Peter M. Vallone,[1] Ph.D. and John M. Butler,[1] Ph.D.

Y-SNP Typing of U.S. African American and Caucasian Samples Using Allele-Specific Hybridization and Primer Extension*

ABSTRACT: Multiplex analysis of genetic markers has become increasingly important in a number of fields, including DNA diagnostics and human identity testing. Two methods for examination of single nucleotide polymorphisms (SNPs) with a potential for a high degree of multiplex analysis of markers are primer extension with fluorescence detection, and allele-specific hybridization using flow cytometry. In this paper, we examined 50 different SNPs on the Y-chromosome using three primer extension multiplexes and five hybridization multiplex assays. For certain loci, the allele-specific hybridization method exhibited sizable background signal from the absent alternate allele. However, 100% concordance (>2000 alleles) was observed in ten markers that were typed using both methods. A total of 18 unique haplogroups out of a possible 45 were observed in a group of 229 U.S. African American and Caucasian males with the majority of samples being assigned into 2 of the 18 haplogroups.

KEYWORDS: forensic science, Y-chromosome, single nucleotide polymorphism, SNP typing, Y-SNPs, SNaPshot, primer extension, Luminex, allele-specific hybridization

In recent years, single nucleotide polymorphisms (SNPs) have become more widely used in a variety of applications, including medical diagnostics, population genetics, and human identity testing (1,2). The ability to analyze a number of SNP markers in parallel relies on multiplex amplification and detection formats (3). Two SNP detection formats capable of multiplex analysis are allele-specific primer extension (ASPE) and allele-specific hybridization (ASH), which we use here to examine variation along the Y-chromosome.

The lack of recombination along most of the Y-chromosome makes it a useful tool in human evolutionary studies and assessing male migration patterns (4–10). Y-chromosome markers have also been used in attempts to address some interesting historical questions (11,12). Analysis of Y-SNP markers has typically been done with manual techniques such as restriction digestion of PCR products (10,13) or by procedures that can only examine one or two markers simultaneously such as allele-specific PCR (7), denaturing high performance liquid chromatography (5), melting curve analysis (14), and real-time PCR (15). More recently, microarrays (16), time-of-flight mass spectrometry (17), and fluorescent primer extension (10,18) have been applied to Y-SNP analysis in order to type multiple markers in parallel.

In an effort to evaluate the usefulness of Y-chromosome SNP markers for human identity applications, we constructed several novel multiplex ASPE assays and utilized a new commercial ASH kit to type 50 Y-SNP markers in more than 200 individuals. In addition, by examining ten of the Y-SNPs with both methods, we were able to assess concordance in more than 2000 allele calls between primer extension and hybridization approaches.

Methods

U.S. African American and Caucasian DNA Samples

Anonymous liquid blood samples with self-identified ethnicities were purchased from Interstate Blood Bank, Inc. (Memphis, TN) and Millennium Biotech, Inc. (Ft. Lauderdale, FL) and extracted using a modified salting out procedure (19). Carll Ladd from the Connecticut Forensic Laboratory (Meriden, CT) kindly provided extracted DNA for 20 U.S. Caucasian and 20 African American samples used as part of this study. The extracted DNA was quantified using ultraviolet (UV) spectrophotometry followed by a PicoGreen assay (20) to adjust concentrations to approximately 1 ng/μL. All samples were examined with 15 autosomal short tandem repeats and the amelogenin sex-typing marker using the AmpFℓSTR Identifiler Kit to verify that each sample was unique (21). A total of 229 male samples were typed (115 African Americans and 114 Caucasians).

Y-SNP Markers

A total of 50 Y-chromosome biallelic markers were used to cover all 18 major haplogroups (A-R) recognized by the Y-chromosome consortium (YCC) (22,23). Table 1 contains a summary of the Y-SNP markers used in this study according to their physical position along the Y-chromosome. These positions were identified using amplicon sequences for the 50 Y-SNPs with the BLAST-Like Alignment Tool (BLAT): http://genome.ucsc.edu/cgi-bin/hgBlat. Of the 50 Y-SNPs, 18 were typed by three hexaplex allele specific primer extension assays. An additional set of 42 Y-SNP loci was typed by an ASH assay described below. Ten of the Y-SNP loci probed were redundant between the two sets.

TABLE 1—*Summary information about the 50 Y-chromosome SNP markers used in this study. The BLAT position designates the exact position along the reference Y-chromosome defined by the Human Genome Project (April 2003). The Y Chromosome Consortium (YCC) haplogroup (Hg) defined by the various Y-SNPs are listed along with the multiplex used to type the samples. The Hg listed for each marker is the Hg defined by typing that Y-SNP alone. Multiplexes 1–5 come from the Marligen Y-SNP ASH kit while multiplexes A, B and C are performed using ASPE assays developed in-house. The polymorphisms are listed with the ancestral allele first followed by the derived allele. The two right-hand columns tabulate the ratio of ancestral:derived allele frequencies for a particular Y-SNP in the 115 African American and 114 Caucasian samples. The sequence immediately surrounding Y-SNP P25 occurs an additional two times on the Y-chromosome at the positions indicated (P25b and P25c).*

Y Position (Mb)	SNP Name	YCC Hg Defined	Multiplex	Polymorphism	African American ($N = 115$)	Caucasian ($N = 114$)
2,562,931	SRY+465	O2b	4	C->T	1.00/0.00	1.00/0.00
2,564,927	SRY$_{10831a,b}$	B-R, R1a	5	A->G, G->A	0.01/0.99, 1.00/0.00	0.00/1.00, 0.95/0.05
2,566,620	SRY$_{9138}$	K1	4	C->T	1.00/0.00	1.00/0.00
2,642,605	M130 (RPS4Y)	C	3	C->T	1.00/0.00	1.00/0.00
13,407,330	M2	E3a	2	A->G	0.42/0.58	1.00/0.00
13,413,670	DYS391	E3	2	C->G	0.40/0.60	0.96/0.04
14,124,138	M168	C-R	1	C->T	0.03/0.97	0.00/1.00
14,157,939	M170	I	3, A	A->C	0.96/0.04	0.79/0.21
14,179,223	M182	B2	2	C->T	0.99/0.01	1.00/0.00
14,232,730	Tat	N3	4	T->C	1.00/0.00	0.99/0.01
14,264,427	M174	D	3, A	T->C	1.00/0.00	1.00/0.00
14,279,781	M172	J2	3, A	T->G	1.00/0.00	0.95/0.05
14,337,676	M201	G	3	G->T	0.99/0.01	0.97/0.03
14,340,899	M198	R1a1	C	C->T	1.00/0.00	0.95/0.05
14,818,875	M175	O	1	−5bp (+/−)	1.00/0.00	1.00/0.00
14,892,152	M207	R	1	A->G	0.73/0.27	0.35/0.65
18,343,057	M3	Q3	4, B	C->T	1.00/0.00	1.00/0.00
20,806,634	M5	M	4	C->T	1.00/0.00	1.00/0.00
20,807,275	P3	A2	2	G->A	1.00/0.00	1.00/0.00
20,903,048	M153	R1b6	5	T->A	1.00/0.00	0.99/0.01
20,926,945	M9	K-R	B	C->G	0.72/0.28	0.33/0.67
20,927,335	M11	L	3, B	A->G	1.00/0.00	1.00/0.00
20,929,850	M18	R1b1	5	+2bp (−/+)	1.00/0.00	1.00/0.00
20,936,442	M31	A1	2	G->C	0.99/0.01	1.00/0.00
20,937,124	M32	A3	2	T->C	1.00/0.00	1.00/0.00
20,937,138	M33	E1	2, B	A->C	0.97/0.03	1.00/0.00
20,938,391	M35	E3b	2, B	G->C	0.98/0.02	0.97/0.03
20,948,239	M37	R1b2	5	C->T	1.00/0.00	1.00/0.00
20,949,887	M52	H	3	A->C	1.00/0.00	1.00/0.00
20,959,373	M119	O1	4, A	A->C	1.00/0.00	1.00/0.00
20,961,146	M137	J2c	C	T->C	1.00/0.00	1.00/0.00
20,961,189	M124	P1	4, C	C->T	1.00/0.00	1.00/0.00
20,961,274	M123	E3b3	C	G->A	1.00/0.00	1.00/0.00
20,961,362	M122	O3	C	T->C	1.00/0.00	1.00/0.00
20,961,382	M166	J2f2	C	G->A	1.00/0.00	1.00/0.00
20,964,647	M112	B2b	A	G->A	1.00/0.00	1.00/0.00
20,975,939	M146	B1	2	A->C	1.00/0.00	1.00/0.00
21,063,528	M42	B-R	1	A->T	0.01/0.99	0.00/1.00
21,064,475	M45	P-R	1	G->A	0.73/0.27	0.35/0.65
21,064,542	M157	R1a1b	5	A->C	1.00/0.00	1.00/0.00
21,066,207	M150	B2a	2	C->T	0.99/0.01	1.00/0.00
21,074,761	M60	B	1	+1bp (−/+)	0.99/0.01	1.00/0.00
21,086,865	M75	E2	2, A	G->A	0.97/0.03	1.00/0.00
21,090,746	M69	H	B	T->C	1.00/0.00	1.00/0.00
21,102,797	M87	R1a1c	5	T->C	1.00/0.00	1.00/0.00
21,114,001	M89	F-R	1	C->T	0.68/0.32	0.04/0.96
21,134,846	M94	B-R	1	C->A	0.01/0.99	0.00/1.00
21,135,132	M95	O2a	4	C->T	1.00/0.00	1.00/0.00
22,653,504	P4	A2	2	C->T	1.00/0.00	1.00/0.00
24,142,855	**P25a**	R1b	5	C->A	0.77/0.23	0.48/0.52
25,868,376	**P25b**	...	...	...	...	...
26,600,000	**P25c**	...	...	...	...	...

