Semi-Automation of mtDNA Arrays: Results from 666 Population Samples and Comparisons.

Margaret C. Kline, Janette W. Redman, Peter M. Vallone, David L. Duewer, and John M. Butler

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

Email:
margaret.kline@nist.gov

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

Acknowledgments: Funding from the National Institute of Justice through the NIST Office of Law Enforcement Standards; LINEAR ARRAYs and supplies provided by Roche Molecular Systems, Alameda, CA

Abstract:

At the National Institute of Standards and Technology (NIST) we have used the LINEAR ARRAYS for mtDNA HVI/HVII region sequence typing (Roche Applied Sciences,Indianapolis, IN) to evaluate our population collection of self declared Hispanic (128), African American (252), and Caucasian (286) samples. The samples were previously evaluated with Short Tandem Repeat Loci (STRs) to assure no sample duplicates exist in the collection. The amplified products were quantified and sized using an Agilent 2100 Bioanalyzer (Agilent Technologies, Wilmington DE). Arrays were processed both manually (120 samples) and using a Tecan Profiblot instrument that applied all the reagents and was "walk away" after the amplified samples were loaded. Frequencies of the mitotypes within and between populations were calculated and are presented along with sequencing results and single nucleotide polymorphism (SNP) results from Caucasian samples exhibiting the most common mitotype.

Whole blood samples were purchased from Interstate Blood Bank Inc. (Memphis, TN). The ethnicities of the blood samples were self-declared. Bloods were extracted by a modified "saltout" procedure [Miller et al. (1988) Nucleic Acids Research, 16, p. 1215].

Portions of the genomic extracts were quantified by UV measurement then diluted to 1 ng/µL, and verified by Pico Green quantitation.

One microliter of each sample extract was amplified using the mtDNA HVI/HVII Primer mix and mtDNA PCR Reaction Mix supplied by Roche Applied Sciences,(Indianapolis, IN) as part of a beta-testing of the HVI/HVII mtDNA LINEAR ARRAYs assay project. Since we used 1 ng of genomic DNA instead of the recommended 5 pg, we reduced the number of cycles from the recommended 34 cycles to 28 cycles.

PCR products were quantified either with agarose yield gels stained with Ethidium bromide and imaged/analyzed with a FMBIO III plus,(MiraiBio Inc. Alameda, CA) or 1 µL of the product was loaded on an Agilent 2100 Bioanalyzer Lab Chip (Agilent Technologies, Wilmington, DE). Quantitation results were imported to an Excel spreadsheet, which calculated the appropriate quantity of the PCR product for the total ng target. Gel-based quantitation is recommended by Roche Applied Sciences (RAS) to ensure that an optimal amount of PCR product is added to the LINEAR ARRAYs. Quantities below the recommended target input may lead to uninterruptible data (too weak) and above may lead to cross-hybridization (especially IIB4 with IIB2).

Figure 1 is a representative yield gel (3% NuSieve® GTG® agarose, 1% SeaKemp® GTG agarose) with two different sizing/quantitation ladders. Imaged FMBIO I II plus with 532 nm Excitation and 598nm Emission filter at 60% PMT sensitivity. The FMBIO ImageAnalysis software was used to quantify the unknowns using Ladder I and R separately. A polynomial best fit line is used to define the regression. While ladders I and R are biased relative to one another, they are in close agreement when plotting the sample concentrations obtained using the different ladders (see Figure 3).

Figure 1. mtDNA Product Yield Gel

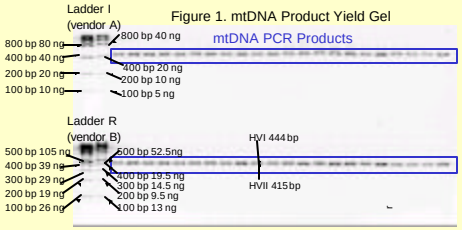


Figure 2 contains representative electropherograms from Agilent 2100 Bioanalyzer lab chip with the quantification information. Differences in the ratio of the HVI to the HVII PCR product peaks can be ascertained.

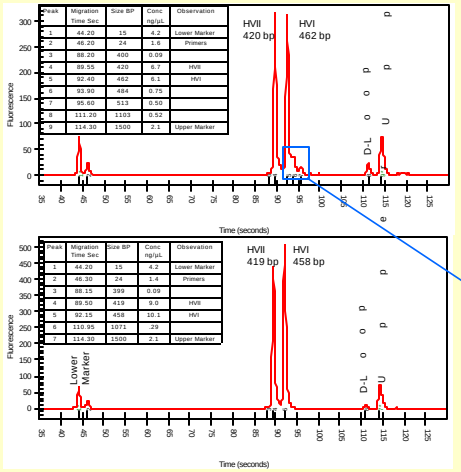


Figure 3 is the comparison of the quantitation results obtained from both gel based ladders (I, R) and the Lab chip(A). The graph represents the average of 3 to 4 measurements of the selected samples. There is good correlation between the gel based sizing standards (linear I:R). However the chip assay does not directly correlate as well with either Standard used in the gel method (I:A, R:A). Both the gel based and the chip based methods appear to be adequate for quantifying the PCR products. However, the variation in the methods indicated that a sensitivity study was required to assess the target quantity of PCR Products for LINEAR ARRAY hybridization.

