

## Capillary Electrophoresis of 38 Noncoding Biallelic Mini-Indels for Degraded Samples and as Complementary Tool in Paternity Testing

Rui Pereira and Leonor Gusmão

### Abstract

This work describes the main advantages and the steps involved in the optimization of a multiplex system able to characterize 38 noncoding biallelic *Insertion Deletion Polymorphisms* (Indels). With this methodology, all markers are amplified in a single PCR, using short amplicons (up to 160 bp) in order to improve its performance in degraded samples. Alleles are easily detected using capillary electrophoresis.

The Indel multiplex typing strategy here described has the same desirable characteristics of forensic SNP assays, including genetic markers (a) with low mutation rates, increasing their usefulness in some kinship cases where few or single incompatibilities can be explained by mutation, and (b) that can be typed using a short amplicon strategy, increasing their usefulness in cases where degraded samples are available. Moreover, this approach uses simple and well-established methodologies already applied in forensic STR assays.

**Key words:** Insertion deletion polymorphism, Indel, DIP, Multiplex PCR, Human identification, Paternity testing, Forensic genetics

---

### 1. Introduction

The human DNA sequence encompasses distinct types of variation, each presenting desirable characteristics to address different biological questions. For example, STRs present high degrees of length polymorphism in human populations and are predominantly multiallelic. Because they are highly diverse, this type of genetic markers is very informative and widely used in population and forensic genetics. Conversely, mutational events such as nucleotide substitutions, insertions, or deletions occur with a very low frequency, possibly only once in the recent evolutionary past of modern humans. By their nature, these mutations create polymorphisms

with two allelic states (ancestral and mutant) and are called binary or biallelic polymorphisms. These genetic markers present lower genetic diversity than STRs, which is a disadvantage regarding forensic applications, but they also have several desirable features as genetic markers in this field. The usefulness of SNPs in forensics is well patent in many publications (e.g., see ref. (1–6)) or in other chapters of this book, and here we dedicate special attention to insertion deletion (Indel) polymorphisms, particularly in the analysis of degraded samples and as complementary tool in paternity testing.

### **1.1. Insertion Deletion Polymorphisms**

The studies by Weber et al. (7) and Mills et al. (8) contributed fundamentally to our knowledge on Indel polymorphisms and highlighted the utility of Indels for most genetic studies given their abundance in the genome and ease of analysis. In 2002, Weber et al. (7) identified and characterized ~2,000 human biallelic Indels with different length variation and measured the allele frequencies in Europeans, Africans, Japanese, and Native Americans. In association with this work, Marshfield Clinic created an Insertion/Deletion Polymorphism online database (<http://www.marshfieldclinic.org/mgs>) including information regarding the type of polymorphism (biallelic or multiallelic), allele length differences, amplification primers, position in the chromosome, as well as allele frequencies in different population groups. Later in 2006, an initial map of Indel variation in the human genome was reported (8), comprising more than 415,000 polymorphisms. According to Mills et al. (8), insertion/deletions represent ~16% of all human DNA polymorphisms and are widely spread throughout the genome, with an average density of one Indel per 7.2 kb. Moreover, a particular class of apparently random DNA sequences comprises about 41% of all Indels and shows a wide range of allele length variation, from 2 up to ~10 kb, with nearly 99% being under 100 bp (8). Of great interest as genetic markers, these small length Indels are abundant in the human genome and amenable to simple analysis through PCR amplification and electrophoresis, or using methodologies already developed for SNPs (7–9).

Another important aspect is that Indel data from these and other studies have been compiled in the free access online database dbSNP from NCBI (under the MARSHFIELD and DEVINE\_LAB submitter handle, respectively, for refs. (7) and (8)), which enables a much easier access to the available information.

In the last few years, there was an increasing interest of the scientific community in the use of Indel polymorphisms in different research projects, as for example when studying the genetic structure of human populations (10–12), as ancestry informative markers (9, 13), in X-chromosome specific panels (14–16), and more recently in human identification studies (17, 18).

### ***1.2. Insertion Deletion Polymorphisms as Genetic Markers in Forensics***

Considering the above-mentioned aspects, it becomes clear that some types of Indels combine interesting characteristics of other genetic markers commonly used in forensics (STRs and SNPs):

- Indels are abundant and widely distributed along the genome.
- The genotyping of small Indels is relatively easy and inexpensive with a simple PCR and electrophoresis approach.
- Small Indels can be analyzed in short amplicons, improving successful amplification of degraded DNA and, in the other hand, opening perspectives for good multiplexing levels.
- Small Indels are also suitable for automation and analysis with high-throughput technologies.
- Indels show very low mutation rates (occurring at low frequency, the polymorphism most likely derives from a single nonrecurrent mutation event in the evolutionary time of modern humans).
- Many Indels have significant divergences in allele frequencies among geographically separated population groups (therefore, they can be used as Ancestry Informative Markers).

### ***1.3. Mini-Indels in the Analysis of Degraded Samples***

Some of these characteristics are the basis behind the recent interest in Indels for forensic applications. The most relevant is the possibility of analysis in much shorter amplicons when compared with standard STRs. This was definitely the main reason for introducing SNPs in forensic practice, aiming to overcome the major drawback of STR analysis: its weak performance with highly degraded samples due to the long template DNA needed for amplification. In this work, we focus on mini-Indels with small length variation between alleles (2–5 bp), which likewise SNPs can be analyzed in very short amplicons (under 160 bp) thus improving the amplification success with challenging DNA (17, 19). The advantage of this approach is that the detection of alleles is accomplished simply by fragment separation with capillary electrophoresis, conversely to laborious SNP typing protocols. In summary, Indels ideally allow combining a short amplicon strategy (see Fig. 1a) with a simple analytic workflow (see Fig. 1b).