TABLE 2—*Primers for three multiplex PCR assays. The orientation of each primer relative to the sequence contained in GenBank is provided. Amplicon size was determined using the given GenBank sequence. *Note that Y-SNP markers M123, M124, M137, and M166 are located in the M122 amplicon. The reverse complement (r&c) of the GenBank reference sequence was used for primer design in some cases to maintain traditional nomenclature for these Y-SNP markers.*

	Markers	GenBank Accession	Forward PCR Primer	Reverse PCR Primer	Amplicon Size (bp)
A	M174	AC002531	TTCTCCGTCACAGCAAAAATG	ATGCAAAAGGAGAAGGACAAGA	186
	M172	AC002531	CACCTGCCTCTCAGTATCAACAG	GCCAGGTACAGAGAAAGTTTGG	138
	M112	AC010889 (r&c)	AAAAGCAAAAGAGAACTGCCTC	TTCAATTCTTGTCTGTTGCAGAA	160
	M75	AC010889	GATTCAGAATTGACTTGTCAAAAGC	GCGTGAATCTCTGCCCAG	178
	M119	AC010889(r&c)	CCGCAGTGCTATGTGTTTATTTG	GAGTATTCTTAATTAATACAAAATTTAAACCG	151
	M170	AC002531	TGTTTGTTCAAATAATTGCAGCTC	AAGACCACACAAAAACAGGTCC	155
B	M3	AC007034	TTGTGAATCTGAAATTTAAGGGC	AATGTGGCCAAGTTTTATCTGC	131
	M35	AC003032(r&c)	CCTGATAATTACTCAGGAGGAAACTG	CACTAAGGAGCTGTGGTGAATG	155
	M9	AC010137	TAGAAAAGTTTGCGCAAGGAAT	TGAAATGCATAATGAAGTAAGCG	192
	M69	AC010889	CGGACTTGAAGGAATCAGCC	GAAAAATTTATCTCCCCTTAGCTCTC	177
	M11	AC003032	CTGAATCCCCCAGTAAGTTCAG	GTGCATGAATATGTTACTGCTGAG	134
	M33	AC003032(r&c)	CACAACTTCATTGGCTACGG	CTTTGCTTATTTGTTGAAGCCC	202
C	M122	AC010889(r&c)	TGTGATCAACTTCTTTCCCTCA	TGCAAATGGTATGCAACTCAG	410
	M123*				N/A
	M124*				N/A
	M137*				N/A
	M166*				N/A
	M198	AC004474	CAGTCATGATGAGGTGGAATGT	CCGTTTGCCTAGGTTAGAATAAAC	127

TABLE 3—*Extension primers for three multiplex ASPE assays. The final strand concentration of each primer is given. The orientation of each extension primer relative to the YCC paper is provided. (M = moles/liter).*

	Markers	Conc. (μM)	Extension Primer with Poly T Tail for Spacing	Length	Strand	SNP (Ancestral/Derived)
A	M174	1.5	(T)$_8$GAATACCTTCTGGAGTGCCC	28	F	T to C
	M172	1.0	(T)$_{11}$CCCAAACCCATTTTGATGCTT	32	F	T to G
	M112	1.5	(T)$_{14}$GAGGTGAGATAAAAACAAAGCAGT	38	F	G to A
	M75	1.0	(T)$_{19}$GAAAAGACAATTATCAAACCACATCC	45	F	G to A
	M119	1.0	(T)$_{18}$AGATAATTCTGCTCTTTAAAATTGCTATTTT	49	F	A to C
	M170	1.5	(T)$_{19}$CAAATTACTATTTTATTTACTTAAAAATCATTGTTC	55	F	A to C
B	M3	0.3	(T)$_6$GGTACCAGCTCTTCCTAATT	26	R	C to T
	M35	1.2	(T)$_9$CCCAATTTTCCTTTGGGACACTG	32	F	G to C
	M9	1.0	(T)$_{18}$ACGGCCTAAGATGGTTGAAT	38	F	C to G
	M69	1.0	(T)$_{19}$CCCTTTGTCTTGCTGAAATATATTT	45	R	T to C
	M11	1.0	(T)$_{16}$TTCTTTCCCAATTATGCAAGTTCTAAACCTTTC	49	R	A to G
	M33	1.0	(T)$_{20}$AGATCTGTTCAGTTTATCTCATAAGTTACTAGTTA	55	F	A to C
C	M122	1.0	(T)$_5$TTCAGATTTTCCCCTGAGAGC	26	R	T to C
	M123	2.0	(T)$_9$CATTTCTAGGTATTCAGGCGATG	32	R	G to A
	M124	1.0	(T)$_{12}$TTGCACAAACTCAGTATTATTAAACC	38	R	C to T
	M137	1.0	(T)$_{18}$TTTATCATATTGAGATTTTGCTTTCCT	45	R	T to C
	M166	1.0	(T)$_{22}$TGGAAATGAATAAATCAAGGTAGAAAA	49	F	G to A
	M198	1.0	(T)$_{21}$TGTATCAGTATACCAATTAATATTTTTGAAAGAG	55	F	C to T

ASH Assay Using Flow Cytometry

Y-SNP typing was accomplished with ASH using the commercially available Signet™ Y-SNP Identification System (Marligen Biosciences, Inc., Ijamsville, MD). This kit enables analysis of 42 Y-SNPs using xMAP™ suspension array technology on the Luminex® 100 flow cytometer (MiraiBio, Alameda, CA). The 42 markers were amplified and analyzed in five different multiplex reactions: Multiplex 1 (M42, M45, M60, M89, M94, M168, M175, and M207), Multiplex 2 (M2, M31, M32, M33, M35, DYS391, M75, M146, M150, M182, P3, and P4), Multiplex 3 (M11, M52, M130, M170, M172, M174, M201), Multiplex 4 (M3, M5, M95, M119, M124, SRY+465, SRY9138, and Tat), and Multiplex 5 (M18, M37, M87, M153, M157, P25, and SRY 10831). In addition, an amelogenin-based SNP was included in Multiplex 1 to determine sex of the tested sample as well as a positive control for the presence of either male or female template DNA. Testing was performed according to the kit protocols using 2 ng of genomic DNA for each multiplex reaction. Note: the sequences of the PCR and hybridization primers are proprietary to Marligen Biosciences.

