Int J Legal Med (2012) 126:725–737
DOI 10.1007/s00414-012-0721-7

ORIGINAL ARTICLE

Forensic performance of two insertion–deletion marker assays

M. Fondevila · C. Phillips · C. Santos · R. Pereira · L. Gusmão · A. Carracedo · J. M. Butler · M. V. Lareu · P. M. Vallone

Received: 8 March 2012 / Accepted: 21 May 2012 / Published online: 20 June 2012

Abstract Improving the amplification and analysis of highly degraded DNA extracts has been a longstanding area of research in forensic genetics. One of the most promising recent developments in analysis of degraded DNA is the availability of short, biallelic insertion–deletion length polymorphisms (InDels) in highly multiplexed assays. InDels share many of the favourable characteristics of single-nucleotide polymorphisms (SNPs) that make them ideal markers for analysis of degraded DNA, including: analysis in short amplicon size ranges, high multiplexing capability and low mutation rates. In addition, as length-based polymorphisms, InDels can be analysed with the same simple dye-labelled PCR primer methods as standard forensic short tandem repeats. Separation and detection of fluorescently dye-labelled PCR products by capillary electrophoresis eliminate the multiple step protocols required by SNP typing with single-base extension assays and provide a closer relationship between the input DNA and the profile peak height ratios. Therefore InDel genotyping represents an effective new approach for human identification that adds informative new loci to the existing battery of forensic markers. To assess the utility of InDels for forensic analysis, we characterised population variation with two InDel identification assays: the 30-plex Qiagen DIPplex panel and a 38-plex panel developed by Pereira et al. in 2009 [1]. Allele frequencies were generated for the 68 markers in US African American, Caucasian, East Asian and Hispanic samples. We made a thorough assessment of the individual and combined performance of the InDel sets, as well as characterising profile artifacts and other issues related to the routine use of these newly developed forensic assays based on artificially degraded DNA and mixed source samples.

Keywords Short amplicon binary markers · Insertion–deletion polymorphisms · InDel · Population variation

Electronic supplementary material The online version of this article (doi:10.1007/s00414-012-0721-7) contains supplementary material, which is available to authorized users.

M. Fondevila · C. Phillips · C. Santos · A. Carracedo · M. V. Lareu
Forensic Genetics Unit, Institute of Legal Medicine,
University of Santiago de Compostela,
Galicia, Spain

R. Pereira · L. Gusmão
IPATIMUP—Institute of Molecular Pathology
and Immunology of the University of Porto,
Porto, Portugal

L. Gusmão
Laboratório de Genética Humana e Médica,
Instituto de Ciências Biológicas, Universidade Federeal do Pará,
Belém, Brazil

M. Fondevila (✉) · J. M. Butler · P. M. Vallone
U.S. National Institute of Standards and Technology,
Biochemical Science Division,
Gaithersburg, MD, USA
e-mail: mafondevila@hotmail.com

Introduction

Insertion–deletion polymorphisms (usually termed InDels or DIPs) are a type of biallelic short DNA length variation [2–6] that in the last 3 years have been subject to a growing interest in the forensic field [1, 7–11] due to a number of advantageous properties shared with the similar binary variation of single-nucleotide polymorphisms (SNPs). These properties include: the ability to type relatively short amplified fragments, much lower mutation rates compared to short tandem repeats (STRs), the potential for large multiplexed sets and a wide choice of loci to build optimum

marker combinations. When DNA is highly degraded, as is commonly encountered in forensic casework, short amplicon PCR has been demonstrated to markedly improve amplification success [12–14], so short allele InDels offer the same potential as SNPs for a near identical amplification efficiency in such conditions. Where InDel typing can offer improved performance compared to SNP analysis is the use of fluorescently dye-labelled PCR primers in genotyping assays that amplify two different length alleles subsequently separated and detected by capillary electrophoresis detectors (PCR-to-CE assays). In contrast, forensic SNP multiplex typing uses single-base extension (SBE) with fluorescently dye-labelled terminating bases that require consecutive reactions increasing the effect of stochastic differences between alleles as well as variation in SBE incorporation efficiencies between the four base terminators. As InDel detection uses a single PCR reaction similar to an STR typing assay, it offers a more direct relationship between input DNA and signal strength thus improving the balance of peak heights in the resulting profile.

Two multiplexed InDel typing assays using dye-labelled PCR products are currently available for human identification purposes: the commercial DIPplex kit of 30 InDels from Qiagen [15] and a 38-InDel multiplex assay developed by Pereira et al. in 2009 [1]. We have performed studies of allele frequency distributions amongst the major US population groups in both marker sets and used these studies to gauge the performance of each assay in routine use. We detected previously uncharted rare variants close to the insertion–deletion motif of certain loci that were then characterised by sequence analysis. We assessed the sensitivity of both short amplicon InDel assays when typing highly degraded DNA. Additionally, since InDel typing with PCR-to-CE assays offers better peak balance that is potentially comparable to STR typing, we analysed the ability of each InDel assay to detect the components of simple artificial mixtures. Lastly, as the low mutation rates of InDels, as with other binary markers, makes them ideal for relationship testing supplements [16] and an accurate system to estimate recombination rates between loci on the same chromosome (syntenic) now exists [17], we added the 68 InDels to an expanded genetic map of identification markers.

Materials and methods

DNA samples

A total of 712 population reference samples from the Biochemical Science Division, National Institute of Standards and Technology (NIST; Gaithersburg, USA) were used for genotyping studies. Sample panels consisted of individuals with self-declared ancestry representative of the four major population groups in the USA, comprising 262 US Caucasians, 260 African Americans, 140 US Hispanics and 50 US East Asians. Population variability data were analysed with Arlequin v2.5 [18] and the Promega Powerstats Excel calculator.

Assessments of routine InDel genotyping were made using the recently issued updated NIST standard reference material DNAs in the SRM2391c panel [19], comprising six component samples (A to F), with component D consisting of a 3:1 mixture of A and C. We also used the previous NIST standard reference material DNAs of SRM2391b.

InDel multiplexes

We examined two InDel multiplexes designed for human identification: the Qiagen Investigator DIPplex [15] (herein DIPlex) and a 38plex InDel set developed by Pereira et al. [1] (herein 38plex). For DIPplex amplification, manufacturer´s guidelines were followed throughout with a 30-cycle PCR amplification and 0.75 ng input DNA. PCR products were diluted 1:25 prior to capillary electrophoresis preparation, 1 μL of such dilution was mixed with a loading buffer containing 8.9 μL of HiDi formamide and 0.3 μL of Qiagen BTO size standard. DIPplex PCR products were run on a 3130xl Genetic Analyzer using POP4 as separation polymer, 36 cm capillary array and the appropriate Qiagen BT Dye matrix. The run module conditions suggested on the DIPplex user's manual was applied. The 38plex PCR amplification used 29 cycles applied in two stages: 10 cycles at 60 °C annealing step then 19 cycles at 58 °C annealing step. This differs from the previously described PCR protocol for this multiplex [1] in two ways: one less final PCR cycling step and a final extension step of 72 °C extended from 60 to 80 min (to promote full adenylation of PCR products). A DNA input of 1 ng was used, and PCR products were diluted 1:10 before capillary electrophoresis. One microlitre of such dilution was mixed with a loading buffer containing 8.9 μL of HiDi formamide and 0.3 μL of LIZ500 size standard. 38plex PCR products were run on a 3130xl Genetic Analyzer using POP4 as separation polymer, 36 cm capillary array, and a preloaded G5 dye fragment analysis run module. Marker details of all 68 InDels [1, 15] are listed in Supplementary Table S1.

