Universidade de São Paulo
Faculdade de Medicina de Ribeirão Preto
Programa de Pós-graduação em Genética

JULIET FIGUEIREDO GONÇALVES DE SOUZA

GENOTIPAGEM DE MARCADORES DO TIPO *INDEL* EM UMA AMOSTRA DA
POPULAÇÃO URBANA BRASILEIRA

Ribeirão Preto
2020

JULIET FIGUEIREDO GONÇALVES DE SOUZA

GENOTIPAGEM DE MARCADORES DO TIPO *INDEL* EM UMA AMOSTRA DA
POPULAÇÃO URBANA BRASILEIRA

Dissertação apresentada ao Programa de Pós-
Graduação em Genética na Faculdade de
Medicina de Ribeirão Preto da Universidade de
São Paulo, para obtenção do título de Mestre em
Ciências

Área de concentração:

Genética

Orientador:

Prof. Dr. Aguinaldo Luiz Simões

Ribeirão Preto

2020

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

Souza, Juliet Figueiredo Gonçalves

Genotipagem de marcadores do tipo *InDel* em uma amostra da população urbana brasileira. Ribeirão Preto, 2020.

72 f.:il.; 30cm.

Dissertação de Mestrado, apresentada à: Faculdade de Medicina de Ribeirão Preto/ Universidade de São Paulo.

Área de concentração: Genética.

Orientador: Simões, Aguinaldo Luiz.

1. *InDel*; 2. misturas de DNA; 3. parâmetros populacionais; 4. parâmetros forenses; 4. distâncias genéticas; 5. Ancestralidade.

FOLHA DE APROVAÇÃO

Souza, Juliet Figueiredo Gonçalves.

Genotipagem de marcadores do tipo *InDel* em uma amostra da população urbana brasileira.

Dissertação apresentada ao Programa de Pós-Graduação em Genética da Faculdade de Medicina de Ribeirão Preto da Universidade de São Paulo, para obtenção do título de Mestre em Ciências, área de concentração: Genética

Aprovado em: _____

BANCA EXAMINADORA

Prof. Dr.: _____ Instituição: _____

Julgamento: _____ Assinatura: _____

Prof. Dr.: _____ Instituição: _____

Julgamento: _____ Assinatura: _____

Prof. Dr.: _____ Instituição: _____

Julgamento: _____ Assinatura: _____

AGRADECIMENTOS

Ao meu orientador, Prof. Dr. Aguinaldo Luiz Simões, por ter me dado a oportunidade de trabalhar em seu laboratório de pesquisa. Agradeço pelas conversas e confiança depositada.

Agradecimentos especiais aos meus colegas de laboratório, Aline Silva Paula Brazorotto, Ana Lúcia Pimentel, Cláudia Emília Vieira Wiesel, Igor Caetano Dias Alcarás, Maria Luísa de Barros Rodrigues, Maria do Carmo Tomitão Canas.

A minha mãe e meu Padrasto, Patrícia Figueiredo Gonçalves e Milton Sampaio pelos incentivos, por serem meu porto seguro e pessoas que me ensinaram ser forte em todos os momentos da vida.

Aos meus avôs que são como pais para mim, Paulo Sérgio Maia Gonçalves e Maria Eliete Figueiredo Gonçalves, por se fazerem presentes na minha vida, mesmo que de longe. Não deixaram de ligar nenhum dia do ano para lembrar que eles estão torcendo e esperando por mim.

Ao meu namorado Rodrigo Ramalho Maciel por aguentar meus momentos de estresses, ansiedades e por acreditar em mim nos momentos em que eu mesma não acreditava.

As minhas amigas de infância, Bárbara Adrião, Clarissa Mesquita e Nina Bemerguy por estarem sempre prontas para me ouvir e me acalmar. A Barbara novamente por tirar as cartas para mim quando eu ligo e digo “Tira as cartas pra mim e veja meu futuro?”

Aos funcionários do departamento de Genética, em especial a Susie Adriana Penha Nalon, por sempre ajudar nos esclarecimentos de dúvidas.

Ao Programa de Excelência Acadêmica da Coordenação de Aperfeiçoamento Pessoal de Nível Superior (CAPES/Proex) e à Fundação de Apoio ao Ensino, Pesquisa e Assistência (FAEPA) do Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto – USP, pelo suporte financeiro, fundamental para a realização deste trabalho.

RESUMO

Souza, J. F. G. Genotipagem de marcadores do tipo *InDel* em uma amostra da população urbana brasileira. [dissertação]. Ribeirão Preto: Universidade de São Paulo, Faculdade de Medicina de Ribeirão Preto; 2020.

Marcadores do tipo *InDel* podem ser utilizados em casos de misturas de DNA entre dois indivíduos para detectar e quantificar diferentes linhagens. Além disso, os locos *InDels* podem apresentar três genótipos, homozigoto para inserção (ii), deleção (dd) e heterozigoto (id). As misturas de DNA mais estudadas são DNA fetal livre de células e quimerismo pós-transplante de células tronco hematopoiéticas. Nesses dois casos, um locus será informativo quando os dois indivíduos apresentarem genótipos diferentes, ou seja, em casos de gestantes, a mãe obrigatoriamente terá que ter uma linhagem diferente do feto. Em casos de quimerismo, o doador e o receptor obrigatoriamente têm que ter genótipos diferentes para que se possa detectar e quantificar o DNA. Desta forma, o objetivo deste trabalho foi descrever um painel de 18 locos *InDels*, com *primers* flanqueadores e alelo-específicos que possa ser útil em detectar linhagens específicas de DNA em uma mesma solução. Dentre os 18 locos, nove já haviam sido descritos anteriormente em pesquisas desenvolvidas em nosso laboratório, outros nove foram descritos pela primeira vez em uma população brasileira neste trabalho. Foram calculadas as frequências alélicas desta população para utiliza-las em dados estatísticos (parâmetros populacionais, forenses, distâncias genéticas e estimativa de ancestralidade). Também foram desenvolvidas duas equações para determinar a informatividade dos locos em casos de misturas em que coexistem diferentes linhagens, ou seja, as equações foram desenvolvidas utilizando as frequências alélicas encontradas na população. Para casos de solução materno-fetal a equação é dada por pq e em casos de quimerismo por $2pq(1-pq)$. Os 18 locos apresentaram uma probabilidade mínima de 98% em que pelo menos um traga a informação para detectar uma linhagem específica. Os locos apresentaram valores altos de parâmetros forenses (Poder de exclusão: 96% e probabilidade de discriminação: 99%) considerando que são locos bialélicos. Nas análises de distâncias genéticas e ancestralidade realizadas em nossa população, a amostra da população brasileira está mais próxima das populações Europeias e Americanas e obteve maior proporção de ancestralidade da população Europeia, seguida de Africana e uma porcentagem baixa com Leste-asiático.

Palavras-chave: *InDel*; misturas de DNA; parâmetros populacionais, parâmetros forenses, distâncias genéticas e ancestralidade.

ABSTRACT

Souza, J. F. G. Genotyping of *InDel* type markers in a Brazilian urban population samples. [dissertation]. Ribeirão Preto: University de São Paulo, Ribeirão Preto Medical School; 2020.

InDel markers can be used to detect and quantify lineages in DNA mixtures from two individuals. Additionally, *InDel* locus can have three genotypes: insertion homozygous (ii), deletion homozygous (dd) or heterozygous (id). The most studied DNA mixtures are cell-free fetus DNA and after stem cell transplant chimerism. In both cases, a locus is informative if the individuals have different genotypes, and this is required to be able to detect and quantify different lineages. In pregnant case, the mother has to have one different lineage than the fetus. Similarly, in the chimerism case, it is required the donor and receptor to have different genotypes. Therefore, the aim of this study is to describe an 18 *InDel* panel, with flanking and allele-specific *primers* to aid specific lineage detection and quantification in DNA mixtures. Nine of these markers were described previously in our laboratory, the other nine were described for the first time for a Brazilian population in this work. Allelic frequencies were estimated to later calculate the statistical data (populational and forensic parameters, genetic distances and ancestry estimates). Two new equations to calculate locus informativeness for both cases of DNA mixtures were developed. These equations have as input the allele frequency. In pregnant-fetus case, the informativeness is given as pq , with p as insertion allele frequency and q as deletion allele frequency. Chimerism cases were calculated as $2pq(1-pq)$. The 18-locus have 98% probability of detecting a different lineage in mixtures on at least one locus. Furthermore, the panel have shown high values of forensic parameters (exclusion power: 96% and discrimination power: 99%), when considering they are only biallelic. In genetic distances and ancestry analysis of our population, it was found that our sample of Brazilian population is close to Europeans and Americans, and had greater fraction of European ancestry, followed by African and a low percentage of East Asian ancestries fraction.

Keywords: *InDel*, DNA Mixtures, populational parameters, forensic parameters, genetic distances and Ancestry.

LISTA DE ILUSTRAÇÕES

Figura 1 - Polimorfismo do tipo <i>InDel</i> bialélico: caracterizado pela presença ou ausência de um ou mais pares de bases.	3
Figura 2 - Os <i>primers F</i> (<i>forward</i>) e <i>R</i> (<i>reverse</i>) são flaqueadores a região de inserção/deleção e ambos alelos são amplificados. Os pares de <i>primers F</i> e <i>Ri</i> ou <i>Fi</i> e <i>R</i> são utilizados para amplificação específica do alelo <i>inserção</i> (A). Para amplificação do alelo <i>deleção</i> são usados os pares de <i>primers F</i> e <i>Rd</i> ou <i>Fd</i> e <i>R</i> (B).....	3
Figura 3. Padrão de bandas formado após migração eletroforética com <i>primers</i> flaqueadores. A seta azul indica o sentido da migração eletroforética para o polo positivo. As colunas estão numeradas de 1 a 4. As amostras em todos os géis estão ordenadas com: Marcador de peso Molecular de 100pb (Coluna 1), indivíduos homozigotos <i>ii</i> (Coluna 2), heterozigotos <i>id</i> (Coluna 3) e homozigoto <i>deleção dd</i> (Coluna 4).	18
Figura 4. Padrão de bandas formado pela inserção (<i>ii</i>) do locus. A seta azul indica o sentido da migração eletroforética para o polo positivo. As colunas estão numeradas de 1 a 3. Todos os géis estão ordenados com Marcador de peso Molecular de 100 a 300 pb (Coluna 1), indivíduos homozigotos <i>ii</i> (Coluna 2) e por fim, indivíduos homozigotos <i>dd</i> (Coluna 3).	20
Figura 5. Padrão de bandas formado pela deleção (<i>dd</i>) do locus. A seta azul indica o sentido da migração eletroforética para o polo positivo. Todos os géis estão ordenados com Marcador de peso Molecular de 100pb (Coluna 1), indivíduos homozigotos <i>dd</i> (Coluna 2) e por fim, indivíduos homozigotos <i>ii</i> (Coluna 3).	21
Figura 6. A seta azul indica o sentido da migração eletroforética para o polo positivo. Em todos os géis, as colunas estão numeradas de 1 a 5, na primeira coluna é visualizado o marcador de peso Molecular (100pb). No gel A, na segunda coluna, é visualizado um indivíduo heterozigoto <i>id</i> e homozigoto <i>dd/ii</i> , na terceira coluna, um indivíduo <i>dd/id/id</i> , na quarta coluna, um indivíduo triplo heterozigoto <i>id/id/id</i> , e por fim, na quinta coluna, um indivíduo <i>ii/id/dd</i> . No gel B, na segunda coluna, é visualizado um indivíduo <i>id</i> e duplo homozigoto <i>ii/dd</i> , na terceira coluna, é visualizado um indivíduo triplo homozigoto <i>ii/ii/dd</i> . Na quarta coluna, um indivíduo <i>id/ii/id</i> e na quinta coluna, um indivíduo cujo genótipo é <i>id/ii/ii</i> . No gel C, na segunda coluna, é visualizado um indivíduo triplo heterozigoto <i>id/id/id</i> . Na terceira coluna, um indivíduo <i>ii/id/id</i> , na quarta coluna, um indivíduo triplo heterozigoto <i>id/id/id</i> e por último, na quinta, um indivíduo <i>ii/dd/ii</i>	24
Figura 7. Gráfico de escala Multidimensional realizado com as distâncias genéticas (<i>Fst</i>) entre a nossa população (BSP) com populações retiradas do projeto 1000 genomas.	26
Figura 8. Resultados do software STRUCTURE utilizando os 17 locus na BSP e populações da referência retirada do projeto 1000. (a) Gráfico de barras - Fração de ancestralidade por indivíduo e (b) Dispersão de indivíduos entre os três grupos dos clusters: AFR (Africana) em cinza EUR (Europeia) em amarelo; em azul EAS (Leste asiático) em azul. c – Dispersão de indivíduos com a BSP (População do interior do estado de São Paulo) em vermelho.	29

LISTA DE TABELAS

Tabela 1 – Lista dos lócus selecionados. Em azul, lócus descritos anteriormente e, em preto, os descritos neste trabalho. Para cada lócus estão listados identificação do lócus (rs), número de indivíduos genotipados (n); número de filhos genotipados (Filhos); frequência do alelo inserção (p_i) e os multiplex no qual foram agrupados...	7
Tabela 2 - Os 18 Lócus selecionados, em azul, lócus descritos anteriormente e em preto, os lócus descritos no presente trabalho e número de identificação (rs) no dbSNP, do NCBI, localização cromossômica e tamanho de amplicons gerados com <i>primers</i> flanqueadores (FL _f /FL _r), inserção (INS) e Deleção (DEL).	10
Tabela 3. Sequências 5'-3' de <i>primers</i> flanqueadores (FL _F direto e FL _R reverso) e alelo-específicos (Ins e Del) para os lócus selecionados. Em azul, são os lócus já descritos em trabalhos anteriores desenvolvidos pelo nosso grupo de pesquisa e em preto, os lócus descritos neste trabalho.	11
Tabela 4 - Temperaturas (°C) de pareamento para cada par de <i>primer</i> . FL _F /FL _R : flanqueadores Direto e Reverso/; Ins/FL _R : inserção específico + flanqueador reverso e Del/FL _R : deleção específico + flanqueador reverso.	16
Tabela 5 – Quantidades de <i>Primers</i> em µL de uma solução com 2,5µM (Micro molar) para as reações com os pares de <i>primers</i> flanqueadores FL _F /FL _R , inserção específicos/flanqueador (Ins/FL _R) e deleção-específicos/flanqueador (Del/FL _R). O lócus rs33941885 não obteve amplificação, por isso não consta na tabela sua quantidade.	17
Tabela 6 – Numero de indivíduos que não obtiveram amplificação (na).	19
Tabela 7– Análise de segregação dos alelos realizada com os nove lócus <i>InDels</i> descritos neste trabalho e o total de trios genotipados para cada lócus.	22
Tabela 8 – Quantidade de <i>primers</i> flanqueadores (FL _f /FL _r), cloreto de Magnésio (MgCl ₂) em µL e temperatura de pareamento dos lócus em Multiplex.	23
Tabela 9 – Número de indivíduos observados, frequência do alelo i , valor de p para o equilíbrio de Hardy-Weinberg e heterozigose, observada e esperada.	25
Tabela 10 - Distâncias genéticas (Fst) entre as populações BSP (População do interior de São Paulo), AMR, EUR, SAS e EAS.	26
Tabela 11 – Frequências do alelo inserção e diferenças alélicas para os 18 lócus em uma amostra do interior de São Paulo e em populações de referência retiradas dos 1000 genomas (AMR, AFR, EAS e EUR). Em azul e vermelho, os lócus que foram descritos anteriormente. Em preto, os nove lócus descritos neste trabalho. O Lócus rs1610941 só consta frequência na População do Interior de São Paulo (BSP), o restante das frequências alélicas não consta no banco de dados.	27
Tabela 12 – Proporções de ancestralidade da BSP (População brasileira de São Paulo) e populações de comparação retiradas do projeto 1000 genomas (AMR, AFR, EUR, EAS e SAS).	28
Tabela 13 – Parâmetros forenses cálculos com o banco de 18 lócus - Valores de PIC , PD e PE , PCD e PCE .	30
Tabela 14 – Calculo da informatividade dos lócus <i>InDels</i> em casos de misturas materno-fetais e quimerismo pós transplante de medula óssea.	31

Sumário

INTRODUÇÃO	1
Misturas de DNA.....	1
InDels	2
Informatividade dos locos <i>InDel</i>	4
JUSTIFICATIVA.....	7
HIPÓTESE	8
OBJETIVOS	8
MATERIAL E MÉTODOS	9
Desenho experimental	9
Seleção dos locos.....	9
Desenho de primers	10
Amostra populacional	13
Aspectos Éticos.....	13
Análise laboratorial / Procedimentos laboratoriais.....	13
Preparação das amostras. Extração de DNA	14
Eletroforese em gel de poliacrilamida	14
Genotipagem	14
Análise estatística	14
RESULTADOS.....	15
Seleção dos locos.....	15
Condições laboratoriais.....	16
Genotipagem: determinação do padrão de bandas	17
Análise de segregação	21
Multiplexação	22
Análises estatísticas.....	24
Frequências alélicas, Heterozigose e aderência ao EHW	24

Distância Genética (Fst) e ancestralidade.....	25
Parâmetros Forenses	29
Informatividade em casos de misturas de DNA	31
DISCUSSÃO.....	32
Seleção dos locos.....	32
Genotipagem.....	34
Multiplexação	34
Frequências alélicas, heterozigoses e eq de H-W	34
Distância Genética (Fst) e Ancestralidade.	35
Parâmetros forenses	36
Informatividade na pesquisa de misturas de DNA	38
CONCLUSÃO	39
REFERÊNCIAS.....	40
Anexo de 1a 6.	45
ANEXO 1. EXTRAÇÃO DE DNA PARA AMOSTRAS DE SANGUE E SWAB BUCAL (ADAPTADO DE HIGUCHI 1989)	45
ANEXO 2. REAÇÃO EM CADEIA DA POLIMERASE (PCR – POLIMERASE CHAIN REACTION)	47
ANEXO 3. ELETROFORESE.....	48
ANEXO 4. COLORAÇÃO E SECAGEM DE GEL	49
ANEXO 5. ÍNDICES DE FREQUÊNCIAS GENOTÍPICAS E ALÉLICAS.....	50
ANEXO 6. Parâmetros forenses	52
APÊNDICE	54

INTRODUÇÃO

MISTURAS DE DNA

Soluções em que coexistem diferentes linhagens de DNA requerem técnicas específicas de detecção e dosagem de cada uma delas. Métodos convencionais utilizando marcadores STRs são limitados (Santurtún *et al.*, 2017), embora sejam os mais utilizados na identificação humana (Benito Ramos-González *et al.*, 2015). Sua utilização na amplificação de apenas uma linhagem de DNA torna-se difícil, uma vez que os *primers* empregados são complementares às sequências flanqueadoras das repetições, fazendo com que ocorra a amplificação do DNA mais prevalente (Liu *et al.*, 2017).