Primer Design for ASPE

The overall strategy for multiplex primer design was used as reported by Schoske et al. (24). All primers (in the primer extension assays) were designed for compatibility (lack of primer-dimer formation) although various subsets were employed for detection purposes. PCR primers used in the primer extension assays are listed in Table 2 along with the GenBank® source of the reference sequences they were designed against.

The extension primers, designed in three multiplex sets labeled A, B, and C for 18 of the 50 Y-SNPs, are listed in Table 3. The remaining

32 Y-SNPs were examined with the ASH approach and ten of the loci were typed with both ASH and ASPE methods. Extension primers were selected to have a predicted annealing temperature of approximately 60°C and were screened for hairpin and primer-dimer interactions using an in-house primer-checking program (24).

PCR Amplification

Desalted oligonucleotides were purchased from Qiagen Operon (Alameda, CA) and primer integrity was confirmed using time-of-flight mass spectrometry (25). Primers were combined in equimolar ratios at a concentration of 0.5 μmole/L and tested with 2 ng DNA template.

Amplifications were performed in reaction volumes of either 7 or 15 μL using a master mix containing 1X GeneAmp® PCR Gold buffer (Applied Biosystems, Foster City, CA), 5 mmol/L MgCl₂, 250 μmol/L deoxynucleotide triphosphates (dNTPs; Promega Corporation, Madison, WI), 0.16 mg/mL bovine serum albumin (BSA) fraction V (Sigma, St. Louis, MO), and 1 unit of AmpliTaq Gold® DNA polymerase (Applied Biosystems). The thermal cycling program was carried out on a GeneAmp 9700 (Applied Biosystems) using the following conditions in 9600-emulation mode (i.e., ramp speeds of 1°C/s) (26):

95°C for 10 min;
3 cycles of {95°C for 30 s, 50°C for 55 s, 72°C for 30 s};
19 cycles of {95°C for 30 s, 50°C for 55 s +0.2°C per cycle, 72°C for 30 s};
11 cycles of {95°C for 30 s, 55°C for 55 s, 72°C for 30 s};
72°C for 7 min;
4°C until removed from thermocycler.

Following PCR amplification, unincorporated primers and dNTPs were removed by adding 3 μL of ExoSAP-IT® (USB Corp., Cleveland, OH) to each 7 μL PCR reaction (or 6 μL of ExoSAP-IT® to a 15 μL PCR reaction). Reactions were mixed briefly and incubated at 37°C for 90 min and then 80°C for 20 min to inactivate the enzymes.

ASPE Assay Using Fluorescence Detection

Multiplex primer extension reactions were carried out in a total volume of 10 μL using 2.5 μL of ABI Prism® SNaPshot™ multiplex kit mix (Applied Biosystems), 0.5 μL of 10X AmpliTaq Gold® PCR buffer, 3 μL of PCR template, 3 μL of water, and 1 μL of a stock solution of extension primers to produce the final primer concentrations indicated in Table 3. Extension reactions were incubated as follows: 25 cycles of 96°C for 10 s, 50°C for 5 s, and 60°C for 30 s. Excess fluorescently labeled ddNTPs were inactivated by addition of 1 unit of shrimp alkaline phosphatase (SAP). Reactions were mixed briefly and incubated at 37°C for 30 min then 80°C for 20 min.

A 0.9 μL aliquot of each SAP-treated primer extension product was diluted in 14 μL Hi-Di™ formamide and 0.4 μL GS120-LIZ internal size standard (Applied Biosystems) and analyzed on the 16-capillary ABI Prism® 3100 Genetic Analyzer (Applied Biosystems) using filter set E5 without prior denaturation of samples. Samples were injected electrokinetically for 13 s at 1 kV. Separations were performed in approximately 30 min on a 36 cm array using POP™-6 (Applied Biosystems). Automated allele calls were made in Genotyper® 3.7 using an in-house macro based on fragment size and dye color.

Results

The 50 markers examined here define 45 of the 153 YCC haplogroups described in a recent publication (22). Figure 1 illustrates the position of each Y-SNP on the YCC tree and highlights markers that are common between the two assay formats (boxed).

Haplogroup Frequencies for African American and Caucasian Samples

We examined 114 U.S. Caucasians and 115 African American samples across these 50 Y-SNPs. A total of 18 different haplogroups were observed out of a possible 45 defined by these 50 Y-SNPs (Fig. 1). A majority of the African American samples (58%) fell in the E3a haplogroup as they were derived at M2 (sometimes designated DYS271 or sY81) (13). The derived allele was observed in a majority of the Caucasian samples for Y-SNP markers M207 and P25 and thus can be placed in the R haplogroup. A number of African Americans were also in haplogroup R(xR1a,R1b) and R1b, demonstrating that admixture may be present at a level of ∼30% as recently reported using Y-STR data (27). YCC haplogroups unique to African American samples in this dataset included A1, B*, B2a, E1, E2, and E3a. Haplogroups unique to the Caucasian samples were E*, E3*, J2, K*, N3, R1a1, and R1b6. Haplogroups shared between these two sample sets included E3b, G, I, R(xR1a,R1b), and R1b (Fig. 1).

Concordance in Allele Calls Between ASPE and ASH

The ten markers in common between the ASPE and ASH methods used here are shown boxed in Fig. 1. They include M174, M33, M75, M35, M170, M172, M11, M119, M124, and M3. We observed full concordance between allele calls made with the primer extension and hybridization methods across almost 2300 allele calls (10 loci × 229 samples). Both methods had an allele calling success rate of over 99%. With 190 samples for the five combined multiplex ASH assays, 8109 allele calls resulted out of a possible 8170 for a success rate of 99.3%. For the three ASPE assays (Fig. 2), 3389 calls were obtained from a possible 3420 allele calls for a success rate of 99.1%. Typically, when a sample failed across all markers in a multiplex, it was due to PCR failure (data not shown). Occasionally 1–3 markers would drop out of an ASPE hexaplex. This could possibly be attributed to polymorphisms in the primer-binding region specific to that method. However, no sequencing experiments have been performed yet to confirm this hypothesis.

Allele Ratios with ASH Assays

With the ASH method, allele calls are made based on the highest signal observed when both alleles are examined for a SNP marker. Single-copy Y-SNPs are expected to be homozygous in all single-source samples because of the haploid nature of the Y-chromosome. Characterizing the amount of signal observed from the alternate allele, which should be absent in a haploid system in single-source samples, is valuable to see how well a system would be able to handle potential mixtures of two male DNA samples (see Table 4).