Artificial DNA degradation and preparation of mixed DNA series

To degrade DNA to a desired maximum oligonucleotide length by controlled fragmentation, we applied the COVARIS commercial DNA shearing system [20]. The COVARIS system uses an adaptive focused acoustics (AFA) method that enables a highly specific and reproducible fragmentation of the DNA molecule. AFA fragments the DNA to a

predetermined length but in a random fashion regarding break points in order to mimic the degradation processes of casework samples, although it was not possible to artificially reproduce the effects of PCR inhibition in these experiments. Several fragmentation runs were made to determine conditions required to generate fragments between 100 and 250 bp from intact genomic DNA extracts. The final AFA conditions developed were: temperature 5 °C, frequency sweeping mode, duty cycle 10 %, intensity 10 %, cycles per burst 1,000, time 20 min, container 100 μL glass tube and total sample volume 100 μL at 3.4 ng/μL. The resulting oligonucleotide fragment lengths were evaluated by agarose gel electrophoresis with ethidium bromide staining and a standard 100-bp size reference ladder (Supplementary Fig. S1). The forensic STR kits of Identifiler and MiniFiler (Applied Biosystems, Foster City CA) were used to evaluate the performance of InDels with core STRs when typing degraded DNA.

For evaluation of PCR-to-CE InDel typing performance with mixed source DNA samples, we made simple artificial mixture series using SRM2391b-A (i.e. component A from the previous NIST SRM panel) and SRM2391c-B. Prior to preparing the mixtures, the SRM components were quantified by Quantifier in duplicate and mixed in ratios: 1:10, 1:5, 1:2, 1:1, 2:1, 5:1 and 10:1. A total of 0.75 ng (DIPplex) to 1 ng (38plex) of DNA was amplified for each mixture ratio point. Average peak height ratios were estimated by analysis of 100 heterozygotes for all 68 InDels, collected from the NIST population panel profiles.

Sequencing of novel DIPplex alleles

Sequence analysis was performed for samples exhibiting four novel or silent InDel alleles observed in the DIPplex set. These comprised allele sizes not previously reported plus samples indicating the possible presence of silent alleles likely to be due to SNPs in the PCR primer-binding sites, detected as consistently low peaks in multiple individuals from particular populations. As the PCR primer sequences of the DIPplex set are not published, we designed sequencing primers to bind outside the PCR fragment segment in each case and these are listed in Supplementary Table S2. Sequencing protocol followed the guidelines of Kline et al. [21].

Addition of InDel locations to an expanded forensic marker genetic map

As InDels are likely to be applied in identification of degraded DNA as supplementary loci for extended relationship testing [16] and may be situated close to core STRs, we expanded the current forensic genetic map to include the 68 InDels of our study. We applied an identical approach to that used for a recently constructed genetic map of 23 core and 16 supplementary forensic STRs that utilized high-density HapMap SNP data [17]. The HapMap recombination map and SNP proxies were used to identify the genetic distances between the above STRs and 68 InDels plus a further 23 of a 26 miniSTR set developed by NIST [22, 23]. An Rc estimate, derived from the Kosambi mapping function [24], of 0.04 or less (i.e. 4 % recombination) was used as a minimum value to highlight those marker pairs that are likely to require care when constructing relationship likelihoods.

Results and discussion

Population data and forensic informativeness of InDels

Genotype estimates, observed and expected heterozygosities plus Hardy–Weinberg equilibrium (HWE) analyses of the four US population groups are summarized in Tables 1 and 2 (DIPplex and 38plex). Loci with exact test *P* values indicating departure from HWE are highlighted in bold type in the tables, but these represent a small proportion of the total data. Therefore they were judged to be within the range of values expected from exact tests and not indicative of any irregular genotyping specific to a population as typified by the reduced peak heights of silent alleles detected in DIPplex InDels D83 and D97 [9–11]. However, we note the lowest *P* value from the most discrepant heterozygosity (observed data underestimating heterozygotes) was obtained for D97 in NIST African Americans. Although no silent allele homozygotes were detected in D97 (revealed as a missing or minimum peak within a normal profile), the possibility of additional, rarer primer-binding site SNPs in this InDel, leading to complete loss of the allele in apparently homozygous individuals, cannot be discounted.

The US Hispanic population shows very similar allele frequency estimates to US Caucasians, and this is in agreement with independent studies of the same samples with ancestry informative SNPs indicating that Hispanics are predominantly European in ancestry, with most individuals showing a much lower Native American co-ancestry component and in some individuals an even smaller African co-ancestry component. The African American panel shows a minor European co-ancestry component that does not exceed ~25–30 % proportion of the genetic variation [25, 26].

The mean random match probabilities (RMP) in each population are shown at the base of Tables 1 and 2, indicating a range of 10^{-11}–10^{-13} for DIPplex and 10^{-14}–10^{-15} for 38plex. While this contrast in match probabilities is expected from the size of each multiplex, there are also some discernible differences between the assays in the balance of allele frequencies, and therefore mean RMPs,

Table 1 Population data for 30 DIPplex InDels in US Caucasians (Cauc), US African Americans (Af-Am), US Hispanics (Hisp) and US East Asians (Asian)