### ***1.4. Indels as Complementary Tool in Paternity Testing***

In paternity or other biological relationship testing situations, it can happen that few mendelian transmission incompatibilities are observed for STRs; and it is necessary to compare the probability of the different hypothesis assuming exclusion vs. the occurrence of a mutation. Nevertheless, after that assessment, the conclusion can sometimes remain unclear or need stronger support. Due to the high mutation rates of most STRs, this is not an uncommon scenario in paternity testing. In such cases, the complementary study of a large set of Indels or SNPs showing low mutation rates and therefore being more stable than microsatellites can help circumvent difficulties (4, 5, 19). Characterizing a high number of markers can

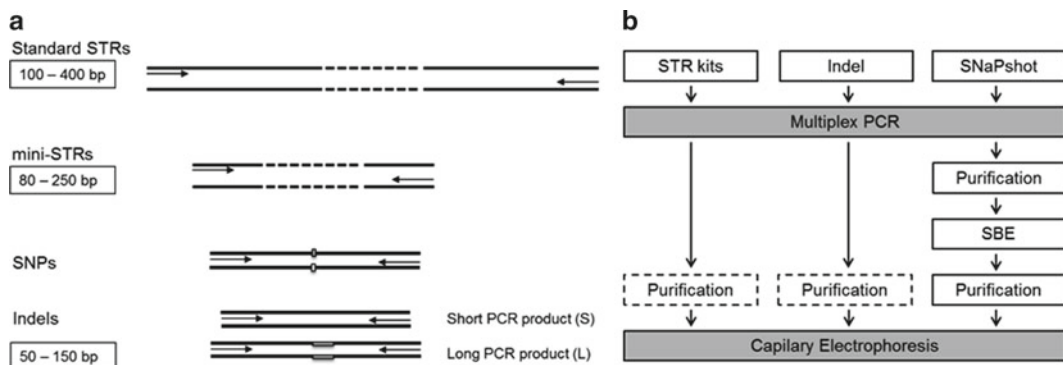


Fig. 1. Insertion deletion (Indel) polymorphism analysis in short amplicon and simple workflow: (a) Indels, like SNPs, are amenable to analysis in much shorter amplicons than standard STRs or even the re-engineered mini-STRs; (b) Indels, like STRs, can be easily typed with a PCR and capillary electrophoresis, conversely to the laborious protocol most widely used in forensic SNPs typing (SNaPshot; Applied Biosystems).

unveil further transmission incompatibilities, favoring the exclusion hypothesis or, in case of inclusion, increase the paternity index values to more confident levels supporting paternity (see Note 1).

### 1.5. General Considerations About the Indel Assay

The multiplex described in the present work includes 38 Indels that were previously found to be polymorphic and very informative in human population groups from Africa, Europe, and East Asia, with an average heterozygosity of 0.45 for the set. No statistically significant deviations from Hardy–Weinberg expectations were found in the studied populations and linkage disequilibrium tests also did not reveal statistically significant allelic association between the different loci pairs in the same populations (17).

Regarding forensic efficiency parameters, the 38 Indel set proved to be highly informative in African, European, and East Asian populations, with accumulated random match probabilities ranging in orders of magnitude from  $10^{-14}$  to  $10^{-15}$  and accumulated discrimination power reaching 99.999999999994% in the less diverse group. Despite being less powerful than standard STR kits (when full profiles are obtained), the Indel set appears to be enough informative in most forensic cases. Even more, it exceeds the efficiency parameters of the mini-STR kit Minifiler (Applied Biosystems), which likewise the mini-Indel multiplex is indicated to the analysis of highly degraded samples (17).

The analytical workflow of the Indel assay is very simple and, in comparison to the most widely available SNP typing methods, reduces considerably the steps needed to genotype a large set of informative biallelic markers for identification purposes. This is advantageous in terms of time and cost-effectiveness of the assay, and more importantly, it minimizes manipulation and risks of contamination or samples' mix up. Furthermore, the use of the same methodology as in standard STR genotyping makes this assay easy to implement in any forensic laboratory.

## 2. Materials

1. A number of free access online databases and computer programs were used in the selection of Indel markers and development of the multiplex here described (see Table 1 for details).

**Table 1**

**Online databases and computer programs used in developing the Indel multiplex**

*Diallelic insertion/deletion polymorphisms database from Marshfield Clinic Mammalian Genotyping Service*

<http://www.marshfieldclinic.org/mgs/>

Contains information regarding the type of polymorphism (biallelic or multiallelic), validation status, allele length differences, amplification primers, position in the chromosome, and allele frequencies in different population groups

*dbSNP or EntrezSNP*

<http://www.ncbi.nlm.nih.gov/projects/SNP/> or <http://www.ncbi.nlm.nih.gov/snp>

dbSNP is the NCBI database of genetic variation. It is now incorporated into NCBI's Entrez system and can be queried using the same approach as the other Entrez databases such as PubMed and GenBank. dbSNP also includes Indels (or DIPs) and integrates data from various submitters. Contains general information about the polymorphism, similarly to the Marshfield database, but with extra features and search possibilities. Using custom queries or the "Limits" tab allows restricting your search by chromosome, SNP class, and other criteria (such as heterozygosity, etc.), for example, to select markers with forensic relevant characteristics