On average the signal from the alternate allele comprised 18% (±14%) of the called "true" allele signal intensity. We observed four instances of "heterozygous" signals from both possible alleles
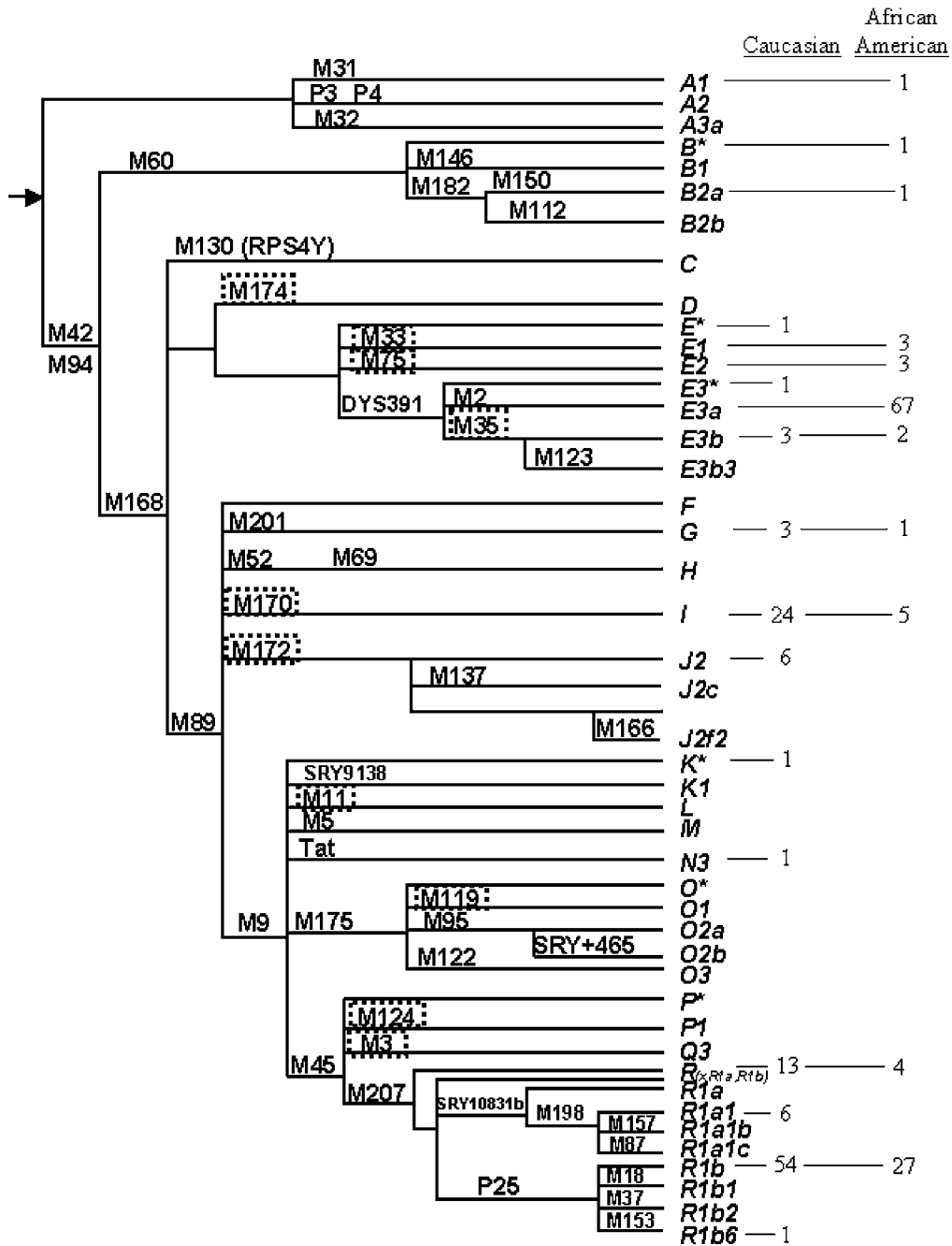
Caucasian | African American

M31
P3   P4
M32
M60
M146
M182   M150
M112
M130 (RPS4Y)
M174
M33
M75
DYS391   M2
M35
M123
M42
M94
M168
M201
M52   M69
M170
M172
M137
M166
M89
SRY9138
M11
M5
Tat
M119
M9   M175   M95
M122   SRY+465
M124
M3
M45
M207
SRY10831b   M198   M157
M87
P25   M18
M37
M153

A1 ——————— 1
A2
A3a
B* ——————— 1
B1
B2a ——————— 1
B2b
C
D
E* —— 1
E1 ———— 3
E2 ———— 3
E3* —— 1
E3a ——————— 67
E3b —— 3 ——— 2
E3b3
F
G —— 3 ——— 1
H
I —— 24 ——— 5
J2 —— 6
J2c
J2f2
K* —— 1
K1
L
M
N3 —— 1
O*
O1
O2a
O2b
O3
P*
P1
Q3
R(xR1a,R1b) — 13 ——— 4
R1a
R1a1 —— 6
R1a1b
R1a1c
R1b —— 54 ——— 27
R1b1
R1b2
R1b6 —— 1

FIG. 1—*Forty-five YCC haplogroups defined by 50 Y-SNPs were examined in this study. Haplogroup designations to the right represent the groups defined by typing all 50 Y-SNPs. Boxed loci indicate Y-SNPs that were typed by both ASPE and ASH methodologies. Only 18 different haplogroups were observed across the 229 males tested from U.S. Caucasian (N = 114) and African American (N = 115) samples. The haplogroup distribution for Caucasian and African American samples is denoted to the right of the tree. Note: The haplogroup R(xR1a,R1b) designation was used in place of R* (this was due to the absence of Y-SNP M173 typing results).*
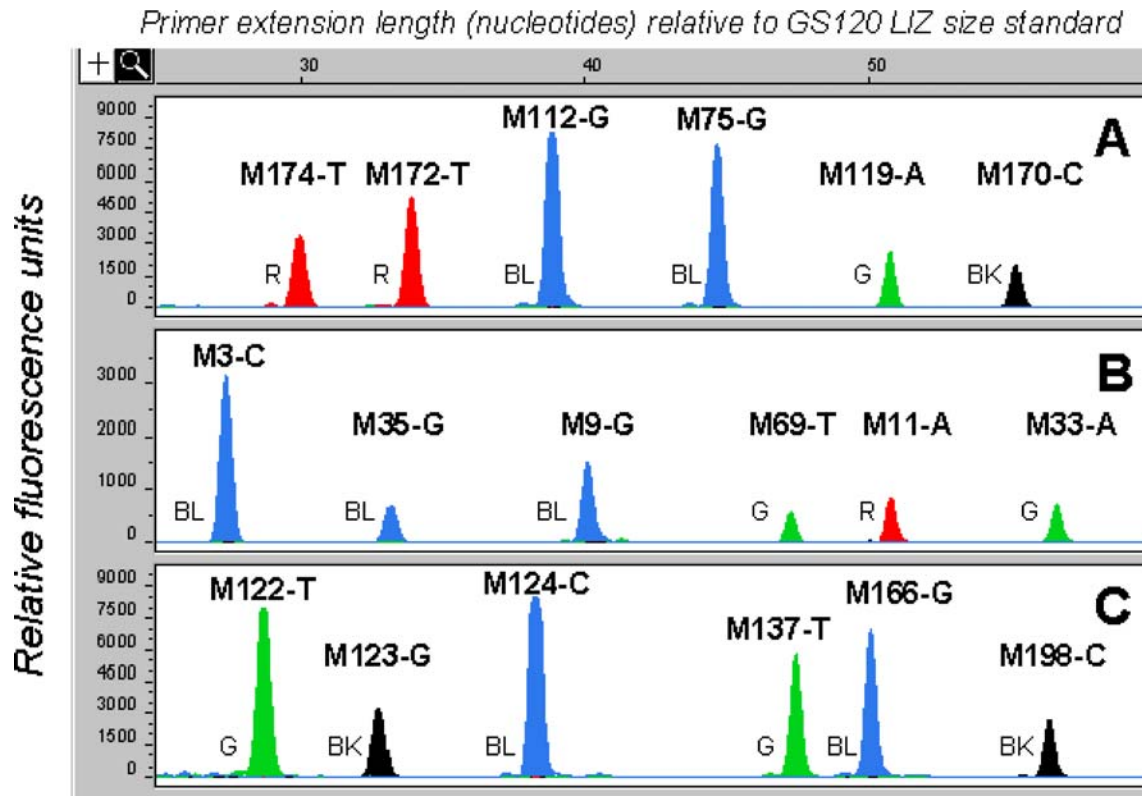
FIG. 2—*Results from three primer extension assays on ABI 3100 using ASPE assays. Typical electropherograms for the three 6-plex Y-SNP assays are shown. The PCR and ASPE primers sequences for multiplexes A, B, and C are contained in Tables 2 and 3. The color of each peak indicates the identity of the enzymatically incorporated ddNTP in the primer extension reaction. Each ddNTP is labeled with a unique fluorescent dye (R = ddTTP, BL = ddGTP, G = ddATP and BK = ddCTP). The Y-SNP locus corresponding to each peak, allele call, and peak color are shown. The length of each extension primer, in nucleotide units, was estimated by comparison to the GS120 LIZ size standard (data not shown).*

