Forensic Performance of Insertion-Deletion Marker Systems

Manuel Fondevila[1,3], Rui Pereira[1,2], Leonor Gusmão[2], Christopher Phillips[1], Maria Victoria Lareu[1], Angel Carracedo[1], John M. Butler[3], and Peter M. Vallone[3]

[1]Instituto de Ciencias Forenses Luis Concheiro, University of Santiago de Compostela, Spain
[2]IPATIMUP – Institute of Molecular Pathology and Immunology of the University of Porto, Portugal
[3]U.S. National Institute of Standards and Technology, Biochemical Science Division, Gaithersburg, MD 20899-8314

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

Instituto de Ciencias Forenses Luis Concheiro. Universidade de Santiago de Compostela. Spain.

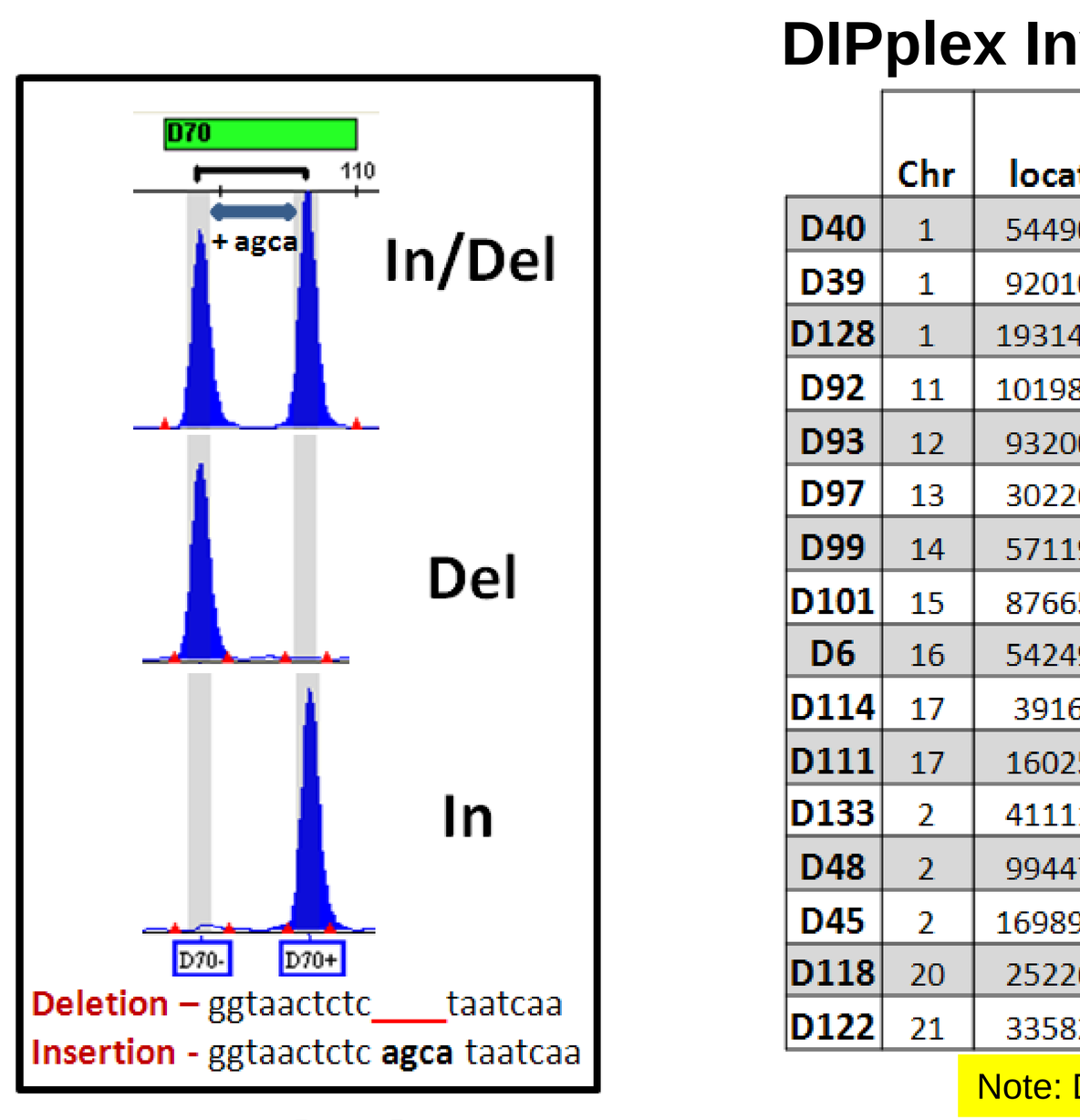
Institute of Molecular Pathology and Immunology of the University of Porto, Portugal.

Email: mafondevila@hotmail.com

P-024

An introduction to InDels

- InDels (insertion-deletion) are a type of short length polymorphism, consisting of the presence or absence of a short (typically 1-50 bp) sequence.
- Closely related to SNPs, as another class of short biallelic polymorphisms. They share most of SNP properties useful for forensic identification applications:
  - Low mutation rate
  - Short amplicon PCR to increase survival in degraded samples
  - High multiplexing capacity
- Relatively common throughout the genome, their total number has been estimated to be nearly 2,000,000. Up to 15% of known polymorphisms in the human genome are InDels [1].
- As length polymorphisms, InDels can be typed with a simple direct PCR-to-CE genotyping strategy, using a single multiplexed PCR with dyed-linked primers immediately followed by capillary electrophoresis.



Since InDel polymorphisms have been found to be relatively common in the human genome [1-2] selecting a high enough number of these biallelic polymorphisms that are selectively neutral, unlinked, and with the highest heterozygosity possible in most populations while keeping short amplicon PCR design, can be considered relatively straightforward.

Regardless of the limited individual informativeness of biallelic markers [3], it is known that the analysis of an affordable number of such polymorphisms may prove applicable [3-4]. In order to assess the informativeness of InDels and their applicability for the forensic community, we have conducted comprehensive genotyping of several U.S. representative population groups with two InDel assays typing in single multiplexes each: 30 and 38 InDel markers, herein: DIPplex and 38plex.

Materials & Methods

The two InDel genotyping assays studied comprised: Qiagen Investigator DIPplex commercial kit and HID InDel 38plex developed by R. Pereira et al. [6] following the protocols and conditions listed below.

We used the guidelines included on the DIPplex investigator kit for PCR and capillary electrophoresis [5]. 1 ng of DNA was used for the PCR reaction. In order to achieve best profile quality with the ABI Prism 3130xl capillary sequencer, we diluted the DIPplex PCR samples to 1:25 before mixing 1 μL of the dilution to the electrophoresis loading mix. POP-4 was used as the electrophoresis polymer in a 36 cm capillary array. A set of bins and panels was developed for the DIPplex results analysis, which were performed using GeneMapper ID-X.