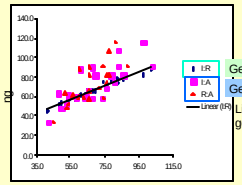


Figure 4a is the sensitivity study used to determine the optimum target (ng of PCR product) to be applied to the arrays based on our quantitation methods. Four samples were loaded at targets of 10, 25, 40, and 55 ng (results of two samples are shown). We selected 50 ng as our optimum target.

Fig. 4a

Figure 4b shows cross hybridization at the HVI C and HVIIB regions at 80 ng.

Fig. 4b

The Agilent electropherogram also gives an indication of the HVI "C-stretch" by the presence of extra peaks. Butler et. al. 1998 Electrophoresis 19, 119-124.

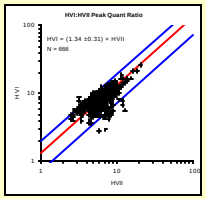


Figure 5. Represents the peak ratio of the HVI/HVII. The average ratio was 1.34 with a standard deviation of 0.31. The figure includes the observed average HVI/HVII ratio (red line) and the 95% confidence interval in blue (2σ).

Figure 6. Tecan ProfiBlot Instrument

The arrays were processed using a Tecan Profiblot DNA probe processor (Tecan®, RTP, NC). This robot was initially used to process the PM/DQA1 reverse dot blot typing strips. The ProfiBlot protocol is listed in Table 1. All reagents required for processing are loaded on the robot. The wash solution is placed in a heated/stirred heat block to maintain the required temperature. After the arrays are added to the tray the instrument dispenses warm wash and pauses for the PCR Products to be added to the appropriate well. The arrays are processed by the instrument through the final water wash. After the arrays were processed, a digital image was taken using a Syngene darkroom (Syngene, Frederick, MD), then mtDNA types are assigned by visual inspection.

Table 1. Profiblot Protocol

Step	File	Channel	Time	Temp	Color	Solution
1	Temp			55 °C		
2	Disp	1		55 °C	Brown	Wash
3	Pause			55 °C		
4	Inc		15 min	55 °C		
5	Asp			55 °C		
6	Disp	1		55 °C	Brown	Wash
7	Asp			55 °C		
8	Disp	3		55 °C	Orange	SA-HRP Conjugate
9	Inc		5 min	55 °C		
10	Asp			55 °C		
11	Disp	2		55 °C	Red	Wash
12	Asp			55 °C		
13	Disp	2		55 °C	Red	Wash
14	Inc		12 min	55 °C		
15	Asp			55 °C		
16	Cool			25 °C		
17	Disp	2		25 °C	Red	Wash
18	Asp			25 °C		
19	Disp	5		25 °C	Green	Citrate
20	Inc		5 min	25 °C		
21	Asp			25 °C		
22	Disp	6		25 °C	Blue	Color Dev
23	Inc		15 min	25 °C		
24	Asp			25 °C		
25	Disp	4		25 °C	Yellow	DI Water
26	Asp			25 °C		
27	Disp	4		25 °C	Yellow	DI Water
28	Inc		5 min	25 °C		
29	Asp			25 °C		
30	Disp	4		25 °C	Yellow	DI Water
31	End			25 °C		

Figure 7. Profiblot Processed mtDNA LINEAR ARRAYs with Typing Results

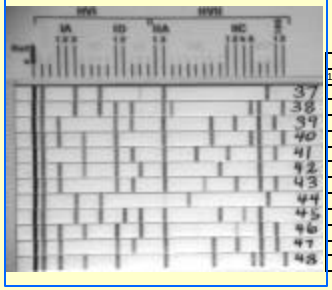


Table 2. Probe/Typings for the mtDNA LINEAR ARRAYs

HVI Probe Designation	Sequence Variation Detected	HVII Probe Designation	Sequence Variation Detected
	16093		73
16093 1	T T T	IIA1	T A T
16093 2	• C •	IIA2	• G •
	16126 16129		146 150 152
IA1	G T A C G G	IIB1	C T C A T C C T A
IA2	• C • • •	IIB2	• C • • • • • • •
IA3	• • • • A •	IIB3	• • • • • • • C •
	16304 16309 16311	IIB4	• C • • • • • C •
IC1	G T A C A T A G T A	IIB5	• • • • • T • • •
IC2	• C • • • • • • •	IIB6	• • • • • T • C •
IC3	• • • • • • • • C •	IIB7	• C • • • T • C •
IC4	• • • • • G • • •		189 195 198 200
IC w2/w3	• C • • • • • C •	IIC1	A A C A T A C T T A C T A A
	16362	IIC2	• • • • • • • C • • • • • •
ID1	G T C	IIC3	• • • • • • • • C • • T • • •
ID2	• C •	IIC4	• G • • • • • • • • • • G •
	16270 16278		247
IE1	A C T A G G A T A C C	IID1	T G A
IE2	• • • • • • • T • •	IID2	• A •
IE3	• T • • • • • • • •		189
		189 1	A A C
		189 2	• G •