*UCSC Genome Browser's "Get DNA in Window" tool*

<http://genome.ucsc.edu/cgi-bin/hgGateway>

The "Get DNA in Window" tool included in UCSC Genome Browser allows obtaining a DNA sequence provided the chromosome coordinates. From inside of your target assembly, set the region and submit. Then click on "DNA" in the top blue navigation bar. Of interest is the possibility of using extended case/color options to highlight relevant annotations in that sequence, for example, all SNPs included in dbSNP. This will allow knowing polymorphic positions to avoid during primer design

*Primer 3*

<http://frodo.wi.mit.edu/primer3/>

*BatchPrimer3*

<http://probes.pw.usda.gov/batchprimer3/index.html>

Primer 3 is a primer design program available online. It allows designing primers for one input sequence following the parameters indicated by the user. If the sequence includes lowercase letters and the option Lowercase masking is checked, candidate primers having lowercase letter exactly at 3' end are rejected

BatchPrimer3 is a comprehensive web primer design program to develop different types of primers in a high-throughput manner. It uses the same algorithm as Primer 3 and has the advantage of allowing primer design for a batch of sequences. This can be useful when designing primers for a high number of markers

(continued)

**Table 1**  
**(continued)**

<i>Autodimer</i> <a href="http://www.cstl.nist.gov/biotech/strbase/AutoDimerHomepage/AutoDimerProgramHomepage.htm">http://www.cstl.nist.gov/biotech/strbase/AutoDimerHomepage/AutoDimerProgramHomepage.htm</a> Autodimer is a screening tool for primer-dimer and hairpin structures. The program evaluates the possible occurrence of secondary structures between the primers included in the input file. It gives a schematic representation, T <sub>m</sub> and ΔG predictions for the possible interactions. It is of particular interest when designing large multiplexes to avoid strong interactions between any primers and assure a successful amplification of all markers
<i>BLAST</i> <a href="http://www.ncbi.nlm.nih.gov/blast/">http://www.ncbi.nlm.nih.gov/blast/</a>
<i>BLAT</i> <a href="http://genome.ucsc.edu/cgi-bin/hgBlat">http://genome.ucsc.edu/cgi-bin/hgBlat</a> BLAST and BLAT are online tools to perform homology searches against sequences deposited in the databases. They allow querying PCR primers to know if other similar sequences are found in the genome, indicating possible mispriming sites that could reduce the efficiency of the PCR amplification
<i>UCSC Genome Browser's In-Silico PCR tool</i> <a href="http://genome.ucsc.edu/cgi-bin/hgPcr">http://genome.ucsc.edu/cgi-bin/hgPcr</a> In-Silico PCR tool, included in UCSC Genome Browser, searches a sequence database with a pair of PCR primers using an indexing strategy for fast performance. It allows choosing a template genome sequence and test in-silico for the PCR products which would be obtained with a given pair of primers. In summary, it permits a quick check for the possible occurrence of unspecific products in the PCR, considering a maximum product size. The default maximum product size is 4,000 bp, but it can be defined by the user

2. Unlabeled and dye-labeled primers. In order to increase the multiplexing level of the technique, a five dye chemistry was used. In this particular case, the forward primers were labeled using one of the fluorochromes 6-FAM™ (blue), VIC® (green), NED™ (yellow), or PET® (red) (Applied Biosystems). Delivered lyophilized primers were brought to 100 μM using molecular grade water and stored in the dark at -20°C. Information on primer sequences is available upon request.
3. Primer mix working solution containing all primers at 1 μM, except for markers rs2308137 (2 μM) and rs3047269 (3 μM). A stock solution for a large number of reactions can be prepared and then divided in smaller aliquots (of volume adequate to process the number of reactions desired per PCR) which are stored in the dark at -20°C (see Note 2).
4. QIAGEN Multiplex PCR Kit (Qiagen).

5. Hi-Di™ Formamide (Applied Biosystems).
6. GeneScan™ 500 LIZ® Size Standard (Applied Biosystems).

Equipment required:

- Capillary Electrophoresis instrument supporting five dye chemistry (e.g., ABI 310, 3100, or 3130 Genetic Analyzers; Applied Biosystems).
- Fragment analysis software (e.g., GeneMapper ID v3.2 or GeneMapper v4.0; Applied Biosystems).

To prevent cross-contaminations, it is recommended that one should use disposable gloves and aerosol-resistant pipette tips. Pre- and post-PCR samples, reagents, and instruments should be used and stored separately. Formamide is potentially hazard (irritant and teratogen) and should be handled accordingly.

---

### 3. Methods

The method described was previously optimized by Pereira et al. (17). Here, it is presented an overview of the stages involved in the development of the Indel multiplex and full details on the genotyping protocol.

#### **3.1. Selecting Indels for Human Identification Purposes**

The Indel markers were chosen from a pool of ~4,000 biallelic Indels initially characterized by Weber et al. (7) and included in the Marshfield Diallelic Insertion/Deletion Polymorphisms database website (<http://www.marshfieldclinic.org/mgs/>). Having in mind their application in genetic identification studies in forensic context, the markers were selected to fulfill the following criteria:

- Autosomal, biallelic Indels.
- Minimum Allele Frequency  $\geq 0.25$  in European, African, and Asian population groups.
- Average Heterozygosity  $\geq 0.40$ .
- Allele length variation of 2–5 bp.
- Localization in intergenic regions (see Note 3).

In summary, the aim was finding noncoding biallelic Indels amenable to analysis in short amplicons, with an expected high degree of polymorphism in most human population groups and widely spread throughout the genome.

