A multiplex panel of short-amplicon insertion-deletion DNA polymorphisms for forensic analysis

V.R.D. Santos[1,2], H.B. Pena[1] and S.D.J. Pena[1,3]

[1]GENE - Núcleo de Genética Médica, Belo Horizonte, MG, Brasil
[2]Polícia Civil de Minas Gerais, Instituto de Criminalística de Minas Gerais,
Belo Horizonte, MG, Brasil
[3]Departamento de Bioquímica e Imunologia,
Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brasil

Corresponding author: S.D.J. Pena
E-mail: spena@gene.com.br

ABSTRACT. We have previously developed a panel of 40 insertion-deletion (INDEL) human DNA polymorphisms that was proven to adequately cover the span of global human genetic diversity. The panel was found to have very low matching probabilities with respect to both the global and Brazilian populations. To optimize the panel for application with degraded DNA samples, which are commonly encountered in forensic analysis, we have significantly reduced the amplicon size of the INDELs and developed a new multiplex panel. The panel has an amplicon size ranging from 50 to 153 base pairs, with a mean of 93 base pairs. It could be amplified by polymerase chain reaction in two multiplex reactions, which were then combined for electrophoretic separation and identification of the individual products in the ABI3130 four-color DNA analyzer. The results of the new panel were fully validated.

Key words: Forensic DNA; Insertion-deletion polymorphisms; DNA; INDELs

INTRODUCTION

The multiplex analysis of microsatellites, also known as short tandem repeats, has become the golden standard for human forensic identification and paternity testing. The hypervariable nature of microsatellites endows them with the highest discrimination power for human DNA profiling among all markers (Butler, 2005). However, the high number of alleles limits the minimum amplicon size of microsatellites. Since the DNA in forensic samples is often degraded, contaminated, or present in low quantities, the relatively large amplicon sizes of microsatellites can cause stochastic effects or failure in DNA amplification, which reduces the possibility of successful identification of samples obtained from crime scenes (Butler et al., 2003).

Two types of DNA markers are more compatible with lower amplicon sizes than microsatellites: single nucleotide polymorphisms (SNPs) and short (1-20 nucleotide pairs) insertions and deletions (INDELs). Both of these polymorphism types are biallelic, which allows them to be amplified by polymerase chain reaction (PCR) with relatively small amplicon sizes. However, their biallelic property also means that these markers are individually less informative than microsatellites, which can be compensated for by using a larger number of markers (Pimenta and Pena, 2010). The typing of SNPs, which depends on qualitative base identification, is best accomplished with microarray hybridization; however, this requires complex equipment that is not generally available in identification laboratories. On the other hand, since the alleles of INDELs differ in size they can be typed using the same methodological approaches and equipment currently utilized for microsatellite genotyping, which are already operational in forensic laboratories.

Several million short INDELs have been described in the human genome, which represent the second most abundant form of genetic variation in humans after SNPs (Mullaney et al., 2010). In 2002, Weber et al. (2002) characterized 2000 human biallelic short INDELs in the human genome. Among them, we identified 40 polymorphisms that fulfilled the following criteria: widespread chromosomal distribution, increasing amplicon sizes to allow multiplex analysis, and allele frequencies close to 0.50 in the European population (Bastos-Rodrigues et al., 2006). We have shown that this set of 40 carefully chosen INDELs was sufficiently informative for adequate characterization of the human population structure at the global level (Bastos-Rodrigues et al., 2006). Moreover, we demonstrated the excellent resolution power of these markers in discriminating among Europeans, Africans, and Amerindians, which could be applied to the analysis of the ancestral roots of more than 1000 individuals of the Brazilian population (Pena et al., 2009, 2011) and in pharmacogenetic studies (Suarez-Kurtz et al., 2007).

During the development of these studies, it soon became clear that our 40-INDEL multiplex panel also had excellent capacity for individual identification, with a combined probability of identity (match probability) of 5.07×10^{-17} for Europeans, which indicated its excellent potential as a forensic tool (Pena, 2005). Extremely low match probabilities were similarly observed for populations from other geographical regions, including Brazil (Pimenta and Pena, 2010; Pena and Pena, 2012).