Table information courtesy of Cassandra Calloway, Roche Molecular Systems, Alameda, CA

Table 3. Typing frequencies of the 666 population samples

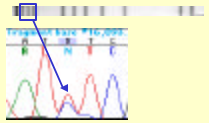
#*	Freq	% Types	% People
1	185	65.6	27.8
2	46	16.3	13.8
3	18	6.4	8.1
4	4	1.4	2.4
5	3	1.1	2.3
6	4	1.4	3.6
7	1	0.4	1.1
8	9	3.2	10.8
9	2	0.7	2.7
10	4	1.4	6.0
11	1	0.4	1.7
12	1	0.4	1.8
18	1	0.4	2.7
23	1	0.4	3.5
28	1	0.4	4.2
51	1	0.4	7.7

*The number of times a specific "type" is seen.

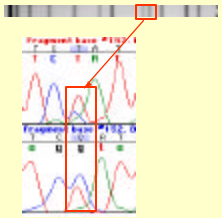
282 different types were seen with these population samples. Of these types 185 were unique(seen once in our population). These unique samples represent 65.6% of the types seen and 27.8% of the people. Several groupings of samples seem to represent specific ethnic groups.

1101126211, 4AA ; 1111112111, 4C
1211121101, 2C,2H;1221120111,4C
1131111111, 5C;1311111111,4C,1H
1331123112, 5AA
1111125111, 2C,2H,2AA; 1111220211, 4AA,1H,1C
1111227410, 6AA; 1131126111, 4C,2H
1112111111, 6C,1AA
1101125410, 8AA; 1111325111, 5C,2AA,1H; 1211120310, 8C;
1211123211, 7C,1H; 1211126211,8C; 1221121111, 7C,1AA
1231023020, 6AA,2H; 1330220020, 7AA,1H; 1101125211, 7AA,1H
1111221111, 8C, 1AA
1131121111, 5C, 3H, 1AA
1121111111, 9C; 1AA; 1141220211, 10AA
1211121111, 7C, 3H; 1231023220, 8AA,2 H
1111121101, 11H
1111113111,11C, 1H
1111121111, 11C, 7H
1102120111, 21H,1AA,1C
1141224211, 27AA, 1H
1111111111, 47C, 4H

Heteroplasmy at the various probe sites can be detected by the LINEAR ARRAYs as seen in the picture of the array with verifying sequencing data.
16093 1,2



HVIIB w5,6



The occurrence of "0" or "blank" type in a probe category indicates that the product or sample did not hybridize to any probe within the region due to highly destabilizing mismatch under the probe. 52% of the population had at least one "0" call; 9.6% of all calls were "0".

Table 4. Number of "0" calls

Locus	Number	%
16093	23	3.5
HVIA	33	5.0
HVIC	76	11.4
HVID	33	5.0
HVIE	60	9.0
HVIIA	3	0.5
HVIIB	96	14.4
HVIIC	122	18.3
HVIID	42	6.3
189	152	22.8

Typing Results 50 ng PCR Product

HVI					HVII				
16093	A	C	D	E	A	B	C	D	189
1	3	3	0	w2	2	0	0	2	0
1	3	3	1	1	2	1	5	1	2
1	1	0	1	1	2	6	2	1	1
1	1	1	1	0	2	6	5	1	2
1	0	1	1	1	2	4	1	1	1
1	1	1	1	2	2	7	0	1	0
1	1	0	1	1	2	5	2	1	1
1	3	3	0	0	w2	3	0	2	0
1	3	3	2	2	2	7	4	1	0
1	1	1	1	2	2	0	w2	1	1
1	1	0	1	1	2	6	2	1	1
1	1	1	1	0	2	6	5	1	2

mtDNA SNP Results

Typing a subset of 51 samples with an 11 plex SNP assay

	3010	4793	10211	5004	7028	7202	16519	12858	4580	477	14470
rCRS	G	A	C	T	C	A	T	C	G	T	T
1	A	A	C	T	C	A	C	T	G	C	T
1	G	A	C	C	C	A	T	C	G	T	T
1	G	A	C	T	C	A	C	C	G	T	A
1	G	A	C	T	T	A	T	C	G	T	T
2	A	A	C	T	C	A	C	C	G	C	T
2	G	A	C	T	T	A	C	C	A	T	T
3	G	A	C	T	T	A	T	C	G	A	T
4	G	A	C	T	T	A	T	C	A	T	T
4	G	G	C	T	C	A	T	C	G	T	T
5	G	A	C	T	C	A	C	C	G	T	T
12	G	A	C	T	C	A	C	C	G	T	T
15	A	A	C	T	C	A	C	C	G	T	T
51											

47 Caucasians
4 Hispanics

The 11 SNP sites are located throughout the mtGenome and are useful for separating the most common Caucasian HV1/HV2 mitotype

Fifty-one samples were determined to be mitotype 1111111111 by LINEAR ARRAYs. After SNP typing the 51 samples could be sub-typed to 12 different SNP types with four samples being unique.