

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

DISSERTAÇÃO DE MESTRADO

ANDRÉ ZORATTO GASTALDO

**FREQUÊNCIA ALÉLICA DE SETE MARCADORES *SHORT TANDEM REPEAT* EM
INDIVÍDUOS DO ESTADO DO RIO GRANDE DO SUL, BRASIL**

PORTO ALEGRE

2012

ANDRÉ ZORATTO GASTALDO

**FREQUÊNCIA ALÉLICA DE SETE MARCADORES *SHORT TANDEM REPEAT* EM
INDIVÍDUOS DO ESTADO DO RIO GRANDE DO SUL, BRASIL**

**Dissertação apresentada como requisito para a
obtenção do grau de Mestre pelo Programa de
Pós-Graduação em Biologia Celular e Molecular
da Pontifícia Universidade Católica do Rio
Grande do Sul**

ORIENTADORA: CLARICE SAMPAIO ALHO

**PORTO ALEGRE
2012**

RESUMO

Dados populacionais de sete *loci Short Tandem Repeat* do kit comercial PowerPlex® CS7 foram obtidos a partir de uma amostra de 401 indivíduos do Rio Grande do Sul (sul do Brasil). Os *loci* são os dois pentanucleotídeos STRs do PowerPlex® 16 (PENTA D e PENTA E), um pentanucleotídeo extra (PENTA C), além dos quatro *loci* STR do kit comercial *FFFL Fluorescent STR System* (LPL, F13B, FES/FPS e F13A01). As frequências alélicas e outros dados estatísticos de relevância forense foram analisados. Todos os *loci* atenderam às expectativas de Hardy-Weinberg e não foram encontrados desequilíbrios de ligação, exceto para o *locus* LPL. A heterozigosidade observada e esperada, o poder de discriminação e exclusão e o conteúdo de informação polimórfica foram calculados. O poder combinado de discriminação e o poder combinado de exclusão foram de 0,999999987 e 0,998159054, respectivamente. As distâncias genéticas entre a população do Rio Grande do Sul e outras populações foram apresentadas neste trabalho.

Palavras-chave: *Loci Short Tandem Repeat*. Frequências alélicas. Dados populacionais. Distâncias genéticas. Rio Grande do Sul.

ABSTRACT

Population data of seven short tandem repeat *loci* of the PowerPlex® CS7 were obtained from a sample of 401 individuals from Rio Grande do Sul (Southern Brazil). The *loci* are the two pentanucleotides STRs in the PowerPlex® 16 (PENTA D and PENTA E), an extra pentanucleotide (PENTA C), and four *loci* STR in the *FFFL Fluorescent STR System* (LPL, F13B, FES/FPS, and F13A01). Allele frequency and other forensically relevant statistics data were generated for these seven *loci*. All of the analyzed *loci* meet Hardy-Weinberg equilibrium expectations and no linkage disequilibrium were found, except for LPL *locus*. The observed and expected heterozygosity, power of discrimination and exclusion and polymorphic information content were calculated. The combined power of discrimination and combined power of exclusion were 0.999999987 and 0.998159054, respectively. Genetic distances between population from Rio Grande do Sul and other populations are presented.

Keywords: Short tandem repeat *loci*. Allele frequencies. Population data. Genetic distances. Southern Brazil.

SUMÁRIO

CAPÍTULO 1 - APRESENTAÇÃO DO TEMA	4
INTRODUÇÃO	4
1 Histórico da Identificação Humana antes do uso do DNA.....	4
2 A análise do DNA para Identificação Humana e Testes de Paternidade.....	5
3 Marcadores Genéticos.....	5
3.1 VNTRs.....	6
3.2 STRs.....	6
3.3 Marcadores STRs LPL, F13B, FES/FPS, F13A01, PENTA C, PENTA D e PENTA E..	8
3.3.1 LPL.....	8
3.3.2 F13B.....	8
3.3.3 FES/FPS.....	8
3.3.4 F13A01.....	9
3.3.5 PENTA C, PENTA D e PENTA E.....	9
3.4 Estudos populacionais e aplicabilidade na genética forense.....	10
OBJETIVOS	13
REFERÊNCIAS BIBLIOGRÁFICAS	13
CAPÍTULO 2 – ARTIGO CIENTÍFICO	16
CAPÍTULO 3 – CONSIDERAÇÕES FINAIS	27

CAPÍTULO 1 - APRESENTAÇÃO DO TEMA

INTRODUÇÃO

1 HISTÓRICO DA IDENTIFICAÇÃO HUMANA ANTES DO USO DO DNA

Ao longo da história, o homem tem buscado, de forma incessante, meios para o aperfeiçoamento na determinação correta de seus vínculos genéticos, bem como ferramentas que o auxiliem a ser diferenciado de seus semelhantes, através da caracterização individual. O problema na determinação da paternidade é um assunto que gera discussões há milhares de anos. A concepção ocorre no interior da mulher, e dessa forma não há provas de qual homem é o espermatozoide que conseguiu penetrar no óvulo, gerando assim um novo ser.

As primeiras formas de estabelecer a paternidade, ou vínculo genético, eram realizadas através da comparação das características fisionômicas e psicológicas entre o filho e o suposto pai, sendo esta caracterização muito subjetiva. A hereditariedade era sinônimo de semelhança e, desse modo, muitas incertezas ainda permaneciam, pois o estabelecimento de vínculo genético ainda era muito superficial.

A partir do início do século XX, com a descrição dos grupos sanguíneos do sistema ABO, a determinação de laços familiares através da herança genética, começou a ser empregada com maior precisão. Com o avanço da ciência, os processos de identificação humana evoluíram de forma significativa, principalmente pela descoberta de outros sistemas sanguíneos.

Os primeiros métodos para avaliação em teste de paternidade, que conseguiam excluir acima de 90% dos supostos pais testados, foram reconhecidos cientificamente em meados da década de 1970, mediante o emprego em conjunto do sistema de antígenos leucocitários humanos (HLA) com outros sistemas sanguíneos. A associação desses métodos atinge um poder exclusão de paternidade de pelo menos 95%, porém longe de ser o ideal para esse fim (SILVER, 1989; WALKER *et al.*, 1987). Esse último tipo de teste de paternidade baseia-se na análise de proteínas presentes em hemácias e leucócitos, as quais são produtos gênicos. Os genes codificantes de tais proteínas sanguíneas, muitas vezes, sofrem mutações que diminuem, alteram ou impedem a produção das mesmas, o que pode ocasionar uma falsa exclusão de paternidade.

2 A ANÁLISE DO DNA PARA IDENTIFICAÇÃO HUMANA E TESTES DE PATERNIDADE

O aumento na confiabilidade nos resultados das investigações de paternidade e a solução para os muitos problemas em vigor surgiram a partir da análise do DNA de cada indivíduo. Ao contrário das análises convencionais de grupos sanguíneos e de sistemas proteicos (que dependem da disponibilidade de sangue ou fluidos corporais relacionados), os resultados das análises de DNA não dependem da natureza do material ou da célula analisada, pois a informação genética está contida integralmente em todas as células somáticas de um indivíduo. Outra vantagem da análise do DNA é que este é fisicamente muito mais resistente à degradação do que as proteínas (SCHNEIDER, 1997).

A investigação de vínculo genético é realizada através da comparação entre locais do DNA (a denominação de local/locais, em genômica utiliza o termo em latim *locus/loci*, ou apenas *lócus*, em português - no presente texto usar-se-á o termo em latim), que variam entre as pessoas da população. As diferentes formas que o DNA pode apresentar para um mesmo *locus* são chamadas alelos, e cada indivíduo possui dois alelos para cada *locus*: um de origem materna e o outro de origem paterna. Ao realizar uma investigação de paternidade, diversos *loci* são analisados, e é verificado se houve o compartilhamento de alelos entre o filho e o suposto pai investigados. É possível confirmar a paternidade com uma probabilidade igual ou maior a 99,99% caso ocorra o compartilhamento de alelos entre filho e suposto pai em todos os *loci* analisados. Da mesma forma, quando os alelos não são compartilhados entre o filho e o suposto pai, este é 100% excluído da possibilidade de ser o pai biológico, ou seja, a probabilidade de paternidade é zero.

