

## Efficient human paternity testing with a panel of 40 short insertion-deletion polymorphisms

J.R. Pimenta<sup>1</sup> and S.D.J. Pena<sup>1,2</sup>

<sup>1</sup>GENE - Núcleo de Genética Médica, Belo Horizonte, MG, Brasil

<sup>2</sup>Departamento de Bioquímica e Imunologia,  
Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brasil

Corresponding author: S.D.J. Pena  
E-mail: spena@dcc.ufmg.br

Genet. Mol. Res. 9 (1): 601-607 (2010)  
Received March 2, 2010  
Accepted March 23, 2010  
Published March 30, 2010

**ABSTRACT.** We developed a panel of 40 multiplexed short insertion-deletion (indel) polymorphic loci with widespread chromosomal locations and allele frequencies close to 0.50 in the European population. We genotyped these markers in 360 unrelated self-classified White Brazilians and 50 mother-child-probable father trios with proven paternity. The average heterozygosity (gene diversity) per locus was 0.48, and the combined probability of identity (matching probability) for the 40-locus set was  $3.48 \times 10^{-17}$ . The combined power of exclusion of the indel panel was 0.9997. The efficiency of the 40 indel set in the exclusion of falsely accused individuals in paternity casework was equivalent to the CODIS set of 13 microsatellites. The geometric mean of the paternity indices of the 50 mother-child-probable father trios was 17,607. This panel of 40 short indels was found to have excellent performance. Thus, especially because of its simplicity and low cost, and the fact that it is composed of genomic markers that have very low mutation rates, it represents a useful new tool for human paternity testing.

**Key words:** Paternity testing; Insertion-deletion polymorphisms; Single nucleotide polymorphisms; DNA; Indels

## INTRODUCTION

From a social standpoint, paternity testing using DNA markers is one of the most revolutionary products of modern human molecular genetics. Indeed, thorny judicial or private paternity disputes, previously unsolved in the pre-DNA era, can now be easily and quickly settled with great scientific reliability. Introduced only 25 years ago (reviewed in Jeffreys and Pena, 1993), DNA testing is now an indispensable tool in family law. In the United States, the American Association of Blood Banks registered 420,740 paternity cases in 2006 (AABB, 2006). Many more were certainly done privately.

Data from the American Association of Blood Banks (AABB, 2006) also indicate the recent occurrence of a remarkable technological convergence in DNA paternity testing - 98.53% of the American cases were performed using multiplex microsatellite (short tandem repeat) genetic markers; these were tested using polymerase chain reaction (PCR).

However, microsatellite analysis has known drawbacks, mostly because of the high mutation rate of these loci, which averages  $1.5 \times 10^{-3}$  (Leopoldino and Pena, 2003). For some of them, such as *ACTBP2* (*SE33*), the mutation rate can approach 1% (Klitsch and Neuhuber, 1998). Also, microsatellite typing is relatively complex, because there are a large number of allelic states, occasionally with irregular sizes. Although tetranucleotide and pentanucleotide repeat loci are commonly used, alleles with non-standard repeat numbers may emerge by mutation, complicating the analysis. For instance, there is a relatively common 9.3 allele at locus *TH01* that differs by only one base from the rarer 10 allele (Mansfield et al., 1998).

Along with microsatellites, other polymorphisms are also abundant in the human genome, especially single base substitutions (single nucleotide polymorphisms - SNPs) and short (1-20 nucleotide pairs) insertions and deletions. These two types of polymorphisms are diallelic and hence are individually less informative than microsatellites. However, they have useful characteristics: i) analysis is simple and quick, since there are only two alleles, which have fixed known sizes; ii) both SNPs and short indels can be typed in very short amplicons (50 bp or less), which is important when DNA is degraded; iii) most importantly, the mutation rate of both SNPs and short indels is roughly  $10^{-8}$ , several orders of magnitude smaller than that of microsatellites (Saitou and Ueda, 1994; Nachman and Crowell, 2000).