	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
	D101				D111				D114				D118				D122			
Het (exp)	0.494	0.498	0.501	0.358	0.496	0.182	0.501	0.47	0.494	0.358	0.501	0.349	0.493	0.149	0.474	0.416	0.493	0.335	0.39	0.464
Het (obs)	0.454	0.56	0.507	0.35	0.542	0.2	0.45	0.488	0.473	0.42	0.464	0.35	0.473	0.12	0.536	0.419	0.496	0.34	0.386	0.419
A1 (del.)	0.559	0.56	0.475	0.233	0.45	0.9	0.518	0.625	0.557	0.77	0.525	0.225	0.565	0.08	0.618	0.706	0.561	0.79	0.736	0.637
A2 (ins.)	0.441	0.44	0.525	0.767	0.55	0.1	0.482	0.375	0.443	0.23	0.475	0.775	0.435	0.92	0.382	0.294	0.439	0.21	0.264	0.363
P value	0.212	0.405	1	0.73	0.14	1	0.242	0.596	0.53	0.417	0.405	1	0.532	0.261	0.153	1	1	1	1	0.142
	D124				D125				D128				D131				D133			
Het (exp)	0.464	0.489	0.495	0.345	0.493	0.46	0.47	0.401	0.5	0.407	0.501	0.451	0.496	0.485	0.502	0.36	0.498	0.5	0.5	0.44
Het (obs)	0.416	0.46	0.529	0.288	0.504	0.34	0.45	0.423	0.458	0.44	0.5	0.477	0.462	0.52	0.507	0.346	0.489	0.5	0.5	0.442
A1 (del.)	0.365	0.41	0.443	0.779	0.439	0.65	0.625	0.722	0.519	0.72	0.514	0.342	0.448	0.6	0.496	0.235	0.462	0.55	0.529	0.675
A2 (ins.)	0.635	0.59	0.557	0.221	0.561	0.35	0.375	0.278	0.481	0.28	0.486	0.658	0.552	0.4	0.504	0.765	0.538	0.45	0.471	0.325
P value	0.108	0.772	0.492	**0.012**	0.801	0.113	0.717	0.439	0.174	0.729	1	0.413	0.319	0.768	1	0.605	0.804	1	1	1
	D136				D39				D40				D45				D48			
Het (exp)	0.501	0.481	0.487	0.345	0.444	0.298	0.484	0.494	0.487	0.453	0.496	0.421	0.499	0.485	0.482	0.487	0.5	0.398	0.407	0.406
Het (obs)	0.519	0.5	0.471	0.35	0.427	0.32	0.443	0.465	0.492	0.6	0.45	0.385	0.515	0.4	0.543	0.508	0.508	0.3	0.421	0.412
A1 (del.)	0.504	0.39	0.414	0.221	0.668	0.82	0.593	0.44	0.582	0.34	0.554	0.7	0.471	0.4	0.4	0.585	0.483	0.73	0.282	0.278
A2 (ins.)	0.496	0.61	0.586	0.779	0.332	0.18	0.407	0.56	0.418	0.66	0.446	0.3	0.529	0.6	0.6	0.415	0.517	0.27	0.718	0.722
P value	0.62	1	0.73	1	0.579	1	0.381	0.377	0.9	**0.027**	0.302	0.185	0.621	0.248	0.158	0.522	0.902	0.144	0.834	0.879
	D56				D58				D6				D64				D67			
Het (exp)	0.449	0.503	0.443	0.498	0.501	0.485	0.443	0.279	0.501	0.5	0.5	0.482	0.497	0.243	0.413	0.349	0.467	0.416	0.487	0.428
Het (obs)	0.443	0.58	0.4	0.473	0.462	0.48	0.4	0.288	0.427	0.5	0.414	0.462	0.481	0.28	0.436	0.35	0.489	0.42	0.529	0.427
A1 (del.)	0.34	0.53	0.329	0.46	0.502	0.6	0.671	0.833	0.511	0.55	0.529	0.596	0.458	0.14	0.289	0.223	0.37	0.29	0.414	0.31
A2 (ins.)	0.66	0.47	0.671	0.54	0.498	0.4	0.329	0.167	0.489	0.45	0.471	0.404	0.542	0.86	0.711	0.777	0.63	0.71	0.586	0.69
P value	0.891	0.392	0.256	0.454	0.218	1	0.255	0.823	**0.019**	1	**0.045**	0.521	0.618	0.573	0.542	1	0.508	1	0.378	1
	D70				D77				D81				D83				D84			
Het (exp)	0.501	0.424	0.459	0.225	0.5	0.505	0.494	0.472	0.489	0.198	0.489	0.5	0.5	0.453	0.487	0.476	0.496	0.432	0.494	0.409
Het (obs)	0.55	0.44	0.464	0.242	0.477	0.5	0.493	0.465	0.469	0.1	0.486	0.531	0.489	0.44	0.486	0.477	0.485	0.34	0.4	0.385
A1 (del.)	0.496	0.3	0.354	0.129	0.483	0.51	0.561	0.621	0.578	0.11	0.421	0.481	0.481	0.66	0.586	0.392	0.452	0.31	0.436	0.389
A2 (ins.)	0.504	0.7	0.646	0.871	0.517	0.49	0.439	0.379	0.422	0.89	0.579	0.519	0.519	0.34	0.414	0.608	0.548	0.69	0.564	0.602
P value	0.138	1	1	0.273	0.458	1	1	0.894	0.526	**0.007**	1	0.384	0.712	1	1	1			A3:	0.009
																P value	0.711	0.184	**0.026**	0.066
	D88				D92				D93				D97				D99			
Het (exp)	0.498	0.502	0.487	0.44	0.498	0.503	0.48	0.424	0.495	0.465	0.477	0.501	0.498	0.368	0.497	0.485	0.492	0.228	0.502	0.485
Het (obs)	0.515	0.56	0.529	0.45	0.462	0.54	0.564	0.438	0.523	0.32	0.507	0.415	0.462	0.36	0.414	0.381	0.42	0.26	0.514	0.438
A1 (del.)	0.46	0.54	0.414	0.325	0.54	0.53	0.604	0.696	0.445	0.36	0.389	0.508	0.464	0.76	0.55	0.59	0.431	0.13	0.457	0.394
A2 (ins.)	0.54	0.46	0.586	0.675	0.46	0.47	0.396	0.304	0.555	0.64	0.611	0.492	0.536	0.24	0.45	0.41	0.569	0.87	0.539	0.591
P value	0.62	0.568	0.385	0.779	0.259	0.775	0.051	0.658	0.381	**0.034**	0.481	**0.007**	0.266	1	0.06	**0.001**		A3:	0.004	0.015
																P value	**0.024**	0.581	0.663	0.177
	Cauc	Asian	Hisp	Af-Am																
Mean DIPplex RMP	2E−13	5E−11	5E−13	6E−12																

Values at the base of the table show the expected average random match probabilities (RMP) in each population. *P* values in Bold represent those outside HWE

Table 2 Four US population panel data and expected average RMPs for the 38plex InDels in US Caucasians (Cauc), US African Americans (Af-Am), US Hispanics (Hisp) and US East Asians (Asian)

	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
	B1				B2				B3				B4				B5			
Het (exp)	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
Het (obs)	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
A1 (del.)	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
A2 (ins.)	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
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
	B6				B7				B8				B9				B10			
Het (exp)	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 (obs)	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
A1 (del.)	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
A2 (ins.)	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	0.599	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			
Het (exp)	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
Het (obs)	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
A1 (del.)	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
A2 (ins.)	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
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
	G6				G7				G8				G9				R1			
Het (exp)	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	0.432	0.492	0.429	0.468
Het (obs)	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	0.419	0.44	0.46	0.376
A1 (del.)	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	0.316	0.42	0.307	0.627
A2 (ins.)	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	0.684	0.58	0.693	0.373
P value	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	0.666	0.562	0.43	**0.002**
	R2				R3				R4				R5				R6			
Het (exp)	0.283	0.358	0.429	0.498	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
Het (obs)	0.278	0.38	0.417	0.492	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
A1 (del.)	0.832	0.77	0.689	0.463	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
A2 (ins.)	0.168	0.23	0.311	0.537	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
P value	0.825	1	0.842	0.901	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
	R7				R8				R9				R10				Y1			
Het (exp)	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	0.494	0.407	0.46	0.383
Het (obs)	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	0.456	0.36	0.453	0.384
A1 (del.)	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	0.443	0.28	0.357	0.258
A2 (ins.)	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	0.557	0.72	0.643	0.742
P value	0.313	1	0.5	0.108	0.778	0.325	1	**0.008**	**0.001**	1	0.107	0.072	0.618	0.384	0.18	0.313	0.213	0.486	0.855	1
	Y2				Y3				Y4				Y5				Y6			
Het (exp)	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
Het (obs)	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
A1 (del.)	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
A2 (ins.)	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

Table 2 (continued)

