Int J Legal Med (2010) 124:589–593
DOI 10.1007/s00414-010-0441-9

ORIGINAL ARTICLE

X-linked insertion/deletion polymorphisms: forensic applications of a 33-markers panel

Natalle S. C. Freitas · Rafael L. Resque · Elzemar M. Ribeiro-Rodrigues · João F. Guerreiro · Ney P. C. Santos · Ândrea Ribeiro-dos-Santos · Sidney Santos

Received: 29 December 2009 / Accepted: 25 February 2010 / Published online: 31 March 2010

Abstract Insertion/deletion (INDEL) polymorphisms are diallelic markers with potential characteristics for use in forensics and biological anthropology, including: the simplicity of laboratory analysis, the possibility of genotyping many markers in a single PCR multiplex reaction, as well as analyzing markers with special inheritance types, such as those linked to the X chromosome (X-INDEL). In this work we developed a laboratory analysis methodology using a 33-INDEL marker panel for the X chromosome in a single PCR multiplex reaction, followed by a capillary electrophoresis run. We employed the panel to genotype a sample of 351 individuals of a mixed population from the Brazilian Amazon. The results demonstrate that the measurement of biostatistical parameters for forensic use in this population is compatible with prior estimates from other populations using current X-STR panels.

Keywords INDEL polymorphism · X chromosome · Amazon · PCR multiplex · X-INDEL · Panel

Electronic supplementary material The online version of this article (doi:10.1007/s00414-010-0441-9) contains supplementary material, which is available to authorized users.

N. S. C. Freitas · R. L. Resque · E. M. Ribeiro-Rodrigues · J. F. Guerreiro · N. P. C. Santos · Â. Ribeiro-dos-Santos · S. Santos (✉)
Laboratório de Genética Humana e Médica, Instituto de Ciências Biológicas, Universidade Federal do Pará,
Cidade Universitária Prof. José da Silva Netto. Av. Augusto Corrêa, 01,
Belém, PA, Brazil
e-mail: sidneysantos@ufpa.br

Introduction

Recent advances in forensic genetics focus on the development of joint genotyping panels for multiple diallelic markers involved in the substitution of a single nucleotide polymorphism (SNP) or those based on the insertion or deletion (INDEL) of a few nucleotides [1–4].

Single nucleotide polymorphism methodologies of analysis typically employ complex genotyping protocols that require many steps and/or the implementation of new and expensive methodologies and technologies [2, 5]. By contrast, the analysis of INDEL markers can be carried out using relatively simple and inexpensive methodologies that are commonly used in STR analysis [1].

INDEL markers have many genetic advantages for analytical use: they are widely spread throughout the genome [6, 7], including the sexual chromosome pair, X and Y; all of the polymorphisms derive from a single mutation event; they have reduced mutation rates; and they can represent significant allelic frequency differences between geographically distinct populations. Additionally, INDEL can be genotyped in small amplicons, increasing the success rate of analysis from degraded DNA [1, 2].

Despite the many potential advantages, studies that employ the variability of INDEL markers in forensic practice [2, 3] and biological anthropology [8, 9] have been published only recently. Considering the potential use of diallelic INDEL as genetic markers in forensic analysis, we attempted to develop and validate a panel of 33 X chromosome INDEL (X-INDEL). All of the chosen markers have a short length in allelic variation (2–30 bp) that can be genotyped in a single multiplex PCR reaction, followed by capillary electrophoresis.

The developed panel was used to genotype a sample of 351 individuals of a mixed population from the Brazilian

Amazon. The data were used to estimate a series of important biostatistical parameters and test the efficiency of this panel in forensic practice, mainly as a complementary tool in complex paternity investigations.

Materials and methods

Samples and DNA extraction

The study was carried out in accordance with the Declaration of Helsinki (2000) of the World Medical Association. Blood samples were collected from healthy, unrelated individuals under informed consent. The samples consisted of 351 individuals (249 males, 102 females) of a population from the Brazilian Amazon, Belém, Brazil. DNA was extracted using a standard phenol–chloroform protocol [10].