Dentre as misturas de DNA que requerem técnicas para dosagem, as mais estudadas são: DNA fetal livre de células em gestantes e quimerismo em pacientes submetidos a transplante de células tronco hematopoiéticas.

A presença de DNA fetal no sangue de gestantes foi verificada há mais de vinte anos (Lo *et al.*, 1997). Por uma diminuição da espessura de uma camada da placenta, o DNA fetal pode ser verificado no sangue materno.

Intercorrências que ocorrem na gravidez como pré-eclampsia e restrição do crescimento fetal podem estar associadas a uma disfunção na placenta. Portanto, a quantidade de DNA fetal pode estar relacionada a essas alterações, fazendo necessário a quantificação de DNA fetal na circulação materna durante a gravidez (Manokhina *et al.*, 2017).

A concentração de DNA fetal no plasma materno também pode variar de acordo com algumas características materno e fetal (Índice de massa corpórea, raça, tabagismo e presença de alterações cromossômicas fetais (Martin *et al.*, 2014).

No entanto, a proporção entre o DNA fetal e da gestante é muito baixa. Com o uso de *primers* complementares a regiões flanqueadoras, a amplificação ocorre preferencialmente pelo DNA materno, pois a competição pelos *primers* faz com que o DNA mais prevalente seja amplificado (Wagner *et al.*, 2009).

Na sexagem fetal, esta limitação é contornada com o uso de *primers* dirigidos a uma sequência exclusiva do feto (sequências do cromossomo Y), o que evita a competição entre as mesmas regiões do genoma materno e fetal (Horinek *et al.*, 2008).

Portanto, faz-se necessário o uso de *primers* alelo- específicos, que nos possibilita

detectar apenas o DNA fetal e quantificar possíveis alterações cromossômicas e complicações que ocorrem na gravidez.

Partindo da ideia de amplificação dirigida a apenas uma linhagem de DNA em uma mesma solução, Santos (2014) propôs o uso de marcadores inserção/deleção (*InDel*) e Brazorotto, 2017 demonstrou sua aplicabilidade utilizando *primers* alelo-específicos na quantificação do DNA fetal livre em parturientes, por qPCR (Brazorotto, 2017) e na quantificação de DNA residual em pacientes submetidos a Transplante de Medula Óssea, para monitoramento de quimerismo hematopoiético (Santos, 2018).

Receptores de aloenxertos apresentam perfil quimérico após transplantes de células tronco hematopoiéticas, sendo classificado como quimerismo completo (Todas as células são de origem do doador) ou misto (coexistência de células do doador e do receptor). Para acompanhamento de recidiva da doença e quantificação de DNA, faz-se necessário analisar regiões polimórficas que diferem entre doador e receptor (Santurtún *et al.*, 2017).

A utilização de marcadores do tipo *InDels* para determinação de quimerismo tem se mostrado uma ferramenta útil, pois o aumento do quimerismo misto pós transplante de células tronco tem sido relacionado a um alto risco de recidiva de leucemia (Velasco *et al.*, 2005).

O método utilizando os *InDels* é capaz em detectar mudanças precoces de quimerismo, monitoramento e previsão de eventos clínicos adversos pós transplante (Gendzekhadze *et al.*, 2013).

Estes marcadores analisados por qPCR representam um novo método para quantificar o DNA livre de células oriundo do doador no plasma do receptor em casos de transplante, sendo uteis para auxiliar no diagnóstico de rejeição (Dauber *et al.*, 2019).

INDELS

Como visto acima, Os *InDels* (figura 1) apresentam características que os tornam polimorfismos úteis para detecção e dosagem de diferentes linhagens de DNA. Estes marcadores são caracterizados pela presença ou ausência da inserção em um ou ambos os cromossomos homólogos. Portanto, podem apresentar apenas três genótipos: homozigoto inserção *ii*, deleção *dd* ou heterozigoto *id*.

Figura 1 - Polimorfismo do tipo *InDel* bialélico: caracterizado pela presença ou ausência de um ou mais pares de bases.

A possibilidade de desenhar *primers* alelo-específicos tanto para inserção quanto para deleção é um diferencial dos *InDels* (Figura 2). Com isso, podemos utilizá-los em análises de misturas de DNA, amplificando linhagens específicas em uma solução.

Figura 2 - Os *primers* *F* (*forward*) e *R* (*reverse*) são flaqueadores a região de inserção/deleção e ambos alelos são amplificados. Os pares de *primers* *F* e *Ri* ou *Fi* e *R* são utilizados para amplificação específica do alelo *inserção* (A). Para amplificação do alelo *deleção* são usados os pares de *primers* *F* e *Rd* ou *Fd* e *R* (B).

Para evitar o pareamento inespecífico do alelo *inserção*, os *primers* *Fi* e *Ri* são

complementares à sequência da inserção e os *primers* alelo – específicos para deleção *Fd* e *Rd*, são complementares à região flanqueadora da inserção, iniciando logo antes do ponto em que acontece a inserção e terminando logo após esse ponto, não incluindo o alelo inserção.

Os *InDels* estão espalhados por todo genoma humano (autossomos e cromossomos sexuais), com a mesma densidade (Francez *et al.*, 2012). São Caracterizados pela elevada estabilidade devido a uma baixa taxa mutacional e por isso são considerados como marcadores de evolução lenta (Bastos-Rodrigues *et al.*, 2006).

A genotipagem de pequenos *InDels* é relativamente fácil e de baixo custo, com uma abordagem eletroforética simples, seguida de eletroforese em gel de poliacrilamida (*PAGE*), mas também adequada para análise com tecnologias de alta produtividade (Francez *et al.*, 2012).

Polimorfismos *InDel* bialélicos são os mais abundantes em número, depois dos SNPs (Huang *et al.*, 2014). No genoma, encontram-se amplamente distribuídos, seja em regiões codificadoras ou não-codificadoras.

Os polimorfismos do tipo *InDel* também são usados em exames de identificação humana, criminal ou de paternidade pela abundância e ampla distribuição desses marcadores (Weber *et al.*, 2002; Mullaney *et al.*, 2010); por serem derivados de um único evento mutacional, tendo taxa de incidência de $2,3 \times 10^{-9}$ (Nachman e Crowell, 2000), quando comparados aos STRs, que têm taxa de incidência igual a $1,2 \times 10^{-3}$ (Weber e Wong, 1993).

Além disso, *InDels* podem ser considerados como marcadores informativos de ancestralidade (AIM) por poderem apresentar significativas diferenças de frequência alélica entre populações geograficamente distintas (Lan *et al.*, 2019) e por terem um número significativo de alelos que possam ser genotipados em pequenos amplicons, aumentando o sucesso das análises de DNA degradado (Cui *et al.*, 2020).

No contexto de ancestralidade da população brasileira, ela se destaca pela sua formação tri-híbrida que ocorreu inicialmente por ameríndios, europeus e africanos e mais recentemente em indivíduos de outras nacionalidades. O estudo de AIM na população brasileira é necessário para minimizar os efeitos da estratificação populacional em estudos genéticos.

Informatividade dos loci *InDel*

O desenho de *primers* alelo-específicos tanto para inserção quanto deleção, nos possibilita detectar e dosar uma linhagem específica, por exemplo, em uma solução de misturas de DNA, se tivermos um indivíduo homozigoto para deleção e o outro

heterozigoto, utiliza-se *primers* dirigidos a inserção para obter uma amplificação específica de apenas um alelo.

Para que os *InDels* sejam informativos em casos de misturas de DNA, é necessário que um indivíduo seja homozigoto para um alelo e o outro tenha genótipo diferente. Se a homozigose for de alelos *d* (deleção), deverão ser usados *primers* '*i*' dirigidos à inserção. Caso a homozigose for de alelos *i*, usa-se *primers* dirigidos à deleção.

Se ambos os indivíduos forem heterozigotos ou homozigotos iguais, o locus não é efetivo na determinação de linhagens diferentes, em outras palavras, não é informativo.

A probabilidade de um locus ser informativo pode ser calculada pelas suas frequências alélicas.

Nos casos de misturas de DNA materno-fetal, um locus é informativo quando a gestante é homozigota para um alelo e o feto heterozigoto. A probabilidade de que esta situação ocorra é facilmente definida pelo produto $p \cdot q$. Isto é, quando a heterozigose é máxima (50%) para um locus bialélico, a probabilidade de um locus ser informativo é de 25%.

No caso de quimerismo, um locus só será informativo se um indivíduo for homozigoto para um alelo e o outro, homozigoto oposto ou heterozigoto. Nesse caso, a probabilidade de um locus ser informativo é dada pelo produto $2pq(1-pq)$. Considerando heterozigose máxima de 50%, a probabilidade máxima de um locus ser informativo, nesses casos, é 37%.

A partir da razão $pq/(2pq(1-pq))$, observa-se que a probabilidade de um locus ser informativo em casos de quimerismo é sempre maior do que em casos de mistura materno-fetal.

Nós estimamos que um conjunto de 18 locus *InDels* com heterozigose entre 30 a 50% fornecem uma probabilidade mínima de 98% de que pelo menos um deles seja informativo em análises de mistura de DNA fetal na circulação materna. Já em pacientes submetidos a transplante de células tronco hematopoiética esta probabilidade é maior que 99%, para o mesmo conjunto de locus.

Histórico do laboratório

Em trabalhos anteriores desenvolvidos em nosso laboratório, foram descritos nove locus do tipo *InDels* com *primers* flanqueadores e alelo-específicos e genotipados em uma população brasileira.

A utilização de três locus *InDels* foi descrita para a detecção de sequências fetais no plasma materno por PCR em Tempo Real (qPCR) (Santos 2014). Nessa técnica, foram utilizados *primers* específicos ao alelo inserção, ou seja, *primers* complementares a

sequências de DNA exclusivas fetais que, ao serem utilizados em conjunto com um *primer* flanqueador correspondente, possibilitaram a detecção de DNA fetal, nas situações em que a mãe é homozigota (*del/del*) e o feto heterozigoto (*in/del*).

Posteriormente, o número de loci *InDels* foi ampliado, sendo descritos 16 novos loci, com *primers* flanqueadores e *primers* específicos a inserção (Ng, 2015), os parâmetros forenses calculados, apresentaram valores dentro do esperado e sugerem que estes loci podem ser úteis na composição de um painel eficiente para ser usado na genotipagem fetal não invasiva.

A aplicabilidade dos loci *InDel* foi demonstrada utilizando *primers* alelo-específicos na quantificação do DNA fetal livre em parturientes, por qPCR (Brazorotto, 2017) e na quantificação de DNA residual em pacientes submetidos a Transplante de Medula Óssea, para monitoramento de quimerismo hematopoiético (Santos, 2018).

Outros sete loci com pares de *primers* flanqueadores e alelo-específicos para inserção e deleção foram descritos e tiveram análise populacional realizada com 150 indivíduos. Os dados populacionais e forenses viabilizam o uso desses loci em projetos posteriores para análises de misturas de DNA, diagnóstico fetal não invasivo e análise de quimerismo em pacientes submetidos a transplantes de células hematopoiéticas (Rodrigues, 2018).

No entanto, para alguns loci que já foram descritos com *primers* flanqueadores em nosso laboratório, não foi possível o desenho dos *primers* alelo-específico inserção ou deleção. Em alguns dos casos o tamanho destes fragmentos continha uma porcentagem elevada de CG ou AT, que impossibilitava o desenho dos mesmos.

Portanto, o objetivo deste trabalho consiste em descrever nove loci, complementando outros nove já selecionados no laboratório que contivessem *primers* com capacidade de desenho alelo-específicos (Tabela 1). O painel completo será um grupo de 18 loci para compor um conjunto adequado de *InDels* que possam ser úteis em análises de misturas de DNA.

Tabela 1 – Lista dos lócus selecionados. Em azul, lócus descritos anteriormente e, em preto, os descritos neste trabalho. Para cada lócus estão listados identificação do lócus (rs), número de indivíduos genotipados (n); número de filhos genotipados (Filhos); frequência do alelo inserção (p_i) e os multiplex no qual foram agrupados.

	Rs	n	Filhos	p_i	Multiplex
1	rs35523917	150	75	0,5067	M1
2	rs148112183	150	75	0,5367	M1
3	rs72083748	150	75	0,5767	M1
4	rs66501513	150	75	0,6400	M2
5	rs11270375	150	75	0,7967	M2
6	rs138730495	150	75	0,7800	M3
7	rs151001596	150	75	0,5900	M3
8	rs55870141	150	75	0,5100	M4
9	rs1610941	150	75	0,2733	M4
10	rs67718373	159	80	0,3396	M5
11	rs201008794	160	80	0,5281	M5
12	rs11271404	160	78	0,5813	M5
13	rs34656661	154	76	0,6753	M6
14	rs113482469	160	80	0,6938	M6
15	rs141059351	159	79	0,4969	M6
16	rs67887232	158	76	0,6551	M7
17	rs149521996	158	80	0,8703	M7
18	rs33941885	150	71	0,5367	M7

JUSTIFICATIVA

Os *InDels* podem obter amplificação alelo específica de apenas uma linhagem em casos de misturas de DNA entre dois indivíduos.

Os *primers* alelo-específicos são úteis para detectar linhagens específicas de DNA presente em mistura, tais como: quimerismo em pacientes que já passaram por

transplantes de células tronco hematopoiéticas, análises forenses e *cffDNA* (*cell-free fetal DNA*).

Além da identificação específica, os *InDels* podem formar um grupo de AIM (Marcadores informativos de ancestralidade) por terem diferenças alélicas entre populações, podendo assim, formar um painel que estime proporção de ancestralidade, o que é indispensável em uma população miscigenada como o Brasil. Os, parâmetros populacionais e forenses nos permitem identificar as frequências alélicas e calcular Heterozigotes dos Locus, pois estas últimas são indispensáveis para obter probabilidade de diferenciar indivíduos. Além de nos auxiliar a verificar se os locus serão informativos em misturas de DNA entre dois indivíduos.

Quando as Heterozigotes dos locus atingem seus valores máximos, os valores de informatividade dos locus também apresentam valores elevados.

Portanto, é indispensável os conhecimentos dos genótipos, frequências alélicas e Heterozigotes para estimar quais locus serão informativos.

Com base no exposto, é aqui proposta a padronização de novos locus *InDel*, de modo que estes, em associação com outros já descritos no laboratório, possam compor um conjunto adequado para a detecção e dosagem de diferentes linhagens de DNA em uma mesma solução.

HIPÓTESE

Este trabalho tem como hipótese que o painel de 18 locus *InDel* seja suficientemente útil para ser empregado em casos de mistura de DNA.

Estima-se que um painel de locus tenha a probabilidade mínima de 98% de informatividade em uma amostra da população brasileira.

OBJETIVOS

Padronizar reações de PCR com *primers* flanqueadores e alelo-específicos para nove locus autossômicos do tipo *InDel*.

Determinar os genótipos correspondentes aos nove locus aqui descritos em 80 trios mãe-filho-pai e estimar as frequências alélicas e genotípicas correspondentes, além de outros parâmetros populacionais e forenses.