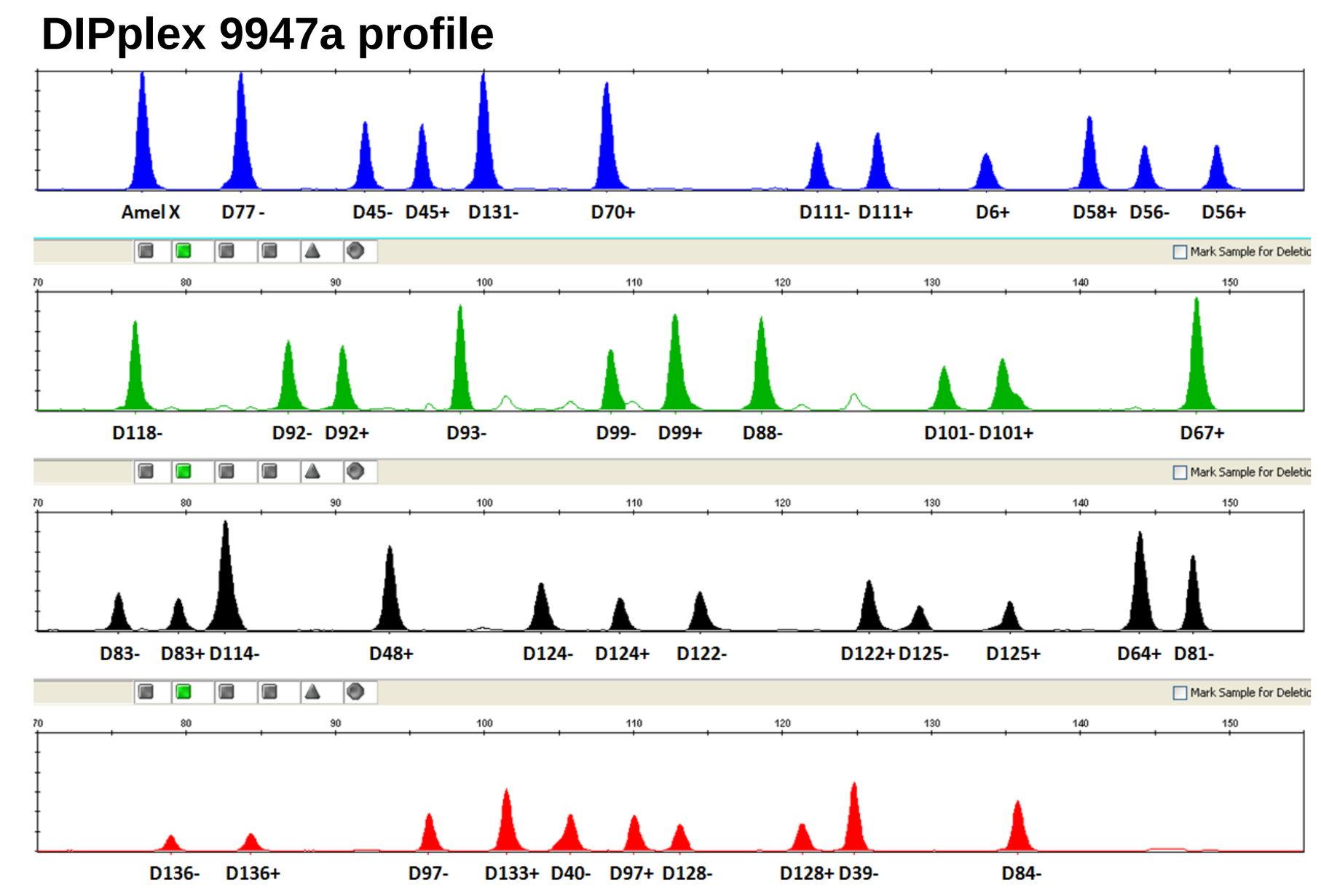
We used, for the HID-38plex, the published guidelines for PCR and CE [6]. 0.60 ng of DNA were used for the PCR reaction. In order to achieve best profile quality with the ABI Prism 3130xl capillary sequencer, we diluted the PCR samples 1:10 before mixing 1 μL of the dilution to the electrophoresis loading mix. POP-4 was used as the electrophoresis polymer in a 36 cm capillary array. A set of bins and panels was made for 38plex results analysis, which were performed using GeneMapper ID-X.

DIPplex Investigator Markers

	Chr	location	Public database ID	Profile position		Chr	location	Public database ID	Profile position
D40	1	54490780	rS2307956	Red 4	D125	22	24080816	rS16388	Yellow 6
D39	1	92010480	rS17878444	Red 6	D124	22	34031900	rS6481	Yellow 4
D128	1	193144970	rS2307924	Red 5	D136	22	35739831	rS16363	Red 1
D92	11	101984628	rS17174476	Green 2	D56	4	108109222	rS2308292	Blue 9
D93	12	93200037	rS2307570	Green 3	D64	5	66250256	rS1610935	Yellow 7
D97	13	30226384	rS17238892	Red 3	D58	5	76780823	rS1610937	Blue 8
D99	14	57119834	rS2308163	Green 4	D67	5	155594834	rS1305056	Green 7
D101	15	87665320	rS2307433	Green 6	D70	6	97564842	rS2307652	Blue 5
D6	16	54249331	rS1610905	Blue 7	D81	7	94885086	rS17879936	Yellow 8
D114	17	3916882	rS2307581	Yellow 2	D77	7	110727223	rS1611048	Blue 2
D111	17	16025713	rS1305047	Blue 6	D131	7	154035495	rS1611001	Blue 4
D133	2	41111225	rS2067235	Red 2	D83	8	19134059	rS2308072	Yellow 1
D48	2	99447993	rS28369942	Yellow 3	D84	8	120016982	rS3081400	Red 7
D45	2	169898519	rS2307959	Blue 3	D88	9	98037732	rS8190570	Green 5
D118	20	25226470	rS16438	Green 1	Amel	X	p22.1-22.3	M55418	Blue 1
D122	21	33582626	rS8177524	Yellow 5	Amel	Y	p11.2	M55419	Blue 1

