Int J Legal Med (2012) 126:533–540
DOI 10.1007/s00414-012-0667-9

ORIGINAL ARTICLE

A validation study of the Qiagen Investigator DIPplex® kit; an INDEL-based assay for human identification

Bobby L. LaRue · Jianye Ge · Jonathan L. King · Bruce Budowle

Received: 10 November 2011 / Accepted: 3 January 2012 / Published online: 15 January 2012

Abstract Marker sets that are based on small insertion/deletion (INDEL) alleles can serve as useful supplementary or stand-alone assays for human identification. A validation study has been performed on a human identification assay based on a panel of 30 INDELs and amelogenin using the Investigator DIPplex® kit (Qiagen). The assay was able to type DNA from a number of forensically relevant sample types and obtain full profiles with 62 pg of template DNA and partial profiles with as little as 16 pg of template DNA. The assay is reproducible, precise, and non-overlapping alleles from minor contributors were detectable in mixture analysis ranging from 6:1 to 19:1 mixtures. Population studies were performed on the 30 indels, and there were no significant departures from Hardy–Weinberg equilibrium or significant linkage disequilibrium between the markers (after correction for sampling). In all populations, the random match probability was 1.43×10^{-11} or less, and the power of exclusion was greater than .999999999. We also discovered several microvariant alleles in our population samples. The data support that the Investigator DIPplex® kit provides a powerful supplement or stand-alone capability for human identity testing.

Keywords INDEL · Forensic DNA analysis · Validation · Degraded DNA · SNP · STR · Genotyping · Population genetics

Electronic supplementary material The online version of this article (doi:10.1007/s00414-012-0667-9) contains supplementary material, which is available to authorized users.

B. L. LaRue · J. Ge · J. L. King · B. Budowle (✉)
Institute of Applied Genetics,
Department of Forensic and Investigative Genetics,
University of North Texas Health Science Center,
3500 Camp Bowie Blvd.,
Fort Worth, TX 76107, USA
e-mail: bruce.budowle@unthsc.edu

Introduction

Typing of short tandem repeat loci (STRs) in forensic human identification testing is robust and reliable for samples containing relatively small quantities (250 pg–1 ng) of challenged quality DNA [1–8]. While STR-based typing performs well, it has some limitations. First, the amplicon sizes of STR-based markers range from approximately 100–400 bp. In highly degraded samples, target amplicons are usually less than 200 bp in length [9–11]; thus, some routinely configured loci will not be typeable. Second, artifacts, such as stutter peaks, can add ambiguity to mixture analysis. Stutter is exacerbated as the number of PCR cycles is increased (or some other method applied) to enhance sensitivity of detection [4, 12, 13] Third, STRs have relatively high mutation rates (approximately 10^{-3}), which can confound as well as reduce the power of kinship analyses [2].

An alternative marker system, i.e., single nucleotide polymorphism (SNP) markers, may perform well where STRs have limitations. SNPs can be captured in smaller amplicons than STRs; they do not produce stutter during the PCR; and their mutation rates are orders of magnitude lower than that of STRs [14]. However, SNP-based detection systems require that the chemistry and instrumentation reliably detect a single base change. To date, complex approaches have been sought which are unwieldy and often not quantitative [15–19].

The presence or absence of a small insertion or deletion (INDEL), technically a SNP, offers the value of a more simplified analytical process [14, 20–23]. The difference between the alleles is based on size rather than detecting a nucleotide substitution, and these differences are readily resolvable using capillary electrophoresis. Thus, the instrumentation for INDEL analyses is commonly found in forensic laboratories. Essentially, indels perform analytically similar to that of STRs.

An INDEL kit has become commercially available. The Investigator DIPplex® kit (Qiagen, Hilden, Germany) is a multiplex five-dye single-tube reaction assay for 30 bi-allelic INDEL markers and amelogenin. The markers are distributed over 18 autosomes (Supplementary Table 1). To evaluate this kit for human identity testing, a validation study was performed. The study included sensitivity and reproducibility studies, mixture analyses, testing of different body fluids or tissues, and population studies of forensic relevance in African-American, Asian-American, Hispanic, and Caucasian population samples.

Methods and materials

Sample acquisition

Buccal, blood, semen, and saliva samples were collected from unrelated individuals residing in Texas. Buccal swabs from unrelated individuals (77 African-American, 80 Asian, 75 Caucasian, and 69 Southwestern Hispanic) were also kindly provided by Genetic Testing Laboratories (Las Cruces, NM). The samples were anonymized and collected in accordance with methods approved by the Institutional Review Board for the University of North Texas Health Science Center in Ft. Worth, Texas. Mock casework samples were created by blotting or dropping collected fluids on cotton cloth and allowing them to dry at least overnight before processing. The sample types and number of samples analyzed are described in Supplementary Table 2.

