Available online at www.sciencedirect.com
SCIENCE DIRECT®
Forensic Science International 149 (2005) 279–286
Forensic Science International
www.elsevier.com/locate/forsciint

Announcement of Population Data

Allele frequencies for 70 autosomal SNP loci with U.S. Caucasian, African-American, and Hispanic samples

Peter M. Vallone*, Amy E. Decker, John M. Butler

Biotechnology Division, National Institute of Standards and Technology, Gaithersburg, MD 20899-8311, USA

Received 1 June 2004; received in revised form 28 July 2004; accepted 30 July 2004
Available online 23 September 2004

Abstract

189 samples from 3 different U.S. sample groups Caucasian (74), African American (71) and Hispanic (44) were typed for 70 autosomal genetic markers. These 70 markers are bi-allelic (C/T) short nucleotide polymorphisms (SNPs). For each sample, the 70 SNP markers were typed in 11 unique 6-plexes and a single 4-plex PCR. A total of 10 of the 210 tests (70 loci × 3 populations) for Hardy-Weinberg equilibrium indicated a statistically significant result. In order to evaluate the minimum number of SNP loci needed to distinguish all 189 samples from one another, we ranked the loci according to their levels of observed heterozygosity and p-values obtained upon testing for Hardy-Weinberg equilibrium. The top 12 loci according to these ranking criteria were tabulated along with the number of unique genotypes observed when combining subsequent SNP markers. The 12 selected SNPs possessed an observed heterozygosity of >0.45 in all three populations examined and thus would be expected to exhibit more differences between samples. All of the 189 samples in this study were individualized with a subset of 12 SNP loci. However, it is likely that the addition of more than 12 SNP loci will be required to resolve larger sets of unrelated individuals from one another. By way of comparison, in these same 189 individuals all but one pair is resolved from one another with three of the traditional short tandem repeat (STR) loci possessing the highest heterozygosity values (D2S1338, D18S51, and FGA) run with the Identifiler kit. The final pair of unrelated samples could be resolved with the combination of 4 STR loci: D2S1338, D18S51, FGA, and VWA.
Published by Elsevier Ireland Ltd.

Keywords: Single nucleotide polymorphism; DNA; SNP; Autosomal markers; Primer extension

1. Population 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). All samples were previously examined with 15 autosomal short tandem repeats and the amelogenin sex-typing marker using the AmpFlSTR Identifiler kit (Applied Biosystems, Foster City, CA) to verify that each sample was unique [1].

N: Seventy-four U.S. Caucasians, 71 African-Americans and 44 U.S. Hispanics were typed for 70 autosomal bi-allelic single nucleotide polymorphism (SNP) markers.

2. DNA extraction

Blood samples were extracted using a modified salting out procedure [2].

* Corresponding author. Tel.: +1 301 975 4872; fax: +1 301 975 8505.
E-mail address: peter.vallone@nist.gov (P.M. Vallone).

0379-0738/$ – see front matter. Published by Elsevier Ireland Ltd.
doi:10.1016/j.forsciint.2004.07.014

3. Quantification

Extracted DNA was quantified using UV spectrophotometry followed by a PicoGreen assay [3] to adjust concentrations to approximately 1 ng/μl.

4. SNP markers

The 70 autosomal SNP markers are listed in Table 1 (see also http://www.cstl.nist.gov/biotech/strbase/SNP.htm). The PCR primer sequences were obtained from Orchid Cellmark (personal communication, Jeanine Baisch, Orchid Cellmark Dallas). The exact chromosomal locations were ascertained using BLAT (http://genome.ucsc.edu/cgi-bin/hgBlat) and dbSNP (http://www.ncbi.nlm.nih.gov/SNP/) and are based on the July 2003 assembly of the human genome. All of the SNPs are C/T transitions.

