

Application of the new insertion–deletion polymorphism kit for forensic identification and parentage testing on the Czech population

Anastasiya Zidkova · Ales Horinek · Vera Kebrdlova · Marie Korabecna

Received: 27 May 2011 / Accepted: 11 November 2011 / Published online: 29 November 2011
© The Author(s) 2011. This article is published with open access at Springerlink.com

Abstract Insertion–deletion polymorphisms (INDELs) are diallelic markers derived from a single mutation event. Their low mutation frequency makes them suitable for forensic and parentage testing. The examination of INDELs thus combines advantages of both short tandem repeats (STR) and single nucleotide polymorphisms (SNP). This type of polymorphisms may be examined using as small amplicon size as SNP (about 100 bp) but could be analyzed by techniques used for routine STR analysis. For our population study, we genotyped 55 unrelated Czech individuals. We also genotyped 11 trios to analyze DIPplex Kit (QIAGEN, Germany) suitability for parentage testing. DIPplex Kit contains 30 diallelic autosomal markers. INDELs in DIPplex Kit were tested with linkage disequilibrium test, which showed that they could be treated as independent markers. All 30 loci fulfill Hardy–Weinberg equilibrium. There were several significant differences between Czech and African populations, but no significant ones within European population. Probability of a match in the Czech population was 1 in 6.8×10^{12} ; combined power of discrimination was 99.9999999999%. Average paternity index was 1.13–1.77 for each locus; combined paternity index reached about 27,000 for a set of 30 loci. We can conclude that DIPplex kit is useful as an

additional panel of markers in paternity cases when mutations in STR polymorphisms are present. For application on degraded or inhibited samples, further optimization of buffer and primer concentrations is needed.

Keywords INDEL polymorphism · Parentage testing · Population study · Forensic genetics · PCR multiplex

Introduction

In the last 20 years, enormous progress was made in forensic science, mainly due to the development of DNA typing technologies. Forensic DNA analysis started from variable number tandem repeats, which, however, turned to be not suitable for degraded DNA. Nowadays, short tandem repeats (STR) are used in most forensic laboratories around the world. Many commercial STR kits with validated population data are currently available for routine forensic identification. Nevertheless, the relatively high size of amplicons (150–500 bp) makes the most of them unsuitable for degraded DNA and low copy number samples. Two approaches are used to analyze such samples: mini STR application (amplicons <150 bp) and single nucleotide polymorphisms (SNP), whose amplicons could be much shorter (<50 bp).

Recently, forensic scientists turn their attention to a different class of polymorphisms in human genome—insertion–deletion polymorphisms (INDELs). These polymorphisms originate from single mutation effect and have low mutation rate in comparison with STRs. They are present in the entire human genome, approximately one INDEL per 7.2 kbp [1]. The exact number of INDELs in human genome is unknown because results differ between studies, most likely due to the use of

Electronic supplementary material The online version of this article (doi:10.1007/s00414-011-0649-3) contains supplementary material, which is available to authorized users.

A. Zidkova · A. Horinek · V. Kebrdlova · M. Korabecna (✉)
First Faculty of Medicine, Charles University in Prague,
Prague, Czech Republic
e-mail: marie.korabecna@lfp.cuni.cz

A. Zidkova · A. Horinek · M. Korabecna
Faculty of Medicine in Pilsen, Charles University in Prague,
Pilsen, Czech Republic

different techniques for INDEL detection [2]. INDELS represent about 20% of all polymorphisms in human genome [3]: high proportion of them (36%) maps within the promoters, introns, and exons of known genes [1]. Mills et al. [1] identified five major INDEL classes: (1) insertions or deletions of single base pairs, (2) monomeric base pair expansions, (3) multi-base pair expansions of 2–15 bp repeat units, (4) transposon insertions, and (5) INDELS containing random DNA sequences ranging from 2 to 9,989 bp in length. Approximately 40% of INDELS belong to the last class; more than 99% of them are shorter than 100 bp [1].