Isolation of DNA and preparation of samples for analysis

DNA was isolated from buccal swabs and various mock casework samples using either the AutoMate *Express*® (Life Technologies, Carlsbad, CA) or the DNA Investigator® Kit (Qiagen, Hilden, Germany) according to the manufacturers' recommendations. The quantity of DNA was determined by qPCR using the Quantifiler® Quantification Kit and 7500 Real-Time PCR® System (Life Technologies). Samples were then normalized to 500 pg/μL and stored at either −20°C or −40°C until amplification.

Amplification and analysis

With the exception of the limit of detection analysis, 500 pg of DNA was amplified with the reagents provided in the DIPplex® Investigator kit (Qiagen) according to the manufacturer's protocol. Following amplification, samples were either immediately analyzed via capillary electrophoresis or stored at −20°C until analysis.

Each sample was prepared immediately prior to electrophoresis analysis and run under the same conditions on either the Applied Biosystems 3130×l or 3500×l Genetic Analyzer® (Life Technologies) according to the manufacturer's recommendations. Electrophoretic data were analyzed using Genemapper IDX® (Life Technologies).

Statistical analyses

Allele frequencies were determined by the counting method. Population genetic parameters were analyzed by using either Genetic Data Analysis software [24] or in-house developed software [25]. Departures from Hardy–Weinberg equilibrium (HWE) and linkage equilibrium were tested using Fisher's exact test. Bonferroni correction for multiple comparisons was performed according to Weir and Cockerham [26].

Results and discussion

General performance

Buccal swabs (5), blood stains (16), semen stains (5), and saliva stains (2) were assayed for the ability to obtain a result from those tissue types. At 500 pg of template DNA, full profiles were obtained from all samples tested. Furthermore, the profiles observed were reproducible and consistent among sample types for a given individual (two individuals for buccal, blood, semen, and saliva stains; three additional individuals for buccal, blood, and semen). A single dye channel example of the typing results from an individual is shown in Supplemental Figure 1.

Sensitivity studies

To determine the limit of detection of the assay, a dilution series was analyzed for two individuals at 2 ng, 1 ng, 500 pg, 250 pg, 125 pg, 62 pg, 31 pg, and 16 pg. The samples were amplified for 30 cycles (Supplemental Figure 2). The results were robust down to the 125 pg. With 62 pg of template, full profiles were obtained, but the quality of the electropherogram was diminished compared with greater quantities of template DNA. Samples amplified for 30 cycles yielded greater than 90% of a full profile at 31 pg of template and generated greater than 60% of a full profile at 16 pg/μL. However, the quality of the electropherogram was lower. The same samples were amplified for 32 cycles as per the manufacturer's recommendation for low copy number samples (Supplemental Table 3). Full profiles were obtained down to 62 pg. Thirty-two cycle amplifications yielded greater than 95% of a full profile with 31 pg of sample and generated greater than 80% of a full profile with 16 pg of sample (Supplemental Table 3).

Ideally, between 250 and 500 pg of template should be used in reactions for the assay. No more than 1 ng of

template should be added to a DIPplex reaction due to saturation issues. Sample overload will likely cause pull-up into the channel of the internal lane standard (ILS). If pull-up of this type occurs, the presence of a number of allele peaks in the 80-, 90-, 100-, and 140-bp range can confound the sizing of the ILS peaks and result in an inability of the analysis software to correctly assign fragment calls to the sample peaks (unless a manual override is performed).

In two instances (one at 2 ng and one at 16 pg), an extra foreign allele was detected with 32 cycles of amplification. No false positive alleles were observed at 30 cycles. Caution should be exercised when increasing the sensitivity of an assay, particularly where the markers are bi-allelic. Contamination may be difficult to detect and could possibly lead to a false interpretation.

Reproducibility and precision assays

Ten separate samples were each amplified in ten individual reactions. Among reactions involving the same sample, there were no differences with regard to allele calls. Heterozygote peak height balance was similar among the replicates of an individual as well as among individuals, noted by the small standard deviations (Supplemental Table 4) except for loci HLD6 (one individual) and HLD97(three individuals). At these loci, examples of greater variance in heterozygote peak balance were observed. These microvariants, which are discussed in the following microvariant subsection, are currently being sequenced to determine if primer binding site or other mutations impacting secondary structure are the cause of the peak imbalance.

To demonstrate the precision of the assay, one sample was amplified twice, and the amplified product was loaded individually into eight separate wells of a 96-well plate and injected three separate times on a 3500XL Genetic analyzer for a total of 24 electrophoretic runs. These replicates essentially were indistinguishable with regard to allele calls, sizing precision, and heterozygote peak balance (Supplemental Table 5).

Mixture studies

Two samples of template DNA (one male and one female) at a total concentration of 500 pg/μL were mixed in the following ratios: 19:1, 9:1, 6:1, 4:1, 2:1, 1:1, 1:2, 1:4, 1:6, 1:9, and 1:19. There were 19 non-overlapping alleles between the two samples, 8 alleles from sample "A" and 11 from sample "B". At a ratio of 19:1 (A/B), 3 of the 11 non-overlapping "B" alleles were detectable in the mixture. This increased to seven alleles at 9:1, with all of the non-overlapping "B" allele peaks visible in the 6:1 mixture. When the ratio was 19:1 (B/A), three of the eight non-overlapping "A" allele peaks were observed in the mixture. Four of the non-overlapping "A" alleles were detectable at 9:1, six at 6:1, and all eight were detectable at a ratio of 4:1 (Table 1; Supplemental Figure 3).

