UNIVERSIDADE DE SÃO PAULO
FACULDADE DE ODONTOLOGIA DE BAURU

PAULO RENATO DE PAULA FREDERICO

**Análise de frequências alélicas de 15 marcadores STR em alunos da
Faculdade de Odontologia de Bauru**

BAURU

2015

PAULO RENATO DE PAULA FREDERICO

**Análise de frequências alélicas de 15 marcadores STR em alunos da
Faculdade de Odontologia de Bauru**

Tese apresentada à Faculdade de Odontologia de Bauru da Universidade de São Paulo para obtenção do título de Doutor em Ciências Odontológicas Aplicadas, na área de concentração de Estomatologia e Biologia Oral.

Orientador: Prof. Dr. Carlos Ferreira dos Santos

Versão Corrigida

BAURU

2015

Frederico, Paulo Renato de Paula

S872a Análise de frequências alélicas de 15 marcadores STR
em alunos da Faculdade de Odontologia de Bauru / Paulo
Renato de Paula Frederico. -- Bauru, 2015.
98 p. : il. ; 31cm.

Tese. (Doutorado) -- Faculdade de Odontologia de
Bauru. Universidade de São Paulo.

Orientador: Prof. Dr. Carlos Ferreira dos Santos

Nota: a versão original desta tese encontra-se disponível no Serviço de Biblioteca e Documentação da Faculdade de Odontologia de Bauru – FOB/USP.

Autorizo, exclusivamente para fins acadêmicos e científicos, a
reprodução total ou parcial desta dissertação/tese, por processos
fotocopiadores e outros meios eletrônicos.

Assinatura:

Data:

Comitê de Ética da FOB-USP
Protocolo nº: 42919615.4.0000.5417
Data: 22/06/2015

AGRADECIMENTOS

Agradecer é reconhecer a importância de uma pessoa em nossa vida.

Para que alguma conquista seja alcançada, a presença de pessoas que estão dispostas a nos ajudar é imprescindível! Portanto, nada mais justo do que reconhecer a grande importância que essas pessoas fizeram em um caminho que foi construído com a ajuda de muitas mãos.

Inicialmente, gostaria de agradecer ao **Dr. Carlos Ferreira dos Santos**, meu orientador e amigo. Tive a grande honra de conhecê-lo e de ser escolhido para ser seu orientando. Com toda a certeza, você deixará seu legado para todos que têm a oportunidade de conviver com você. Se pudesse descrevê-lo em poucas palavras, as escolhidas seriam competência, retidão e generosidade. Sem essa última qualidade, eu não estaria escrevendo essas linhas. Muito obrigado, querido amigo!

Dra. Elza Araújo Torres, muito obrigado! Tivemos a oportunidade de trabalhar na mesma instituição no passado, mas trabalhando agora nessa pesquisa pude comprovar sua competência acadêmica. Sem sua presença, esse trabalho não existiria. Seu amor pela área da genética me contagiou e me fez ver um mundo novo .

Thiago Dionísio. Tão jovem e tão brilhante. Seu conhecimento e seu companheirismo foram cruciais ao longo da pesquisa. Na reta final, sua liderança orquestrou um belo trabalho em equipe. Muito obrigado!

Adriana Calvo. Sempre disposta a ajudar. A Plataforma Brasil, uma ilustre complicação para todo pesquisador, é algo simples nas suas mãos! Muito obrigado pela ajuda e pela presença sempre proativa.

Daniel Brozowski, o gringo do laboratório. Sempre calado e com a fala mansa, mas disposto a decifrar programas que seriam cruciais para a análise dos resultados. Muito obrigado!

Cynthia Bachir Moises e Wesley Tsutsumida, diretores executivos, CEO – Chief Executive Officer - da empresa **Thermo Fisher Brasil**, que não pouparam esforços no momento das análises dos resultados obtidos ao longo da pesquisa.

Alessandro Alves Pereira, aluno do programa de pós-graduação da ESALQ, Piracicaba e Dra. **Maria Imaculada Zucchi**, pesquisadora da da Universidade Estadual de Campinas e professora credenciada no programa de pós-graduação da ESALQ, que tiveram a grande disposição de compartilhar seus conhecimentos sobre programas de análises estatísticas de dados de Genética de População. Muito obrigado!

Gostaria de agradecer aos **funcionários do laboratório de Farmacologia**, especialmente à **Viviane A. Parisi Santos**, pela valiosa ajuda na coleta de saliva dos alunos e na extração de DNA e a **todos os funcionários da FOB/USP**. Nos anos que estive nessa instituição, percebi que o clima organizacional propicia uma fluidez na equipe que parece trabalhar como uma única grande engrenagem. Obrigado a todos por facilitarem minha vida acadêmica.

Agradeço a minha mãe, **Zuza**, e a meu padrasto, **Clodoaldo**, por estarem sempre ao meu lado não me deixando em momento algum desistir dessa empreitada e cuidando dos meus filhos nos momentos em que tinha de me dedicar aos estudos.

Obrigado **Leda e Ricardo** que sempre estiveram ao meu lado!

Obrigado **Valdete** por cuidar da minha casa e da minha família!

Finalmente, gostaria de fazer um agradecimento especial a três pessoas que completam a minha vida: **Renata, Leonardo e Isabela**. Obrigado por me fazerem entender o verdadeiro sentido do amor e por estarem comigo em todos os momentos. A presença de vocês dá sentido a minha constante busca em ser uma pessoa melhor!

RESUMO

RESUMO

Entre as muitas aplicações das tecnologias de identificação biológica humana, estão as finalidades forenses. O objetivo desta pesquisa foi verificar frequências alélicas de “Short Tandem Repeat” (STR) e os parâmetros estatísticos de interesse em genética de populações e forense para desenvolver o primeiro banco de dados populacional de DNA na Faculdade de Odontologia de Bauru, Universidade de São Paulo, (FOB/USP) para futuros usos forenses. Frequências alélicas de 15 locos autossômicos e do marcador de gênero amelogenina foram determinadas utilizando amostras de 200 µL de saliva doados por 296 alunos de graduação da FOB/USP, com idade ≥ 18 anos, após aprovação ética. Os testes laboratoriais foram feitos com kits comerciais. Resultados e parâmetros estatísticos foram obtidos por meio de programas clássicos: GeneMapper-ID-X, MS Excel 2002 versão 10.6871.6870, GenAlEx 6.5 e Arlequin 3.5, comparando quatro populações (brasileira, portuguesa, norte-americana e a população deste estudo). Os locos mais polimórficos foram D18S51 (17 alelos) e FGA (15 alelos), seguidos pelo D21S11 (13 alelos) e os menos polimórficos foram D16S539 e TH01 (7 alelos cada). A análise comparativa com amostra da população brasileira proveniente de estudos anteriores ($n > 100.000$) pelo teste goodness of fit χ^2 index não mostrou diferenças significativas entre estes grupos ($p = 0,9999$). Outros parâmetros estatísticos foram calculados comparando as populações: local (deste estudo), portuguesa e norte-americana. A análise de variância molecular (AMOVA) entre as três populações, entre as pessoas da mesma população e para cada pessoa de cada população mostrou que existe uma elevada variância individual (99%), que esta variância é mantida uniformemente entre as pessoas da mesma amostra/região (1%) e entre as três populações estudadas (0%). O estudo confirmou o elevado grau de polimorfismo e a alta heterozigosidade (96,5%) da população. Houve diferença significativa quanto ao gênero (79,7% mulheres) quando comparado à população brasileira em geral (50,4%), explicada pelas características do corpo discente da FOB/USP composto por 80,6% de pessoas do gênero feminino. Interessante foi a observação de uma microvariante alélica no loco D18S51, fora da escada padrão e da escala de abrangência do kit, correspondente ao alelo 29, ainda não definida na base de dados internacional (STRBase, atualizada em 07/08/2015). Esta microvariante deverá ser confirmada por testes familiares e sequenciamento de DNA para verificar a possibilidade de outra ocorrência familiar ou duplicação de nucleotídeos. No futuro, os dados obtidos neste estudo devem ser incorporados ao banco de dados da população brasileira e podem ser considerados como referência genética da população regional, ajudando a elucidar casos forenses. Após a confirmação, a potencial nova microvariante alélica contribuirá para a base de dados internacional STRBase.

Palavras-chave: Identificação humana. Frequência alélica. STR. Banco de dados.

ABSTRACT

ABSTRACT

There are many ways of applying biological human identification technologies, among these are forensic applications. The objective of this study was to verify allele frequencies for 15 autosomal short tandem repeat (STR) loci and develop the first human DNA population database at the Bauru School of Dentistry, University of São Paulo (FOB/USP), for future forensic uses. Allele frequencies for these STR loci and an amelogenin gender marker were determined using 200 µL samples of saliva donated by 296 undergraduates from FOB/USP who were ≥ 18 years old at the time of the sample collect after signing a consent form with ethical approval. For laboratory tests, commercial kits were used. Results and statistical parameters were obtained using the following software: GeneMapper IDX (version 1.5), MS Excel 2002 (version 10.6871.6870), GenAlEx (version 6.5) and Arlequin (version 3.5) to compare four populations (Brazil, Portugal, U.S. and this study population). Results indicated that the most polymorphic loci were D18S51 (17 alleles) and FGA (15 alleles), followed by D21S11 (13 alleles); the least polymorphic loci were D16S539 and TH01 (7 alleles each). Various Brazilian populations ($n > 100,000$) from other studies were compared with this study's Brazilian population using a goodness-of-fit chi-squared test, and no significant differences in these frequencies were observed between these two population groups ($p = 0.9999$). Other forensic and population genetic statistical parameters were calculated comparing this study's population with Portugal and U.S. populations. For example, an analysis of molecular variance (AMOVA) among all populations, among people of the same population and for each person for each population, showed that people have high individual variance (99%) and that this variance is maintained evenly between people of the same sample/region (1%) and among the three populations studied (0%). This study reinforced the conclusion of other allele frequency population studies for the 15 autosomal STR loci tested, confirming high polymorphic grade and high heterozygosity (96.5%). There were significantly more women in the study (79.7%) when compared to the general Brazilian population (50.4%) since the student body of FOB/USP is 80.6% female. Interestingly, an off-ladder D18S51 allele micro-variance corresponding to the allele 29, not yet identified, was found which should be confirmed using paternity and sequencing tests to verify the possibility of either familial occurrence or nucleotide duplication. In the future, due to the small differences found, the parameters obtained in this study should be incorporated into the Brazilian population database and be considered for a regional population genetic prototype database potentially aiding forensic cases with comparisons and calculations. After confirmation, the potentially new micro variant allele found could be included in the international STRBase.

Keywords: Human identification. Allele frequency. STR. Database.