Note: D125, D124 and D136 are within 12 Mb from one another on Chr 22

DIPplex 9947a profile

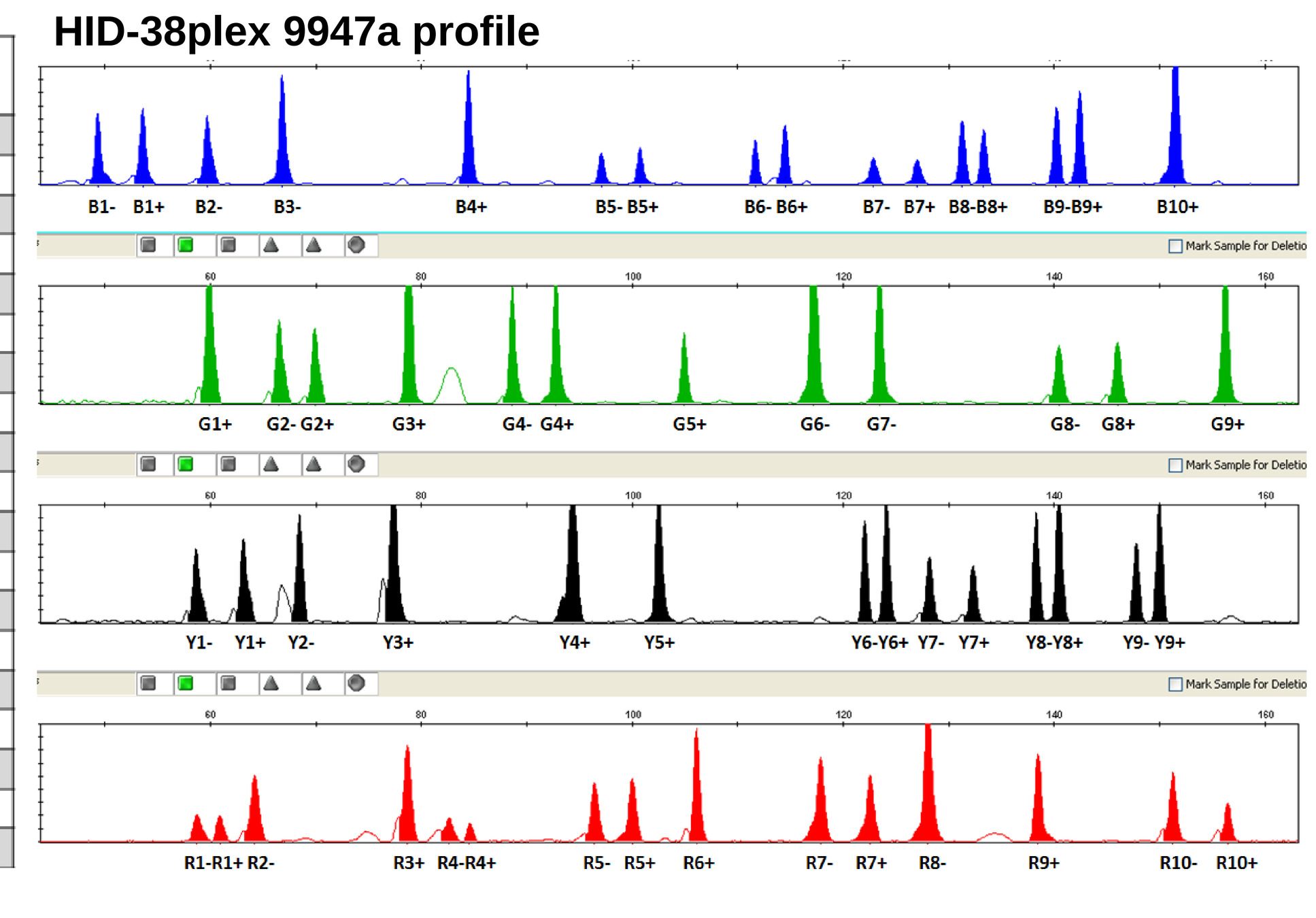


Marker	type
Amel	XX
D77	De
D45	De/In
D131	De
D70	In
D111	De/In
D6	In
D58	In
D56	De/In
D118	De
D92	De/In
D93	De
D99	De/In
D88	De
D101	De/In
D67	In
D83	De/In
D114	De
D48	In
D124	De/In
D122	De/In
D125	De/In
D64	In
D81	De
D136	De/In
D133	In
D97	De/In
D40	De
D128	De/In
D39	De
D84	De

HID-38plex Markers

	Chr	location	Public database ID	Profile position		Chr	location	Public database ID	Profile position
B7	1	161077452	rs3047269	Blue 7	B9	20	11643625	rs33917182	Blue 9
Y5	1	245878706	rs2307579	Yellow 5	B1	20	30165066	rs34541393	Blue 1
B6	10	6027167	rs140809	Blue 6	R10	21	14556736	rs35605984	red 10
R3	10	54112392	rs1160886	Red 3	Y2	21	30294208	rs10629077	Yellow 2
Y3	11	258180	rs10688868	Yellow 3	B5	22	25120901	rs2307700	Blue 5
R6	11	30134266	rs34811743	Red 6	G5	3	8591709	rs2308242	Green 5
B8	11	125794082	rs33972805	Blue 8	R4	4	119404855	rs2308026	Red 4
Y8	12	14801263	rs1610919	Yellow 8	R5	5	5178112	rs2307526	Red 5
G2	12	113772931	rs2067238	Green 2	G7	5	65414216	rs1160956	Green 7
R9	13	43778155	rs2308171	Red 9	G1	5	171020572	rs1610871	Green 1
R7	14	28106508	rs2308189	Red 7	G4	6	47929222	rs2307710	Green 4
Y6	15	51268809	rs2308020	Yellow 6	Y9	6	117600251	rs2307839	Yellow 9
Y4	16	83139788	rs2067208	Yellow 4	R1	6	149655891	rs2308137	Red 1
Y1	17	10076666	rs3051300	Yellow 1	G9	7	83121850	rs2307978	Green 9
Y7	18	21507205	rs3080855	Yellow 7	B4	8	76681235	rs35769550	Blue 4
G8	18	34677042	rs34511541	Green 8	R8	8	138489776	rs589447	Red 8
R2	19	1353662	rs36040336	Red 2	B10	9	38396788	rs16402	Blue 10
B3	19	48896180	rs2307689	Blue 3	G3	9	70504241	rs2067294	Green 3
B2	2	234681130	rs16624	Blue 2	G6	9	104626014	rs2307580	Green 6

HID-38plex 9947a profile



Marker	type
B1	De/In
B2	De
B3	De
B4	In
B5	De/In
B6	De/In
B7	De/In
B8	De/In
B9	De/In
B10	In
G1	In
G2	De/In
G3	In
G4	De/In
G5	In
G6	De
G7	De
G8	De/In
G9	In
Y1	De/In
Y2	De
Y3	In
Y4	In
Y5	In
Y6	De/In
Y7	De/In
Y8	De/In
Y9	De/In
R1	De/In
R2	De
R3	In
R4	De/In
R5	De/In
R6	De
R7	De/In
R8	De
R9	In
R10	De/In

InDels less than 10 Mb from a core STR locus

	CHR	STR	DIPplex	Physical Distance
DIPplex	5	CSF1PO	rS1305056	6,158,834
	6	SE33	rS2307652	8,521,842
	8	D8S1179	rS3081400	5,959,018
	15	PentaE	rS2307433	7,509,680
	22	D22S1045	rS6481	1,747,100
	22	D22S1045	rS16363	39,169

	CHR	STR	DIPplex	Physical Distance
HID-38plex	7	D7S820	rs2307978	311,150
	11	TH01	rs10688868	1,890,820
	12	D12S391	rs1610919	2,352,263
	12	vWA	rs1610919	8,838,263
	16	D16S539	rs2067208	1,804,212
	21	D21S11	rs35605984	4,919,264

Left: When contemplating the possibility of combining the information contained in these InDel markers systems with each other or with core STR loci, we should keep in mind that the proximity between some of these markers could lead to a linkage disequilibrium state. The table on the left identifies 6 loci from each InDel assay that are less than 10 Mb from a core STR locus.

Results

Allele frequency analysis

We performed population allele frequency analysis with both InDel multiplexes typing the NIST collection of population samples consisting of a total of 712 male DNA samples plus two female samples.

Comprising:

- 262 African Americans
- 260 U.S. Caucasians
- 140 U.S. Hispanics
- 50 U.S. Asians

The results of the frequency analysis are shown in the tables on the right.

Several markers in each of the assays have shown a deviation from HWE (marked in red), however the number of such markers falls into the expected error of the calculation.

There are a number of systems (marked in yellow), that have shown a reduced heterozygosity in certain populations. However, all systems remain polymorphic in all populations. Most of these reduced heterozygosity systems are still contained within the limit 0.2 frequency [4] values and remain informative enough in all populations.