We further explored the potential application of the 40-INDEL panel for forensic studies by genotyping the markers in 360 unrelated self-classified White Brazilians and 50 mother-child-probable father trios with proven paternity (Pimenta and Pena, 2010). The average heterozygosity (gene diversity) per locus was 0.48 and the combined probability of identity (matching probability) for the 40-locus set in Brazilian individuals self-classified as White was 3.5×10^{-17}. For paternity testing, the combined power of exclusion of the INDEL panel

was 0.9997. The geometric mean of the paternity indices of the 50 mother-child-probable father trios was 17,607. Thus, the 40-INDEL panel represents a useful new tool as an adjunct to routine microsatellite multiplex sets for human paternity testing, especially given that it is composed of genomic markers with very low mutation rates.

Despite the excellent capacity of our 40-INDEL multiplex panel for individual identification (Pena, 2005), the average amplicon size was 162 base pairs (bp), with a range of 95-336 bp. These amplicons are excessively large, especially considering that the DNA of forensic samples is often degraded and/or available in limited quantities. We here describe the successful development of short amplicons for the 40-locus INDEL set. The new average amplicon size is 93 bp, with a range of 50-153 bp.

MATERIAL AND METHODS

The 40 INDELs used in the multiplex panel were: rs3917, rs4181, rs4183, rs16343, rs16381, rs16394, rs16415, rs16416, rs16430, rs16438, rs16448, rs16695, rs16715, rs140709, rs140733, rs140757, rs140759, rs1305047, rs1610874, rs1610942, rs1610997, rs1611001, rs1611084, rs2067180, rs2067188, rs2067217, rs2067373, rs2307548, rs2307624, rs2307733, rs2307745, rs2307782, rs2307838, rs2307850, rs2307955, rs2307959, rs2308043, rs2308057, rs2308135, and rs2308144 (Bastos-Rodrigues et al., 2006, Pena et al., 2011).

The short-amplicon INDELs were separated into four groups: "SET1" (12 amplicons fluorescently labeled with NED), "SET2" (9 amplicons fluorescently labeled with 6-FAM), "SET3" (10 amplicons fluorescently labeled with PET), and "SET4" (9 amplicons fluorescently labeled with VIC) (Applied Biosystems).

The fluorescent labeling of the amplicons was performed using Schuelke (2000) economic nested protocol, as follows: for SET1, the forward primer had a tail of the M13(-40) sequence (5'-GTTTTCCCAGTCACGAC-3') and the reaction was performed in the presence of an M13(-40) primer labeled with NED; for SET2, the forward primer had a tail of an M13-universal sequence (5'-CACGACGTTGTAAAACGAC-3') and the reaction was performed in the presence of an M13-universal primer labeled with 6-FAM; for SET3, the forward primer had a tail of the M13-reverse sequence (5'-GTAAAACGACGGCCAGT-3') and the reaction was performed in the presence of an M13-universal primer labeled with PET; for SET4, the forward primer had a tail of the T7 sequence (5'-TAATACGACTCACTATAGGG-3') and the reaction was performed in the presence of a T7 primer labeled with VIC. The "pigtail" extension 5'-GTTTCTT-3' was added to the 5' end of all of the reverse primers to promote complete adenylation at the 3' end (Brownstein et al., 1996).

PCRs were performed on a VERITI 96-well thermal cycler (Applied Biosystems, Foster City, CA, USA) using the following program: 1) initial denaturation at 95°C for 5 min; 2) 94°C for 1 min; 3) annealing at 60-50°C for 1 min (reducing by 1°C per cycle); 4) extension at 72°C for 2 min. Once the annealing temperature reached 50°C, this cycle was repeated 28 times. The program was terminated with a final extension at 72°C for 5 min. Amplified products were stored at 4°C.

Electrophoresis and typing were performed on an ABI 3130 Genetic Analyzer (Applied Biosystems) in total volume of 11 µL, containing 8.8 µL highly deionized formamide with 0.2 µL GeneScan 500 LIZ size standard (ABI) and 2 µL PCR product. The tube was denatured for 5 min at 95°C and immediately cooled to 4°C. POP4 polymer was used for electrophoresis (Applied Biosystems). The fragments were analyzed using the GeneMapper

I.D. Software 3.2 (Applied Biosystems). Typing quality and allele designation were verified by simultaneous electrophoretic analysis of a control sample of known size and sequence.