Testar reações de PCR multiplex (mais que dois pares de *primers* na mesma reação).

MATERIAL E MÉTODOS

DESENHO EXPERIMENTAL

Foram selecionados 18 loci em trabalhos anteriores a partir do banco de dados do *NCBI (dbSNP)*.

Nove desses loci já foram descritos nos trabalhos realizados em nosso laboratório (Santos 2014; Ng 2015; Brazorotto, 2017 e Rodrigues 2018]. Portanto, a descrição dos outros nove constitui o objetivo do presente trabalho.

Selecionamos, loci concordantes com os seguintes critérios: loci bialélicos, validados pelo Projeto 1000 Genomes, com heterozigose de 30 a 50% e tamanho da inserção entre 20 e 50 pb. Destes, foram selecionados aqueles que para os quais foi possível o desenho de ambos os *primers* alelo-específicos.

A genotipagem foi realizada utilizando os pares de *primers* flanqueadores para conhecimento das frequências alélicas.

Foram desenvolvidas duas equações que foram definidas utilizando as frequências alélicas para estimar quais loci serão úteis em análises de misturas de DNA (Solução materno-fetal e quimerismo pós-transplante).

Para cada loci foram determinadas as condições laboratoriais para diagnóstico por PCR convencional seguida de *PAGE*, as frequências alélicas (com teste de equilíbrio de H-W) e parâmetros populacionais (diversidade genética, *PIC*, *PE* e *PD*).

Foram estimadas as distâncias genéticas entre populações (*Fst*) e os cálculos de ancestralidade de uma população do interior de São Paulo (BSP) para estimar proporções de ancestralidade em uma população miscigenada.

SELEÇÃO DOS LÓCUS

Em projetos anteriores desenvolvidos pelo nosso grupo de pesquisa, foram realizadas buscas de loci *InDels* no banco de dados de *SNPs (dbSNP)* do National Center for Biotechnology Information – *NCBI* (<http://www.ncbi.nlm.nih.gov/SNP/>). Foi salvo um arquivo para cada cromossomo contendo todas os *InDels* registrados. Em seguida, os arquivos FASTA do banco de dados foram convertidos ao excel para posterior seleção de sequências a serem estudadas. Os dados foram organizados com o número de identificação (rs). Muitos loci não eram bialélicos, aqueles que possuíam, além da deleção, mais de uma inserção foram excluídos, bem como aqueles com inserção menor do que 17 pares de bases e mais de 4 repetições na

sequência.

As sequências então, foram escolhidas manualmente visando o desenho dos iniciadores. Por fim, 18 loci foram selecionados em projetos anteriores para compor o painel de *InDels*.

Na Tabela 2 estão apresentados os 18 loci selecionados para compor um painel de *InDels* que seja possa ser informativo em casos de misturas de DNA.

Tabela 2 - Os 18 Loci selecionados, em azul, loci descritos anteriormente e em preto, os loci descritos no presente trabalho e número de identificação (rs) no dbSNP, do *NCBI*, localização cromossômica e tamanho de amplicons gerados com *primers* flangeadores (FL_F/FL_R), inserção (INS) e Deleção (DEL).

n	Rs	Localização	Tamanho dos amplicons gerados (pb)		
		Cromossômica	FL _F /FL _R (in/del)	INS	DEL
1	rs35523917	Chr2 (p25.2)	234/210	197	189
2	rs148112183	Chr19(p13.3)	171/145	80	73
3	rs72083748	Chr17 (q24.3)	127/101	67	60
4	rs66501513	Chr6 (q27)	245/215	190	176
5	rs11270375	Chr12 (q21.31)	130/100	76	75
6	rs138730495	Chr1 (q23.3)	194/168	110	104
7	rs151001596	Chr5(p13.2)	115/75	74	58
8	rs55870141	Chr22(q12.2)	103/71	73	57
9	rs1610941	Chr5(q12.3)	138/112	104	97
10	rs11271404	Chr3 (p24.2)	225/186	169	153
11	rs67718373	Chr4 (q13.3)	160/136	71	68
12	rs201008794	Chr14 (q21.1)	115/85	75	63
13	rs113482469	Chr5 (p15.33)	224/200	194	188
14	rs34656661	Chr10 (p14)	168/139	130	124
15	rs141059351	Chr21 (q22.3)	121/99	82	75
16	rs67887232	Chr7 (q35)	221/186	175	170
17	rs33941885	Chr18 (q23)	153/132	81	76
18	rs149521996	Chr11 (p15.5)	114/93	89	85

DESENHO DE PRIMERS

Os *primers* foram desenhados com auxílio do *software GeneRunner* (<http://www.generunner.net/>); na tabela 3, estão listados o tamanho dos fragmentos gerados após amplificação. As temperaturas de pareamento foram verificados pelo

software online *In-Silico* PCR (<https://genome.ucsc.edu/cgi-bin/hgPcr>).

Tabela 3. Sequências 5'-3' de *primers* flangeadores (FL_F direto e FL_R reverso) e alelo-específicos (Ins e Del) para os lócus selecionados. Em azul, são os lócus já descritos em trabalhos anteriores desenvolvidos pelo nosso grupo de pesquisa e em preto, os lócus descritos neste trabalho.

Lócus	Primer	Sequência
rs35523917	FL Direto	AGGATTGACTCTTCTGATCAG
	Ins Direto	TTAGAGAACAGAAGAGGCAGG
	Del Direto	AGAGCAATGCCTAGTCCGC
	FL Reverso	ATCACTGTTGTTAGGACGCA
rs148112183	FL Direto	CCTCTCTGCTGTGTGGACTT
	Ins Direto	TCTCCCTGTCCGCACTGAT
	Del Direto	TGGAATCAGTTCCCCGCT
	FL Reverso	CTGGACCTCACTCTCCTCAC
rs72083748	FL Direto	CCATTTCTCAGTGTTGCTTG
	Ins Direto	ATTGCACATCACACATACCC
	Del Direto	TTCAGGGTGCAATGCTGTT
	FL Reverso	CCATCGAAGCATCAGGAGA
rs66501513	FL Direto	GAATGGCTTGAACCTGGGA
	Ins Direto	ACTCCAGTCTGGCAACAGAGA
	Del Direto	GCGATGGTGCCACTGTAAAA
	FL Reverso	AGGACTGCGTAGTTTTGAGC
rs11270375	FL Direto	GGTATCCCATCTTCAAAAGG
	Ins Direto	CTCCCCCAGTCTATTCCAA
	Del Direto	TCTGAGCTGCTGCTTCTCTG
	FL Reverso	CATAACCATGGGATCCACTAGT
rs138730495	FL Direto	GAACAGGCATGGATGCTCA
	Ins Direto	TGACCCAATCACACACTAAGG
	Del Direto	TGACAGGTGAGGTTAAGCTGG
	FL Reverso	AGCTTGGGCGACAAAGTGA
rs55870141	FL Direto	GTGAGTGAGGGGACAGTCGT
	Ins Direto	TGGCCTGATAACAGGGTAAGA
	Del Direto	AGTCGTGTTGAGTGCCCTC
	FL Reverso	TGCAGGCCTGACATAGACC
rs1610941	FL Direto	AACATGGGAAGTGTCTATTA
	Ins Direto	TTCGTTGTGATTAGAAGATTC
	Del Direto	CTGGGTCGACACTTTTATG
	FL Reverso	TATTGTGCTCATTTTCTGGG
rs151001596	FL Direto	TGGCGTTTCATCTCTTTGG

rs11271404	Ins Direto	GTTTCATTGTGCTGTGCCTGT
	Del Direto	CTGCACCATGTTAGTTAATACTA
	FL Reverso	CACAGGCTTGCAAATCTCTG
	FL Direto	TTCTTCACGTGGTGGATGG
rs67718373	Ins Direto	ATAGCAAGCCAACGCCACT
	Del Direto	GAGAAGACAATCACCATGCTCA
	FL Reverso	CTTTGTCACTTCTCCCCTTAG
	FL Direto	CCCACCACCTTCAATTCCCT
rs201008794	Ins Direto	CATGGATGTTGGTGCCTCC
	Del Direto	TAGGTTTGGTGTAGGTTACTGG
	FL Reverso	GTATGCGAAGTAGTGCCTGC
	FL Direto	TCAGAGTGATAAGCCACGGT
rs113482469	Ins Direto	CCCAAACCAGGATACCCT
	Del Direto	AGGCAGCTGCGCTGTAAA
	FL Reverso	ACAGAGAGAAGGCCAGTCTT
	FL Direto	TGATTTCCGACTGTGCCG
rs34656661	Ins Direto	TGGGGAAGGGTTCAAAGTG
	Del Direto	GTGCCGATCTCTGTCACAGT
	FL Reverso	GAAACCTCGTGTGTGCTGTTA
	FL Direto	TGACACTTTGTACTCAGCGA
rs141059351	Ins Direto	GAACATACTCCACTCTGTATAGC
	Del Direto	CAGCGATAGTGGTAGGACTTGA
	FL Reverso	GCTTTTGATGCTAATTCCCATG
	FL Direto	ACAAAGCCCAGGCATTCA
rs67887232	Ins Direto	AGCCCCCACCAAAATCTT
	Del Direto	ATAAAGGGGGAACGCAGAC
	FL Reverso	GGGCTTTATGGGACTCGTATT
	FL Direto	TGTCCTGAGGGTTTTGCTT
rs33941885	Ins Direto	ACTCCAGTCTGGCAACAGAGA
	Del Direto	GCGATGGTGCCACTGTAAAA
	FL Reverso	AAGGAATGCCTGGAATGAG
	FL Direto	TGGAGAATCTCAAACCTGGG
rs149521996	Ins Direto	ATGAAGCAAGGCTCCATCT
	Del Direto	CACACCAGGCTGGACAAAA
	FL Reverso	TTTGCGACTCCTGTAGTTGT
	FL Direto	CTCCCAAGGCCAAGTATGAA
	Ins Direto	ATTCCCCCAGACAACCTT
	Del Direto	GCCAAGTATGAAGTCTCCTCC
	FL Reverso	GACTGCCTTCATGGACAAAA

Após a seleção dos *lôcus* e desenho dos *primers*, foram estabelecidas as condições laboratoriais de amplificação dos iniciadores flanqueadores e alelo-específicos em PCR (Alguns com Temperaturas diferentes entre os *primers* flanqueadores, o que dificulta a padronização em PCR Multiplex)

AMOSTRA POPULACIONAL

Foram selecionados 80 trios (mãe-filho-pai) aleatoriamente com mãe e pai não aparentados e que tiveram a paternidade confirmada por testes de rotina do Serviço de Investigação de Paternidade do HCFMRP-USP, vinculado ao Departamento de Genética da FMRP-USP. As amostras coletadas para investigação de paternidade nos anos de 2003 e 2004 foi aprovado pelo CEP do Hospital das Clínicas da Faculdade de Medicina de Ribeirão Preto (HCFMRP) – USP, processo HCRP no8079/2004, e inseridas no banco de amostras Populações Urbanas para fins de pesquisa

ASPECTOS ÉTICOS

O presente projeto foi aprovado pelo Comitê de Ética em Pesquisa (CEP) do HCFMRP no parecer: 3.474.036

Número de aprovação do CEP (CAAE): 17588219.2.0000.5440

ANÁLISE LABORATORIAL / PROCEDIMENTOS

LABORATORIAIS

Os procedimentos laboratoriais e estatísticos foram executados de acordo com a descrição contida nos Protocolos do Laboratório de Genética Bioquímica, reproduzidos, para facilitar sua consulta, ao final deste texto na sessão ANEXOS:

ANEXO 1: Extração de DNA para amostras de sangue e swab bucal;

ANEXO 2: Reação em cadeia da Polimerase;

ANEXO 3: Eletroforese;

ANEXO 4: Coloração e secagem de Gel;

ANEXO 5: Índices de frequências genotípicas e alélicas;

ANEXO 6: Parâmetros forenses.

Serão apresentados aqui apenas as modificações específicas do presente trabalho.

Preparação das amostras. Extração de DNA

Foram utilizadas amostras de DNA já extraídas segundo protocolo descrito no ANEXO I de Higuchi (1989), a partir de amostras de sangue ou *swab* bucal, pelo Serviço de Investigação de Paternidade do HCFMRP-USP.

As amostras pertencentes ao banco *Populações Urbanas* encontram-se estocadas a temperatura de -20. Quando houve necessidade de uma nova extração das amostras selecionadas, foi utilizado o mesmo método de Higuchi.

Os ensaios de padronização da amplificação dos loci *inDel* por PCR foram realizados no termociclador *GeneAmp® PCR System 9700* (*Applied Biosystems*). Os reagentes e procedimento estão descritos no ANEXO 2).

Eletroforese em gel de poliacrilamida

Os fragmentos amplificados foram analisados por eletroforese em gel de poliacrilamida (PAGE) de acordo com o descrito nos ANEXOS 3 e 4.

Genotipagem

A genotipagem é realizada com os 80 trios (Mãe-Filho-Pai), em eletroforese de poliacrilamida (PAGE), que é uma técnica que nos possibilita visualizar as bandas de DNA de acordo com o tamanho dos amplicons de cada locus.

Para os loci *InDels*, pode-se visualizar duas bandas para indivíduos heterozigotos (id), já para indivíduo homozigoto inserção (ii) e deleção (dd), visualiza-se apenas uma banda.

Para o alelo específico-inserção, as bandas são visualizadas mais a cima no gel, pois elas apresentam uma sequência maior e mais pesada, com amplicons maiores em relação ao alelo específico-deleção.

O alelo específico para deleção apresenta a sequência deletada, ou seja, as bandas que são visualizadas no gel são mais leves e apresentam amplicons menores, que faz com que as bandas migrem mais rápido.

Através da genotipagem, pode-se conhecer os genótipos dos indivíduos, ou seja, conhecer a frequência dos alelos, para estimar homozigose, Heterozigoses e estimar parâmetros populacionais e forenses.

ANÁLISE ESTATÍSTICA

Para cada um dos nove loci descritos neste trabalho, foram estimados Parâmetros

Populacionais e Forenses considerando a amostra populacional parental. foram calculados valores combinados dos parâmetros forenses, informatividade em casos de misturas de DNA e estimativa de ancestralidade utilizando o painel de 18 loci.

As frequências alélicas (p_i), o número de genótipos esperados, Heterozigose Observada e Esperada (H_{obs} e H_{esp}) e aderência ao equilíbrio de Hardy-Weinberg foram calculadas pelo software Genepop 4.7.0 (Rousset, 2008).

O Conteúdo de Informação do Polimorfismo (PIC - Polymorphism Information Content) (Botstein et al., 1980); Poder de Exclusão (PE) (Jamieson, 1965) e Probabilidade Cumulativa de Exclusão (PCE) (Evetts and Weir, 1998); Poder de Discriminação (PD) e Poder Cumulativo de Discriminação. (PCD) (Jones, 1972) foram estimados usando o software Microsoft Office Excel. Todas as equações estão disponíveis nos ANEXOS 5 e 6.

A estimativa de Ancestralidade e distâncias genéticas (F_{st}) foram calculadas com indivíduos geneticamente independentes, utilizando o software STRUCTURE v2.3.4 (Falush; Stephens; Pritchard, 2003) e do software Genepop 4.7.0 (Rousset, 2008), respectivamente.

O gráfico Multidimensional utilizando as distâncias genéticas (F_{st}) foi realizado utilizando o software R v3.5.3.

RESULTADOS

SELEÇÃO DOS LÓCUS

A seleção dos loci foi realizada a partir do banco de dados *dbSNP* do *NCBI*, que apresentou um número inicial de *InDels* de mais de 12 milhões de sequências. Para reduzir este valor, foram aplicados filtros: sequências validadas pelo projeto 1000 genomas, heterozigose de 30% a 50%, loci bialélicos, com sequência de inserção maiores que 20 pb, eliminando sequências repetitivas. Destes filtros, resultaram 688 loci, buscou-se então aqueles que fosse possível o desenho de pares de *primers* flanqueadores e alelo inserção e deleção. Foi também considerado os tamanhos dos amplicons para possível descrição dos loci em multiplex.

O locus rs1610941 é o único locus que não foi validado pelo projeto 1000 genomas, pois a seleção do mesmo foi realizada em 2014. Ainda não havia validação pelo projeto dos loci. Entretanto, este locus foi escolhido pela possibilidade de desenho de

primers flanqueadores e alelo específicos inserção e deleção para compor o banco de 18 loci. Além disso, os cálculos dos parâmetros populacionais viabilizavam seu uso em análises forenses.

CONDIÇÕES LABORATORIAIS

As temperaturas de pareamento dos pares de *primers* Flanqueadores (FL_F/FL_R), Inserção (Ins/FL_i) e Deleção (Del/FL_d) estão listadas na tabela 4, para descrever estes pares a diferença entre temperaturas do *primer* Direto e reverso não podem ser maiores que 5°C.