LISTA DE ILUSTRAÇÕES

Figura 1:	Sequência da região de repetição do loco D8S1179 (GenBank: G08710.1 e GI939260).	18
Figura 2:	Posição de Marcadores STR nos Cromossomos Humanos	20
Figura 3:	Perfil alélico do DNA 9947A controle positivo do AmpF ℓ STR® Identifiler® PCR Amplification Kit (Life Technologies, Foster City, CA, USA).	47
Figura 4:	Eletoferograma do perfil genético STR e AMEL da amostra 69.	53
Figura 5:	Distribuição da frequência alélica dos 15 locos entre as populações: FOB, HILL, RIB.	64
Figura 6:	Heterozigosidade individual para os 15 marcadores STR na amostra FOB	67
Figura 7:	Probabilidade de exclusão de acordo com as combinações dos 15 locos em três populações comparadas (FOB, HILL, RIB).	68
Figura 8:	Porcentagens de variação molecular entre as populações FOB, HILL e RIB (AMOVA)	69
Figura 9:	Eletoferograma do loco D18S51 da amostra 194.	70
Figura 10:	Eletoferograma do loco D18S51 da amostra 194: alelo 12 e microvariante do alelo 29 com seus respectivos “stutters”, 11 e 28.	71

LISTA DE TABELAS

Tabela 1:	Composição do KIT AmpFISTR® Identifiler® (Life Technologies, Foster City, CA, USA).....	45
Tabela 2:	Faixa de variação dos alelos, números de alelos encontrados, os alelos mais e menos frequentes em cada marcador genético e o marcador de gênero AMEL.	55
Tabela 3:	Comparações entre as frequências alélicas para o loco D8S1179	56
Tabela 4:	Comparações entre as frequências alélicas para o loco D21S11	56
Tabela 5:	Comparações entre as frequências alélicas para o loco D7S820	57
Tabela 6:	Comparações entre as frequências alélicas para o loco CFS1PO	57
Tabela 7:	Comparações entre as frequências alélicas para o loco D3S1358	57
Tabela 8:	Comparações entre as frequências alélicas para o loco TH01	58
Tabela 9:	Comparações entre as frequências alélicas para o loco D13S317	58
Tabela 10:	Comparações entre as frequências alélicas para o loco D16S539	58
Tabela 11:	Comparações entre as frequências alélicas para o loco vWA.....	59
Tabela 12:	Comparações entre as frequências alélicas para o loco TPOX.....	59
Tabela 13:	Comparações entre as frequências alélicas para o loco D5S818	59
Tabela 14:	Comparações entre as frequências alélicas para o loco D18S51	60

Tabela 15:	Comparações entre as frequências alélicas para o loco FGA	60
Tabela 16:	Frequências alélicas e parâmetros estatísticos de interesse forense e de genética de populações de 15 STR e AMEL na amostra FOB.	62
Tabela 17:	Equilíbrio de Hardy-Weinberg X^2 de acordo com as frequências alélicas de 15 STR na amostra FOB.	65
Tabela 18:	Análise da diversidade genética medida pelo coeficiente de consanguinidade das populações: FOB, HILL e RIB	65
Tabela 19:	Comparação da variância molecular entre as três populações e dentro de cada população FOB, HILL e RIB.	66

LISTA DE ABREVIATURAS E SIGLAS

BD	Banco de dados
CEP	Comitê de Ética em Pesquisa
CODIS	Combined DNA Index System
CONEP	Comissão Nacional de Ética em Pesquisa
FBI	Federal Bureau of Investigation
FIS	Coeficiente de endogamia
GFI	Goodness-of-fit X^2 index
HE	Heterozigosidade esperada
HID	Human identification (identificação humana)
HO	Heterozigosidade observada
HWE	Equilíbrio de Hardy-Weinberg
ISFG	International Society for Forensic Genetics
N	Número de indivíduos
P	Probabilidade do teste exato de Hardy-Weinberg
PCR	Polymerase Chain Reaction
PD	Poder de discriminação
PE	Poder de exclusão
PIC	Conteúdo de informações do polimorfismo
PM	Probabilidade de correspondência aleatória
RFLP	Restriction fragment length polymorphism
RIBPG	Rede Integrada de Bancos de Perfis Genéticos
SENASP	Secretaria Nacional de Segurança Pública
SNP	Single nucleotide polymorphism
SSR	Simple sequence repeats
SSRP	Simple sequence repeats polymorphism
STR	Short tandem repeat
SWGDM	Scientific Working Group on DNA Analysis Methods
TPI	Índice típico de paternidade
VNTR	Variable number tandem repeat

SUMÁRIO

1	INTRODUÇÃO.....	13
1.1	Identificação Humana	15
1.2	Polimorfismos genéticos e a identificação humana	16
1.3	Histórico no Brasil e Aspectos Éticos	23
1.4	Publicações na área	25
1.5	Parâmetros estatísticos de interesse forense – Genética de populações	30
2	JUSTIFICATIVA	33
3	OBJETIVOS	37
4	MATERIAIS E MÉTODOS	41
4.1	Amostras e extração de DNA	43
4.2	Genotipagem PCR-STR.....	44
4.3	Amplificação.....	46
4.4	Eletroforese capilar	46
4.5	Análise Estatística.....	48
5	RESULTADOS	51
6	DISCUSSÃO	73
6.1	Por que formar e incrementar os Bancos de Dados?	75
6.2	Características genéticas da população estudada.....	77
7	CONCLUSÃO	81
	REFERÊNCIAS.....	85

1 INTRODUÇÃO

1 INTRODUÇÃO

1.1 Identificação Humana

Desde o início dos tempos, quando o ser humano ganhou consciência, tem tido uma meta incansável de estabelecer a sua identidade. Entre outros, passou a fazer a determinação do nome próprio, títulos, marcas, adornos, impressão digital e, nos dias de hoje, com o desenvolvimento e a revolução tecnológica, é possível a identificação laboratorial e odontológica de uma pessoa.

As impressões digitais foram o primeiro método biológico de identificação pessoal descoberto em 1858 e publicamente demonstrado em 1877. Mas, foi em 1880 que o pesquisador escocês Henry Faulds publicou o artigo descrevendo que, depois de reunir algumas impressões digitais de seus alunos, teve a ideia de usar estas impressões para encontrar os seus proprietários e assim demonstrou com total êxito esta possibilidade (FAULDS, 1880). Federico Olóriz Aguilera, antropólogo espanhol, foi o criador do Sistema Datiloscópico Espanhol, que seguido por um número pessoal no Brasil, tornou-se o Registro de Identidade Civil (RG).

A ciência forense, que objetiva elucidar litígios de diversas naturezas (civil, trabalhista, criminal) e identificar os culpados, teve início no século XVI na Europa por médicos do exército e universitários que reuniram informações sobre alterações patológicas, causas e circunstâncias da morte. Atualmente, século XXI, a ciência forense aplica métodos científicos para identificar, juntar e analisar informações para a aplicação justa da lei em relação ao direito penal ou civil. Dentre estes métodos, os mais utilizados são as impressões digitais, análise da arcada dentária e de elementos dentais e os testes de DNA.

Os conhecimentos de genética e de técnicas de biologia molecular constituem a base para a identificação de uma pessoa de modo inequívoco, tornando-se fundamental no auxílio à justiça e trazendo uma nova ferramenta: a Genética Forense ou DNA Forense. Este ramo da ciência é definido como “a aplicação de material genético humano e não humano, com o propósito de estudar as características hereditárias para a análise das variações inter e intraespecíficas em populações para a resolução de conflitos legais”, conforme definido no prefácio da conceituada revista denominada “Forensic Science International: Genetics”,

revista científica de abrangência mundial, dedicada à aplicação da genética na administração da justiça e representativa da Sociedade Internacional de Genética Forense (ISFG).

1.2 Polimorfismos genéticos e a identificação humana

A Genética Forense estuda os marcadores conhecidos como polimorfismos de DNA identificados a partir de amostras biológicas de tecidos e fluidos corpóreos (manchas de sangue, sêmen, cabelo, ossos, dentes, unhas, saliva e urina). O pioneiro no uso dos perfis de DNA foi o geneticista britânico Alec Jeffreys. Ele observou a presença de sequências repetitivas de DNA no gene da mioglobina e verificou que estas repetições se espalham ao longo de todo o genoma com comprimentos variáveis, de modo a formar bandas distintas em um gel. Este padrão, com número de repetições em série altamente variável na população, portanto polimórfico, é transmitido de forma hereditária (JEFFREYS, BROOKFIELD e SEMEONOFF, 1985; JEFFREYS, WILSON e THEIN, 1985a; b).

O método, conhecido como impressões digitais de DNA ou DNA “fingerprinting”, é considerado uma das maiores descobertas do final do século XX e revolucionou as investigações forenses, tornando-se fundamental para elucidação de crimes, na resolução de disputas de paternidade, imigração e estudos antropológicos (LEWIS, REYNOLDS e GAGG, 2004). O princípio é simples: conforme a sequência de DNA, cada pessoa é caracterizada como um indivíduo singular. Quanto mais polimórficos os sistemas utilizados, maior a probabilidade de identificação. Um loco é considerado polimórfico quando o seu alelo mais raro tem frequência mínima de 1%.

As principais técnicas laboratoriais para identificar, comparar e avaliar o DNA são as análises de polimorfismo de comprimento dos fragmentos de restrição (RFLP), a reação em cadeia da polimerase (PCR), o sequenciamento e, mais recentemente, o pirosequenciamento (DIVNE, EDLUND e ALLEN, 2010).

O genoma humano eucromático é composto por aproximadamente 2,85 bilhões de nucleotídeos interrompidos por apenas 341 lacunas (gaps), sendo que menos de 3% representa regiões codificadoras - apenas 20.000 a 25.000 genes codificam proteínas, de acordo com o International Human Genome Sequencing Consortium (CONSORTIUM, 2004). As sequências não codificadoras compreendem cerca de 98%, sendo que 5% destas são altamente

conservadas, em escalas de tempo calculadas em centenas de milhões de anos, o que implica que estas regiões não codificadoras conservadas estão sob forte pressão evolutiva e de seleção positiva (CONSORTIUM, 2001). As regiões não codificadoras do DNA são mais polimórficas, sendo, por isso, muito utilizadas na identificação humana desde os anos 1990 (URQUHART *et al.*, 1994; REGATEIRO, 2007).

Regiões não codificadoras repetitivas são geradas por repetição sequencial da mesma unidade e formam as sequências satélites classificadas como minissatélites ou VNTR (Variable Number Tandem Repeat), com unidades repetitivas entre 9 e 100 bases, e microsatélites, conhecidas como STR em humanos (SSR – single sequence repeats - em plantas e animais), com unidades repetitivas entre 2 a 7 bases. As sequências STR constituem cerca de 3% do genoma humano e apresentam-se flanqueadas por sequências únicas, sendo unilocos e com herança codominante (TAUTZ, 1993; PHILLIPS, FONDEVILA e LAREAU, 2012).

Os polimorfismos de repetição, conhecidos genericamente como SSRP (Simple Sequence Repeat Polymorphisms), ocorrem em vários cromossomos e são formados devido à continuidade de ação da DNA polimerase que “desliza” no momento da replicação, sem um mecanismo de reparo eficiente, gerando inserções ou deleções no número original de repetições, levando ao polimorfismo alélico (ELLEGREN, 2004).

Em 1997, a International Society of Forensic Haemogenetics (ISFH) padronizou as regras para a definição e a nomenclatura dos alelos. A nomenclatura é pré-requisito para a reprodutibilidade e comparação de dados, sendo feita de acordo com a localização no cromossomo e a caracterização. Por exemplo, D8S1179 significa: D = molécula de DNA; 8 = cromossomo no qual se encontra o gene; S = sequência de cópia única; 1179 = ordem de descoberta e caracterização desta sequência no cromossomo (BAR *et al.*, 1997).