The allele frequencies displayed by these markers in U.S. populations allow us to assess the power of the assays, with Random Match Probability (RMP) values nearing those of the combined current STR markers (as shown in the table below).

	U.S. Cauc	U.S. Asian	U.S. Hisp	Af-Am
Mean DIPplex RMP	1.86E-13	4.67E-11	4.88E-13	5.88E-12
Mean HID-38plex RMP	3.67E-15	5.11E-14	1.47E-15	4.74E-15
Mean Combined RMP*	6.79E-28	2.43E-24	7.20E-28	2.54E-26

* Assuming all markers are completely independent

DIPplex U.S. population data

	D111				D131				D45				D56				D58				D6				D70				D77				D101				D118			
	Cauc	Asian	Hisp	Af-Am	Cauc	Asian	Hisp	Af-Am	Cauc	Asian	Hisp	Af-Am	Cauc	Asian	Hisp	Af-Am	Cauc	Asian	Hisp	Af-Am	Cauc	Asian	Hisp	Af-Am	Cauc	Asian	Hisp	Af-Am	Cauc	Asian	Hisp	Af-Am	Cauc	Asian	Hisp	Af-Am	Cauc	Asian	Hisp	Af-Am
Het (expected)	0.496	0.182	0.501	0.47	0.496	0.485	0.502	0.36	0.499	0.485	0.482	0.487	0.449	0.503	0.443	0.498	0.501	0.485	0.443	0.279	0.501	0.5	0.5	0.482	0.501	0.424	0.459	0.225	0.5	0.505	0.494	0.472	0.494	0.498	0.501	0.358	0.493	0.149	0.474	0.416
Het (observed)	0.542	0.2	0.45	0.488	0.462	0.52	0.507	0.346	0.515	0.4	0.543	0.508	0.443	0.58	0.4	0.473	0.462	0.48	0.4	0.288	0.427	0.5	0.414	0.462	0.55	0.44	0.464	0.242	0.477	0.5	0.493	0.465	0.454	0.56	0.507	0.35	0.473	0.12	0.536	0.419
1 (deletion)	0.450	0.900	0.518	0.625	0.448	0.600	0.496	0.235	0.471	0.400	0.400	0.585	0.340	0.530	0.329	0.46	0.502	0.600	0.671	0.833	0.511	0.550	0.529	0.596	0.496	0.300	0.354	0.129	0.483	0.510	0.561	0.621	0.559	0.560	0.475	0.233	0.565	0.080	0.618	0.706
2 (insertion)	0.550	0.100	0.482	0.375	0.552	0.400	0.504	0.765	0.529	0.600	0.600	0.415	0.660	0.470	0.671	0.54	0.498	0.400	0.329	0.167	0.489	0.450	0.471	0.404	0.504	0.700	0.646	0.871	0.517	0.490	0.439	0.379	0.441	0.440	0.525	0.767	0.435	0.920	0.382	0.294
3 (variant)																																								
P-value	0.14	1	0.242	0.596	0.319	0.768	1	0.605	0.621	0.248	0.158	0.522	0.891	0.392	0.256	0.454	0.218	1	0.255	0.823	0.019	1	0.045	0.521	0.138	1	1	0.273	0.458	1	1	0.894	0.212	0.405	1	0.73	0.532	0.261	0.153	1

	D67				D88				D92				D93				D99				D128				D133				D136				D39				D40			
	Cauc	Asian	Hisp	Af-Am	Cauc	Asian	Hisp	Af-Am	Cauc	Asian	Hisp	Af-Am	Cauc	Asian	Hisp	Af-Am	Cauc	Asian	Hisp	Af-Am	Cauc	Asian	Hisp	Af-Am	Cauc	Asian	Hisp	Af-Am	Cauc	Asian	Hisp	Af-Am	Cauc	Asian	Hisp	Af-Am	Cauc	Asian	Hisp	Af-Am
Het (expected)	0.467	0.416	0.487	0.428	0.498	0.502	0.487	0.44	0.498	0.503	0.48	0.424	0.495	0.465	0.477	0.501	0.492	0.228	0.502	0.485	0.5	0.407	0.501	0.451	0.498	0.5	0.5	0.44	0.501	0.481	0.487	0.345	0.444	0.298	0.484	0.494	0.487	0.453	0.496	0.421
Het (observed)	0.489	0.42	0.529	0.427	0.515	0.56	0.529	0.45	0.462	0.54	0.564	0.438	0.523	0.32	0.507	0.415	0.42	0.26	0.514	0.438	0.458	0.44	0.5	0.477	0.489	0.5	0.5	0.442	0.519	0.5	0.471	0.35	0.427	0.32	0.443	0.465	0.492	0.6	0.45	0.385
1 (deletion)	0.370	0.290	0.414	0.31	0.460	0.540	0.414	0.325	0.540	0.530	0.604	0.696	0.445	0.360	0.389	0.508	0.431	0.130	0.457	0.394	0.519	0.720	0.514	0.342	0.462	0.550	0.529	0.675	0.504	0.390	0.414	0.221	0.668	0.820	0.593	0.44	0.582	0.340	0.554	0.7
2 (insertion)	0.630	0.710	0.586	0.69	0.540	0.460	0.586	0.675	0.460	0.470	0.396	0.304	0.555	0.640	0.611	0.492	0.569	0.870	0.539	0.591	0.481	0.280	0.486	0.658	0.538	0.450	0.471	0.325	0.496	0.610	0.586	0.779	0.332	0.180	0.407	0.56	0.418	0.660	0.446	0.3
3 (variant)																			0.004	0.015																				
P-value	0.508	1	0.378	1	0.62	0.568	0.385	0.779	0.259	0.775	0.051	0.658	0.381	0.034	0.481	0.007	0.024	0.581	0.663	0.177	0.174	0.729	1	0.413	0.804	1	1	1	0.62	1	0.73	1	0.579	1	0.381	0.377	0.9	0.027	0.302	0.185

	D84				D97				D114				D122				D124				D125				D48				D64				D81				D83			
	Cauc	Asian	Hisp	Af-Am	Cauc	Asian	Hisp	Af-Am	Cauc	Asian	Hisp	Af-Am	Cauc	Asian	Hisp	Af-Am	Cauc	Asian	Hisp	Af-Am	Cauc	Asian	Hisp	Af-Am	Cauc	Asian	Hisp	Af-Am	Cauc	Asian	Hisp	Af-Am	Cauc	Asian	Hisp	Af-Am	Cauc	Asian	Hisp	Af-Am
Het (expected)	0.496	0.432	0.494	0.409	0.498	0.368	0.497	0.485	0.494	0.358	0.501	0.349	0.493	0.335	0.39	0.464	0.464	0.489	0.495	0.345	0.493	0.46	0.47	0.401	0.5	0.398	0.407	0.406	0.497	0.243	0.413	0.349	0.489	0.198	0.489	0.5	0.5	0.453	0.487	0.476
Het (observed)	0.485	0.34	0.4	0.385	0.462	0.36	0.414	0.381	0.473	0.42	0.464	0.35	0.496	0.34	0.386	0.419	0.416	0.46	0.529	0.288	0.504	0.34	0.45	0.423	0.508	0.3	0.421	0.412	0.481	0.28	0.436	0.35	0.469	0.1	0.486	0.531	0.489	0.44	0.486	0.477
1 (deletion)	0.452	0.310	0.436	0.389	0.464	0.760	0.550	0.59	0.557	0.770	0.525	0.225	0.561	0.790	0.736	0.637	0.365	0.410	0.443	0.779	0.439	0.650	0.625	0.722	0.483	0.730	0.282	0.278	0.458	0.140	0.289	0.223	0.578	0.110	0.421	0.481	0.481	0.660	0.586	0.392
2 (insertion)	0.548	0.690	0.564	0.602	0.536	0.240	0.450	0.41	0.443	0.230	0.475	0.775	0.439	0.210	0.264	0.363	0.635	0.590	0.557	0.221	0.561	0.350	0.375	0.278	0.517	0.270	0.718	0.722	0.542	0.860	0.711	0.777	0.422	0.890	0.579	0.519	0.519	0.340	0.414	0.608
3 (variant)							0.009																																	
P-value	0.711	0.184	0.026	0.066	0.266	1	0.06	8E-04	0.53	0.417	0.405	1	1	1	1	0.142	0.108	0.772	0.492	0.012	0.801	0.113	0.717	0.439	0.902	0.144	0.834	0.879	0.618	0.573	0.542	1	0.526	0.007	1	0.384	0.712	1	1	1