#### **3.2. Constructing the Multiplex Assay**

The stages involved in constructing the Indel multiplex are summarized in Fig. 2 (see ref. (17) for details). Here the workflow is reviewed highlighting relevant aspects to consider when constructing large Indel multiplexes for forensic applications (see Table 1 for

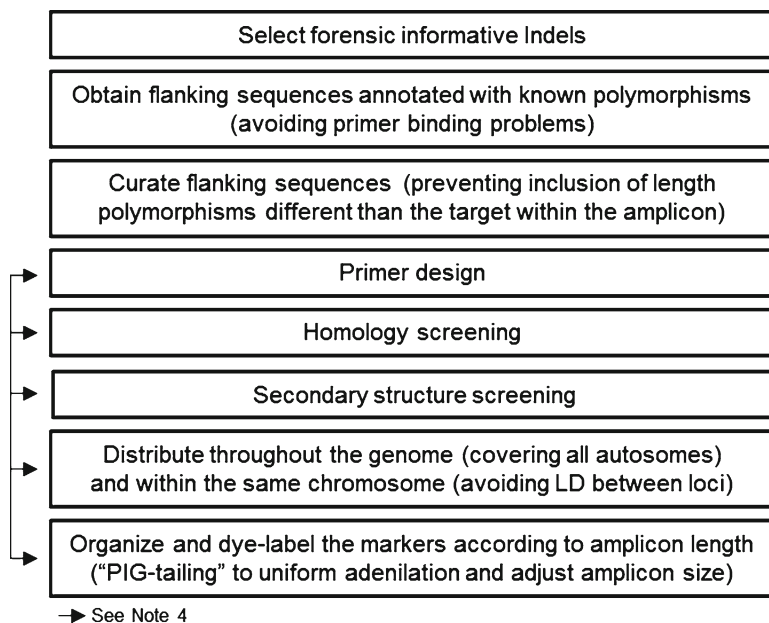


Fig. 2. Schematic representation of the different stages involved in the development of Indel multiplexes: from marker selection to PCR multiplex setup.

details on online databases and computer programs used; on the topic “constructing multiplex assays” see also ref. (20)).

The flanking sequences ( $\pm 150$  bp) of the preselected Indels, including annotations of sequence variants within that region, were obtained using the “Get DNA in Window” tool included in UCSC Genome Browser (human Mar. 2006 assembly; table snp128). This template was used for primer design, and knowing polymorphic positions helped avoiding primer-binding problems and potential null alleles. Furthermore, flanking sequences showing length polymorphisms other than the target, as well as mononucleotide tracts ( $\geq 7$  bp), were curated in order to prevent its inclusion in the amplicon range and get unexpected length variations.

Primers were designed with Primer3 using parameters aiming short amplicon sizes (60–160 bp), optimum  $T_m$  around  $60^\circ\text{C}$ , and 45% as minimum CG content. After this, all primer pairs were checked for nonspecific hybridizations in other genome regions using NCBI’s BLAST, and markers with primers presenting 100% homology in unexpected locations were excluded.

At this stage, it was necessary to multiplex as many Indels as possible, having simultaneously in mind several aspects of forensic interest:

- Secondary structures screening between primers, allowing an efficient coamplification.
- Balanced distribution throughout the human genome, covering all autosomes.



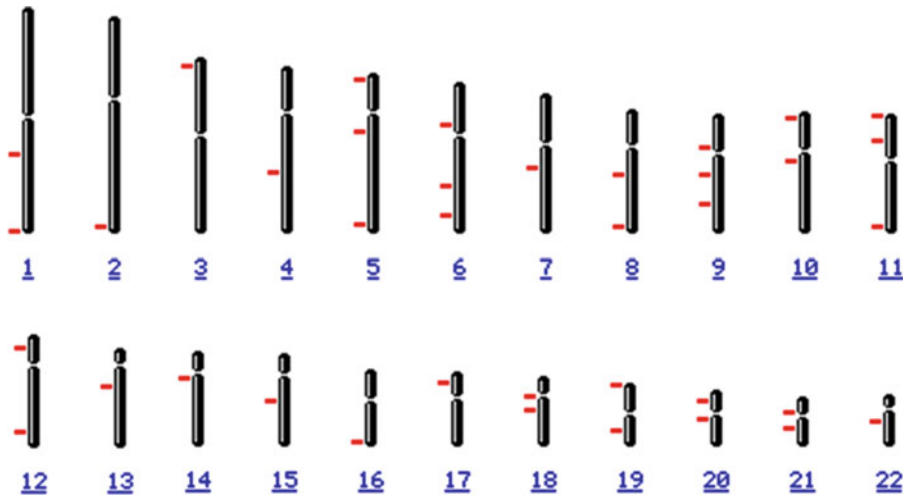


Fig. 3. Distribution of the 38 Indel polymorphisms along the 22 human autosomes.

- Position along the chromosomes, avoiding Linkage disequilibrium between markers at the same chromosome.
- Multicolor dye-labeling and distribution of the markers according to the expected amplicon sizes so that they can be typed in a single electrophoretic run.

At the end of the process, 38 Indels distributed across the 22 human autosomes (see Fig. 3) were successfully multiplexed in a single genotyping reaction. All markers were organized by the expected amplicon length and assigned to four different dye-labeling fluorochromes, enabling a genotyping analysis via a single PCR and electrophoresis. When necessary, tails of random nucleotides were added to the designed primers, aiming to adjust the final amplicon size (see Note 5).

### 3.3. Loci Information

The 38 Indels included in the assay are presented in Table 2.

