Automated Genotyping of a Highly Informative Panel of 40 Short Insertion-Deletion Polymorphisms Resolved in Polyacrylamide Gels for Forensic Identification and Kinship Analysis

Heloisa B. Pena^a Sérgio D. J. Pena^{a,b}

^a GENE – Núcleo de Genética Médica, Belo Horizonte,

^b Departamento de Bioquímica e Imunologia, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil

Keywords

Insertion-deletion polymorphisms · Indels · DNA · Identification · Genome · Polyacrylamide gels

Summary

Objective: Short insertion-deletion polymorphisms (indels) are the second most abundant form of genetic variations in humans after SNPs. Since indel alleles differ in size, they can be typed using the same methodological approaches and equipment currently utilized for microsatellite genotyping, which is already operational in forensic laboratories. We have previously shown that a panel of 40 carefully chosen indels has excellent potential for forensic identification, with combined probability of identity (match probability) of 7.09×10^{-17} for Europeans. **Methods:** We describe the successful development of a multiplex system for genotyping the 40-indel panel in long thin denaturing polyacrylamide gels with silver staining. We also demonstrate that the system can be easily fully automated with a simple large scanner and commercial software. **Results and Conclusion:** The great advantage of the new system of typing is its very low cost. The total price for laboratory equipment is less than EUR 10,000.–, and genotyping of an individual patient will cost less than EUR 10.– in supplies. Thus, the 40-indel panel described here and the newly developed 'low-tech' analysis platform represent useful new tools for forensic identification and kinship analysis in laboratories with limited budgets, especially in developing countries.

Schlüsselwörter

Insertion-Deletion-Polymorphismen · Indels · DNA · Identifizierung · Genom · Polyacrylamid-Gele

Zusammenfassung

Hintergrund: Kurze Insertion-Deletion-Polymorphismen (Indels) sind nach SNPs die zweithäufigste Form von genetischen Variationen beim Menschen. Da Indel-Allele in der Größe variieren, können sie mit denselben methodischen Ansätzen und Equipment typisiert werden, das für die Mikrosatelliten-Genotypisierung verwendet wird, die bereits in forensischen Laboren etabliert ist. Wir haben bereits früher gezeigt, dass ein Panel von 40 sorgfältig ausgewählten Indels ein ausgezeichnetes Potential für eine forensische Identifizierung hat; die kombinierte Wahrscheinlichkeit der Identität (Match-Wahrscheinlichkeit) liegt bei Europäern bei 7.09×10^{-17} . **Methoden:** Die erfolgreiche Entwicklung eines Multiplex-Systems zur Genotypisierung des 40-Indel-Panels in langen dünnen denaturierenden Polyacrylamid-Gels mit Silberfärbung wird beschrieben. Wir konnten außerdem zeigen, dass das System mit Hilfe eines einfachen großen Scanners und kommerzieller Software leicht vollautomatisiert werden kann. **Ergebnisse und Schlussfolgerung:** Der größte Vorteil des neuen Typisierungssystems sind seine sehr geringen Kosten. Die Gesamtkosten für das Laborequipment liegen unter 10 000.– EUR, und die Genotypisierung eines einzelnen Patienten wird weniger als 10.– EUR für Verbrauchsmaterialien kosten. Das hier beschriebene 40-Indel-Panel und die neu entwickelte Low-Cost-Analyse-Plattform sind nützliche neue Tools für die forensische Identifizierung und Verwandtschaftsanalyse in Laboratorien mit begrenztem Budget, besonders in Entwicklungsländern.

Introduction

In the past few years, several million short insertion-deletion polymorphisms (indels) have been described in the human genome – they are the second most abundant form of genetic variation in humans after SNPs [1]. The typing of SNPs, which depends on qualitative base identification, is best done by microarray hybridization, which requires complex equipment that is not generally available in forensic laboratories. On the other hand, indel alleles differ in size and can be typed using the same methodological approaches and equipment currently utilized for microsatellite genotyping, which is already operational in paternity laboratories.