5. PCR amplification

For each sample, the 70 SNP markers were typed in 11 unique 6-plexes and a single 4-plex PCR. The final concentrations of the six (or 4) PCR primer pairs were present at 0.5 μM for all multiplex PCRs. Amplifications were performed in reaction volumes of 10 μl using a master mix containing 1X GeneAmp® PCR Gold buffer (Applied Biosystems, Foster City, CA), 4.5 mmol/l $MgCl_2$, 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 0.5 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) [4]:

95 °C for 10 min
Three cycles of {95 °C for 30 s, 50 °C for 55 s, 72 °C for 30 s}
18 cycles of {95 °C for 30 s, 50 °C for 30 s +0.2 °C per cycle, 72 °C for 30 s}
11 cycles of {95 °C for 30 s, 55 °C for 30 s, 72 °C for 30 s}
72 °C for 7 min
25 °C until removed from thermocycler

Following PCR amplification, unincorporated primers and dNTPs were removed by adding 4 μl of a Exo-SAP enzyme cocktail consisting of 1.4 μl Exonuclease I (10 U/μl) and 2.6 μl (1 U/μl) of shrimp alkaline phosphatase (SAP; USB Corp., Cleveland, OH) to each 10 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.

6. Alelle specific primer extension (ASPE)

ASPE reactions were also carried out in eleven 6-plexes and a single 4-plex. Multiplex primer extension reactions were conducted 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, which contained empirically balanced primers (approximately 1 μM each). 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 μl of SAP (1 U/μl). Reactions were mixed briefly and incubated at 37 °C for 40 min then 90 °C for 5 min.

7. Electrophoresis and typing

A 1.0 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.

8. Analysis of data

The data were analyzed with PowerMarker v3.07 [5]. Allele frequencies, expected heterozygosity values and p-values (based on an exact test with 1000 reshufflings) for each marker are provided in Tables 2–4 for the three U.S. sample groups.

9. Access to the data

SNP marker information is available on the forensic SNP site: http://www.cstl.nist.gov/biotech/strbase/SNP.htm and genotyping results are posted at http://www.cstl.nist.gov/biotech/strbase/NISTpop.htm.

10. Results and discussion

Tables 2–4 contain the observed allele frequencies for U.S. Caucasian, African-American, and Hispanic samples, respectively. The C and T allele frequencies were used to calculate the expected heterozygosities that were then compared to the observed frequencies of heterozygotes. A total of 10 of the 210 tests (70 loci × 3 populations) for Hardy–Weinberg equilibrium indicated a deviation from the expected result. As has been noted before, it is reasonable to expect approximately 5%, or 10 to 11 out of 210, of

Table 1
Information on 70 autosomal SNP loci examined in this study sorted by chromosome position. Locus numbers are arbitrary. Chromosome positions were determined using the July 2003 human genome reference sequence. GenBank and dbSNP information can be retrieved from http://www.ncbi.nlm.nih.gov. The SNP Consortium numbers (TSC #) from http://snp.cshl.org are also listed