HID-38plex U.S. population data

	B1				B2				B3				B4				B5				B6				B7				B8				B9				B10			
	Cauc	Asian	Hisp	Af-Am	Cauc	Asian	Hisp	Af-Am	Cauc	Asian	Hisp	Af-Am	Cauc	Asian	Hisp	Af-Am	Cauc	Asian	Hisp	Af-Am	Cauc	Asian	Hisp	Af-Am	Cauc	Asian	Hisp	Af-Am	Cauc	Asian	Hisp	Af-Am	Cauc	Asian	Hisp	Af-Am	Cauc	Asian	Hisp	Af-Am
Het (expected)	0.491	0.416	0.474	0.477	0.311	0.217	0.497	0.458	0.423	0.243	0.481	0.496	0.466	0.481	0.499	0.274	0.495	0.44	0.489	0.407	0.474	0.46	0.373	0.494	0.491	0.495	0.5	0.465	0.5	0.258	0.501	0.471	0.501	0.495	0.435	0.498	0.412	0.481	0.448	0.424
Het (observed)	0.441	0.46	0.507	0.485	0.292	0.204	0.482	0.416	0.421	0.24	0.511	0.527	0.498	0.5	0.496	0.233	0.487	0.32	0.482	0.388	0.456	0.58	0.364	0.55	0.502	0.38	0.413	0.477	0.471	0.3	0.5	0.508	0.513	0.5	0.393	0.461	0.425	0.46	0.514	0.407
1 (deletion)	0.431	0.71	0.382	0.39	0.808	0.122	0.55	0.353	0.303	0.14	0.399	0.45	0.368	0.61	0.464	0.163	0.554	0.32	0.421	0.283	0.385	0.35	0.246	0.442	0.431	0.57	0.536	0.634	0.473	0.15	0.479	0.378	0.498	0.43	0.682	0.537	0.289	0.39	0.336	0.304
2 (insertion)	0.569	0.29	0.618	0.61	0.192	0.878	0.45	0.647	0.697	0.86	0.601	0.55	0.632	0.39	0.536	0.837	0.446	0.68	0.579	0.717	0.615	0.65	0.754	0.558	0.569	0.43	0.464	0.366	0.527	0.85	0.521	0.622	0.502	0.57	0.318	0.463	0.711	0.61	0.664	0.696
P-value	0.1	0.512	0.475	0.796	0.324	0.535	0.734	0.167	1	1	0.488	0.316	0.288	1	1	0.023	0.803	0.1	0.865	0.446	0.599	0.071	0.821	0.078	0.802	0.148	0.059	0.788	0.384	0.577	1	0.236	0.711	1	0.248	0.263	0.653	0.774	0.088	0.558

	G1				G2				G3				G4				G5				G6				G7				G8				G9				Y1			
	Cauc	Asian	Hisp	Af-Am	Cauc	Asian	Hisp	Af-Am	Cauc	Asian	Hisp	Af-Am	Cauc	Asian	Hisp	Af-Am	Cauc	Asian	Hisp	Af-Am	Cauc	Asian	Hisp	Af-Am	Cauc	Asian	Hisp	Af-Am	Cauc	Asian	Hisp	Af-Am	Cauc	Asian	Hisp	Af-Am	Cauc	Asian	Hisp	Af-Am
Het (expected)	0.498	0.492	0.502	0.499	0.435	0.46	0.491	0.365	0.453	0.272	0.433	0.276	0.464	0.421	0.39	0.486	0.32	0.379	0.338	0.447	0.501	0.424	0.501	0.41	0.221	0.503	0.435	0.501	0.48	0.481	0.497	0.48	0.322	0.503	0.383	0.47	0.494	0.407	0.46	0.383
Het (observed)	0.49	0.48	0.493	0.407	0.462	0.46	0.453	0.346	0.46	0.28	0.4	0.26	0.473	0.388	0.343	0.523	0.322	0.3	0.329	0.485	0.494	0.32	0.475	0.395	0.238	0.46	0.393	0.477	0.456	0.46	0.496	0.48	0.318	0.367	0.4	0.469	0.456	0.36	0.453	0.384
1 (deletion)	0.536	0.42	0.489	0.529	0.678	0.35	0.568	0.238	0.345	0.16	0.314	0.165	0.362	0.296	0.264	0.413	0.199	0.25	0.214	0.335	0.5	0.7	0.475	0.287	0.874	0.53	0.682	0.506	0.398	0.39	0.546	0.391	0.201	0.469	0.257	0.376	0.443	0.28	0.357	0.258
2 (insertion)	0.464	0.58	0.511	0.471	0.322	0.65	0.432	0.762	0.655	0.84	0.686	0.835	0.638	0.704	0.736	0.587	0.801	0.75	0.786	0.665	0.5	0.3	0.525	0.713	0.126	0.47	0.318	0.494	0.602	0.61	0.454	0.609	0.799	0.531	0.743	0.624	0.557	0.72	0.643	0.742
P-value	0.805	1	0.867	0.004	0.391	1	0.388	0.494	0.891	1	0.433	0.366	0.789	0.731	0.19	0.249	1	0.149	0.801	0.207	0.903	0.097	0.611	0.645	0.394	0.579	0.25	0.458	0.436	0.774	1	1	0.848	0.084	0.662	1	0.213	0.486	0.855	1