In 2002, Weber et al. [2] characterized 2,000 human diallelic short indels in the human genome. Among them, we identified 40 polymorphisms that fulfilled the following criteria: widespread chromosomal location, increasing amplicon sizes to allow multiplex analysis, and allele frequencies close to 0.50 in the European population [3]. We have shown that this set of 40 carefully chosen indels was sufficiently informative for an adequate characterization of human population structure at the global level [3]. Moreover, we demonstrated that the excellent resolution power of these markers in discriminating among Europeans, Africans, and Amerindians could be used in the analysis of the ancestral roots of the Brazilian population in more than 1,000 individuals [4, 5] and in pharmacogenetic studies [6].

During these studies it became clear that our 40-indel panel also had excellent capacity for individual identification, with combined probability of identity (match probability) of 7.09×10^{-17} for Europeans [7]. Extremely low match probabilities were likewise observed for other geographical regions (table 1).

We explored further the potential application of the 40-indel panel in forensic studies by genotyping the markers in 360 unrelated self-classified White Brazilians and 50 mother-child-probable father trios with proven paternity [8]. 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 Whites was 3.48×10^{-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, especially because it is composed of genomic markers that have very low mutation rates, it represents a useful new tool as adjunct to routine microsatellite multiplex sets in human paternity testing. Other groups have also had good results in developing informative panels of insertion/deletion polymorphisms for forensic use [9, 10].

Developing countries such as Brazil constitute technological mosaics. Thus, while forensic services in some regions of the country have state-of-the-art laboratories, in other poorer regions they may lack budget for the large expenses involved in purchasing and operating equipment such as fluorescent

Table 1. Match probabilities of the 40-indel panel in individuals from different regions of the globe (calculated from the data of Bastos-Rodrigues et al. [3])

Region	Number of populations	Number of individuals	Match probability
Europe	8	161	7.09×10^{-17}
Middle East	4	178	1.01×10^{-16}
Central Asia	9	210	1.01×10^{-16}
East Asia	17	241	1.02×10^{-14}
Africa	7	127	1.20×10^{-12}
Oceania	2	19	8.73×10^{-13}
Americas	5	108	4.66×10^{-13}

DNA analyzers. In such ‘low-tech’ laboratories, the simplicity of running and analyzing the simple diallelic indel multiplexes would be very helpful.

In the present article we describe the successful development of a simple multiplex system for genotyping the 40-indel panel in long thin denaturing polyacrylamide gels with silver staining, which can be highly useful as a low-cost adjunct to the more technologically complex fluorescent typing of the same loci in automatic DNA sequencers. We also demonstrate that the system can be easily fully automated with a simple large scanner and commercial software.

Material and Methods

Populations Studied

DNA samples from 1,044 individuals were obtained from HGPD-CEPH Human Genome Diversity Cell Line Panel [11]. The individuals had been sampled across all five continents and assigned to 52 different populations from seven regional groups (Africa, Europe, Middle East, Central Asia, East Asia, Oceania, and America).

DNA Analysis

DNA from each individual was independently typed for the 40 diallelic short indels listed in table 2. PCR amplification in four multiplex reaction and electrophoretic analyses in a MegaBace 1000 sequencer (Applied Biosystems, Carlsbad, CA, USA) or an ABI 3130 Genetic Analyzer (Applied Biosystems) were performed exactly as described previously [3, 4]. A representative run is shown in figure 1.

For analysis in long polyacrylamide gels and silver staining, four modified multiplex amplifications were used using the primers listed in table 2. The amplification reactions were performed in a volume of 9 μ l containing 2mmol/l dNTPs, 1.4mmol/l $MgCl_2$, 1.5 U of Taq Polymerase Platinum (Invitrogen, Foster City, CA, USA), and primer mixes (table 2) in a 20 mmol/l Tris-HCl buffer, pH 8.4, containing 50 mmol/l KCl. After an initial 5-min denaturation at 95 °C we used a 10-cycle step-down protocol from 60 °C to 51 °C, with elongation at 72 °C for 2 min and denaturation at 94 °C for 1 min followed by 28 cycles of 95 °C, 50 °C and 72 °C and a final extension step at 72 °C for 10 min.