3 MARCADORES GENÉTICOS

A identificação humana através da estrutura do DNA é a metodologia mais moderna aplicada mundialmente e está baseada em diferenças na sequência de DNA dos *loci* entre uma pessoa e outra; tais diferenças genéticas são conhecidas como polimorfismos.

Os polimorfismos genéticos correspondem a formas alternativas de sequências de nucleotídeos, as quais, quando analisadas no nível de DNA, diferem entre si por alterações no comprimento do segmento de DNA ou alterações do tipo troca de base. Dentro do primeiro grupo, identifica-se os polimorfismos originados pela alteração do número de sequências de repetições consecutivas (*in tandem*), que pode ser dividido em duas classes: as Repetições Sucessivas de Número Variável (*Variable Number Tandem Repeats* – VNTRs) e as Repetições Sucessivas Pequenas (*Short Tandem Repeats* – STRs) (JEFFREYS *et al.*, 1985a). Já no segundo grupo, pode-se destacar uma classe de marcadores polimórficos conhecidos como *Single Nucleotide Polimorphism* (SNPs), amplamente distribuídos pelo genoma humano (BUTLER, 2005).

3.1 VNTRs

Os marcadores polimórficos conhecidos como VNTRs (ou minissatélites) compreendem várias unidades de repetições sucessivas, cada uma com oito a oitenta pares de base (pb) de comprimento. O número exato de repetições, assim como o comprimento da região de VNTR, varia de um alelo para outro, e diferentes alelos podem ser identificados pelo seu comprimento (JEFFREYS *et al.*, 1985b).

3.2 STRs

STRs (ou microssatélites) consistem em sequências de um a seis pares de base que estão repetidas várias vezes ao longo do genoma e são altamente polimórficas (TAUTZ, 1989). Os STRs tornaram-se uma poderosa ferramenta para identificação humana, uma vez que, para sua detecção, pode-se utilizar a técnica da Reação em Cadeia da Polimerase (PCR), pela qual o DNA é amplificado a partir de oligonucleotídeos (*primers*) situados em regiões não repetitivas que flanqueiam os marcadores polimórficos (WEIR, 1992).

Treze STRs, localizados em cromossomos diferentes, foram selecionados para formar o banco de dados conhecido como CODIS (*Combined DNA Indexing System*). Atualmente, há diversos produtos comerciais disponíveis para amplificar todos esses marcadores em uma só reação, sendo esta chamada de *PCR multiplex*. São amplamente utilizados por laboratórios forenses e, dessa forma, contribuem para a padronização e o estabelecimento de bancos de dados das mais variadas populações, espalhadas ao redor do mundo. Cada marcador do CODIS apresenta ao redor de 10 a 20 alelos. Estes *loci* foram justamente escolhidos como marcadores para o sistema CODIS por serem multialélicos. Cada *locus* do CODIS é uma sequência de poucos nucleotídeos, em geral no número de quatro, que aparecem repetidas sucessivamente. A grande variabilidade alélica de cada *loci* e a análise conjunta dos *locus* tornam a identificação pessoal e/ou estudo de vínculo de parentesco matematicamente exata (RAPLEY; WHITEHOUSE, 2007).

Entre os principais *kits* comerciais que possibilitam a amplificação em *PCR multiplex* de marcadores STRs para genotipagem, utilizados na rotina da maioria dos laboratórios forenses, destacam-se o *AmpFISTR® Identifier™*, da empresa *Applied Biosystems, Life Technologies* e o *PowerPlex® 16*, da *Promega Corporation*. Em ambos os *kits* estão inclusos os treze marcadores STRs do CODIS. Além destes, os *kits* possuem a amelogenina (marcador genético que identifica o cromossomo sexual) e os marcadores genéticos D2S1338 e D19S433, no kit *AmpFISTR® Identifier™*, e os marcadores PENTA D e PENTA E, no kit *PowerPlex® 16* (BUTLER, 2005).

Com um total de 15 marcadores genéticos em uma só reação de PCR, esses dois *kits* satisfazem as exigências necessárias para a maioria dos casos de identificação e/ou testes de paternidade denominados 'casos típicos', ou seja, naqueles em que há a possibilidade de análise de vínculo genético entre mãe, filho e o suposto pai (BETZ *et al.*, 2007).

Contudo há casos de vínculo de parentesco mais complexos que a utilização destes 15 STRs não garante o poder de discriminação desejado, podendo resultar em um caso sem solução. Testes de paternidade chamados 'casos atípicos', em que o suposto pai é falecido e não se dispõe de parentes suficientes para a reconstrução do genótipo do mesmo, ou nos casos em que os pretensos genitores são relacionados e possuem semelhança genética, estão entre as situações em que a utilização de 15 marcadores genéticos pode não apresentar resultados conclusivos. Até mesmo 'casos típicos' de investigação de paternidade podem ser dificultados, quando mutações são encontradas (BETZ *et al.*, 2007).

Em situações complexas, como citado acima, é necessária a inclusão de mais marcadores genéticos além dos disponíveis nos *kits* comerciais, a qual possibilitará o aumento do poder de discriminação, a fim de se obter um resultado conclusivo. Uma estratégia para aumentar o poder de discriminação pode ser o uso conjunto dos dois principais *kits* comerciais de identificação humana (*AmpFISTR® Identifiler™* e *PowerPlex® 16*), pois dessa forma ocorre a inclusão de dois STRs (PENTA D e PENTA E / D2S1338 e D19S433), além dos 15 STRs convencionais.

Todavia, para que um determinado STR possa ser utilizado em casos de investigação de vínculo de parentesco é necessário que possua características como: 1) ser altamente polimórfico (elevado número de alelos); 2) possuir herança mendeliana clássica; 3) apresentar elevado índice de discriminação e de heterozigose; 4) apresentar baixa taxa de mutação (FIGINI *et al.*, 2003).

Adicionalmente, dados populacionais devem ser reunidos para se ter o valor da frequência para cada um dos alelos nas diferentes populações. Assim, apenas após analisar STRs que possuam as quatro características citadas acima e a determinação da frequência alélica deles na população em investigação, torna-se possível determinar a probabilidade de um genótipo particular ocorrer de forma aleatória em uma população. Contudo, ainda que a frequência alélica de certo STR já tenha sido descrita em uma região, é necessário que cada laboratório utilize a frequência de alelos da sua própria população, se tal dado não estiver presente corre-se o risco de obtenção de dados não fidedignos. Ainda que o levantamento deste dado particular não seja um fator dificultoso, pois segundo a literatura, a análise de amostras oriundas de 100 indivíduos é suficiente para fazer projeções confiáveis a respeito da frequência de um genótipo em uma grande população (CHAKRABORTY, 1992), a maioria dos laboratórios não o realiza.

Com base no descrito, sete marcadores genéticos STR, encontrados no *kit* comercial *PowerPlex® CS7 System, Custom* (Promega Corporation), para amplificação de DNA através da técnica de *PCR multiplex*, foram aqui selecionados com o intuito de estabelecer as frequências

alélicas em indivíduos do Rio Grande do Sul, para inseri-los na rotina dos laboratórios forenses desse estado.

3.3 MARCADORES STRs LPL, F13B, FES/FPS, F13A01, PENTA C, PENTA D PENTA E

Os sete marcadores genéticos aos que se dará detalhes no presente estudo são: LPL, F13B, FES/FPS, F13A01, PENTA C, PENTA D e PENTA E. Destes, dois STR (PENTA D e PENTA E) já são amplamente estudados na rotina de laboratórios forenses ao redor do mundo, pois são componentes do *kit* comercial *PowerPlex® 16*, da *Promega Corporation*. Outros quatro STR (LPL, F13B, FES/FPS e F13A01) compunham o *kit* *FFFL Fluorescent STR System*, também da *Promega Corporation*, utilizado largamente no início da era dos *kits* comerciais para fins de identificação humana através do DNA.