Os alelos são definidos de acordo com o número de repetições de uma determinada sequência. A unidade repetida ou sequência molde é chamada de motivo, podendo ser dinucleotídica a até hexanucleotídica. Assim, o alelo 5 tem 5 repetições da sequência motivo; o alelo 6 tem 6 repetições e assim sucessivamente. Quando ocorre uma repetição incompleta, usa-se um ponto decimal para separar as repetições completas dos nucleotídeos adicionais. Exemplo: o alelo 9.3 do loco TH01 contém nove tetranucleotídeos (AATG) e uma repetição incompleta de 3 nucleotídeos, diferindo do alelo 10 por uma única base adenina suprimida na sétima posição de repetição (PUERS *et al.*, 1993). A designação dos alelos é feita por

comparação com uma escada alélica padrão que deve ser abrangente e representativa dos alelos dos locos a serem analisados (BAR *et al.*, 1997; GILL *et al.*, 1997; LINCOLN, 1997).

Outra característica é que o marcador pode ter uma variação na sequência repetida. Por exemplo, ainda em relação ao marcador D8S1179, este é classificado como uma STR composta porque originalmente é formado pela repetição do tetranucleotídeo TCTA (Figura 1). Porém, em casos de alelos com mais de 13 repetições, o loco é constituído pelas repetições TCTG, desde a segunda ou terceira posição a partir da extremidade 5' da região TCTA (BUTLER, 2005).

```

1  tggcaactta tatgtatattt tgtatttcat gtgtacattc gtatctatct atctatctat
61 ctatctatct atctatctat ctatctatct attccccaca gtgaaaataa tctacaggat
121 aggtaaaataa attaaggcat attcacgcaa tgggatacgn tacagtgatg aaaatgaact
181 aattatagct acgtgaaact atactcatgn acacaatttg gtaaaagaaa ctgggaacaa
241 gaatacatac ggtttttgnc agctgtgcta ttttacattc ccaacaacaa tgcacagggt
301 ttcagnttct ccacatnctt gtcaacattn tgttattttg

```

Figura 1: Sequência da região de repetição do loco D8S1179 (GenBank: G08710.1 e GI939260).

Os métodos aceitos para detectar estes marcadores STR são a separação de tamanho por eletroforese capilar, o sequenciamento de segunda geração e sequenciamento da próxima geração, o NGS (Next-generation sequencing). Além disso, o sequenciamento é um meio eficaz de revelar variações de uma base no DNA conhecidas como polimorfismos de nucleotídeo único (SNPs) (NIELSEN *et al.*, 2011). Enquanto o método tradicional baseado em detecção de STR por eletroforese capilar revela apenas o comprimento de alelos, o pirosequenciamento pode aumentar a resolução e detectar variação de nucleotídeos dentro de regiões de repetição e em regiões próximas de flanco (BERGLUND, KIIALAINEN e SYVANEN, 2011).

As novas tecnologias como, por exemplo, os miniprotocolos padrões mais rápidos e menos dispendiosos foram validados em laboratórios de excelência. Atualmente, a amplificação por PCR seguida de eletroforese capilar compõem a principal metodologia usada para amplificar, separar e detectar alelos STR em laboratórios de DNA forenses em todo o mundo. Para se alcançarem resultados confiáveis, três condições são fundamentais: a resolução espacial é necessária para separar alelos que diferem em tamanho por um único nucleotídeo; a resolução espectral de cores dos corantes fluorescentes é necessária para separar uns dos outros de modo que os produtos de PCR a partir de locos marcados com

diferentes corantes possam ser verificados; a precisão de dimensionamento do DNA deve ser consistente o suficiente para que as amostras sejam relacionadas com escadas alélicas corridas para fins de calibração (BUTLER e HILL, 2012).

A identificação dos alelos por meio da comparação com uma escada alélica é rotineira em humanos e todos os kits de STR comercialmente disponíveis fornecem sua própria escada alélica padrão. As empresas que fornecem estes produtos têm investido muito tempo em pesquisa para garantir que a escada seja tão abrangente quanto o possível, mas a construção dessa escada alélica é complexa e deve permitir a comparação dos alelos entre os diversos laboratórios com denominações precisas que tornem todos os resultados de todos os kits, máquinas e laboratórios diretamente comparáveis (BUTLER, 2014).

Mais de 3.500 referências de aplicação dos locos STR para identificação de DNA forense com detalhes sobre a tecnologia em centenas de diferentes grupos populacionais têm sido publicadas. A página *NIST STRBase*, no endereço eletrônico <http://www.cstl.nist.gov/biotech/strbase>, pode ser consultada para as principais publicações. Esta lista crescente de publicações evidencia que os marcadores STR tornaram-se importantes ferramentas para testes de identidade humana no mundo todo e a expansão do número de locos assegura que estes serão usados por muitos anos vindouros (BUTLER e HILL, 2012).

Perfis de DNA forense são atualmente realizados utilizando um painel de marcadores STR multialélicos estruturalmente análogos aos minissatélites originais, mas com curtos intervalos de repetição e, portanto, mais fáceis para amplificar e “multiplexar” com a reação em cadeia da polimerase. Até 30 STR podem ser detectadas numa única eletroforese capilar, gerando para cada indivíduo um código único (ROEWER, 2013).

Basicamente, existem dois conjuntos de marcadores STR que satisfazem as normas solicitadas pelos bancos de dados criminais em todo o mundo: o conjunto padrão Europeu de 12 STR marcadores (GILL *et al.*, 2004) e o padrão norte-americano CODIS (Combined DNA Index System) de 13 marcadores (COLLINS *et al.*, 2004). Devido à sobreposição parcial, formam em conjunto um padrão de 18 marcadores STR. Atualmente, estão identificados e bem caracterizados 47 locos autossômicos STR (BUTLER e HILL, 2012). Estima-se que em 2017 sejam utilizados 20 marcadores para finalidades forenses. A Figura 2 abaixo ilustra os marcadores e suas posições cromossômicas (BUTLER, 2015).

Figura 2: Posição de marcadores STR nos cromossomos humanos. Copiado de Butler, 2015. http://www.cstl.nist.gov/strbase/pub_pres/Butler-DNA-interpretation-AAFS2015.pdf. Acessado em 24/10/2015.

O CODIS é um programa criado em 1998, de propriedade exclusiva do Federal Bureau of Investigation (FBI), que permite a comparação de perfis genéticos. É o primeiro passo para a criação de uma rede integrada de dados de DNA. Nos USA, o CODIS não é acessível ao público em geral ou para a maioria dos profissionais da justiça criminal. Apenas o administrador do banco de dados em laboratório do governo e o pessoal responsável pela entrada de perfis de DNA para CODIS têm acesso ao banco de dados de DNA em todos os níveis (local, estadual e nacional).

A produção de kits comerciais melhorou muito a padronização destes marcadores para todos os tipos de provas de DNA, com resultados reprodutíveis mesmo com poucas células nucleadas e amostras gravemente comprometidas. A probabilidade dos indivíduos não aparentados terem os mesmos marcadores nos 13 diferentes locos STR contemplados pelo CODIS é de 1/1 bilhão. Esta singularidade e a codificação numérica simples levou ao estabelecimento de bancos de dados de DNA para investigações controlados pelo governo nas nações desenvolvidas em todo o mundo (ROEWER, 2013).

A validação dos dados para finalidades forenses requer duas diretrizes principais que envolvem os vários aspectos de testes biológicos de DNA neste campo: "Orientações de

validação dos Métodos de Análise de DNA" emitida pelo Grupo de Trabalho Científico sobre Métodos de Análise de DNA (SWGDM - Scientific Working Group on DNA Analysis Methods), e "Critérios Mínimos Recomendados para a Validação", emitido pela Rede Européia de Institutos de Ciência Forense - Grupo de Trabalho DNA (ENFSI - European Network of Forensic Science Institutes) (GILL *et al.*, 2004). Em resumo, os seguintes estudos são necessários: amostras conhecidas, repetibilidade, reprodutibilidade, sensibilidade, efeitos estocásticos, misturas e avaliação de contaminação.

A análise de STR rotineiramente usada envolve analistas forenses altamente treinados que usam múltiplos instrumentos para processar e analisar amostras no laboratório. O fluxo de trabalho básico de referência é: (1) extração de DNA; (2) amplificação de locos STR pela PCR multiplex com primers fluorescentes marcados; (3) separação do tamanho dos fragmentos por matriz de eletroforese capilar com detecção de fluorescência induzida por laser; e (4) análise dos eletroferogramas resultantes usando programas de genotipagem. (JOVANOVIĆ *et al.*, 2015).

A interpretação dos dados STR começa pela análise minuciosa do perfil de DNA obtido no eletroferograma. A quantificação e qualidade do DNA são exigidas para estudos de amostras forenses, caso contrário, podem surgir perdas de sinais por inibidores de PCR ou de DNA degradado, ou picos extras fora da escada (off-ladder) (EGYED *et al.*, 2000; ALLOR, EINUM e SCARPETTA, 2005; BECKER *et al.*, 2007).

Como até seis cores fluorescentes são usadas para criar os perfis de DNA, cores adjacentes podem ser reproduzidas se picos de dados aparecem fora da escada ou a calibração espectral não estiver devidamente feita. Estes picos extras (pull-up) fora da escada de dados podem cair em faixas de tamanho específico para algum alelo de loco STR em outra cor e, neste caso, precisam ser deletados através de edição do perfil para adequada interpretação (GRUBWIESER *et al.*, 2005; HEINRICH *et al.*, 2005).

Alternativamente, a amostra de DNA pode ser reamplificada com menos DNA, ou os produtos PCR originais podem ser reinjectados provocando a obtenção de picos extras sobre a escada. Assim, estes picos podem complicar a interpretação do perfil STR. Outras complicações na interpretação dos dados STR que podem surgir incluem a presença de padrões trialélicos, exclusões de alelos ou alelos nulos (HUEL *et al.*, 2007; KLINE *et al.*, 2011).

A análise dos eletroferogramas também revela a presença de pequenos picos com uma unidade de repetição a menos do que o pico alélico propriamente dito. São os chamados “stutters” ou picos de gagueira resultantes do processo de repetição pela polimerase de cópias com pareamento incorreto, caracterizando um artefato técnico bem definido. Na literatura, estes produtos também têm sido referidos como banda de sombra (HAUGE e LITT, 1993; KLINTSCHAR e WIEGAND, 2003).

O rápido crescimento nas aplicações de identificação humana por meio de STR e o consequente aumento de dados investigativos têm levado à descoberta de microvariantes alélicas. É o caso de alelos STR que se estendem para além das suas gamas de tamanho esperado, os chamados “off-scale” ou fora da escala estabelecida e contemplada pelos kits, mas que podem ser verificados e confirmados por critérios técnicos. Para contemplar esta situação e garantir a atualização dos dados em tempo real, estão disponíveis na rede mundial de computadores páginas contendo uma base de dados que podem ser atualizados por usuários competentes para tal, a chamada STRBase, fundada oficialmente em julho de 1997 e mantida pela “Applied Genetics Group of the National Institute of Standards and Technology” (RUITBERG, REEDER e BUTLER, 2001; BUTLER, 2012). As microvariantes são mais comumente encontradas nos locos STR mais polimórficos, como FGA, D21S11 e D18S51, que possuem as maiores e mais complexas estruturas de repetição em comparação com os locos mais simples como TPOX e CSF1PO (BUTLER, 2014).

Particularmente, em alusão ao loco D18S51, durante esta pesquisa, encontramos uma microvariante que será apresentada e discutida posteriormente. Sendo assim, torna-se pertinente uma descrição pormenorizada sobre este loco.