Mixtures also were created by combining body fluids on cotton cloth. A blood-soaked cotton cloth sample from a single individual was allowed to dry overnight, and then 100 μL of freshly thawed and vortexed semen was placed directly on top of the blood stain. Five separate mixed stains were produced using five different semen donors. Two samples of saliva from a single donor were placed on cotton and allowed to dry overnight and followed by 100 μl of

Table 1 Investigator DIPplex® assay allele calls for non-overlapping alleles of two 500-pg/μL samples mixed at varying ratios[a,b]

Mixture ratio	Amel	HLD 125	HLD 128	HLD 133	HLD 136	HLD 40	HLD 45	HLD 48	HLD 56	HLD 64	HLD 70	HLD 81	HLD 84	HLD 92	HLD 93	HLD 99
A1	X/X	−/+	−/−	−/−	+/+	−/−	+/+	−/+	−/+	+/+	−/+	+/+	+/+	−/+	+/+	+/+
A19:1	X/X	−/+	−/−	−/+	+/+	−/−	−/+	−/+	−/+	+/+	−/+	+/+	+/+	−/+	−/+	+/+
A9:1	X/X	−/+	−/+	−/+	−/+	−/−	−/+	−/+	−/+	−/+	−/+	+/+	+/+	−/+	−/+	−/+
A6:1	**X/Y**	−/+	−/+	−/+	−/+	−/+	−/+	−/+	−/+	−/+	−/+	−/+	−/+	−/+	−/+	−/+
A4:1	**X/Y**	−/+	−/+	−/+	−/+	−/+	−/+	−/+	−/+	−/+	−/+	−/+	−/+	−/+	−/+	−/+
A2:1	**X/Y**	−/+	−/+	−/+	−/+	−/+	−/+	−/+	−/+	−/+	−/+	−/+	−/+	−/+	−/+	−/+
A1:1B	**X/Y**	−/+	−/+	−/+	−/+	−/+	−/+	−/+	−/+	−/+	−/+	−/+	−/+	−/+	−/+	−/+
B2:1	**X/Y**	−/+	−/+	−/+	−/+	−/+	−/+	−/+	−/+	−/+	−/+	−/+	−/+	−/+	−/+	−/+
B4:1	**X/Y**	−/+	−/+	−/+	−/+	−/+	−/+	−/+	−/+	−/+	−/+	−/+	−/+	−/+	−/+	−/+
B6:1	**X/Y**	−/+	−/+	−/+	−/+	−/+	−/+	−/−	−/−	−/+	−/+	−/+	−/+	−/+	−/+	−/+
B9:1	**X/Y**	−/−	−/+	−/+	−/+	−/+	−/+	−/−	−/−	−/+	+/+	−/+	−/+	−/+	−/+	−/+
B19:1	**X/Y**	−/−	−/+	−/+	−/+	−/+	−/+	−/−	−/−	−/+	+/+	−/+	−/+	+/+	−/+	−/+
B1	**X/Y**	−/−	−/+	+/+	−/−	−/+	−/+	−/−	−/−	−/+	+/+	−/+	−/+	+/+	−/−	−/+

Entries in bold denote where minor contributor is observed. Only loci with non-overlapping alleles between the two individuals are shown

Table 2 Allele distribution, observed heterozygosity, Fisher's exact test for Hardy–Weinberg equilibrium, random match probability, and power of exclusion for the 30 indels in Investigator DIPplex® kit in four major North American populations