SNP locus #	Chromosome	Chromosome position	GenBank reference	Reference allele	dbSNP reference	TSC #	PCR Product size
58	1	14,359,351	AL034395.6	T	rs734664	TSC0026510	64
18	1	33,657,974	AL161643.2	T	rs732889	TSC0022461	65
66	1	53,564,936	AL049745.9	T	rs702490	TSC0243017	73
48	1	109,978,841	AL160006.1	C	rs924181	TSC0239374	63
24	1	164,087,184	AL009182.1	T	rs2013526	TSC0013419	65
17	1	181,421,773	AC009481.4	C	rs997568	TSC0002990	64
50	2	7,855,560	AC092580.3	C	rs772436	TSC0013222	69
35	2	12,145,530	AC018866	T	rs896499	TSC0170191	78
22	2	33,160,796	AL133244.1	C	rs1020636	TSC0224883	61
28	2	53,012,403	AC018713.7	C	rs2015632	TSC0006100	70
30	2	59,987,336	AC007100.3	C	rs1019264	TSC0222995	80
38	2	155,136,334	AC008166.2	T	rs1079861	TSC0066071	78
52	2	205,040,634	AC009965.9	T	rs3096741	TSC0211925	64
1	2	221,198,958	AC008064	T	rs734295	TSC0025620	62
13	4	182,888,181	AC019235.6	T	rs716360	TSC0005559	62
19	5	118,134,847	AC008444.4	T	rs730907	TSC0018534	62
16	5	132,731,841	AC010307.7	T	rs733023	TSC0022796	61
67	5	153,889,557	AC026688.7	C	rs880083	TSC0214907	73
14	5	164,661,137	AC008644	C	rs1024997	TSC0248350	89
26	6	3,295,184	AL160398.2	T	rs730488	TSC0017645	68
54	6	15,118,209	AL050335.3	C	rs927628	TSC0244208	70
44	6	39,929,605	AL136089.9	T	rs716856	TSC0009879	62
10	6	65,000,740	AL078597.1	C	rs1028484	TSC0253002	68
3	6	88,362,198	AL133211.9	T	rs1075665	TSC0021586	63
4	6	119,778,606	AL360215.1	C	rs924397	TSC0239700	63
45	6	124,123,520	AL354936.2	T	rs765533	TSC0069344	77
56	6	166,939,350	Z98049.1	C	rs916388	TSC0201398	61
41	7	15,262,947	AC012061.4	T	rs1072292	TSC0010262	79
8	7	51,706,494	AC022458.3	T	rs997556	TSC0002961	61
7	7	102,804,023	AC006316.2	T	rs123714	TSC0105900	64
33	8	57,612,000	AC009597.5	T	rs919023	TSC0232293	81
49	8	63,755,655	AC018398	C	rs734701	TSC0026586	69
37	8	103,506,597	AP002852.3	C	rs892503	TSC0163923	76
39	8	117,079,183	AF130343.1	T	rs6469629	TSC0218207	69
32	9	17,592,496	AL133214.1	T	rs1008730	TSC0087190	69
51	10	27,923,937	AC024606.1	T	rs997750	TSC0003290	63
12	10	82,436,151	AC013242.8	T	rs922992	TSC0237737	63
43	10	113,292,473	AL136119.1	T	rs585070	TSC0231103	81
20	11	19,942,027	AC079361.1	C	rs729999	TSC0016655	60
63	11	35,506,187	AL354921.1	T	rs627119	TSC0016429	59
46	11	103,380,259	AP003043.2	C	rs1021290	TSC0225784	80
62	11	131,628,725	AP004248.2	C	rs921269	TSC0235383	76
40	12	30,160,004	AC068811.8	T	rs959566	TSC0261969	75
60	12	100,652,449	AC063950.3	T	rs730013	TSC0016681	72
36	13	37,899,740	AL158194.1	C	rs730249	TSC0017130	65
2	13	68,098,515	AL162378.1	C	rs2018205	TSC0062893	108
59	13	98,078,430	AL445184.1	T	rs1105576	TSC0135614	67
69	13	107,113,189	AL136132.1	C	rs729549	TSC0015745	72
23	14	53,115,754	AL133444.4	T	rs911621	TSC0193791	60
57	14	82,658,064	AL583743.3	C	rs734656	TSC0026489	64
5	15	20,547,981	AC016204.1	T	rs999842	TSC0014520	64
27	15	52,239,965	AC022302.7	T	rs719211	TSC0043836	82
70	15	58,792,647	AC022898.1	T	rs877228	TSC0209754	62
25	17	32,063,659	AC011824.8	C	rs727206	TSC0061444	68

Table 1 (*Continued*)

SNP locus #	Chromosome	Chromosome position	GenBank reference	Reference allele	dbSNP reference	TSC #	PCR Product size
21	17	39,523,608	AC007455.7	C	rs2010209	TSC0005110	70
6	17	55,624,843	AC007114.8	C	rs917927	TSC0230724	74
42	17	79,577,854	AC127496.5	C	rs868432	TSC0124845	70
9	18	22,615,097	AC018371.1	T	rs1017415	TSC0220510	70
47	18	32,376,937	AC131053.2	T	rs4105107	TSC0133660	71
65	19	1,126,396	AC120982.2	T	rs873289	TSC0202573	65
53	19	16,310,517	AC020917.4	C	rs1000329	TSC0015380	67
64	20	23,525,035	AL121894.2	C	rs1003204	TSC0038789	68
55	20	25,048,105	AL080312.1	C	rs743018	TSC0113727	71
68	20	43,388,976	AL034419.2	T	rs4467339	TSC0215392	66
29	20	60,743,636	AL160412.1	T	rs1000322	TSC0015372	75
15	21	17,486,896	AP001669	T	rs18579	TSC0003543	67
31	22	34,220,248	AL022334.1	C	rs736210	TSC0029823	62
61	22	35,362,839	AL049749.2	C	rs738518	TSC0112989	64
11	22	41,403,208	AL049757.1	C	rs738532	TSC0113010	71
34	22	41,810,649	Z82214.2	T	rs138952	TSC0117592	64