### 3.4. Multiplex PCR Amplification

The amplification of the 38 Indels is performed in a single multiplex PCR using 1× Qiagen Multiplex PCR kit (Qiagen), 0.1 μM of all primers, except rs2308137 (0.2 μM) and rs3047269 (0.3 μM), and 0.3–5 ng of genomic DNA in a 10 μL final reaction volume (17). In Table 3, volumes are indicated for all PCR reagents per reaction, according to the concentration of stock solutions. If a different final volume is used, adjust all reagents accordingly.

The use of a negative control and a positive control sample with known genotype is recommended in every PCR. As reference, the genotypes of 9947A and 9948 human cell lines DNA samples are presented in Table 4.

Thermal cycling conditions are: initial incubation at 95°C for 15 min; followed by 10 cycles at 94°C for 30 s, 60°C for 90 s, and

**Table 2**  
**Indel polymorphisms included in the multiplex**

Int. code	MID	rs Number	Chromosome	Position (bp)		Alleles	Amplicon size (S–L)
				Build 36.3	Build 37.1		
B1	MID-2719	rs34541393	20	30165066	30701405	–/AACT	57–61
B2	MID-185	rs16624	2	234681130	235016391	–/GT	65–67
B3	MID-1493	rs2307689	19	48896180	44204340	–/TTC	74–77
B4	MID-2946	rs35769550	8	76681235	76518680	–/TGAC	89–93
B5	MID-1504	rs2307700	22	25120901	26790901	–/TCAC	101–105
B6	MID-520	rs140809	10	6027167	5987163	–/CAA	115–118
B7	MID-2305	rs3047269	1	161077452	162810828	–/CTGA	126–130
B8	MID-3221	rs33972805	11	125794082	126288872	–/CT	135–137
B9	MID-2698	rs33917182	20	11643625	11695625	–/CA	142–144
B10	MID-116	rs16402	9	38396788	38406788	–/TTAT	150–154
G1	MID-785	rs1610871	5	171020572	171087970	–/TAGG	61–65
G2	MID-1151	rs2067238	12	113772931	115288548	–/GCT	71–74
G3	MID-1209	rs2067294	9	70504241	71314421	–/CTT	80–83
G4	MID-1514	rs2307710	6	47929222	47821263	–/AGGA	92–96
G5	MID-2050	rs2308242	3	8591709	8616709	–/CT	106–108
G6	MID-1384	rs2307580	9	104626014	105586193	–/AATT	120–124
G7	MID-743	rs1160956	5	65414216	65378460	–/AGA	128–131
G8	MID-3097	rs34511541	18	34677042	36423040	–/CTCTT	143–148
G9	MID-1782	rs2307978	7	83121850	83283913	–/GA	156–158
Y1	MID-2648	rs3051300	17	10076666	10135941	–/GTAT	63–67
Y2	MID-2890	rs10629077	21	30294208	31372337	–/AT	74–76
Y3	MID-3277	rs10688868	11	258180	268180	–/CT	81–83
Y4	MID-1120	rs2067208	16	83139788	84582287	–/GCCAG	93–98
Y5	MID-1383	rs2307579	1	245878706	247812083	–/ATG	104–107
Y6	MID-1824	rs2308020	15	51268809	53481517	–/TT	127–129
Y7	MID-3114	rs3080855	18	21507205	23253207	–/AATT	133–137
Y8	MID-834	rs1610919	12	14801263	14909996	–/AT	142–144
Y9	MID-1643	rs2307839	6	117200251	117093558	–/GA	152–154
R1	MID-1945	rs2308137	6	149655891	149614198	–/GA	61–63
R2	MID-2592	rs36040336	19	1353662	1402662	–/AT	65–67
R3	MID-649	rs1160886	10	54112392	54442386	–/ACT	75–78
R4	MID-1830	rs2308026	4	119404855	119185407	–/CA	83–85
R5	MID-1330	rs2307526	5	5178112	5125112	–/ACAC	95–99
R6	MID-3220	rs34811743	11	30134266	30177690	–/TG	108–110
R7	MID-1997	rs2308189	14	28106508	29036757	–/AACTA	119–124
R8	MID-3031	rs5895447	8	138489776	138420594	–/CA	128–130
R9	MID-1979	rs2308171	13	43778155	44880155	–/TCTG	135–139
R10	MID-2806	rs35605984	21	14556736	15634865	–/TAAAG	151–156

**Table 3**  
**PCR amplification mix**

Water	3 $\mu$ L
Qiagen Multiplex PCR master mix (2 $\times$ )	5 $\mu$ L
Primer mix (10 $\times$ )	1 $\mu$ L
Template DNA 0.3–5 ng/ $\mu$ L	1 $\mu$ L
Total reaction volume	10 $\mu$ L

*Note:* To prevent contamination, it is recommended to use gloves and pipette tips with filter

**Table 4**  
**Genetic profiles for the reference samples 9947A and 9948**

9947A										
DNA marker	1	2	3	4	5	6	7	8	9	10
B	SS	SL	SL	SL	SL	LL	SL	LL	SL	SL
G	SL	SL	SL	LL	SL	SL	SS	SL	LL	
Y	LL	LL	SL	SL	SL	SL	SL	SL	SS	
R	SS	SL	SL	LL	LL	SS	SL	SL	SL	SL

9948										
DNA marker	1	2	3	4	5	6	7	8	9	10
B	SL	SS	SS	SS	SL	SL	SL	SL	SL	LL
G	LL	SL	LL	SL	LL	SS	SS	SL	LL	
Y	SL	SS	LL	LL	LL	SL	SL	SL	SL	
R	SL	SS	LL	SL	SL	SS	SL	SS	LL	SL

S Short allele; L long allele

72°C for 60 s; plus 20 cycles at 94°C for 30 s, 58°C for 90 s, and 72°C for 60 s (see Note 6); and a final extension at 72°C for 60 min (see Note 7).