with the ASH methodology (out of 9600 possible allele calls) for single-copy Y-SNPs. At locus M168 both the C and T alleles were observed with almost equal intensities (C:T = 0.88, 0.94, 0.98, data not shown) in three African American samples that fell in YCC haplogroups A and B (ancestral state for M168). In addition, the M31 marker exhibited both G and C alleles (G:C = 0.82, data not shown) at the one African American sample that was in haplogroup A1 (derived state for M31). In all other cases with the exception of the multi-copy Y-SNP P25 (22), there was a dominant single allele observed with the ASH approach. Two female samples were tested and yielded a negligible signal above background with 2 ng of DNA template. No attempts were made in this study to examine higher amounts of female DNA. Y-SNPs M168 and M31 were typed using only the ASH approach. The sequences of the PCR primers and ASH probes are currently unknown (proprietary to Marligen). This makes it difficult to determine the exact nature of "heterozygous" signal from M168 and M31. It may be possible that the sequence surrounding these SNP sites has been duplicated somewhere else on the Y-chromosome, but BLAST queries of the most recent human genome reference sequence have not indicated such a problem.

Analysis of X and Y allele ratios for amelogenin across the 229 male samples revealed an average ratio of 0.99 with a standard deviation of 0.06. Only six males fell outside the allele signal ratio of 0.80–1.20. One of these samples exhibited loss of the X amelogenin allele when typed by Identifiler (21). The loss of the amelogenin X allele with PCR primers used for sex typing in Applied Biosystem's kits has been reported previously (28). Shewale et al. observed three out of 7220 males tested to have this loss of the amelogenin X al-

lele, which was recovered when these samples were amplified with a different amelogenin primer set. Although we do not know where the Marligen amelogenin PCR and SNP ASH probes are located, it is possible that some of these more variable X/Y ratio samples have one or more polymorphisms in the template DNA that impact hybridization of the PCR primers or the hybridization probe. By contrast, the X-to-Y allele ratio for the two females examined in our study using the Signet kit was 13.8 and 19.4. Thus, the amelogenin X and Y alleles, which should be approximately equal in amount in males, do in fact show up that way in our data. The histograms plotted in Fig. 3 indicate the average signal intensity for each locus when typed by the ASH assay (for samples residing the R1b haplogroup). As can be seen in Fig. 3, there is a sizable background in many cases for the alternate Y-SNP allele. This is likely due to the difficulty of distinguishing single-base mismatches by hybridization alone under uniform temperature conditions across all SNP alleles.

Multi-Copy Y-SNP P25

The ability to accurately type the Y-SNP marker P25 is important due to the fact that it defines the most common Y haplogroup, R1b, observed within our sample groups. P25 is a multi-copy locus that is present three times within a highly palindromic region (29) of the Y-chromosome at 24.1, 25.9, and 26.6 megabases (Table 1). When comparing the sequences of the three GenBank entries (AC016911, AC016698, and AF257064) the allele identity of P25 SNP position for all three is "C." This is consistent with ASH and

TABLE 4—*Average signal-to-noise ratios (S/N) observed for Y-SNP ASH experiments. "Signal" is considered the fluorescence intensity of the called allele while "noise" is the fluorescence from the alternate (absent) allele. Only samples with a minimum of 3000 bead counts were included in these calculations. The standard deviation for each allele measurement varies for each locus depending on the number of alleles called (N). Dotted regions indicate alleles that were not observed. The ancestral and derived alleles are designated as A1 and A2, respectively.*

Y-SNP Locus	A1 S/N Avg		N	A2 S/N Avg		N
M175(−/ +)	. . .	. . .	0	25.3	± 5.4	179
M207(A/G)	12.2	± 2.7	97	12.9	± 2.5	82
M42(A/T)	2.2	± 0.0	1	3.6	± 0.4	178
M45(A/G)	8.1	± 1.0	82	4.9	± 0.7	97
M60(−/ +)	18.6	± 4.2	177	3.7	± 0.3	2
M89(C/T)	4.8	± 0.8	65	17.4	± 3.8	114
M94(A/C)	14.3	± 3.1	178	7.7	± 0.0	1
DYS391(C/G)	6.0	± 0.7	130	4.8	± 0.9	53
M146(A/C)	3.7	± 0.7	183	. . .	. . .	0
M150(C/T)	4.0	± 0.6	182	40.5	± 0.0	1
M182(C/T)	11.6	± 5.8	182	11.3	± 0.0	1
M2(A/G)	1.6	± 0.1	136	5.5	± 0.6	47
M31(C/G)	1.6	± 0.0	1	14.7	± 17.4	182
M32(T/A)	. . .	. . .	0	27.0	± 12.4	183
M33(A/C)	27.1	± 23.3	180	11.7	± 2.5	3
M35(C/G)	2.5	± 0.6	6	4.6	± 6.5	177
M75(A/G)	2.6	± 0.1	3	7.1	± 1.3	180
P3(C/T)	3.3	± 0.8	183	. . .	. . .	0
P4(A/G)	. . .	. . .	0	8.2	± 0.9	183
M11(A/G)	2.7	± 0.3	187	. . .	. . .	0
M130(C/T)	13.5	± 4.5	187			0
M170(A/C)	26.5	± 6.1	24	6.2	± 1.4	163
M172(G/T)	3.7	± 0.6	6	3.0	± 0.5	181
M174(C/T)	. . .	. . .	0	7.2	± 6.2	187
M201(G/T)	17.0	± 4.0	184	2.7	± 0.1	3
M52(A/C)	5.5	± 2.4	187	. . .	. . .	0
M119(A/C)	3.4	± 0.7	135	. . .	. . .	0
M124(C/T)	4.5	± 0.4	135	. . .	. . .	0
M3(C/T)	13.9	± 2.6	135	. . .	. . .	0
M5(C/T)	3.5	± 0.4	135	. . .	. . .	0
M95(C/T)	7.0	± 0.8	135	. . .	. . .	0
SRY465(C/T)	4.3	± 2.2	135	. . .	. . .	0
SRY9138(C/T)	4.2	± 0.9	135	. . .	. . .	0
Tat(C/T)	43.4	± 14.6	135	. . .	. . .	0
M153(A/T)	. . .	. . .	0	2.6	± 0.4	116
M157(A/C)	8.8	± 1.1	116	. . .	. . .	0
M18(−/ +)	19.7	± 5.0	116	. . .	. . .	0
M37(C/T)	3.6	± 0.7	116	. . .	. . .	0
M87(C/T)	. . .	. . .	0	27.4	± 5.4	116
SRY10831(A/G)	8.3	± 0.6	4	63.1	± 32.4	112
M168(C/T)	na	± na	3	17.6	± 3.3	176
P25(A/C)	1.7	± 0.1	61	3.2	± 0.2	55