Table 2
Allele frequencies observed for 74 U.S. Caucasians listed by SNP locus number (see Table 1)

Caucasian (N = 74)												
	1	2	3	4	5	6	7	8	9	10	11	12
CC	0.243	0.405	0.068	0.581	0.311	0.149	0.486	0.108	0.203	0.068	0.257	0.054
TT	0.243	0.135	0.514	0.135	0.189	0.338	0.122	0.378	0.284	0.459	0.216	0.365
CT	0.514	0.459	0.419	0.284	0.500	0.514	0.392	0.514	0.514	0.473	0.527	0.581
He	0.500	0.463	0.401	0.401	0.493	0.482	0.433	0.463	0.497	0.423	0.499	0.452
P	0.816	1.000	1.000	**0.008**	0.816	0.475	0.413	0.305	1.000	0.269	0.818	**0.021**
	13	14	15	16	17	18	19	20	21	22	23	24
CC	0.068	0.243	0.392	0.446	0.243	0.162	0.473	0.365	0.270	0.108	0.270	0.203
TT	0.514	0.311	0.122	0.149	0.284	0.324	0.054	0.203	0.189	0.432	0.176	0.270
CT	0.419	0.446	0.486	0.405	0.473	0.514	0.473	0.432	0.541	0.459	0.554	0.527
He	0.401	0.498	0.463	0.456	0.499	0.487	0.412	0.487	0.497	0.447	0.496	0.498
P	1.000	0.220	0.805	0.183	0.663	0.818	0.163	0.348	0.485	1.000	0.362	0.650
	25	26	27	28	29	30	31	32	33	34	35	36
CC	0.243	0.176	0.162	0.068	0.257	0.432	0.419	0.527	0.122	0.581	0.257	0.108
TT	0.203	0.446	0.432	0.689	0.284	0.149	0.122	0.122	0.311	0.068	0.203	0.243
CT	0.554	0.378	0.405	0.243	0.459	0.419	0.459	0.351	0.568	0.351	0.541	0.649
He	0.499	0.463	0.463	0.307	0.500	0.460	0.456	0.418	0.482	0.368	0.499	0.491
P	0.480	0.135	0.327	0.119	0.350	0.302	0.797	0.092	0.088	0.538	0.642	**0.008**
	37	38	39	40	41	42	43	44	45	46	47	48
CC	0.378	0.095	0.378	0.149	0.297	0.311	0.257	0.473	0.122	0.189	0.162	0.351
TT	0.122	0.473	0.149	0.514	0.216	0.149	0.216	0.095	0.446	0.284	0.351	0.162
CT	0.500	0.432	0.473	0.338	0.486	0.541	0.527	0.432	0.432	0.527	0.486	0.486
He	0.467	0.428	0.474	0.433	0.497	0.487	0.499	0.428	0.447	0.496	0.482	0.482
P	0.444	0.790	0.806	0.060	0.827	0.491	0.822	0.791	0.790	0.635	0.803	0.802
	49	50	51	52	53	54	55	56	57	58	59	60
CC	0.135	0.203	0.176	0.365	0.095	0.284	0.446	0.419	0.149	0.081	0.081	0.351
TT	0.568	0.378	0.257	0.176	0.541	0.189	0.135	0.108	0.392	0.662	0.527	0.203
CT	0.297	0.419	0.568	0.459	0.365	0.527	0.419	0.473	0.459	0.257	0.392	0.446
He	0.407	0.485	0.497	0.482	0.401	0.496	0.452	0.452	0.470	0.331	0.401	0.489
P	**0.008**	0.229	0.244	0.638	0.566	0.645	0.472	0.612	0.618	0.068	0.785	0.331

Table 2 (*Continued*)