	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
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
	Y7				Y8				Y9											
Het (exp)	0.418	0.465	0.457	0.407	0.5	0.424	0.47	0.49	0.4	0.485	0.354	0.36								
Het (obs)	0.425	0.48	0.457	0.38	0.448	0.36	0.464	0.496	0.375	0.52	0.3	0.283								
A1 (del.)	0.301	0.36	0.35	0.283	0.523	0.3	0.625	0.574	0.28	0.4	0.229	0.234								
A2 (ins.)	0.699	0.64	0.65	0.717	0.477	0.7	0.375	0.426	0.72	0.6	0.771	0.766								
P value	0.881	1	1	0.285	0.106	0.321	1	0.899	0.351	0.768	0.09	**0.001**								
	Cauc	Asian	Hisp	Af-Am																
Mean 38plex RMP	4E−15	5E−14	1E−15	5E−15																
Mean combined 68-InDel RMP	7E−28	2E−24	7E−28	3E−26																

Values at the base of the table show the expected average random match probabilities (RMP) in each population

amongst the four population groups. The characteristic of population balance has been previously debated for forensic SNP sets [27], and it has been suggested that binary markers should ideally have very similar frequencies amongst the major population groups such that typical profile frequencies do not vary markedly between populations. Figure 1 shows the categorisation of composite InDels into five allele frequency bins for the four US populations studied and reveals better comparability between populations for the 38plex assay than the DIPplex assay with the latter showing a slightly disproportionate number of informative InDels in Caucasians compared to Africans and East Asians. However as binary marker informativeness only drops marginally between 0.5 to 0.3 allele frequencies, the interpopulation informativeness differences are not particularly strong.

Performance of InDel assays typing artificially fragmented DNA sample

Table 3 shows the allele and locus dropout rates observed with Identifiler, Minifiler and the two InDel sets analysing the artificially degraded DNA. Only Identifiler STRs performed poorly with two thirds of loci dropping out altogether and half of the alleles dropping out amongst the five heterozygous STRs detected. Both Minifiler and DIPplex assays gave complete profiles though peaks from fragments rising above 150 bp in size were progressively weaker. No discernible peak height ratio skew was observed for the longest allele size differences in DIPplex (up to a maximum 22 bp). Five of the longest 38plex loci dropped out, and these were all above 135 bp in length. The value of complete or near complete profiles obtained from large binary marker multiplexes is underlined by the RMP values obtained for this highly challenging DNA sample, where both InDel sets gave combined match probabilities (Table 3) two orders of magnitude higher than the eight STRs of Minifiler. InDel sets currently comprise ~35 % extra markers (comparing the SNP*for*ID 29plex [28] with this 38plex and the SNP*for*ID 34plex ancestry SNP panel with an equivalent 46plex set recently developed for forensic use [29, 30]). It is important to remark that these results have been generated with artificially degraded samples mimicking only DNA fragmentation. Caution is then recommended in extrapolating these results to naturally degraded samples, since as a result of a more complex degradation process and the accumulation of PCR inhibitors [12–14], these may behave differently from artificially degraded DNA.

Peak height ratio measurements and artificial mixture analysis

Normal variation in peak height ratios (PHR) was measured in heterozygous NIST population panel DNAs with the

Table 3 Profile completeness and resulting profile RMPs for two forensic STR sets and InDel multiplexes typing artificially degraded DNA (profile RMPs constructed from European frequencies)

Assay	Expt. alleles	Obs. alleles	Total loci	Amplified loci	% success	Profile RMP
Identifiler	10	5	15	5	33	–
Minifiler	16	16	9	9	100	2.89×10^{-12}
DIPplex	49	49	30	30	100	4.77×10^{-14}
38plex	43	43	38	33	87	1.03×10^{-14}

DIPplex InDel assay. Analysis of the artificial mixture series therefore concentrated on the DIPplex assay, and Fig. 2 plots the spread of PHR values obtained from 100 DIPplex heterozygotes. The normal value ranges create clusters of data points (black diamonds) around the 1:1 chart midline representing a control range of PHRs that can be expected from routine genotyping. The subclusters formed in D83 and D97 by the reduced signal variants are immediately apparent and contrast with the normal alleles in these InDels giving the same tight clustering around the 1:1 chart midline as the other DIPplex InDels. Sporadic outlier PHR values are evident in a small proportion of normal unmixed samples in certain InDels, and the log scale used exaggerates the deflected positions of these outliers. The outlier PHR values represent about 1 % of total data and while not randomly distributed with regard to the InDels (e.g. there are 7 out of 100 outliers in D125, position 8 from the left), they can be expected to form a small fraction of PHR values in any one profile. It is possible the outlier values observed in the controls may signify additional signal variants that would merit further study, but if these observations just indicate the extremes of normal variation, they still only comprise a small part of the expected PHR values in a normal DIPplex profile. The artificial mixture PHRs are shown as light grey elements for each InDel, and a large proportion of them fall outside the normal range suggesting each of the mixture ratios made would be detected as sets of multiple outliers, whereas extreme PHR values in a normal unmixed profile would be rarer events confined to just one or two loci at a time. Note that although the ratios are spread across three plots to minimize overlap, the data in Fig. 2 are complicated slightly by a lack of data points for certain loci in some ratios and no replication in ratios 10:1 and 5:1. Not all replicates were successfully genotyped for some loci in the other ratios. Three InDels, D133, D77 and D83, were not analysed as they did not show sufficiently informative peak height ratios in the mixture experiments.

To simplify the above data and more precisely gauge the proportion of mixed sample PHRs that can be expected to fall beyond the normal range, we measured how many PHR points in any one ratio lay outside the full maximum–minimum range and how many outside the 25 and 75 % quartile range derived from the 100 control heterozygotes. These results are shown for each ratio individually in supplementary Fig. S2. The plots show that generally almost half (average 48 %) of the PHRs in the artificial mixtures fall outside the full control range and a further ~29 % fall outside the control lower–upper quartile range, though this latter value is more variable. Values were: 10:1 30 % outside full range, 52 % outside quartile range; 5:1 54 and 27 %; 2:1 57 and 17 %; 1:1 60 and 23 %; 1:2 57 and 16 %; 1:5 44 and 24 %; and 1:10 41 and 43 %. Although these experimental mixture studies are best accomplished with a wide range of homozygous–heterozygous mixtures and must be extended to typical forensic

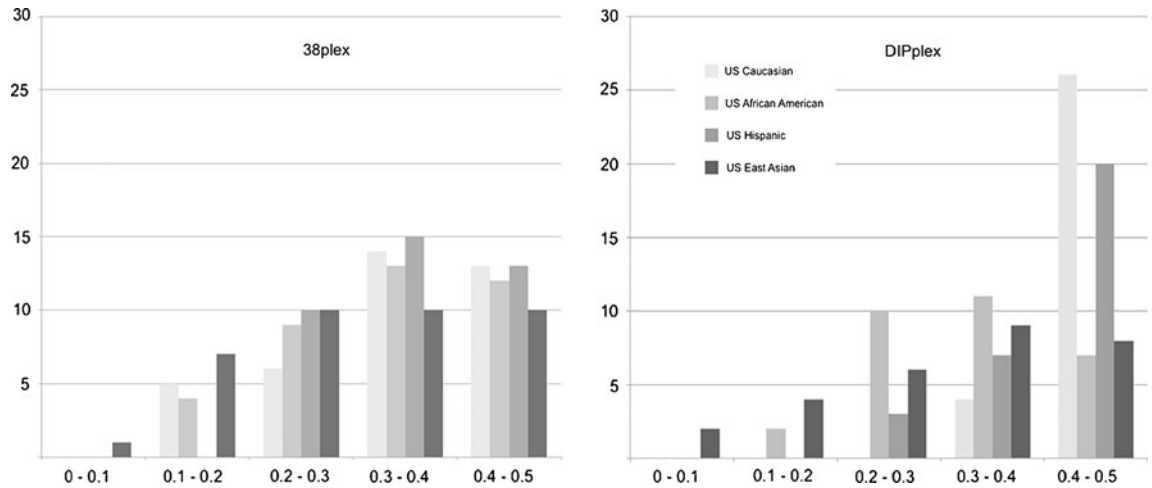

Fig. 1 Balance of InDel allele frequency distributions assessed across different population groups. Charts show the number of loci in each allele frequency bin for each assay and indicate 38plex has a more regular comparability of allele frequencies across the four populations