O loco D18S51, descrito em 1993, é composto pela repetição de um tetranucleotídeo simples encontrado no braço longo do cromossoma 18. O motivo de repetição é AGAA, mas algumas variantes alélicas são esperadas devido à inserção do dinucleotídeo AG na região flanqueadora 3'. Inicialmente, foram identificados 21 alelos com tamanhos variando entre 271 a 343 pares de bases (BARBER e PARKIN, 1996). Em 2009, foram referidos 12 novos alelos (MORALES-VALVERDE *et al.*, 2009). Atualmente, mais de 70 alelos e microvariantes têm sido relatados, tornando-o um dos mais polimórficos do núcleo CODIS, com elevada heterozigosidade e muito informativo para detecção de misturas de amostras. Todos os kits comercializados contêm os mesmos iniciadores de PCR para D18S51 (BUTLER, 2014).

Na STRBase, é apresentada uma tabela com 53 microvariantes alélicas do D18S51, atualizada em 07/08/2015, contendo: designação alélica estimada, tamanho do alelo, modelo do sequenciador, nome dos pesquisadores que contribuíram com a identificação, método de confirmação do resultado, observações pertinentes e a frequência até então verificada. Consideramos oportuna a apresentação de trechos ilustrativos específicos da referida tabela na forma de anexo, contendo alguns dados existentes sobre este loco, objetivando dar ideia de como o trabalho colaborativo está sendo construído no mundo todo, agregando pesquisadores das mais diversas regiões e culturas. Para embasamento à discussão dos dados obtidos nesta pesquisa, foram incluídas propositadamente, neste anexo, algumas microvariantes alélicas encontradas fora da escada, mas dentro da faixa esperada, acrescida de algumas, que além de fora da escada, também estão fora da gama de tamanhos esperados, próximas ao pico extra observado nesta pesquisa. (Anexo 1). A STRBase pode ser acessada via <http://www.cstl.nist.gov/biotech/strbase>.

O conhecimento acumulado a partir de observações experimentais do comportamento destes e de outros artefatos técnicos é primordial para a correta leitura e interpretação dos resultados.

1.3 Histórico no Brasil e Aspectos Éticos

Testes laboratoriais para fins judiciais de identificação humana são utilizados no Brasil de maneira informal desde os anos 1980. A princípio, foram os grupos sanguíneos e depois, a partir de 1983, o sistema HLA. Nos anos 1990, surgiram os primeiros testes de DNA em alguns laboratórios detentores de tecnologia de ponta.

Em 1997, por meio do Decreto nº 2.315, de 04/09/1997, a SENASP (Secretaria Nacional de Segurança Pública) passou a auxiliar o Ministro de Estado da Justiça na definição e implementação da política nacional de segurança pública, sendo uma de suas prerrogativas a modernização e ampliação do sistema nacional de informações de justiça e segurança pública, efetivar o intercâmbio de experiências técnicas e operacionais entre os serviços policiais federais e estaduais. No ano seguinte, o Artigo 17 do Decreto nº 2.802 de 13 de Outubro de 1998, determinou "... implementar o registro de identidade civil (ação de investigação de

paternidade através do DNA); estimular a capacitação dos profissionais da área de segurança pública; realizar estudos e pesquisas e consolidar estatísticas nacionais...”.

No ano 2000, a Sociedade Brasileira de Medicina Legal publicou as “Recomendações para Laboratórios – Testes de Paternidade por meio de DNA” estabelecendo um código de conduta ética e laboratorial para laboratórios que realizam os referidos testes, cujas normas são seguidas até hoje.

A criação de um programa cooperativo nacional e o apoio financeiro da SENASP, em 2004, permitiu a melhoria dos laboratórios existentes, a criação de novos laboratórios, a capacitação e treinamento de peritos e a formação de um grupo de especialistas que estabeleceu as principais regras a serem adotadas.

Em 2009, a SENASP assinou um acordo de cooperação técnica para o uso do banco de dados de perfis genéticos CODIS na identificação de suspeitos de crimes. Seis Estados (Amazonas, Amapá, Ceará, Mato Grosso, Paraíba e Paraná) assinaram o acordo de cooperação técnica com o Ministério da Justiça para o uso do banco de dados, com o objetivo de acesso ao programa auxiliar, principalmente, na solução dos casos de estupro e pedofilia. A utilização do programa foi cedida, pelo FBI, gratuitamente para o Brasil.

Em 2010, ocorreu a solenidade da implantação do CODIS em território nacional. A seguir, alguns estados começaram a criar BD a partir do programa CODIS: Amazonas, Amapá, Bahia, Ceará, Espírito Santo, Minas Gerais, Mato Grosso do Sul, Pará, Paraíba, Paraná, Rio de Janeiro, Rio Grande do Sul, Santa Catarina, Mato Grosso e São Paulo, contemplando a área criminal e de identificações civis, havendo um comitê gestor responsável pela fiscalização e pelas regras de utilização dos perfis genéticos de criminosos para o cruzamento de amostras de DNA. (BONACCORSO, 2010).

Desde então, o Brasil celebrou um termo de compromisso com o FBI e acordos de cooperação técnica com as secretarias de segurança estaduais para formar a sua própria Rede Integrada de Bancos de Perfis Genéticos (RIBPG), de criminosos.

Em 2012, foi sancionada a LEI 12.654/2012 publicada no Diário Oficial da União – DOU- no dia 29/05/2012 que criou um banco de DNA de condenados por crimes violentos. A lei torna obrigatória a identificação genética, por meio de DNA, de condenados por crimes hediondos ou crimes violentos contra a pessoa, como homicídio, extorsão mediante sequestro, estupro, entre outros. O objetivo é utilizar os dados colhidos nas investigações de crimes

cometidos por ex-detentos, ou seja, os reincidentes. De acordo com o texto, os condenados "serão submetidos, obrigatoriamente, à identificação do perfil genético, mediante extração de DNA. A identificação do perfil genético será armazenada em banco de dados sigiloso, conforme regulamento expedido pelo Poder Executivo". A polícia poderá requisitar ao juiz o acesso ao banco de dados. A lei prevê punição civil, penal e administrativa àquele que permitir ou promover a utilização dos dados para outros fins.

Em 2013, por meio do decreto 7.950, foram instituídos, formalizados e regulamentados no âmbito do Ministério da Justiça, o Banco Nacional de Perfis Genéticos e a Rede Integrada de Bancos de Perfis Genéticos – RIBPG. Nesta ocasião, existiam 17 laboratórios cadastrados para análises genéticas com o sistema CODIS, registrados na RIBPG.

Estas iniciativas foram um grande passo no combate ao crime. Quanto ao banco de dados de informações genéticas de pessoas desaparecidas, embora sendo ferramenta investigativa eficiente e servindo para armazenamento e busca, ainda há carência para esta finalidade (FIGUEIREDO e PARADELA, 2006).

Tratando-se de Genética Forense, além de BD com perfis de criminosos, suspeitos, acusados ou condenado por um crime, para a correta interpretação e cálculos estatísticos de probabilidades, é necessário o conhecimento prévio do perfil genético da população em geral, residente na região. Por isso, a indicação de estudos e da criação de BD da população regional.

1.4 Publicações na área de marcadores STR

Sem dúvida, a investigação de semelhanças e diferenças genéticas entre populações humanas e a compreensão sobre a magnitude e o significado da variabilidade representam um instigante campo, suscitando a curiosidade científica em todo o mundo e, de certa forma, justificando os inúmeros artigos científicos abordando estudos de frequências gênicas nas populações, independente das finalidades a que se propõem.

De modo geral, considerando que somos praticamente iguais porque compartilhamos 99,7% do mesmo DNA e diferimos em apenas 0,3% que nos torna seres únicos (BUTLER, 2005), estudos populacionais são importantes e assumem determinados interesses justamente pelas diferenças, quando estas são encontradas. Sendo assim, é de se esperar que a maioria esmagadora dos cerca de 3.500 artigos científicos abordando frequências alélicas nas populações não traga muitos dados

dignos de nota ou que representem achados científicos novos, mas são imprescindíveis para a atualização e complementação de dados obrigatórios para a composição de Banco de Dados comparativos para finalidades forenses, enquanto ajudam a responder outras questões sobre distribuição geográfica, história evolutiva, seleção natural, entre outras (ROSENBERG, 2011; ALGEE-HEWITT *et al.*, 2016).

As publicações na área seguem um formato padrão devido às finalidades às quais se destinam. A partir dos anos 1990, artigos referentes às frequências gênicas de marcadores STR em amostras populacionais de diversas regiões do planeta e em regiões brasileiras têm sido publicados.

Uma das revistas escolhidas pelos cientistas da área para a publicação de dados referentes às frequências populacionais, padronização de amostras, número de marcadores e demais critérios é a “Forensic Science International: Genetics”, que contém a maioria expressiva dos artigos científicos na área. Geralmente, estes artigos apresentam e discutem os seguintes dados: descrição da população, questões de genética de populações de interesse e relevância forense; análise e controle de qualidade; relato ou inclusão de novos marcadores ou simplesmente relato de frequências alélicas ou de haplótipos. Também podem ser na forma de comunicação sucinta, entendendo que tanto a qualidade dos dados sobre a população e a relevância dos resultados são cruciais, ou ainda, comunicações no formato de carta ao editor e tabelas com atualizações do tamanho da amostra, caso o ineditismo dos dados não seja suficiente para um artigo original ou uma curta comunicação. Estes dados de atualizações podem ser baixados como arquivos eletrônicos suplementares (CARRACEDO *et al.*, 2013; 2014).

Quanto ao controle de qualidade, tanto para STR e SNP, os autores devem declarar que seguem estritamente as recomendações sobre a análise dos polimorfismos de DNA utilizados, a nomenclatura e análises estatísticas, e ainda testes controles de amostras historicamente conhecidas (CARRACEDO *et al.*, 2010).

Conforme já versado anteriormente, são inúmeros os artigos publicados, mas, na impossibilidade de mencioná-los todos, consideramos importante a citação de alguns destes referentes à população brasileira com uma descrição sucinta sobre o seu conteúdo.

Os primeiros trabalhos brasileiros surgiram nos anos 2.000 na literatura internacional e seus dados foram quase que invariavelmente extraídos de casos forenses de investigação de paternidade (FERREIRA DA SILVA *et al.*, 2002; FIGUEIREDO *et al.*, 2004; GOES *et al.*, 2004; WHITTLE, ROMANO e NEGREIROS, 2004; RODRIGUES, PALHA TDE e DOS SANTOS, 2007; DEL

CASTILLO *et al.*, 2009; DE ASSIS POIARES *et al.*, 2010; AGUIAR *et al.*, 2012; VIEIRA *et al.*, 2013; GASTALDO *et al.*, 2015; SANTOS *et al.*, 2015)

Merece destaque um estudo envolvendo os mesmos locos estudados no presente trabalho que revelou similaridade da frequência alélica entre brasileiros e ameríndios, europeus (portugueses principalmente) e africanos (VIEIRA *et al.*, 2013). Esta mesma similaridade foi encontrada entre habitantes do nordeste e centro-oeste brasileiro e da cidade do Rio Janeiro, o que pode explicar o padrão de imigração durante a colonização no Brasil (WOLFGRAMM EDE *et al.*, 2011).

Um estudo com mais de 12 mil pessoas brasileiras revelou que os locos autossômicos STR mais polimórficos são o D2S1338 e o D18S51 (DE ASSIS POIARES *et al.*, 2010). Estas informações são muito valiosas, pois quanto mais diferenças se encontram em um loco, mais singularidade se confere ao indivíduo.