The use of SNPs has been growing exponentially in human genetics and, recently, both experimental and theoretical studies have addressed their advantages and possible usefulness for paternity testing (Inagaki et al., 2004; Ayres, 2005; Phillips et al., 2008; Børsting et al., 2009). Among the diallelic polymorphic loci, SNPs are the most abundant and best studied; more than 15 million human SNPs have been identified and mapped (see, <http://www.ncbi.nlm.nih.gov/projects/SNP/>). The typing of SNPs, which depends on qualitative base identification, is best done by microarray hybridization; however, this requires complex equipment that is not generally available in forensic laboratories. On the other hand, the alleles of short indels 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.

Weber et al. (2002) characterized 2000 human diallelic short indels in the human genome. Among them, we identified 40 polymorphisms that fulfilled the following criteria: widespread chromosomal location, increased amplicon sizes to allow multiplex analysis and allele frequencies close to 0.50 in the European population (Bastos-Rodrigues et al., 2006; Pena et al., 2009). We decided to evaluate these 40 indel markers for paternity testing.

## MATERIAL AND METHODS

### DNA samples

Among paternity casework from GENE - Núcleo de Genética Médica, we randomly selected 50 mother-child-probable father trios with a paternity index above 100,000 based on multiplex microsatellite analysis (Pena, 1999). We also studied 360 unrelated self-classified White Brazilians from different geographical regions of Brazil (Pena et al., 2009).

### Genotyping with the panel of 40 short indels

DNA from each individual was independently typed for the 40 diallelic short insertion/deletion polymorphisms (short indels; Table 1). The composition of the primer pairs is available at <http://www.ncbi.nlm.nih.gov/projects/SNP/>. To each forward PCR primer, we added a tail of the M13-40 17-mer oligonucleotide.

The PCR amplifications were made with four multiplex reaction systems, each consisting of a mix of 7-13 primer pairs (Bastos-Rodrigues et al., 2006). The multiplex PCR assay was performed in a 10- $\mu$ L final volume of the following: 1X PCR buffer (10 mM Tris-HCl, pH 8.3 or pH 9.2, 75 mM KCl, 3.5 mM MgCl<sub>2</sub>), 200  $\mu$ M dNTPs, 1.0 U Platinum Taq DNA polymerase (Invitrogen), 20 ng genomic DNA, 1.5  $\mu$ M M13-40 forward primer labeled with the Cy5 dye, 1.5  $\mu$ M of each unlabeled reverse primer, and 0.1  $\mu$ M of each unlabeled forward primer.

Two microliters of labeled PCR products was denatured in formamide solution and subjected to fragment analysis using an ALF-Express (GE Healthcare), according to manufacturer instructions. Analyses of allele sizes were scored using AlleleLinks (GE Healthcare).

### Data analysis

Allele frequencies were calculated by the simple counting method, and the probabilities of identity (matching probability) for each locus and for the whole panel were computed as described by Butler (2005). Hardy-Weinberg equilibrium was evaluated using the exact test (Guo and Thompson, 1992).

The average power of exclusion  $Z$  of the indel polymorphisms was estimated as  $Z = pq(1 - pq)$ , and gene diversity was determined as  $h = 2pq$  (Krawczak, 1999). The number  $N(p)$  of power-effective maximally informative SNPs was estimated as twice the number of diallelic loci multiplied by their average gene diversity (Krawczak, 1999).

The paternity index for individual markers was calculated in a standard fashion (Gjertson et al., 2007) as a likelihood ratio. The combined paternity index was determined as the product of all the individual paternity indices (Gjertson et al., 2007).

## RESULTS

We had very good success in genotyping all paternity trios and all unrelated Brazilian individuals at all 40 short indel loci with only four multiplex electrophoretic runs in an ALF-Express fluorescent automatic sequencer. Similar success has been obtained in a MegaBace 1000 sequencer (Bastos-Rodrigues et al., 2006) and ABI PRISM 3130.

Allele frequencies for each of the 40 insertion-deletion (indel) polymorphisms were calculated from 360 unrelated White Brazilians (Table 1). There were no deviations from expected Hardy-Weinberg proportions.