**3.5. Analysis  
of Amplified Products**

PCR products are prepared for capillary electrophoresis by adding 1  $\mu$ L of each amplified product to 10  $\mu$ L Hi-Di™ Formamide (Applied Biosystems) and 0.3  $\mu$ L of internal size standard GeneScan™ 500 LIZ® (Applied Biosystems).

Separation and detection are performed by a capillary electrophoresis system. During the optimization of the present method, an Applied Biosystems 3130 Genetic Analyzer was used, with G5 filter set and Performance Optimized Polymer 7 (POP-7™; Applied Biosystems). Analysis conditions were: Module: “FragmentAnalysis 36\_POP7”; Oven temperature: 60°C; Injection Voltage: 1.2 kV; Injection Time: 16 s; Run Voltage: 15 kV; Run time: 1,200 s. Data collected are analyzed using a multicomponent matrix (see Note 8).

Finally, the electropherograms are visualized with a fragment analysis software and fragment sizes are determined in order to genotype the samples. In this case, GeneMapper ID v3.2 or GeneMapper v4.0 (Applied Biosystems) were used. It is recommended to create genotyping panels including all markers with the respective allelic bin sets in order to automate allele calling and facilitate the genotyping of a large number of samples. The genotype data can be easily exported to text documents for further treatment and databasing.

An example electropherogram is shown in Fig. 4.

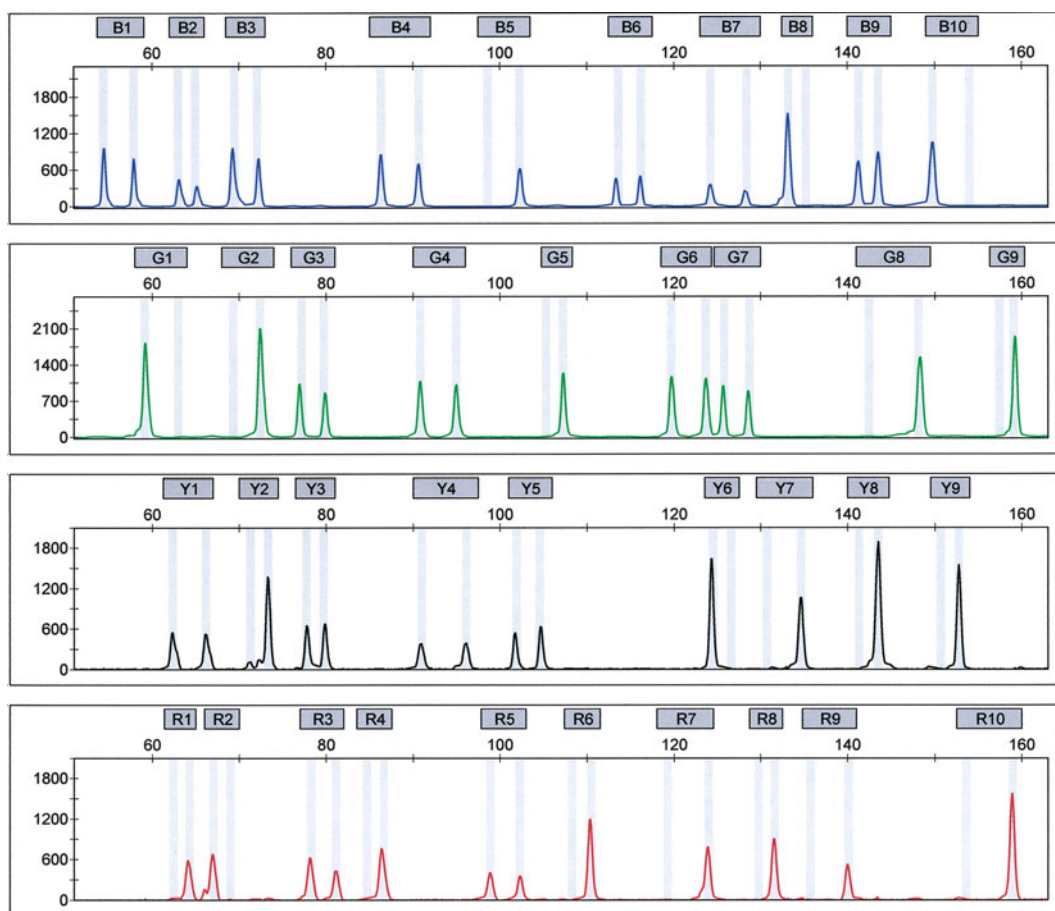


Fig. 4. Example of an electropherogram of the 38 Indel multiplex.

### 3.6. Additional Considerations

Depending on the particular characteristics of a case, there are important aspects to consider when selecting the best analytical approach, namely the amount of sample available for analysis, the DNA quality (good quality or highly degraded?), and/or the presence of inhibitors. Some recommendations/tips are suggested that might help improving the results, particularly when dealing with challenging DNA:

- In cases of very low DNA quantity or when analyzing highly degraded samples (low template DNA), test increasing the total number of PCR cycles to 32 and/or increasing the primer mix concentration to 1.5 or 2× in the PCR reaction (see Note 9).
- Using larger final volumes in the PCR reaction (e.g., 25 µL) may also improve the results when analyzing highly degraded samples.
- In the presence of PCR inhibition, if enough DNA is available use dilution series of the original sample (e.g., 1:2, 1:5, 1:10; 1:20; 1:50, 1:100). While diminishing the quantity of template DNA in the reaction, it also dilutes possible inhibitors, which is sometimes the only way to obtain a successful amplification.