Tabela 4 - Temperaturas (°C) de pareamento para cada par de *primer*. FL_F/FL_R: flanqueadores Direto e Reverso/; Ins/FL_R: inserção específico + flanqueador reverso e Del/FL_R: deleção específico + flanqueador reverso.

Locus	Pares de primers		
	FL _F /FL _R	Ins/FL _R	DEL/FL _R
rs11271404	58	58	56
rs201008794	55	58	55
rs67718373	55	55	55
rs113482469	60	60	60
rs34656661	55	62	57
rs141059351	60	60	60
rs67887232	59	59	55
rs33941885	59	-	57
rs149521996	59	57	59

As quantidades de reagentes que variaram foram de *primers* e cloreto (MgCl₂), o restante dos reagentes mantiveram as mesmas concentrações como descritos em anexo 2.

Para amplificação do locus rs34656661, utilizando os *primers* inserção específicos, foram necessários 0,80 µL de uma solução de cloreto de Magnésio 20mM.

Para amplificação do locus rs67718373, utilizando os *primers* deleção específicos, foram necessários 1,50 µL da mesma solução de Cloreto de Magnésio.

Todos as outras reações de amplificação foram feitas com 2,0 µL da mesma solução de Cloreto de Magnésio.

As concentrações necessárias de *primers* para as diferentes combinações de *primers* estão listadas na Tabela 5.

Tabela 5 – Quantidades de *Primers* em μL de uma solução com $2,5\mu\text{M}$ (Micro molar) para as reações com os pares de *primers* flanqueadores $\text{FL}_\text{F}/\text{FL}_\text{R}$, inserção específicos/flanqueador ($\text{Ins}/\text{FL}_\text{R}$) e deleção-específicos/flanqueador ($\text{Del}/\text{FL}_\text{R}$). O locus rs33941885 não obteve amplificação, por isso não consta na tabela sua quantidade.

Lócus	Par de <i>primers</i>		
	$\text{FL}_\text{F}/\text{FL}_\text{R}$	$\text{Ins}/\text{FL}_\text{r}$	$\text{Del}/\text{FL}_\text{r}$
rs11271404	2,00	1,00	2,00
rs201008794	2,00	1,00	1,00
rs67718373	1,00	1,00	1,00
rs113482469	2,00	1,00	1,00
rs34656661	2,00	1,00	2,00
rs141059351	1,00	2,00	2,00
rs67887232	2,00	1,00	2,00
rs33941885	2,00	-	2,00
rs149521996	2,00	2,00	2,00

*Em todas as reações de PCR foi utilizada uma temperatura inicial de aquecimento a 95°C por 11 minutos, seguidos de 30 ciclos de acordo com as temperaturas expostas acima por 1 minuto e uma extensão final 70°C por 10 minutos.

GENOTIPAGEM: DETERMINAÇÃO DO PADRÃO DE BANDAS

Para todos os lócus, as visualizações das bandas foram em *PAGE* e corados com nitrato de prata.

Os genótipos dos 225 indivíduos estão listados no apêndice.

Em todos os lócus amplificados com *primers* flanqueadores, como se vê na Fig. 3, o genótipo heterozigoto se caracteriza pela presença de duas bandas. A banda com migração mais lenta refere-se ao alelo inserção e a banda com migração mais rápida, ao alelo deleção. A presença apenas da banda inserção indica genótipo homozigoto ii e apenas da banda deleção, homozigose dd.

Figura 3. Padrão de bandas formado após migração eletroforética com *primers* flanqueadores. A seta azul indica o sentido da migração eletroforética para o polo positivo. As colunas estão numeradas de 1 a 4. As amostras em todos os géis estão ordenadas com: Marcador de peso Molecular de 100pb (Coluna 1), indivíduos homozigotos ii (Coluna 2), heterozigotos id (Coluna 3) e homozigoto deleção dd (Coluna 4).

Dos 225 indivíduos (Mão-filho-Pai) genotipados, ou seja, 160 pais e 80 filhos, 17 indivíduos não amplificaram para o locus rs33941885, isto não influenciou nas análises estatísticas, pois apenas 6 pais e 4 mães não amplificaram, o restante não amplificado foram números de filhos que não são inclusos nas análises estatísticas. O mesmo ocorre para o locus rs34656661, no qual 10 pessoas não apresentaram amplificação, dentre elas, 6 pais e 5 filhos.

O objetivo foi obter pelo menos 150 indivíduos parentais estudados em cada locus, isto foi obtido, mesmo com indivíduos não amplificados.

Os números de indivíduos não amplificados durante a genotipagem e seus

respectivos locus estão listados na tabela 6.

Tabela 6 – Numero de individuos que não obtiveram amplificação (na).

Locus	na
rs67718373	1
rs34656661	10
rs11271404	2
rs201008794	0
rs113482469	0
rs141059351	0
rs67887232	5
rs33941885	17
rs149521996	1

Com exceção do *primer* inserção do locus rs33941885, todos os alelos inserção-específicos apresentam amplificação nos indivíduos id ou ii, os indivíduos dd não apresentam amplificação (Figura 4).

Figura 4. Padrão de bandas formado pela inserção (ii) do locus. A seta azul indica o sentido da migração eletroforética para o polo positivo. As colunas estão numeradas de 1 a 3. Todos os géis estão ordenados com Marcador de peso Molecular de 100 a 300 pb (Coluna 1), indivíduos homozigotos ii (Coluna 2) e por fim, indivíduos homozigotos dd (Coluna 3).

O desenho de *primer* foi revisado e não encontramos erros referente ao seu desenho. As temperaturas e reagentes foram reajustados, houve troca amostras, ou seja, troca de indivíduos para testa-lo, e mesmo assim persistiu a não amplificação. Para os locus deleção específicos, a amplificação acontece apenas para indivíduos id ou dd, os indivíduos ii não apresentam amplificação. (Figura 5).

Figura 5. Padrão de bandas formado pela deleção (dd) do locus. A seta azul indica o sentido da migração eletroforética para o polo positivo. Todos os géis estão ordenados com Marcador de peso Molecular de 100pb (Coluna 1), indivíduos homozigotos dd (Coluna 2) e por fim, indivíduos homozigotos ii (Coluna 3).

ANALISE DE SEGREGAÇÃO

Foi realizada a análise de segregação paralelamente a genotipagem (Tabela 7), com os nove loci descritos no presente trabalho.

No locus rs113482469 não foi observado um trio M(dd)-P(dd)-F(dd), no rs34656661 não ocorreu M(id)-P(id)-F(ii), M(id)-P(id)-F(dd) e M(dd)-P(id) e F(id); no rs67887232 não foram vistos M(dd)-P(id)-F(id) e M(dd)-P(dd)-F(dd) e no rs149521996 não foram observados M(dd)-P(ii)-F(id), M(dd)-P(id) e F(dd) e M(dd)-P(dd)-F(dd).

Os genótipos analisados entre trios (Mãe-filho-Pai), corroboram com a segregação de marcadores co-dominantes, no qual pode-se distinguir indivíduos homozigotos e heterozigotos.

Tabela 7– Análise de segregação dos alelos realizada com os nove loci *InDels* descritos neste trabalho e o total de trios genotipados para cada locus.

Mãe	Pai	Fi- lho	rs67718373	rs11271404	rs201008794	rs113482469	rs34656661	rs141059351	rs67887232	rs149521996	rs33941885
ii	ii	ii	1	10	11	17	15	2	11	47	3
		id	3	6	6	10	10	3	7	1	6
		id	3	7	5	12	8	5	11	9	4
		dd	4	3	7	4	4	4	3	1	1
id	ii	ii	2	8	2	6	11	3	6	11	5
		id	2	4	5	7	7	9	5	4	5
		id	3	2	4	3	1	5	2	0	6
		id	13	14	11	3	3	10	8	1	5
		dd	2	3	4	6	3	9	5	1	8
		dd	7	4	7	2	2	6	4	1	6
		dd	3	3	4	1	1	1	2	1	3
dd	ii	id	1	4	1	5	4	9	7	0	3
		id	9	3	3	2	1	4	1	1	2
		dd	8	3	3	2	2	4	2	0	5
		dd	18	4	7	0	2	5	0	0	3
Total de trios			79	78	80	80	74	79	74	78	65

MULTIPLEXAÇÃO

Os nove loci descritos neste trabalho foram agrupados, de acordo com o tamanho dos amplicons para testar reações de multiplex (Figura 6).

Os outros nove loci já descritos, foram agrupados em triplex e duplex em pesquisas anteriores e demonstraram amplificação adequada, isto é, os amplicons entre eles os loci agrupados em multiplex são diferentes e não ficaram intercalados, o que facilita a visualização dos genótipos.

Para os 18 loci, o agrupamento não poderia ser em quadriplex ou mais loci, pois em muitos loci, o tamanho dos amplicons eram os mesmos ou muito próximos, fazendo com que as bandas entre os loci ficassem intercaladas.

As reações em Multiplex foram visualizadas em *PAGE*. Foi necessário ajustar a

temperatura de pareamento e concentrações de reagentes para evitar a competição entre os *primers* quando agrupados.

Na tabela 8 estão apresentadas as quantidades de reagentes e temperatura de pareamento dos lócus em Multiplex.

Tabela 8 – Quantidade de *primers* flanqueadores (FL_F/FL_R), cloreto de Magnésio (MgCl₂) em µL e temperatura de pareamento dos lócus em Multiplex.

	Multiplex 1		
	rs11271404	rs67718373	rs201008794
<i>Primer</i> FL _F /FL _R (µl)	2,00	2,00	2,00
MgCl ₂ (µl)	2,00		
Temperatura de pareamento (°C)	55		
	Multiplex 2		
	rs141059351	rs34656661	rs113482469
<i>Primer</i> FL _F /FL _R (µl)	1,00	1,00	1,00
MgCl ₂ (µl)	2,00		
Temperatura de pareamento (°C)	60		
	Multiplex 3		
	rs67887232	rs33941885	rs149521996
<i>Primer</i> FL _F /FL _R (µl)	2,00	2,00	1,00
MgCl ₂ (µl)	2,00		
Temperatura de pareamento (°C)	59		

Na figura 6 – A, B e C estão apresentados os lócus em Multiplex.

No gel A, correspondente ao Multiplex rs11271404 (225pb e 186pb), rs67718373 (160pb e 136pb) e rs201008794 (115pb e 85pb), na mesma ordem da migração eletroforética, o rs11271404 é o que migra mais próximo da origem e o rs201008794, o mais distante. Na primeira coluna é visualizado o Marcador de peso Molecular de 100pb e, nas outras colunas 4 indivíduos diferentes com as leituras indicadas na legenda da Fig. 6.

No gel B, correspondente ao Multiplex rs113482469 (224pb e 200pb), rs34656661 (168pb e 139pb) e rs141059351 (121pb e 99pb), na mesma ordem da migração eletroforética, o rs113482469 é o que migra mais próximo da origem e o rs141059351, o mais distante. Na primeira coluna é visualizado o Marcador de peso Molecular de 100pb e, nas outras colunas, 4 indivíduos diferentes com leituras indicadas na legenda.

No gel C, correspondente ao Multiplex rs67887232 (221pb e 186pb), rs33941885 (153 pb e 132pb) e rs149521996 (114pb e 93pb), na mesma ordem da migração eletroforética, o rs67887232 é o que migra mais próximo da origem e o rs149521996,

o mais distante. Na primeira coluna é visualizado o Marcador de peso Molecular de 100pb e, nas outras colunas, 4 indivíduos diferentes com leituras indicadas na legenda.

Figura 6. A seta azul indica o sentido da migração eletroforética para o polo positivo. Em todos os géis, as colunas estão numeradas de 1 a 5, na primeira coluna é visualizado o marcador de peso Molecular (100pb). No gel A, na segunda coluna, é visualizado um indivíduo heterozigoto id e homozigoto dd/ii, na terceira coluna, um indivíduo dd/id/id, na quarta coluna, um indivíduo triplo heterozigoto id/id/id, e por fim, na quinta coluna, um indivíduo ii/id/dd. No gel B, na segunda coluna, é visualizado um indivíduo id e duplo homozigoto ii/dd, na terceira coluna, é visualizado um indivíduo triplo homozigoto ii/ii/dd. Na quarta coluna, um indivíduo id/ii/id e na quinta coluna, um indivíduo cujo genótipo é id/ii/ii. No gel C, na segunda coluna, é visualizado um indivíduo triplo heterozigoto id/id/id. Na terceira coluna, um indivíduo ii/id/id, na quarta coluna, um indivíduo triplo heterozigoto id/id/id e por último, na quinta, um indivíduo ii/dd/ii.

ANÁLISES ESTATÍSTICAS

Frequências alélicas, Heterozigose e aderência ao EHW

Os indivíduos geneticamente independentes foram utilizados para o cálculo das frequências alélicas, Heterozigose e aderência ao EHW (Tabela 9).

A frequência do alelo inserção foi elevada, ou seja, acima de 50 % para todos os loci, com exceção do locus rs67718373(34%).

A frequência genotípica de homozigotos deleção do locus rs149521996 foi menor quando comparada aos demais loci, mas isso não afetou o equilíbrio de Hardy Weinberg.

Os valores de Heterozigose esperada variaram de 22,6% (rs149521996) 50%. Com exceção do locus (rs149521996), os demais loci apresentaram He dentro dos valores esperados de 30 a 50%.

Todos os loci apresentaram proporções genotípicas esperadas segundo o Equilíbrio de Hardy-Weinberg.

Tabela 9 – Número de indivíduos observados, frequência do alelo i , valor de p para o equilíbrio de Hardy-Weinberg e heterozigose, observada e esperada.

Lócus	n	Número de Indivíduos			p_i	p (EHW)	H_{obs}	H_{esp}
		observados						
		li	id	Dd				
rs67718373	159	17	74	68	0,3396	0,7297	0,4654	0,4486
rs34656661	154	74	60	20	0,6753	0,1976	0,3896	0,4385
rs11271404	160	54	78	28	0,5813	1,0000	0,4875	0,4868
rs201008794	160	48	73	39	0,5281	0,3393	0,4563	0,4984
rs113482469	160	78	66	16	0,6938	0,7135	0,4125	0,4249
rs141059351	159	37	84	38	0,4969	0,5263	0,5283	0,5000
rs67887232	158	69	69	20	0,6551	0,7286	0,4367	0,4519
rs33941885	150	41	79	30	0,5367	0,5122	0,5267	0,4973
rs149521996	158	121	33	4	0,8703	0,3045	0,2089	0,2258

Distância Genética (F_{st}) e ancestralidade.

Foram utilizados 17 loci para os cálculos de distância genética (F_{st}) e ancestralidade. As frequências alélicas do locus rs1610941 disponíveis são exclusivas da população brasileira do interior de São Paulo (Santos, 2014) e da população britânica (disponível no banco de dados UK 40K) segundo a base de pesquisa EnsemblGenome Browser 93, por isso não foram consideradas nos cálculos de distância genética (F_{st}) e ancestralidade.

Contudo, para as análises de parâmetros forenses e informatividade em casos de gestantes e quimerismo pós TMO, este locus foi considerado, pois não há comparação entre frequências de alelos ou genótipos de outras populações nestas análises.

Nas análises de distância genética e ancestralidade foram consideradas para comparação cinco superpopulações referenciadas no *dbSNP* do *NCBI*, disponibilizadas pelo projeto 1000 genomas: Africanas (AFR), americana (AMR), leste asiática (EAS), europeia (EUR) e sul asiática (SAS).

A análise de distância genética (F_{st}) entre as populações mostrou uma menor distância entre a população do presente estudo (BSP) com americana (AMR), com valor de 0,0191 e entre BSP e EUR (0,0195) e uma maior distância entre AFR e EAS

(0,2484) (Tabela 10).

Tabela 10 - Distâncias genéticas (Fst) entre as populações BSP (População do interior de São Paulo), AMR, EUR, SAS e EAS.

Popula- ção	SAS	EAS	EUR	AFR	AMR
EAS	0,0696				
EUR	0,0391	0,0941			
AFR	0,1873	0,2484	0,1791		
AMR	0,0252	0,0549	0,0191	0,1589	
BSP	0,0450	0,1108	0,0195	0,0960	0,0191

As distâncias genéticas também foram representadas em um gráfico Multidimensional (Figura 7). Nesta figura, foi possível representar os dados mostrados na tabela 10, onde a BSP mostrou ser mais próxima geneticamente das populações EUR, AMR e SAS.

Figura 7. Gráfico de escala Multidimensional realizado com as distâncias genéticas (Fst) entre a nossa população (BSP) com populações retiradas do projeto 1000 genomas.

Em suma, a BSP apresentou menor distância genética com AMR e EUR. A SAS possui a terceira menor distância genética (F_{st}) com a BSP que pode ser explicada por recentes migrações de orientais para o estado de São Paulo.

As diferenças entre frequências alélicas podem indicar que os loci sejam AIM, ou seja, loci que sejam capazes de estimar ancestralidade.