ASPE typing results that exhibit "background" signal for the "C" allele. The multi-copy nature of P25 presents problems for simple interpretation of ASH data because the signal for the "A" allele is never greater than that of the "C" allele. This is illustrated in a histogram plot shown in panel A of Fig. 4. Samples were therefore typed by calculating the ratio between the A and C allele signal intensity. The peak ratios fell into two categories of 1.7 (±0.1) and 3.2 (±0.2) with well-defined signal variation. An "A" call is made when the ratio of C-to-A signal is 1.7 (±0.1) while a "C" allele call is assigned to a 3.2 (±0.2) ratio for C-to-A signal. It is therefore possible to call single-source samples accurately by examining the C-to-A allele signal ratio. Panel B of Fig. 4 is an example of ASPE data for the P25 locus. A "C" allele is represented by a single peak (black) while the "A" allele is called when both A (green) and C (black) peaks are observed.

Discussion

Experimental Aspects of ASH and ASPE Assays

The primer extension technique described here provided a cleaner background than ASH. In spite of high background with the alternate alleles in ASH, allele calls were fully concordant in overlapping markers between ASPE and ASH SNP typing technologies with single-source DNA samples. The three hexaplex ASPE assays illustrated in Fig. 2 did not exhibit a background signal (i.e., signal from the alternate allele) when evaluating over 4100 allele calls. In forensic situations where the potential of mixed-source samples exists, a technique with little to no background for an absent allele is advantageous. The use of single-copy Y-SNPs is also desirable over multi-copy ones such as P25 where result interpretation is more complicated.

The ability to detect multiple SNPs in a single reaction with the ASPE format is accomplished by employing tailed extension primers. Extension primers for this purpose typically range in size up to 60 bases. Higher levels of multiplexing require longer tailed primers, which usually necessitate high-performance liquid chromatography (HPLC) purification (30). Purified primers add additional cost to the assay, especially when the initial goal may be to only screen/evaluate for useful markers.

An important characteristic of a methodology in forensic science can be the ability to reanalyze a sample if it fails the first time through a procedure (due to an interruption in electrical service, a computer malfunction, etc.). In the case of a poor capillary injection, a sample can be easily reinjected onto the ABI 3100 for reanalysis. However, resampling by the Luminex/flow cytometry instrumentation requires a repeat of the PCR labeling step and bead hybridization because the original sample is consumed and diluted during the analysis. Thus, reanalysis is not as easily performed with the flow cytometry technique.

Multiplex ASPE Development

Several other groups have used ASPE for Y-SNP analysis. Inagaki and co-workers (18) examined 15 Y-SNPs in two SNaPshot multiplexes and observed 13 different haplogroups in 159 Japanese males tested. Kayser and co-workers (10) also used ASPE assays to examine M95, M104, M173, M210, and M217 as part of a study of New Guinea populations. In this study, the rationale for developing multiple ASPE 6-plex assays versus a single 18-plex was to decrease overall assay development time. Typically, increasing the number of loci probed results simultaneously in an increase in multiplex assay development time. Issues pertaining to PCR primer compatibility and nonspecific template binding can complicate development of robust multiplex assays. The PCR and extension primers for the three 6-plex assays shown here provided successful typing results in a single round of design using our strategy described previously (24,31). The goal of typing the 50 Y-SNP loci over a cohort of U.S. samples was to determine their diversity. The typing results indicated that 26 of the 50 Y-SNP loci did not exhibit any variation (see Table 1). Designing medium-size (e.g., 6-plex) PCR and primer extension assays allowed rapid evaluation of candidate Y-SNP loci without investing great effort in developing a larger multiplex that might contain noninformative loci in the sample set of interest. Once the most informative Y-SNP loci are identified, efforts will be directed towards developing a higher multiplexed primer extension assay, similar to the 35-plex demonstrated by Sanchez et al. (30).
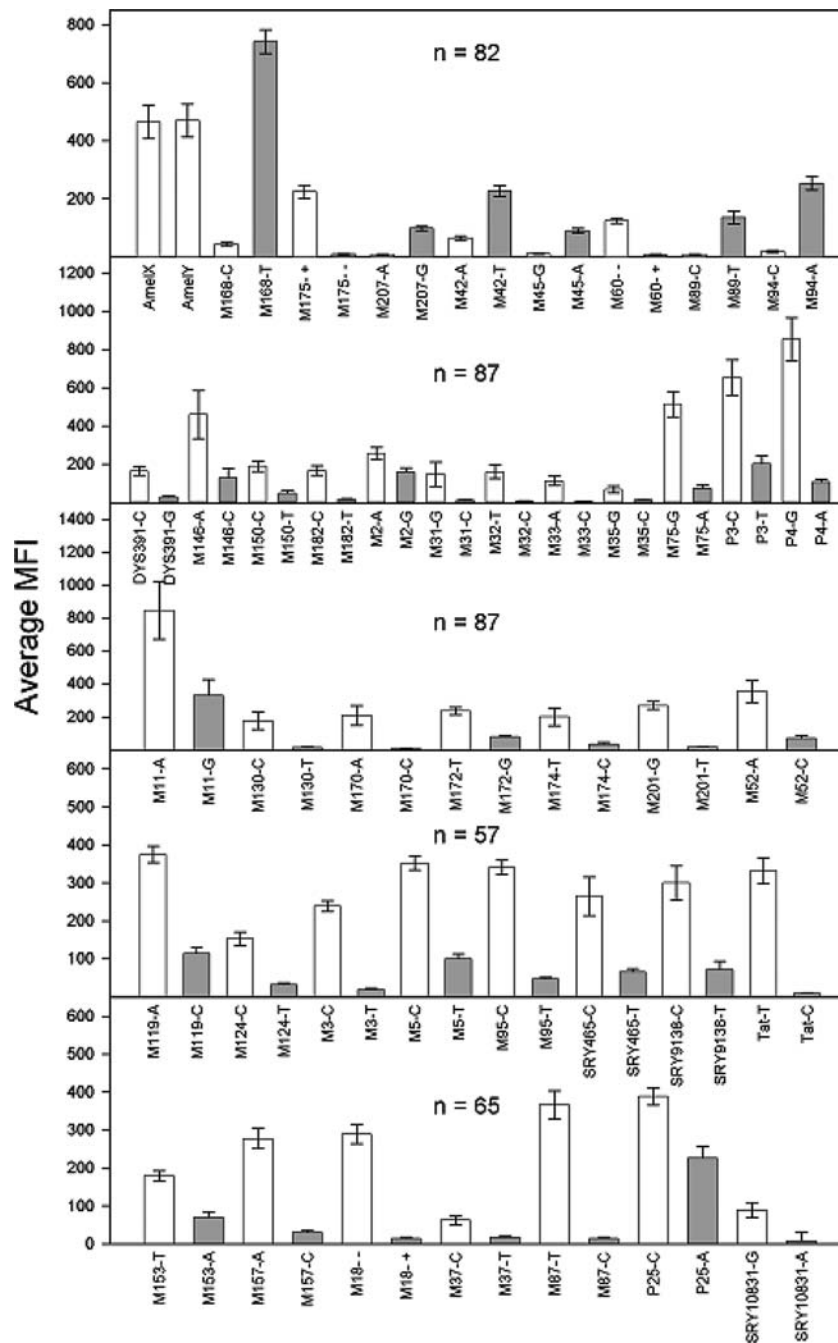
FIG. 3—*Average signal intensity for 42 Y-SNPs and Amelogenin for samples observed in haplogroup R1b typed by ASH methodology. The mean fluorescence intensity (MFI) is plotted for each marker. The MFI for each marker was averaged for samples with a minimum bead count of 3000 events for each multiplex. Error bars indicate the standard deviation for each set of* n *results, which varies for each multiplex.*

