenclature should be the DYS number if available. For example, laboratories reporting results for Y-GATA-A7.1 [96] should use its new name DYS460 [8]. This ISFG group also recommended that allelic ladders should span the distance of known allelic variants within each locus with rungs that are one repeat unit apart wherever possible. Ladders should be widely available and contain alleles that have been sequenced.

Regarding allele nomenclature, the ISFG guidelines state that the number of complete repeat units should be counted with partial repeats (variant alleles) being designated by the number of complete repeats separated by a dot followed by the number of bases in the incomplete repeat as is commonly done with autosomal STR markers.

Unfortunately, the designation of some locus nomenclatures take into account the total number of repetitive units (nonvariant plus variant) while others report only the variable repetitive stretches. This presents problems for some markers, such as DYS439. At the Porto meeting in November 2002, it was decided to refer to repeats whenever possible by only the repeats that are immediately adjacent to one another or within a single repeat unit of the core variable repeat. Thus, DYS439 alleles should be called solely by their core repeat unit as done by Ayub et al. [3]. In addition, sequence analysis with DYS439 in chimpanzees has revealed that flanking repeats do not vary, arguing for use of only the core repeat [28,30].

Another potentially problematic locus with future database compatibility is the Y-STR marker GATA-H4 [96]. PCR primers have been published [12] that are internal to some of the invariant repeats reported by Gonzales-Neira et al. [25] and Gusmao et al. [28]. Methods for converting genotypes back and forth when using different primer sets with GATA-H4 need to be carefully considered [28].

C. Validation and Interlaboratory Studies

Validation studies help provide laboratories with performance characteristics for a particular DNA test prior to implementation in forensic casework. Several validation studies have been published or presented on in-house [49,67] and commercial Y-STR kits, such as Y-Plex™ 6 [88]. In addition, interlaboratory studies have been performed to verify that Y-STR systems can be reliably typed among multiple forensic DNA laboratories [14,65,82].

CONCLUSIONS

The field of Y-chromosome analysis and its application to forensic science has undergone rapid improvement in recent years. Male-specific amplification and its use in the analysis of sexual assault DNA evidence as well as missing persons and paternity investigations will likely play an important role in the future of forensic DNA typing. Commercially available kits now enable the forensic practitioner to easily perform Y-STR typing. Validation and interlaboratory studies have demonstrated that Y-STR typing is reliable. With more than 200 Y-STRs and 250 Y-SNPs now available, much remains to be done to understand the value of these new markers relative to the ones widely used today. **Table 9** includes some Internet resources where more information on Y-chromosome research, population data and applications of the techniques described here may be found.

ACKNOWLEDGMENTS

Funding for this work was provided by the National Institute of Justice through an interagency agreement between NIJ and the NIST Office of Law Enforcement Standards. Richard Schoske kindly provided the data for the Y-STR multiplex figures, Peter Vallone developed the Y-SNP assays involving SNaPshot and generated the Y-SNP data using Luminex technology, Margaret Kline and Jan Redman helped prepare many of the population samples used in our studies, and David Duewer provided valuable review of manuscript drafts. Ben Krenke from Promega Corporation kindly supplied the data used for the PowerPlex® Y allelic ladders figure. Alan Redd, Michael Hammer, David Carlson, Mecki Prinz, Debang Liu, Del Price, and Clem Smetana have provided helpful insights or valuable collaborations over the course of our Y-chromosome work at NIST.

