

**PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO GRANDE DO SUL
FACULDADE DE BIOCÊNCIAS
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOLOGIA CELULAR E MOLECULAR**

Cecília Helena Fricke Matte

Indels autossômicos e do cromossomo X em uma amostra da população do Rio Grande do Sul para possíveis aplicações forenses

Porto Alegre

2011

PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO GRANDE DO SUL
FACULDADE DE BIOCIÊNCIAS
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOLOGIA CELULAR E MOLECULAR

Indels autossômicos e do cromossomo X na população do Rio Grande
do Sul para possíveis aplicações forenses

Dissertação apresentada à Faculdade de Biociências da Pontifícia Universidade Católica do Rio Grande do Sul, como requisito para obtenção do título de Mestre em Biologia Celular e Molecular.

Autora

Cecília Helena Fricke Matte

Orientador

Dr. Sandro Luis Bonatto

Porto Alegre

2011

AGRADECIMENTOS

Ao IGP, por possibilitar a realização deste mestrado através do convênio com a PUCRS.

Ao professor Sandro Luis Bonatto, por aceitar me orientar neste trabalho, me auxiliando e também pelos ensinamentos trocados.

Ao meu grande professor, incentivador e orientador Sidney Santos, pelas idéias, auxílio e estímulo para o planejamento e desenvolvimento do projeto de pesquisa.

Aos colegas do Setor de Genética Forense do Instituto-Geral de Perícias pela amizade e companheirismo, e, claro, disponibilização das amostras.

Aos amigos do Laboratório de Genética Humana e Médica da Universidade Federal do Pará (especialmente Elze, Teka e Nei), por me receberem e por me auxiliarem na parte prática.

A Tatiane, pelo grande auxílio no final das análises.

A minha mãe, doutora Ruth Marilda Fricke, pelo apoio, incentivo e ensinamentos, fundamentais para que eu conseguisse concluir este trabalho.

Ao meu marido Daniel, pelo apoio, paciência e companheirismo...

Aos meus amados filhos Matheus e Vinícius, pelo tempo que deixei de dedicar a eles para a execução deste, e pelas madrugadas que me fizeram companhia enquanto eu escrevia....

RESUMO

A genética forense tem se desenvolvido muito nos últimos anos devido à sua reconhecida importância no auxílio para resolver os mais diversos casos criminais. Este reconhecimento levou a um aumento na demanda e, conseqüentemente, na necessidade de se obter resultados em amostras que anteriormente nem eram trabalhadas (devido ao alto grau de degradação) ou geravam laudos inconclusivos. O desenvolvimento de novas metodologias de análise e busca de novos marcadores genéticos para o uso forense é constante, e os marcadores bialélicos têm mostrado grande capacidade de individualização e identificação de uma amostra. Neste estudo foram analisadas as frequências de 47 marcadores autossômicos e 13 marcadores do cromossomo X de polimorfismos bialélicos do tipo inserção-deleção (Indel) em uma amostra de 90 indivíduos da população do Rio Grande do Sul, para avaliar sua aplicabilidade em casos forenses. Diversidade genética haplotípica (DH) e haplóide (h), e informações de índice (I) para marcadores do cromossomo X, e heterozigosidade observada (H_o), poder de discriminação (PD), poder de exclusão (PE), probabilidade de correspondência (MP), índice de paternidade (TPI) e conteúdo de informação de polimórfica (PIC) para os marcadores autossômicos foram calculados. Para os 47 indels autossômicos estudados, o poder combinado de discriminação e o poder combinado de exclusão foram de 99,999999999999998956% e 99,64%, respectivamente. Não foi observada uma subestruturação significativa na população estudada, e os componentes da mistura global estimados para a amostra foram 72,35%, 18,10% e 9,54% de DNA de europeus, nativo-americanos e africanos.

Palavras-chave

Polimorfismo de inserção deleção (Indel), genética de populações, genética forense, Rio Grande do Sul.

ABSTRACT

Forensic genetics has developed in recent years due to its recognized importance in helping to solve a wide variety of criminal cases. This recognition has led to an increase in demand and hence the need to achieve results on samples that previously were not worked (due to the high degree of degradation) or inconclusive reports generated. The development of new methods of analysis and search for new genetic markers for forensic use is constant, and the biallelic markers of have shown great capacity for individualization and identification of a sample. We studied the frequencies of 47 autosomal and 13 X-chromosome insertion-deletion polymorphism (InDel) markers in a sample of 90 individuals in the population of Rio Grande do Sul state, to evaluate its applicability in forensic cases. Haplotype (DH) and haploid genetic (h) diversity, and information index (I) for X-chromosome markers, and observed heterozygosity (Ho), power of discrimination (PD), power of exclusion (PE) matching probability (MP), typical paternity index (TPI) and polymorphism information content (PIC) for autosomal markers, were calculated. For the 47 autosomes indels studied, the combined power of discrimination and the combined power of exclusion were 99.999999999999998956% and 99.64%, respectively. We did not observe a significant substructure in the population studied, and the global estimated admixture components for the sample were 72.35%, 18.10%, and 9.54% for European, Native American, and African DNA.