Os polimorfismos de 9 STR, incluindo frequências alélicas dos locos STR 16S539, D7S820, D13S317, CSF1PO, TPOX, TH01, F13A01, FESFPS e vWA, relativas à população de Alagoas foram estudados em amostras de 598 indivíduos da população de Alagoas no Nordeste do Brasil. Foram determinadas as frequências alélicas assim como vários índices estatísticos utilizados em testes de paternidade e casos forenses. Os locos mais polimórficos encontrados foram YH01 e o D7S317, estando as amostras em equilíbrio de Hardy-Weinberg ($P > 0.05$) (FERREIRA DA SILVA *et al.*, 2002).

No Rio de Janeiro e Mato Grosso do Sul, amostras foram coletadas de 208 e 234 indivíduos não aparentados. A PCR foi feita com o kit PowerPlex 2.1 e PowerPlex 1.1. Análises estatísticas foram feitas utilizando o programa DNAtyping fornecido por Ranajid Chakraborty. Os 14 locos STR analisados ofereceram um sistema de alta discriminação para utilização em testes de paternidade e em casos forenses. Não foram constatadas diferenças nas frequências alélicas em relação a outras populações brasileiras (SILVA *et al.*, 2004).

No estado de Pernambuco, foram estudadas as frequências alélicas de 13 locos STR obtidos de 546 amostras de indivíduos não aparentados. Esta população representa uma mistura de caucasianos, africanos e nativos americanos. Todos os locos estudados estavam HWE. Os parâmetros forenses calculados demonstraram ser úteis para a solução de problemas forenses na região nordeste do Brasil (DELLALIBERA *et al.*, 2004).

Frequências alélicas de 13 locos STR foram analisados em amostras de sangue de 325 indivíduos não aparentados da população do Amazonas. O poder de discriminação e a

probabilidade de exclusão para os 13 STR foram de 99.9999999992% e 99.9998% respectivamente. Os dados foram analisados utilizando os programas Arlequin 2.000, GENEPOP e Powerstatsv 12.xls. Todos os locos alcançaram o equilíbrio Hardy-Weinberg. Concluiu-se que os marcadores foram apropriados para utilização em casos forenses e de paternidade na população do Amazonas (RODRIGUES, PALHA TDE e DOS SANTOS, 2007).

No Centro de Doação de Sangue da Paraíba, foram estudadas STR de 323 indivíduos não aparentados. Foram calculadas as frequências alélicas, polimorfismo (PIC), poder de discriminação, poder de exclusão, probabilidade de correspondência e índice típico de paternidade utilizando o programa Powerstats V12. Os locos observados estavam em HWE. Foram feitas comparações envolvendo outras 5 populações do Nordeste brasileiro, uma de Portugal e uma da Andaluzia. Para índices forenses, as comparações foram dificultadas por conta das diferenças abordagens estatísticas utilizadas nos diversos estudos comparados. Entretanto, para todos os estudos foram calculados o poder de discriminação, o qual não pareceu variar significativamente entre as populações (GOMES *et al.*, 2007).

Em Minas Gerais, foram estudados DNA de uma média de 12.000 indivíduos não aparentados. Os dados necessários para os cálculos estatísticos foram extraídos do banco de dados de casos de paternidade do Laboratório de Biologia Genética e Molecular do Núcleo de Ações e Pesquisa em Apoio Diagnóstico (NUPAD) da Universidade Federal de Minas Gerais. A PCR foi feita utilizando o kit AmpFISTR® Identifiler™ PCR Amplification Kit (Applied Biosystems). Os índices observados foram: frequências alélicas, teste HWE, heterozigosidade observada, heterozigosidade esperada, poder de exclusão, poder de discriminação e polimorfismo e foram calculados para todos os 15 locos analisados. Os cálculos estatísticos foram feitos utilizando GenePop 3.4, GDA 1.0, Arlequin 3.01 e Cervus 2.0. Todos os locos alcançaram o equilíbrio de Hardy-Weinberg. Os resultados demonstraram o alto poder de discriminação observados da população de Minas Gerais (DEL CASTILLO *et al.*, 2009)

Frequências alélicas de 15 locos STR foram obtidas de 2.038 indivíduos oriundos de testes de paternidade na população do Rio Grande do Sul, analisados pelo kit comercial AmpFISTR1 Identifiler (Applied Biosystems). O loco mais polimórfico dessa amostra foi o D2S1338. As taxas de mutação foram incertas a partir dessa amostra. Todos os locos estavam em HWE. Os resultados de distância genética foram consistentes com a descendência europeia da população do Rio Grande do Sul (CHULA *et al.*, 2009).

No Distrito Federal, frequências alélicas de 18 locos STR foram obtidas de 915 indivíduos e comparados a outras amostras do Brasil. As amostras foram retiradas de sangue, saliva e swab. Todos os locos estavam em equilíbrio de Hardy-Weinberg, exceto o F13B. Índices estatísticos foram utilizados para validar o estudo (DALTON *et al.*, 2009).

Em Goiânia, amostras de sangue foram coletadas de 429 indivíduos. A amplificação foi feita com o kit PowerPlex BIO 16, Promega Corporation. Frequências alélicas e outros parâmetros estatísticos foram calculados pelo programa Power Stats v1.2.xls (Promega Corporation, USA). Os cálculos do equilíbrio de Hardy-Weinberg e da heterozigosidade esperada foram feitos utilizando o programa Arlequin versão 3.1.1. Todos os locos alcançaram o equilíbrio de Hardy-Weinberg. Os autores concluíram que os 15 locos oferecem um poder de discriminação para casos de paternidade e casos forenses e que não houve diferença na comparação com outras populações brasileiras (SILVA *et al.*, 2010).

Outra pesquisa objetivou construir um banco de dados com frequências alélicas de 13 locos STR com os resultados disponibilizados pelo Instituto de Ciência da Saúde (ICS), Universidade Federal da Bahia (UFBA) e pela Coordenação de Genética Forense do Laboratório Central de Polícia Técnica (LCPT) para dar maior força de discriminação aos indivíduos autores de eventos criminais com as técnicas de identificação humana por DNA na Bahia. Frequências alélicas foram calculadas em uma amostra de 1.000 indivíduos não aparentados. As amostras foram escolhidas randomicamente de acordo com sua origem na Bahia. Vários índices estatísticos foram utilizados para validar as amostras (Ho (heterozigosidade observada), He (heterozigosidade esperada), PD (poder de discriminação), PE (probabilidade de exclusão), PIC (conteúdo de informação polimórfica), HWE (equilíbrio de Hardy-Weinberg) e valor de P (NASCIMENTO, CERQUEIRA e GUSMAO, 2011)).

No Paraná, foram calculadas as frequências alélicas de 15 locos STR, seus parâmetros forenses e estatísticas utilizando AmpFISTR Identifier. Todos os locos estavam em HWE. O poder de discriminação e o de exclusão foram de 0,999999999999999 e 0,9999993 respectivamente. Concluiu-se que população do Paraná é, na maioria, de descendência europeia. A combinação dos 15 locos STR representa uma estratégia poderosa para a identificação individual e análise parental para a população do Paraná (ALVES *et al.*, 2014).

O maior e mais completo estudo brasileiro realizado até este ano de 2015, foi feito por Aguiar e colaboradores em 2012. Os autores analisaram os dados de 15 STR obtidas de até

137.161 indivíduos não aparentados da geração F1 dos casais submetidos a testes de paternidade no laboratório Hermes Pardini, em Minas Gerais. Estes indivíduos foram representativos de todas as 5 regiões geográficas e do Distrito Federal. O total de indivíduos genotipados para cada loco foi: CSF1PO (113.526); D3S1358 (135,133); D5S818 (135,181); D7S820 (137,136); D8S1179 (134,211); D13S317 (137,161); D16S539 (136,942); D18S51 (136,739); D21S11 (130,014); FGA (135.839); Penta D (110.333); Penta E (128.055); TH01 (112,695); TPOX (123102); SVA (127.415). O kit utilizado foi o PowerPlex 116 da Promega. Os parâmetros estatísticos foram calculados por meio de análise de variância molecular (AMOVA) e análise de pareamento (FST) usando o programa Arlequin (v3.11). O tamanho dos alelos variou até 48,2 e os parâmetros estatísticos (PD, PIC e Ho) considerando valores de $P \geq 0,75$, sugeriram que os marcadores D13S317, D16S539, D18S51, D21S11, D7S820, D8S1179, Penta D, Penta E, TH01, FGA e SVA foram mais informativos para fins de identificação genética na população brasileira (AGUIAR *et al.*, 2012).

1.5 Parâmetros estatísticos de interesse forense – Genética de Populações

Considerando a aplicação forense, um desafio comum para a apresentação de dados de STR nos tribunais é que as estimativas das frequências utilizadas para construir o perfil alélico baseiam-se no grupo populacional foco que nem sempre é representativo de populações menores, muitas vezes isoladas, ou subpopulações a partir das quais um réu pode ter origem. Portanto, deve haver estudos regionais com ajustes para o efeito de estratificação de frequências alélicas derivadas de pesquisa de um grupo populacional foco (PHILLIPS *et al.*, 2011).

Para a análise estatística dos dados, os parâmetros clássicos de estudos de genética de populações devem ser contemplados: N = número de indivíduos; frequências alélicas; PD = poder de discriminação; PM = probabilidade de correspondência aleatória; PE = poder de exclusão; PIC = conteúdo de informações do polimorfismo; HO = heterozigosidade observada; HE = heterozigosidade esperada; FIS = coeficiente de endogamia ou consanguinidade; HWE = equilíbrio de Hardy-Weinberg; TPI = índice típico de paternidade (BALDING e NICHOLS, 1995; SONG, DEY e HOLSINGER, 2011).

Um grande avanço para a compreensão da genética de populações foi o desenvolvimento do método de análise de variância molecular (AMOVA) e de programas de computador. AMOVA

permite o particionamento hierárquico da variação genética entre as populações e regiões e a estimativa da amplamente utilizada estatística F ou de seus análogos. Foi um marco porque permitiu pela primeira vez a análise hierárquica da estrutura da população para todos os tipos de marcadores genéticos (haplóides, haplótipo, binário, codominantes e sequência), bem como a oferta de opções de teste estatísticos por permutação aleatória.

Os pacotes de programas Arlequin e GenAlEx, entre outros, oferecem o teste de AMOVA. O GenAlEx fornece exportação de dados para estatística F, estatística G e distância genética entre populações (PEAKALL e SMOUSE, 2006; EXCOFFIER e LISCHER, 2010; PEAKALL e SMOUSE, 2012). Os algoritmos para esses cálculos seguem o código disponível publicamente no programa MaRQ (LYNCH e RITLAND, 1999).

A hipótese de nulidade (H_0) típica na biologia significa que não há diferença entre os grupos comparados e no teste AMOVA, também: H_0 = nenhuma diferença genética entre as populações ($F_{ST} = 0$) e H_1 = há uma diferença genética entre as populações ($F_{ST} > 0$). Desta forma, sob hipótese H_0 as subpopulações podem ser consideradas parte de uma única grande população genética com acasalamento aleatório.

O teste estatístico por pareamento aleatório embaralha os indivíduos entre as populações e regiões. Para calcular as probabilidades, os indivíduos são embaralhados dentro da população e entre as populações e sub populações. Os seguintes parâmetros são calculados:

FIS = Coeficiente de endogamia: proporção da variância em uma sub população contida na população total. Mede a redução de heterozigotia de um indivíduo devido à inexistência de cruzamentos aleatórios na sua subpopulação. Quanto maior o valor de FIS, maior é o grau de endogamia.