REFERENCES

1. Alvarez S, Soledad MM, Lopez AM, de las Heras J, de

amplification of 20 Y-chromosome STR markers; *Forensic Sci Int* 129:10; 2002.
13. Cali F, Forster P, Kersting C, Mirisola MG, D'Anna R, De Leo G, Romano V: DXYS156: a multi-purpose short tandem repeat locus for determination of sex, paternal and maternal geographic origins and DNA fingerprinting; *Int J Legal Med* 116:133; 2002.
14. Carracedo A, Beckmann A, Bengs A, Brinkmann B, Caglia A, Capelli C, Gill P, Gusmao L, Hagelberg C, Hohoff C, Hoste B, Kihlgren A, Kloosterman A, Myhre DB, Morling N, O'Donnell G, Parson W, Phillips C, Pouwels M, Scheithauer R, Schmitter H, Schneider PM, Schumm J, Skitsa I, Stradmann-Bellinghausen B, Stuart M, Syndercombe CD, Vide C: Results of a collaborative study of the EDNAP group regarding the reproducibility and robustness of the Y-chromosome STRs DYS19, DYS389 I and II, DYS390 and DYS393 in a PCR pentaplex format; *Forensic Sci Int* 119:28; 2001.
15. Carvalho-Silva DR, Pena SD: Molecular characterization and population study of an X chromosome homolog of the Y-linked microsatellite DYS391; *Gene* 247:233; 2000.
16. Corach D, Filgueira RL, Marino M, Penacino G, Sala A: Routine Y-STR typing in forensic casework; *Forensic Sci Int* 118:131; 2001.
17. Dau P, Liu D: DNA bracketing locus compatible standards for electrophoresis; U.S. Patent 6,013,444.
18. Dekairelle AF, Hoste B: Application of a Y-STR-pentaplex PCR (DYS19, DYS389I and II, DYS390 and DYS393) to sexual assault cases; *Forensic Sci Int* 118:122; 2001.
19. de Knijff P, Kayser M, Caglia A, Corach D, Fretwell N, Gehrig C, Graziosi G, Heidorn F, Herrmann S, Herzog B, Hidding M, Honda K, Jobling M, Krawczak M, Leim K, Meuser S, Meyer E, Oesterreich W, Pandya A, Parson W, Penacino G, Perez-Lezaun A, Piccinini A, Prinz M, Schmitt C, Schneider PM, Szibor R, Teifel-Greding J, Weichhold GM, Roewer L: Chromosome Y microsatellites: population genetic and evolutionary aspects; *Int J Legal Med* 110:134; 1997.

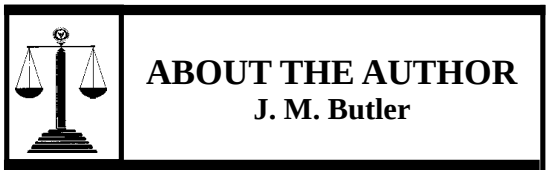
polymorphisms on the human Y-chromosome by microchip electrophoresis; *Electrophoresis* 23:1537; 2002.
43. Jobling MA, Tyler-Smith C: Fathers and sons: the Y-chromosome and human evolution; *Trends Genet* 11:449; 1995.
44. Jobling MA, Samara V, Pandya A, Fretwell N, Bernasconi B, Mitchell RJ, Gerelsaikhan T, Dashnyam B, Sajantila A, Salo PJ, Nakahori Y, Disteche CM, Thangaraj K, Singh L, Crawford MH, Tyler-Smith C: Recurrent duplication and deletion polymorphisms on the long arm of the Y-chromosome in normal males; *Hum Mol Genet* 5:1767; 1996.
45. Jobling MA, Pandya A, Tyler-Smith C: The Y-chromosome in forensic analysis and paternity testing; *Int J Legal Med* 110:118; 1997.
46. Jobling MA, Tyler-Smith C: New uses for new haplotypes the human Y-chromosome, disease and selection; *Trends Genet* 16:356; 2000.
47. Jobling MA: Y-chromosomal SNP haplotype diversity in forensic analysis; *Forensic Sci Int* 118:158; 2001.
48. Jobling MA: In the name of the father — Surnames and genetics; *Trends Genet* 17:353; 2001.
49. Johnson CL, Eisenberg A, Warren JE, Planz J, Warren JH, Staub RW: Validation of a Y-chromosome STR 10-plex; Poster presented at 13th International Symposium on Human Identification, Phoenix, Arizona, October 2002. http://www.promega.com/geneticidproc/ussymp13proc/abstracts/35_staub.pdf