Keywords

Insertion-deletion polymorphism (InDel), population genetics, forensic genetics, Rio Grande do Sul.

ÍNDICE

Introdução.....	7
Objetivos Gerais e Específicos.....	9
Artigo.....	10
Considerações inais.....	19
Referências Bibliográficas.....	21
Anexos.....	24

1. INTRODUÇÃO

A Genética Forense está sendo cada vez mais solicitada para a resolução de crimes de diversas naturezas e nos casos cíveis. A vasta maioria dos polimorfismos de DNA humano pode ser dividida em dois grupos: aqueles baseados em substituições nucleotídicas e aqueles baseados em inserção ou deleção de um ou mais nucleotídeos (*indels*). Os *indels* podem ser divididos naqueles com múltiplos alelos (multialélico) e aqueles com somente dois alelos (bialélico) (Weber *et al.*, 2002). Quase todos os *indels* multialélicos são baseados em *tandem repeats*, ou seja, pequenas repetições de DNA em sequência, a maioria STRs (*short tandem repeats*). Esses chamados “microsatélites” são atualmente os principais marcadores usados nos estudos em genética forense (Pena, 2005).

Atualmente vem sendo investigando mais profundamente outros marcadores genéticos que possam ser úteis em situações fora do padrão. Os SNPs (polimorfismos de um único nucleotídeo) e os *indels* bialélicos (que representam um terço de todos os *indels* identificados) se destacam por apresentarem a vantagem de se utilizarem de sequências muito curtas de DNA, facilitando sua utilização em amostras velhas ou muito degradadas – casos frequentes na área forense (Pereira *et al.*, 2009).

Os polimorfismos do tipo *indels* são abundantes no genoma humano. Estima-se que eles sejam responsáveis por aproximadamente 16-25% de todos os polimorfismos de sequência do genoma (Mills *et al.*, 2006). Apesar disto, eles têm recebido pouca atenção quando comparado aos SNPs, tanto nos estudos forenses como nos de variação na população (Mills *et al.*, 2006). Petrovski *et al.*, (2003) realizaram uma validação de 11 SNPs informativos para a identificação individual. Estudos comparativos realizados entre os dois tipos de marcadores concluíram que os SNPs servem como uma estratégia importante nas análises onde ocorrem problemas estatísticos com o uso somente dos STRs (Amorim e Pereira, 2005). Recentemente começaram a aparecer trabalhos específicos de *indels*, demonstrando sua utilidade como marcadores de ancestralidade e prevendo seu uso forense, cujos resultados mostram-se animadores (Bastos-Rodrigues *et al.*, 2006; Pereira *et al.*, 2009; Ribeiro-Rodrigues *et al.*, 2009; Freitas *et al.*, 2010; Santos *et al.*, 2010; Li *et al.*, 2011 a e b; Francez *et al.*, 2011 a e b). Além de comparar resultados em amostras forenses,

com baixa quantidade de DNA, e em marcadores de padrões de transmissão especial – como os marcadores de cromossomo X - ,os cientistas também estão investindo na padronização de conjuntos multiplex que facilitem a utilização na rotina da grande quantidade de marcadores exigida para resultados cada vez mais satisfatórios.

Algumas pesquisas atualmente têm verificado a utilidade dos *indels* nas mais diversas situações onde os demais polimorfismos não conseguiram sozinhos, chegar a um resultado conclusivo. Gibbon *et al.* (2009) utilizaram *indels* e 14 regiões polimórficas no gene da amelogenina para determinação de sexo de amostras de esqueletos arqueológicos, obtendo resultados muito satisfatórios.

A população brasileira é etnicamente complexa, tendo sido formada por sucessivas ondas migratórias, incluindo os ameríndios, africanos e europeus, sendo que a contribuição relativa de cada um destes grupos é muito heterogênea entre as diversas regiões do país. A população do Estado do Rio Grande do Sul é caracterizada por ter tido uma história de povoamento e imigração única, com uma relativamente baixa contribuição africana e indígena e um grande influxo de imigrantes europeus (alemães e italianos) no final do século 19 e início do século 20 (Salzano e Bortolini, 2002).

Antes de utilizar novos marcadores nas análises forenses, é importante coletar dados das populações e construir base de dados de referência, caracterizando, assim, a variação do genoma humano nas diferentes populações de interesse (Bastos-Rodrigues *et al.*, 2006; Rodrigues *et al.*, 2009) e permitindo avaliar o poder discriminatório dos marcadores entre as populações alvo. Bastos-Rodrigues *et al.* (2006), realizaram um estudo com 40 marcadores *indels*, escolhidos dentro da base de dados de Weber *et al.* (2002), que seguissem os critérios de estarem distribuídos em diferentes localizações cromossômicas, tamanho de amplicons que permitisse uma análise multiplex e que apresentassem frequência alélica na população européia próxima de 0,5. Com este conjunto de 40 *indels* conseguiram reproduzir com alto valor discriminatório a identificação dos cinco clusters populacionais principais (maiores regiões geográficas do mundo) identificados por Cann *et al.* (2002) com o uso de microssatélites autossômicos.