RESULTS AND DISCUSSION

We obtained good multiplex amplification of SET1 and SET2 as well as SET3 and SET4 in a single reaction tube, respectively. The amplification products were then combined and subjected to a single electrophoretic run. The size distribution of the several INDELs in the four sets is shown in Figure 1. The panel has an amplicon size ranging from 50 to 153 bp, with a mean of 93 bp.
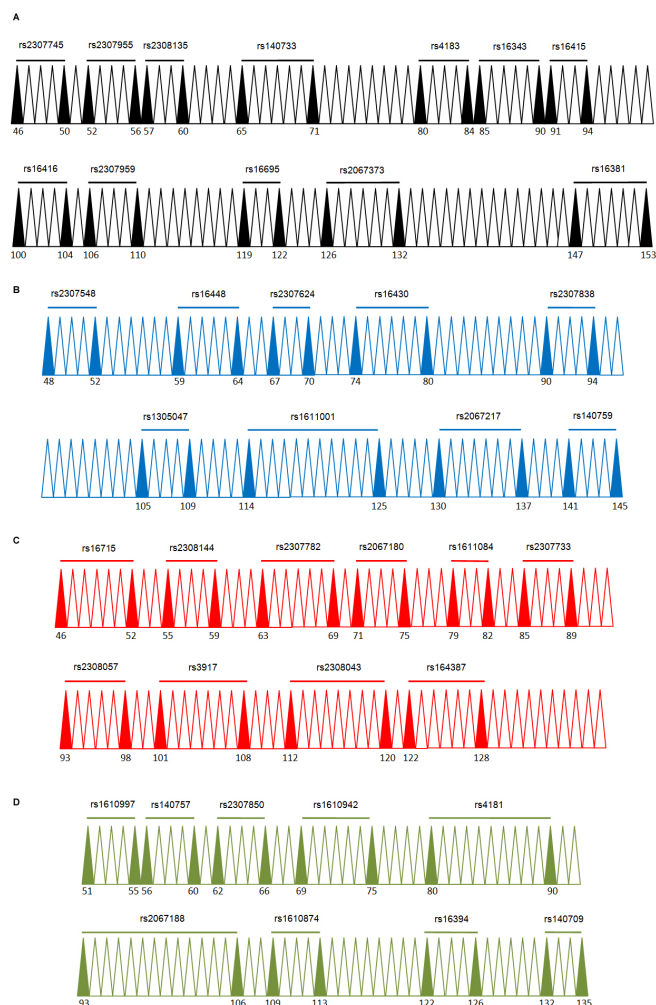

Figure 1. Size distribution of the several INDELs in the four short-amplicon sets: **A.** SET1, labeled with NED; **B.** SET2, labeled with 6-FAM; **C.** SET3, labeled with PET, and **D.** SET4, labeled with VIC. The colors in the figure are arbitrary, and do not represent the actual fluorophore colors. Each set is shown in two continuous lines. Amplicons are represented by filled triangles, with the RS identification indicated on top and the size in base pairs indicated below.

SET1 was labeled with the fluorophore NED and contained 12 INDELs: rs2307745, rs2307955, rs2308135, rs140733, rs4183, rs16343, rs16415, rs16416, rs2307959, rs16695, rs2067373, and rs16381. SET2 was labeled with the fluorophore 6-FAM and contained nine INDELs: rs2307548, rs16448, rs2307624, rs16430, rs2307838, rs1305047, rs1611001, rs2067270, and rs140759. SET3 was labeled with the fluorophore PET and contained 10 INDELs: rs16715, rs2308144, rs2307782, rs2067180, rs1611084, rs2307733, rs2308057, rs3917, rs2308043, and rs16438. SET4 was labeled with the fluorophore VIC and contained nine INDELs: rs1610997, rs140757, rs2307850, rs1610942, rs4181, rs2067188, rs1610874, rs16394, and rs140709. Two representative runs of SET2 using respective DNA samples from a pool of 200 individuals and from a single human subject are shown in Figure 2.
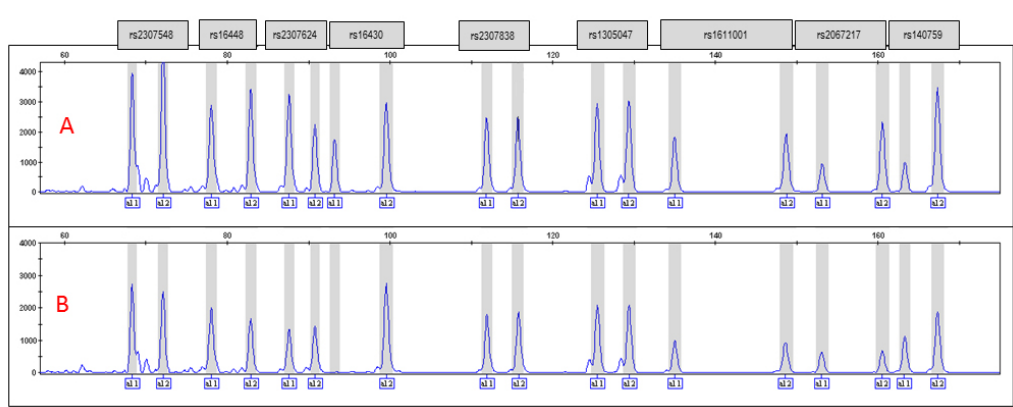