3.3.1 LPL

A primeira vez que foi relatada a ocorrência de polimorfismos neste marcador genético foi em 1990, por Zuliani e Hobbs, no qual se observou a ocorrência de três diferentes formas de apresentação (alelos) em 20 indivíduos caucasianos, com repetições [TTTA]_n no íntron 6 do gene da lipoproteína lipase. Atualmente, sabe-se que o LPL localiza-se no braço curto do cromossomo 8 na posição 22 (8p22) e apresenta, normalmente, alelos (7-14), com tamanho de fragmento entre 105-133 pares de base.

3.3.2 F13B

Nishimura e Murray descreveram, no ano de 1992, tetranucleotídeos repetitivos em uma região no gene do fator XIII B da coagulação sanguínea, com 10 repetições TTTA e que apresentou herança mendeliana em três famílias estudadas. Localizado no braço longo do cromossomo 1, entre as posições 31 e 32.1 (1q31-1q32.1), apresenta, normalmente, 10 alelos (6-12, além de microvariantes) com tamanho de fragmento entre 169-193 pares de base

3.3.3 FES/FPS

Descrito pela primeira vez em 1991 por Polymeropoulos *et al.*, o marcador genético FES/FPS localiza-se no braço longo do cromossomo 15, entre as posições 25 e terminal (15q25-*qter*). Os polimorfismos repetitivos [ATTT]_n iniciam-se no par de base 4713 do íntron 5

do proto-oncogene humano c-fes/fps. Normalmente, são encontradas 11 formas de repetição (alelos 7-15) neste marcador, mas também já foram relatadas microvariâncias nos alelos 7, 10, 11, 12 e 13, de acordo com o *STR Fact Sheet*, disponível no *site* do *STR Base*. O tamanho de fragmento pode variar de 222–250 pb. Este marcador genético integrou um dos primeiros kits de amplificação através de *PCR multiplex*, juntamente com os marcadores TH01, VWA e F13A01 (BUTLER, 2005).

3.3.4 F13A01

O *locus* F13A01 foi descrito pelo mesmo grupo de pesquisadores (POLYMEROPOULOS *et al.*, 1991a; 1991b) que publicou os dados referentes ao marcador genético FEP/FPS na mesma edição da revista científica *Acid Nucleic Research*. Localiza-se no braço curto do cromossomo 6 entre as posições 24 e 25 (6p24-p25) e apresenta repetições polimórficas [AAAG]_n que começam no par de base 248 do íntron A da subunidade A no gene do fator XIII da coagulação sanguínea humana. Diversos alelos já foram descritos neste marcador genético (3-17), sendo que o tamanho do fragmento está entre 279-331 pb. Conforme mencionado, este marcador genético integrou um dos primeiros kits de amplificação através de *PCR multiplex*, juntamente com os marcadores TH01, VWA e FES/FPS.

3.3.5 PENTA C, PENTA D e PENTA E

Uma das principais desvantagens em utilizar regiões microssatélites de DNA é a ocorrência de artefatos (*stutters*; Figura 1) nas amplificações das regiões repetitivas, ocasionados por deslizamentos (*slippage*) da DNA polimerase durante os eventos de replicação. Interessantemente, o deslizamento ou *slippage* da DNA polimerase é motivo pelo qual se geraram ao longo da evolução as variantes polimórficas em cada *locus* (em geral, uma nova mutação a cada 300 gerações, em humanos) e, por isso, tais *loci* são os de interesse forense. Contudo, é por este mesmo motivo que nesses *loci* se podem encontrar mutações ocasionais que alteram o alelo original, podendo inferir na análise e a interpretação do dado do *locus*.

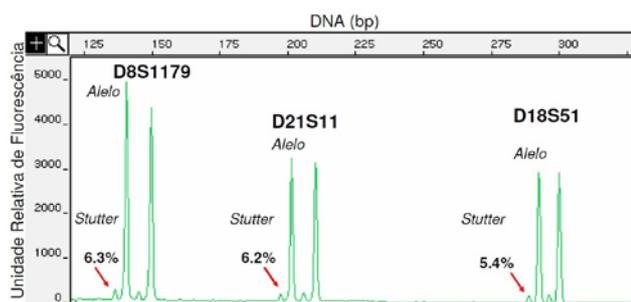


Figura 1 – Ocorrência de *stutters* em STRs (Fonte: *Applied Biosystems*, 2009).

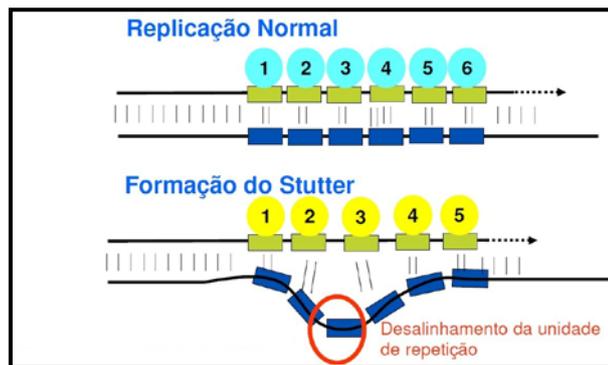


Figura 2– Processo de formação dos *slippage* (Fonte: *Applied Biosystems*, 2009).

Com o intuito de diminuir a ocorrência de *stutter* nas ampliações de DNA, *Bacher et al.* (1998) realizaram teste de validação de alguns *loci* que se caracterizavam por apresentar cinco unidades repetitivas (pentanucleotídeos), estando entre estes os marcadores genéticos utilizados no presente estudo (PENTA C, PENTA D e PENTA E).

O *loci* PENTA C está localizado no braço curto do cromossomo 9 na posição 13 (9p13). Os alelos encontrados habitualmente são 4-15 e 17. O tamanho do fragmento pode ser de 104-169 pb e sua unidade repetitiva é [AAAAC]_n. Não há muita informação em relação a este marcador genético, tendo em vista que poucos estudos populacionais foram realizados. Foi componente do *kit* de amplificação *PENTA BEC*, que também incluía outros dois pentanucleotídeos (PENTA B e E). Esse produto não permaneceu no mercado por muito tempo, o que impossibilitou a realização e a publicação de muitos trabalhos relacionados.

O marcador genético PENTA D está localizado no braço longo do cromossomo 21 (21q). Os alelos encontrados habitualmente são 2.2, 3.2, 5-17. O tamanho do fragmento pode ser de 373-446 pb e sua unidade repetitiva é [AAAGA]_n.

Já o marcador genético PENTA E está localizado no braço longo do cromossomo 15 (15q). Os alelos encontrados habitualmente são 5-24. Assim como o PENTA D, o tamanho do fragmento de DNA do PENTA E é bem maior que os fragmentos dos demais marcadores genéticos do *kit* comercial *PowerPlex® CS7 System, Custom*, podendo ser de 376-471 pb, sendo que a unidade repetitiva também é [AAAGA]_n.

3.4 ESTUDOS POPULACIONAIS E APLICABILIDADE NA GENÉTICA FORENSE

Estudos referentes a esses marcadores genéticos foram anteriormente publicados em algumas populações. Visto que os marcadores genéticos PENTA D e PENTA E são componentes do *kit* comercial *PowerPlex® 16*, alguns laboratórios forenses, de diferentes localidades, publicaram a frequência dos alelos mais comuns destes STRs. Entre esses trabalhos, citam-se as publicações de países como Japão (*HARA et al.*, 2004), Paquistão (*RAKHA et al.*, 2009),

Libano (CHOUERY *et al.*, 2010), Tunísia (CHERNI *et al.*, 2005), Bélgica (DECORTE *et al.*, 2006), Áustria (STEINLECHNER *et al.*, 2002), Costa Rica (RODRÍGUEZ *et al.*, 2007) e El Salvador (LOVO-GÓMEZ *et al.* 2007). Além disso, existem publicações outras regiões próximas ao Estado do Rio Grande do Sul. Entre essas, podemos citar as publicações de frequências alélicas dos marcadores genéticos PENTA D e PENTA E do centro e norte da Argentina (MARINO *et al.*, 2006a; 2006b) e dos estados brasileiros Santa Catarina (CAINÉ *et al.*, 2005; OCAMPOS *et al.*, 2009) e Rio de Janeiro (GOÉS *et al.*, 2004).