A investigação da frequência de marcadores de inserção-deleção em populações de diferentes origens, tem caráter preliminar para permitir a padronização da técnica de identificação humana através deste tipo de polimorfismos, sendo fundamental para uma posterior aplicação em casos forenses.

2. OBJETIVOS GERAIS E ESPECÍFICOS

Objetivo Geral:

Realizar estudo de marcadores de inserção-deleção autossômicos e do cromossomo X em amostra randomizada da população do Rio Grande do Sul, com o objetivo de avaliar sua aplicabilidade em casos forenses.

Objetivos Específicos:

- 1- Estimar a frequência alélica de 60 locos *indel* em uma amostra de 90 indivíduos do Rio Grande do Sul.
- 2- Utilizando as frequências dos marcadores *indels*, estimar a existência de estruturação genética na população sul-rio-grandense.
- 3- Estimar a contribuição genética relativa dos três grupos étnicos ascendentes na população sul-rio-grandense atual.

3. ARTIGO CIENTÍFICO

Os resultados do presente trabalho possibilitaram a elaboração do artigo "Autosomal and X-linked InDel polymorphisms for forensic use in the population of Rio Grande do Sul state, southern Brazil" apresentado a seguir, o qual foi submetido ao periódico:

FORENSIC SCIENCE INTERNATIONAL: GENETICS

ISSN: 1872-4973

Fator de Impacto 2009 (ISI Web of Knowledge - © Thomson Reuters Journal Citation Reports 2010): 2,421

Ms. Ref. No.: FSIGEN-D-12-00022

Forensic Science International: Genetics é especificamente dedicado à Genética Forense.

O artigo apresentado segue as normas de publicação de dados populacionais requeridas pela revista, de acordo com o editorial Publication of population data for forensic purposes (Forensic Sci. Int. Genet. 4 (2010) 145-147).

Short Communications

"Autosomal and X-linked InDel polymorphisms for forensic use in the population of Rio Grande do Sul state, southern Brazil"

Cecília Helena Fricke Matte^{1,2}, Solange Pereira Schwengber¹, Bianca de Almeida Carvalho¹, Trícia Cristine Kommers Albuquerque¹, Elzemar Martins Ribeiro-Rodrigues³, Fábio Pereira das Neves Leite¹, Sidney Emanuel Batista dos Santos³ e Sandro Luis Bonatto²

1- Instituto-Geral de Perícias do Rio Grande do Sul

2- Faculdade de Biociências, Pontifícia Universidade Católica do Rio Grande do Sul

3- Laboratório de Genética Humana e Médica, Instituto de Ciências Biológicas, Universidade Federal do Pará

*Corresponding author: Cecília Helena Fricke Matte

E-mail address: cecilia-matte@igp.rs.gov.br

Phone: 55 51 32336477

Adress: Av. Azenha, 255 – térreo – Setor de Genética Forense. Porto Alegre/RS Brazil
90160-000

Abstract

The Brazilian population was formed by successive waves of migration (European and African) over Native American autochthonous population. This led to formation of a multi-ethnic population mixed with three main ethnic groups. However, the percentage contribution of each of these parental populations is very different among the different regions of the country. The population of Rio Grande do Sul state is peculiar in which the African and Native American relative contribution is low in most areas given the large

number of immigrants from Germany and later Italy during the 19th century and the beginning of the 20th century. 47 autosomal and 13 X-chromosome insertion-deletion (InDel) loci were studied in a sample of 90 individuals in the population of Rio Grande do Sul state, to evaluate their applicability in forensic cases. The combined power of discrimination and the combined power of exclusion were 99.99999999999998956% and 99.64%, respectively for the autosomal indels. The average information index across the X-linked loci was relatively high ($I = 0.507$). We did not observe a significant substructure in the population studied, and the global estimated admixture components for the sample were 72.35%, 18.10%, and 9.54% for European, Native American, and African DNA.

Keywords

Insertion-deletion polymorphism (InDel), population genetics, forensic genetics, Rio Grande do Sul.

Population: The Brazilian population was formed by successive waves of migration (European and African) over Native American autochthonous population. This led to formation of a multi-ethnic population mixed with three main ethnic groups. However, the percentage contribution of each of these parental populations is not uniform among the different regions of the country. The population of Rio Grande do Sul state is peculiar in which the African and Native American relative contribution is low in most areas given the large number of immigrants from Germany and later Italy during the 19th century and the beginning of the 20th century [1]. A total of 90 unrelated males living in the Rio Grande do Sul state, southern Brazil, were randomly selected from individuals seen for routine examinations at the Forensic Genetics Laboratory – Instituto-Geral de Perícias do Rio Grande do Sul (IGP-RS). All the samples were collected after signing of an informed consent.

DNA extraction: The DNA from peripheral blood samples were extracted by the salting-out method [2] and from oral mucosa cells with a NaOH DNA extraction method [3].