Table 4 Population frequency estimates of four variant alleles detected and suggested cluster SNP identifiers

Mobility shift alleles	US Caucasian	US African American	US Hispanics	US East Asian
D99	0	0.077	0.016	0
D84	0	0.044	0	0
Imbalanced signal alleles	US Caucasian	US African American	US Hispanics	US East Asian
D97	0.044	0.220	0.062	0.060
D83	0	0.080	0.015	0
Suggested SNP	1,000 genomes CEU	1,000 genomes ASW	1,000 genomes MXL	1,000 genomes CHB
D97: rs17245568	0.047	0.172	0.015	0.036
D83: rs73588849	0	0.049	0.008	0

Note that sample sizes vary, comprising 262 US Caucasians, 260 US African Americans, 140 US Hispanics and 50 US East Asians [25, 26]

casework material (when stochastic effects influence PHR variation much more), our early tests suggest that for the majority of DIPplex component loci, the average PHR range is compact and about half of the peak pairs in any one mixed sample would fall outside the range of extreme outlier PHRs detected in unmixed controls. Therefore, despite their binary nature, the prospects of mixture detection with InDels typed with the PCR-to-CE system are good and a hindrance to secure interpretation of forensic SNP profiles typed with SBE is largely overcome using InDel assays such as DIPplex.

Characterisation of previously unreported variants in the DIPplex assay

Multiple examples of four different novel variants were detected in the NIST population panels—all were heterozygous samples. There were two distinct types of variant allele, and these are detailed in Table 4 with examples of sequence analysis of each type shown in Fig. 3a, b. The first variant type was observed as a consistent pattern of non-stochastic allele signal imbalance in D83 (rs2308072) and D97 (rs17238892). Among additional similar variants, these findings have been independently observed and reported in similar works by Larue et al. and Friis et al. [9, 10]. Second, an off-ladder allele in DIPplex InDels D84 (rs2308163) and D99 (rs3081400) [9, 11] was observed as reproducible third alleles with consistent mobility shifts and comparable signal strength to the accompanying allele in each case.

The signal imbalance was strong, reproducible and allele consistent, i.e. in both InDels, only one of the two peaks showed reduced signal strength. The skew in peak height ratios was considerable (as seen on Fig. 2). At the time of writing, the D97 variant has also been reported in previous publications [9–11] as the effect over the primer binding of an SNP referenced as rs17245568 (sequence results on Fig. 3) in public databases.

The D83, D84 and D99 variants were almost exclusively confined to African Americans with a much lower frequency in US Hispanics in D83 and D99. Sequencing proved, in both cases, that the mobility shift was due to secondary InDel variation within the amplicon. In the case of D99, the mobility variant was just 1 bp shorter than the standard insertion allele.

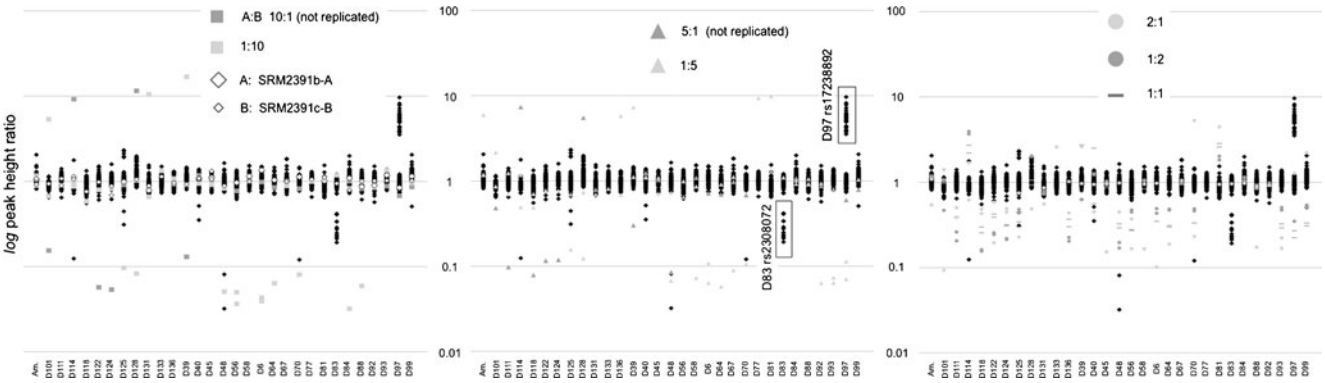

Fig. 2 Analysis of PHR distributions in DIPplex profiles from seven artificial mixtures (five with replicated analyses) and 100 control heterozygotes. A log scale is used to exaggerate the deflected values in most of the mixture signals. The PHR values for controls are shown in each chart as *black diamonds* and component SRM samples are shown as *white diamonds*. Artificial mixture values are spread across three identical charts for clarity. The effect of primer binding site SNPs in D83 and D97 are highlighted in the middle chart as clearly separated PHR subclusters. *Am* amelogenin

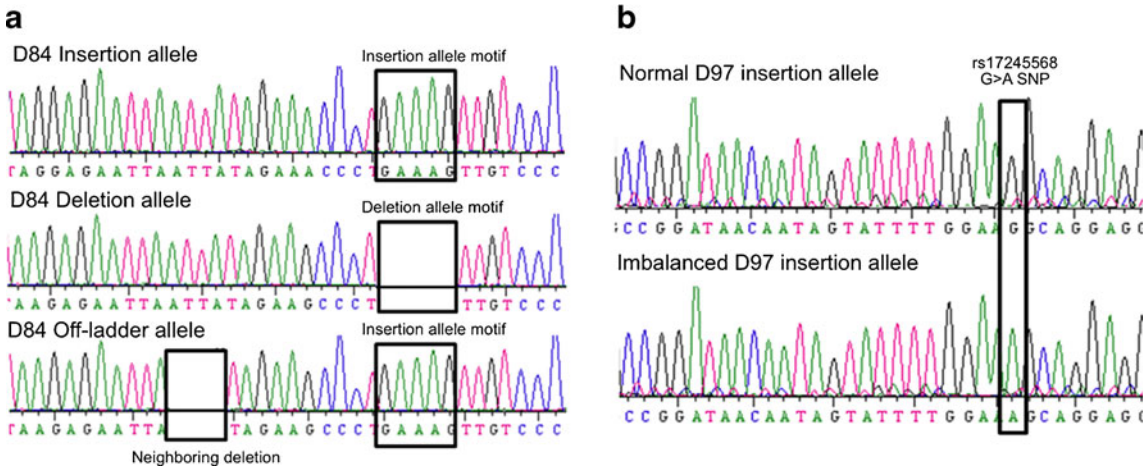

Fig. 3 Sequence analysis examples for **a** D84 mobility shift variant and **b** imbalanced signal variant D97