	African-American (n=77)					Asian(n=80)					Caucasian (n=75)					Southwestern Hispanic (n=69)				
Marker	Frequency of deletion allele	H_o	HWE (p)	RMP	PE	Frequency of deletion allele	H_o	HWE (p)	RMP	PE	Frequency of deletion allele	H_o	HWE (p)	RMP	PE	Frequency of deletion allele	H_o	HWE (p)	RMP	PE
HLD 77	0.6039	0.4545	0.6591	0.3770	0.6135	0.6500	0.4500	1.0000	0.3988	0.5995	0.4800	0.4800	0.8125	0.3664	0.6246	0.4380	0.5507	0.3169	0.4220	0.6157
HLD 45	0.5779	0.5065	0.8278	0.3905	0.6187	0.3250	0.4750	0.6141	0.4247	0.5887	0.4333	0.4667	0.6513	0.3689	0.6204	0.4310	0.4493	0.6244	0.3741	0.6141
HLD 131	0.2338	0.2857	0.1078	0.4785	0.5240	0.6250	0.4250	0.4788	0.3772	0.6079	0.4267	0.4533	0.6472	0.3657	0.6194	0.4560	0.5797	0.2409	0.4245	0.6249
HLD 70	0.1169	0.2078	1.0000	0.6505	0.3490	0.3938	0.4625	0.8006	0.3809	0.6129	0.4467	0.5467	0.4869	0.4073	0.6221	0.4810	0.4493	0.4709	0.3640	0.6196
HLD 6	0.5779	0.4545	0.6419	0.3675	0.6187	0.4813	0.4375	0.2644	0.3503	0.6246	0.5133	0.4133	0.1525	0.3433	0.6248	0.4500	0.4348	0.3259	0.3573	0.6206
HLD 111	0.6883	0.4416	1.0000	0.4218	0.5820	0.8500	0.2750	1.0000	0.5834	0.4125	0.4867	0.4400	0.3706	0.3508	0.6248	0.5060	0.4928	0.6025	0.4178	0.5986
HLD 58	0.8312	0.2857	1.0000	0.5561	0.4432	0.6438	0.4375	0.6163	0.3909	0.6018	0.4867	0.6000	0.1059	0.4404	0.6248	0.4880	0.3768	1.0000	0.4799	0.5221
HLD 56	0.4610	0.5065	1.0000	0.3813	0.6235	0.5188	0.5125	1.0000	0.3822	0.6246	0.3600	0.5600	0.0847	0.4496	0.6031	0.3440	0.4203	1.0000	0.4392	0.5643
HLD 118	0.6688	0.4545	1.0000	0.4124	0.5916	0.1563	0.2875	0.6941	0.5728	0.4231	0.5867	0.4000	0.1522	0.3550	0.6172	0.5620	0.3188	0.1319	0.4346	0.5593
HLD 92	0.7143	0.4156	1.0000	0.4353	0.5664	0.4688	0.5375	0.6563	0.3978	0.6240	0.5533	0.5467	0.4775	0.4073	0.6221	0.5500	0.4783	0.8000	0.3653	0.6248
HLD 93	0.5844	0.4416	0.4825	0.3652	0.6176	0.4500	0.4000	0.1200	0.3450	0.6225	0.5200	0.5333	0.6541	0.3941	0.6246	0.4500	0.5072	0.4091	0.4295	0.5957
HLD 99	0.3896	0.5714	0.0978	0.4427	0.6119	0.2063	0.2625	0.0994	0.5134	0.4940	0.4267	0.4533	0.6309	0.3657	0.6194	0.4810	0.5362	0.6378	0.3989	0.6231
HLD 88	0.3052	0.4545	0.7809	0.4313	0.5784	0.4813	0.4625	0.5116	0.3591	0.6246	0.4467	0.5200	0.8163	0.3913	0.6221	0.5560	0.4203	0.3156	0.3598	0.6171
HLD 101	0.1883	0.2987	0.7322	0.5294	0.4712	0.5563	0.4625	0.6478	0.3647	0.6218	0.4600	0.5200	0.8113	0.3888	0.6234	0.4560	0.4783	0.8138	0.3686	0.6231
HLD 67	0.2662	0.3766	0.7622	0.4454	0.5524	0.2875	0.4250	1.0000	0.4363	0.5676	0.3733	0.4267	0.4672	0.3785	0.6074	0.3000	0.6377	0.012[a]	0.4900	0.6157
HLD 83	0.3377	0.4156	0.6022	0.3962	0.5945	0.6250	0.4750	1.0000	0.3947	0.6079	0.4800	0.4267	0.2534	0.3472	0.6246	0.6060	0.4203	1.0000	0.4157	0.5819
HLD 114	0.2013	0.3506	0.7106	0.5122	0.4880	0.7688	0.2625	0.0228[a]	0.4853	0.5215	0.5933	0.5200	0.6406	0.4030	0.6158	0.6310	0.4783	0.7969	0.3665	0.6242
HLD 48	0.2727	0.4156	0.7803	0.4468	0.5573	0.5688	0.4375	0.3753	0.3591	0.6201	0.4867	0.5200	0.8188	0.3860	0.6248	0.4500	0.4928	0.4028	0.4371	0.5857
HLD 124	0.7403	0.3117	0.1391	0.4495	0.5473	0.4250	0.5250	0.6463	0.3997	0.6192	0.3733	0.5067	0.6175	0.4105	0.6074	0.3750	0.5797	0.2044	0.4329	0.6206
HLD 122	0.6494	0.4416	0.8047	0.3955	0.5997	0.7938	0.3125	0.7153	0.5066	0.4940	0.5400	0.5200	0.8231	0.3888	0.6234	0.6060	0.4493	0.7884	0.3875	0.6063

Table 2 (continued)