INDELS like most of SNPs are diallelic. The size range of amplicons used for forensic examination of SNPs and INDELS is also comparable (50–150 bp) [4–6]. INDELS with their reduced mutation rates [7] are particularly appropriate in paternity cases where mutation in STR loci may occur [6]. INDELS can also be analyzed as STRs with no change of laboratory workflow and can be highly multiplexed. Despite indisputable INDEL advantages, the application of INDELS in routine forensic investigation and paternity testing remains still rare.

Here we present validation study of new INDEL kit—DIPplex (QIAGEN, Germany). DIPplex kit contains 30 INDEL markers, which are located almost on all chromosomes, and amelogenin locus as sex-informative marker (see [Electronic supplementary material](#)). This paper contains population data obtained on 55 unrelated Czech population members. These data were used for calculation of forensic and population parameters. Finally, we analyzed 11 trios (mother–child–alleged father) to test DIPplex kit suitability for parentage testing.

Materials and methods

Samples and DNA extraction

The study was carried out according to the SWGDAM guidelines [8]. Only samples with complete DNA profile were used for population study. For forensic parameters determinations, complete genotypes of 55 unrelated Czech population members were used. For paternity casework 11 mother–child–alleged father trios with proven paternity based on microsatellite analysis (PowerPlex 16 HS, Promega, USA) were randomly selected. DNA was isolated from buccal swabs using QIAmp Blood Mini Kit (QIAGEN, Germany).

PCR amplification and genotyping

PCR reactions were prepared according to the protocol recommended by the manufacturer of the DIPplex Kit in a

Veriti 96-Well Thermal Cycler (Applied Biosystems, USA). Several concentrations of DNA were tested (0.2, 0.5, and 1 ng); best results were obtained with 0.5 ng of DNA. XY5 was used as a control DNA to test performance of the DIPplex Kit.

PCR products were subsequently analyzed on capillary electrophoresis. Samples were mixed with Hi-Di Formamide (Applied Biosystems) and BTO 550 size standard (QIAGEN). Fragment detection and separation were performed using ABI PRISM 310 Genetic Analyzer, G5 filter set, and POP4 polymer. Samples were then genotyped using Genotyper v3.7 software (Applied Biosystems).

Statistical analysis

Allele frequencies were determined by counting. Pearson's χ^2 test for R×C and Fisher's exact test with subsequent Bonferroni correction for multilocus testing were employed to test the deviation of allele frequencies from Hardy–Weinberg equilibrium [9]. Gene diversity, polymorphism information content, homozygosity, heterozygosity, and within-population inbreeding coefficient were determined [10]. Probability of a match (PM), power of discrimination (PD), and average paternity index (API) were calculated for each locus and subsequently for whole multiplex as well. Chi-square test was used to calculate differences of allele frequency between Czech populations and other populations (German, European, Asian, African, and Native American). All statistical parameters mentioned above were calculated using in-house-developed applications. Linkage disequilibrium analysis was performed using SNPAnalyzer v1.2 (Istech, South Korea).

Results and discussion

In this study we tested: (1) DIPplex Kit eligibility for routine use in forensic laboratories, (2) suitability of DIPplex Kit for application in Czech population, and (3) INDELS applicability for parentage testing.

Multiplex performance

We realized that DIPplex PCR multiplex is very sensitive and requires very low amount of DNA. However, samples with high DNA concentrations (more than 0.5 ng per PCR reaction) cause pull-up peaks and abnormally shaped peaks. One of the most frequent reasons for sample exclusion from population data were allele and locus dropouts occurred in several loci (rs1610905, rs2307956, and rs1610937). Interestingly, there was no correlation between amplicon size and allele or locus dropout. Further optimization of

Table 1 Comparison of combined probability of a match for different INDEL sets and STR kits

Multiplex	Type of markers	Number of markers	Population	PM
PowerPlex 16	STR	15	Caucasian	1 in 1.8×10^{17}
Pimenta and Pena, 2010 [6]	INDEL	40	Caucasian	1 in 2.9×10^{16}
Pereira et al. 2009 [5]	INDEL	38	Caucasian	1 in 2.8×10^{14}
Li et al. 2011 [4]	INDEL	29	Asian	1 in 1.1×10^{11}
This study	INDEL	30	Caucasian (Czech)	1 in 6.8×10^{12}

PM probability of a match

buffer content and primer concentration would be very helpful in this context.