Caucasian (N = 74)										
	61	62	63	64	65	66	67	68	69	70
CC	0.068	0.473	0.189	0.162	0.243	0.378	0.486	0.324	0.216	0.284
TT	0.608	*0.027*	0.486	0.284	0.216	0.162	0.054	0.108	0.351	0.149
CT	0.324	0.500	0.324	0.554	0.541	0.459	0.459	0.568	0.432	0.568
He	0.354	0.401	0.456	0.493	0.500	0.477	0.407	0.477	0.491	0.491
P	0.326	**0.043**	**0.011**	0.232	0.493	0.805	0.401	0.142	0.346	0.263

He = expected heterozygosity; *P*: Fisher's exact test for Hardy–Weinberg equilibrium, based on 1000 shufflings. Values that are below the minimum allele frequency of 5/2 N (0.034) are italicized.

Table 3
Allele frequencies observed for 71 African-Americans listed by SNP locus number

African-American (N = 71)												
	1	2	3	4	5	6	7	8	9	10	11	12
CC	0.648	0.113	0.141	0.352	0.141	0.127	0.648	0.183	0.225	*0.014*	0.070	0.394
TT	0.070	0.352	0.437	0.141	0.338	0.563	0.070	0.408	0.338	0.662	0.549	0.155
CT	0.282	0.535	0.423	0.507	0.521	0.310	0.282	0.408	0.437	0.324	0.380	0.451
He	0.333	0.471	0.456	0.478	0.481	0.405	0.333	0.475	0.494	0.290	0.385	0.471
P	0.278	0.207	0.441	0.808	0.468	**0.043**	0.275	0.226	0.225	0.680	0.760	0.624
	13	14	15	16	17	18	19	20	21	22	23	24
CC	0.141	0.296	0.239	0.479	0.113	0.113	0.634	0.197	0.070	*0.028*	0.282	0.268
TT	0.451	0.113	0.338	0.085	0.479	0.451	0.042	0.366	0.606	0.648	0.239	0.239
CT	0.408	0.592	0.423	0.437	0.408	0.437	0.324	0.437	0.324	0.324	0.479	0.493
He	0.452	0.483	0.495	0.422	0.433	0.443	0.325	0.486	0.357	0.308	0.499	0.500
P	0.298	0.093	0.158	0.784	0.396	1.000	0.720	0.469	0.326	1.000	0.644	1.000
	25	26	27	28	29	30	31	32	33	34	35	36
CC	0.099	0.394	0.239	0.225	0.113	0.352	0.380	0.113	0.197	0.493	0.113	0.352
TT	0.394	0.155	0.254	0.282	0.451	0.169	0.183	0.423	0.338	0.042	0.535	0.197
CT	0.507	0.451	0.507	0.493	0.437	0.479	0.437	0.465	0.465	0.465	0.352	0.451
He	0.456	0.471	0.500	0.498	0.443	0.483	0.481	0.452	0.490	0.398	0.411	0.488
P	0.310	0.606	1.000	1.000	1.000	0.805	0.315	0.799	0.645	0.234	0.157	0.622
	37	38	39	40	41	42	43	44	45	46	47	48
CC	0.324	0.211	0.113	0.141	0.113	0.225	0.056	0.606	0.380	0.211	0.099	0.380
TT	0.183	0.366	0.592	0.465	0.549	0.310	0.479	0.056	0.183	0.366	0.465	0.169
CT	0.493	0.423	0.296	0.394	0.338	0.465	0.465	0.338	0.437	0.423	0.437	0.451
He	0.490	0.488	0.385	0.448	0.405	0.496	0.411	0.349	0.481	0.488	0.433	0.478
P	1.000	0.335	0.061	0.182	0.159	0.639	0.256	0.488	0.331	0.336	0.786	0.625
	49	50	51	52	53	54	55	56	57	58	59	60
CC	0.085	0.465	0.310	0.099	0.423	0.282	0.211	0.408	0.042	0.239	0.282	0.352
TT	0.549	0.070	0.099	0.648	0.169	0.141	0.310	0.141	0.535	0.225	0.183	0.211
CT	0.366	0.465	0.592	0.254	0.408	0.577	0.479	0.451	0.423	0.535	0.535	0.437
He	0.392	0.422	0.478	0.349	0.468	0.490	0.495	0.464	0.378	0.500	0.495	0.490
P	0.369	0.401	0.081	**0.030**	0.307	0.221	0.626	0.616	0.532	0.633	0.475	0.337
	61	62	63	64	65	66	67	68	69	70		
CC	0.310	0.352	0.268	0.479	0.183	0.423	0.592	0.380	0.338	0.296		
TT	0.225	0.141	0.183	0.099	0.310	0.169	0.113	0.113	0.197	0.141		
CT	0.465	0.507	0.549	0.423	0.507	0.408	0.296	0.507	0.465	0.563		
He	0.496	0.478	0.496	0.428	0.492	0.468	0.385	0.464	0.490	0.488		
P	0.634	0.803	0.469	0.786	1.000	0.296	0.059	0.459	0.626	0.226		