	Y2				Y3				Y4				Y5				Y6				Y7				Y8				Y9				R1				R2			
	Cauc	Asian	Hisp	Af-Am	Cauc	Asian	Hisp	Af-Am	Cauc	Asian	Hisp	Af-Am	Cauc	Asian	Hisp	Af-Am	Cauc	Asian	Hisp	Af-Am	Cauc	Asian	Hisp	Af-Am	Cauc	Asian	Hisp	Af-Am	Cauc	Asian	Hisp	Af-Am	Cauc	Asian	Hisp	Af-Am	Cauc	Asian	Hisp	Af-Am
Het (expected)	0.296	0.232	0.38	0.338	0.43	0.489	0.416	0.362	0.412	0.311	0.383	0.281	0.5	0.358	0.494	0.5	0.406	0.505	0.488	0.4	0.418	0.465	0.457	0.407	0.5	0.424	0.47	0.49	0.4	0.485	0.354	0.36	0.432	0.492	0.429	0.468	0.283	0.358	0.429	0.498
Het (observed)	0.291	0.224	0.379	0.306	0.446	0.58	0.414	0.395	0.425	0.3	0.429	0.283	0.487	0.3	0.429	0.453	0.456	0.38	0.507	0.403	0.425	0.48	0.457	0.38	0.448	0.36	0.464	0.496	0.375	0.52	0.3	0.283	0.419	0.44	0.46	0.376	0.278	0.38	0.417	0.492
1 (deletion)	0.18	0.133	0.254	0.215	0.31	0.41	0.293	0.236	0.289	0.19	0.257	0.169	0.481	0.23	0.436	0.517	0.718	0.51	0.582	0.725	0.301	0.36	0.35	0.283	0.523	0.3	0.625	0.574	0.28	0.4	0.229	0.234	0.316	0.42	0.307	0.627	0.832	0.77	0.689	0.463
2 (insertion)	0.82	0.867	0.746	0.785	0.69	0.59	0.707	0.764	0.711	0.81	0.743	0.831	0.519	0.77	0.564	0.483	0.282	0.49	0.418	0.275	0.699	0.64	0.65	0.717	0.477	0.7	0.375	0.426	0.72	0.6	0.771	0.766	0.684	0.58	0.693	0.373	0.168	0.23	0.311	0.537
P-value	0.834	1	1	0.139	0.566	0.243	1	0.166	0.654	1	0.187	1	0.709	0.254	0.125	0.138	0.049	0.094	0.73	1	0.881	1	1	0.285	0.106	0.321	1	0.899	0.351	0.768	0.09	0.001	0.666	0.562	0.43	0.002	0.825	1	0.842	0.901

	R3				R4				R5				R6				R7				R8				R9				R10			
	Cauc	Asian	Hisp	Af-Am	Cauc	Asian	Hisp	Af-Am	Cauc	Asian	Hisp	Af-Am	Cauc	Asian	Hisp	Af-Am	Cauc	Asian	Hisp	Af-Am	Cauc	Asian	Hisp	Af-Am	Cauc	Asian	Hisp	Af-Am	Cauc	Asian	Hisp	Af-Am
Het (expected)	0.472	0.495	0.461	0.425	0.494	0.258	0.465	0.198	0.46	0.465	0.407	0.471	0.44	0.368	0.44	0.472	0.485	0.407	0.502	0.5	0.442	0.437	0.434	0.43	0.367	0.059	0.327	0.491	0.496	0.495	0.502	0.493
Het (observed)	0.458	0.62	0.471	0.447	0.504	0.18	0.429	0.191	0.421	0.52	0.364	0.492	0.437	0.28	0.464	0.457	0.454	0.4	0.471	0.45	0.432	0.51	0.432	0.358	0.288	0.06	0.276	0.435	0.513	0.42	0.443	0.461
1 (deletion)	0.38	0.57	0.357	0.304	0.443	0.15	0.364	0.114	0.358	0.64	0.282	0.378	0.674	0.76	0.675	0.62	0.416	0.28	0.493	0.516	0.328	0.316	0.314	0.31	0.241	0.03	0.205	0.571	0.448	0.43	0.493	0.564
2 (insertion)	0.62	0.43	0.643	0.696	0.557	0.85	0.636	0.886	0.642	0.36	0.718	0.622	0.326	0.24	0.325	0.38	0.584	0.72	0.507	0.484	0.672	0.684	0.686	0.69	0.759	0.97	0.795	0.429	0.552	0.57	0.507	0.436
P-value	0.691	0.087	0.854	0.464	0.801	0.061	0.366	0.531	0.223	0.541	0.219	0.504	1	0.118	0.565	0.691	0.313	1	0.5	0.108	0.778	0.325	1	0.008	9E-04	1	0.107	0.072	0.618	0.384	0.18	0.313

Artificially degraded DNA assay

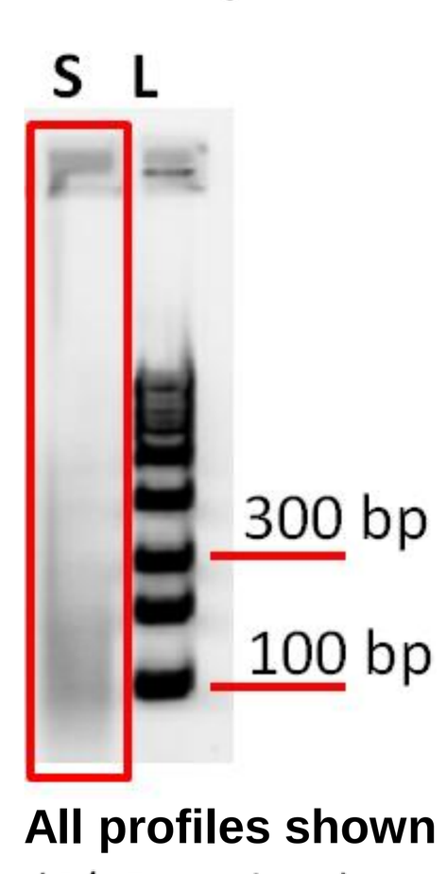
We conducted several experiments to examine degraded DNA samples in a controlled way [8]. Our objective was to compare the short-amplicon InDel typing reaction performance to established short amplicon STRs kits.

We selected the COVARIS DNA shearing system to perform the fragmentation since this approach enables precise control of the process in isothermal conditions. The system uses a process called Adaptive Focused Acoustics (AFA) that works by creating shock waves from a conical shaped transducer focused to converge on a small localized area [9].