Sequencing revealed this to be due to a single-base deletion of an A residue, located 4 bp upstream from the D99 InDel and catalogued in dbSNP as rs11346981. D84 has a standard 5-bp insertion allele and also displayed a 1-bp shorter variant allele. Sequencing revealed this to be an ATTA deletion located 10 bp beyond the D84 site. The secondary ATTA deletion is catalogued in dbSNP as rs11573892, with a very low reported frequency of 0.021 for the deletion allele (from a combined African-non-African SNP discovery sample set). An example of the detected 4-bp deletion is shown in the sequence analysis examples for D84 in Fig. 3a.

The high frequency of the signal variant in D97 in all population groups and its independent observation in the Danish DIPplex study [10] provides a strong argument for the reformulation of the reverse amplification primer for this component InDel. In some ways the discovery of additional polymorphisms within the DIPplex amplicons suggests the promise of raised discrimination power for analysis of most or all populations. However the possibility of additional undetected variants with identical insertion–deletion lengths within other InDel amplicons may lead to increasing incidences of genotype disequilibrium in a proportion of loci. This extra, undetected variation could be difficult to accommodate if it occurs at significant frequencies in any one population. Another important factor is the risk of complete loss of peaks in highly degraded DNA samples carrying signal variants that will lead to mistyping of genotypes. D97 variants are present at a high enough frequency to produce a significant number of homozygotes, and although we did not detect D97 variant homozygotes, these can be expected in ~5 % of Africans.

Expanding the forensic genetic map to include InDels and mini-STRs

As the identification of missing persons and victims of mass disasters often involves pairwise comparisons in deficient pedigrees as well as the frequent need to type highly degraded DNA, it is likely that InDels will become popular as relationship testing supplements and will be chosen to improve identifications based on incomplete STR profiles [14, 16]. Although physical distance is a guide to linkage between marker pairs, the genetic distance is a better value to use when variation data are combined from very closely sited loci [17]. Therefore we have summarized data from a previously constructed genetic map we made [31], consisting of the 23 NIST mini-STRs [22, 23], 21 core forensic STRs and the 68 InDels of this study. Supplementary Table S1 lists the rs-number identifiers of the 68 InDels [1, 15] in dbSNP and their genomic positions that were used to locate the most closely linked marker pairs. Table 5 lists the 31 forensic marker pairs with the lowest recombination rates between them. Of these, 11 pairs (highlighted in bold) showed recombination rates below 4 %, the previous lowest value observed for forensic STRs: SE33-D6S1043 [17]. When adding 91 extra forensic loci, only three InDels should be highlighted as too close to established STRs to allow their use as independent supplements: 38plex rs1610919-D12S391 with ~2.7 % recombination, 38plex rs2307978-D7S820 ~0.45 % and DIPplex rs16363-D22S1045 ~0.04 %. Therefore only three InDels require exclusion from likelihood calculations when combined with established STRs due to very close linkage and this will have a minimal effect on kinship likelihood calculations. Removing the one DIPplex InDel close to D22S1045 and the two 38plex InDels close to core STRs reduces the resulting mean InDel set RMPs by one order of magnitude in both cases.

Remarkably, only five 38plex-miniSTR pairs and a single DIPplex-miniSTR pair have less than 4 % recombination; furthermore, none of the 23 mini-STRs show close linkage to established STRs. Lastly, making an adjustment for two InDel–InDel pairs on chromosome 22, and one each on

Table 5 Closely linked forensic marker pairs in a genetic map expanded to add 68 InDels and 23 NIST mini-STRs [21, 22] to 39 forensic STRs (ext. core=supplementary STRs)

Marker	Chromosome	Position	Assay	Type	Proxy SNP position (bp)	Physical distance in nucleotides	cM interval of closest HapMap SNP proxies	Rc from Kosambi MF
D1S1677	**1**	**160,747,000**	**NIST miniplex**	**STR**	**160,747,192**			
rs3047269	**1**	**161,077,452**	**HID-38plex**	**InDel**	**161,077,635**	**330,452**	**0.17608285**	**0.0018**
D2S1776	**2**	**169,471,000**	**NIST miniplex**	**STR**	**169,471,044**			
rs2307959	**2**	**169,898,519**	**DIPplex**	**InDel**	**169,899,211**	**427,519**	**0.96805650**	**0.0097**
D5S2500	**5**	**58,735,000**	**NIST miniplex**	**STR**	**58,734,281**			
rs1160956	**5**	**65,414,216**	**HID-38plex**	**InDel**	**65,414,138**	**6,679,216**	**3.29744874**	**0.0329**
rs1610935	**5**	**66,250,256**	**DIPplex**	**InDel**	**66,250,032**	**836,040**	**1.01442709**	**0.0101**
CSF1PO	**5**	**149,436,000**	**Identifiler/Powerplex**	**core STR**	**149,435,808**			
rs1305056	**5**	**155,594,834**	**DIPplex**	**InDel**	**155,594,321**	**6,158,834**	**7.06327427**	**0.0702**
D6S1017	6	41,785,000	NIST miniplex	STR	41,784,860			
rs2307710	6	47,929,222	HID-38plex	InDel	47,929,963	6,144,222	9.23875395	0.0914
SE33	6	89,043,000	Powerplex	ext core STR	89,043,366			
rs2307652	6	97,564,842	DIPplex	InDel	97,564,902	8,521,842	7.88452168	0.0782
D6S474	**6**	**112,986,000**	**NIST miniplex**	**STR**	**112,986,187**			
rs2307839	**6**	**117,600,251**	**HID-38plex**	**InDel**	**117,601,805**	**4,614,251**	**2.30250490**	**0.0230**
rs2307978	**7**	**83,121,850**	**HID-38plex**	**InDel**	**83,121,677**			
D7S820	**7**	**83,433,000**	**Identifiler/Powerplex**	**core STR**	**83,433,888**	**311,150**	**0.44400794**	**0.0044**
rs3081400	8	120,016,982	DIPplex	InDel	120,016,919			
D8S1179	8	125,976,000	Identifiler/Powerplex	core STR	125,977,380	5,959,018	8.34507763	0.0827
rs2067294	9	70,504,241	HID-38plex	InDel	70,506,274			
D9S1122	9	76,918,000	NIST miniplex	STR	76,917,975	6,413,759	5.52164750	0.0550
rs8190570	9	98,037,732	DIPplex	InDel	98,037,631			
rs2307580	9	104,626,014	HID-38plex	InDel	104,625,899	6,588,282	5.76410091	0.0574
D10S1435	10	2,233,000	NIST miniplex	STR	2,231,511			
rs140809	10	6,027,167	HID-38plex	InDel	6,027,136	3,794,167	9.29481510	0.0919
rs10688868	11	258,180	HID-38plex	InDel	258,219			
TH01	11	2,149,000	Identifiler/Powerplex	core STR	2,149,374	1,890,820	4.29551813	0.0428
rs33972805	11	125,794,082	HID-38plex	InDel	125,794,166			
D11S4463	11	130,338,000	NIST miniplex	STR	130,333,023	4,543,918	9.08990696	0.0899
VWA	12	6,093,104	Identifiler/Powerplex	core STR	5,963,801			
D12S391	**12**	**12,450,134**	**NGM/Powerplex**	**ext core STR**	**12,948,424**	**6,357,030**	**11.94138000**	**0.1194**
rs1610919	**12**	**14,801,263**	**HID-38plex**	**InDel**	**14,801,244**	**2,351,129**	**2.67496477**	**0.0267**
D12ATA63	12	106,825,000	NIST miniplex	STR	106,824,516			
D12ATA63	12	106,825,000	NIST miniplex	STR	106,824,516			
rs2067238	12	113,772,931	HID-38plex	InDel	113,772,984	6,947,931	7.36902815	0.0732
rs2307433	15	87,665,320	DIPplex	InDel	87,665,757			
PentaE	15	95,175,000	Powerplex	ext core STR	95,174,856	7,509,680	20.60811590	0.1952
rs2067208	16	83,139,788	HID-38plex	InDel	83,139,269			
D16S539	16	84,944,000	Identifiler/Powerplex	core STR	84,943,868	1,804,212	7.69258483	0.0763
rs2307581	17	3,916,882	DIPplex	InDel	3,916,925			
rs3051300	**17**	**10,076,666**	**HID-38plex**	**InDel**	**10,076,340**	**6,159,784**	**15.54811352**	**0.1507**
D17S974	**17**	**10,459,000**	**NIST miniplex**	**STR**	**10,459,517**	**382,334**	**0.67448392**	**0.0067**
rs1305047	17	16,025,713	DIPplex	InDel	16,025,829	5,566,713	14.62410283	0.1422
D18S51	18	591,000	Identifiler/Powerplex	core STR	592,081			
D18S853	18	3,981,000	NIST miniplex	STR	3,981,340	3,390,000	10.22313265	0.1008