Marker	African-American (n=77)					Asian(n=80)					Caucasian (n=75)					Southwestern Hispanic (n=69)				
	Frequency of deletion allele	H_o	HWE (p)	RMP	PE	Frequency of deletion allele	H_o	HWE (p)	RMP	PE	Frequency of deletion allele	H_o	HWE (p)	RMP	PE	Frequency of deletion allele	H_o	HWE (p)	RMP	PE
HLD 125	0.7403	0.3377	0.3803	0.4488	0.5473	0.6063	0.4125	0.2475	0.3653	0.6129	0.5333	0.5067	1.0000	0.3806	0.6239	0.4440	0.5507	0.025[a]	0.4808	0.5779
HLD 64	0.2922	0.4545	0.5813	0.4417	0.5706	0.1375	0.2250	0.6228	0.6138	0.3900	0.4733	0.6267	0.0328 [f]	0.4638	0.6243	0.4630	0.3623	0.7203	0.5026	0.4986
HLD 81	0.4740	0.5844	0.1734	0.4292	0.6243	0.2688	0.4375	0.4019	0.4566	0.5544	0.6267	0.4800	1.0000	0.3977	0.6074	0.5060	0.5362	0.1981	0.4371	0.6013
HLD 136	0.2143	0.3247	0.7541	0.4967	0.5034	0.4375	0.5500	0.3609	0.4116	0.6210	0.4933	0.5600	0.3622	0.4105	0.6250	0.5060	0.3188	0.0084[a]	0.3716	0.6039
HLD 133	0.5649	0.5584	0.3519	0.4178	0.6207	0.6500	0.5250	0.2325	0.4334	0.5995	0.4600	0.4133	0.1644	0.3461	0.6234	0.4560	0.4493	0.8122	0.3804	0.6105
HLD 97	0.5260	0.5065	1.0000	0.3797	0.6243	0.5563	0.5625	0.2584	0.4184	0.6218	0.5067	0.4800	0.8153	0.3657	0.6250	0.4880	0.5797	0.2344	0.4245	0.6249
HLD 40	0.6883	0.4156	0.7969	0.4144	0.5820	0.4375	0.5000	1.0000	0.3828	0.6210	0.5600	0.4800	0.8084	0.3728	0.6213	0.5810	0.4928	1.0000	0.3892	0.6157
HLD 128	0.3052	0.4545	0.7772	0.4313	0.5784	0.6625	0.3750	0.1394	0.3888	0.5944	0.4800	0.4800	0.8266	0.3664	0.6246	0.5060	0.5217	0.8234	0.3917	0.6224
HLD 39	0.4610	0.5065	1.0000	0.3813	0.6235	0.8500	0.2250	0.3681	0.5959	0.4125	0.6667	0.4000	0.4269	0.3956	0.5926	0.6120	0.4928	0.7872	0.4094	0.6039
HLD 84	0.2792	0.4026	1.0000	0.4380	0.5620	0.2375	0.3500	0.7700	0.4716	0.5276	0.4400	0.4800	0.8194	0.3728	0.6213	0.4060	0.4493	0.4747	0.3535	0.6250
Total				1.43×10^{-11}	0.9999				7.62×10^{-12}	0.9999				3.65×10^{-13}	0.9999				2.12×10^{-12}	0.9999

α-level of .05 is adjusted from .05 to .0016667 when corrected for multiple tests (Bonferroni's correction) [26]. Hardy–Weinberg equilibrium was calculated with GDA software [24]

H_o observed heterozygosity, *RMP* random match probability, *PE* denotes power of exclusion

[a] Loci with p values less than .05

semen (from three different donors) being placed on separate dried stains. These two-person mixtures were allowed to dry overnight. The next day, the DNA was extracted (from (~0.0625 cm^2), quantified, and normalized. Approximately 500 pg total DNA was used for amplification. While not as stochiometrically precise of a mixture as the samples involving extracted DNA, all of the mock casework-type mixture samples returned full expected profiles for both contributors. An example of a resultant mixture is shown in Supplemental Figure 4.

Population studies

Four sample populations, African-Americans (n=77), Asians (n=80), Caucasians (n=75), and southwestern Hispanics (n=69), were typed for the 30 indels. The allele frequency for the deletion allele, observed heterozygosity, power of exclusion, and random match probability (RMP) for each locus are shown in Table 2. All loci were highly polymorphic in the four populations. The cumulative RMP of all 30 loci was 1.43×10^{-11}, 7.63×10^{-12}, 3.65×10^{-13}, and 2.12×10^{-12} for African-American, Asian, Caucasian, and Southwestern Hispanic populations, respectively. The cumulative probabilities of exclusion are greater than 0.999999999 for all four populations. There were five instances of detected departures from HWE (HLD114 in Asians; HLD64 in Caucasians; HLD67, HLD125, and HLD136 in Southwestern Hispanics). For African-Americans, Asians, and Caucasians, the number of observed departures is no more than would be expected by chance. There were more departures from HWE observed in Southwestern Hispanics than would be expected by chance alone. However, after Bonferroni correction for multiple comparisons [26], no significant deviation from HWE was observed for any of the loci in any of the populations (Table 2).

