NIST
National Institute of Standards and Technology
Technology Administration, U.S. Department of Commerce

Email:
petev@nist.gov

The Evaluation of an Autosomal SNP 12-plex Assay

Peter M. Vallone, Amy E. Decker, Michael D. Coble and John M. Butler

Biotechnology Division, National Institute of Standards and Technology, 100 Bureau Drive MS 8311, Gaithersburg, MD 20899-8311

SNPs (single nucleotide polymorphisms) have potential to play a useful role in human identification testing. Small PCR amplicon sizes associated with SNP typing technologies make SNPs attractive for typing degraded DNA or other low copy number situations. SNP markers can be useful in combination with STRs for resolving complex paternity issues (e.g. incest), identifying victims of mass disasters where insufficient family references are available and possibly inferring population of origin. Important considerations for SNP markers are the larger number required to equal the discriminatory power compared to traditional STRs, their inability to resolve complex mixtures, issues related to databasing new loci, and the availability of a standard analysis platform. However, in appropriate situations SNPs can be useful as a supplementary tool complementary to STR markers.

Forensic Interest in Using SNP Markers

- Use on degraded samples, low copy number, or telogenic (shed) hairs
- Lower mutation rate (Paternity testing)
- Easier data interpretation (no microvariants or stutter)
- Amenable to high throughput analysis

General issues that need to be addressed

- How many SNPs = STR?
- PCR multiplexing (50-plex < 1ng DNA)
- Databases
- Platform for SNP typing? (many available)
- Unique interpretation issues (Mixtures)
- Validation
- Sensitivity
- Cost

Previously: 70 SNP markers were typed for 189 U.S. samples: 74 Caucasians, 71 African Americans and 44 Hispanics. The 70 SNPs were typed using a panel of 6-plex PCR/ASPE (allele specific primer extension) reactions.

Results: The allele distribution ranged from (0.25 – 0.74).The exact p-value was < 5% for 10 of the loci. A subset (12) of the 70 SNP markers were selected for developing a 12-plex autosomal SNP assay. The 12 loci selected had an observed heterozygosity of >0.45 in each of the 3 U.S. sample groups.

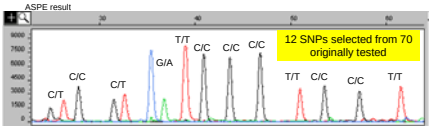
Vallone, P.M., Decker, A.E., Butler, J.M. (2005) Allele frequencies for 70 autosomal SNP loci with U.S. Caucasian, African American, and Hispanic Samples. *Forensic Sci. Int.* 149:279-286.

Candidate Loci for 12-plex

Locus	Chromosome	Chr. Position	Reference Allele	dbSNP Reference	TSC #	PCR Product Size
1	13	68,098,515	C	rs2018205	TSC0062893	108
2	15	20,547,981	T	rs999842	TSC0014520	64
3	10	82,436,151	T	rs922992	TSC0237737	63
4	1	164,087,184	T	rs2013526	TSC0013419	65
5	17	32,063,659	C	rs727206	TSC0061444	68
6	13	37,899,740	C	rs730249	TSC0017130	65
7	17	79,577,854	C	rs868432	TSC0124845	70
8	1	109,978,841	C	rs924181	TSC0239374	63
9	6	15,118,209	C	rs927628	TSC0244208	70
10	11	131,628,725	C	rs921269	TSC0235383	76
11	20	43,388,976	T	rs4467339	TSC0215392	66
12	15	58,792,647	T	rs877228	TSC0209754	62

Three loci (1,6 & 10) are located on chromosomes that already contain CODIS loci.
Locus 1 and 6 are 12.5 and 43.5 cM apart from D13S317, respectively.
Locus 10 and TH01 are 100 cM apart.
PCR amplicons range from 62 – 108 base pairs.