Table 5 (continued)

Marker	Chromosome	Position	Assay	Type	Proxy SNP position (bp)	Physical distance in nucleotides	cM interval of closest HapMap SNP proxies	Rc from Kosambi MF
D20S482	20	4,454,000	NIST miniplex	STR	4,453,538			
rs33917182	20	11,643,625	HID-38plex	InDel	11,643,946	7,189,625	17.76035799	0.1705
rs16438	**20**	**25,226,470**	**DIPplex**	**InDel**	**25,224,680**			
rs34541393	**20**	**30,165,066**	**HID-38plex**	**InDel**	**30,165,461**	**4,938,596**	**0.67179959**	**0.0067**
rs35605984	21	14,556,736	HID-38plex	InDel	14,556,175			
D21S11	21	19,476,000	Identifiler/Powerplex	core STR	19,476,386	4,919,264	11.45578199	0.1126
rs10629077	21	30,294,208	HID-38plex	InDel	30,294,077			
rs8177524	21	33,582,626	DIPplex	InDel	33,582,670	3,288,418	5.12986658	0.0511
rs16388	**22**	**24,080,816**	**DIPplex**	**InDel**	**24,809,745**			
rs2307700	**22**	**25,120,901**	**HID-38plex**	**InDel**	**25,120,999**	**1,040,085**	**0.74801115**	**0.0075**
rs6481	22	34,031,900	DIPplex	InDel	34,032,068			
rs16363	**22**	**35,739,831**	**DIPplex**	**InDel**	**35,739,741**	**1,707,931**	**4.32384116**	**0.0431**
D22S1045	**22**	**35,779,000**	**NGM/Powerplex**	**ext core STR**	**35,778,806**	**39,169**	**0.04034871**	**0.0004**

An extract of data previously published [31]

chromosomes 20 and 5 reduces the total of independent InDels to 64 when combining the two sets but this only reduces the mean RMP in US Caucasians from 6.79×10^{-28} to 8.32×10^{-25} (equivalent to values obtained from 20 core STRs).

Concluding remarks This study did not aim to compare the relative performance and qualities of the two InDel multiplexes now available for forensic identification. Rather we concentrated on the forensic advantages that short amplicon InDel typing with PCR-to-CE methods will provide in general when choosing either set. In fact, as each multiplex uses 1 ng or less of DNA, a large proportion of casework can be analysed with both InDel sets and this can bring levels of discrimination well in excess of those provided by STRs. Furthermore this high discrimination power is maintained whether 38plex profiles are incomplete from dropout of the longest alleles or when three plus four InDels are removed to adjust for linkage with STRs or between component InDels in relationship testing. This suggests typing of ~60 InDels in two simple and robust multiplexes has considerable potential in the identification of missing persons, when incomplete STR profiles can be expected from the most challenging sources of DNA [14].

It is evident that the commercially prepared dye-labelled primer components of the DIPplex kit lead to more balanced profiles than those of the 38plex and this aids the detection of mixtures using DIPplex. However it is likely that InDel loci as a whole create much more balanced peak heights in heterozygous samples than SNPs typed by SBE, therefore binary markers are now better able to add complimentary data to mixed STR profiles than was previously achievable.

The biggest problem remains the hitherto undetected variants that in certain cases reach high frequencies in Africans and in one case has been detected in an independent study and across all our population panels. Luckily the addition of redundant primers is a common and effective approach to overcome the influence of similar rare primer-binding site SNPs in core STRs and will be equally applicable to adjust for the observed reduced signal variation in the two DIPplex loci. The mobility shift variants while rare suggest a slight increase in discrimination can be obtained in at least two population groups once these have been identified by the user and can be anticipated in the profile analysis process.

Acknowledgments The authors wish to thank Margaret Kline and Becky Hill at NIST as well as Anke Prochnow at Qiagen. The work of MF at NIST has been supported by the Fundacion Barrie de la Maza postgrade grant program (2010). MVL was supported by funding from Xunta de Galicia INCITE 09 208163PR and Ministerio de Educación y Ciencia BIO2006-06178. The support by funding from the Fundacion Marcelino Botin to the Institute of Legal Medicine of the USC. CS was partially supported through a PhD grant (SFRH/BD/75627/2010) awarded by the Portuguese Foundation for Science and Technology (FCT) and co-financed by the European Social Fund (Human Potential Thematic Operational Program). The Institute of Molecular Pathology and Immunology of the University of Porto (IPATIMUP) is an Associate Laboratory of the Portuguese Ministry of Education and Science

and is partially supported by FCT. RP holds an FCT fellowship (SFRH/BPD/81986/2011). LG is supported by a grant from CAPES/Brazil.

Disclaimer This work was funded in part through an interagency agreement between the National Institute of Justice and the NIST Office of Law Enforcement Standards as well as funding from the FBI Biometric Center of Excellence: 'Forensic DNA Typing as a Biometric Tool'. Points of view in this document are those of the authors and do not necessarily represent the official position or policies of the U.S. Department of Justice. 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.

Conflicts of interest None.