Electrophoresis was performed in long denaturing thin polyacrylamide gels (300 $\times$ 340 $\times$ 0.4 mm) using ‘shark tooth’ sample combs for sample application. The gel has the capacity for 64 samples (i.e. 16 individual sets of 4 lanes). For the preparation and running of the gels we followed the Promega GenePrint STR Systems (Promega, Madison, WI, USA) technical manual [12].

Table 2. List of the 40 indels used in the panel described, together with the primer sequences and amplicon sizes of the multiplexes analyzed in long thin denaturing polyacrylamide gels with silver staining

Indel	Forward primer	Reverse primer	Amplicon sizes
Multiplex 1			
rs2067373	AAAGGGGAAAAACATAAATGA	CACTCTCAGCCTCAATAGATTT	72–78
rs2307745	AACTGTCAGCTGGCTTTTCC	GAATAGTGCTTTAGTTCTTGTTGTCA	83–97
rs140733	AGCATCTTGGGAGTAGGACT	CCCATTTACCCACCTAAATT	93–98
rs16343	TCCGATACAATTTTGATTGG	CCTCCACCTTCTTTTCTGT	114–118
rs2307955	TGCAATAGAAAATTTCTGGG	CAATGTCTAACCTTTTGGA	130–134
rs2307548	AGGCTCTTATTTTCCATTC	ATTAATGGGGAAAAGTGCTC	139–143
rs16381	TCTACCCAACTACCTCCACAT	GATGGCTTATGTCATCAGTAAAG	152–158
rs4183	ACAGTTTCACCTAGAACTGCAA	TGCTTGTAACAGATGTGATAGA	166–170
rs2307733	TGAGTCCTTCCTCTATTAACACAGTT	ACTGCAAAATATGACCCAACG	190–194
rs2307959	TGAACATGGTCTGTGAGGCTCT	CCAGAAAAAGCCTAGATGAATTG	201–206
rs1611001	TGAATTCAACTGCTGGGTTG	CACCAGAAATGGACCATGC	313–326
Multiplex 2			
rs16448	TCATGATGCAGTCTCCTCTT	AAATACAGGTCTCTGACCC	75–80
rs2307624	CCAATGGAGAAATGTACCTG	TCCATACACCTGTGAGTCCT	89–92
rs16415	GTGAAGTTTGCATTGCTTT	AGCTGGAAATCACTTGCTTT	93–96
rs16430	GACGAATGCAGAACACTTCT	AATCTGAGGGAGCTGAGTAAC	104–110
rs16416	GCCTTCAAAAATGTGGTTAA	ATCACATTGGCAAAAATGTT	115–118
rs1305047	TGACAGATATGTTCACTGGCT	TGTCCTGTTTTGTAGAGCC	133–136
rs2307838	AGAGGAGAACTGAGCCATT	GCTTATGGTGCTATCCTCAA	142–146
rs16695	TTCTCTTTTGCTTTCC	CCCATTTTGGTAAAACCACT	154–157
rs140759	TCAATCATTTAAGGCAGTATTTT	TGGTTGAGAAAAGAAGAGAAATA	175–179
rs16394	ATCATGTCATTTACCTGTAAAT	TGTGGCCTTTGTTTGAATC	186–190
rs3917	TTTAATTTTCTGCTTGCCC	GAATGTTGATGGTGCTAGAAGA	243–250
Multiplex 3			
rs2308144	CAGCATTAGTCATGACCACC	GCTGCAAGAGAATGTGATCT	80–83
rs1610874	CAGTGCAGGGAGCTCAGA3'	CACTTAAAATTCAAATCCATTGC3'	89–93
rs2308135	ATTCGATCTTATGACCGTGA	AAGAGAAGCAAGGGAAAGAG	94–97
rs2307782	GTTATGTGGGCAGTGTTTTC	CCACAGAAGGCTCAGTCTTA	103–109
rs2067180	CAGATGACACAGTTGACCTTC	CCTGGTTCTGATCCAACTA	121–125
rs16715	TTGCAATTTGAGTGTGTCAG	TGAAAGCAACTCTTCTCGTC	140–146
rs16438	GTAACAATGAGGGGCACA	GAGCACTACAGCCTTTTATTGA	157–162
rs1611084	AGGTAAAGTCGACCTGGATAA	AATTCAGTGTGAATGAAACCA	183–186
rs2308057	AGGTGAAGTATTGCCTGCAT	GTGGGACCATGGACCAGAG	204–208
rs2308043	TGCAAGAAACAACCCTCACA	CACCACCCAATTTGCTACCT	254–263
Multiplex 4			
rs2067217	GGCCTTAGCAATGGAGTAGGA	TGGGCTGAGCATTCTGTTAGG	88–95
rs140757	GGACACACTTTTGTGTTTGTG	AGCTTACATAGGCATGCAAC	111–115
rs1610997	GTCGAGGCTGTGATAGTGAG	AAGTCTCTTAGTGCAGGGGT	121–124
rs2307850	GTAATAGCCCCATTAGGGAG	AAATGGAGGTTAAATGGCTT	134–138
rs140709	TCTTCACCATCAAAGAGTCTTC	CCAAGGAAAAATTATTTAAGGAG	185–188
rs1610942	TTGTTGCCTATGTGGGGAAT	CCTCAAAGTCAACTGAGGAA	204–206
rs2067188	CCAAAAAGGAGGGAAAGTC	CAGCATCAAAGGCACACACT	253–272
rs4181B	TTGTGGCACAGCTGCATTA	TGTTAAATTAATGTAGCCTGCC	280–290