**Table 1.** Nomenclature and genetic characteristics of the 40 short insertion-deletion polymorphic loci used for paternity testing.

Indel MID#	SNP-ID*	Allele frequency**	Gene diversity	Matching probability	Exclusion power
1	rs3917	0.5000	0.5000	0.3924	0.1793
15b	rs4181	0.5719	0.4897	0.3803	0.1849
17	rs4183	0.6126	0.4747	0.3886	0.1810
51	rs16343	0.4972	0.5000	0.3750	0.1875
89	rs16381	0.4072	0.4828	0.3841	0.1831
107	rs16394	0.6288	0.4668	0.3932	0.1789
131	rs16415	0.4802	0.4992	0.3754	0.1873
132	rs16416	0.4637	0.4974	0.3763	0.1868
150	rs16430	0.6164	0.4729	0.3897	0.1805
159	rs16438	0.3675	0.4649	0.3944	0.1784
170	rs16448	0.5480	0.4954	0.3773	0.1863
258	rs16695	0.4955	0.5000	0.3750	0.1875
278	rs16715	0.3743	0.4684	0.3923	0.1793
420	rs140709	0.5481	0.4954	0.3774	0.1863
444	rs140733	0.6954	0.4237	0.4219	0.1670
468	rs140757	0.4880	0.4997	0.3751	0.1874
470	rs140759	0.6277	0.4674	0.3929	0.1791
663	rs1305047	0.4951	0.5000	0.3750	0.1875
788	rs1610874	0.4762	0.4989	0.3756	0.1872
857b	rs1610942	0.3111	0.4286	0.4183	0.1684
914	rs1610997	0.5953	0.4818	0.3846	0.1829
918	rs1611001	0.4950	0.5000	0.3750	0.1875
1002	rs1611084	0.4856	0.4996	0.3752	0.1874
1092	rs2067180	0.5072	0.4999	0.3751	0.1875
1100	rs2067188	0.4790	0.4991	0.3754	0.1873
1129	rs2067217	0.6922	0.4261	0.4201	0.1677
1291	rs2067373	0.5887	0.4843	0.3832	0.1835
1352	rs2307548	0.5712	0.4899	0.3802	0.1849
1428	rs2307624	0.4824	0.4994	0.3753	0.1873
1537	rs2307733	0.5621	0.4923	0.3789	0.1856
1549	rs2307745	0.4374	0.4922	0.3790	0.1855
1586	rs2307782	0.5315	0.4980	0.3760	0.1870
1642	rs2307838	0.4845	0.4995	0.3752	0.1874
1654	rs2307850	0.6978	0.4218	0.4233	0.1664
1759	rs2307955	0.6653	0.4454	0.4068	0.1731
1763	rs2307959	0.5026	0.5000	0.3750	0.1875
1847	rs2308043	0.3011	0.4209	0.4239	0.1662
1861	rs2308057	0.6830	0.4330	0.4152	0.1696
1943	rs2308135	0.5207	0.4991	0.3754	0.1873
1952	rs2308144	0.6927	0.4257	0.4204	0.1676

\*Entrez SNP (<http://www.ncbi.nlm.nih.gov/snp>). \*\*Allele frequency in 360 self-declared White Brazilians.

The average heterozygosity (gene diversity) per locus was 0.48, very close to the theoretical maximum, which is 0.50. The combined probability of identity (matching probability) for the 40-locus panel was  $3.48 \times 10^{-17}$ .

The power of exclusion  $Z$  of a genetic polymorphism is defined as the prior probability that in a trio of mother, child, and falsely alleged father, the latter will not carry the necessary paternal allele deduced from the genotypes of mother and child (Krawczak, 1999). This calculation was performed for each of the 40 loci studied (Table 1). The combined power of exclusion of the whole panel was 0.9997.

Krawczak (1999) created a measure of the informativity of a battery of diallelic polymorphic loci, which he called the number  $N(p)$  of power-effective maximally informative SNPs. This number  $N(p)$  can be calculated as twice the number of diallelic loci multiplied by their average gene diversity. In our case, the 40 loci correspond to 38.2 maximally effective SNPs.