Trabalhos envolvendo a determinação das frequências alélicas dos marcadores LPL, F13B, FES/FPS e F13A01, que compunham o *kit* comercial *FFFL Fluorescent STR System*, foram publicados em países como: Butão (KRAAIJENBRINK *et al.*, 2007), Bélgica (DECORTE *et al.*, 2006), Polônia (JACEWICZ *et al.*, 2008), Portugal (ABRANTES *et al.*, 2006) e Colômbia (BRAVO *et al.*, 2004; CASTILLO *et al.*, 2009). No Brasil, foram publicados dados de frequências alélicas dos mesmos quatro marcadores genéticos, entre os quais estão as populações de São Paulo (BYDLOWSKI *et al.*, 2003), Rio Grande do Norte (SILVA *et al.*, 2003), Rio de Janeiro (SILVA *et al.*, 2004), Sergipe (PIMENTEL *et al.*, 2004), Paraíba (VIEIRA *et al.*, 2007) e Distrito Federal (DALTON *et al.*, 2009).

A utilização dessas frequências pode auxiliar na resolução de casos complexos de vínculo de parentesco, como as situações de supostos pais falecidos e poucos parentes disponíveis para a reconstrução genética, pretensos genitores relacionados e casos típicos de paternidade (mãe, filho e suposto pai) com a ocorrência de mutação.

Abrantes *et al.* (2006) descreve uma investigação de parentesco entre dois indivíduos, para verificar se estes poderiam ser irmãos bilaterais ou apenas unilaterais. Para esse estudo, foram utilizados os dois principais *kits* comerciais dos laboratórios forenses (*AmpFISTR® Identifier™* e *PowerPlex® 16*), além da inclusão do SE33, um marcador genético muito informativo altamente polimórfico. Não foram obtidos resultados satisfatórios que elucidassem tal investigação. Somente após a inclusão e a análise dos quatro marcadores do *kit FFL Fluorescent STR System* (LPL, F13B, FES/FPS e F13A01) foi possível atingir valores que praticamente provavam que os dois indivíduos eram irmãos bilaterais.

Balloch *et al.* (2008) relata um caso de investigação de paternidade com resultado de falsa exclusão, apresentado por outro laboratório, devido a duas inconsistências observadas entre o pretenso filho e o suposto pai. Após a repetição do exame, utilizando os mesmos marcadores da primeira investigação, as duas inconsistências foram novamente observadas. Adicionando seis marcadores STR (entre estes LPL, F13B, FES/FPS e F13A01), nenhuma inconsistência a mais foi verificada, o que apontou para uma provável ocorrência de mutação em dois marcadores genéticos. Foram, então, incorporados ao cálculo estatístico os valores das taxas de mutação nestes dois STRs inconsistentes. Os resultados obtidos apontaram um Índice de Paternidade Combinado (IPC) final com valor acima de 10 milhões com uma probabilidade de paternidade acima de 99,99999%.

Sánchez *et al.* (2008) apresenta outro caso de investigação de paternidade entre dois supostos pais e um pretenso filho sem a genitora. Os resultados obtidos nas primeiras

análises, através dos marcadores genéticos do *kit PowerPlex® 16*, não foram capazes de excluir a paternidade de nenhum dos dois indivíduos, pois não foi observada qualquer inconsistência. Os dois supostos pais não eram relacionados entre si. Após a inclusão dos quatro marcadores do *kit FFFL Fluorescent STR System*, foi observada uma inconsistência entre o pretense filho e um dos indivíduos no marcador LPL. Somente através da análise do cromossomo Y é que foi possível a elucidação do caso, pois se confirmou que o pretense filho pertencia à mesma linhagem patrilinea do segundo suposto pai. Para esse caso, a ausência da genitora contribuiu enormemente para a ocorrência de duas inclusões de paternidade para o mesmo filho.

Portanto, ainda que na maioria das vezes as investigações de vínculo de parentesco tenham resultados conclusivos utilizando apenas os principais *kits* de identificação humana, há situações mais complexas, como as descritas acima, em que o uso de mais marcadores genéticos pode auxiliar a elucidação dos casos.

Atualmente, o estado do Rio Grande do Sul possui alguns dados populacionais já publicados referentes à frequência alélica de outros marcadores genéticos autossômicos mais largamente utilizados em laboratórios de identificação humana. Esses dados são de grande valia, principalmente para os estudos de vínculo genético, como as investigações de paternidade. Cita-se, como exemplo, a publicação do grupo de peritos do Laboratório de Investigação de Paternidade (LIP) da Fundação Estadual de Produção e Pesquisa em Saúde (FEPPS), que estabeleceu a frequência alélica de quinze marcadores genéticos STR, utilizados pelo *kit* comercial *AmpFISTR® Identifiler™* (CHULA *et al.*, 2007). Os dados fornecidos por essa publicação oferecem aos laboratórios forenses desse estado informações fidedignas sobre a frequência dos alelos mais comuns na população local, permitindo o seu uso na identificação humana e nas análises de vínculo genético, obtendo maior exatidão nos resultados. Contudo não há qualquer publicação no Rio Grande do Sul referente à frequência alélica dos sete marcadores STR a serem estudados no presente trabalho.

Sendo assim, espera-se que o estabelecimento dessas frequências, determinadas através deste estudo, permita a utilização de tais dados nos laboratórios forenses, para auxiliar na resolução de diversos casos complexos ou sem solução, quando as informações contidas nos marcadores genéticos dos *kits* de rotina forense não garantem a conclusão da investigação.

OBJETIVOS

O objetivo deste trabalho foi estabelecer, em uma amostra representativa da população do estado do Rio Grande do Sul, a frequência de sete STRs, utilizando o *kit* comercial *PowerPlex® CS7 System, Custom* (Promega Corporation) para a amplificação dos fragmentos de DNA específicos. Especificamente, a nos propusemos a:

- 1- Estabelecer, na população local, a frequência de sete *loci* autossômicos: LPL, F13B, FES/FPS, F13A01, PENTA C, PENTA D e PENTA E.
- 2- Determinar e comparar as frequências de cada marcador nesta população com as frequências já relatadas em outras populações.
- 3- Calcular a diversidade gênica que cada *locus* apresenta.

REFERÊNCIAS BIBLIOGRÁFICAS

ABOUD, Maurice J.; GASSMANN, Marcus; MCCORD, Bruce R. The development of mini pentameric STR loci for rapid analysis of forensic DNA samples on a microfluidic system. **Electrophoresis**, Goettingen, v. 31, n. 15, p. 2672-2679, Jul. 2010.

ABRANTES, D. *et al.* Complex paternity investigations: the need for more genetic information. **International Congress Series**, v. 1288, p. 465-467, 2006.

BALLOCH, K. J. D. *et al.* Reporting paternity testing results when 2 exclusions are encountered. **Forensic Science International: Genetics Supplement Series 1**, p. 492-493, 2008.

BETZ, Thomas *et al.* "Paterniplex", a highly discriminative decaplex STR multiplex tailored for investigating special problems in paternity testing. **Electrophoresis**, Goettingen, v. 28, n. 21, p. 3868-3874, Nov. 2007.

BRAVO, M. L. J. *et al.* Analysis of 12 STR loci in Antioquia (Colombia) population sample. **International Congress Series**, v. 1261, p. 151-153, 2004.

BUTLER, John M. **Forensic DNA typing: biology, technology, and genetics of STR markers**. Oxford: Elsevier, 2005. p. 85-121.