Selected Markers and PCR Amplification Conditions: A set of 13 X-linked and 47 autosomal InDel markers were used, which were tested and optimized [4, 5] as ancestry-informative-markers (AIMs) (markers whose allelic frequencies vary significantly among populations from different ethnic origins, in this case, specifically European, African and Native American). The PCR primer sequences and DNA amplification conditions used were described by Santos et al. [5].

Genotyping: Genotyping of the fragments was performed using the genetic sequencers ABI PRISM[®] 3100-Avant Genetic Analyzer and ABI PRISM[®] 3130-Avant Genetic Analyzer (Applied Biosystems), with the assistance of the Data Collection v2.0 and the GeneMapperID v3.2 softwares (Applied Biosystems), using the GS LIZ-500 reference ladder (Applied Biosystems) as size standard.

Quality control: Laboratory internal control standards and kit control were employed.

Statistical Analysis: Basic statistics such as allele and haplotype frequencies, expected (H_e) and observed heterozygosity (H_o), haplotype (DH) and haploid genetic diversity (h), Hardy–Weinberg equilibrium (HWE) test, and information index (I) were estimated using the GenAlEx software [6]. Polymorphism information content (PIC), power of discrimination (PD), power of exclusion (PE), matching probability (MP) and typical paternity index (TPI) were calculated using the Powerstats V12 software [7]. The existence of substructure in this population (the number of clusters K , from 1 to 5) was tested using Structure v2.2 software [8] with admixture model, and a burn-in period of 200,000 interactions followed by 500,000 additional interactions.

Admixture components were estimated by the gene identity method [9] using Admix95 (available at www.genetica.fmed.edu.uy/software.htm). For this analysis, we used the genotype frequencies of the 47 InDel autosome markers for Africans, Europeans, and Native Americans presented by [5] as the parental populations.

Results: Allele frequencies and statistical parameters for the 60 InDel loci of the population of Rio Grande do Sul state are shown in Tables 1 and 2.

Other remarks: 46 of the 47 autosomal loci analyzed did not differ significantly from the HWE in the study population, considering $p > 0.05$, after Bonferroni correction to eliminate type I errors. The exception was marker MID1098 that nonetheless was highly informative, while marker MID913 was the less informative in this population (Table 1). The averaged parameter values across the autosomal markers were: PIC = 0.319; PD = 0.554; PE = 0.111; Ho = 0.375; I = 0.590. For the 47 InDels studied, the combined power of discrimination and the combined power of exclusion were 99.999999999999998956% and 99.64%, respectively, that support the usefulness of the joint use of these markers in routine forensic in this population.

Of the 13 X-linked loci analyzed, 10 were very informative, the less informative being MID2657 (Table 2). The average information index across these 13 markers was relatively high (I = 0.507). The average haploid genetic diversity (h) observed was 0.338. Sixty three different haplotypes were found in the 90 samples studied, demonstrating the great power of discrimination of this set of X-linked InDel markers (whole sample haplotype diversity = 99.62%).

The analyses of population genetic structure by the program Structure did not find evidence for substructure (division into different clusters) in this population, both when the autosomal and X-linked markers were analyzed separately or together. The average percentage contribution of each of the three main ethnic groups in the parental population of Rio Grande do Sul, as estimated by Admix95 using the 47 autosomal markers, were 72.35%, 18.10%, and 9.54% for European, Native American, and African DNA, respectively. These results differ from a previous publication [5, 10], in which the European contribution was 91.3% and the African was 0%, much likely because that study used samples of individuals of known prevalent European ancestry, whereas here our Rio Grande do Sul sample was random in relation to the ethnic ancestry of the individuals.

The results of this study were corroborated with other studies, whose publications have shown similar results with respect to the power of discrimination obtained, even in populations with very different backgrounds, observing the proportion of markers used (11).

This publication followed the guidelines for population data requested by the journal [12].

Conflict of interest: None.

Acknowledgements

This study was supported by Instituto-Geral de Perícias do Rio Grande do Sul (IGP-RS), Secretaria Nacional de Segurança Pública (SENASP) and Laboratório de Genética Humana e Médica (LGHM-UFPA).

References

- [1] Salzano FM, Bortolini MC (2002) The evolution and genetics of Latin American populations. Cambridge University Press, Cambridge.
- [2] Miller SA, Dykes DD & Polesky HF (1998) A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Research* 6:1215.
- [3] Richards B, Skoletsky J, Balfour R, Stern RC, Dorkin HL, Witt D & Klinger KW (1993) Multiplex PCR amplification from the CFTR gene using DNA prepared from buccal brushes/swabs. *Hum Mol Genet* 2(2):159-63.
- [4] Ribeiro-Rodrigues EM, Santos NP, Santos AK, Pereira R, Amorim A, Gusmão L, Zago MA & Santos SE (2009) Assessing interethnic admixture using an X-linked insertion-deletion multiplex. *Am J Hum Biol*, 21:707-709.
- [5] Santos NPC, Ribeiro-Rodrigues EM, Ribeiro-dos-Santos AKC, Pereira R, Gusmão L, Amorim A, Guerreiro JF, Zago MA, Matte C, Hutz MH & Santos SEB (2010) Assessing individual interethnic admixture and population substructure using a 48 insertion-deletion ancestry-informative marker panel. *Hum Mutat*. 31:184-190.
- [6] Peakall, R. and Smouse P.E. (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes*. 6, 288-295. Available in <http://www.anu.edu.au/BoZo/GenAIEx/>.