Genetic variation

Allele frequencies and percentages of homozygotes and heterozygotes in each locus were tested to detect whether loci are in Hardy–Weinberg equilibrium. We used Pearson's test which showed no deviation from Hardy–Weinberg expectation. We also used Fisher's exact test; *P* value was adjusted by Bonferroni correction for 30 loci and $P < 0.00167$ was considered statistically significant (see [Electronic supplementary material](#)). There was no deviation for all 30 loci from Hardy–Weinberg proportions. Average heterozygosity was 0.457. Average gene diversity reached 0.496. All 30 loci have minor allele frequency higher than 0.25. Obtained data indicate that all selected loci are highly polymorphic.

Population comparison

Allele frequencies of 24 loci in DIPplex kit in four populations (general European, German, American/Indian, Asian/Japanese, African) were compared with our results using chi-square test at 0.05 significance level (see [Electronic supplementary material](#)). Loci rs17879936, rs8190570, rs17174476, rs17878444, rs3081400, and rs2307433 were tested only in our study and on the German population. We detected statistically nonsignificant differences between Czech and European populations and Czech and Asian ones. After comparison of Czech and American/Indian population, statistically significant difference was found in rs16388. In African population, there were two loci (rs1610937, rs6481) in which allele frequencies are statistically different from the Czech population.

We also compared four INDEL loci in DIPplex kit (rs2307959, rs1305047, rs16438, and rs1611001); they were involved in the study by Pimenta and Pena [6]. The differences between our study and published results were nonsignificant. These findings indicate that selected markers have high discrimination power among different populations. This is an essential assumption for a robust forensic kit.

Linkage disequilibrium

To test whether loci that are located on the same chromosome show allelic association, we applied linkage disequilibrium test. SNPAnalyzer v1.2 as an internet-based software was used. This test showed no significant allelic association between all 30 markers, so they could be treated as independent ones.

Forensic efficiency

Forensic parameters were also evaluated in this study. We tested the efficiency of DIPplex Kit by PM and PD. PM was 1 in 6.8×10^{12} , which is lower than PM for STR kits, but it is comparable with other INDEL-based studies (Table 1) [4–6]. Combined PD reached 99.999999999% (see [Electronic supplementary material](#)). In every locus API was higher than one, therefore proving usefulness of selected loci in paternity testing. Combined PI was about 27,000, so probability of paternity assuming 50% prior probability was 99.99%. Although obtained value is less than probability of paternity for Combined DNA Index System (CODIS) loci, it is still sufficient for paternity to be practically proven. For example, probability of paternity in one case increased from 99.99999 to 99.9999999999 by combining both STR (PowerPlex 16 HS) and INDEL (DIPplex) analysis (Table 2). Despite that we did not perform linkage disequilibrium test between STR markers in PowerPlex 16 HS and INDEL markers in DIPplex Kit, minimal 10-Mb distance between INDEL and STR markers is in our opinion enough for treating all markers as

Table 2 Example of combining STR and INDEL analysis for paternity testing

Multiplex	CPI	W, % ^a
PowerPlex 16	54116364	99.99999
DIPplex	49366	99.99
Combined	2.67×10^{12}	99.9999999999

CPI combined paternity index, W probability of paternity

^a Assuming 50% prior probability

independent ones. Therefore DIPplex kit could be used independently in simple trios but more probably as an additional panel of markers in complicated kinship analysis [11]. With regard to the problem of paternity tests, we also measured informativity of this marker set for kinship testing. Since marker set informativity equals twice the amount of INDELs multiplied by their average gene diversity [12], we received the value 29.76 in our case, which equals to the amount of markers in the DIPplex kit. Marker set informativity of the DIPplex kit in the Czech population reveal an exclusion power at a similar level of 29.76 maximally informative SNPs. The published studies employed similar numbers of INDEL markers, Pimenta and Pena [6] used 40 markers, Pereira et al. [5] 38, and Li et al. [4] 29 markers. As the markers involved in these studies were different with no or minimal overlap, it seems that the forensic parameters are dependent mainly on the total number of examined INDELs (Table 1).