BYDLOWSKI, Sergio Paulo *et al.* Genetic data on 12 STRs (F13A01, F13B, FESFPS, LPL, CSF1PO, TPOX, TH01, vWA, D16S539, D7S820, D13S317, D5S818) from four ethnic groups of Sao Paulo, Brazil. **Forensic Science International**, v. 135, p. 67-71, 2003.

CAINÉ, Laura M. *et al.* Genetic data of five STR loci in a population sample of Santa Catarina, Brazil. **Journal of Forensic Sciences**, v. 50, p. 1-2, 2005.

CASTILLO, Adriana *et al.* Population genetic data for F13A01, FES/FPS, F13B and LPL in Colombia (Department of Santander). **Forensic Science International: Genetics Supplement Series 2**, p. 357-358, 2009.

CHAKRABORTY, Ranajit. Sample size requirements for addressing the population genetic issues of forensic use of DNA typing. **Human Biology**, v. 64, n. 2, p. 141-159, Apr. 1992.

CHERNI, L. *et al.* Data for 15 autosomal STR markers (Powerplex 16 System) from two Tunisian populations: Kesra (Berber) and Zriba (Arab). **Forensic Science International**, v. 147, p. 101-106, 2005.

- CHOUERY, Eliane *et al.* Population genetic data for 17 STR markers from Lebanon. **Legal Medicine**, v. 12, p. 324-326, 2010.
- CHULA, Fernanda Goulart Lanes *et al.* 15 STR loci frequencies with mutation rates in the population from Rio Grande do Sul, Southern Brazil. **Forensic Science International: Genetics** 3, e35-e38, 2009.
- DALTON, Gustavo C. *et al.* Genetic profile of Federal District of Brazil based on 18 STR autosomal loci. **Forensic Science International: Genetics Supplement Series** 2, p. 373-375, 2009.
- DECORTE, Ronny *et al.* Allele frequency data for 19 short tandem repeats (PowerPlex 16 and FFFL) in a Belgian population sample. **Journal of Forensic Sciences**, v. 51, p. 436-437, 2006.
- FIGINI, Adriano R. L. *et al.* **Identificação Humana**. Campinas: Millenium, 2003. p. 256-257.
- GÓES, Andréa Carla de Souza *et al.* Allele frequencies data and statistic parameters for 16 STR loci – D19S433, D2S1338, CSF1PO, D16S539, D7S820, D21S11, D18S51, D13S317, D5S818, FGA, Penta E, TH01, vWA, D8S1179, TPOX, D3S1358 – in the Rio de Janeiro population, Brazil. **Forensic Science International**, v. 140, p. 131-132, 2004.
- HARA, Masaaki *et al.* Population data for 15 STR loci D3S1358, TH01, D21S11, D18S51, Penta E, D5S818, D13S317, D7S820, D16S539, CSF1PO, Penta D, vWA, D8S1179, TPOX and FGA in Japanese. **International Congress Series**, v. 1261, p. 204-206, 2004.
- JACEWICZ, R. *et al.* Population data of STR loci included in FFFL tetraplex in central Poland. **Forensic Science International: Genetics Supplement Series** 1, p. 337-339, 2008.
- JEFFREYS A. J.; WILSON, V.; THEIN, S. L.. Hypervariable 'minisatellite' regions in human DNA. **Nature**, v. 314, p. 67-73, 1985a.
- JEFFREYS A. J.; WILSON, V.; THEIN, S. L.. Individual-specific 'fingerprints' of human DNA. **Nature**, v. 316, p. 76-79, 1985b.
- KRAAIJENBRINK, Thirsa *et al.* Allele frequency distribution for 21 autosomal STR loci in Bhutan. **Forensic Science International**, v. 170, p. 68-72, 2007.
- LOVO-GÓMEZ, José; SALAS, Antonio; CARRACEDO, Ángel. Microsatellite autosomal genotyping data in four indigenous populations from El Salvador. **Forensic Science International**, v. 170, p. 86-91, 2007.
- MARINO, Miguel; SALA, Andrea; CORACH, Daniel. Genetic analysis of the populations from Northern and Mesopotamian provinces of Argentina by means of 15 autosomal STRs. **Forensic Science International**, v. 160, p. 224-230, 2006a.
- MARINO, Miguel; SALA, Andrea; CORACH, Daniel. Population genetic analysis of 15 autosomal STRs loci in the central region of Argentina. **Forensic Science International**, v. 161, p. 72-77, 2006b.
- NISHIMURA, D. Y.; MURRAY, J. C. A tetranucleotide repeat for the F13B locus. **Nucleic Acids Research**, v. 20, n. 5, p. 1167, 1992.
- PIMENTEL, B. J.; AZEVEDO, D. A. de; SILVA, L. A. F. da. Population genetics of eleven STR loci in the State of Sergipe Northeastern Brazil. **Journal of Forensic Sciences**, v. 49, p. 402-403, 2004.
- POLYMEROPOULOS, Mihael H. *et al.* Tetranucleotide repeat polymorphism at the human c-fes/fps proto-oncogene (FES). **Nucleic Acids Research**, v. 19, n. 14, p. 4018, 1991a.

POLYMEROPOULOS, Mihael H. *et al.* Tetranucleotide repeat polymorphism at the human coagulation factor XIII A subunit gene (F13A1). **Nucleic Acids Research**, v. 19, n. 15, p. 4306, 1991b.

OCAMPOS, Maristela *et al.* 15 STR loci frequencies in the population from Santa Catarina, Southern Brazil. **Forensic Science International: Genetics** 3, e129–e131, 2009.

RAKHA, Allah *et al.* Population genetic data on 15 autosomal STRs in a Pakistani population sample. **Legal Medicine**, v. 11, p. 305-307, 2009.

RAPLEY, Ralph; WHITEHOUSE, David. **Molecular Forensics**. West Sussex: Wiley, 2007. p. 91-126.

RODRÍGUEZ, A. *et al.* Population genetic data for 18 STR loci in Costa Rica. **Forensic Science International**, v. 168, p. 85-88, 2007.

SÁNCHEZ, Dora *et al.* False inclusion in a deficient paternity case with two alleged fathers. **Forensic Science International: Genetics Supplement Series** 1, p. 525-527, 2008.

SILVA, F. F. M. da *et al.* Microsatellite markers in the population from Rio Grande do Norte, Northeastern Brazil. **Journal of Forensic Sciences**, v. 48, p. 1189-1190, 2003.

SILVA, Rosane; MOURA-NETO, Rodrigo Soares de. Genetic diversity and admixture data on 11 STRs (F13B, TPOX, CSF1PO, F13A01, D7S820, LPL, TH01, vWA, D13S317, FESFPS, and D16S539) in a sample of Rio de Janeiro European-descendants population, Brazil. **Forensic Science International**, v. 142, p. 51-53, 2004.

SILVER, H. Paternity testing. **Critical Reviews in Clinical Laboratory Sciences**, v. 27, n. 5, p. 391-408, 1989.

SCHNEIDER, P. M. Basic issues in forensic DNA typing. **Forensic Science International**, v. 88, n. 1, p. 17-22, 1997.

STEINLECHNER, M. *et al.* STR loci Penta D and Penta E: Austrian Caucasian population data. **International Journal of Legal Medicine**, v. 116, p. 174-175, 2002.

TAUTZ, D. Hypervariability of simple sequences as a general source for polymorphic DNA markers. **Nucleic Acids Research**, v. 17, p. 6463-6417, 1989.

VIEIRA, A. *et al.* 13 STR loci frequencies in the population from Paraíba Northeast Brazil. **Forensic Science International**, v. 173, p. 231-234, 2007.

WALKER, R. H.; MEYERS, M. A.; PHILLIPS, L. M. The probability of exclusion of the HLA-A,B system in North American whites and blacks in parentage tests. **Transfusion**, v. 27, n. 1, p. 75-79, 1987.

WEIR, B. S. Independence of VNTR alleles defined as fixed bins. **Genetics**, v. 130, p. 873-887, 1992.