Results from COVARIS DNA fragmentation

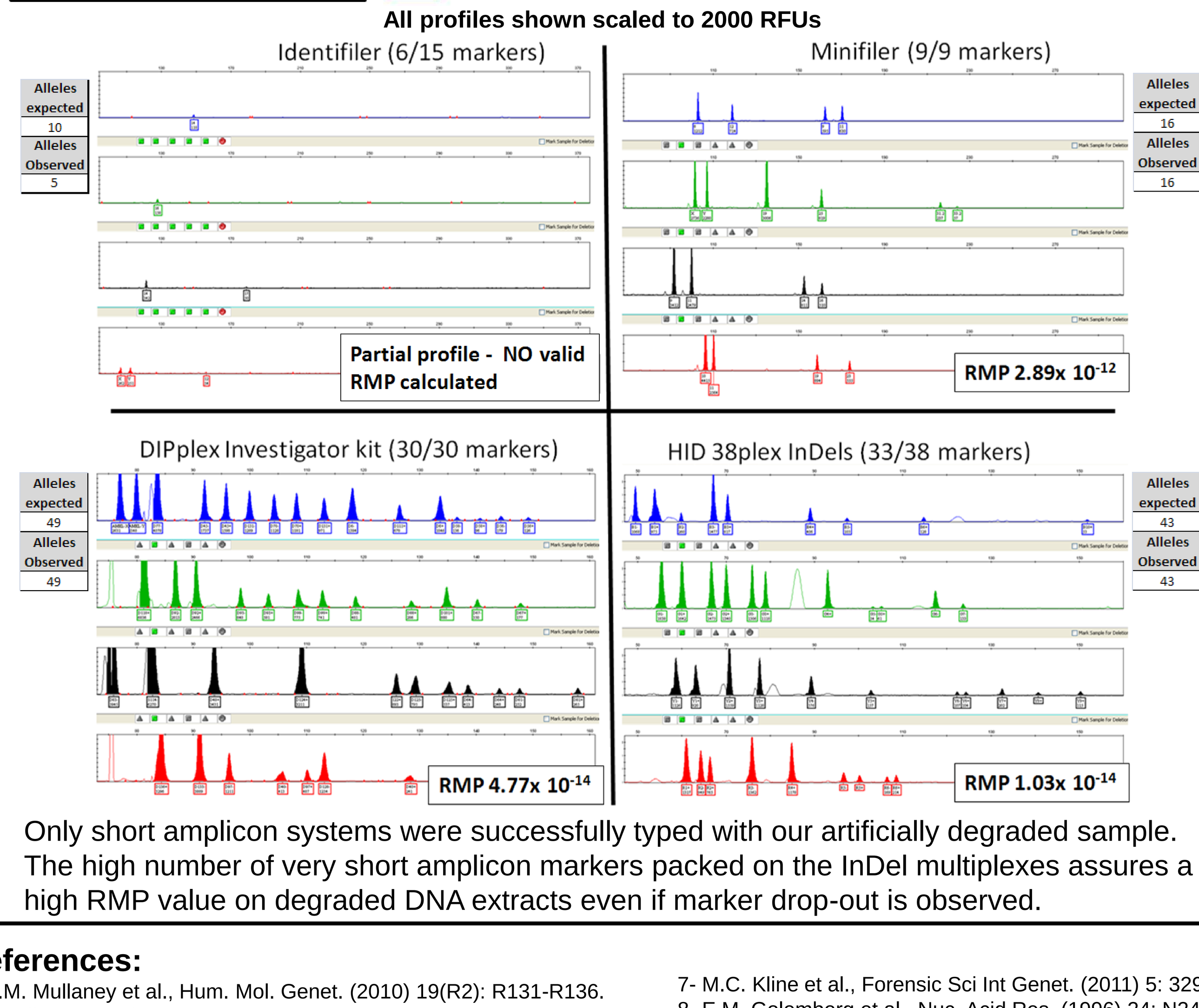
Final shearing protocol

Temperature: 5 ºC
Mode: Frequency sweeping
Duty Cycle: 10%
Intensity: 10%
Cycle/Burst: 1000
Time: 20 minutes
DNA: 50ng
Dilution volume: 100uL
Tube: glass walled- 100uL tube



We performed several trials with different assay conditions until we achieved the desired DNA fragmentation between 100 and 200 bp. A highly characterized DNA extract from NIST was used as the degradation target. The same degraded DNA aliquot (**S**) was used with all kits. All kit amplification and electrophoresis was performed following the manufacturer's guidelines for usual sample treatment.

All profiles shown scaled to 2000 RFUs

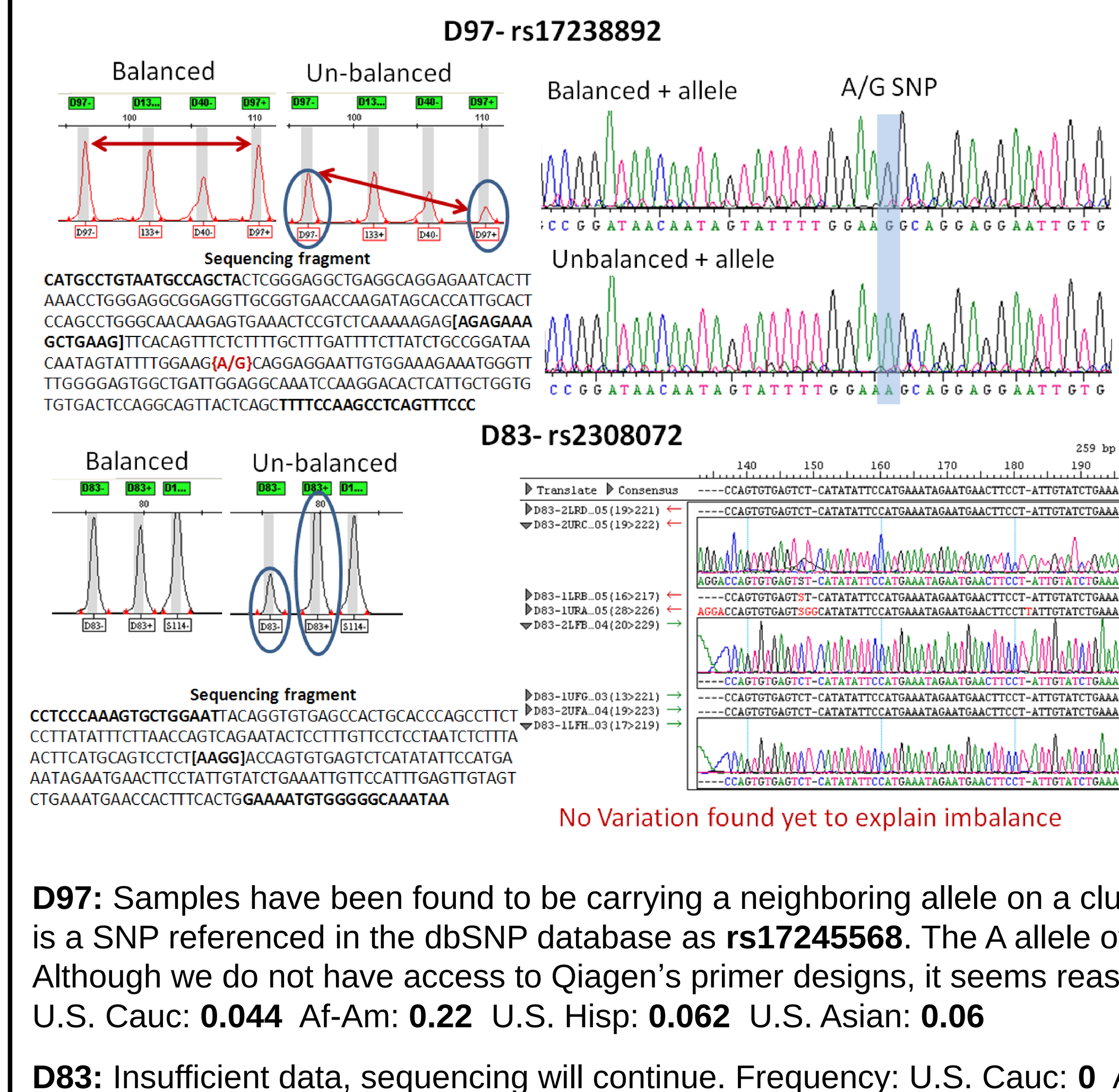


Only short amplicon systems were successfully typed with our artificially degraded sample. The high number of very short amplicon markers packed on the InDel multiplexes assures a high RMP value on degraded DNA extracts even if marker drop-out is observed.