As diferenças entre frequências alélicas nas populações foram maiores que 30% (Tabela 11), para os loci rs148112183 (AFR/EUR); rs66501513 (AFR/EAS); rs138730495 (AFR/EAS); rs55870141 (EAS/EUR); rs151001596 (SAS/EUR); rs67718373 (AMR/AFR); rs113482469 (EAS/SAS); rs34656661 (EAS/AFR); rs141059351 (AFR/EAS); rs33941885 (AFR/EAS) e rs149521996 (EUR/EAS).

Tabela 11 – Frequências do alelo inserção e diferenças alélicas para os 18 loci em uma amostra do interior de São Paulo e em populações de referência retiradas dos 1000 genomas (AMR, AFR, EAS e EUR). Em azul e vermelho, os loci que foram descritos anteriormente. Em preto, os nove loci descritos neste trabalho. O Locus rs1610941 só consta frequência na População do Interior de São Paulo (BSP), o restante das frequências alélicas não consta no banco de dados.

Locus	BSP	AMR	AFR	EAS	EUR	SAS	Amplitude entre as frequências
rs35523917	0,5067	0,4870	0,4580	0,2660	0,4780	0,4290	0,2210
rs148112183	0,5367	0,522	0,8410	0,643	0,3610	0,3810	0,4800
rs72083748	0,5767	0,5200	0,6760	0,4600	0,5860	0,549	0,2160
rs66501513	0,6400	0,5190	0,9440	0,447	0,4910	0,4820	0,4620
rs11270375	0,7967	0,7520	0,8750	0,7850	0,7640	0,5900	0,2850
rs138730495	0,7800	0,6540	0,874	0,4380	0,7620	0,5890	0,3240
rs55870141	0,5100	0,5230	0,6900	0,7880	0,4740	0,6520	0,3140
rs1610941	0,2733	-	-	-	-	-	-
rs151001596	0,5900	0,5750	0,8230	0,5480	0,4180	0,8250	0,4070
rs67718373	0,3396	0,5190	0,0998	0,4812	0,4195	0,3960	0,4192
rs34656661	0,6753	0,8530	0,2496	0,9980	0,8976	0,9480	0,7484
rs11271404	0,5813	0,4710	0,5640	0,3200	0,5159	0,4410	0,2440
rs201008794	0,5281	0,5430	0,6278	0,4177	0,4284	0,4430	0,2101
rs113482469	0,6938	0,6510	0,7927	0,8899	0,6779	0,5790	0,3109
rs141059351	0,4969	0,3650	0,7254	0,1220	0,4583	0,3190	0,6034
rs67887232	0,6551	0,5980	0,7988	0,5704	0,5099	0,5550	0,2889
rs33941885	0,5367	0,5200	0,6172	0,3165	0,5358	0,5010	0,3007
rs149521996	0,8703	0,6920	0,8336	0,5556	0,9006	0,8320	0,3450

Para o restante dos loci, a menor diferença entre as frequências alélicas foi de 21%. Para locus rs1610941 foi somente descrito a frequência alélica na População do

Interior de São Paulo (BSP). Ainda não há frequência alélica deste locus na literatura. Na análise de proporções de ancestralidade realizado com a população do presente estudo e as populações Americana (AMR), Africana (AFR), Europeia (EUR) e asiática (EAS) foi observado que todas elas se agruparam em 3 grandes clusters (K=3), formados pelas populações EAS AFR e EUR (Tabela 12).

Tabela 12 – Proporções de ancestralidade da BSP (População brasileira de São Paulo) e populações de comparação retiradas do projeto 1000 genomas (AMR, AFR, EUR, EAS e SAS).

População	Cluster de população		
	1 - EAS	2 - AFR	3 - EUR
BSP	0,1720	0,2930	0,5350
AMR	0,3760	0,1240	0,4990
AFR	0,0510	0,8750	0,0750
EUR	0,2110	0,0800	0,7090
EAS	0,8270	0,0340	0,1390
SAS	0,3770	0,0740	0,5490

Na BSP 53,5% dos indivíduos se agruparam no cluster EUR enquanto que 29,3% se agruparam no AFR e 17,2% no EAS.

As proporções de ancestralidade de cada indivíduo, representado por uma barra, podem ser observados na figura 8(a). As populações AFR, EUR e EAS representadas pelas cores verde, azul e vermelha, respectivamente, compuseram os três clusters. Na figura 8(b) o triangulo plot mostra apenas as populações dos três clusters, ou seja, AFR (cinza), EUR (amarelo) e EAS (azul), com a maioria dos indivíduos alocada em cada extremidade do triângulo.

Na figura 8 (c), foi adicionado ao triangulo plot a população do presente estudo (BSP), representada pela cor vermelha. Neste triangulo, é nítido a miscigenação da BSP, com uma maior proporção de indivíduos se agrupando com EUR, seguida de AFR.

Figura 8. Resultados do software STRUCTURE utilizando os 17 loci na BSP e populações da referência retirada do projeto 1000. (a) Gráfico de barras - Fração de ancestralidade por indivíduo e (b) Dispersão de indivíduos entre os três grupos dos clusters: AFR (Africana) em cinza EUR (Europeia) em amarelo; em azul EAS (Leste asiático) em azul. c – Dispersão de indivíduos com a BSP (População do interior do estado de São Paulo) em vermelho.

Parâmetros Forenses

Os parâmetros forenses foram calculados segundo equações constantes no ANEXO 6. Os parâmetros forenses foram calculados utilizando o painel de 18 loci e apenas os indivíduos parentais.

Na tabela 13 são apresentados os cálculos dos parâmetros forenses individuais para cada loci (Poder de informação do Polimorfismo (*PIC*), Probabilidade de discriminação (*PD*), Probabilidade de exclusão (*PE*)) e as probabilidades combinadas de discriminação e exclusão (*PCD* e *PCE*).

Tabela 13 – Parâmetros forenses cálculos com o banco de 18 loci - Valores de *PIC*, *PD* e *PE*, *PCD* e *PCE*.

Locus	<i>PIC</i>	<i>PD</i>	<i>PE</i>
rs35523917	0,3750	0,6250	0,1875
rs148112183	0,3736	0,6236	0,1868
rs72083748	0,3690	0,6189	0,1845
rs66501513	0,3546	0,6031	0,1773
rs11270375	0,2715	0,4905	0,1357
rs138730495	0,2843	0,5097	0,1422
rs55870141	0,3749	0,6249	0,1874
rs1610941	0,3183	0,5578	0,1592
rs151001596	0,3668	0,6165	0,1834
rs67718373	0,3480	0,5953	0,1740
rs34656661	0,3683	0,6181	0,1842
rs11271404	0,3683	0,6181	0,1842
rs201008794	0,3742	0,6242	0,1871
rs113482469	0,3346	0,5790	0,1673
rs141059351	0,3750	0,6250	0,1875
rs67887232	0,3498	0,5975	0,1749
rs33941885	0,3737	0,6236	0,1868
rs149521996	0,2003	0,3752	0,1002
Combinados	0,9999999	0,9659947	

Os valores calculados de *PIC*, *PD* e *PE* permitem avaliar a eficácia dos marcadores moleculares quanto ao seu uso na identificação humana e forense.

Considerando o painel de 18 loci, os valores mais baixos de *PIC* encontrados foram 0,20; 0,27 e 0,28 para os loci rs149521996, rs11270375 e rs138730495 respectivamente.

O restante dos loci apresentou valores de *PIC* elevados, próximo ao valor máximo para um locus bialélico, sendo este valor 0,37.

Os loci rs149521996, rs11270375 e rs138730495 apresentaram os valores mais baixos de *PE*. Os outros 15 loci apresentaram valores entre 0,15 e 0,18, que são próximos aos máximos para um locus bialélico (0,18).

Os menores valores de *PD* foram encontrados para os loci rs149521996, rs11270375 e rs138730495. Os restantes dos loci tiveram valores elevados que variaram entre 0,55 a 0,62.

A probabilidade cumulativa (*PCD* e *PCE*) considerando este painel de 18 loci foram de 0,99 e 0,96, respectivamente.

Informatividade em casos de misturas de DNA

Propomos neste trabalho o cálculo de informatividade dos locos *InDels* em casos de misturas de DNA. Em casos de solução materno-fetal a probabilidade para detectar a linhagem que difere entre a mãe e feto foi calculada como pq . Em casos de quimerismo a probabilidade de diferenciar linhagens entre doador e receptor é definida por $2pq(1-pq)$.

As frequências alélicas encontrados em nossa população foram utilizadas nas equações mencionadas acima. Na Tabela 14 estão dispostos os resultados dos 18 locos individualmente e em conjunto.

Tabela 14 – Calculo da informatividade dos locos *InDels* em casos de misturas materno-fetais e quimerismo pós transplante de medula óssea.

Locus	Informatividade de Gestantes	Informatividade de Quimerismo
rs35523917	0,2500	0,3750
rs148112183	0,2487	0,3736
rs72083748	0,2441	0,3690
rs66501513	0,2304	0,3546
rs11270375	0,1620	0,2715
rs138730495	0,1716	0,2843
rs55870141	0,2499	0,3749
rs1610941	0,1986	0,3183
rs151001596	0,2419	0,3668
rs67718373	0,2243	0,3480
rs34656661	0,2193	0,3424
rs11271404	0,2434	0,3683
rs201008794	0,2492	0,3742
rs113482469	0,2125	0,3346
rs141059351	0,2500	0,3750
rs67887232	0,2260	0,3498
rs33941885	0,2487	0,3737
rs149521996	0,1129	0,2003
Combinado	0,9891272	0,9994851

O locus rs11270375 apresentou o menor valor de informatividade em casos de gestantes e quimerismo (0,11 e 0,16, respectivamente). Os locos rs35523917 e rs141059351 apresentaram valores máximos de 0,25 e 0,37.

O restante dos locos variou entre 0,19 a 0,24 em soluções materno-fetal e 0,20 a 0,37 em casos de quimerismo. Por fim, a probabilidade cumulativa dos locos nos dois casos de misturas de DNA foi de 0,98 e 0,99, respectivamente.

DISCUSSÃO

SELEÇÃO DOS LÓCUS

No presente trabalho, foram selecionados um conjunto de 18 lócus *InDel* que obtiveram uma probabilidade de 99% que pelo menos um seja informativo em casos de quimerismo, já em casos de mistura de DNA fetal na circulação materna, essa probabilidade é de 98% de que pelo menos um deles seja informativo.

Todos os lócus foram validados pelo projeto 1000 genomas, apresentaram uma heterozigose na população global de pelo menos 30% e foram escolhidos considerando a possibilidade do desenho de *primers* alelo – específicos tanto pra deleção quanto inserção, além do tamanho dos amplicons para que fosse possível o agrupamento dos lócus em Multiplex.

Nossos *InDels* possuem amplificação alelo-específica, permitindo-se amplificar apenas o DNA fetal, com isto não ficamos limitados apenas às gestações de sexo masculino como em abordagens tradicionais.

A comparação dos nossos lócus com a Literatura, mais especificamente, a comparação com kits *InDels* comercializados é difícil, pelo fato que os autores não disponibilizam os rs (identificação dos lócus no banco de dados) e o tamanho dos amplicons do painel estudado.

Em comparação aos nossos Lócus, o kit comercial mais utilizado (Investigator DIPplex (Qiagen)) não apresenta amplificação alelo-específica, portanto, nosso trabalho se destaca pela originalidade dos novos lócus selecionados com a presença de *primers* específicos, além de serem validados pelo projeto 1000 genomas, um banco de dados amplo, que torna possível a comparação das frequências alélicas com outras populações.

O Kit Investigator DIPplex (Qiagen) é o mais utilizado em estudos populacionais e já foi utilizado para avaliar as diferenças genéticas em populações da Alemanha, Malásia, Catar (Majid Bashir et al., 2016), Tibet (Guo et al., 2016), Mexicanas (Cortés et al., 2015), Vietnamitas e Nigerianas (Du et al., 2017) Japão, Bangladesh, Indonésia (Nagai et al., 2019) e do Paquistão (Shahzad et al., 2019). Por outro lado, apenas algumas investigações populacionais utilizando este kit foram realizadas em populações Sul- americanas (Saiz et al., 2014) e apenas duas em populacionais brasileiras (Torres et al., 2014 e Palha et al., 2015).

Um painel contendo 30 *InDels* foi desenvolvido para estudos de ancestralidade de populações (Pereira et al., 2012). Dos 30 marcadores estudados, 12 apresentaram heterozigoses abaixo de 30% em uma população do Equador (Zambrano et al., 2019), indicando excesso de homozigotos de um determinado alelo na população e perdendo assim, o poder de diferenciação dos indivíduos.

Diversos outros painéis foram desenvolvidos utilizando marcadores do tipo *InDel* para quantificação de misturas de DNA (Sartuntun et al., 2014; Barret et al., 2017). Entretanto, todos os painéis até então mencionados não possuem amplificação alelo-específica, sendo então limitados em detectar o DNA que esteja menos prevalente na amostra.

Ao comparar os loci selecionados no presente trabalho aos 14 loci propostos por Liu et al. (2017), nota-se que ambos os conjuntos são compostos por loci distintos. Os autores foram os únicos até então que selecionaram loci com *primers* flanqueadores e alelo específicos, entretanto, em quatro loci, a heterozigose foi abaixo de 30%, isto diminui a probabilidade de discriminar indivíduos. Além disso, não há garantias que este painel seja informativo em outras populações, por não estarem disponíveis as frequências alélicas em banco de dados amplo para comparação. Para que se possa estimar a informatividade de um conjunto de loci, é de muito interesse que as frequências alélicas estejam disponíveis em um banco de dados de confiança e abrangente.

Além disso, os kits comerciais de marcadores mencionados acima compõem um painel de multiplex, o que é útil para genotipar de forma rápida e eficiente. Em comparação com nosso painel, não possuímos um grupo de 18 loci projetados para multiplex. Apesar de não termos este agrupamento, nosso painel será o primeiro a ser desenvolvido em uma população brasileira com fins de investigações populacionais e misturas de DNA desbalanceadas.

Portanto, faz-se necessário obter um kit de marcadores que possam quantificar com precisão a fração de DNA fetal tanto em gravidez como em casos de quimerismo pós TMO.

De modo geral, a seleção dos nossos loci é adequada tanto na identificação forense quanto em misturas desbalanceadas de DNA, pois apresentam *primers* flanqueadores e alelo – específicos necessários na identificação Humana.

GENOTIPAGEM

Todos os *primers* flanqueadores tiveram amplificação adequada, isto é, amplificação de acordo com o tamanho dos amplicons.

17 amostras não apresentaram amplificação para o locus rs33941885, isto pode ser explicado pelo congelamento e descongelamento dos *primers* que podem degradar neste processo. Porém, a não amplificação dessas amostras não influenciou nas análises estatísticas, pois ainda assim, foi obtido 150 indivíduos genotipados que participaram das análises estatísticas.

O *primer* alelo específico do locus rs33941885 não obteve amplificação, entretanto, o iniciador reverso que é usado em conjunto ao iniciador inserção está funcionando, estima-se que este *primer* inserção possa ter vindo com algum erro de confecção.

MULTIPLEXAÇÃO

O Multiplex é proposto para reduzir o tempo de genotipagem em análises de estudos populacionais, forenses e nos dois casos de misturas: quimerismo e DNA fetal.

Em casos de quimerismo, a genotipagem é realizada utilizando os *primers* flanqueadores para o conhecimento dos genótipos de doador e receptor, ou seja, para saber em quais locus eles diferem, e, possa aplicar os *primers* alelo específicos na quantificação quimerismo.

Em casos de gestante, o Multiplex será aplicado para conhecer os genótipos maternos, isto é, saber quais *primers* serão informativos e aplicar os *primers* específicos para quantificação do DNA amplificação fetal.

O Multiplex dos nove locus descritos neste trabalho foi escolhido respeitando o tamanho dos amplicons de cada locus. Foram agrupados em triplex para que o alelo inserção e deleção de cada locus ficasse um em baixo do outro, ou seja, sem ter alelos de outro locus intercalando.

FREQUÊNCIAS ALÉLICAS, HETEROZIGOSES E EQ DE H-W

Como visto nos itens acima, há muitos kits de marcadores para identificação Humana, entretanto, há pouca informação das frequências alélicas destes bancos de *InDels* em uma população brasileira.

Todos os locus e sua frequências alélicas foram descritos pela primeira vez na população brasileira, com exceção dos iniciadores flanqueadores de dois locus (rs34656661 e rs67718373) que tiveram sua padronização iniciada por Ng (2015).

Ao comparar as frequências alélicas das populações de referência do projeto 1000 genomas com a nossa população, vimos que as frequências de oito loci (rs35523917, rs148112183, rs66501513, rs55870141, rs151001596, rs34656661, rs201008794, rs67887232) estão mais próximas com a População Americana (AMR), em outros seis loci (rs72083748, rs138730495, rs113482469, rs141059351, rs33941885, rs149521996) com populações Europeias (EUR), um locus (rs11270375) com Leste asiático, um locus (rs11271404) com Africana (AFR) e por fim, um locus (rs67718373) com a população Sul-asiática (SAS).