The acrylamide/biscrylamide gels (T = 6%, C = 5%) contained 7% urea in TBE buffer (50 mmol/l Tris / 82 mmol/l borate / 1 mmol/l EDTA). Samples were denatured in a 95% formamide in 10 mmol/l NaOH containing 12.5% sucrose and bromphenol blue and xylene cyanol tracer dyes. The run was in TBE at 45 W until the xylene cyanol marker reached 19 cm before the end of the gel. Silver stain was performed with the gel attached at the glass plate according to the same PROMEGA technical manual [12].

After silver staining, the gel was dried at room temperature still attached at the glass plate overnight, was digitalized in a Kodak i1210 Plus Scanner (46.5 cm × 31 cm; Eastman Kodak, Rochester, NY, USA), and then was analyzed with the software UN-SCAN-IT gel version 6.1 (Silk Scientific, Orem, UT, USA), which converts the gel image to an electropherogram. The peaks of the electropherogram were inspected visually and adjusted when needed; finally their mobility was listed in a table that was exported in text format. A proprietary program (Sysgene) was writ-

Fig. 1. Demonstrative run of Multiplex 2 of the indel panel in the Applied Biosystems Genetic Analyzer 3130. Multiplex 2 is composed of 12 indels, which are individually identified.

ten in the Delphi Programming Language to compare the relative mobility data from each lane of the electropherogram with the relative mobility of the allele ladder and finally output the genotypes at each lane.

Data Analysis

Basic statistics and histogram were obtained using the program MedCalc (MedCalc Software, Mariakerke, Belgium).

Results and Discussion

This set of 40 indels was previously validated as useful in ancestry estimation through the study of the HGDP-CEPH Diversity Panel, which is composed of 1,064 individuals from 52 different worldwide populations distributed in seven geographical regions [11]. The individual results have been deposited in the CEPH Genotype Database (www.cephb.fr/en/hgdp/main.php), from which they are available. Such study was performed using PCR amplification with fluorescently labeled markers and automated analysis in a MegaBace 1000 DNA sequencer as described previously [3].