ZULIANI, G.; HOBBS, H. H. Tetranucleotide repeat polymorphism in the LPL gene. **Nucleic Acids Research**, v. 18, n. 16, p. 4958, 1990.

CAPÍTULO 2 - ARTIGO CIENTÍFICO

[Submetido ao periódico científico *International Journal of Legal Medicine*].

TITLE

Allele frequencies of seven STR *loci* (LPL, F13B, FES/FPS, F13A01, PENTA C, PENTA D and PENTA E) in the population from Rio Grande do Sul, Southern Brazil

AUTHORS

André Zoratto Gastaldo - Rodrigo Rodenbusch

Laboratório de Investigação de Paternidade
Fundação Estadual de Produção e Pesquisa em Saúde
Av. Ipiranga, 5400, 90610-000
Porto Alegre, RS, Brazil

André Zoratto Gastaldo - Rodrigo Rodenbusch

Programa de Pós-graduação em Biologia Celular e Molecular
Pontifícia Universidade Católica
Av. Ipiranga, 6681, 90619-900
Porto Alegre, RS, Brazil

Clarice Sampaio Alho

Departamento de Biologia Celular e Molecular
Faculdade de Biociências
Pontifícia Universidade Católica
Av. Ipiranga, 6681, 90619-900
Porto Alegre, RS, Brazil

ABSTRACT

Population data of seven short tandem repeat *loci* of the PowerPlex® CS7 were obtained from a sample of 401 individuals from Rio Grande do Sul (Southern Brazil). The *loci* are the two pentanucleotide STRs in the PowerPlex® 16 (PENTA D and PENTA E), an extra pentanucleotide (PENTA C), and four *loci* STR in the *FFFL Fluorescent STR System* (LPL, F13B, FES/FPS, and F13A01). Allele frequency and other forensically relevant statistics data were generated for these seven *loci*. All of the analyzed *loci* meet Hardy-Weinberg equilibrium expectations and no linkage disequilibrium were found, except for LPL *locus*. The observed and expected heterozygosity, power of discrimination and exclusion and polymorphic information content were calculated. The combined power of discrimination and combined power of exclusion were 0.999999987 and 0.998159054, respectively. Genetic distances between population from Rio Grande do Sul and other populations are presented.

KEYWORDS

Short tandem repeat *loci*, Allele frequencies, Population data, Genetic distances, Southern Brazil.

POPULATION

Rio Grande do Sul State is one of the 27 federal units in Brazil, located in the extreme south of the country (Figure 1). With a population of more than 10 million inhabitants, it is characterized by descendents from Europeans (especially portuguese, german and italian) and small number with African and Amerindian origins, mainly [1]. Blood samples from 401 unrelated individuals of this population (males and females) were obtained from cases of paternity tests. All participants signed a consent form.

DNA EXTRACTION

Genomic DNA was purified from dried blood samples preserved in FTA cards (Whatman Bioscience, Cambridge, UK) following the manufacturer's instructions.

PCR AMPLIFICATION AND TYPING

The extracted DNA was amplified using the PowerPlex[®] CS7 System, Custom PCR Amplification Kit (Promega Corporation, USA) in the GeneAmp 9700 PCR System (Applied Biosystems, Foster City, CA, USA) according to the manufacturer's recommendations. The analysis of the amplified PCR product was performed on the ABI 3130xl Genetic Analyzer (Applied Biosystems, Foster City, USA) capillary electrophoresis instrument. The data analysis and allele identification were performed using GeneMapper[®] ID (version 3.2) software.

QUALITY CONTROL

Proficiency testing of the GEP-ISFG Working Group.

ANALYSIS OF DATA

Calculations of allele frequencies, observed and expected heterozygosity and p -values of the Hardy-Weinberg equilibrium tests for all seven *loci* were assessed using CERVUS version 3.0.3 [2, 3]. Bonferroni's correction was used for Hardy-Weinberg equilibrium (HWE) test, which assumes that a 0.05 significance level obtained for 7 tests (one per locus) yields an actual significance threshold of 0.007 [4]. Power of discrimination (PD) and power of exclusion (PE) were estimated with PowerStats version 12 (Promega Corporation) [5]. Analyses of molecular variance (AMOVA, based on F -statistics; default values used for permutation tests) for four *loci* (LPL, F13B, FES/FPS, F13A01) were performed with ARLEQUIN version 3.1 [6]. UPGMA tree was built from Nei's genetic distance matrix using the PHYLIP software package version 3.69 [7] and visualized with Treeview software version 1.6.6 [8].

RESULTS

The allele frequencies and statistical parameters of all seven STR *loci* are available in **Table 1**, while the genetic distances of LPL, F13B, FES/FPS, F13A01 *loci* (Fst analysis) between the studied sample and other populations are available in **Table 2**. The observed heterozygosity varies between 0.7007 for FES/FPS and 0.9127 for Penta E. The power of discrimination was lowest for LPL: 0.846, and highest for PENTA E: 0.982. The power of exclusion ranges from 0.4290 for FES/FPS to 0.8210 for PENTA E. The combined power of discrimination (PD) and combined power of exclusion (PE) were 0.999999987 and 0.998159054, respectively. All the *loci* analyzed reached the Hardy-Weinberg equilibrium in the population studied ($P > 0.05$), except LPL *locus* ($P = 0.0009$). Even after application of the Bonferroni's correction, using the number of *loci* analysed, the differences observed were statistically significant for this *locus*. Almost half the studied sample has allele 10 in genotypes (0.4564). Kraaijenbrink *et al.* [9] also found HWE imbalance in the population of Bhutan, when analyzing the frequencies of alleles of the LPL *locus*. The frequency of allele 10 at this *locus* reached more than 64%. Rodrigues *et al.* [10] when discussing the results obtained in a study of 15 STR *loci* in a population of Rio de Janeiro, attributed the HWE imbalance in TH01 *locus* by a random excess of some genotypes, which caused the observed heterozygosity higher than expected, considering this marker unsuitable for independent testing in population samples. In the present study, the observed imbalance, probably, occurs for the same reason, and thus, the isolated use of LPL for population studies is also not recommended. This paper presents the allele frequencies of seven STR *loci*, which six of them (LPL, F13B, FES/FPS, F13A01, PENTA D and PENTA E) have not been published yet in the population of Rio Grande do Sul, but are used in forensic genetics for a long time with many published studies with diverse populations. The presentation of the allele frequencies of *locus* PENTA C, in this work, is one of the first publications in the scientific community, since there are no other published articles that used this marker beyond the manufacturer's kit [11,12]. Allelic frequencies for four *loci* in Rio Grande do Sul (Southern Brazil) population were compared with four populations: Belgium [13], European-derived [14], African-derived [14] and Asian-derived [14]. Based on gene frequencies of four STRs (LPL, F13B, FES/FPS, F13A01), to which data are available [13,14], pairwise genetic distances were calculated between populations using the Nei's formulas implemented in Phylip software. The analysis (see Figure 2) showed a clear separation between Southern Brazil and other populations studied, although, as expected, Southern Brazil is closer to European-derived and Belgium populations. In conclusion, the seven STR *loci* analyzed are a powerful tool for forensic identification and paternity testing and are very helpful in many cases, especially to assist in solving complex cases like paternity cases with mutations, single-parent paternity cases and family reconstructions cannot be sufficiently resolved with standard STR kits.

ACKNOWLEDGEMENTS

This work was supported by Laboratório de Investigação de Paternidade, CDCT/FEPPS, Promega Corporation and Human and Molecular Genetic Laboratory of PUCRS. We would like to thank Dr. Rodrigo Soares Moura-Neto (Laboratório de Biologia Molecular Forense – Instituto de Biologia/UFRJ, Brazil) and Dr. Ronny Decorte (Laboratory for Forensic Genetics and Molecular Archaeology - Leuven, Belgium) for providing the population sample.