After genotyping the 50 mother-child-probable father trios, the paternity index for each of 40 insertion-deletion polymorphic loci studied was calculated in a standard fashion (Gjertson et al., 2007) as a likelihood ratio. In other words, the paternity index is the ratio of the likelihood of obtaining the results observed under the hypothesis that the tested man is the biological father, divided by the likelihood of obtaining the results observed under the hypothesis that a random man from the population is the biological father. Based on the typing results of the panel of 40 insertion-deletion polymorphisms, the 50 paternity indices were calculated and their geometric mean was 17,607. Assuming an *a priori* paternity probability of 0.5, we can use Bayes' theorem to calculate a posterior probability of paternity (W) of 99.994%.

## DISCUSSION

Given the results obtained in this study, especially the average exclusion power of 0.9997 and the average paternity index for 50 mother-child-father trios of 17,677, the panel of 40 insertion-deletion polymorphisms seems to be very efficient and apt for paternity testing. However, it would be desirable to make a comparison of its effectiveness with presently used microsatellite sets.

We chose for comparison the widely used CODIS battery of 13 tetrameric microsatellites, which was designed for criminal investigations and contains the following loci: CSF1PO, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D21S11, FGA, TH01, TPOX, and VWA (Budowle et al., 1999). Using the expected heterozygosity data published by Budowle et al. (1999) for American Caucasians, we can calculate an average gene diversity of 0.789. From that it is possible to estimate the average power of exclusion of the 13-locus CODIS set as  $h^2[1 - (1 - h) / h^2]$  (Butler, 2005), which numerically computes as 0.541.

If the CODIS panel is used for a paternity testing study, the probability of obtaining at least one incompatibility in the panel of 13 microsatellites can be computed as  $[1 - (1 - 0.541)^{13}] = 0.99996$ , which is considerably larger than the 0.9997 calculated for the panel of 40 indels. However, we have to take into account the fact that the mutation rates of microsatellites are several orders of magnitude higher than those of short indels (Saitou and Ueda, 1994). Indeed, because of the very low mutation rates of the insertion-deletion polymorphisms, inconsistencies are much stronger when typing is done with them; a single inconsistency would in theory be sufficient to exclude paternity, as was the case when coding genes (blood groups, HLA antigens and polymorphic proteins) were employed in paternity testing (Cifuentes et al., 2006).

The American Association of Blood Banks has issued standards for parentage testing laboratories regarding mutations. For microsatellite testing, American Association of Blood Banks standard 6.4.1 states that a conclusion of non-paternity shall not be made on the basis of exclusion at a single DNA locus, i.e., a single inconsistency (Butler, 2005). Thus, most parentage testing laboratories using microsatellite testing employ a "two-exclusion" rule (Butler, 2005) or even a "three-exclusion" rule, since it is not uncommon to see two inconsistencies between a child and the true biological father (Cifuentes et al., 2006). Indeed, in its annual report for 2006, AABB reported 47 instances of double mutations in microsatellite tests (AABB, 2006).

Thus, for microsatellites the criterion of power of exclusion should be changed to at least two, or even three, genetic incompatibilities (Cifuentes et al., 2006). The probability of obtaining at least two exclusions for the CODIS panel can be calculated as  $\{1 - (1 - 0.541)^{13} - 13[0.541(1 - 0.541)^{12}]\} = 0.9994$ , which is actually a little lower than that of obtaining at least one exclusion with indels, which is 0.9997. Thus, the efficiency of the 40 indel panel in the exclusion of falsely accused individuals in paternity casework is equivalent to the CODIS set of 13 microsatellites.

In conclusion, the panel of 40 short indels that we developed gave excellent performance. Thus, especially because of its simplicity and low cost, and the fact that it is composed of genomic markers that have very low mutation rates, it represents a useful new tool for human paternity testing.