FIT = Coeficiente de endogamia total: mede a redução de heterozigotia de um indivíduo em relação à população total, tendo em conta os efeitos da subdivisão populacional e a inexistência de cruzamentos aleatórios nas subpopulações. Quanto maior o valor de FIT, maior é o grau de endogamia.

F_{ST} = Índice de fixação: proporção de variância total contida em uma sub população (S) em relação à variância genética total (T). Mede a redução de heterozigotia de uma subpopulação relativamente à população total, isto é, determina a diferenciação genética entre subpopulações. Reflete o coeficiente de endogamia dentro de subpopulações, em relação ao total, fornecendo a

proporção da divergência genética total que separa as populações em estudo. Varia de zero a um. Quanto maior o valor de FIT, maior é o grau de endogamia.

F = índices de fixação: se todas as subpopulações estão em equilíbrio de Hardy-Weinberg com as mesmas frequências alélicas $F \approx 0$.

PI = Probabilidade de inclusão ou de identidade: fornece uma estimativa da probabilidade de que dois indivíduos não relacionados, extraídos aleatoriamente na população, por acaso tenham o mesmo genótipo multilocus.

TPI = Probabilidade típica de parentesco (ou paternidade): fornece uma estimativa da probabilidade média de que dois indivíduos não aparentados, extraídos de acasalamentos aleatórios na população, por acaso tenham o mesmo genótipo.

PE = Probabilidade de exclusão: fornece uma estimativa da probabilidade de que duas pessoas extraídas aleatoriamente na população, por acaso, não tenham o mesmo genótipo.

A hipótese do presente estudo é que a distribuição alélica dos voluntários da pesquisa seja a mesma encontrada por Aguiar e colaboradores (2012) e que os parâmetros estatísticos de interesse forense sejam compatíveis com os encontrados por Hill e colaboradores (2013) e Ribeiro e colaboradores (2013).

2 JUSTIFICATIVA

2 JUSTIFICATIVA

Considerando as necessidades de Identificação Humana (HID) nos casos abaixo, justifica-se a realização de um trabalho que visa estabelecer este tipo de serviço na FOB/USP após as devidas etapas de credenciamento do laboratório.

- Casos forenses – correspondência entre suspeito de crime e evidência.
- Testes de paternidade – identificar o pai; troca em maternidades; rapto; tráfico.
- Pessoas desaparecidas.
- Imigrações.
- Investigações históricas.
- Desastres em massa – para identificar partes de corpos e juntá-los.
- Banco de dados de DNA de criminosos (USA/ Europa).

Os BD de DNA podem ser referentes à população em geral (populacional), ou procedentes de criminosos, suspeitos, acusados ou ainda de pessoas desaparecidas. Neste trabalho, tratamos exclusivamente de BD populacional para conhecimento prévio do perfil genético regional. Como um BD da população regional pode ser semelhante ao nacional ou internacional, mas também pode conter diferenças que devem ser contempladas em casos forenses, são de fundamentais importâncias a criação, a complementação e a manutenção de BD de DNA da população, tanto o nacional (já em processo) quanto os regionais, contendo frequências alélicas das populações, índices de mutações, novos alelos identificados, além de ser de grande importância para a ciência biológica como um todo.

Além disso, deve-se destacar que trabalhos como este podem contribuir com a ampliação do quadro de pesquisadores nessa área, contribuir com a rede RIBPG no Brasil e tornar mais ágeis os processos judiciais na região de Bauru/SP.

3 OBJETIVOS

3 OBJETIVOS

O principal objetivo desta pesquisa está embasado na proposta pioneira e ousada de criar um banco de dados de perfis STR populacional, representativo da região de Bauru/SP, para atender a finalidades forenses. Deste modo, as propostas são:

- Iniciar o protótipo de Banco de Dados de DNA populacional para posteriores usos em casos forenses de Bauru/SP e região.
 - Estudar as frequências alélicas de 15 marcadores autossômicos STR incluídos no CODIS (D8S1179, D21S11, D7S820, CSF1PO, D3S1358, TH01, D13S317, D16S539, D2S1338, D19S433, vWA, TPOX, D18S51, D5S818, FGA) e o marcador de gênero amelogenina (AMEL) em alunos de graduação da Faculdade de Odontologia de Bauru/SP, não aparentados, comparando com dados de artigo publicado em amostra de mais de 100.000 brasileiros (AGUIAR *et al.*, 2012).
 - Comparar a distribuição alélica e os parâmetros estatísticos de genética de populações aplicáveis a finalidades forenses das STR dos voluntários desta pesquisa com dados de amostra de população americana referência em kits de identificação (HILL *et al.*, 2013) e com amostra de população portuguesa (RIBEIRO *et al.*, 2013).
 - Analisar eventuais características genéticas distintas referentes a estes marcadores na região, por exemplo, a existência de alelos raros e microvariantes na população estudada.
-

4 MATERIAIS E MÉTODOS

4 MATERIAIS E MÉTODOS

O presente estudo foi aprovado pelo CEP da FOB/USP, sob o número de protocolo 1.116.974, de 22/06/2015.

Para fins de esclarecimentos frente aos questionamentos éticos, é de fundamental importância ressaltar que nos bancos de dados não são arquivados nomes, números de documentos ou quaisquer outros dados que permitam identificar as pessoas, mas somente códigos; os registros são baseados em STR que não revelam características físicas, comportamentais e nem de saúde das pessoas, exceto a determinação genética de gênero, conforme determinam as normas constitucionais nacionais, do CEP (Comitê de Ética em Pesquisa) e da CONEP (Comissão Nacional de Ética em Pesquisa), e internacionais sobre direitos humanos, genoma humano e dados genéticos.

Enfatizando que conforme a alínea “b” do item IV.6 da Resolução 466/12 do CNS, ficou particularmente garantida a liberdade do consentimento para aqueles participantes de pesquisa que, embora plenamente capazes, estejam expostos a condicionamentos específicos, ou à influência de autoridade, caracterizando situações passíveis de limitação da autonomia, como estudantes, militares, empregados, presidiários e internos em centros de readaptação, em casas-abrigo, asilos, associações religiosas e semelhantes, assegurando-lhes inteira liberdade de participar, ou não, da pesquisa, sem quaisquer represálias”, e alínea “d” do item IV.1 - “garantia de plena liberdade ao participante da pesquisa, de recusar-se a participar ou retirar seu consentimento, em qualquer fase da pesquisa, sem penalização alguma”.

4.1 Amostras e extração de DNA

DNA genômico foi extraído de amostras de 200 µL de saliva não estimulada coletadas de 296 doadores voluntários, alunos de graduação maiores de 18 anos da Faculdade de Odontologia de Bauru, que tiveram ciência dos objetivos da pesquisa, firmando sua participação por meio do TCLE.

A coleta da saliva foi realizada nos períodos diurno e vespertino ao término das aulas. No início dos turnos os alunos foram convidados, para de livre e espontânea vontade, participarem da referida pesquisa. Eles permaneceram em sala de aula onde a equipe de

pesquisadores orientou a coleta, levando o material necessário para o procedimento que durou cerca de 15 minutos.

Por ser um procedimento não invasivo, não causou riscos em potencial para o aluno que aceitou fazer parte do estudo, não trazendo nenhum prejuízo ou benefício individual. Houve garantia plena, total e irrestrita de privacidade e sigilo profissional, além de plena liberdade de recusa ou retirada do consentimento em qualquer fase do estudo, sem penalização alguma. Não houve despesas de quaisquer naturezas para os participantes com garantia de indenização diante de eventuais danos decorrentes da pesquisa.

Para a extração de DNA a partir da saliva foi utilizado o QIAamp DNA Mini Kit (número de catálogo 51104, Qiagen®, Alemanha) conforme orientações do fabricante. Descrevendo brevemente, 200 µL da saliva descongelada foram colocados em contato com a solução de lise celular do kit, o que provocou a lise da membrana celular e nuclear expondo o material genético. Em seguida, essa solução foi colocada na coluna de sílica do kit, a qual reteve o DNA e permitiu a passagem dos restos celulares. Para finalizar houve a eluição desse DNA com água ultra-pura em um novo tubo, que foi armazenado a -20°C.

4.2 Genotipagem PCR-STR

A amplificação de DNA genômico, sendo 15 locos STR autossômicos polimórficos que variam em polimorfismo entre 8 e 28 alelos e o marcador de gênero AMEL, foi feita pelo método de PCR multiplex, utilizando o sistema AmpFℓSTR® Identifiler® PCR Amplification Kit (Life Technologies, Foster City, CA, USA), conforme protocolo do fabricante. Este Kit foi validado para tal finalidade de acordo com as normas SWGDAM e diretrizes do FBI. Os 16 locos amplificados contemplaram o núcleo CODIS e são os seguintes: D8S1179, D21S11, D7S820, CSF1PO, D3S1358, TH01, D13S317, D16S539, D2S1338, D19S433, vWA, TPOX, D18S51, D5S818, FGA e AMEL. Foram seguidos na íntegra os protocolos do guia do usuário desse kit (Life Technologies Corporation, 2012).

A tabela a seguir (Tabela 1) mostra os locos amplificados com as sequências repetidas (SR), suas localizações cromossômicas, os alelos contidos na escada alélica, os corantes correspondentes ao marcador fluorescente e o genótipo do DNA controle AmpFℓSTR® 9947A. A escada AmpFℓSTR® Identifiler® Allelic é utilizada para a comparação com as amostras analisadas.

Tabela 1: Composição do KIT AmpFlSTR® Identifiler® (Life Technologies, Foster City, CA, USA)

Locos/ SR	Localização	Alelos incluídos no AmpFlSTR® Identifiler® Allelic Ladder	Corantes	DNA Controle 9947A
D8S1179 TCTA/TCTG	8q24.13	8, 9 10, 11, 12, 13, 4, 15, 16, 17, 18,19	6-FAM™	13†
D21S11 TCTA/TCTG	21q11.2-q21	24, 24.2, 25, 26, 27, 28, 28.2, 29, 29.2, 30, 30.2, 31, 31.2, 32, 32.2, 33, 33.2, 34, 34.2, 35, 35.2, 36, 37, 38		30‡
D7S820 GATA	7q11.21-22	6, 7, 8, 9, 10, 11, 12, 13, 14, 15		10, 11
CSF1PO TAGA	5q33.3-34	6, 7, 8, 9, 10, 11, 12, 13, 14, 15		10,12
D3S1358 TCTG/TCTA	3p	12, 13, 14, 15, 16, 17, 18, 19	VIC 6™	14, 15
TH01 AATG	11p15. 5	4, 5, 6, 7, 8, 9, 9.3, 10, 11, 13.3		8, 9.3
D13S317 TATC	13q22-31	8, 9, 10, 11, 12, 13, 14, 15		11§
D16S539 GATA	16q24-qter	5, 8, 9, 10, 11, 12,13, 14, 15		11, 12
D2S1338 TGCC/TTCC	2q35-37.1	15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28	NED™	19, 23
D19S433 AAGG/TAGG	19q12-13.1	9, 10, 11, 12, 12.2, 13, 13.2, 14, 14.2, 15, 15.2, 16, 16.2, 17, 17.2		14, 15
vWA TCTG/TCTA	12p12-pter	11,12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24		17, 18
TPOX GAAT	2p23-2per	6,7, 8, 9, 10, 11, 12, 13		8††
D18S51 AGAA	18q21.3	7, 9, 10, 10.2, 11, 12, 13, 13.2, 14, 14.2, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27	PET®	15, 19
AMEL X	X: p22.1-22.3	X		X
AMEL Y	Y: p11.2	Y		
D5S818 AGAT	5q21-31	7, 8, 9, 10, 11, 12, 13, 14, 15, 16		11‡‡
FGA CTTT	4q28	17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 26.2, 27, 28, 29, 30, 30.2, 31.2, 32.2, 33.2, 42.2, 43.2, 44.2, 45.2, 46.2, 47.2, 48.2, 50.2, 51.2		23, 24

† Para fins CODIS, perfil relatado como 13, 13.