REFERENCES

- [1] Instituto Brasileiro de Geografia e Estatística (IBGE): www.ibge.gov.br
- [2] CERVUS 3.0.3: <http://en.bio-soft.net/other/CERVUS.html>
- [3] Kalinowski ST, Taper ML, Marshall TC (2007) Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Molecular Ecology* 16: 1099-1106 doi:10.1111/j.1365-294X.2007.03089.x
- [4] Weir BS (1996) "Multiple tests" in *Genetic Data Analysis II*. Sinauer Associates Inc., Sunderland, Massachusetts, p 134.
- [5] PowerStats v.12 Software: www.promega.com/geneticidtools/powerstats
- [6] Excofier L, Laval GG, Schneider S (2005) Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* (1):47-50. Available on <http://cmpg.unibe.ch/software/arlequin3>
- [7] Felsenstein J (2009) PHYLIP: Phylogeny Inference Package version 3.69. Available on <http://evolution.gs.washington.edu/phylip.html>
- [8] Page RDM (1996) Treeview: an application to display phylogenetic trees on personal computers. *Comput. Appl. Biosci.* (12): 357–358.
- [9] Kraaijenbrink T, van Driem GL, Gaselô KT, Knijff P (2007) Allele frequency distribution for 21 autosomal STR *loci* in Bhutan. *Forensic Science International* 170: 68–72 doi: 10.1016/j.forsciint.2006.04.006
- [10] Rodrigues EL, Machado FB, Arruda MM, Moura-Neto RS, Medina-Acosta E (2009) Genetic data on 15 STR autosomal *loci* for a sample population of the Northern Region of the State of Rio de Janeiro, Brazil. *Forensic Science International: Genetics* 4: e25–e26 doi: 10.1016/j.fsigen.2009.01.008
- [11] Bacher J, Schumm JW (1998) Development of Highly Polymorphic Pentanucleotide Tandem Repeat *Loci* with Low Stutter. Profiles in DNA – Feature Article. Available on www.promega.com.

[12] Gorman K, Slinker Z, Johnson S, Beckwith M (2004) The Penta BEC Multiplex Primers from Promega: Additional *Loci* Available for Identity Testing. Profile in DNA – March, 2004. Available on www.promega.com.

[13] Decorte R, Verhoeven E, Vanhoutte E, Knaepen K, Cassiman JJ (2006) Allele Frequency Data for 19 Short Tandem Repeats (PowerPlex® 16 and FFFL) in a Belgian Population Sample. *J Forensic Sci* 51(2): 436-437 doi: 10.1111/j.1556-4029.2006.00089.x

[14] Bydlowski SP, Moura-Neto RS, Soares RPS, Silva R, Debes-Bravo AA, Morganti L (2003) Genetic data on 12 STRs (F13A01, F13B, FESFPS, LPL, CSF1PO, TPOX, TH01, vWA, D16S539, D7S820, D13S317, D5S818) from four ethnic groups of São Paulo, Brazil. *Forensic Science International* 135: 67-71 doi: 10.1016/S0379-0738(03)00174-9

LEGENDS OF TABLE AND FIGURES

Table 1 – Allele frequencies and statistical parameters measures for seven STR *loci* of PowerPlex® CS7 from Rio Grande do Sul, Southern Brazil, DNA samples.

Table 2 – Matrix of Nei's genetic distances based on the allelic frequencies of Southern Brazil (this study), Belgium [13], European-derived [14], African-derived [14] and Asian-derived [14] populations.

Figure 1 – Map of Rio Grande do Sul state (Southern Brazil) and the amount of individuals analyzed (N).

Figure 2 – Neighbor-joining tree of Southern Brazil (this study), Belgium [13], European-derived [14], African-derived [14] and Asian-derived [14] populations, using Nei's matrix as a measure of genetic distances.

Table 1

Allele	LPL	F13B	FES/FPS	F13A01	PENTA D	PENTA C	PENTA E
2.2					0.0162		
3.2				0.1160			
4				0.0686			
5				0.2244	0.0087	0.0162	0.0623
6		0.1301		0.2494			
6.4					0.0012		
7	0.0062	0.0301		0.2781	0.0037	0.0050	0.1197
8		0.2123	0.0224	0.0237	0.0299	0.0262	0.0349
9	0.0511	0.2616	0.0037		0.2145	0.1783	0.0137
9.2					0.0025		
9.3							
10	0.4564	0.3616	0.2731	0.0012	0.1496	0.0574	0.0698
10.2							
10.3			0.0062				
11	0.2357	0.041	0.4140	0.0025	0.1546	0.3466	0.1259
11.3							
12	0.2057		0.2332	0.0025	0.1883	0.2431	0.1633
12.2							
12.3							
13	0.0424		0.0449	0.0100	0.1509	0.1072	0.1035
13.2							
13.3							
14	0.0025		0.0025	0.0037	0.0611	0.0175	0.0586
14.2							
14.3							
15				0.0162	0.0150		0.0748
15.2							
15.3							
16				0.0037	0.0037	0.0025	0.0449
16.1							
16.2							
16.3							
17							0.0461
17.2							
17.3							
18							0.0212
18.2							
18.3							
19							0.0175
19.1							
19.2							
19.3							
20							0.0200
20.3							
21							0.0112
21.2							
22							0.0087
22.2							
23							0.0025
23.2							
24							
24.2							
24.3							
25							
26							0.0012
	LPL	F13B	FES/FPS	F13A01	PENTA D	PENTA C	PENTA E
N	802	730	802	802	802	802	802
Ho	0.7307	0.7205	0.7007	0.7805	0.8529	0.7506	0.9127
He	0.6903	0.7388	0.6980	0.7920	0.8453	0.7738	0.9088
PCI	0.6412	0.6932	0.6433	0.7599	0.8250	0.7404	0.9006
PD	0.8460	0.8850	0.8550	0.9270	0.9540	0.9160	0.9820
PE	0.4770	0.4610	0.4290	0.5630	0.7010	0.5110	0.8210
P	0.0009*	0.5586	0.7376	0.2430	0.4666	0.2819	0.5924

N: number of chromosomes; Ho: observed Heterozygosity; He: expected Heterozygosity; PIC: polymorphism information content; PD: power of discrimination; PE: power of exclusion; P: Hardy-Weinberg equilibrium exact probability tests; *P: statistically significant after Bonferroni's correction.