- [7] A. Tereba, Tools for Analysis of Population Statistics, Profiles in DNA, vol. 2, no. 3, Home page at <http://www.promega.com/geneticidtools>
- [8] Pritchard JK, Stephens M & Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 1239:33-36. Available in http://pritch.bsd.uchicago.edu/software/structure2_1.html
- [9] Chakraborty R. (1985) Gene identity in racial hybrids and estimation of admixture rates. In: Neel JV, Ahuja Y, editors. *Genetic Micro-differentiation in Man and other Animals*. New Delhi: Indian Anthropological Association 171-180.
- [10] Francez PAC, Ribeiro-Rodrigues EM & Santos SEB (2011) Allelic frequencies and statistical data obtained from 48 AIM INDEL loci in an admixed population from the Brazilian Amazon. *Forensic Sci. Int. Genet.*, doi:10.1016/j.fsigen.2011.04.002
- [11] Li C, Zhao S, Zhang S, Li L, Liu Y, Chen J & Xue J (2011) Genetic polymorphism of 29 highly informative InDel markers for forensic use in the Chinese Han population. *Forensic Science International: Genetics* 5:e27-e30
- [12] Publication of population data for forensic purposes. Editorial / *Forensic Science International: Genetics* 4 (2010) 145–147.

Table 1 - Allele frequencies distribution of the 47 autosomal InDel loci investigated.

Locus	Deletion (short allele)	N (Total Alleles)	Short Allele Frequency	Ho	He	MP	PD	PE	PIC	I	TPI
MID1172	113	154	0.182	0.234	0.301	0.551	0.449	0.04	0.25	0.478	0.65
MID1176	149	158	0.367	0.405	0.467	0.376	0.624	0.117	0.36	0.660	0.84
MID1271	144	158	0.734	0.354	0.393	0.444	0.556	0.089	0.31	0.582	0.77
MID1357	95	158	0.709	0.329	0.416	0.421	0.579	0.076	0.33	0.606	0.75
MID1358	235	134	0.799	0.313	0.325	0.512	0.488	0.069	0.27	0.507	0.73
MID1379	150	156	0.910	0.179	0.165	0.705	0.295	0.024	0.15	0.305	0.61
MID1684	157	158	0.386	0.468	0.476	0.387	0.613	0.161	0.36	0.669	0.94
MID1726	243	158	0.563	0.468	0.492	0.369	0.631	0.161	0.37	0.685	0.94
MID1785	95	158	0.165	0.228	0.269	0.575	0.425	0.038	0.24	0.440	0.65
MID2011	192	158	0.690	0.443	0.431	0.423	0.577	0.142	0.34	0.622	0.9
MID273	137	158	0.696	0.532	0.421	0.469	0.531	0.217	0.33	0.612	1.07
MID494	227	158	0.570	0.481	0.490	0.376	0.624	0.171	0.37	0.683	0.96
MID625	115	158	0.677	0.316	0.431	0.397	0.603	0.071	0.34	0.622	0.73
MID789	161	158	0.842	0.190	0.269	0.598	0.402	0.027	0.23	0.440	0.62
MID818	222	158	0.241	0.380	0.369	0.471	0.529	0.102	0.3	0.555	0.81
MID473	83	168	0.435	0.536	0.491	0.403	0.597	0.221	0.37	0.685	1.08
MID619	134	170	0.382	0.506	0.475	0.406	0.594	0.193	0.36	0.668	1.01
MID1448	162	168	0.321	0.357	0.436	0.398	0.602	0.09	0.34	0.628	0.78
MID1923	198	170	0.182	0.271	0.298	0.541	0.459	0.052	0.25	0.475	0.69
MID856	256	170	0.694	0.400	0.425	0.415	0.585	0.114	0.33	0.616	0.83
MID99	132	170	0.582	0.482	0.486	0.38	0.62	0.173	0.37	0.680	0.97
MID93	146	168	0.571	0.357	0.490	0.344	0.656	0.09	0.37	0.683	0.78
MID682	182	170	0.824	0.235	0.291	0.557	0.443	0.04	0.25	0.466	0.65
MID1039	211	168	0.530	0.512	0.498	0.383	0.617	0.198	0.37	0.691	1.02
MID1780	104	166	0.620	0.470	0.471	0.39	0.61	0.162	0.36	0.664	0.94
MID1470	133	166	0.416	0.494	0.486	0.386	0.614	0.182	0.37	0.679	0.99
MID132	190	168	0.500	0.476	0.500	0.364	0.636	0.168	0.38	0.693	0.95
MID1098	340	134	0.522	0.299	0.499	0.336	0.664	0.063	0.37	0.692	0.71
MID1558	145	170	0.524	0.365	0.499	0.336	0.664	0.094	0.37	0.692	0.79
MID217	171	168	0.470	0.440	0.498	0.352	0.648	0.141	0.37	0.691	0.89
MID568	201	168	0.250	0.405	0.375	0.466	0.534	0.117	0.3	0.562	0.84
MID1952	96	160	0.419	0.463	0.486	0.372	0.628	0.157	0.37	0.679	0.93
MID476	167	180	0.339	0.411	0.447	0.394	0.606	0.121	0.35	0.639	0.85
MID275	210	180	0.661	0.411	0.451	0.394	0.606	0.121	0.35	0.643	0.85
MID20.1	255	178	0.348	0.494	0.446	0.418	0.582	0.183	0.35	0.638	0.99
MID196	112	160	0.350	0.425	0.459	0.391	0.609	0.13	0.35	0.652	0.87
MID1716	173	160	0.656	0.363	0.449	0.383	0.617	0.093	0.35	0.641	0.78
MID778	197	180	0.717	0.433	0.407	0.442	0.558	0.136	0.32	0.598	0.88
MID1386	327	178	0.258	0.382	0.378	0.454	0.546	0.103	0.31	0.566	0.81
MID10.3	109	180	0.617	0.456	0.470	0.383	0.617	0.151	0.36	0.663	0.92
MID575	175	180	0.161	0.233	0.273	0.578	0.422	0.04	0.23	0.444	0.65
MID216	226	174	0.741	0.333	0.392	0.45	0.55	0.078	0.31	0.581	0.75
MID481	283	178	0.225	0.292	0.360	0.487	0.513	0.06	0.29	0.546	0.71
MID913	331	178	0.028	0.056	0.057	0.894	0.106	0.003	0.05	0.133	0.53
MID350	102	180	0.106	0.189	0.187	0.676	0.324	0.027	0.17	0.335	0.62
MID988	149	178	0.404	0.360	0.482	0.353	0.647	0.091	0.37	0.675	0.78
MID184	243	172	0.442	0.419	0.494	0.351	0.649	0.126	0.37	0.687	0.86