References

1. Pereira R, Phillips C, Alves C, Amorim A, Carracedo Á, Gusmão L (2009) A new multiplex for human identification using insertion/deletion polymorphisms. Electrophoresis 30:3682–3690
2. Weber JL, David D, Heil J, Fan Y, Zhao C, Marth G (2002) Human diallelic insertion/deletion polymorphisms. Am J Hum Genet 71:854–862
3. Mills RE, Luttig CT, Larkins CE, Beauchamp A, Tsui C, Pittard WS, Devine SE (2006) An initial map of insertion and deletion (INDEL) variation in the human genome. Genome Res 16:1182–1190
4. Mullaney JM, Mills RE, Pittard WS, Devine SE (2010) Small insertions and deletions (INDELs) in human genomes. Hum Mol Genet 19(R2):131–136
5. Mills RE, Pittard WS, Mullaney JM, Farooq U, Creasy TH, Mahurkar AA, Kemeza DM, Strassler DS et al (2011) Natural genetic variation caused by small insertions and deletions in the human genome. Genome Res 21:830–839
6. Väli Ü, Brandström M, Johansson M, Ellegren H (2008) Insertion-deletion polymorphisms (InDels) as genetic markers in natural populations. BMC Genet 9:8
7. Romanini C, Catelli ML, Borosky A, Pereira R, Romero M, Salado Puerto M, Phillips C, Fondevila M et al (2011) Typing short amplicon binary polymorphisms: supplementary SNP and InDel genetic information in the analysis of highly degraded skeletal remains. Forensic Sci Int Genet. doi:10.1016/j.fsigen.2011.10.006
8. Zidkova A, Horinek A, Kebrdlova V, Korabecna M (2011) Application of the new insertion–deletion polymorphism kit for forensic identification and parentage testing on the Czech population. Int J Legal Med. doi:10.1007/s00414-011-0649-3
9. Larue BL, Ge J, King JL, Budowle B (2012) A validation study of the Qiagen Investigator DIPplex® kit; an INDEL-based assay for human identification. Int J Legal Med (in press)
10. Friis SL, Børsting C, Rockenbauer E, Poulsen L, Fredslund SF, Tomas C, Morling N (2011) Typing of 30 insertion/deletions in Danes using the first commercial InDel kit—Mentype® DIPplex. Forensic Sci Int Genet. doi:10.1016/j.fsigen.2011.08.002
11. Fondevila M, Pereira R, Gusmao L, Phillips C, Lareu MV, Carracedo A, Butler JM, Vallone PM (2011) Forensic performance of insertion–deletion marker systems. Forensic Sci Int Genet Suppl Ser 3:e443–e444
12. Golenberg E, Bickel A, Weihs P (1996) Effect of highly fragmented DNA on PCR. Nucleic Acids Research Nucleic Acids Res 24:5026–5033
13. Sikorsky JA, Primerano DA, Fenger TW, Denvir J (2007) DNA damage reduces Taq DNA polymerase fidelity and PCR amplification efficiency. Biochem Biophys Res Commun 355:431–437
14. Fondevila M, Phillips C, Naverán N, Fernandez L, Cerezo M, Salas A, Carracedo Á, Lareu MV (2008) Case report: identification of skeletal remains using short-amplicon marker analysis of severely degraded DNA extracted from a decomposed and charred femur. Forensic Sci Int Genet 2:212–218
15. Investigator DIPplex Kit. http://www.qiagen.com/products/investigatordipplexkit.aspx
16. Phillips C, Fondevila M, García-Magariños M, Rodriguez A, Salas A, Carracedo Á, Lareu MV (2008) Resolving relationship tests that show ambiguous STR results using autosomal SNPs as supplementary markers. Forensic Sci Int Genet 2:198–204
17. Phillips C, Ballard D, Gill P, Syndercombe Court D, Carracedo Á, Lareu MV (2012) The recombination landscape around forensic STRs: accurate measurement of genetic distances between syntenic STR pairs using HapMap high density SNP data. Forensic Sci Int Genet 6:354–365
18. Schneider S, Roessli D, Excoffier L (2000) Arlequin: a software for population genetics data analysis User manual ver 2.00. Genetics and Biometry Lab, Dept. of Anthropology, University of Geneva; Geneva
19. Kline MC, Butts ELR, Hill CR, Coble MD, Duewer DL, Butler JM (2011) The new Standard Reference Material 2391c: PCR-based DNA profiling standard. Forensic Sci Int Genet Suppl Ser 3:e355–e356
20. Eschelbach JW, Zhuome D, Grady B, Goetzinger W (2011) Characterization of short-term temperature, exposure, and solubilization effects on library compound quality. J Biomol Screen 16:1112–1118
21. Kline MC, Hill CR, Decker AE, Butler JM (2011) STR sequence analysis for characterizing normal, variant, and null alleles. Forensic Sci Int Genet 5:329–332
22. Coble MD, Butler JM (2005) Characterization of 26 miniSTR loci for improved analysis of degraded DNA samples. J Forensic Sci 50:43–53
23. Hill CR, Kline MC, Coble MD, Butler JM (2008) Characterization of 26 miniSTR loci for improved analysis of degraded DNA samples. J Forensic Sci 53:73–80
24. Kosambi DD (1943) The estimation of map distances from recombination value. Ann Eugen 12:172–175
25. Phillips C, Fondevila M, Vallone PM, Santos C, Freire-Aradas A, Butler JB, Lareu MV, Carracedo Á (2011) Characterization of U.S. population samples using a 34plex ancestry informative SNP multiplex. Forensic Sci Int Genet. Suppl Ser 3 (2011) e182–e183
26. Lao O, Vallone PM, Coble MD, Diegoli TM, van Oven M, van der Gaag KJ, Pijpe J, de Knijff P et al (2010) Evaluating self-declared ancestry of U.S. Americans with autosomal, Y-chromosomal and mitochondrial DNA. Hum Mutat 12:1875–1893
27. Ryckman K, Williams SM (2008) Calculation and use of the Hardy–Weinberg model in association studies. Curr Protoc Hum Genet 1:1.18
28. Sanchez JJ, Phillips C, Borsting C, Balogh K, Bogus M, Fondevila M, Harrison CD, Musgrave-Brown E et al (2006) A multiplex assay with 52 single nucleotide polymorphisms for human identification. Electrophoresis 27:1713–1724
29. Phillips C, Salas A, Sánchez JJ, Fondevila M, Gómez-Tato A, Álvarez-Dios J, Calaza M, Casares de Cal M, Ballard D, Lareu MV, Carracedo Á (2007) The SNP*for*ID Consortium, Inferring ancestral origin using a single multiplex assay of ancestry-informative marker SNPs. Forensic Sci Int Genet 1:273–280
30. Pereira R, Phillips C, Pinto N, Santos C, dos Santos SEB, Amorim A, Carracedo Á, Gusmão L (2012) Straightforward

inference of ancestry and admixture proportions through ancestry-informative insertion deletion multiplexing. PLoS One 7:e29684
31. Fondevila M, Santos C, Phillips C, Carracedo A, Butler JM, Lareu MV, Vallone PM (2011) An assesment of linkage between forensic markers: core STRs, mini-STRs and InDels. Proccedings of 22nd International Symposium of Human Identification, Washington. https://www.promega.com/~/media/files/resources/conference%20-proceedings/ishi%2022/oral%20presentations/fondevila.ashx?la=en. Accessed 1 February 2012