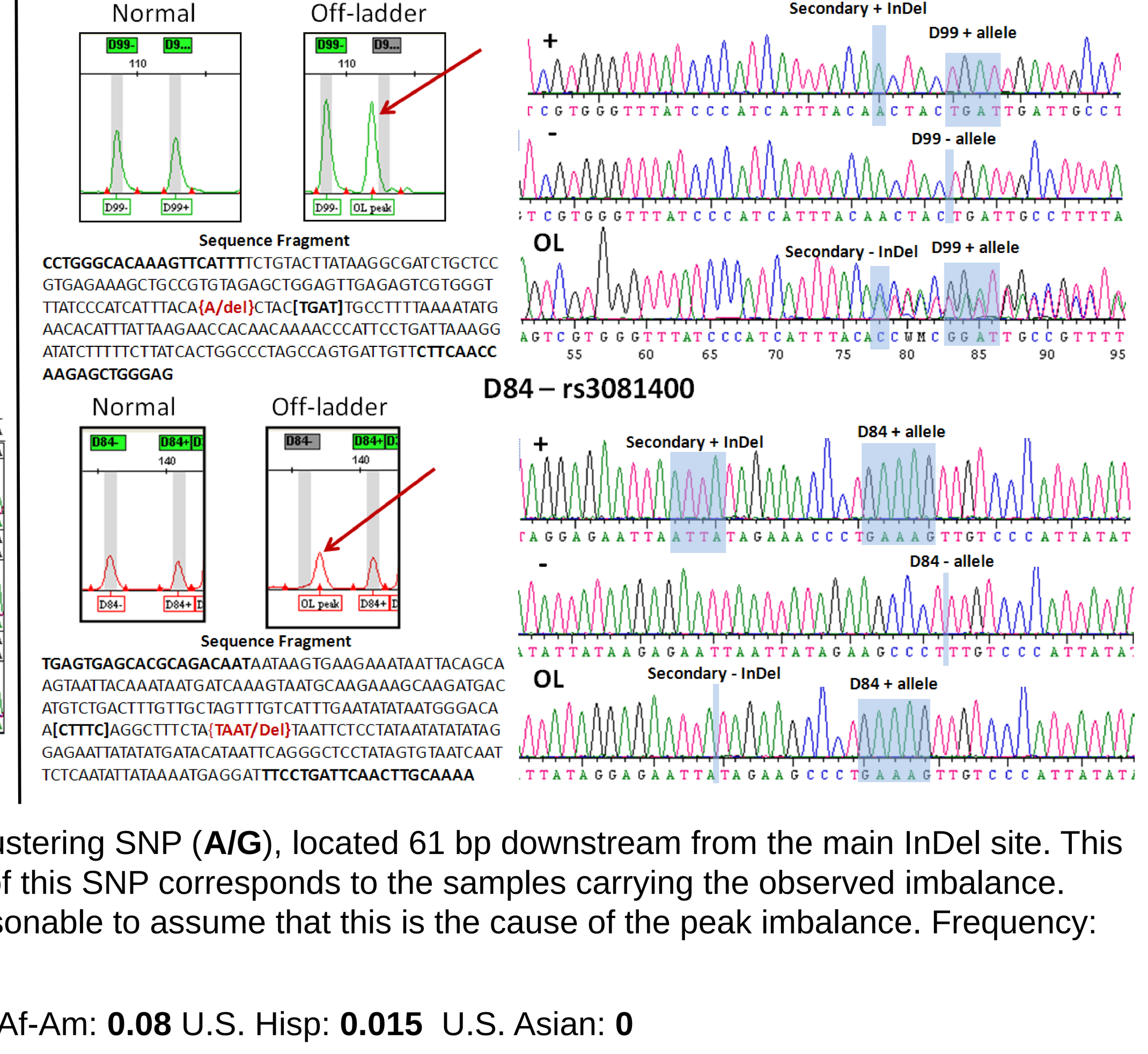
Sequencing of previously unreported variation

Sequencing was performed following the published guidelines of M.C. Kline et al. [7]

Two DIPplex markers (D97 and D83) showed a higher degree of allele imbalance in heterozygotes than expected. This suggested the presence of a SNP within the primer binding site potentially disrupting primer annealing in samples carrying the minor allele



A second observed feature was the presence of a third off-ladder allele for two DIPplex markers (D99 and D84). Two explanations can account for these features: A different size deletion/insertion allele at the locus or an additional neighboring InDel site with a rare minor allele within the amplicon range.



D97: Samples have been found to be carrying a neighboring allele on a clustering SNP (**A/G**), located 61 bp downstream from the main InDel site. This is a SNP referenced in the dbSNP database as **rs17245568**. The A allele of this SNP corresponds to the samples carrying the observed imbalance. Although we do not have access to Qiagen's primer designs, it seems reasonable to assume that this is the cause of the peak imbalance. Frequency: U.S. Cauc: **0.044** Af-Am: **0.22** U.S. Hisp: **0.062** U.S. Asian: **0.06**

D83: Insufficient data, sequencing will continue. Frequency: U.S. Cauc: **0** Af-Am: **0.08** U.S. Hisp: **0.015** U.S. Asian: **0**

D99: A locus with an off-ladder allele observed. This allele was determined to be a regular **Insertion** allele with a neighboring single-base deletion of an **A** residue, located 4 bp upstream from the InDel site. This is a referenced InDel on dbSNP database as **rs11346981**. No official allele frequency data were available on dbSNP for this marker. Frequency: U.S. Cauc: **0** Af-Am: **0.0766** U.S. Hisp: **0.0156** U.S. Asian: **0**

D84: A locus with an off-ladder allele in African American samples. This allele was determined to be a regular **Insertion** allele with a neighboring 4 bp (**ATTA**) deletion located 10 bases downstream the main InDel site. This is a referenced InDel on dbSNP database as **rs11573892**. Frequency: U.S. Cauc: **0** Af-Am: **0.0443** U.S. Hisp: **0** U.S. Asian: **0**

Conclusions:

- InDel genotyping constitutes a robust, sensitive and very informative identification system highly applicable to routine forensic casework.
- Population frequency data for U.S. representative human populations have been determined.
- A number of non-standard allelic variants were observed in the Qiagen DIPplex InDel set. These have been analyzed and show stable mobility variants for two of the markers that have been sequence characterized and their population frequencies estimated. The characterization of such rarer mobility variants can further contribute to the informative power of InDel typing.
- A successful protocol for artificial DNA fragmentation mimicking challenging DNA has been devised using COVARIS DNA shearing technology. InDel assays appear to be much better performers for these samples than any of the other assays compared in the study.

Acknowledgements: **Margaret Kline and Becky Hill** for assistance with allele sequencing, **Jennifer McDaniel** for assistance with the COVARIS system, **QIAGEN** for providing access to DIPplex kits.

References:

1- J.M. Mullaney et al., Hum. Mol. Genet. (2010) 19(R2): R131-R136.
2- H. Ellegren et al., BMC Genetics (2008), 9:8 doi:10.1186/1471-2156-9-8.
3- P. Gill, Int J legal Med, (2001) 114: 204-210.
4- C. Phillips, Methods Mol Biol. (2005) 297: 83-106.
5- http://www.qiagen.com/products/investigatordipplexkit.aspx
6- R. Pereira et al., *Electrophoresis (2009) 30: 3682–3690.*
7- M.C. Kline et al., Forensic Sci Int Genet. (2011) 5: 329-32.
8- E.M. Golemberg et al., Nuc. Acid Res. (1996) 24: N24.
9- Covaris Application manual. Adaptive focused Acoustics (AFA). Covaris, Inc. January 2007