Na maioria dos loci a Heterozigose observada foi maior que 30%, com exceção do locus rs149521996 que apresentou um valor de 21%. Do ponto de vista da Genética forense, valores altos de Heterozigose são vantajosos, pois apresentam alto poder de discriminação (Pepinski et al., 2013).

Em conclusão, através das frequências alélicas nossa população está mais próxima da população Americana, seguida de Europeia.

DISTÂNCIA GENÉTICA (FST) E ANCESTRALIDADE.

A população do presente estudo (BSP) apresentou baixos valores de distâncias corroborando os resultados de Succi et al., 2016 e Rodrigues 2018, que mostraram que a população do Sudeste do Brasil tem mais proximidade com a população europeia.

Através das distâncias genéticas, nossa população está mais próxima com a AMR, isto pode ser explicado pelo fato desta também possuir miscigenação com Europeus e Africanos. Já a proximidade com SAS também pode ser explicada pela migração no século XIV de Asiáticos que se instalaram em São Paulo. (Pereira et al, 2019)

A maioria dos loci deste trabalho apresentou um diferencial entre frequências alélicas alto, ou seja, em 11 loci as diferenças entre a maior e menor frequência do alelo inserção foi maior que 30%. Portanto, estes loci podem ser considerados AIMS (Marcadores informativos de ancestralidade).

O painel de 17 loci *InDels* desenvolvidos em nosso laboratório foram utilizados para cálculo de ancestralidade, com exceção do locus rs1610941. Esse locus não foi considerado, pois o mesmo não foi validado pelo projeto 1000 genomas.

Nas análises de ancestralidade, os genótipos dos 17 loci da BSP foram comparados com os das populações de referência do projeto 1000 genomas e foram definidos três clusters formados por EUR, AFR e EAS (Figura 8 a -b-).

Como esperado, a população do presente estudo apresentou maior composição

ancestral com europeus (53,5%), seguida de africanos (29,3%). Trabalhos prévios descreveram a composição ancestral da região Sudeste com marcadores *InDels* mostrando uma contribuição européia de 67% e 23% de africana (Moura., et al 2015). A ausência de dados de ameríndios na literatura para os lócus aqui analisados não permitiu a comparação com esse grupo ancestral que faz parte da composição étnica da população brasileira desta forma comparamos os nossos dados com EAS que mostrou uma contribuição de 17,2% com esse grupo. Vale- Silva et al., (2019) utilizaram a população Asiática em uma estimativa de ancestralidade, pois de acordo com os autores a diferença entre os ameríndios e asiáticos para estimar ancestralidade é menor que 5%.

Rodrigues (2018) também realizou um estudo de ancestralidade com marcadores do tipo *InDels* autossômicos com a BSP e este mostrou maior semelhança com europeus, resultados concordantes com os nossos.

Em conclusão, apesar do painel de 17 *InDels* aqui desenvolvido, ser menor quanto ao número de lócus em comparação com os da Literatura, ele foi capaz de diferir indivíduos e estimar proporções de mistura em uma população miscigenada.

PARÂMETROS FORENSES

Atualmente, já foram desenvolvidos kits de *InDels* para identificação humana em populações europeias; asiáticas, chinesas e japonesas. Então aqui propomos um kit de *InDels* para identificação humana em uma população brasileira que possa ser usado como um complemento aos marcadores usualmente utilizados (STRs) em casos forenses.

Os STRs apresentam valores de *PIC*, *PD* e *PE* mais elevados devido a diversidade alélica destes marcadores. Entretanto, os *InDels* possuem vantagens pela sua abundancia pelo genoma e facilidade de análise, o que os torna uteis em análises de casos forenses.

Quando a Heterozigose de um lócus bialélico é máxima (50%), os lócus atingem os valores máximos de *PIC* (Conteúdo de informação polimórfica), *PE* (Probabilidade de exclusão) e *PD* (Probabilidade de discriminação).

PIC, indica o quanto a prole é informativa para dedução dos genótipos parentais, isto é, quão um lócus é polimórfico. Valores de *PIC* entre 0,25 e 0,50 indicam uma informatividade mediana (Botstein *et al.*, 1980), porém o valor máximo para um lócus bialélicos é 0,3750 (Hildebrand *et al.*, 1994).

O locus rs149521996 apresentou os menores valores esperados em cálculos forenses para um locus bialélico. Devido a sua Heterozigose baixa de 0,22, seus valores de *PIC*, *PD* e *PE* são considerados baixos: 0,20 ;0,37 e 0,10 respectivamente. Apesar disso, este locus não será excluído do painel destes marcadores para uso forenses, pois mesmo com valores dos parâmetros forenses baixos, ele ainda contribui nos cálculos combinados de discriminação e exclusão.

Os loci rs141059351; rs35523917 e rs55870141 atingiram os valores máximos de 0,3750 de *PIC*.

O restante dos loci, totalizando 18 apresentaram valores entre 0,2715 a 0,3742. Ainda estão entre 0,25 e 0,50 e, portanto, são considerados com informatividade mediana.

Em comparação com a literatura, um painel de 47 *InDels* em um estudo populacional com um grupo de Tibetanos, Wang et al., 2019 obtiveram valores de *PIC* entre 0,2750 a 0,3750, os 18 loci, com exceção de um locus apresentou valores de *PIC* entre 0,2715 a 0,3750.

Deste modo, os loci aqui estudados se mostraram bastante polimórficos considerando que são marcadores bialélicos.

O *PE*, indica a probabilidade de excluir um indivíduo erroneamente acusado, os loci apresentaram valores entre 0,1357 a 18,75, para 17 loci estudados, valores estes próximos ou igual ao máximo teórico de 0,1875 para loci bialélicos (Jamieson e Taylor, 1997).

O *PD* indica a probabilidade de dois indivíduos aleatoriamente escolhidos apresentarem genótipos diferentes. Foram obtidos valores de *PD* entre 0,5097 a 0,6250, valores estes próximos ao máximo teórico de 0,6250 (Butler, 2005).

O *PCE* é a probabilidade combinada dos loci excluir um indivíduo que não seja pai em casos forenses (Evetts e Weir, 1998). Utilizando os nove loci que foram descritos em trabalhos anteriores ao presente trabalho, obtivemos: *PCE*: 0,8166% e *PDC*: 0,9996. Fica nítido que havia necessidade em completar um painel de *InDels* adequado, ou seja, com maiores probabilidades de exclusão e discriminação.

Com o conjunto de 18 loci os resultados de *PCE* foram de 0,9658 e *PCD*: 0,9999 incluindo o locus rs149521996 que apresentou os menores valores dos parâmetros.

Jin et al., 2019 fizeram uma seleção de 35 *InDels* para desenvolver um novo kit que tenha discriminação elevada em populações chinesas. Em comparação com o painel desenvolvido neste trabalho, com apenas 18 loci atingiu o valor combinado da probabilidade de discriminação igual para um painel acima de 30 loci, que é 0,99.

Torres et al., 2014 obtiveram valores de *PE* baixos (Entre 0,09 a 0,15) em cinco de 30

lôcus estudados, e obtiveram valores de *PCE* de 0,99. O painel proposto aqui, obteve o valor de *PCE* de 0,96 e mesmo sendo um valor mais baixo quando comparado com a literatura, ainda assim, possui uma alta probabilidade de excluir um indivíduo que seja acusado de forma errada em casos de paternidade.

Em suma, o painel de 18 lôcus mesmo com um número pequeno de *InDels* é considerado adequado em análises forenses e pode ser utilizado como complemento aos STRs.

INFORMATIVIDADE NA PESQUISA DE MISTURAS DE DNA

O uso de lôcus do tipo *InDel* em análises de misturas de DNA é útil pois com uso de *primers* alelo-específicos podemos amplificar o DNA de apenas uma linhagem em análises de misturas de DNA entre dois indivíduos.

Devido ao critério de seleção dos lôcus, esperava-se heterozigose entre 30 e 50%, portanto, as equações desenvolvidas neste trabalho foram realizadas considerando Heterozigose média dos lôcus de 40%.

Com exceção do lôcus rs149521996, que apresentou os menores valores de informatividade (0,09 e 0,17), o restante dos lôcus apresentaram valores altos nos dois casos de misturas de DNA.

A baixa informatividade do lôcus rs149521996 é facilmente explicada pela baixa heterozigose encontrada na população brasileira.

Em comparação a outros trabalhos que utilizaram os Locus *InDels* em casos de misturas de DNA, não foi desenvolvida uma equação que estimasse a informatividade destes lôcus na população de estudo.

Está equação é necessária para sabermos quais lôcus serão uteis na quantificação de DNA em uma solução que coexistem linhagens.

Para quantificar a fração de DNA fetal, foi desenvolvido um painel de *InDels* que foram uteis em 65% dos casos em uma população de Singapura (Barret et al., 2012). Já em casos de quimerismo em pacientes pós-TMO, o kit de *InDels* AlleleSEQR Chimerism Assay foi muito utilizado na literatura, obtendo informatividades de 99,9% (Kim et al., 2014).

Em suma, a probabilidade combinada dos lôcus desenvolvidos neste trabalho foi de acordo com o esperado de 0,98 e 0,99 em casos de gestantes e quimerismo. Ou seja, os lôcus apresentam probabilidade altas para serem utilizados na nossa amostra da população brasileira.

CONCLUSÃO

Os nove locos descritos neste trabalho em conjunto com outros nove já descritos, apresentaram probabilidade mínima de 98% e 99% em que pelo menos um locus seja informativo em casos de soluções materno-fetal e quimerismo pós-TMO, respectivamente.

Todos os nove locos descritos neste trabalho, tiveram seus *primers* flanqueadores e genotipados em uma população brasileira.

Todos os *primers* alelo-específicos dos nove locos foram descritos neste trabalho

As frequências alélicas descritas neste trabalho foram descritas pela primeira vez em uma população brasileira.

17 locos do painel foram uteis para estimar ancestralidade na amostra populacional do sudeste do Brasil. Nossa população obteve a maior proporção de ancestralidade com Europeia, seguida de Africana e Sul-asiática.

As frequências alélicas encontradas na população brasileira são semelhantes a populações Americanas (AMR) e Europeias (EUR), isto corrobora com as distâncias genéticas (F_{st}) entre populações, no qual europeia e americana estão mais próximos da nossa população.

No conjunto, o poder de exclusão é menor do que o conjunto de STRs normalmente utilizado para identificação humana. Entretanto, constitui-se em um conjunto confiável e suficientemente informativo para ser empregado como adjuvante de outros kits.

REFERÊNCIAS

- BARRETT, A. N. et al. Digital PCR Analysis of Maternal Plasma for Noninvasive Detection of Sickle Cell Anemia. **Clinical Chemistry**, v. 58, n. 6, p. 1026–1032, jun. 2012.
- BARRETT, A. N. et al. Measurement of fetal fraction in cell-free DNA from maternal plasma using a panel of insertion/deletion polymorphisms. **PLOS ONE**, v. 12, n. 10, p. e0186771, out. 2017.
- BASHIR, M.; HASSAN, N. H. B. Analysis of 30 Biallelic INDEL Markers Using the Investigator DIPplex® Kit. **Methods in Molecular Biology**, n. 1420, p. 135–142, 2016.
- BASTOS-RODRIGUES, L.; PIMENTA, J. R.; PENA, S. D. J. The Genetic Structure of Human Populations Studied Through Short Insertion-Deletion Polymorphisms. **Annals of Human Genetics**, v. 70, n. 5, p. 658–665, set. 2006.
- BOTSTEIN, D. et al. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. **American journal of human genetics**, v. 32, n. 3, p. 314–31, maio 1980.
- BRAZOROTTO, A. S. P. **Quantificação de DNA fetal por marcadores inserção/deleção no plasma de parturientes**. Dissertação (Mestrado) – Faculdade de Medicina de Ribeirão Preto Universidade de São Paulo, São Paulo, 2017.
- BUTLER, J. M. M. **Forensic DNA typing: biology, technology and genetics of STR Markers**. 1. ed. London: Academic Press, 2005.
- CUI, W. et al. An innovative panel containing a set of insertion/deletion loci for individual identification and its forensic efficiency evaluations in Chinese Hui ethnic minority. **Molecular Genetics & Genomic Medicine**, v. 8, n. 2, fev. 2020.
- DAUBER, E. et al. Quantitative PCR of INDELs to measure donor-derived cell-free DNA-a potential method to detect acute rejection in kidney transplantation: a pilot study. **Transplant International**, v. 33, n. 3, p. 298–309, mar. 2020.
- DU, W. et al. Forensic efficiency and genetic variation of 30 InDels in Vietnamese and Nigerian populations. **Oncotarget**, v. 8, n. 51, 24 out. 2017.

- EVETT IW, W. B. **Interpreting DNA evidence: statistical genetics for forensic scientists**. [s.l.] Sinauer, 1998.
- FERREIRA PALHA, T. DE J. B. et al. Population genetic analysis of insertion–deletion polymorphisms in a Brazilian population using the Investigator DIPplex kit. **Forensic Science International: Genetics**, v. 19, p. 10–14, nov. 2015.
- FRANCEZ, P. A. DA C.; RIBEIRO-RODRIGUES, E. M.; DOS SANTOS, S. E. B. Allelic frequencies and statistical data obtained from 48 AIM INDEL loci in an admixed population from the Brazilian Amazon. **Forensic Science International: Genetics**, v. 6, n. 1, p. 132–135, jan. 2012.
- GENDZEKHADZE, K.; GAIDULIS, L.; SENITZER, D. Chimerism Testing by Quantitative PCR Using Indel Markers. In: [s.l: s.n.]. p. 221–237.
- GUO, Y. et al. Population Differentiations and Phylogenetic Analysis of Tibet and Qinghai Tibetan Groups Based on 30 InDel Loci. **DNA and Cell Biology**, v. 35, n. 12, p. 787–794, 2016.
- HILDEBRAND, C. et al. Informativeness of polymorphic DNA markers. In: **The Human Genome Project: deciphering the blueprint of heredity**. California, USA: University ScienceBooks, 1994. p. 100.
- HORINEK, A. et al. Cell-Free Fetal DNA in Maternal Plasma during Physiological Single Male Pregnancies: Methodology Issues and Kinetics. **Fetal Diagnosis and Therapy**, v. 24, n. 1, p. 15–21, 2008.
- HUANG, Y.-T.; VANDERWEELE, T. J.; LIN, X. Joint analysis of SNP and gene expression data in genetic association studies of complex diseases. **The Annals of Applied Statistics**, v. 8, n. 1, p. 352–376, mar. 2014.
- JAMIESON, A.; TAYLOR, S. C. Comparisons of three probability formulae for parentage exclusion. **Animal genetics**, v. 28, n. 6, p. 397–400, dez. 1997.
- JIMÉNEZ-VELASCO, A. et al. Reliable quantification of hematopoietic chimerism after allogeneic transplantation for acute leukemia using amplification by real-time PCR of null alleles and insertion/deletion polymorphisms. **Leukemia**, v. 19, n. 3, p. 336–343, mar. 2005.

- JIN, X. et al. Development of a novel multiplex polymerase chain reaction system for forensic individual identification using insertion/deletion polymorphisms. **ELECTROPHORESIS**, v. 40, n. 12–13, p. 1691–1698, jul. 2019.
- KIM, S. Y. et al. Chimerism Monitoring after Allogeneic Hematopoietic Stem Cell Transplantation Using Quantitative Real-Time PCR of Biallelic Insertion/Deletion Polymorphisms. **The Journal of Molecular Diagnostics**, v. 16, n. 6, p. 679–688, nov. 2014.
- LAN, Q. et al. Distinguishing three distinct biogeographic regions with an in-house developed 39-AIM-InDel panel and further admixture proportion estimation for Uyghurs. **ELECTROPHORESIS**, v. 40, n. 11, p. 1525–1534, jun. 2019.
- LIU, J. et al. A mixture detection method based on separate amplification using primer specific alleles of INDELs-a study based on two person's DNA mixture. **Journal of Forensic and Legal Medicine**, v. 46, p. 30–36, fev. 2017.
- LO, Y. M. D. et al. Presence of fetal DNA in maternal plasma and serum. **The Lancet**, v. 350, n. 9076, p. 485–487, ago. 1997.
- MANOKHINA, I. et al. Review: placental biomarkers for assessing fetal health. **Human Molecular Genetics**, v. 26, n. R2, p. R237–R245, out. 2017.
- MARTIN, A. et al. Can the quantity of cell-free fetal DNA predict preeclampsia: a systematic review. **Prenatal Diagnosis**, p. n/a-n/a, jun. 2014.
- MARTÍNEZ-CORTÉS, G. et al. Genetic structure and forensic parameters of 38 Indels for human identification purposes in eight Mexican populations. **Forensic Science International: Genetics**, v. 17, p. 149–152, jul. 2015.
- MULLANEY, J. M. et al. Small insertions and deletions (INDELs) in human genomes. **Human Molecular Genetics**, v. 19, n. R2, p. R131–R136, out. 2010.
- NACHMAN, M. W.; CROWELL, S. L. Estimate of the mutation rate per nucleotide in humans. **Genetics**, v. 156, n. 1, p. 297–304, set. 2000.
- NAGAI, A. et al. Genetic polymorphisms of 30 insertion/deletion markers in the populations of Japan, Bangladesh, and Indonesia. **Forensic Science International: Genetics Supplement Series**, v. 7, n. 1, p. 866–868, dez. 2019.