See Table 1. He = expected heterozygosity; *P*: Fisher's exact test for Hardy–Weinberg equilibrium, based on 1000 shufflings. Values that are below the minimum allele frequency of 5/2 N (0.035) are highlighted.

Table 4
Allele frequencies observed for 44 U.S. Hispanics listed by SNP locus number (see Table 1)

Hispanic (N = 44)												
	1	2	3	4	5	6	7	8	9	10	11	12
CC	0.455	0.477	0.114	0.364	0.364	*0.045*	0.432	0.182	0.227	0.091	0.182	0.205
TT	0.068	*0.045*	0.341	0.136	0.136	0.432	0.114	0.386	0.273	0.409	0.341	0.341
CT	0.477	0.477	0.545	0.500	0.500	0.523	0.455	0.432	0.500	0.500	0.477	0.455
He	0.425	0.407	0.474	0.474	0.474	0.425	0.449	0.479	0.499	0.449	0.487	0.491
P	0.723	0.441	0.522	1.000	1.000	0.177	1.000	0.545	1.000	0.741	1.000	0.724
	13	14	15	16	17	18	19	20	21	22	23	24
CC	0.068	0.341	0.500	0.386	0.114	0.295	0.477	0.273	0.136	0.091	0.250	0.273
TT	0.568	0.205	0.091	0.159	0.545	0.318	0.068	0.205	0.364	0.455	0.364	0.273
CT	0.364	0.455	0.409	0.455	0.341	0.386	0.455	0.523	0.500	0.455	0.386	0.455
He	0.375	0.491	0.416	0.474	0.407	0.500	0.416	0.498	0.474	0.434	0.494	0.500
P	1.000	0.777	1.000	0.782	0.311	0.135	0.719	1.000	1.000	1.000	0.226	0.565
	25	26	27	28	29	30	31	32	33	34	35	36
CC	0.318	0.205	0.227	0.091	0.227	0.114	0.523	0.477	0.136	0.568	0.182	0.318
TT	0.114	0.455	0.227	0.682	0.364	0.295	0.182	0.091	0.432	0.091	0.318	0.227
CT	0.568	0.341	0.545	0.227	0.409	0.591	0.295	0.432	0.432	0.341	0.500	0.455
He	0.479	0.469	0.500	0.325	0.491	0.483	0.442	0.425	0.456	0.386	0.491	0.496
P	0.326	0.097	0.766	0.063	0.365	0.198	**0.036**	1.000	0.752	0.435	1.000	0.558
	37	38	39	40	41	42	43	44	45	46	47	48
CC	0.523	*0.045*	0.455	0.114	0.295	0.250	0.205	0.523	0.091	0.136	0.205	0.477
TT	*0.045*	0.568	0.091	0.477	0.159	0.136	0.432	*0.023*	0.455	0.318	0.477	*0.023*
CT	0.432	0.386	0.455	0.409	0.545	0.614	0.364	0.455	0.455	0.545	0.318	0.500
He	0.386	0.363	0.434	0.434	0.491	0.494	0.474	0.375	0.434	0.483	0.463	0.397
P	0.694	1.000	1.000	0.727	0.542	0.135	0.119	0.248	1.000	0.522	**0.048**	0.143
	49	50	51	52	53	54	55	56	57	58	59	60
CC	0.068	0.318	0.227	0.29545	0.159	0.250	0.409	0.500	0.182	*0.023*	0.136	0.182
TT	0.636	0.227	0.318	0.25	0.341	0.295	0.136	0.068	0.455	0.636	0.500	0.250
CT	0.295	0.455	0.455	0.45455	0.500	0.455	0.455	0.432	0.364	0.341	0.364	0.568
He	0.339	0.496	0.496	0.499	0.483	0.499	0.463	0.407	0.463	0.312	0.434	0.498
P	0.381	0.562	0.563	0.576	1.000	0.558	1.000	1.000	0.168	1.000	0.294	0.564
	61	62	63	64	65	66	67	68	69	70		
CC	0.068	0.409	0.205	0.205	0.386	0.432	0.295	0.295	0.273	0.318		
TT	0.455	0.091	0.341	0.205	0.227	0.182	0.227	0.227	0.227	0.205		
CT	0.477	0.500	0.455	0.591	0.386	0.386	0.477	0.477	0.500	0.477		
He	0.425	0.449	0.491	0.500	0.487	0.469	0.498	0.498	0.499	0.494		
P	0.721	0.744	0.777	0.383	0.191	0.336	0.797	0.754	1.000	1.000		