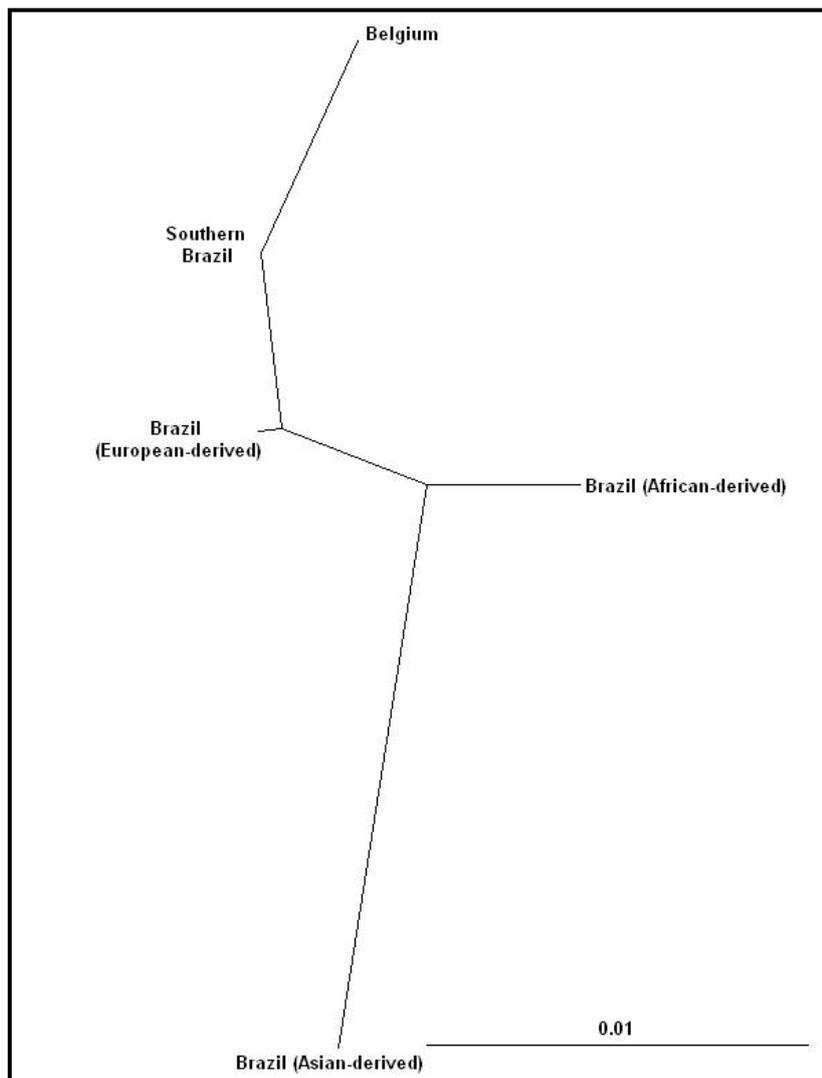
Table 2

Microsatellite markers	Southern Brazil x Belgium	Southern Brazil x Asia-derived	Southern Brazil x African-derived	Southern Brazil x European-derived
LPL	0.00533	0.01419	0.00889	0.01230
F13B	0.01028	0.08724	0.02473	-0.00102
FESFPS	-0.00089	0.03569	0.01261	0.00263
F13A01	0.00845	0.00243	0.01690	0.00659

Figure 1



Figure 2



CAPÍTULO 3 - CONSIDERAÇÕES FINAIS

A população avaliada para os fins deste estudo consiste em amostras oriundas de 401 indivíduos adultos, não relacionados, do sexo feminino e masculino, residentes no Rio Grande do Sul. Essas amostras de DNA foram coletadas, com o consentimento informado dos sujeitos doadores, para os fins da identificação molecular e vínculo de parentesco, tendo sido incluídas no estudo amostras aleatórias pertencentes a indivíduos provenientes de diferentes cidades gaúchas, de forma a serem representativas da mistura étnica do Estado. Os procedimentos técnicos e éticos desta pesquisa foram submetidos, respectivamente, à Comissão Científica da Faculdade de Biociências da PUCRS e ao Comitê de Ética em Pesquisa da PUCRS, a fim de cumprir as Normas Internacionais de Pesquisa em Seres Humanos. Seguindo estritamente o recomendado pelo CONSELHO NACIONAL DE SAÚDE na sua RESOLUÇÃO Nº 196, de 10 de outubro de 1996, item III - ASPECTOS ÉTICOS DA PESQUISA ENVOLVENDO SERES HUMANOS, esta pesquisa se comprometeu a atender às exigências éticas e científicas fundamentais adotando como eixo o subitem III.1 do documento citado, levou-se em consideração o que está definido no parágrafo da letra "i", o qual diz "prever procedimentos que assegurem a confidencialidade e a privacidade, a proteção da imagem e a não estigmatização, garantindo a não utilização das informações em prejuízo das pessoas e/ou das comunidades, inclusive em termos de auto-estima, de prestígio e/ou econômico – financeiro". Para isso, ainda dentro da mesma resolução, no seu item VI - PROTOCOLO DE PESQUISA, o presente trabalho foi concebido atendendo aos requisitos considerados adequados pelo CEP-PUCRS constantes na *"Proposta de Normas e Diretrizes para a Institucionalização e Utilização de Dados e Amostras de Bancos de Material Biológico Humano em Projetos de Pesquisa"* (BioBanco-PUCRS), o qual prevê proteger a dignidade e a privacidade dos sujeitos que doaram amostras de material biológico, de forma que todo material biológico armazenado estará disponível para uso dos pesquisadores, sempre que autorizado pelo sujeito doador, tendo a identificação preservada. Assim, as amostras foram mascaradas, de maneira que os pesquisadores não tiveram acesso aos dados de identificação pessoal. O uso das amostras anônimas preservou a privacidade de identificação dos participantes.

Em relação às análises genéticas, as regiões adjacentes aos STRs de interesse foram amplificadas, simultaneamente, através da reação em cadeia da polimerase (*PCR multiplex*), os fragmentos de PCR amplificados foram separados por eletroforese capilar para, por fim, se obter o genótipo de cada indivíduo. O laboratório onde foi realizado o estudo possui um rigoroso sistema de controle de qualidade interna com P.O.P. (Procedimento Operacional Padrão) o que permitiu a confiabilidade do resultado obtido. Como resultado se determinou as frequências alélicas e os parâmetros estatísticos para todos os sete *loci* STR de interesse, bem como, as distâncias genéticas dos *loci* LPL, F13B, FESFPS, F13A01 (análise F_{ST}) entre a população do Rio Grande do Sul e outras populações com dados publicados anteriormente. Os resultados mostraram que o menor poder de discriminação foi em LPL: 0,846, e o maior em

PENTA E: 0,982. O intervalo do poder de exclusão ficou entre 0,4290 para FES/FPS e 0,8210 para PENTA E. O poder combinado de discriminação (PD) e o poder combinado de exclusão (PE) foram 0,999999987 e 0,998159054, respectivamente. A heterozigotidade observada variou entre 0,7007 para FESFPS e 0,9127 para Penta E. Todos os *loci* analisados atingiram o equilíbrio de Hardy-Weinberg na população estudada ($P > 0,05$), exceto o *locus* LPL ($P = 0,0009$). Mesmo após a aplicação da correção de Bonferroni, utilizando o número de *loci* analisados, as diferenças observadas foram estatisticamente significativas para este marcador genético. Quase metade da população estudada possui o alelo 10 em seus genótipos (0,4564). Citou-se que outros autores também encontraram desequilíbrio de HW na população do Butão, ao analisar as frequências dos alelos do *locus* LPL: a frequência do alelo 10 neste marcador genético atingiu mais de 64%. Isso pode ser devido a um excesso aleatório de alguns genótipos.

O presente trabalho apresenta as frequências alélicas de seis *loci* STR (LPL, F13B, FES/FPS, F13A01, PENTA D e PENTA E) que ainda não tem publicações relacionadas à população do Rio Grande do Sul. Contudo, estes seis marcadores tem sido usados em genética forense há muito tempo e muitos trabalhos já foram publicados em outras populações. A apresentação das frequências alélicas do *locus* PENTA C, neste trabalho, é uma das primeiras publicações na comunidade científica, uma vez que não existem outros artigos descritos que utilizaram este marcador genético para estudos populacionais. Aqui, ainda, se comparou as frequências alélicas de quatro *loci* (LPL, F13B, FESFPS, F13A01) na população do Rio Grande do Sul com outras populações: da Bélgica, de brasileiros descendentes de europeus, de brasileiros descendentes de africanos e de brasileiros descendentes de asiáticos. Com base nessas frequências, as distâncias genéticas foram calculadas notando-se uma clara separação entre o sul do Brasil e outras populações estudadas, embora, como esperado, a população do Rio Grande do Sul esteja mais próxima de brasileiros descendentes de europeus e da Bélgica.

Em conclusão, os sete *loci* STR analisados são uma ferramenta poderosa para a identificação forense e testes de paternidade e podem ser muito úteis em diversos casos, especialmente para ajudar na resolução de casos complexos, como investigações de paternidade com mutações, casos monoparentais (quando é feita a análise genética somente entre filho e pretensão genitor) e casos de reconstruções de família que podem não ser suficientemente resolvidos com os *kits* de STR convencionais.

Considerando os objetivos iniciais deste trabalho, julga-se que os mesmos foram atingidos uma vez que:

- 1- Foi estabelecida na população do RS a frequência de sete *loci* autossômicos: LPL, F13B, FES/FPS, F13A01, PENTA C, PENTA D e PENTA E.
- 2- Foram determinadas e foram comparadas com sucesso as frequências de cada marcador da população do RS com as frequências já registradas em outras populações.
- 3- Foi calculada a diversidade gênica que cada um destes *locus* apresenta.