I=information index; Ho=observed heterozygosity; He=expected heterozygosity; MP=matching probability; PD=power of discrimination; PE=power of exclusion; PIC= polymorphism information content; TPI=typical paternity index.

Table 2 - Allele frequencies distribution of the 13 X-linked InDel loci investigated.

InDel identification	Deletion (short allele)	N (Total Alleles)	Short Allele Frequency	I	h	uh
MID1445	112	70	0.786	0.520	0.337	0.342
MID184	130	67	0.448	0.688	0.495	0.502
MID357	154	70	0.386	0.667	0.474	0.481
MID1326	192	72	0.417	0.679	0.486	0.493
MID193	115	72	0.375	0.662	0.469	0.475
MID2610	139	64	0.078	0.274	0.144	0.146
MID2652	167	71	0.803	0.496	0.317	0.321
MID2657	196	71	0.028	0.128	0.055	0.056
MID1540	231	70	0.329	0.633	0.441	0.448
MID2694	114	72	0.139	0.403	0.239	0.243
MID2600	160	71	0.070	0.255	0.131	0.133
MID358	174	71	0.775	0.534	0.349	0.354
MID1705	223	65	0.354	0.650	0.457	0.464

I=information index; h=haploid genetic diversity; uh=unbiased haploid genetic diversity.

4. CONSIDERAÇÕES FINAIS

Os resultados obtidos para os parâmetros forenses usuais demonstraram que, tanto os marcadores indels autossômicos quanto os ligados ao X, apresentam características que possibilitam seu uso na rotina de análises da genética forense. Devido a sua natureza bialélica, um polimorfismo bialélico possui um menor poder discriminatório individual se comparado com um STR multialélico. Nesse sentido, Krawczak (1999) calculou que são necessários de 4 a 8 BPs para se ter uma equivalência ao poder de discriminação de um único locus STR. Segundo Kidd *et al.* (2006), é necessário entre 40 e 50 BPs para se alcançar o mesmo poder de discriminação individual que é fornecido pelo conjunto de STRs já em uso pela comunidade científica, compartilhando os dados obtidos neste trabalho (vide anexos). Ainda não há um levantamento matemático que comprove com exatidão o número necessário de BPs para se obter a probabilidade similar à obtida pelos perfis resultantes dos kits comerciais que utilizam marcadores STRs. No entanto, os trabalhos que vêm sendo desenvolvidos (Pereira *et al.*, 2009; Freitas *et al.*, 2010, Li *et al.*, 2011 b e Pereira *et al.* 2011) mostram que o uso de um sistema multiplex com vários marcadores de inserção-deleção tende a fornecer resultados estatisticamente mais significativos que os obtidos com os marcadores STRs atualmente utilizados no âmbito de identificação humana.