He = expected heterozygosity; *P*: Fisher's exact test for Hardy–Weinberg equilibrium, based on 1000 shufflings. Values that are below the minimum allele frequency of 5/2 N (0.057) are highlighted.

the comparisons to deviate from Hardy–Weinberg equilibrium (see [6,7]). Those *P*-values significant at the 95% confidence level are those less than 0.05 and bolded in Tables 2–4. Six were observed in Caucasian samples and two in both the African-American and Hispanic data sets.

Typically the minimum number of samples needed to provide a robust estimate for allele frequencies with loci containing 5–15 alleles is 100–150 samples for each population [8]. Since we are measuring bi-allelic markers in this study that only have three possible genotypes (CC, TT or CT), a smaller number of samples should be sufficient [8] provided that we utilize a minimum allele frequency, such as 5/2 N [9]. An examination of the data in Tables 2–4 finds a total of 10 allele frequency measurements (out of 630 total) below the 5/2 N threshold across the three populations. Thus, to be conservative a minimum allele frequency of 0.034, 0.035, and 0.057 should be used with those infrequently observed alleles in our Caucasian (Table 2), African-American (Table 3), and Hispanic (Table 4) population data sets, respectively.

In order to evaluate the minimum number of SNP loci needed to distinguish all 189 samples from one another, we ranked the loci according to their levels of observed heterozygosity and *P*-values obtained upon testing for Hardy–Weinberg equilibrium. The top 12 loci according

Table 5
Number of unique genotypes observed in the set of 189 unrelated individuals examined in this study with the addition of each SNP from a set of the top 12 SNPs ranked by observed heterozygosity and *P*-value. Locus number correlates to Table 1

No. of markers combined	Locus # (see Table 1)	No. of unique genotypes
1	5	3
2	24	9
3	2	27
4	62	64
5	48	107
6	70	145
7	54	160
8	12	175
9	68	182
10	42	186
11	36	188
12	25	189

to these ranking criteria are listed in Table 5 along with the number of unique genotypes observed when combining subsequent SNP markers. The 12 selected SNPs possessed an observed heterozygosity of >0.45 in all three populations examined and thus would be expected to exhibit more differences between samples. All of the 189 samples in this study were individualized with these 12 SNP loci (Table 5). However, it is likely that the addition of more than 12 SNP loci will be required to resolve larger sets of unrelated individuals from one another. By way of comparison, in these same 189 individuals all but one pair is resolved from one another with three of the STR loci possessing the highest heterozygosity values (D2S1338, D18S51, and FGA) run with the Identifiler kit. The final pair of unrelated samples could be resolved with the combination of 4 STR loci: D2S1338, D18S51, FGA, and VWA (data not shown).

While not as polymorphic as multiallelic STRs, it appears that biallelic SNPs are still able to separate unrelated and related individuals from one another with a reasonable number of loci. However, the construction of large robust multiplexes, such as demonstrated by Sanchez et al. [10], with these and other SNPs will be necessary to enable routine recovery of genetic data from degraded and low copy number DNA present in forensic evidence [11]. The small PCR product sizes from these particular autosomal SNPs (see Table 1) have enabled recovery of information from severely degraded DNA samples and assisted in the identification of some of the World Trade Center victims with data generated by Orchid Cellmark (personal communication, Bob Shaler, Office of the Chief Medical Examiner, New York City).