Autosomal 12-plex SNP Assay



12-plex PCR followed by 12-plex ASPE (ABI SNaPshot)
Fragments separated on a ABI 3100 in 35 minutes
The 12-plex assay has been run on over 1000 samples
Works well on 0.5 to 1 ng of template
12-plex genotypes fully concordant with those obtained with previous 6-plex assays

32 cycle PCR
1.5 U Taq Gold
15 uL volume

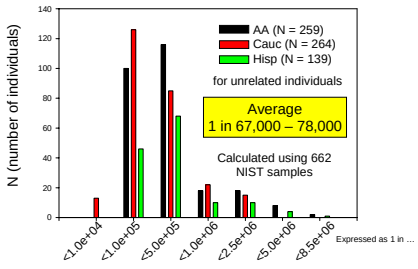
Frequency Data for 662 NIST samples

AA														Average Het
Locus		1	2	3	4	5	6	7	8	9	10	11	12	
n=259														
	CC	0.154	0.178	0.405	0.255	0.154	0.328	0.172	0.510	0.337	0.432	0.425	0.371	
	TT	0.351	0.351	0.162	0.251	0.359	0.208	0.336	0.104	0.143	0.108	0.112	0.116	
	CT	0.494	0.471	0.432	0.494	0.486	0.463	0.492	0.386	0.519	0.459	0.463	0.514	0.473
	He	0.481	0.485	0.470	0.500	0.479	0.493	0.487	0.418	0.481	0.447	0.451	0.468	
	Exact p-value	0.705	0.695	0.150	0.815	0.813	0.324	0.891	0.191	0.234	0.776	0.794	0.151	
Cauc														
Locus		1	2	3	4	5	6	7	8	9	10	11	12	
n=264														
	CC	0.360	0.333	0.125	0.197	0.295	0.167	0.330	0.311	0.311	0.534	0.386	0.333	
	TT	0.136	0.186	0.420	0.303	0.193	0.280	0.167	0.205	0.174	0.057	0.125	0.170	
	CT	0.504	0.481	0.455	0.500	0.511	0.553	0.504	0.485	0.515	0.409	0.489	0.496	0.492
	He	0.475	0.489	0.456	0.494	0.495	0.494	0.487	0.494	0.491	0.386	0.466	0.487	
	Exact p-value	0.380	0.707	0.885	0.795	0.543	0.066	0.604	0.712	0.393	0.373	0.521	0.805	
Hisp														
Locus		1	2	3	4	5	6	7	8	9	10	11	12	
n=139														
	CC	0.374	0.295	0.173	0.173	0.259	0.230	0.266	0.410	0.288	0.475	0.324	0.273	
	TT	0.101	0.230	0.338	0.317	0.194	0.230	0.173	0.115	0.245	0.065	0.230	0.273	
	CT	0.525	0.475	0.489	0.511	0.547	0.540	0.561	0.475	0.468	0.460	0.446	0.453	0.496
	He	0.463	0.498	0.486	0.490	0.498	0.500	0.496	0.456	0.499	0.416	0.496	0.500	
	Exact p-value	0.098	0.501	0.858	0.604	0.241	0.412	0.146	0.587	0.491	0.303	0.154	0.339	

All 662 profiles were unique with the 12-plex assay
The exact p-value for all loci > 5 %

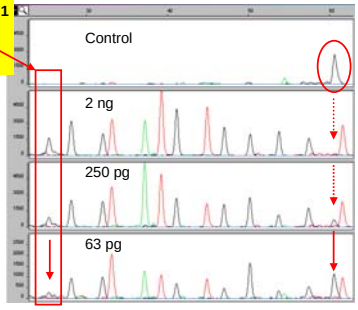
These projects were and are supported by **Grant Number 1999-IJ-R-A094 and 2003-IJ-R-029**, which is an interagency agreement between NIJ and the NIST Office of Law Enforcement Standards, awarded by the National Institute of Justice, Office of Justice Programs, US Department of Justice. Points of view in this document are those of the authors and do not necessarily represent the official position or policies of the US Department of Justice. 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.

Probability of a Random Match using 12-plex



Initial Sensitivity Studies

Very weak signal intensity for **locus 1** as the amount of input DNA decreases



An artifact specific to the ASPE primer for **locus 12** obscures accurate genotyping (especially at low amounts of input DNA)

Alternative PCR primers for Locus 1 were tried, but the signal intensity did not improve

The "reverse" ASPE primer for Locus 12 also exhibited the same artifact obscuring accurate genotyping

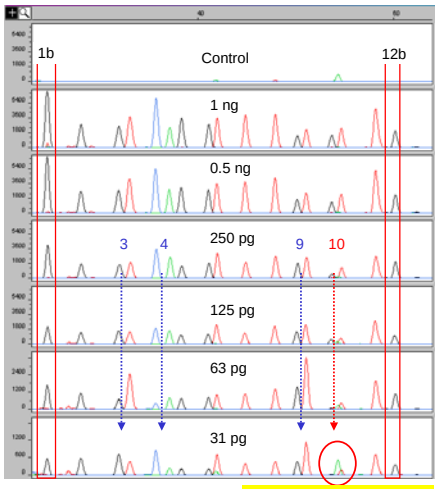
Two alternate loci from the set of 70 were chosen to replace loci 1 & 12