Este projeto de pesquisa foi desenvolvido com a finalidade de obtenção de título de mestre e utilizando a parceria firmada entre as Instituições PUCRS e IGP, através do Convênio Nº 043/04, o qual visa à conjugação de esforços entre os partícipes para cooperação, intercâmbio técnico-científico e desenvolvimento de recursos humanos, não existindo conflito de interesses que envolva aspectos econômicos, científicos, pessoais ou qualquer outro entre as partes. O presente estudo tem o principal objetivo de compilar e analisar estatisticamente os dados de frequências genotípicas gerados para um estudo populacional de marcadores neutros. O IGP e/ou a PUCRS não têm nenhum interesse financeiro nos resultados específicos deste estudo. Os resultados deste estudo são apenas e exclusivamente para se entender melhor a estruturação genética da população do Rio

Grande do Sul e avaliar se estes novos marcadores poderão ter aplicação forense localmente.

Este estudo foi aprovado pelo IGP para ser desenvolvido através do convênio IGP/RS-PUCRS e foi aprovado pelo Comitê de Ética da UFPA, onde as genotipagens foram realizadas. Posteriormente, o Projeto de Pesquisa também passou por aprovação do Comitê de Ética da PUCRS.

5. REFERÊNCIAS BIBLIOGRÁFICAS

- Amorim A & Pereira L (2005) Pros and cons in the use of SNPs in forensic kinship investigation: a comparative analysis with STRs. *Forensic Science International* 150:17–21.
- Bastos-Rodrigues L, Pimenta JR & Pena SDJ (2006) The genetic structure of human populations studied through short insertion-deletion polymorphisms. *Annals of Hum Genet*, 70: 658-665.
- Francez PAC, Ribeiro-Rodrigues EM & Santos SEB (2011) Allelic frequencies and statistical data obtained from 48 AIM INDEL loci in an admixed population from the Brazilian Amazon. *Forensic Sci. Int. Genet.*, doi:10.1016/j.fsigen.2011.04.002
- Francez PAC, Ribeiro-Rodrigues EM, Velasco AF & Santos SEB (2011) Insertion-deletion polymorphisms. *Int J Legal Med.*, doi:10.1007/s00414-011-0588-z
- Freitas NSC, Resque RL, Ribeiro-Rodrigues EM, Guerreiro JF, Santos NPC, Ribeiro-dos-Santos A & Santos SEB (2010) X-linked insertion/deletion polymorphisms: forensic applications of a 33-markers panel. *Int J Legal Med* 124:589-593.
- Gibbon V, Paximadis M, Štrkalj G, Ruff P & Penny C (2009) Novel methods of molecular sex identification from skeletal tissue using the amelogenin gene. *Forensic Science International: Genetics* 3 (2009) 74–79.
- Kidd KK., Pakstis AJ, Speed WC, Grigorenko EL, Kajuna SL, Karoma NJ, Kungulilo S, Kim JJ, Lu RB, Odunsi A, Okonofua F, Parnas J, Schulz LO, Zhukova OV & Kidd JR. J (2006) Developing a SNP panel for forensic identification of individuals. *Forensic Sci. Int.* 2006; 164: 20-32. *Forensic Sci. Int.* 164:20-32.
- Krawczak M (1999) Informativity assessment for biallelic single nucleotide polymorphisms. *Electrophoresis* 1999; 20: 1676–1681
- Laval LG & Schneider S (2005) Arlequin ver. 3.0: An integrated software package for population genetic data analysis. *Evolutionary Bioinformatics Online*, 1:47-50.
- Li C, Zhao S, Zhang S, Li L, Liu Y, Chen J & Xue J (2011) Genetic polymorphism of 29 highly informative InDel markers for forensic use in the Chinese Han population. *Forensic Science International: Genetics* 5:e27-e30.
- Li C, Zhang S, Li L, Chen J, Liu Y & Zhao S (2011) Selection of 29 highly informative InDel markers for human identification and paternity analysis in Chinese Han population by the SNPlex genotyping system. *Mol Biol Rep*.
- Miller SA, Dykes DD & Polesky HF (1998) A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Research* 6:1215.

- 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.
- Peakall, R. and Smouse P.E. (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes*. 6, 288-295.
- Pena SDJ (2005) Segurança pública: determinação de identidade genética pelo DNA. *Parcerias Estratégicas*, 20:447-460.
- Pereira R, Pereira V, Gomes I, Tomas C, Morling N, Amorin A, Prata MJ, Carracedo A & Gusmão L (2011) A method for the analysis of 32 X chromosome insertion deletion polymorphisms in a single PCR. *Int J Legal Med.*
- Pereira R, Phillips C, Alves C, Amorin A, Carracedo A & Gusmão L (2009) A new multiplex for human identification using insertion/deletion polymorphisms. *Electrophoresis*, 30:3682-3690.
- Petkovski E, Keyser C, Ludes B & Hienne R (2003) Validation of SNPs as markers for individual identification. *International Congress Series 1239*: 33–36.
- Pritchard JK, Stephens M & Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 1239:33-36.
- Richards B, Skoletsky J, Balfour R, Stern RC, Dorkin HL, Witt D, Klinger KW (1993) Multiplex PCR amplification from the CFTR gene using DNA prepared from buccal brushes/swabs. *Hum Mol Genet* 2(2):159-63.
- Rodrigues EMR, Santos NPC, Santos AKCR, Marinho AN, Zago MA, Gomes I, Amorim A, Gusmão L & Santos EB (2009) An INDEL polymorphism at the X-STR GATA172D05 flanking region. *Int J Legal Med*, 123:89-94.
- Ribeiro-Rodrigues EM, Santos NP, Santos AK, Pereira R, Amorim A, Gusmão L, Zago MA & Santos SE (2009) Assessing interethnic admixture using an X-linked indsertinn-deletion multiplex. *Am J Hum Biol*, 21:707-709.
- Salzano FM, Bortolini MC (2002) *The evolution and genetics of Latin American populations*. Cambridge University Press, Cambridge.
- Santos NPC, Ribeiro-Rodrigues EM, Ribeiro-dos-Santos AKC, Pereira R, Gusmão L, Amorim A, Guerreiro JF, Zago MA, Matte C, Hutz MH & Santos SEB (*in press*) Assessing individual interethnic admixture and population substructure using a 48 insertion-deletion ancestry-informative marker panel. *Hum Mut.*
- Tereba A, Tools for Analysis of Population Statistics, Profiles in DNA, vol. 2, no. 3, Home page at <http://www.promega.com/geneticidtools>