- NG, A. M. **Validação de Marcadores Inserção/Deleção para Genotipagem Fetal não Invasiva**. Dissertação (Mestrado) – Faculdade de Medicina de Ribeirão Preto Universidade de São Paulo, São Paulo, 2015.
- PEPINSKI, W. et al. Population genetics of 30 INDELs in populations of Poland and Taiwan. **Molecular Biology Reports**, v. 40, n. 7, p. 4333–4338, jul. 2013.
- PEREIRA, F. DOS S. C. F. et al. A systematic literature review on the European, African and Amerindian genetic ancestry components on Brazilian health outcomes. **Scientific Reports**, v. 9, n. 1, p. 8874, dez. 2019.
- PEREIRA, R. et al. Straightforward Inference of Ancestry and Admixture Proportions through Ancestry-Informative Insertion Deletion Multiplexing. **PLoS ONE**, v. 7, n. 1, p. e29684, jan. 2012.
- RODRIGUES, M. L. DE B. **Parâmetros populacionais e forenses de polimorfismos indel e detecção alelo-específica**. Dissertação (Mestrado) – Faculdade de Medicina de Ribeirão Preto Universidade de São Paulo, São Paulo, 2018.
- SAIZ, M. et al. Allelic frequencies and statistical data from 30 INDEL loci in Uruguayan population. **Forensic Science International: Genetics**, v. 9, p. e27–e29, mar. 2014.
- SANTOS, M. D. DOS. **Marcadores Inserção/Deleção na quantificação de quimerismo hematopoiético**. Tese (Doutorado) – Faculdade de Medicina de Ribeirão Preto Universidade de São Paulo, São Paulo, 2018.
- SANTOS, M. D. **Marcadores Inserção/Deleção na identificação humana pré-natal por método não invasivo**. Dissertação (Mestrado) – Faculdade de Medicina de Ribeirão Preto Universidade de São Paulo, São Paulo, 2014.
- SANTURTÚN, A. et al. Analysis of post-transplant chimerism by using a single amplification reaction of 38 Indel polymorphic loci. **Bone Marrow Transplantation**, v. 49, n. 11, p. 1432–1435, 11 nov. 2014.
- SANTURTÚN, A. et al. Indel analysis by droplet digital PCR: a sensitive method for DNA mixture detection and chimerism analysis. **International Journal of Legal Medicine**, v. 131, n. 1, p. 67–72, jan. 2017.

- SHAHZAD, M. et al. Population genetic data of 30 insertion–deletion markers in Punjabi population of Pakistan. **International Journal of Legal Medicine**, v. 134, n. 2, p. 511–512, mar. 2020.
- SUCCI, I. B. et al. Genetic ancestry of patients with porphyria cutanea tarda in a country with mixed races: a cross-sectional study (Rio de Janeiro - Brazil). **Anais Brasileiros de Dermatologia**, v. 93, n. 1, p. 148–150, fev. 2018.
- TORRES, S. R. R. et al. Population genetic data and forensic parameters of 30 autosomal InDel markers in Santa Catarina State population, Southern Brazil. **Molecular Biology Reports**, v. 41, n. 8, p. 5429–5433, ago. 2014.
- VALLE-SILVA, G. DO et al. Applicability of the SNPforID 52-plex panel for human identification and ancestry evaluation in a Brazilian population sample by next-generation sequencing. **Forensic Science International: Genetics**, v. 40, p. 201–209, maio 2019.
- WAGNER, J. et al. Non-invasive prenatal paternity testing from maternal blood. **International Journal of Legal Medicine**, v. 123, n. 1, p. 75–79, jan. 2009.
- WANG, M. et al. Forensic characteristics of Tibeto-Burman-speaking Tibetans revealed by 50 InDels. **Forensic Science International: Genetics Supplement Series**, v. 7, n. 1, p. 758–759, dez. 2019.
- WEBER, J. L. et al. Human Diallelic Insertion/Deletion Polymorphisms. **The American Journal of Human Genetics**, v. 71, n. 4, p. 854–862, out. 2002.
- WEBER, J. L.; WONG, C. Mutation of human short tandem repeats. **Human molecular genetics**, v. 2, n. 8, p. 1123–8, ago. 1993.
- ZAMBRANO, A. K. et al. The three-hybrid genetic composition of an Ecuadorian population using AIMs-InDels compared with autosomes, mitochondrial DNA and Y chromosome data. **Scientific Reports**, v. 9, n. 1, p. 9247, dez. 2019.

Anexo de 1a 6.

ANEXO 1. EXTRAÇÃO DE DNA PARA AMOSTRAS DE SANGUE E SWAB BUCAL (ADAPTADO DE HIGUCHI 1989)

Reagentes e soluções

Tampão de lise de eritrócitos (Lise I): Tris/HCl 0,01 M pH 7,6; Sacarose 0,32 M; MgCl₂ 5,0 mM; Triton X-100 1%.

Tampão de lise de leucócitos (Lise II): Tris/HCl 0,01 M pH 8,5; KCl 0,05 M; MgCl₂ 2,5 mM; NP-40 (Detergente Nonidet) 1%; Tween 20 – 0,45%.

3. Proteinase K (SIGMA-ALDRICH): 10 mg/mL.

Procedimento

Reagentes de Soluções.

1. TAMPÃO PARA LISE DE ERITRÓCITOS (LISE I)

Tris/HCL 0,01M pH 7,6

Sacarose 0,32M

MgCl₂ 5 mM

Triton X-100 1%

2. TAMPÃO PARA LISE DE LEUCÓCITOS (LISE II)

Tris/HCL 0,01 M pH 8,5

KCL 50 mM

MgCL₂ 2,5 mM

NP-40 0,45%

Tween 20 (ou 80) 0,45%

3. PROTEINASE K

10 mg/ml

PROCEDIMENTOS PARA: Swab Bucal/Líquido Amniótico / Meio de Cultura

1. (a) swab bucal

Transferir 700 μ para um microtubo de 1,5 ml

2. Acrescentar 700 μ do tampão de lise de eritrócitos (lise 1) (independente do material) e homogeneizar bem (não utilizar vortex).

3. Centrifugar a 6000 G por 2 minutos e descartar o sobrenadante.

4. Os procedimentos 1,2 e 3 devem ser repetidos até que o precipitado seja visível.

5. Ressuspender o precipitado em 300 μ de tampão de lise de leucócitos (LISE II). Adicionar 5 μ l de proteinase K (10mg/ml) e deixar por pelo menos 1 hora a 65° C.(ou seguir à temperatura recomendada pelo fabricante).

6. Aquecer a 94 °C por 10 minutos para inativar a proteinase K.

7. Estocar a -20 °C

ANEXO 2. REAÇÃO EM CADEIA DA POLIMERASE (PCR – POLIMERASE CHAIN REACTION)

Reagentes e soluções

- Tampão de reação 10X livre de MgCl₂: Tris/HCl 75 mM pH 9,0; KCl 50 mM; (NH₄)₂SO₄ 20 mM (Biotools-B&M Labs, SA).
- MgCl₂: 50mM (Biotools -B&M Labs, SA).
- Solução de *primers*: solução trabalho a 2,5 µM, sendo 10 µL do *primer* direto, 10 µL do *primer* inverso e 180 µL de água destilada.
- dNTP: 20 mM cada (Biotools -B&M Labs, SA).
- DNA polimerase (Taq): 5 U/µL (Biotools -B&M Labs, SA).
- Água destilada ultrafiltrada (MilliQ - Millipore Corporation).

Procedimento

Cada tubo de reação teve volume final de 25 µL, sendo 4 µL de DNA genômico previamente extraído (aproximadamente 200 ng/µL) e 21 µL do mix de reação. O mix é preparado pela adição de 2,5 µL de tampão de reação; 1 µL de MgCl₂; 0,25 µL de dNTP; 0,1 µL de Taq e a quantidade da solução de *primer* padronizada. Para completar o volume final de 21 µL para o mix de reação, adiciona-se a quantidade necessária de água. Após a adição de 4 µL de DNA genômico de cada amostra em seu microtubo de 0,5 mL correspondente, adicionou-se os 21 µL do mix de reação. Para cada reação houve um controle negativo, com a adição de 4 µL de água destilada ao invés de DNA.

ANEXO 3. ELETROFORESE

Reagentes e soluções

- Solução de acrilamida/bisacrilamida (29:1): 29 g de acrilamida (PM= 71,08); 1 g de bisacrilamida (N,N'-metileno-bis-acrilamida; PM= 154,17), 100 mL de água destilada.
- Solução de EDTA 0,5 M pH 8,0: 186 g de EDTA (PM = 372,24); 1 L de água destilada; pastilhas de hidróxido de sódio (NaOH) para ajuste do pH. - Tampão TBE 10X 0,9 M pH 8,0: 108 g de Tris (PM = 121,1); 53 g de ácido bórico (PM = 61,83); 40 mL de solução de EDTA 0,5 M pH 8,0; água destilada para completar volume final de 1 L.
- Tampão de corrida TBE 1X 0,9 M pH 8,0: 100 mL de Tampão TBE 10X; 900 mL de água destilada.
- TEMED (N,N,N',N'-tetrametiletenodiamina) (Sigma-Aldrich).
- Solução hipersaturada de persulfato de potássio: 650 mg de persulfato de potássio (PM = 270,32); 6,5 mL de água destilada.
- Tampão de amostra: 900 µL de bromofenol; 900 µL de xilenocianol; 900 µL de TBE 10X; 4,5 mL de Ficol 30% diluído em água; 1,8 mL de EDTA 0,5 M pH 8,0; 3,6 g de sacarose (completamente diluída na solução final).
- Mix para gel não-desnaturante 10%: 6,66 mL de solução acrilamida/bisacrilamida; 1,4 mL de glicerol; 9,7 mL de água destilada; 2,0 mL de TBE 10%.
- Padrão de peso molecular 100 pb (Promega): 5 µL de solução estoque do padrão de peso molecular 500 µg/µL; 150 µL de água destilada; 50 µL de tampão de amostra.

ANEXO 4. COLORAÇÃO E SECAGEM DE GEL

Reagentes e soluções

Solução fixadora: 166,67 mL de álcool etílico; 6,67 mL de ácido acético glacial e 826,66 mL de H₂O.

Solução reveladora: 22,5g de NaOH (PM= 40,00); 1L de H₂O. No momento da coloração, adicionar 2 mL de formaldeído (PM = 30,05) para cada 100 mL de solução reveladora utilizada.

Solução de nitrato de prata: 100 g de nitrato de prata; 1 L de água (armazenada sob abrigo da luz).

Procedimento

Retirando os espaçadores e separando cuidadosamente as placas de vidro com ajuda de uma espátula, remove-se o gel e mergulha em 100 mL da solução fixadora, adicionando 4 mL da solução de nitrato de prata. Após manter sob agitação por 5 minutos, a solução é descartada e lava-se rapidamente com 100 mL de água destilada a 60 C. A água é então descartada e adiciona-se 100 mL da solução reveladora levemente aquecida e adicionada de 2 mL de formaldeído. O recipiente é agitado por alguns minutos até aparecimento de bandas nítidas e a solução reveladora é então, descartada. Para bloqueio da reação de revelação adiciona-se 100 mL da solução fixadora. A documentação dos resultados é feita com a secagem do gel forrando-se uma placa de vidro com papel celofane culinário umedecido. Sobre a placa forrada, coloca-se o gel e, por cima, outra folha de papel celofane umedecido, de maneira que não haja bolhas de ar entre as camadas de celofane sobre o vidro e sobre o gel. Deixa o conjunto, em condições ambientes, até secar completamente, cerca de 2 dias.

ANEXO 5. ÍNDICES DE FREQUÊNCIAS GENOTÍPICAS E ALÉLICAS

As frequências genotípicas (p_{ii}) e alélicas (p_i) foram calculadas pelo *software Genepop 4.7.0* (Rousset, 2008) encontrado em <http://kimura.univ-montp2.fr/~rousset/Genepop.htm>.

$$p_i = \frac{2n_{ii} + n_{id}}{2n}$$

$$P_{ii} = \frac{n_{ii}}{n}$$

Em que:

p_i = frequência do alelo inserção

P_{ii} = frequência do genótipo inserção/inserção;

n_{ii} = número observado de homozigotos para o alelo inserção;

n_{id} = número observado de heterozigotos

n = número de indivíduos

Aderência ao Equilíbrio de Hardy- Weinberg

No Equilíbrio de Hardy-Weinberg (EHW), espera-se que as frequências genotípicas possam ser estimadas a partir do binômio expandido:

$$(p + q)^2 = p^2 + 2pq + q^2$$

Em que:

p = frequência do alelo inserção;

q = frequência do alelo deleção;

p^2 = frequência esperada de homozigotos inserção/inserção;

$2pq$ = frequência esperada de heterozigotos;

q^2 = frequência esperada de homozigotos deleção/deleção.

Heterozigose

A heterozigose esperada H_e estima a diversidade gênica intra-populacional, definida como a probabilidade de dois alelos escolhidos aleatoriamente em uma população serem diferentes (Weir, 1996). A heterozigose observada é estimada por contagem direta. Ambas podem ser calculadas segundo as equações:

$$H_o = \frac{n_{id}}{n} \qquad H_e = 2pq$$

Em que:

n_{id} = número observado de heterozigotos;

n = número total de indivíduos analisados;

p = frequência do alelo inserção;

q = frequência do alelo deleção.

ANEXO 6. Parâmetros forenses

Conteúdo de informação do polimorfismo

O conteúdo de informação do polimorfismo pode ser calculado de acordo com a fórmula expandida (Botstein *et al.*, 1980):

PIC é definido como uma fração de prole informativa quanto aos genótipos parentais em um cruzamento, ou em outras palavras, é a probabilidade de uma dada prole permitir a dedução dos genótipos parentais (Botstein *et al.*, 1980, Hildebrand *et al.*, 1994)

$$PIC = 1 - \sum_{i=1}^a p_i^2 - \left(\sum_{i=1}^a p_i^2 \right)^2 + \sum_{i=1}^a p_i^4$$

Em que:

a = número de alelos;

p_i = frequência do i -ésimo alelo

Quando *PIC* é inferior a 0,25, o polimorfismo é considerado pouco informativo. Valores entre 0,25 e 0,5 indicam informatividade mediana e acima de 0,5 o polimorfismo pode ser considerado muito informativo.

Poder de discriminação

$$PD = 1 - \sum_{i=1}^n x_i^2$$

Em que:

n = número de genótipos possíveis;

x_i = frequência do i -ésimo genótipo.

$PD = 1 - P_i$, onde $\sum$ (Probabilidade de Identidade) é a probabilidade de que dois indivíduos selecionados aleatoriamente tenham genótipos idênticos no mesmo locus (Butler, 2005).

Podemos calcular, para um conjunto de marcadores, o Poder Cumulativo de Discriminação (*PCD*):

$$PCD = 1 - \prod_{i=1}^n (1 - PD_i)$$

Em que:

n = número de loci avaliados;

PD_i = PD do i -ésimo alelo

Poder de exclusão

PE é a probabilidade de um locus excluir um indivíduo não relacionado à paternidade (Evetts and Weir, 1998)

$$PE = \sum_{i=1}^n p_i(1 - p_i)^2 - \frac{1}{2} \sum_{i=1}^n \sum_{j \neq i}^n p_i^2 p_j^2 (4 - 3p_i - 3p_j)$$

Em que:

n = número de alelos;

p_i = frequência do i -ésimo alelo;

p_j = frequência do j -ésimo alelo diferente de i -ésimo

O valor de PE será aumentado ao utilizar um conjunto de loci, sendo aqui identificado como Probabilidade Cumulativa de Exclusão (PCE) e corresponde à probabilidade de exclusão em ao menos um locus do conjunto (Evetts and Weir, 1998).

$$PCE = 1 - \prod_{i=1}^n (1 - PE_i)$$

Em que:

n = número de loci avaliados;

PE_i = PE do i -ésimo alelo

APÊNDICE

Genótipos dos 80 trios (Mãe-Filho-Pai) para nove marcadores *InDel*..