Forensic Sci Int 130:97; 2002.
73. Roewer L, Epplen JT: Rapid and sensitive typing of forensic stains by PCR amplification of polymorphic simple repeat sequences in case work; *Forensic Sci Int* 53:163; 1992.
74. Roewer L, Kayser M, Dieltjes P, Nagy M, Bakker E, Krawczak M, de Knijff P: Analysis of molecular variance (AMOVA) of Y-chromosome-specific microsatellites in two closely related human populations; *Hum Mol Genet* 5:1029; 1996.
75. Roewer L, Krawczak M, Willuweit S, Nagy M, Alves C, Amorim A, Anslinger K, Augustin C, Betz A, Bosch E, Caglia A, Carracedo A, Corach D, Dekairelle A, Dobosz T, Dupuy BM, Furedi S, Gehrig C, Gusmao L, Henke J, Henke L, Hidding M, Hohoff C, Hoste B, Jobling MA, Kargel HJ, de Knijff P, Lessig R, Liebeherr E, Lorente M, Martinez-Jarreta B, Nievas P, Nowak M, Parson W, Pascali VL, Penacino G, Ploski R, Rolf B, Sala A, Schmidt U, Schmitt C, Schneider PM, Szibor R, Teifel-Greding J, Kayser M: Online reference database of European Y-chromosomal short tandem repeat (STR) haplotypes; *Forensic Sci Int* 118:106; 2001.
76. Rolf B, Keil W, Brinkmann B, Roewer L, Fimmers R: Paternity testing using Y-STR haplotypes: assigning a probability for paternity in cases of mutations; *Int J Legal Med* 115:12; 2001.
77. Ruitberg CM, Butler JM: New primer sets for Y-chromosome and CODIS STR loci; *11th International Symposium on Human Identification*; Biloxi, MS; 2000.
78. Ruiz LA, Nayar K, Goldstein DB, Hebert JM, Seielstad MT, Underhill PA, Lin AA, Feldman MW, Cavalli Sforza LL: Geographic clustering of human Y-chromosome haplotypes; *Ann Hum Genet* 60:401; 1996.
79. Sanchez JJ, Hallenberg C, Borsting C, Buchard A, Hernadez A, Morling N: A Y-chromosome SNP multiplex for forensic genotyping; Presentation at 3rd International Forensic Y-User Workshop, Porto, Portugal, November 7, 2002.
80. Santos FR, Epplen JT, Pena SDJ: Testing deficiency paternity cases with a Y-linked tetranucleotide repeat polymorphism; Pena SDJ, Chakraborty R, Epplen JT, Jeffreys AJ (Eds) *DNA Fingerprinting: State of the Science*, pp 261-265; 1993.
81. Schneider PM, Meuser S, Waiyawuth W, Seo Y, Rittner C: Tandem repeat structure of the duplicated Y-chromosomal STR locus DYS385 and frequency studies in the German and three Asian populations; *Forensic Sci Int* 97:61; 1998.
82. Schneider PM, d'Aloja E, Dupuy BM, Eriksen B, Jangblad A, Kloosterman AD, Kratzer A, Lareu MV, Pfitzinger H, Rand S, Scheithauer R, Schmitter H, Skitsa I, Syndercombe-Court D, Vide MC: Results of collaborative study regarding the standardization of the Y- linked STR system DYS385 by the European DNA Profiling (EDNAP) group; *Forensic Sci Int* 102:159; 1999.
83. Schoske R, Vallone PM, Ruitberg CM, Butler JM: Multiplex PCR design strategy used for the simultaneous amplification of 10 Y-chromosome short tandem repeat (STR) loci; *Anal Bioanal Chem* 375:333; 2003.
84. Seielstad MT, Hebert JM, Lin AA, Underhill PA, Ibrahim M, Vollrath D, Cavalli-Sforza LL: Construction of human Y-chromosomal haplotypes using a new polymorphic A to G transition; *Hum Mol Genet* 3:2159; 1994.
85. Shewale JG, Sikka SC, Schneida E, Sinha SK: DNA profiling of azoospermic semen samples from vasectomized males by using Y-PLEX 6 amplification kit; *J Forensic Sci* 48:127; 2003.
86. Shin DJ, Jin HJ, Kwak KD, Choi JW, Han MS, Kang PW, Choi SK, Kim W: Y-chromosome multiplexes and their potential for the DNA profiling of Koreans; *Int J Legal Med* 115:109; 2001.
87. Sibille I, Duverneuil C, Lorin dlG, Guerrouache K, Teissiere F, Durigon M, de Mazancourt P: Y-STR DNA amplification as biological evidence in sexually assaulted female victims with no cytological detection of spermatozoa; *Forensic Sci Int* 125:212; 2002.
88. Sinha SK, Budowle B, Arcot SS, Richey SL, Chakrabor R, Jones MD, Wojtkiewicz PW, Schoenbauer DA, Gross AM, Sinha SK, Shewale JG: Development and validation of a multiplexed Y-chromosome STR genotyping system, Y-PLEX 6, for forensic casework; *J Forensic Sci* 48:93; 2003.
89. Sykes B, Irven C: Surnames and the Y-chromosome; *Am J Hum Genet* 66:1417; 2000.
90. Tilford CA, Kuroda-Kawaguchi T, Skaletsky H, Rozen S, Brown LG, Rosenberg M, McPherson JD, Wylie K, Sekhon M, Kucaba TA, Waterston RH, Page DC: A physical map of the human Y-chromosome; *Nature* 409:943; 2001.
91. Uchihi R, Yamamoto T, Usuda K, Yoshimoto T, Tanaka M, Tokunaga S, Kurihara R, Tokunaga K, Katsumata Y: Haplotype analysis with 14 Y-STR loci using 2 multiplex amplification and typing systems in 2 regional populations in Japan; *Int J Legal Med* 117:34; 2003.
92. Underhill PA, Li J, Zemans R, Oefner PJ, Cavalli-Sforza LL: A pre-Columbian Y-chromosome-specific transition and its implications for human evolutionary history; *Proc Natl Acad Sci USA* 93:196; 1996.
93. Underhill PA, Jin L, Lin AA, Mehdi Q, Jenkins T, Vollrath D, Davis RW, Cavalli-Sforza LL, Oefner PJ: Detection of numerous Y-chromosome biallelic polymorphisms dy denaturing high-performance liquid chromatography; *Genome Res* 7:996; 1997.
94. Underhill PA, Shen P, Lin AA, Jin L, Passarino G, Yang WH, Kauffman E, Bonne-Tamir B, Bertranpetit J, Francalacci P, Ibrahim M, Jenkins T, Kidd JR, Mehdi SQ, Seielstad MT, Wells RS, Piazza A, Davis RW, Feldman MW, Cavalli-Sforza LL, Oefner PJ: Y-chromosome sequence variation and the history of human populations; *Nat Genet* 26:358; 2000.
95. Wells RS, Yuldasheva N, Ruzibakiev R, Underhill PA, Evseeva I, Blue-Smith J, Jin L, Su B, Pitchappan R, Shanmugalakshmi S, Balakrishnan K, Read M, Pearson NM, Zerjal T, Webster MT, Zholoshvili I, Jamarjashvili E,