Sample Consumption

A hierarchic genotyping approach in which only a subset of Y-SNP markers is examined once the branch of the tree has been determined (17) is usually necessitated by limited amounts of DNA available on many of the population samples examined (5). However, with multiplex assays such as described here, rapid analysis of large numbers of markers is possible while consuming only a small amount of genomic DNA. We consumed only 16 ng of each DNA sample in obtaining the results from the 50 Y-SNPs. With further

optimization of reaction conditions, consumption of less than 5 or 10 ng of DNA should be easily achievable as has been demonstrated by Sanchez et al. (30). In addition, if an optimized set of markers with fewer loci were assayed, the correct Y-SNP haplogroup could be ascertained while consuming less than a nanogram of template DNA. Typically 10–20 ng (8), 30 ng (7), and even 50 ng (5) are needed for each marker when run with singleplex analyses. Even multiplex assays that do not employ fluorescent detection methods required as much as 60 ng DNA for each reaction (17). Thus, the assays reported in this study represent an improvement in reducing the
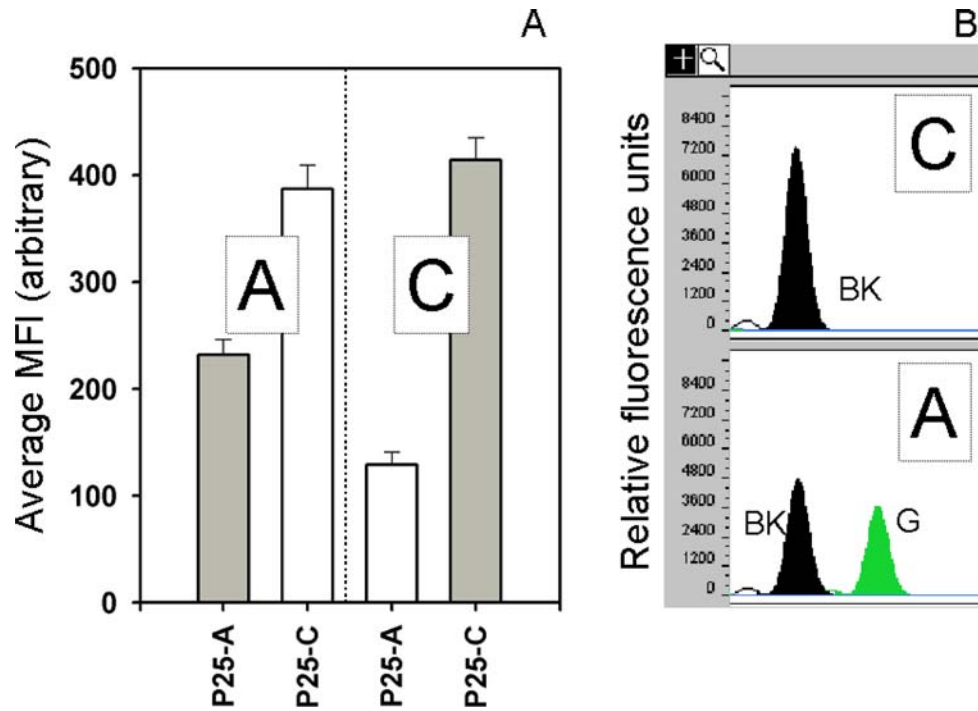
FIG. 4—*Typing the Y-SNP locus P25 with ASH (panel A) and ASPE (panel B) methodologies. Panel A shows the average MFI (mean fluorescence intensity) for A and C allele calls from ASH experiments. Note that the signal intensity for C is always greater than that of A. The shaded bars and labels indicate the proper allele call. Panel B illustrates ASPE data for typing the P25 locus. A single black peak is observed for a C allele call (top of panel B) while two peaks, black and green, are observed for an A allele call (bottom of panel B).*

consumption of DNA template while providing a rapid examination of a number of Y-SNP markers covering the major haplogroups in the YCC tree (22).

Reliability of Marker Positions of YCC Phylogenetic Tree

In all cases with our 229 samples tested, the YCC tree marker positions were robust across the 50 Y-SNPs we examined. If a derived allele was observed at a particular Y-SNP that took one down a branch of the YCC tree, no other derived alleles were observed outside of that branch of the tree. The redundancy of many of these Y-SNPs in terms of tree position permits a degree of internal quality control. For example, a sample should not possess the derived allele for both M9 and M170 since M170 defines YCC haplogroup I and M9 defines non-overlapping haplogroups K through R.

Y-SNPs Specific to a U.S. Sample Group

While there is a degree of admixture between U.S. African American and Caucasian samples that has been noted with Y-STR data (27), Y-SNPs have been proposed as genetic markers for inferring the population-of-origin for a crime scene stain should that ability be desired in the future (2). For our purposes a good candidate marker for inferring population-of-origin should be specific to a group and should exist in most, if not all, individuals associated with the group. However, our data suggest that currently available Y-SNPs will have little value for reliably inferring populations-of-origin from U.S. male DNA samples. For example, in our dataset the derived allele for M2 (defining haplogroup E3a) was exclusive to African Americans. However, only ~58% of the African American males examined were derived at M2 with 23% of the total

samples falling into a common Caucasian haplogroup (R1b). Other than M2, no other Y-SNP markers were uniquely observed to an extent of greater than 6% in either our Caucasian or African American samples.

Towards Optimal Marker Sets

Most studies conducted to date with Y-SNPs have been from diverse populations with sample sizes of less than 100 samples per population. For example, although Underhill and co-workers (5) examined 1062 males, only two of the 21 different populations analyzed contained more than 100 individuals. In another survey of 1935 men with 23 Y-SNPs, 49 different Eurasian populations were studied (7). The purpose of these studies is typically to examine similarities or differences between populations rather than to evaluate the ability of Y-SNP markers to distinguish between individuals in the same population. Without typing an appropriate number of samples using additional Y-SNP markers, it may be difficult to assess which markers will be useful in forensically relevant datasets if these markers are going to be used for human identity applications (2).

The major haplogroups seen in our U.S. samples (i.e., E3a for African Americans and R1b for U.S. Caucasians) are in agreement with unpublished results collected by Chris Tyler-Smith's lab at Oxford University (Chris Tyler-Smith, personal communication). It is interesting to note that simple analysis of only two Y-SNPs, M2 and P25, would have accurately reflected over 60% of the 229 samples in these U.S. sample datasets. Additional information from the other 48 markers typed did not further refine the position of these samples on the YCC haplogroup tree. The results summarized in Fig. 1 demonstrate that a majority of the markers were not useful in classifying the 18 observed haplogroups. Rather, the same results could have been obtained with far fewer markers. Over 77% of the

samples could have been assigned into specific Y haplogroups using only the markers M2, P25, and M170. Taking the entire dataset into account, only eight of 229 samples were observed once for a specific haplogroup. Unfortunately, as can be seen by the results summarized, the Y-SNPs examined here do not appear very useful for application to forensic problems where it would be advantageous to separate all individuals tested into unique types. Additional Y-SNPs will need to be discovered in individuals residing in the R1b and E3a haplogroups to potentially enhance the forensic utility of Y-SNP markers. If Y-SNPs are used at all in future forensic applications, it will likely be in conjunction with Y-STR information (31).