## ACKNOWLEDGMENTS

Research supported by CNPq of Brazil. J.R. Pimenta was the recipient of a fellowship from the Rhae Program of CNPq.

## REFERENCES

- American Association of Blood Banks (AABB) (2006). Annual Report Summary for Testing in 2006. Available at [<http://www.gep-isfg.org/documentos/AABB%202006.pdf>]. Accessed March 1, 2010.
- Ayres KL (2005). The expected performance of single nucleotide polymorphism loci in paternity testing. *Forensic Sci. Int.* 154: 167-172.
- 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.
- Børsting C, Rockenbauer E and Morling N (2009). Validation of a single nucleotide polymorphism (SNP) typing assay with 49 SNPs for forensic genetic testing in a laboratory accredited according to the ISO 17025 standard. *Forensic Sci. Int. Genet.* 4: 34-42.
- Budowle B, Moretti TR, Baumstark AL, Defenbaugh DA, et al. (1999). Population data on the thirteen CODIS core short tandem repeat loci in African Americans, U.S. Caucasians, Hispanics, Bahamians, Jamaicans, and Trinidadians. *J. Forensic Sci.* 44: 1277-1286.
- Butler JM (2005). *Forensic DNA Typing: Biology, Technology and Genetics of STR Markers*. 2nd edn. Elsevier Academic Press, Amsterdam.
- Cifuentes LO, Martinez EH, Acuna MP and Jonquera HG (2006). Probability of exclusion in paternity testing: time to reassess. *J. Forensic Sci.* 51: 349-350.
- Gjertson DW, Brenner CH, Baur MP, Carracedo A, et al. (2007). ISFG: Recommendations on biostatistics in paternity testing. *Forensic Sci. Int. Genet.* 1: 223-231.
- Guo SW and Thompson EA (1992). Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics* 48: 361-372.
- Inagaki S, Yamamoto Y, Doi Y, Takata T, et al. (2004). A new 39-plex analysis method for SNPs including 15 blood group loci. *Forensic Sci. Int.* 144: 45-57.
- Jeffreys AJ and Pena SDJ (1993). A Brief Introduction to Human DNA Fingerprinting. In: *DNA Fingerprinting: State of the Science* (Pena SDJ, Chakraborty R, Eppel JT and Jeffreys AJ, eds.). Birkhäuser Verlag, Basel, 1-20.
- Klitschar M and Neuhuber F (1998). A study on the short tandem repeat system ACTBP2 (SE33) in an Austrian population sample. *Int. J. Legal Med.* 111: 46-48.
- Krawczak M (1999). Informativity assessment for biallelic single nucleotide polymorphisms. *Electrophoresis* 20: 1676-1681.
- Leopoldino AM and Pena SD (2003). The mutational spectrum of human autosomal tetranucleotide microsatellites. *Hum. Mutat.* 21: 71-79.
- Mansfield ES, Robertson JM, Vainer M, Isenberg AR, et al. (1998). Analysis of multiplexed short tandem repeat (STR) systems using capillary array electrophoresis. *Electrophoresis* 19: 101-107.
- Nachman MW and Crowell SL (2000). Estimate of the mutation rate per nucleotide in humans. *Genetics* 156: 297-304.
- Pena SDJ (1999). Single-tube single-colour multiplex PCR amplification of ten polymorphic microsatellites (ALF10): a

- new powerful tool for DNA profiling. *Pure Appl. Chem.* 71: 1683-1690.
- Pena SD, 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.
- Phillips C, Fondevila M, Garcia-Magarinos M, Rodriguez A, et al. (2008). Resolving relationship tests that show ambiguous STR results using autosomal SNPs as supplementary markers. *Forensic Sci. Int. Genet.* 2: 198-204.
- Saitou N and Ueda S (1994). Evolutionary rates of insertion and deletion in noncoding nucleotide sequences of primates. *Mol. Biol. Evol.* 11: 504-512.
- Weber JL, David D, Heil J, Fan Y, et al. (2002). Human diallelic insertion/deletion polymorphisms. *Am. J. Hum. Genet.* 71: 854-862.