PCR amplification and genotyping

DNA amplification was performed in a single multiplex reaction (primer pairs and primer concentrations are shown in Supplementary Table ST1) under the following conditions: preincubation (11 min at 95°C); 10 cycles (1 min at 94°C, 1 min at 60°C, and 2 min at 70°C); 17 cycles (1 min at 90°C, 1 min at 60°C, and 2 min at 70°C); and a final extension (60 min at 60°C). PCR amplification was performed in a 12.5 μL reaction volume, including 1 X PCR buffer with 3 mM $MgCl_2$, 125 μM dNTPs, 2 U AmpliTaq Platinum DNA Polymerase (Invitrogen Life Technologies, Carlsbad, CA, USA), primers, and 5–10 ng DNA.

PCR products were subsequently prepared for capillary electrophoresis [11]. Separation and detection were performed with an ABI PRISM 3130 Genetic Analyzer, using a GS-500 LIZ standard size, a G5 filter set and POP7 polymer (Applied Biosystems, Foster City, CA, USA). The samples were genotyped with GeneMapper v4.0 software (Applied Biosystems).

Statistical analysis

The allele frequency for each X-INDEL was determined by counting. The heterozygosity (Het) and linkage disequilibrium (LD) exact test were assessed using Arlequin v3.1 [12]. Hardy–Weinberg equilibrium and possible deviations were tested using the female genotypes only [12]. Power of discrimination in females (PD_F), power of discrimination in males (PD_M), probability of exclusion in trios involving daughters (PE_T), and in father/daughter duos lacking maternal genotype information (PE_M) were computed according to Desmarais et al. [13].

Results and discussion

The pool of markers investigated in this study was chosen based on previously published studies [6, 7]. We selected the markers from a list of INDEL polymorphisms previously characterized in populations from different geographical origins, available on the website (http://www.marshfieldclinic.org/mgs/pages/default.aspx?page=DIDP).

Thirteen of the 33 investigated markers are part of the X-INDEL marker panel previously described [1]. The new markers were chosen based on their physical location along the X chromosome and if the following information were true when available: mean heterozygosity over 30% in European, African, and Asian individuals, as well as an allele length variation of 2-30 base pairs.

We chose physically close markers in an attempt to identify INDEL blocks with pronounced linkage disequilibrium in allele transmission. The criterion we used required markers to have a maximum distance of 250 Kb, as suggested by Zhu et al. [14].

Using these criteria, we selected six possible marker blocks: Group I (MID357 and MID356—distance of 5.3 Kb), Group II (MID 3780, M448804, MID3703, MID218, and MID3774—greatest distance of 173 Kb), Group III (MID3705, MID 3706, M304737, and MID3692—greatest distance of 142.6 Kb), Group IV (M197147, MID3754, and MID3756—greatest distance of 166 Kb), Group V (MID3764, M284601, and M103547—greatest distance of 40.8 Kb), and Group VI (MID358, MID3763, and MID3728—greatest distance of 224 Kb).

Primer and multiplex assay design

Primers were designed using the software PRIMER3 (http://www.genome.wi.mit.edu/cgi-bin/primer/primer3) and screened for hairpins and primer dimers with Auto Dimer heck (http://www.cstl.nist.gov/biotech/strbase/AutoDimerHomepage). The primer pairs were checked for non-specific hybridizations in other genome regions using the BLAT software (BLAST-like Alignment Tool) at http://genome.brc.mcw.edu/cgi-bin/hgBlat. All markers were then organized by expected amplicon length and assigned to four different dye-labeling fluorochromes.

The final panel was composed of 33 X-INDEL markers that can be amplified with demonstrated reproducibility in a fast and easy PCR multiplex reaction followed by electrophoresis (Fig. 1). The panel developed here proves to be capable of amplifying all of the markers in samples extracted by different methods and those containing different amounts of largely variable DNA (0.5–30 ng). The best results were obtained for DNA samples of 5 ng in a final volume of 12.5 μL.