Conclusions

In this study we presented validation of a new DIPplex Kit (QIAGEN) for the use in the Czech population and we compared our data with few existing INDEL-based studies published till today [4–6]. The kit contains a panel of 30 INDEL loci. Since we found no LD between all 30 markers, the whole markers panel can be treated as an independent loci set. No significant differences in allele frequencies from European and Asian populations were detected. The API and CPI values achieved in our study led to the conclusion that the DIPplex Kit could be applied as an addition to STR markers for solution of complicated kinship cases, thus increasing power of DNA evidence. The DIPplex Kit could be used as an independent assay for paternity testing with the reduced risk of mutation events during the gametogenesis. However, occasionally paternity indices obtained using the DIPplex kit only may not reach levels comparable to microsatellite CODIS loci. After optimization of buffer content and primer design by the manufacturer, the DIPplex Kit could be used also for analysis of degraded and/or inhibited samples. Data obtained from DIPplex kit could be analyzed by tools used for routine STR analysis and do not require any change in laboratory workflow.

Acknowledgments This study was supported by the Grant from Ministry of Trade and Industry of Czech Republic no. I/398.

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

1. Mills RE, Luttig CT, Larkins CE, Beauchamp A, Tsui C, Pittard WS et al (2006) An initial map of insertion and deletion (INDEL) variation in the human genome. *Genome Res* 16(9):1182–1190. doi:10.1101/gr.4565806
2. Mullaney JM, Mills RE, Pittard WS, Devine SE (2010) Small insertions and deletions (INDELs) in human genomes. *Hum Mol Genet* 19(2):131–136. doi:10.1093/hmg/ddq400
3. 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. doi:10.1086/342727
4. Li C, Zhao S, Zhang S, Li L, Liu Y, Chen J et al (2011) Genetic polymorphism of 29 highly informative InDel markers for forensic use in the Chinese Han population. *For Sci Int Genet* 5(1):e27–e30. doi:10.1016/j.fsigen.2010.03.004
5. Pereira R, Phillips C, Alves C, Amorim A, Carracedo A, Gusmão L (2009) A new multiplex for human identification using insertion/deletion polymorphisms. *Electrophoresis* 30(21):3682–3690. doi:10.1002/elps.200900274
6. Pimenta JR, Pena SDJ (2010) Efficient human paternity testing with a panel of 40 short insertion–deletion polymorphisms. *Genet Mol Res* 9(1):601–607. doi:10.4238/vol9-1gmr838
7. Nachman MW, Crowell SL (2000) Estimate of the mutation rate per nucleotide in humans. *Genetics* 156:297–304
8. FBI Web (2004) Forensic Science Communications—July 2004. Available at http://www.fbi.gov/about-us/lab/forensic-science-communications/fsc/july2004/standards/2004_03_standards02.htm/ Accessed July 2004
9. Guo SW, Thompson EA (1992) Performing the exact test of Hardy–Weinberg proportion for multiple alleles. *Biometrics* 48(2):361–372
10. Buckleton J, Triggs C, Walsh S (2005) Forensic DNA evidence interpretation. CRC Press, Boca Raton
11. Phillips C, Fondevila M, García-Magariños M, Rodríguez A, Salas A, Carracedo A et al (2008) Resolving relationship tests that show ambiguous STR results using autosomal SNPs as supplementary markers. *Forensic Sci Int Genet* 2(3):198–204. doi:10.1016/j.fsigen.2008.02.002
12. Krawczak M (1999) Informativity assessment for biallelic single nucleotide polymorphisms. *Electrophoresis* 20(8):1676–1681. doi:10.1002/(SICI)1522-2683(19990101)20:8<1676::AID-ELPS1676>3.0.CO;2-D