Sensitivity studies performed on pristine samples (concentration determined using ABI Quantifiler)

Locus	Chromosome	Chr. Position	Reference Allele	dbSNP Reference	TSC #	PCR Product Size
1b	5	164,661,137	C	rs1024997	TSC0248350	89
12b	11	35,506,187	T	rs627119	TSC0016429	59

With the 2 replacement loci all amplicons are now under 90 base pairs
Locus 1b is on a non-CODIS chromosome and Locus 12b is ~33 cM from TH01

Further Sensitivity Study



Green artifact peak will not obscure genotype calls

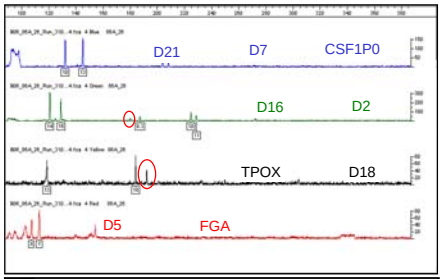
The signal for Locus 1b does not drop out at lower concentrations
The ASPE primer for Locus 12b does not exhibit any artifacts

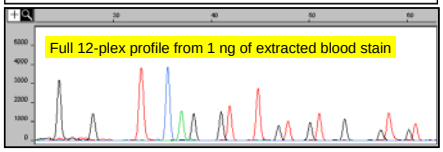
Below approximately 100 pg of input DNA various issues arise due to stochastic amplification:

Blue Arrows: For Locus 3,4 & 9, an imbalance in the heterozygous peaks occurs at low levels of DNA.

Red arrow: For Locus 10, overall signal intensity drops and potential allele drop out is observed.

Typing an Aged Blood Stain



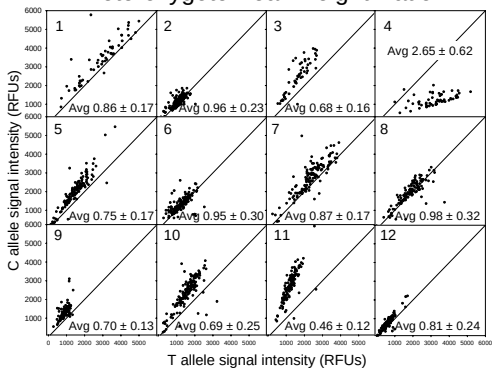


Above: Identifiler genotyping result from a blood stain aged 15 years stored at room temperature. (stored on 903 paper, Chelex extracted)

Below: The same sample extract as above typed by the 12-plex SNP assay. 11 different samples that gave partial profiles with Identifiler gave full profiles typed with the 12-plex assay.

The experiments shown above used 1 ng of input genomic DNA. (as determined by the ABI Quantifiler kit)

Heterozygote Peak Height Ratio



In order to evaluate the robustness *and* limitations of the assay it is important to characterize signal intensity and balance. The reproducibility of peak ratios will impact the ability to decipher mixtures.

The black line represents a 1:1 ratio of the heterozygote peak intensities. The average ratio ± the standard deviation is given for each locus. Note that Locus 4 (G/A) incorporates the blue/green dyes versus the black/red.