Figure 2. Representative runs of SET2 using respective DNA samples from a pool of 200 individuals (**A**) and from a single human subject (**B**). The INDELs are identified by their RS numbers. In the pooled sample, every allele was amplified and the peak height was influenced by the allele frequency.

We compared the typing of four unrelated individuals using the newly developed short-amplicon INDELs and the previously described multiplex INDEL set (Bastos-Rodrigues et al., 2006) and the results were identical. This, of course, matched our expectations, since the reduction of amplicon size occurred only in the region surrounding the INDEL and did not interfere with its sequence. Thus, we could apply the same allele frequency properties and match probabilities obtained for the regular-sized amplicons to the short-amplicon INDELs, which had been extensively validated and analyzed in several studies (Bastos-Rodrigues et al., 2006; Pimenta and Pena, 2010; Pena et al., 2009, 2011; Pena and Pena, 2012).

In conclusion, we were able to successfully generate a panel of 40 highly informative INDELs with short amplicons, varying from 50 to 153 bp, with a mean of 93 bp. We have also demonstrated that this panel is able to efficaciously generate useful profiles from degraded forensic samples (unpublished data). We firmly believe that the multiplex panel of 40 short amplicon INDELs described herein represents a useful new tool for forensic identification.

ACKNOWLEDGMENTS

The specific set of 40 INDEL polymorphisms and its genotyping methodologies are the subject of a Patent Application made by GENE - Núcleo de Genética Médica at the Institute Nacional de Propriedade Industrial (INPI) of Brazil.

REFERENCES

Bastos-Rodrigues L, Pimenta JR and Pena SD (2006). The genetic structure of human populations studied through short insertion-deletion polymorphisms. *Ann. Hum. Genet*. 70: 658-665.

Brownstein MJ, Carpten JD and Smith JR (1996). Modulation of non-templated nucleotide addition by Taq DNA polymerase: primer modifications that facilitate genotyping. *Biotechniques* 20: 1004-1010.

Butler JM (2005). Forensic DNA typing. 2nd edn. Elsevier, Amsterdam.

Butler JM, Shen Y and McCord BR (2003). The development of reduced size STR amplicons as tools for analysis of degraded DNA. *J. Foren. Sci*. 48: 1054-1064.

Mullaney JM, Mills RE, Pittard WS and Devine SE (2010). Small insertions and deletions (INDELs) in human genomes. *Hum. Mol. Genet*. 19: R131-R136.

Pena SDJ (2005). Segurança pública: determinação de identidade genética pelo DNA. *Parcer. Estrateg*. 20: 447-460.

Pena HB and Pena SDJ (2012). Automated genotyping of a highly informative panel of 40 short insertion-deletion polymorphisms resolved in polyacrylamide gels for forensic identification and kinship analysis. *Transfus. Med. Hemother*. 39: 211-216.

Pena SDJ, Bastos-Rodrigues L, Pimenta JR and Bydlowski SP (2009). DNA tests probe the genomic ancestry of Brazilians. *Braz. J. Med. Biol. Res*. 42: 870-876.

Pena SDJ, Di Pietro G, Fuchshuber-Moraes M, Genro JP, et al. (2011). The genomic ancestry of individuals from different geographical regions of Brazil is more uniform than expected. *PLoS One* 16: e17063.

Pimenta JR and Pena SDJ (2010). Efficient human paternity testing with a panel of 40 short insertion-deletion polymorphisms. *Genet. Mol. Res.* 30: 601-607.

Schuelke M (2000). An economic method for the fluorescent labeling of PCR fragments. *Nat. Biotechnol.* 18: 233-234.

Suarez-Kurtz G, Vargens DD, Struchiner CJ, Bastos-Rodrigues L, et al. (2007). Self-reported skin color, genomic ancestry and the distribution of GST polymorphisms. *Pharmacogenet. Genomics* 17: 765-771.

Weber JL, David D, Heil J, Fan Y, et al. (2002). Human diallelic insertion/deletion polymorphisms. *Am. J. Hum. Genet.* 71: 854-862.