In this article we describe the development of a new ‘low-tech’ platform for obtaining the genotypes at the 40-indel panel using simple and inexpensive long thin denaturing polyacrylamide gels with silver staining (fig. 2). Following a work-intensive phase in which the gels are prepared, run, and stained, the actual analysis of the gels and reading of the genotypes can be obtained quickly and effortlessly using the automation with a simple large scanner, application of a commercial software to convert the gel bands into an electropherogram, and finally the use of a simple proprietary software to obtain the genotypes (fig. 3).

Fig. 2. Demonstrative run of PCR amplification products of the 40-indel panel with five unrelated individuals (numbered 1–5) from the HGDP-CEPH set. Lanes labeled ‘L’ are allelic ladders prepared by PCR amplification of a mix of the DNA of 200 individuals.

A histogram of the amplicon sizes for the large allele of the indels is shown in figure 4. They range from 78 to 326 base pairs, with a mean value of 154. 20 of the 40 amplicons have size below 140 base pairs. We validated the newly developed

Sample	rs3917	rs4181	rs4183	rs16343	rs16381	rs16394	rs16415	rs16416	rs16430	rs16438
1	1	2	1	2	2	2	1	2	1	1
2	1	2	1	2	2	2	1	1	1	2
3	1	2	1	2	2	2	2	2	1	2
4	1	1	1	1	2	2	1	1	2	1
5	2	2	1	2	2	2	1	2	1	1

Sample	rs16448	rs16695	rs16715	rs140709	rs140733	rs140757	rs140759	rs1305047	rs1610874	rs1610942
1	1	2	2	2	1	1	2	2	1	2
2	2	2	1	2	1	1	2	2	2	1
3	2	2	1	2	1	1	2	2	2	1
4	2	2	1	1	1	1	2	1	2	1
5	2	2	1	2	1	2	1	2	1	1

Sample	rs1610997	rs1611001	rs1611084	rs2067180	rs2067188	rs2067217	rs2067373	rs2307548	rs2307624	rs2307733
1	2	2	2	2	1	1	2	2	2	1
2	1	2	1	2	2	2	1	2	1	1
3	1	2	1	2	1	1	2	2	2	1
4	1	2	1	2	1	1	2	2	1	2
5	2	2	2	2	1	2	1	2	1	2

Sample	rs2307745	rs2307782	rs2307838	rs2307850	rs2307955	rs2307959	rs2308043	rs2308057	rs2308135	rs2308144
1	1	1	1	2	1	2	2	1	1	2
2	1	2	1	2	1	2	1	2	2	2
3	1	2	2	2	2	2	1	1	1	2
4	1	2	1	1	2	2	1	2	2	2
5	2	2	2	2	1	2	1	2	2	1

Fig. 3. Genotypes of individuals 1–5 shown in figure 2 at the 40 indel loci, calculated automatically using the computer programs described in the article.

Fig. 4. Histogram of the amplicon sizes of the large allele of the 40-indel panel obtained with the primers listed in table 2.

system by comparing the genotypes of 50 individuals in both platforms. The results were 100% concordant for all samples.

The great advantage of the new system of typing is its very low cost. Besides a PCR thermocycler, all the special laboratory equipment needed is an electrophoresis DNA sequencing apparatus for long thin denaturing gels, a high wattage power supply, and a large flat scanner. The total cost for this equipment is less than EUR 10,000.–. The unlabeled primers and all other supplies are very inexpensive so that the genotyping of an individual patient will cost less than EUR 10.–.

Thus, the 40-indel panel described here and the newly developed ‘low-tech’ analysis platform represent useful new tools for forensic identification and kinship analysis in laboratories with limited budgets, especially in developing countries.

Acknowledgements

This work was supported by a grant from FINEP (Financiadora de Estudos e Projetos), a funding agency of the Brazilian Ministry of Science and Technology (Contract No. 01.07.0635.00). The specific set of 40 insertion-deletion (indel) polymorphisms and its genotyping methodologies are the subject of a Patent Application made by GENE – Núcleo de Genética Médica at the Instituto Nacional de Propriedade Industrial (INPI) of Brazil.