References

[PubMed] 1. Brookes AJ. The essence of SNPs. Gene 1999;234:177–86.
2. Jobling MA. Y-chromosomal SNP haplotype diversity in forensic analysis. [PubMed] Forensic Sci Int 2001;118:158–62.
3. Gut IG. Automation in genotyping of single nucleotide polymorphisms. [PubMed] Hum Mutat 2001;17:475–92.
4. de Knijff P. Messages through bottlenecks: On the combined use of slow and fast evolving polymorphic markers on the human Y chromosome. [PubMed] Am J Hum Genet 2000;67:1055–61.
5. Underhill PA, Shen P, Lin AA, Jin L, Passarino G, Yang WH, et al. Y chromosome sequence variation and the history of human populations. [PubMed] Nat Genet 2000;26:358–61.
6. Underhill PA, Passarino G, Lin AA, Shen P, Mirazon LM, Foley RA, et al. The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. Ann Hum Genet 2001;65:43– [PubMed] 62.
7. Wells RS, Yuldasheva N, Ruzibakiev R, Underhill PA, Evseeva I, Blue-Smith J, et al. The Eurasian heartland: a continental perspective on [PubMed] Y-chromosome diversity. Proc Natl Acad Sci USA 2001;98:10244–9.
8. Hurles ME, Nicholson J, Bosch E, Renfrew C, Sykes BC, Jobling MA. Y chromosomal evidence for the origins of oceanic-speaking peoples. [PubMed] Genetics 2002;160:289–303.
9. Qamar R, Ayub Q, Mohyuddin A, Helgason A, Mazhar K, Mansoor A, et al. Y-chromosomal DNA variation in Pakistan. Am J Hum Genet [PubMed] 2002;70:1107–24.
10. Kayser M, Brauer S, Weiss G, Schiefenhovel W, Underhill P, Shen P, et al. Reduced Y-chromosome, but not mitochondrial DNA, diversity in human [PubMed] populations from West New Guinea. Am J Hum Genet 2003;72:281–302.
11. Foster EA, Jobling MA, Taylor PG, Donnelly P, de Knijff P, Mieremet [PubMed] R, et al. Jefferson fathered slave's last child. Nature 1998;396:27–8.
12. Zerjal T, Xue Y, Bertorelle G, Wells RS, Bao W, Zhu S, et al. The genetic [PubMed] legacy of the mongols. Am J Hum Genet 2003;72:717–21.
13. Seielstad MT, Hebert JM, Lin AA, Underhill PA, Ibrahim M, Vollrath D, et al. Construction of human Y-chromosomal haplotypes using a new [PubMed] polymorphic A to G transition. Hum Mol Genet 1994;3:2159–61.
14. Ye J, Parra EJ, Sosnoski DM, Hiester K, Underhill PA, Shriver MD. Melting curve SNP (McSNP) genotyping: a useful approach for diallelic [PubMed] genotyping in forensic science. J Forensic Sci 2002;47:593–600.

15. Lareu M, Puente J, Sobrino B, Quintans B, Brion M, Carracedo A. The use of the LightCycler for the detection of Y chromosome SNPs. Forensic Sci Int 2001;118:163–8. [PubMed]
16. Raitio M, Lindroos K, Laukkanen M, Pastinen T, Sistonen P, Sajantila A, et al. Y-chromosomal SNPs in Finno-Ugric-speaking populations analyzed by minisequencing on microarrays. Genome Res 2001;11:471–82. [PubMed]
17. Paracchini S, Arredi B, Chalk R, Tyler-Smith C. Hierarchical high-throughput SNP genotyping of the human Y chromosome using MALDI-TOF mass spectrometry. Nucleic Acids Res 2002;30:e27. [PubMed]
18. Inagaki S, Yamamoto Y, Doi Y, Takata T, Ishikawa T, Yoshitome K, et al. Typing of Y chromosome single nucleotide polymorphisms in a Japanese population by a multiplexed single nucleotide primer extension reaction. Legal Medicine 2002;4:202–6. [PubMed]
19. Miller SA, Dykes DD, Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res 1988;16:1215. [PubMed]
20. Singer VL, Jones LJ, Yue ST, Haugland RP. Characterization of PicoGreen reagent and development of a fluorescence-based solution assay for double-stranded DNA quantitation. Anal Biochem 1997;249:228–38. [PubMed]
21. Butler JM, Schoske R, Vallone PM, Redman JW, Kline MC. Allele frequencies for 15 autosomal STR loci on U.S. Caucasian, African American, and Hispanic populations. J Forensic Sci 2003;48:908–11. [PubMed]
22. Y chromosome consortium. A nomenclature system for the tree of human Y-chromosomal binary haplogroups. Genome Res 2002;12:339–48. [PubMed]
23. Dugan K, Olivastro E, Underhill P, Budowle B. SNP analysis of the Y chromosome. Poster presented at the 13th International Symposium on Human Identification, October 2002; http://www.promega.com/geneticidproc/ussymp13proc/abstracts/38_dugan.pdf
24. 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 2003;375:333–43. [PubMed]
25. Butler JM, Devaney JM, Marino MA, Vallone PM. Quality control of PCR primers used in multiplex STR amplification reactions. Forensic Sci Int 2001;119:87–96. [PubMed]
26. Bell PA, Chaturvedi S, Gelfand CA, Huang CY, Kochersperger M, Kopla R, et al. SNPstream® UHT: Ultra-high throughput SNP genotyping for pharmacogenomics and drug discovery. Biotechniques 2002;Suppl: 70–7. [PubMed]
27. Kayser M, Brauer S, Schadlich H, Prinz M, Batzer MA, Zimmerman PA, et al. Y chromosome STR haplotypes and the genetic structure of U.S. populations of African, European, and Hispanic ancestry. Genome Res 2003;13:624–34. [PubMed]
28. Shewale JG, Richey SL, Sinha SK Anomalous amplification of the amelogenin locus typed by AmpFℓSTR Profiler Plus amplification kit. Forensic Sci Comm 2000;2(4); available at http://www.fbi.gov/hq/lab/fsc/backissu/oct2000/shewale.htm
29. Skaletsky H, Kuroda-Kawaguchi T, Minx PJ, Cordum HS, Hillier L, Brown LG. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. Nature 2003;423:825–37. [PubMed]
30. Sanchez JJ, Borsting C, Hallenberg C, Buchard A, Hernandez A, Morling N. Multiplex PCR and minisequencing of SNPs: a model with 35 Y chromosome SNPs. Forensic Sci Int 2003;137:74–84. [PubMed]
31. Butler JM, Schoske R, Vallone PM, Kline MC, Redd, AJ, Hammer MF. A novel multiplex for simultaneous amplification of 20 Y chromosome STR markers. Forensic Sci Int 2002;129:10–24. [PubMed]

Additional information and reprint requests:
Peter Vallone, Ph.D.
National Institute of Standards and Technology
100 Bureau Dr., Mail Stop 8311
Gaithersburg, MD 20899-8311
E-mail: petev@nist.gov