Typing results from 7 populations

Biaka Pygmys	**n=70**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	Average Het
	CC	0.186	0.243	0.471	0.100	0.200	0.629	0.157	0.329	0.757	0.214	0.443	0.814	
	TT	0.271	0.200	0.071	0.529	0.257	0.029	0.329	0.229	not obs.	0.300	0.100	not obs.	
	CT	0.543	0.557	0.457	0.371	0.543	0.343	0.514	0.443	0.243	0.486	0.457	0.186	0.429
	He	0.496	0.499	0.420	0.408	0.498	0.320	0.485	0.495	0.213	0.496	0.441	0.168	
	Exact p	0.492	0.465	0.582	0.395	0.491	0.723	0.610	0.356	0.253	1.000	1.000	0.463	
Cambodians	**n=25**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	
	CC	0.640	0.480	0.120	0.640	0.400	0.200	0.040	0.560	0.280	0.880	0.440	0.360	
	TT	0.120	0.240	0.440	not obs.	0.160	0.120	0.520	0.040	0.360	not obs.	0.120	0.040	
	CT	0.240	0.280	0.440	0.360	0.440	0.680	0.440	0.400	0.360	0.120	0.440	0.600	0.400
	He	0.365	0.471	0.449	0.295	0.471	0.497	0.385	0.365	0.497	0.113	0.449	0.449	
	Exact p	0.099	0.089	1.000	0.576	0.705	0.109	1.000	1.000	0.215	1.000	1.000	0.161	
Chinese_Taiwan	**n=49**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	
	CC	0.429	0.306	0.082	0.490	0.265	0.551	0.327	0.592	0.184	0.551	0.673	0.429	
	TT	0.122	0.163	0.327	0.163	0.143	0.082	0.224	0.020	0.429	0.020	0.020	0.143	
	CT	0.449	0.531	0.592	0.347	0.592	0.367	0.449	0.388	0.388	0.429	0.306	0.429	0.439
	He	0.453	0.490	0.470	0.447	0.493	0.390	0.495	0.337	0.470	0.359	0.287	0.459	
	Exact p	1.000	0.803	0.108	0.112	0.236	0.729	0.571	0.408	0.231	0.279	1.000	0.783	
European American	**n=92**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	
	CC	0.250	0.315	0.098	0.109	0.293	0.196	0.413	0.293	0.272	0.385	0.359	0.196	
	TT	0.250	0.152	0.478	0.326	0.239	0.217	0.163	0.207	0.152	0.055	0.141	0.228	
	CT	0.500	0.533	0.424	0.565	0.467	0.587	0.424	0.500	0.576	0.560	0.500	0.576	0.518
	He	0.500	0.487	0.428	0.476	0.499	0.500	0.469	0.496	0.493	0.446	0.476	0.499	
	Exact p	1.000	0.506	1.000	0.128	0.532	0.132	0.278	1.000	0.091	0.018	0.821	0.142	
Hausa	**n=39**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	
	CC	0.641	0.128	0.538	0.231	0.077	0.410	0.231	0.487	0.538	0.308	0.513	0.538	
	TT	0.051	0.333	0.154	0.179	0.487	0.103	0.256	0.103	0.103	0.103	0.103	0.051	
	CT	0.308	0.538	0.308	0.590	0.436	0.487	0.513	0.410	0.359	0.590	0.385	0.410	0.444
	He	0.326	0.479	0.426	0.499	0.416	0.453	0.500	0.426	0.405	0.479	0.416	0.381	
	Exact p	0.661	0.536	0.141	0.340	1.000	0.730	1.000	1.000	0.462	0.185	0.683	1.000	
Ibo	**n=48**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	
	CC	0.479	0.083	0.375	0.313	0.208	0.333	0.250	0.646	0.333	0.313	0.438	0.542	
	TT	0.021	0.292	0.125	0.167	0.271	0.167	0.292	0.083	0.125	0.125	0.083	0.104	
	CT	0.500	0.625	0.500	0.521	0.521	0.500	0.458	0.271	0.542	0.563	0.479	0.354	0.486
	He	0.395	0.478	0.469	0.489	0.498	0.486	0.520	0.362	0.478	0.482	0.437	0.404	
	Exact p	0.133	0.068	0.742	0.760	1.000	1.000	0.012	0.002	0.551	0.362	0.740	0.458	
Maya	**n=52**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	
	CC	0.827	0.096	0.096		0.192	0.404	0.077	0.769	0.462	0.615	0.173	0.212	
	TT	0.038	0.365	0.519	0.692	0.269	0.135	0.365	0.038	0.077	0.038	0.135	0.269	
	CT	0.135	0.538	0.385	0.308	0.538	0.462	0.558	0.192	0.462	0.346	0.692	0.519	0.428
	He	0.189	0.464	0.411	0.260	0.497	0.464	0.458	0.233	0.426	0.334	0.499	0.498	
	Exact p	0.006	0.359	0.735	0.332	0.571	1.000	0.233	0.214	0.517	0.667	0.006	1.000	

As part of a collaboration we are typing global populations with our SNP panels. F_{st} values will be estimated to evaluate allele frequency variation between populations.
Exact p-values < 5% are highlighted in yellow.
Population samples were provided by Dr. Kenneth K. Kidd (Yale University School of Medicine)

Poster available for download from STRBase http://www.cstl.nist.gov/biotech/strbase/pub_pres/ValloneISFG2005.pdf