This paper follows the guidelines for publication of population data requested by the journal [12].

Acknowledgements

This work was funded by the National Institute of Justice (NIJ) through an interagency agreement with the NIST Office of Law Enforcement Standards. Jeanine Baisch and Bob Giles from Orchid Cellmark (Dallas, TX) enabled this work with these 70 markers through providing their PCR primer information. The first tests with these SNP markers were done as part of a concordance study with Orchid Cellmark at the request of NIJ's World Trade Center Kinship and Data Analysis Panel in July 2002. Gordon Spangler, a graduate student from American University, helped with some of the early tests. Initial preparation of the population samples by Margaret Kline, Jan Redman, and Richard Schoske is gratefully acknowledged. Certain commercial equipment, instruments and materials are identified in order to specify experimental procedures as completely as possible. In no case does such identification imply a recommendation or endorsement by the National Institute of Standards and Technology nor does it imply that any of the materials, instruments or equipment identified are necessarily the best available for the purpose.

References

[1] J.M. Butler, R. Schoske, P.M. Vallone, J.W. Redman, M.C. Kline, Allele frequencies for 15 autosomal STR loci on U.S. Caucasian, African American, and Hispanic populations, J. Forensic Sci. 48 (2003) 908–911.
[2] S.A. Miller, D.D. Dykes, H.F. Polesky, A simple salting out procedure for extracting DNA from human nucleated cells, Nucl. Acids Res. 16 (1988) 1215.
[3] V.L. Singer, L.J. Jones, S.T. Yue, R.P. Haugland, Characterization of PicoGreen reagent and development of a fluorescence-based solution assay for double-stranded DNA quantitation, Anal. Biochem. 249 (1997) 228–238.
[4] P.A. Bell, S. Chaturvedi, C.A. Gelfand, C.Y. Huang, M. Kochersperger, R. Kopla, F. Modica, M. Pohl, S. Varde, R. Zhao, X. Zhao, M.T. Boyce-Jacino, A. Yassen, SNPstream® UHT: Ultra-high throughput SNP genotyping for pharmacogenomics and drug discovery, Biotech. Suppl. (2002) 70–77.
[5] K. Liu, S. Muse, PowerMarker: new genetic data analysis software. Version 3.07. Free program distributed by the author over the internet from http://www.powermarker.net/.
[6] B. Budowle, T.R. Moretti, A.L. Baumstark, D.A. Defenbaugh, K.M. Keys, Population data on the thirteen CODIS core short tandem repeat loci in African Americans, U.S. Caucasians, Hispanics, Bahamians, Jamaicans, and Trinidadians, J. Forensic Sci. 44 (1999) 1277–1286.
[7] P. Gill, L. Foreman, J.S. Buckleton, C.M. Triggs, H. Allen, A comparison of adjustment methods to test the robustness of an STR DNA database comprised of 24 European populations, Forensic Sci. Int. 131 (2003) 184–196.
[8] R. Chakraborty, Sample size requirements for addressing the population genetic issues of forensic use of DNA typing, Hum. Biol. 64 (1992) 141–159.
[9] Council Report (NRC II), The Evaluation of Forensic DNA Evidence, Washington, DC, 1996.

[10] J.J. Sanchez, C. Borsting, C. Hallenberg, A. Buchard, A. Hernandez, N. Morling, Multiplex PCR and minisequencing of SNPs-a model with 35 Y chromosome SNPs, Forensic Sci. Int. 137 (2003) 74–84.
[11] P. Gill, D.J. Werrett, B. Budowle, R. Guerrieri, An assessment of whether SNPs will replace STRs in national DNA databases–joint considerations of the DNA working group of the European Network of Forensic Science Institutes (ENFSI) and the Scientific Working Group on DNA Analysis Methods (SWGDAM), Sci. Justice 44 (2004) 51–53.
[12] P. Lincoln, A. Carracedo, Publication of population data of human polymorphisms, Forensic Sci. Int. 110 (2000) 3–5.