Occasionally, dye artifacts (“dye-blobs”) can appear in electropherograms. Such artifacts are usually due to a poor quality of the primers, produced by an incomplete fluorochrome attaching during manufacturing (20). Also, the primers can suffer degradation with time and freeze–thaw cycles, leading to a reduced performance and the presence of more residual dye artifacts. Minor “dye-blobs” are frequently observed in electropherograms, but usually do not interfere with genotyping. Nevertheless, sometimes they can appear in regions with possible alleles thus hindering genotyping. In these cases, a complementary purification step prior to electrophoresis (e.g., with Sephadex™ G-50 DNA Grade; GE Healthcare) can improve considerably the overall electropherogram by eliminating exceeding primers and dye residues.

If electropherograms present unusually strong peaks, with pull-ups that can be confused with alleles and/or interfere with the internal size standard alignment, testing one of the following is recommended: further dilute the DNA samples; reduce the number of PCR thermal cycles (see Note 6); and apply less PCR product or use less injection time in the capillary electrophoresis system. Good results are usually obtained within the range of 1,000–5,000 rfu, in average.

---

## 4. Notes

1. The simultaneous use of STRs and Indels in calculating PI must consider the LD status between the markers under study. Therefore, at least for markers at the same chromosome, an LD test must be performed in the reference population to

discard possible associations between markers (and therefore safely use the product rule in PI calculations); otherwise, linked loci pairs must be treated as haplotypes.

2. Dye-labeled primers are light-sensitive and frequent freeze–thaw cycles can accelerate breakdown of the dye attachment to the oligonucleotide and cause an overall deterioration of the primers leading to weaker performance with time.
3. The noncoding Indels localized only in intergenics regions cannot be found directly in dbSNP. In order to do that, first query the Entrez SNP database for all human Indels fulfilling the criteria: previously validated by assessing allele frequencies, heterozygosity values between 0.40 and 0.50, mapping only once in the human genome, and submitted to the database under the MARSHFIELD handle. Afterwards, obtain the intragenic Indels (obtained by checking-in all *Function Class* search field limits “coding nonsynonymous, intron, coding synonymous, locus region, mrna utr, splice site”). Finally, go to the “History” tab and subtract the results of the first search minus the second (#1 NOT #2). This delivers a list of Indels located only in intergenic regions. To know more about searching for markers of forensic interest in Online databases, see ref. (21).
4. The final stages in Fig. 2 (signaled with arrows) are interconnected and the process runs somehow simultaneously: markers can be excluded for different reasons at different stages and there may be the possibility to rescue an old marker because its prior excluding reason no longer applies (e.g., marker A has conflicting primers with markers B, C, and D and is preliminarily excluded to keep these three; markers B, C, D, E are closely located in a chromosome and only one can be selected to avoid LD; Marker E is selected from the cluster, so Marker A can be rescued). The final aim is obtaining the maximum markers in a single reaction and fulfilling all criteria.
5. More recently, we started to employ the Brownstein et al. (22) “PIG tailing” approach which was originally developed to overcome genotyping problems created by the nonspecific adenylation of amplified products by DNA polymerases, which generates confusing double peaks appearing in the electropherogram (–A,+A). This strategy can be used aiming a dual function of adjusting amplicon sizes while at the same time homogenizing nontemplate adenylation of all products in order to obtain a uniform electropherogram. The unlabeled primers of amplicons that do not need mobility adjustments remain unchanged, or a G is added at the 5′ end if there is not already a G in that position. In the cases requiring little mobility adjustments, a 5′ tail GTTT or partial tails of the original sequences GTTTCTT or GTGTCTT are attached to the unlabeled primer.

6. The number of thermal cycles can be adjusted for different input DNA, depending on the quality and/or concentration of the samples being processed. Samples with a good quantity/quality DNA can provide well-balanced results with a total of 27 cycles, avoiding unusually strong peaks and possible pull-ups, and minimizing adenylation issues. On the contrary, when dealing with very low amounts of DNA (e.g., low copy number DNA) or highly degraded samples, the reaction can benefit of using up to 32 cycles.
7. Double peaks differing in a single bp ( $-A,+A$ ) can appear due to an incomplete adenylation, usually when high amounts of DNA are present in the PCR reaction. In these cases, apart from reducing the amount of DNA, the final extension step may be extended to 80 min in order to complete adenylation of the amplified products.
8. DS-33 (Dye Set G5) Matrix Standard (Applied Biosystems) is used to generate the multicomponent matrix required when analyzing 6-FAM<sup>TM</sup>, VIC<sup>®</sup>, NED<sup>TM</sup>, PET<sup>®</sup>, and LIZ<sup>®</sup> dye-labeled DNA (for details see the Applied Biosystems 3130 Genetic Analyzer User's Manual).
9. Regularly, concentrations of 0.1  $\mu$ M are used for all primers in the PCR reaction (except for the two primer pairs needing reinforcement rs2308137 and rs3047269) following the Qiagen Multiplex PCR kit User's Manual, which suggests that for multiplex amplification reactions with more than ten PCR products, a decrease of primer concentration to 0.1  $\mu$ M may lead to a more uniform product yield (Appendix F: Optimization of Reaction Conditions for Special Multiplex PCR Applications; Large number of PCR products). Nonetheless, the 0.2  $\mu$ M concentration of each primer, originally suggested at the User's Manual, can be used in cases where the template DNA is scarce or highly degraded, since having more primer available enhances the probability of molecules to encounter and anneal to attain amplification.