‡ Para fins CODIS, o perfil relatado como 30, 30.

§ Para fins CODIS, perfil classificado como 11, 11.

†† Para fins CODIS, o perfil relatado como 8, 8.

‡‡ Para fins CODIS, o perfil relatado como 11, 11.

4.3 Amplificação

As reações de amplificação por PCR foram preparadas num volume final de 25 µL contendo 10 µL de AmpFℓSTR® Identifiler® STR Plus Master Mix com AmpliTaq Gold DNA polimerase, 5 µL de AmpFℓSTR® Identifiler® STR Plus Primer Set e um volume máximo de 10 µL de DNA alvo na concentração final de 1 ng. As amostras foram amplificadas em placas de 96 poços (Life Technologies, Foster City, CA, USA) no GeneAmp Veriti (Life Technologies, Foster City, CA, USA).

As condições de ciclos térmicos padrão constituem em: ativação da enzima a 95°C durante 11 minutos; 30 ciclos de desnaturação a 94°C durante 20 segundos e anelamento a 59°C por 1 minuto; e extensão a 72°C por 1 minuto. O último passo é realizado a 60°C durante 30 minutos para estabilizar o produto resultante desse processo, denominado produto da PCR, mantido a 4°C até o próximo passo.

4.4 Eletroforese capilar

Para a separação e identificação dos fragmentos obtidos, foi utilizada a metodologia de eletroforese capilar, conhecida como PCR-CE (PCR com eletroforese capilar). Foram seguidos os protocolos do Guia do Usuário do Kit AmpFℓSTR® Identifiler® PCR Amplification Kit (Life Technologies, Foster City, CA, USA). As amostras foram preparadas em placas de 96 poços por adição de 1 µL do produto da PCR para 9 µL de solução de formamida / LIZ® (0,3 µL de GeneScan™ 600 LIZ® de tamanho padrão e 8,7 µL de formamida deionizada Hi-DI™; Applied Biosystems).

A cada 14 amostras aplicadas, foi acrescentada uma escada alélica de fragmentos com tamanhos conhecidos e mais um controle positivo de DNA (9947A) por placa (Figura 3) que são necessários para obtenção de informações precisas e garantia da correta genotipagem por comparação. As amostras foram desnaturadas a 95°C durante 5 minutos, em seguida acondicionadas no gelo durante 5 minutos.

Utilizando o sequenciador Applied Biosystems 3130xl Genetic Analyzer, os produtos da PCR foram injetados por meio dos 16 capilares de 36 cm a 3 kV durante 10 segundos. A seguir, foram submetidos à eletroforese a 15 kV durante 15 segundos em polímero otimizado-4 (POP-4™ Applied Biosystems) a 60°C.

Figura 3: Perfil alélico do DNA 9947A controle positivo do AmpF ℓ STR® Identifiler® PCR Amplification Kit (Life Technologies, Foster City, CA, USA).

Ao passar por um feixe de laser que excitou os fluoróforos (FAN, VIC, NED e PET), a luz emitida foi detectada. A informação foi processada pelo programa de computador (software ABI Prism 3100 Data Collection), gerando dados brutos convertidos em um gráfico do sinal analítico em função do tempo, denominado eletroferograma. A análise dos eletroferogramas e as designações dos perfis alélicos foram estabelecidas por meio de processamento com o programa de análise GeneMapper ID-X v.1.1.2 (Applied Biosystems), no Centro de Treinamento da Thermo Fisher Scientific do Brasil Ltda, de acordo com as recomendações da Comissão da Sociedade Internacional de Genética Forense (ISFG) para a nomenclatura de STR humano.

Para a correta atribuição de picos aos perfis, de acordo com o preconizado pelo “Interpretation Guidelines for Autosomal STR Typing by Forensic DNA” (SWGDAM, 2010), além dos parâmetros de qualidade do próprio programa GeneMapper ID-X, foram consideradas e analisadas as seguintes características: stutters (pico alélico anterior), off-ladder e off-scale (picos fora da escada e da escala) e os picos –A e A. Quando um marcador mostrou picos desbalanceados, o pico mais alto foi considerado como válido. Quanto à presença de picos fora da escada, as amostras foram reanalisadas e amplificações foram repetidas, conforme a exigência do caso. Picos cujos resultados não se mantiveram, foram excluídos.

4.5 Análise Estatística

As análises dos resultados obtidos na identificação dos 16 locos gênicos foram calculadas para cada marcador em comparação com outras amostras populacionais. Para a comparação das frequências alélicas foi escolhida uma amostra bastante representativa da população brasileira contendo as frequências verificadas em mais de 100.000 pessoas (AGUIAR *et al.*, 2012). Foi aplicado o teste goodness of fit χ^2 index (GFI), para verificar a adequação dos dados, sendo que o valor de p próximo de 1,0 indicou similaridade entre os grupos analisados.

Sendo assim, foram realizadas as demais análises de genética de populações, especialmente as de interesse forense. Para estas análises comparativas são necessários dados brutos de identificação dos alelos. Por este motivo, foram escolhidas duas outras populações cujos dados estão disponíveis como suplemento eletrônico: HILL, com 1036 pessoas (HILL *et al.*, 2013) e RIB, com 302 pessoas (RIBEIRO *et al.*, 2013).

Para facilidade de denominação das amostras nos programas utilizados e para codificação das populações comparadas, foram utilizadas as seguintes siglas: FOB = amostra de 296 alunos de graduação da FOB/USP, HILL = amostra de 1036 pessoas de origem americana estudadas por HILL e colaboradores (HILL *et al.*, 2013) e RIB = amostra de 302 pessoas de origem portuguesa estudadas por Ribeiro e colaboradores (RIBEIRO *et al.*, 2013).

As frequências alélicas, heterozigosidade observada (HO), heterozigosidade esperada (HE), equilíbrio de Hardy-Weinberg (HWE) e outros parâmetros estatísticos de interesses forenses como o poder de exclusão (PE), índice de paternidade típica (TPI) e índice de fixação (F) foram analisados utilizando o programa GenAlEx 6.5 (PEAKALL e SMOUSE, 2006; PEAKALL e SMOUSE, 2012). O coeficiente de consanguinidade (Fis), coeficiente de consanguinidade total (Fit) e índice de fixação total (Fst), bem como os testes AMOVA e de pareamento (pairwise) contemplando as análises de variação entre as pessoas, dentro de cada população e entre as três populações estudadas, o respectivo índice de fixação e a distância molecular entre as três populações foram calculados pelo programa Arlequin, versão 3.5 (EXCOFFIER e LISCHER, 2010).

Os praticantes de genética forense usam banco de dados de população com informações relevantes de frequências alélicas de genes na população em geral para as comparações e cálculos estatísticos de probabilidades do genótipo em questão. Em

populações endogâmicas, os indivíduos são mais propensos a compartilhar perfis de DNA, diferente de indivíduos não aparentados. Esta é uma questão que deve ser levada em conta para a realização de cálculos e sua correta interpretação. Frequentemente, uma correção é aplicada para acomodar as variações existentes entre subpopulações.

5 RESULTADOS

5 RESULTADOS

Do total de 324 alunos de graduação da FOB/USP, sendo 194 do curso de Odontologia e 130 do curso de Fonoaudiologia, houve adesão para participar da pesquisa de 296, ou seja, 91,35%. Quanto ao gênero, do total de alunos, havia 261 femininos e 63 masculinos.

Eletroferogramas das 296 amostras de DNA extraído de salivas dos alunos foram analisados quanto à presença de picos relativos aos 16 marcadores (15 locos STR e o loco AMEL), perfazendo um total de 9.472 alelos. Destes, 9.307 foram identificados e 164 foram considerados alelos brancos, representados pela falta de algum pico. A presença de um único pico pode significar homozigose ou um alelo ainda não identificado e não contemplado pelo kit. Dois picos alélicos indicam heterozigose. Para a finalidade de análises estatísticas, quando pelo menos um alelo foi identificado no loco e não houve indícios técnicos representativos de outras inferências, os alelos brancos foram considerados homozigose e duplicados, por exigência dos programas utilizados. Houve uma exceção: um pico tecnicamente bem definido, fora da escada padrão e da escala (off-ladder e off-scale), que por tratar de caso raro, será apresentado no final.

A Figura 4 ilustra o eletroferograma com o perfil genético dos 15 STR e da AMEL de uma pessoa do gênero feminino (amostra 69) para os diferentes locos analisados. A barra verde superior a cada loco indica a qualidade ideal do experimento.

Figura 4: Eletroferograma do perfil genético STR e AMEL da amostra 69.

Os perfis genéticos dos alunos estudados foram organizados por meio da utilização do programa Microsoft Excel 2010. A Tabela 2 mostra, para cada loco analisado, a faixa de variação dos alelos, os números de alelos obtidos em cada loco, os alelos mais e os menos frequentes e a distribuição de gênero identificada pelo loco AMEL. Os locos mais polimórficos foram D18S51 (17 alelos) e FGA (15 alelos) e os menos polimórficos foram D16S539 e THO1 (7 alelos cada).

Tabela 2: Faixa de variação dos alelos, números de alelos encontrados, os alelos mais e menos frequentes em cada marcador genético e o marcador de gênero AMEL.

Locos	Faixa de variação dos Alelos	N ^{os} de Alelos	Alelos mais Frequentes		Alelos menos Frequentes	
			Alelo	Frequência	Alelo	Frequência
D8S1179	8 a 17	10	13	0,2755	17	0,0075
			14	0,2302	11	0,0774
D21S11	23.2 a 35	13	30	0,2434	23.2	0,0019
			29	0,2004	35	0,0037
D7S820	7 a 14	8	10	0,2589	14	0,0075
			11	0,2477	7	0,0094
CSF1PO	7 a 15	8	11	0,3160	15	0,0019
			12	0,2852	7	0,0077
D3S1358	12 a 20	9	15	0,2619	18.2	0,0019
			16	0,2486	20	0,0019
TH01	5 a 11	7	9.3	0,2237	5	0,0018
			6	0,2181	11	0,0037
D13S317	8 a 15	8	11	0,2839	15	0,0019
			12	0,2597	14	0,0390
D16S539	8 a 14	7	11	0,3141	8	0,0231
			12	0,2466	14	0,0270
D2S1338	14 a 27	13	17	0,2359	14	0,0018
			20	0,1234	27	0,0091
D19S433	10 a 18.2	9	14	0,2987	10	0,0019
			13	0,2430	11	0,0019
vWa	13 a 21	9	16	0,2524	13	0,0038
			17	0,2260	21	0,0038
TPOX	5 a 13	9	8	0,4747	5	0,0020
			11	0,2424	13	0,0040
D18S51	9 a 30	17	15	0,1652	9	0,0018
			13	0,1436	10.2	0,0036
D5S818	7 a 15	9	12	0,3672	15	0,0020
			11	0,3066	7	0,0059
FGA	17.2 a 32.2	15	22	0,1817	17.2	0,0036
			24	0,1601	28	0,0036
AMEL	X/X	2	X	0,7973		
	X/Y	2			Y	0,2027

As tabelas 3 a 15 representam as comparações entre as frequências alélicas de cada um de 13 locos analisados nos 296 alunos da FOB/USP comparados com a amostra variada entre 112.695 a 137.136 brasileiros descritos por Aguiar e colaboradores (2012), por meio do teste goodness of fit X^2 index (GFI) e seus respectivos valores de p, que para todos os locos foi $p=0,99$. Foram comparados apenas os 13 locos estudados em comum na amostra FOB/USP e na amostra de Aguiar e colaboradores (2012) que foram os seguintes: D8S1179, D21S11, D7S820, CSF1PO, D3S1358, TH01, D13S317, D16S539, vWA, TPOX, D18S51, D5S818, FGA, faltando no trabalho de Aguiar e colaboradores (2012) o estudo dos locos D2S1338 e D19S433, que por este motivo não puderam ser comparados.