Linkage disequilibrium (LD) between the loci was also tested. Thirteen pairs in African-Americans, 26 pairs in Asians, 11 pairs in Caucasians, and 30 pairs in Hispanics were detected demonstrating significant LD (p<0.05). A number of the pairs of loci demonstrating LD also were ones where one or both of the loci showed departures from HWE before correction. This observation could be due to substructure leading to significant LD or could be due to an association with the locus departing from HWE as described by Chakraborty and Falush et al. [27–29]. When the critical value for LD was corrected for multiple tests (via Bonferroni's correction), only the loci pair HLD67/HLD136 in the Southwestern Hispanic population displayed significant LD (Supplemental Table 6). These two markers are not on the same chromosome (Supplemental Table 1), but both departed from HWE at the non-corrected critical value of 0.05.

Several pairs of loci that did not demonstrate significant LD (after correction) reside on the same chromosome and were physically close to each other (i.e., <50 Mb); they were HLD40 and HLD39, HLD45 and HLD48, HLD64 and

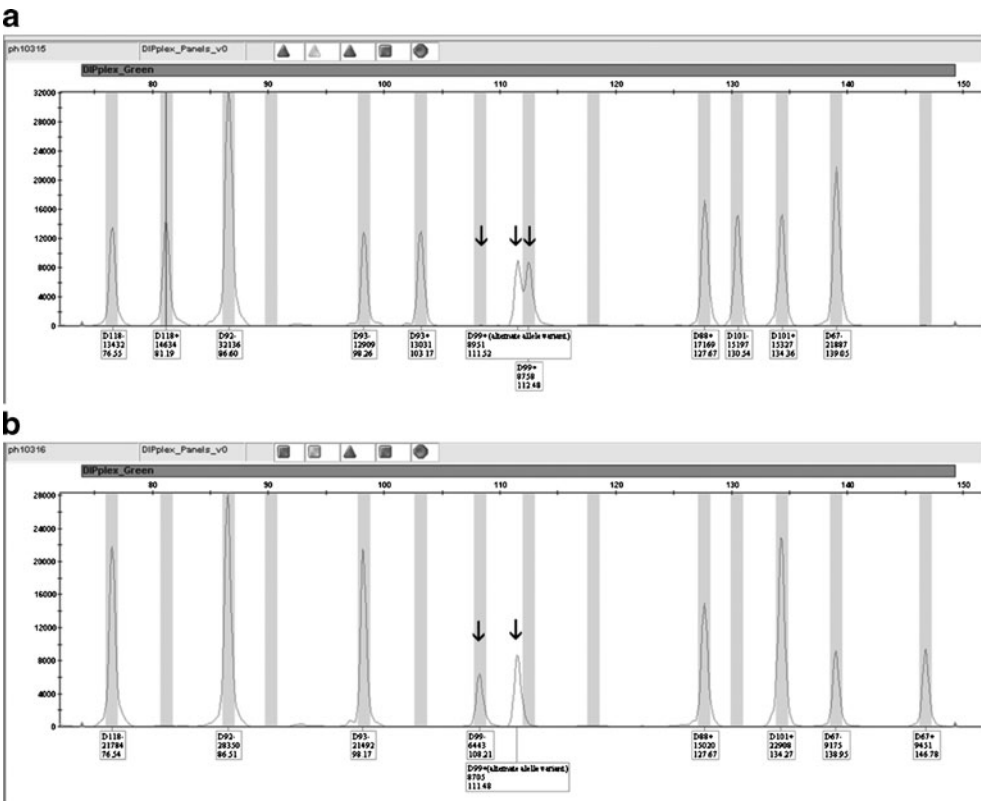

Fig. 1 Example of HLD99+ microvariant allele. Presence of apparent one base deletion microvariant of the HLD99+ allele. **a** A father who is heterozygous for the microvariant allele and the reported allele for HLD99+. The allele is passed to his offspring (**b**) along with a maternal HLD99− allele

HLD58, HLD81 and HLD77, HLD77 and HLD131, HLD114 and HLD111, HLD125 and HLD136, and HLD136 and HLD124 (Supplemental Tables 1 and 6). These loci may be genetically linked, and the recombination fractions between these loci may be less than 0.5. Sets of family data are needed to perform linkage analysis and determine the recombination fractions between these loci.

Wright's F_{ST} was analyzed to measure population substructure effects [26]. For all four populations combined, the F_{ST} value over all 30 loci was 0.073. A few loci contributed more so than others to these values. For example, the loci HLD114 and HLD118 had F_{ST} values of 0.216 and 0.252, respectively. Typically, such disparate populations would not be combined for forensic applications. Thus, these values should be considered as an upper bound (Supplemental Table 7). More subpopulation data within a major population group are needed to determine population-specific values.

Microvariants

The initial data gleaned from the NCBI's dbSNP database [30] were from samples of no more than 21 individuals in the case of the African-American samples and no more than 25 individuals for Asians. While such a small sampling of populations might be adequate for allele frequency estimates for high heterozygosity bi-allelic loci, larger population samples are needed to detect microvariants in populations. Indeed, microvariants, not previously described, were observed in our study as well as by Friis et al. [31] and Alvarez et al. [32].