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.

6. ANEXOS

Tabela 1 – Informações sobre o conjunto de *Indels* Europeu, incluindo: cromossomo no qual está localizado, localização física do *indel* no cromossomo, comprimento do *indel* e tamanho esperado do amplicon para cada alelo.

Marcador	Cromossomo	Localização cromossômica	Comprimento <i>Indel</i> (pb)	Tamanho Amplicon (pb)
MID473	6	95.3	8	83/91
MID1923	6	12.1	3	134/137
MID1448	3	192.6	4	162/166
MID619	7	27.9	4	198/202
MID856	5	65.3	26	256/282
MID1780	11	35.1	3	104/107
MID1470	11	64.5	5	133/138
MID132	15	22.9	4	190/194
MID1098	10	78.8	10	340/350
MID1558	3	132.1	4	145/149
MID217	7	145.5	3	171/174
MID568	13	92.7	3	201/204
MID99	22	24.1	6	132/138
MID93	22	40.3	3	146/149
MID682	9	94.6	4	173/177
MID1716	6	83.9	3	182/185
MID1039	5	34	40	211/251

Tabela 2 – Informações sobre o conjunto de *Indels* Africano, incluindo: cromossomo no qual está localizado, localização física do *indel* no cromossomo, comprimento do *indel* e tamanho esperado do amplicon para cada alelo.

Marcador	Cromossomo	Localização cromossômica	Comprimento <i>indel</i> (pb)	Tamanho amplicon (pb)
MID1357	14	79.2	4	95/99
MID273	17	5.3	4	137/141
MID1684	14	37.9	6	157/163
MID818	16	2.1	24	222/246
MID1172	16	2.13	5	113/118
MID1176	5	55.6	3	149/152
MID1358	5	35.6	17	235/252
MID1785	3	12	3	95/98
MID1271	5	109.8	5	144/149
MID780	14	78.5	4	161/165
MID494	6	6.6	10	227/237
MID625	2	31.3	3	115/118
MID1379	19	54.4	3	150/153
MID2011	2	108.8	5	192/197
MID1726	1	39.6	13	243/256

Tabela 3 – Informações sobre o conjunto de *Indels* Ameríndio, incluindo: cromossomo no qual está localizado, localização física do *indel* no cromossomo, comprimento do *indel* e tamanho esperado do amplicon para cada alelo.

Marcador	Cromossomo	Localização cromossômica	Comprimento <i>indel</i> (pb)	Tamanho amplicon (pb)
MID1952	15	60.95	4	96/100
MID476	9	19.70	3	167/170
MID275	17	66.39	5	210/215
MID152	20	2.07	9	255/264
MID196	6	99.90	3	112/115
MID778	16	82.40	3	197/200
MID1603	5	70.86	4	234/238
MID1386	1	244.10	21	327/348
MID660	10	17.40	3	109/112
MID575	1	36.06	3	175/178
MID216	2	191.38	4	226/230
MID481	2	63.80	7	283/290
MID913	6	38.40	14	331/345
MID350	2	211.2	3	102/105
MID988	5	113.8	4	149/153
MID1184	16	82.3	10	243/253

Tabela 4 – Informações sobre o conjunto de *Indels* do cromossomo X, incluindo: localização física do *indel* no cromossomo, comprimento do *indel* e tamanho esperado do amplicon para cada alelo.

Marcador	Localização cromossômica	Comprimento <i>indel</i> (pb)	Tamanho amplicon (pb)
MID357	12.822	5	154/159
MID1445	41.622	4	112/116
MID2694	85.400	5	114/119
MID2657	97.307	4	196/200
MID1326	103.116	3	192/195
MID2600	110.073	6	160/166
MID2610	110.397	3	139/142
MID184	116.469	3	131/134
MID1705	127.269	4	223/227
MID193	141.100	4	114/118
MID358	147.606	7	174/181
MID1540	149.838	20	231/251
MID2652	154.215	3	167/170