Shriver MD: Melting curve SNP (McSNP) genotyping — A useful approach for diallelic genotyping in forensic science; *J Forensic Sci* 47:593; 2002.

100. Zerjal T, Wells RS, Yuldasheva N, Ruzibakiev R, Tyler-Smith C: A genetic landscape reshaped by recent events: Y-chromosomal insights into Central Asia; *Am J Hum Genet* 71:466; 2002.

101. Zerjal T, Beckman L, Beckman G, Mikelsaar AV, Krumina A, Kucinskas V, Hurles ME, Tyler-Smith C: Geographical, linguistic, and cultural influences on genetic diversity: Y-chromosomal distribution in Northern European populations;



ABOUT THE AUTHOR

J. M. Butler

John M. Butler received his Ph.D. in chemistry from the University of Virginia (Charlottesville, VA) in 1995. Dr. Butler is currently a research chemist at the U.S. National Institute of Standards and Technology.

Dr. Butler's dissertation work was performed at the FBI Laboratory's Forensic Science Research and Training Center (Quantico, VA). He worked there with Bruce McCord and Bruce Budowle on developing capillary electrophoresis methods for DNA typing involving short tandem repeat (STR) DNA markers and mitochondrial DNA. Dr. Butler then went to the National Institute of Standards and Technology (NIST) where he did postdoctoral research with Dennis Reeder. At NIST, he developed STRBase, which is an Internet-accessible database of information on STR markers used in human identity testing (http://www.cstl.nist.gov/biotech/strbase/). After completing his postdoctoral work, Dr. Butler went to a start-up company in California named GeneTrace Systems where he was a staff scientist and project leader developing the capabilities of time-of-flight mass spectrometry for STR and SNP genotyping assays. In 1999, Dr. Butler returned to NIST where he is currently a project leader developing new methods and technologies to aid forensic DNA typing through funding from the National Institute of Justice.

Dr. Butler has published more than 40 book chapters and peer-reviewed papers and made numerous presentations at national and international scientific conferences with a primary focus on improving technologies for DNA typing. He is a member of the American Society of Human Genetics and the International Society of Forensic Genetics. His recent textbook from Academic Press entitled "Forensic DNA Typing: Biology and Technology behind STR Typing" is gaining wide acceptance as a tool for training students and forensic scientists. Dr. Butler and his wife have four children (all of which have been proved to be theirs through DNA testing).