Two "off-ladder" microvariants were observed—an apparent one base deletion at the HLD99+ allele (Fig. 1) and an apparent one base insertion at the HLD84− allele. These microvariants occur at frequencies of 0.039 in the African-American population for the HLD99+ allele and 0.052 in the African-American and 0.14 in Hispanic populations for the HLD84− allele. Alvarez et al. [32] observed these microvariants at similar frequencies in the same populations. Friis et al. [31] did not describe these two microvariants; however, they did not assay African-American or Hispanic populations.

A reduced peak height variant (i.e., peak height imbalance where an allele had reduced amplicon yield) for the HL97+ allele was observed at 0.28 in African-American, 0.04 in Caucasian, 0.075 in Asian, and 0.043 in Southwestern Hispanic sample populations. Also, a low peak height variant for the HL97− allele was observed in the African-American and Asian populations at frequencies of 0.013 and 0.012, respectively. The low peak height insertion allele variant was observed by both Alvarez et al. [32] at similar frequencies (0.22, 0.044, 0.06, and 0.062, respectively) and Friis et al. [31] (although no frequencies were provided), but neither observed the low peak height deletion allele variant.

Low peak height variants for both the HLD83− and HLD83+ alleles were observed at frequencies of 0.04 and 0.026 in the African-American population and 0.025 and 0.025 in the Asian population, respectively. These microvariants also were described by Alvarez et al. [32] at similar frequencies, but not by Friis et al. [31].

Finally, three peak height microvariants were observed in the Asian population that have not been described elsewhere. At the HLD40+, HLD125+, and HLD6+ alleles, low peak height microvariants were observed at frequencies of 0.113, 0.113, and 0.073, respectively. These variants along with the others described above are being sequenced to determine the underlying cause of the variation. Additional variants are likely to be observed as population sample sizes increase or as additional populations are sampled.

Conclusions

The Investigator DIPplex® kit enables the simultaneous amplification and analysis of 30 small indels in a manner similar to that of STR analyses. Indels have several features that make them highly desirable for forensic applications. They are: small amplicon size, no stutter, low mutation rates, simplified analytical procedures, and the ability to be analyzed using current technology in crime laboratories. At the population level, the loci overall meet HWE and LE expectations. The data herein support that the Investigator DIPplex® kit provides a powerful supplement or stand-alone capability for human identity testing.

Special Acknowledgments This work was supported in part by the IC Postdoctoral Fellowship program. We also specially thank Qiagen (Hilden, GE) for providing the Investigator DIPplex kits.

Ethical standards The work described above was performed in accordance with all laws (both Federal and State) that apply to research, researcher conduct, and the protection of human test subjects. We also are operating under the guidance of and in accordance with policies of the UNTHSC Institutional Review Board.

References

1. Andersen JF, Greenhalgh M, Butler H, Kilpatrick S, Piercy R, Way K, Myhill H, Wright J, Hallett R, Parkin B (1996) Further validation of a multiplex STR system for use in routine forensic identity testing. Forensic Sci Int 78(1):47–64
2. Brinkmann B, Klintschar M, Neuhuber F, Hühne J, Rolf B (1998) Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat. Am J Hum Genet 62(6):1408–1415
3. Collins PJ, Hennessy LK, Leibelt CS, Roby RK, Reeder DJ, Foxall PA (2004) Developmental validation of a single-tube Amplification of the 13 CODIS STR loci, D2S1338, D19S433, and amelogenin: the AmpFSTR® Identifiler® PCR Amplification Kit. J Forensic Sci 49(6):1265–1277