Disclosure Statement

GENE – Núcleo de Genética Médica is a private medical genetics institute and the senior author, Sérgio Danilo Junho Pena is its President. The specific set of 40 insertion-deletion (indel) polymorphisms and its genotyping methodologies are the subject of a Patent Application made by GENE – Núcleo de Genética Médica at the Instituto Nacional de Propriedade Industrial (INPI) of Brazil.

References

- Mullaney JM, Mills RE, Pittard WS, Devine SE: Small insertions and deletions (INDELs) in human genomes. *Hum Mol Genet* 2010;19:R131–R136.
- Weber JL, David D, Heil J, Fan Y, Zhao C, Marth G: Human diallelic insertion/deletion polymorphisms. *Am J Hum Genet* 2002;71:854–862.
- Bastos-Rodrigues L, Pimenta JR, Pena SDJ: The genetic structure of human populations studied through short insertion-deletion polymorphisms. *Ann Hum Genet* 2006;70:658–665.
- Pena SDJ, Bastos-Rodrigues L, Pimenta JR, Bydlowski SP: DNA tests probe the genomic ancestry of Brazilians. *Braz J Med Biol Res* 2009;42:870–876.

- 5 Pena SDJ, Di Pietro G, Fuchshuber-Moraes M, Genro JP, Hutz MH, Kehdy Fde S, Kohlrausch F, Magno LA, Montenegro RC, Moraes MO, de Moraes ME, de Moraes MR, Ojopi EB, Perini JA, Racciopi C, Ribeiro-Dos-Santos AK, Rios-Santos F, Romano-Silva MA, Sortica VA, Suarez-Kurtz G: The genomic ancestry of individuals from different geographical regions of Brazil is more uniform than expected. *PLoS One* 2011;16:e17063.
- 6 Suarez-Kurtz G, Vargens DD, Struchiner CJ, Bastos-Rodrigues L, Pena SDJ: Self-reported skin color, genomic ancestry and the distribution of GST polymorphisms. *Pharmacogenet Genomics* 2007;17:765–771.
- 7 Pena SDJ: Segurança pública: determinação de identidade genética pelo DNA. *Parcerias Estratégicas* 2005;20:447–460.
- 8 Pimenta JR, Pena SDJ: Efficient human paternity testing with a panel of 40 short insertion-deletion polymorphisms. *Genet Mol Res* 2010;30:601–607.
- 9 Pereira R, Phillips C, Alves C, Amorim A, Caracado A, Gusmão L: A new multiplex for human identification using insertion/deletion polymorphisms. *Electrophoresis* 2009 30:3682–3690.
- 10 Freitas NS, Resque RL, Ribeiro-Rodrigues EM, Guerreiro JF, Santos NP, Ribeiro-dos-Santos A, Santos S: X-linked insertion/deletion polymorphisms: forensic applications of a 33-markers panel. *Int J Legal Med* 2010 124:589–593.
- 11 Cann HM, de Toma C, Cazes L, Legrand MF, Morel V, Piouffre L, Bodmer J, Bodmer WF, Bonne-Tamir B, Cambon-Thomsen A, Chen Z, Chu J, Carcassi C, Contu L, Du R, Excoffier L, Ferrara GB, Friedlaender JS, Groot H, Gurwitz D, Jenkins T, Herrera RJ, Huang X, Kidd J, Kidd KK, Langaney A, Lin AA, Mehdi SQ, Parham P, Piazza A, Pistillo MP, Qian Y, Shu Q, Xu J, Zhu S, Weber JL, Greely HT, Feldman MW, Thomas G, Dausset J, Cavalli-Sforza LL: A human genome diversity cell line panel. *Science* 2002;296:261–262.
- 12 PROMEGA: GenePrint STR Systems (Silver Stain Detection) Technical Manual. www.promega.com/~media/Files/Resources/Protocols/Technical%20Manuals/101/GenePrint%20STR%20Systems%20Protocol.ashx.