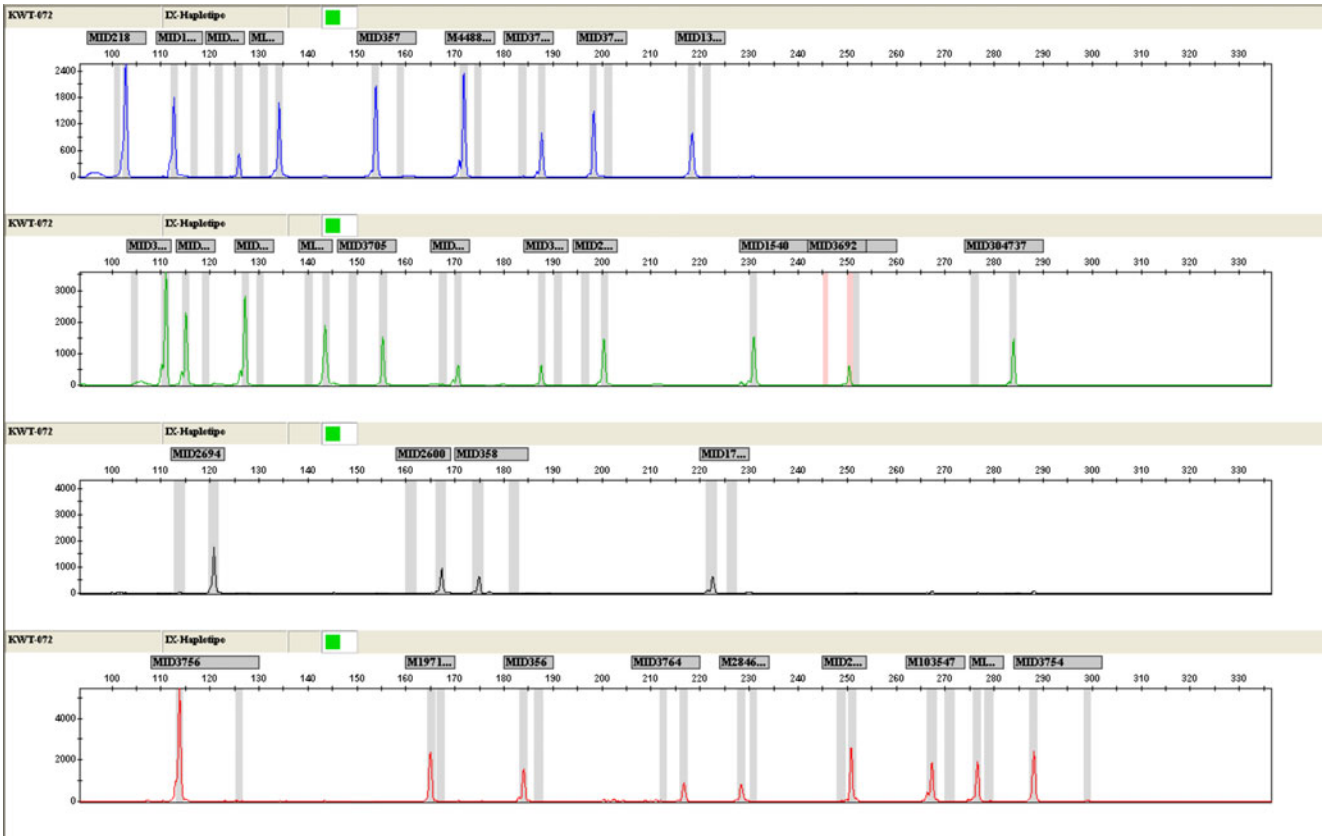

Fig. 1 Electropherogram of the 33 X-INDELs amplified using a 5-ng DNA sample from an individual of the Amazon population

Genetic variation

The genetic variability of the 33 INDEL was investigated in a sample of 351 individuals from a mixed population from the Brazilian Amazon. The data from the sample of investigated females were used to test the exact Hardy–Weinberg equilibrium. After correcting for multiple analyses [15], only the marker MID3756 ($p<0.003$) showed significant H-W disequilibrium.

The genotyping of this system revealed a heterozygote deficit on the order of 20%, which may indicate the occurrence of a null allele. To address this possibility, we investigated some SNP in the flanking regions of this marker using a recent dbSNP build (129). We could not identify any polymorphism in these regions to explain the presence of a null allele. Most likely, the heterozygote deficit resulted from the small sample size of investigated females. Therefore, this marker should not be included in the panel for future forensic investigations until we can clarify the cause of this H-W disequilibrium.

Linkage disequilibrium

The data were used to test the LD between all marker pairs within the supposed linkage groups described above (marker pairs with a maximum distance of 250 Kb). Among the 20 possible marker pairs, only 10 presented significant LD. The physical distance between the 10 pairs of markers with a significant p value varied from 4.5 Kb (MID3705/MID3706) to 115 Kb (MID3763/MID3728).