---

## Acknowledgments

IPATIMUP is an Associate Laboratory of the Portuguese Ministry of Science, Technology and Higher Education and is partially supported by FCT, the Portuguese Foundation for Science and Technology.



## References

1. Dixon LA, Murray CM, Archer EJ, Dobbins AE, Koumi P, and Gill P. (2005) Validation of a 21-locus autosomal SNP multiplex for forensic identification purposes, *Forensic Sci. Int.* **154**, 62–77.
2. Sobrino B, and Carracedo A. (2005) SNP typing in forensic genetics: a review, *Methods Mol. Biol.* **297**, 107–126.
3. 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, and Morling N. (2006) A multiplex assay with 52 single nucleotide polymorphisms for human identification, *Electrophoresis* **27**, 1713–1724.
4. Børsting C, Sanchez JJ, Hansen HE, Hansen AJ, Bruun HQ, and Morling N. (2008) Performance of the SNPforID 52 SNP-plex assay in paternity testing, *Forensic Sci Int Genet* **2**, 292–300.
5. Phillips C, Fondevila M, Garcia-Magarinos M, Rodriguez A, Salas A, Carracedo A, and Lareu MV. (2008) Resolving relationship tests that show ambiguous STR results using autosomal SNPs as supplementary markers, *Forensic Sci Int Genet* **2**, 198–204.
6. Krjutskov K, Viltrop T, Palta P, Metspalu E, Tamm E, Suvi S, Sak K, Merilo A, Sork H, Teek R, Nikopensius T, Kivisild T, and Metspalu A. (2009) Evaluation of the 124-plex SNP typing microarray for forensic testing, *Forensic Sci Int Genet* **4**, 43–48.
7. Weber JL, David D, Heil J, Fan Y, Zhao C, and Marth G. (2002) Human diallelic insertion/deletion polymorphisms, *Am. J. Hum. Genet.* **71**, 854–862.
8. Mills RE, Luttig CT, Larkins CE, Beauchamp A, Tsui C, Pittard WS, and Devine SE. (2006) An initial map of insertion and deletion (INDEL) variation in the human genome, *Genome Res.* **16**, 1182–1190.
9. Yang N, Li H, Criswell LA, Gregersen PK, Alarcon-Riquelme ME, Kittles R, Shigeta R, Silva G, Patel PI, Belmont JW, and Seldin MF. (2005) Examination of ancestry and ethnic affiliation using highly informative diallelic DNA markers: application to diverse and admixed populations and implications for clinical epidemiology and forensic medicine, *Hum. Genet.* **118**, 382–392.
10. Rosenberg NA, Mahajan S, Ramachandran S, Zhao C, Pritchard JK, and Feldman MW. (2005) Clines, clusters, and the effect of study design on the inference of human population structure, *PLoS Genet* **1**, e70.
11. 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.
12. Tishkoff SA, Reed FA, Friedlaender FR, Ehret C, Ranciaro A, Froment A, Hirbo JB, Awomoyi AA, Bodo JM, Doumbo O, Ibrahim M, Juma AT, Kotze MJ, Lema G, Moore JH, Mortensen H, Nyambo TB, Omar SA, Powell K, Pretorius GS, Smith MW, Thera MA, Wambebe C, Weber JL, and Williams SM. (2009) The genetic structure and history of Africans and African Americans, *Science* **324**, 1035–1044.
13. Santos NP, Ribeiro-Rodrigues EM, Ribeiro-Dos-Santos AK, Pereira R, Gusmão L, Amorim A, Guerreiro JF, Zago MA, Matte C, Hutz MH, and Santos SE. (2010) Assessing individual interethnic admixture and population substructure using a 48-insertion-deletion (INDEL) ancestry-informative marker (AIM) panel, *Hum. Mutat.* **31**, 184–190.
14. Ribeiro-Rodrigues EM, dos Santos NP, dos Santos AK, Pereira R, Amorim A, Gusmão L, Zago MA, and dos Santos SE. (2009) Assessing interethnic admixture using an X-linked insertion-deletion multiplex, *Am J Hum Biol* **21**, 707–709.
15. Edlmann J, Hering S, Augustin C, and Szibor R. (2009) Indel polymorphisms--An additional set of markers on the X-chromosome, *Forensic Science International: Genetics Supplement Series* **2**, 510–512.
16. Freitas NS, Resque RL, Ribeiro-Rodrigues EM, Guerreiro JF, Santos NP, Ribeiro-dos-Santos A, and Santos S. (2010) X-linked insertion/deletion polymorphisms: forensic applications of a 33-markers panel, *Int J Legal Med* **124**, 589–593.
17. Pereira R, Phillips C, Alves C, Amorim A, Carracedo A, and Gusmão L. (2009) A new multiplex for human identification using insertion/deletion polymorphisms, *Electrophoresis* **30**, 3682–3690.
18. Pimenta JR, and Pena SD. (2010) Efficient human paternity testing with a panel of 40 short insertion-deletion polymorphisms, *Genet. Mol. Res.* **9**, 601–607.



19. Pereira R, Phillips C, Alves C, Amorim A, Carracedo Á, and Gusmão L. (2009) Insertion/deletion polymorphisms: A multiplex assay and forensic applications, *Forensic Science International: Genetics Supplement Series* 2, 513–515.
20. Butler JM. (2005) Constructing STR multiplex assays, *Methods Mol. Biol.* 297, 53–66.
21. Phillips C. (2005) Using online databases for developing SNP markers of forensic interest, *Methods Mol. Biol.* 297, 83–106.
22. 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, 1004–1006, 1008–1010.