4. Gill P, Whitaker J, Flaxman C, Brown N, Buckleton J (2000) An investigation of the rigor of interpretation rules for STRs derived from less than 100 pg of DNA. Forensic Sci Int 112(1):17–40
5. Junge A, Lederer T, Braunschweiger G, Madea B (2003) Validation of the multiplex kit genRESMPX-2 for forensic casework analysis. Int J Leg Med 117(6):317–325
6. LaFountain MJ, Schwartz MB, Svete PA, Walkinshaw MA, Buel E (2001) TWGDAM validation of the AmpFeSTR Profiler Plus and AmpFeSTR COfiler STR multiplex systems using capillary electrophoresis. J Forensic Sci 46(5):1191–1198
7. Moretti T, Baumstark A, Defenbaugh D, Keys K, Smerick J, Budowle B (2001) Validation of short tandem repeats (STRs) for forensic usage: performance testing of fluorescent multiplex STR systems and analysis of authentic and simulated forensic samples. J Forensic Sci 46(3):647
8. Micka KA, Sprecher CJ, Lins AM, Comey CT, Koons BW, Crouse C, Endean D, Pirelli K, Lee SB, Duda N (1996) Validation of multiplex polymorphic STR amplification sets developed for personal identification applications. J Forensic Sci 41:582–590
9. Fondevila M, Phillips C, Naverán N, Cerezo M, Rodríguez A, Calvo R, Fernández L, Carracedo Á, Lareu M (2008) Challenging DNA: assessment of a range of genotyping approaches for highly degraded forensic samples. Forensic Sci Int: Genet Suppl Ser 1 (1):26–28
10. Burger J, Hummel S, Herrmann B, Henke W (1999) DNA preservation: a microsatellite DNA study on ancient skeletal remains. Electrophoresis 20(8):1722–1728
11. Golenberg EM, Bickel A, Weihs P (1996) Effect of highly fragmented DNA on PCR. Nucleic Acids Res 24(24):5026–5033. doi:10.1093/nar/24.24.5026
12. Forster L, Thomson J, Kutranov S (2008) Direct comparison of post-28-cycle PCR purification and modified capillary electrophoresis methods with the 34-cycle "low copy number" (LCN) method for analysis of trace forensic DNA samples. Forensic Sci Int: Genet 2(4):318–328
13. Kloosterman AD, Kersbergen P (2003) Efficacy and limits of genotyping low copy number DNA samples by multiplex PCR of STR loci. Int Congr Ser 1239:795–798
14. Budowle B, van Daal A (2008) Forensically relevant SNP classes. BioTechniques 44:603–610
15. Dixon LA, Murray CM, Archer EJ, Dobbins AE, Koumi P, Gill P (2005) Validation of a 21-locus autosomal SNP multiplex for forensic identification purposes. Forensic Sci Int 154(1):62–77
16. Kidd KK, Pakstis AJ, Speed WC, Grigorenko EL, Kajuna SLB, Karoma NJ, Kungulilo S, Kim J-J, Lu R-B, Odunsi A, Okonofua F, Parnas J, Schulz LO, Zhukova OV, Kidd JR (2006) Developing a SNP panel for forensic identification of individuals. Forensic Sci Int 164(1):20–32
17. Phillips C, Fang R, Ballard D, Fondevila M, Harrison C, Hyland F, Musgrave-Brown E, Proff C, Ramos-Luis E, Sobrino B, Carracedo A, Furtado MR, Court DS, Schneider PM (2007) Evaluation of the Genplex SNP typing system and a 49plex forensic marker panel. Forensic Sci Int: Genet 1(2):180–185
18. Divne A-M, Allen M (2005) A DNA microarray system for forensic SNP analysis. Forensic Sci Int 154(2–3):111–121
19. Sanchez JJ, Phillips C, Børsting C, Balogh K, Bogus M, Fondevila M, Harrison CD, Musgrave-Brown E, Salas A, Syndercombe-Court D, Schneider PM, Carracedo A, Morling N (2006) A multiplex assay with 52 single nucleotide polymorphisms for human identification. Electrophoresis 27(9):1713–1724. doi:10.1002/elps.200500671
20. Edelmann J, Hering S, Augustin C, Szibor R (2009) Indel polymorphisms—an additional set of markers on the X-chromosome. Forensic Sci Int: Genet Suppl Ser 2(1):510–512
21. Francez PAdC, Ribeiro-Rodrigues EM, dos Santos SEB (2012) Allelic frequencies and statistical data obtained from 48 AIM INDEL loci in an admixed population from the Brazilian Amazon. Forensic Sci Int: Genet 6(1):132–135
22. Li C, Zhao S, Zhang S, Li L, Liu Y, Chen J, Xue J (2011) Genetic polymorphism of 29 highly informative InDel markers for forensic use in the Chinese Han population. Forensic Sci Int: Genet 5(1): e27–e30
23. Pereira R, Phillips C, Alves C, Amorim A, Carracedo Á, Gusmão L (2009) Insertion/deletion polymorphisms: a multiplex assay and forensic applications. Forensic Sci Int: Genet Suppl Ser 2(1):513–515
24. Lewis PO, Zaykin D (1999) Genetic Data Analysis: computer program for the analysis of allelic data, version 1.0 (d13). Free program distributed by the authors over the internet from the GDA homepage at http://chee.unm.edu/gda/
25. Jones DA (1972) Blood samples: probability of discrimination. J Forensic Sci Soc 12(2):355–359
26. Weir B, Cockerham CC (1984) Estimating *F*-statistics for the analysis of population structure. Evolution 38(6):1358–1370
27. Chakraborty R (2000) Linkage disequilibrium: concept, utility and evolutionary dynamics in the context of the human genome variation. In: Destobio 2000, West Lafayette, IND
28. Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics 164(4):1567
29. Weir BS (1996) Genetic data analysis II, vol 2. Sinauer Associates, Sunderland
30. Weber JL, David D, Heil J, Fan Y, Zhao C, Marth G (2002) Human diallelic insertion/deletion polymorphisms. Am J Hum Genet 71 (4):854–862
31. Friis SL, Børsting C, Rockenbauer E, Poulsen L, Fredslund SF, Tomas C, Morling N (2012) Typing of 30 insertion/deletions in Danes using the first commercial indel kit—Mentype® DIPplex. Forensic Sci Int: Genet (in press)
32. Alvarez MF PR, Gusmao L, Phillips C, Butler J, Lareu MV, Carracedo A, Vallone PM (2011) Forensic Performance of Short Amplicon Insertion-Deletion (INDEL) Markers. In: 22nd International Symposium on Human Identification, National Harbor, MD, 5 October 2011