We tried to estimate the LD distance in X chromosomes based on these investigated diallelic markers. Only one out of the 10 pairs of markers in LD had a physical distance larger than 50 Kb (MID3763/MID3728, 105 Kb). Therefore, the data for X-INDEL markers in the mixed population of the Amazon indicate that most of the basic linkage blocks within the X chromosome are 50 Kb in size. These results must be taken with caution, as the number of investigated blocks is limited and not representative of all regions of the X chromosome, which is known to have different rates of recombination along its length [16, 17].

The LD data were used to define the linkage blocks of the investigated INDEL markers. The linkage blocks were identified based on pairwise D' values, as described by Zhu et al. [18]. Intervals in which all INDEL have a pairwise D' value >0.8 are identified and assumed that they constitute the basic blocks. Following this criterion, we identified six linkage blocks: Block I (MID357 and MID356), Block II (MID3780 and M448804), Block III (MID218 and MID3774), Block IV (MID3705, MID3706, and

M304737), Block V (MID3764, M284601, and M103547), and Block VI (MID3763 and MID3728).

In the identified linkage blocks, a set of alleles from different markers were inherited as haplotypes. We attempted to identify the linkage blocks, assuming that these haplotypes can be informative and constitute an excellent tool for forensic cases, especially in paternity investigations of female children when the alleged father cannot be genotyped.

Forensic parameters

The genotyping data were used to estimate the allele frequencies and a series of biostatistical parameters of forensic interest (see Supplementary Table ST2). The mean heterozygosity of the INDEL panel was 0.373, with most markers (25/33) with heterozygosity values greater than 30%. When the six LD blocks were considered, as previously described, the mean heterozygosity was 0.402 (25 systems with independent segregation).

The 33 X-INDEL marker panel used in this work was evaluated for potential application in human identification and kinship testing in the Belém population: high female (99.99999995%) and male (99.99988%) overall discrimination power values were obtained, as well as high overall exclusion power values in father/mother/daughter trios (99.9971%) and in father/daughter duos (99.92%). The calculations for the markers that formed linkage blocks were performed based on the haplotype frequencies rather than the independent allele frequencies. This procedure reduced the panel to 25 non-linked markers. The estimates obtained with the X-INDEL panel in Belém are compatible with those observed in other populations in the world using eight [19] and 10 Short Tandem Repeat panels of the X chromosome [20].

Additionally, due to the high mutation rate of STR, Mendelian incompatibility situations between the alleged father and daughter (incompatibility of one or more X-STR or even autosomal STR) is a common finding in paternity investigations. Therefore, the use of panels with a large number of markers with a low mutation rate, such as the X-INDEL panel developed here is preferable.

Conclusions

In this work, we describe a panel of 33 diallelic X-INDEL and demonstrate its application to the human identification processes and kinship testing, based on data of a mixed population from the Amazon.

We identified six linkage blocks among the investigated markers that constituted either two (four blocks) or three (two blocks) INDEL. The extension of the LD between the diallelic markers of the X chromosome may be smaller than 50 Kb, although we observed LD between markers with a distance of 105 Kb. Confirmation of this estimate requires further investigation, preferably in non-mixed populations, such as European, African, Asian, and Native American populations.

In accordance with other recent studies on INDEL markers, we propose that the approach adopted in this study is efficient, robust, and can be easily adopted in forensic laboratories without the requirement of additional equipment or software, while using the current STR analysis technology.

Finally, we have demonstrated that this marker panel affords estimates with forensic use (Het, PD_M, PD_F, PE_T, and PE_M) very similar to the currently used X-STR panels, which are extensively described in the literature, with the additional advantage of working with reduced mutation rate systems. The estimates led to our conclusion that the X-INDEL panel proposed here can be used as an auxiliary of the STR marker panel in forensic practice, especially when the alleged father cannot be genotyped or in cases of probable Mendelian inconsistencies due to mutation.

Acknowledgments This work was supported by the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), and Financiadora de Estudos e Projetos (FINEP).