		Locus								
ID		rs67718373	rs11271404	rs201008794	rs113482469	rs34656661	rs141059351	rs67887232	rs149521996	rs33941885
4711	mãe	id	id	ii	id	ii	dd	ii	id	dd
4711	filho	id	ii	ii	id	ii	dd	ii	id	dd
4711	pai	dd	id	id	ii	ii	dd	ii	na	dd
4714	mãe	dd	id	id	ii	ii	id	ii	ii	ii
4714	filho	dd	id	id	id	ii	id	ii	ii	ii
4714	pai	dd	id	dd	id	ii	dd	id	ii	id
4717	mãe	dd	id	ii	dd	ii	id	id	ii	dd
4717	filho	dd	id	ii	dd	ii	dd	ii	ii	id
4717	pai	id	id	id	id	ii	id	ii	ii	ii
4723	mãe	id	ii	dd	ii	id	ii	dd	ii	ii
4723	filho	id	ii	dd	ii	id	ii	id	ii	ii
4723	pai	dd	ii	dd	id	dd	ii	ii	ii	ii
4725	mãe	dd	dd	id	ii	dd	id	ii	ii	dd
4725	filho	dd	dd	ii	ii	dd	id	id	id	dd
4725	pai	dd	id	ii	ii	id	dd	id	id	id
4728	mãe	dd	id	dd	ii	id	id	ii	id	id
4728	filho	dd	id	id	ii	id	id	ii	id	id
4728	pai	dd	id	id	ii	ii	id	id	id	dd
4731	mãe	id	ii	id	ii	ii	dd	id	id	ii
4731	filho	id	ii	id	id	id	id	dd	ii	id
4731	pai	dd	id	dd	id	dd	id	id	ii	id
4763	mãe	ii	dd	dd	ii	ii	dd	id	ii	id
4763	filho	ii	id	dd	ii	ii	id	id	ii	id
4763	pai	id	ii	dd	id	id	id	dd	ii	ii
4764	mãe	id	id	ii	ii	id	ii	ii	ii	id
4764	filho	id	id	ii	ii	id	id	id	ii	dd

4764	pai	dd	id	id	ii	ii	dd	id	ii	id
4766	mãe	dd	dd	ii	ii	dd	ii	ii	ii	id
4766	filho	dd	id	ii	id	na	ii	id	ii	ii
4766	pai	dd	id	ii	id	na	id	dd	ii	id
4769	mãe	dd	id	ii	ii	ii	d	id	ii	dd
4769	filho	dd	dd	id	id	ii	d	ii	ii	dd
4769	pai	dd	id	dd	id	id	id	ii	ii	id
4771	mãe	id	ii	id	id	ii	ii	id	id	ii
4771	filho	id	ii	ii	ii	id	id	id	ii	id
4771	pai	id	ii	id	id	id	dd	ii	ii	dd
4773	mãe	id	id	ii	ii	id	ii	ii	ii	id
4773	filho	id	dd	ii	ii	ii	id	ii	ii	id
4773	pai	id	dd	ii	id	ii	id	id	ii	ii
4774	mãe	id	id	id	ii	ii	id	id	ii	ii
4774	filho	id	id	id	id	ii	id	id	ii	na
4774	pai	ii	ii	ii	id	id	ii	ii	ii	dd
4775	mãe	id	ii	ii	ii	ii	id	ii	id	ii
4775	filho	id	ii	id	ii	ii	dd	ii	ii	ii
4775	pai	id	id	id	ii	id	id	id	ii	id
4776	mãe	id	id	dd	id	ii	id	id	ii	dd
4776	filho	dd	id	id	dd	ii	ii	dd	id	na
4776	pai	id	dd	ii	id	ii	id	dd	id	id
4778	mãe	id	id	ii	dd	id	id	ii	ii	id
4778	filho	id	ii	ii	dd	id	id	ii	ii	dd
4778	pai	id	ii	ii	id	ii	id	ii	ii	id
4779	mãe	dd	id	id	ii	ii	id	ii	ii	id
4779	filho	id	id	id	id	id	ii	ii	ii	ii
4779	pai	id	ii	dd	id	dd	ii	ii	ii	id
4800	mãe	ii	ii	ii	ii	dd	dd	id	ii	id
4800	filho	id	ii	id	ii	id	dd	ii	ii	ii
4800	pai	dd	ii	dd	ii	ii	id	ii	ii	ii
4801	mãe	ii	dd	ii	ii	id	id	id	ii	na
4801	filho	ii	dd	id	ii	ii	id	ii	ii	id
4801	pai	id	dd	id	id	ii	ii	ii	ii	ii

4851	mãe	dd	id	ii	ii	ii	dd	id	ii	ii
4851	filho	dd	id	ii	ii	ii	dd	id	ii	ii
4851	pai	dd	id	id	ii	ii	dd	ii	ii	id
4860	mãe	dd	id	ii	id	ii	ii	id	id	dd
4860	filho	dd	id	ii	id	id	ii	dd	id	id
4860	pai	id	id	ii	id	dd	ii	id	ii	id
4862	mãe	id	id	id	ii	id	dd	dd	ii	ii
4862	filho	id	id	id	ii	id	dd	id	ii	ii
4862	pai	dd	dd	ii	ii	id	id	ii	ii	ii
4877	mãe	ii	id	id	dd	ii	ii	ii	ii	dd
4877	filho	ii	ii	dd	id	ii	id	ii	ii	dd
4877	pai	id	ii	id	ii	ii	id	ii	ii	dd
4906	mãe	dd	id	dd	id	dd	id	id	ii	id
4906	filho	dd	dd	dd	ii	id	id	id	ii	id
4906	pai	dd	dd	dd	id	ii	dd	id	ii	dd
4910	mãe	id	ii	dd	id	id	ii	ii	ii	id
4910	filho	dd	ii	dd	id	id	ii	ii	ii	ii
4910	pai	dd	id	id	ii	ii	id	ii	ii	ii
4957	mãe	id	dd	id	ii	dd	id	id	ii	id
4957	filho	ii	dd	id	ii	dd	id	id	ii	dd
4957	pai	id	dd	id	id	id	id	id	ii	id
4960	mãe	id	dd	id	id	ii	id	dd	ii	id
4960	filho	id	id	ii	id	ii	dd	id	ii	dd
4960	pai	id	ii	id	id	ii	dd	id	ii	id
4969	mãe	dd	ii	ii	ii	ii	dd	ii	ii	id
4969	filho	dd	id	ii	id	ii	id	ii	ii	ii
4969	pai	id	id	ii	id	id	ii	ii	ii	ii
4973	mãe	ii	id	dd	ii	id	dd	ii	ii	ii
4973	filho	id	ii	id	id	ii	id	id	ii	ii
4973	pai	id	ii	id	dd	ii	ii	na	na	na
4975	mãe	dd	dd	dd	id	id	id	dd	ii	id
4975	filho	dd	id	dd	id	dd	id	id	ii	dd
4975	pai	id	ii	dd	dd	dd	ii	ii	ii	dd
4978	mãe	dd	dd	ii	id	dd	id	id	ii	id

4978	filho	dd	id	id	id	id	id	id	id	id
4978	pai	id	ii	dd	ii	ii	id	dd	id	dd
4981	mãe	ii	id	id	id	ii	id	id	id	id
4981	filho	id	ii	id	ii	id	ii	dd	id	id
4981	pai	dd	ii	ii	ii	id	id	id	ii	ii
4985	mãe	dd	id	id	id	id	dd	id	ii	na
4985	filho	dd	na	id	ii	ii	id	dd	ii	id
4985	pai	dd	id	ii	ii	ii	ii	dd	ii	ii
4986	mãe	id	ii	dd	ii	id	id	ii	ii	id
4986	filho	id	ii	dd	id	ii	ii	id	ii	id
4986	pai	id	id	dd	id	ii	id	id	ii	id
4994	mãe	dd	id	id	id	dd	id	ii	ii	id
4994	filho	dd	id	dd	dd	id	id	ii	ii	dd
4994	pai	dd	id	dd	id	ii	id	ii	ii	id
5010	mãe	id	id	id	id	id	id	ii	ii	ii
5010	filho	id	ii	ii	dd	ii	ii	id	ii	ii
5010	pai	id	ii	ii	id	ii	ii	id	ii	ii
5017	mãe	dd	id	ii	id	ii	dd	id	ii	dd
5017	filho	id	dd	id	id	ii	dd	id	ii	na
5017	pai	id	id	id	ii	id	dd	id	ii	id
5023	mãe	ii	id	id	ii	ii	ii	id	ii	dd
5023	filho	ii	id	ii	ii	id	id	id	id	na
5023	pai	ii	dd	id	ii	id	id	dd	id	ii
5024	mãe	ii	ii	ii	ii	id	id	ii	ii	dd
5024	filho	id	id	id	id	id	dd	id	ii	dd
5024	pai	id	id	id	id	id	id	id	ii	id
5025	mãe	dd	ii	id	ii	id	dd	ii	ii	id
5025	filho	dd	id	dd	ii	ii	dd	na	ii	id
5025	pai	dd	id	id	ii	ii	dd	dd	ii	ii
5028	mãe	id	ii	id	ii	ii	id	ii	ii	id
5028	filho	dd	ii	id	ii	ii	id	id	id	dd
5028	pai	dd	ii	id	id	id	id	id	dd	id
5029	mãe	id	dd	ii	ii	ii	dd	dd	ii	id
5029	filho	ii	id	id	ii	id	id	dd	ii	id

5029	pai	id	id	dd	ii	id	ii	id	id	id
5030	mãe	id	dd	ii	ii	ii	dd	id	id	id
5030	filho	ii	dd	ii	id	ii	id	ii	ii	dd
5030	pai	ii	id	id	dd	id	id	ii	ii	id
5031	mãe	ii	ii	id	ii	ii	ii	ii	ii	id
5031	filho	id	id	dd	ii	id	id	id	id	ii
5031	pai	dd	dd	dd	ii	id	dd	dd	id	id
5032	mãe	dd	ii	ii	ii	dd	ii	ii	id	id
5032	filho	dd	id	ii	ii	dd	id	id	dd	id
5032	pai	dd	dd	ii	ii	dd	id	id	dd	ii
5034	mãe	ii	id	ii	id	id	id	ii	ii	id
5034	filho	id	ii	ii	ii	id	id	na	ii	ii
5034	pai	dd	id	ii	ii	dd	id	ii	ii	id
5036	mãe	id	id	id	dd	id	id	dd	ii	ii
5036	filho	ii	id	id	id	id	id	dd	ii	na
5036	pai	ii	ii	id	ii	ii	id	id	ii	ii
5038	mãe	id	dd	dd	id	dd	dd	id	id	ii
5038	filho	ii	dd	dd	dd	dd	dd	id	ii	id
5038	pai	id	dd	dd	id	dd	dd	id	ii	na
5045	mãe	id	id	id	dd	ii	id	id	id	na
5045	filho	id	id	id	id	id	id	id	dd	id
5045	pai	dd	id	dd	ii	id	ii	id	id	id
5049	mãe	dd	ii	ii	ii	ii	id	na	ii	ii
5049	filho	dd	ii	ii	ii	ii	id	ii	ii	id
5049	pai	dd	ii	ii	id	ii	dd	ii	ii	id
5050	mãe	id	id	id	ii	id	id	ii	id	dd
5050	filho	id	id	dd	ii	id	ii	id	id	dd
5050	pai	dd	id	dd	ii	ii	ii	id	ii	id
5051	mãe	id	dd	id	ii	id	id	ii	ii	ii
5051	filho	dd	dd	id	ii	ii	ii	ii	ii	ii
5051	pai	na	dd	id	ii	na	id	ii	ii	id
5053	mãe	dd	ii	dd	ii	id	id	ii	ii	ii
5053	filho	dd	ii	id	ii	ii	dd	na	ii	ii
5053	pai	dd	ii	id	id	ii	id	ii	ii	id

5054	mãe	dd	id	ii	id	ii	dd	dd	ii	id
5054	filho	id	id	ii	id	id	dd	id	id	ii
5054	pai	id	id	id	ii	dd	id	ii	id	id
5055	mãe	id	id	id	id	ii	dd	id	id	id
5055	filho	id	dd	id	ii	ii	dd	dd	ii	ii
5055	pai	id	id	dd	id	ii	id	id	ii	id
5058	mãe	id	id	ii	id	ii	ii	id	id	id
5058	filho	dd	ii	ii	id	ii	id	id	id	id
5058	pai	id	ii	ii	ii	ii	dd	ii	ii	dd
5059	mãe	id	id	id	ii	id	id	id	id	na
5059	filho	id	id	dd	id	ii	id	ii	ii	ii
5059	pai	id	id	id	id	ii	dd	id	ii	ii
5060	mãe	dd	ii	ii	id	ii	id	ii	ii	id
5060	filho	dd	na	id	ii	id	id	id	ii	dd
5060	pai	dd	ii	dd	ii	id	ii	id	ii	dd
5062	mãe	dd	ii	id	ii	ii	id	ii	id	na
5062	filho	id	ii	id	ii	ii	dd	ii	ii	id
5062	pai	id	ii	dd	ii	ii	id	id	ii	na
5065	mãe	id	ii	ii	ii	id	id	ii	ii	id
5065	filho	id	ii	ii	id	ii	id	ii	ii	id
5065	pai	id	ii	ii	dd	id	dd	ii	ii	dd
5069	mãe	dd	ii	id	ii	dd	ii	ii	ii	id
5069	filho	id	ii	dd	ii	id	ii	ii	id	id
5069	pai	ii	ii	dd	id	id	id	id	id	id
5070	mãe	dd	id	id	ii	id	ii	id	ii	id
5070	filho	dd	id	id	ii	dd	id	id	ii	id
5070	pai	dd	ii	id	ii	id	id	id	ii	id
5072	mãe	dd	ii	id	ii	ii	id	id	ii	id
5072	filho	id	ii	ii	ii	ii	id	ii	ii	dd
5072	pai	id	id	id	id	ii	ii	id	ii	id
5076	mãe	dd	id	id	dd	na	id	ii	ii	na
5076	filho	id	id	id	id	na	id	ii	ii	na
5076	pai	id	id	ii	id	na	ii	id	ii	na
5078	mãe	dd	dd	id	ii	ii	dd	dd	ii	id

5078	filho	dd	id	dd	id	ii	id	id	ii	id
5078	pai	dd	id	id	id	ii	ii	ii	ii	dd
5079	mãe	id	id	ii	id	ii	id	id	id	dd
5079	filho	id	dd	id	ii	ii	dd	dd	ii	id
5079	pai	id	dd	id	ii	id	id	id	ii	ii
5080	mãe	dd	ii	dd	ii	ii	id	ii	ii	dd
5080	filho	dd	ii	dd	ii	ii	id	id	ii	id
5080	pai	id	id	id	ii	ii	id	id	ii	id
5082	mãe	id	ii	id	id	id	id	ii	ii	id
5082	filho	id	id	id	ii	dd	id	na	ii	dd
5082	pai	ii	dd	id	ii	id	ii	id	ii	dd
5083	mãe	id	ii	ii	id	ii	id	dd	ii	ii
5083	filho	dd	ii	id	dd	ii	ii	id	ii	ii
5083	pai	dd	ii	dd	id	id	id	ii	ii	id
5084	mãe	dd	id	dd	ii	id	id	id	ii	dd
5084	filho	id	ii	dd	id	dd	dd	id	id	dd
5084	pai	id	ii	dd	dd	id	id	id	id	id
5087	mãe	id	dd	ii	id	na	dd	ii	ii	id
5087	filho	id	dd	ii	dd	na	id	ii	ii	na
5087	pai	id	id	ii	id	id	ii	ii	ii	ii
5088	mãe	id	id	id	ii	id	dd	ii	ii	dd
5088	filho	id	id	id	id	id	id	id	ii	id
5088	pai	id	id	id	id	ii	ii	id	ii	ii
5089	mãe	dd	id	id	id	id	id	ii	id	dd
5089	filho	dd	id	id	id	ii	id	id	ii	dd
5089	pai	id	dd	id	id	ii	ii	dd	ii	dd
5090	mãe	dd	ii	dd	id	ii	id	id	ii	id
5090	filho	dd	id	dd	id	id	dd	ii	id	id
5090	pai	id	id	id	dd	id	id	ii	id	id
5091	mãe	dd	ii	id	dd	dd	dd	id	dd	ii
5091	filho	dd	id	id	id	na	id	id	id	na
5091	pai	dd	id	id	id	id	id	id	id	id
5094	mãe	dd	ii	id	dd	ii	dd	id	ii	ii
5094	filho	id	id	id	id	ii	id	id	ii	id

5094	pai	id	id	dd	ii	ii	ii	dd	ii	id
5095	mãe	dd	id	ii	id	id	dd	ii	id	ii
5095	filho	dd	id	id	id	id	id	ii	id	id
5095	pai	dd	id	dd	ii	id	ii	ii	dd	id
5097	mãe	ii	ii	id	id	id	id	dd	id	id
5097	filho	id	id	id	dd	ii	dd	id	ii	ii
5097	pai	id	id	id	dd	ii	id	ii	ii	ii
5100	mãe	dd	id	id	dd	na	id	id	ii	id
5100	filho	id	ii	id	id	id	id	id	ii	ii
5100	pai	id	ii	id	ii	id	id	ii	ii	ii