Reference

1. Ribeiro-Rodrigues EM, Santos NPC, Ribeiro-dos-Santos AKC, Pereira R, Amorim A, Gusmão L, Zago MA, Santos SEB (2009) Assesing interethnic admixture using an X-linked insertion deletion multiplex. Am J Hum Biol 21:707–709
2. Pereira R, Phillips C, Alves C, Amorim A, Carracedo A, Gusmão L (2009) Insertion/deletion polymorphisms: a multiplex assay and forensic applications. Forensic Sci Int Genet 2:513–515
3. Edelmann J, Hering S, Augustin C, Szibor R (2009) Indel polymorphisms an additional set of markers on the X-chromosome. Forensic Sci Int Genet 2:510–512
4. Tomas C, Stangegaard M, Borsting C, Hansen AJ, Morling N (2008) Typing of 48 autosomal SNPs and amelogenin with Genplex SNP genotyping system in forensic genetics. Forensic Sci Int 3:1–6
5. Sanchez JJ, Phillips C, Borsting C, Balogh K, Bogus M, Fondevila M, Harrison CD, Brown EM, Salas A, Court DS, Schneider PM, Carracedo A, Morling N (2006) A multiplex assay with 52 single nucleotide polymorphisms for human identification. Electrophoresis 27:1713–1724
6. Weber JL, David D, Heil J, Fan Y, Zhao C, Marth G (2002) Human diallelic insertion/deletion polymorphisms. Am J Hum Genet 71:854–862
7. Mills RE, Luttig CT, Larkins CE, Beauchamp A, Tsui C, Pittard WS, Devine SE (2006) An initial map of insertion and deletion (INDEL) variation in the human genome. Genome Res 16:1182–1190
8. Bastos-Rodrigues L, Pimenta JR, Pena SD (2006) The genetic structure of human populations studied through short insertion-deletion polymorphisms. Ann Hum Genet 70:658–665
9. Santos NPC, Ribeiro-Rodrigues EM, Ribeiro-dos-Santos AKC, Pereira R, Gusmão L, Amorim A, Guerreiro JF, Zago MA, Matte C,

Hutz M, SEB S (2010) Assessing individual interethnic admixture and population substructure using a 48 insertion-deletion (INDEL) ancestry-informative marker (AIM) panel. Human Mutation 31 (2):184–190
10. Sambrook J, Fritsch EF, Maniatis T (1989) Isolations of DNA From mammalian cells. In: Ford N, Nolan C, Ferguson M (eds) Molecular cloning. Cold Spring Harbor Laboratory, New York, pp 916–919
11. Ribeiro-Rodrigues EM, Santos NPC, Ribeiro-dos-Santos AKC, Marinho AN, Zago MA, Gomes I, Amorim A, Gusmão L, Santos SEB (2009) An INDEL polymorphism at the X-STR GATA172D05 flanking region. Int J Legal Med 123:89–94
12. Excoffier L, Laval G, Schneider S (2005) Arlequin (version 3.0): an integrated software package for population genetics data analysis. Evol Bioinform Online 1:47–50
13. Desmarais D, Zhong Y, Chakraborty R, Perreault C, Busque L (1998) Development of a highly polymorphic STR marker for identity testing purposes at the human androgen receptor gene (HUMARA). J Forensic Sci 43:1046–1049
14. Zhu X, Zhang S, Kan D, Cooper R (2004) Haplotype block definition and its application. Pacific Symposium on Biocomputing 9:152–163
15. Bonferroni CE (1936) Teoria statistica delle classi e calcolo delle probabilit á. Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commerciali di Firenze 8:3–62
16. Katoh T, Mano S, Ikuta T, Munkhbat B, Tounai K, Ando H, Munkhtuvshin N, Imanishi T, Inoko H, Tamiya G (2002) Genetic isolates in East Asia: a study of linkage disequilibrium in the X chromosome. Am J Hum Genet 71:395–400
17. Schaffner SF (2004) The X chromosome in population genetics. Nat Rev Genet 5:43–51
18. Zhu X, Yan D, Cooper RS et al (2003) Linkage disequilibrium and haplotype diversity in the genes of the pressure program renin–angiotensin system: findings from the family blood. Genome Res 13:173–181
19. Becker D, Rodig H, Augustin C, Edelmann J, Gotz F, Hering S, Szibor R, Brabetz W (2008) Population genetic evaluation of eight X-chromosomal short tandem repeat loci using Mentype Argus X-8 PCR amplification kit. Forensic Sci Int 2:69–74
20. Gomes I, Pereira V, Gomes V, Prata MJ, Pinto N, Carracedo A, Amorim A, Gusmão L (2009) The Karimojong from Uganda: genetic characterization using na X-STR decaplex system. Forensic Sci Int 3:127–128