Tabela 3: Comparações entre as frequências alélicas para o loco D8S1179

D8S1179		
Alelos	296 alunos FOB/USP	134.211 Brasileiros (Aguiar 2012)
8	0,019	0,008
9	0,011	0,009
10	0,092	0,068
11	0,077	0,077
12	0,132	0,130
13	0,275	0,279
14	0,230	0,251
15	0,119	0,138
16	0,036	0,034
17	0,008	0,004

p = 0,99

Tabela 4: Comparações entre as frequências alélicas para o loco D21S11

D21S11		
Alelos	296 alunos FOB/USP	130.014 Brasileiros (Aguiar 2012)
26.2	0,0019	0,00003
27	0,0225	0,0300
28	0,1423	0,1594
29	0,2004	0,2081
30	0,2434	0,2317
30.2	0,0412	0,0284
31	0,0674	0,0769
31.2	0,1049	0,0899
32	0,0150	0,0285
32.2	0,1030	0,0751
33.2	0,0393	0,0284
34.2	0,0075	0,0038
35	0,0037	0,0064

p = 0,99

Tabela 5: Comparações entre as frequências alélicas para o loco D7S820

D7S820		
Alelos	296 alunos FOB/USP	137.136 Brasileiros (Aguiar 2012)
7	0,0094	0,0150
8	0,1689	0,1571
9	0,1163	0,1205
10	0,2589	0,2750
11	0,2477	0,2364
12	0,1632	0,1614
13	0,0281	0,0304
14	0,0075	0,0038
p = 0,99		

Tabela 6: Comparações entre as frequências alélicas para o loco CFS1PO

CFS1PO		
Alelos	296 alunos FOB/USP	113.526 Brasileiros (Aguiar 2012)
7	0,0077	0,0134
8	0,0173	0,0200
9	0,0270	0,0329
10	0,2678	0,2624
11	0,3160	0,2936
12	0,2852	0,3025
13	0,0771	0,0637
15	0,0019	0,0012
p = 0,99		

Tabela 7: Comparações entre as frequências alélicas para o loco D3S1358

D3S1358		
Alelos	296 alunos FOB/USP/USP	135.133 Brasileiros (Aguiar 2012)
12	0,0057	0,0024
13	0,0038	0,0042
14	0,1195	0,0932
15	0,2619	0,2910
16	0,2486	0,2703
17	0,2011	0,2033
18	0,1404	0,1239
19	0,0152	0,0099
20	0,0019	0,0008
p = 0,99		

Tabela 8: Comparações entre as frequências alélicas para o loco TH01

TH01		
Alelos	296 alunos FOB/USP	112.695 Brasileiros (Aguar 2012)
5	0,0018	0,0016
6	0,2181	0,2112
7	0,2015	0,2405
8	0,1386	0,1494
9	0,1719	0,1626
9.3	0,2237	0,2262
10	0,0185	0,0080
11	0,0037	0,0005
p = 0,99		

Tabela 9: Comparações entre as frequências alélicas para o loco D13S317

D13S317		
Alelos	296 alunos FOB/USP	137.161 Brasileiros (Aguar 2012)
8	0,1373	0,0998
9	0,0909	0,0817
10	0,0668	0,0526
11	0,2839	0,2952
12	0,2597	0,2933
13	0,1206	0,1264
14	0,0390	0,0482
15	0,0019	0,0015
p = 0,99		

Tabela 10: Comparações entre as frequências alélicas para o loco D16S539

D16S539		
Alelos	296 alunos FOB/USP	136.942 Brasileiros (Aguar 2012)
8	0,0231	0,0241
9	0,1445	0,1603
10	0,0559	0,0884
11	0,3141	0,2887
12	0,2466	0,2560
13	0,1888	0,1545
14	0,0270	0,0250
p = 0,99		

Tabela 11: Comparações entre as frequências alélicas para o loco vWA

vWA		
Alelos	296 alunos FOB/USP	127.415 Brasileiros (Aguiar 2012)
13	0,0038	0,0058
14	0,1205	0,0844
15	0,1337	0,1380
16	0,2524	0,2600
17	0,2260	0,2548
18	0,1733	0,1702
19	0,0716	0,0657
20	0,0151	0,0141
21	0,0038	0,0022
p = 0,99		

Tabela 12: Comparações entre as frequências alélicas para o loco TPOX

TPOX		
Alelos	296 alunos FOB/USP	123.102 Brasileiros (Aguiar 2012)
5	0,0020	0,0006
6	0,0101	0,0177
7	0,0061	0,0084
8	0,4747	0,4559
9	0,1414	0,1241
10	0,0747	0,0673
11	0,2424	0,2755
12	0,0444	0,0481
13	0,0040	0,0022
p = 0,99		

Tabela 13: Comparações entre as frequências alélicas para o loco D5S818

D5S818		
Alelos	296 alunos FOB/USP	135.181 Brasileiros (Aguiar 2012)
7	0,0059	0.017458
8	0,0059	0.017610
9	0,0430	0.038123
10	0,0781	0.061854
11	0,3066	0.329063
12	0,3672	0.346376
13	0,1816	0.174473
14	0,0098	0.013071
15	0,0020	0.001646
p = 0,99		

Tabela 14: Comparações entre as frequências alélicas para o loco D18S51

D18S51		
Alelos	296 alunos FOB/USP	136.739 Brasileiros (Aguar 2012)
9	0,0018	0,0012
10	0,0090	0,0099
10.2	0,0036	0,0006
11	0,0233	0,0108
12	0,1095	0,1211
13	0,1436	0,1085
14	0,1364	0,1472
15	0,1652	0,1522
16	0,1329	0,1445
16.2	0,0018	0,0000
17	0,1113	0,1219
18	0,0682	0,0776
19	0,0395	0,0486
20	0,0287	0,0291
21	0,0126	0,0130
22	0,0054	0,0055
24	0,0036	0,0008
25	0,0018	0,0004
p = 0,99		

Tabela 15: Comparações entre as frequências alélicas para o loco FGA

FGA		
Alelos	296 alunos FOB/USP	135.839 Brasileiros (Aguar 2012)
17.2	0,0036	0,00003
18	0,0072	0,0093
19	0,0540	0,0754
20	0,1277	0,1139
21	0,1439	0,1511
22	0,1817	0,1617
22.2	0,0036	0,0024
23	0,1565	0,1493
23.2	0,0036	0,0011
24	0,1601	0,1509
24.2	0,0054	0,0003
25	0,0845	0,1034
25.2	0,0018	0,0002
26	0,0468	0,0474
27	0,0126	0,0160
28	0,0036	0,0051
31.2	0,0018	0,0001
32.2	0,0018	0,0002
p = 0,99		

A Tabela 16 representa a distribuição das frequências alélicas dos 296 alunos da FOB/USP, de acordo com os marcadores de 15 locos STR e amelogenina, bem como os resultados estatísticos de parâmetros de genética de populações muito utilizados na análise de casos forenses, feitos por meio dos programas GenAlEx 6.5 e Arlequin 3.5.

Para a facilidade de nomenclatura no texto, nas tabelas e figuras, a amostra de 296 alunos da FOB/USP de Bauru analisada neste estudo, daqui para frente, será simplesmente denominada amostra FOB.

Os parâmetros analisados foram: N: tamanho amostral; NA: número de alelos; HO: heterozigosidade observada; HE: heterozigosidade esperada; I: índice de Shannon; HWE: equilíbrio de Hardy-Weinberg; PE: probabilidade de exclusão; TPI: índice de paternidade típico. Estão em destaques coloridos os dados que chamaram a atenção e que serão discutidos.

[illegible][illegible]

32	0,117														0,002
33	0,034														
34	0,007														
35	0,003														
36	0,002														
37	0,000														
38	0,000														
39	0,000														
43															0,000
44															0,002
X															0,797
Y															0,203
NA	10	13	8	8	9	7	8	7	13	9	9	9	17	9	15
HO	0,787	0,814	0,801	0,753	0,787	0,828	0,821	0,750	0,861	0,824	0,791	0,672	0,885	0,726	0,882
HE	0,813	0,808	0,798	0,736	0,791	0,722	0,796	0,768	0,871	0,748	0,811	0,656	0,881	0,715	0,862
I	1,878	1,822	1,707	1,47	1,665	1,43	1,743	1,614	2,234	1,511	1,782	1,341	2,283	1,444	2,11
HWE	0,102	0,000	0,917	0,934	0,256	0,000	0,720	0,135	0,731	0,844	0,954	0,900	0,460	0,804	0,998
PE	0,463	0,447	0,423	0,321	0,409	0,309	0,428	0,377	0,591	0,341	0,450	0,249	0,610	0,302	0,561
TPI	0,058	0,063	0,071	0,117	0,076	0,121	0,070	0,089	0,029	0,105	0,062	0,162	0,026	0,129	0,035

NA: número de alelos; HO: heterozigosidade observada; HE: heterozigosidade esperada; I: índice de Shannon; HWE: equilíbrio de Hardy-Weinberg; PE: probabilidade de exclusão; TPI: índice de paternidade típico.

FOB = amostra de 296 alunos de graduação da FOB/USP.

Em destaque colorido, os dados representativos de valores máximos ou mínimos.

Na Figura 5, estão representadas as comparações da distribuição da frequência alélica de 15 locos STR entre três populações denominadas FOB, HILL, RIB feitas e copiadas do programa GenAlEx 6.5.

Figura 5: Distribuição da frequência alélica dos 15 locos entre as populações: FOB, HILL, RIB.

FOB = amostra de 296 alunos de graduação da FOB/USP, HILL = amostra de 1036 pessoas de origem americana estudadas por HILL e colaboradores (HILL *et al.*, 2013) e RIB = amostra de 302 pessoas de origem portuguesa estudadas por Ribeiro e colaboradores (Ribeiro *et al.*, 2013).

A Tabela 17 representa o teste exato de X^2 para o equilíbrio de Hardy-Weinberg, mostrando o grau de liberdade (DF), o qui-quadrado e os valores de p para cada loco da amostra FOB, assumindo nível de significância de $p < 0,05$. *D21S11 e TH01 apresentaram p significativo.

Tabela 17: Equilíbrio de Hardy-Weinberg X^2 de acordo com as frequências alélicas de 15 STR na amostra FOB.

	D8S1179	D21S11	D7S820	CSF1PO	D3S1358	TH01	D13S317	D16S539	D2S1338	D19S433	vWa	TPOX	D18S51	D5S818	FGA
DF	45	78	28	28	36	21	28	21	78	36	36	36	136	36	105
X^2	57,375	637,591	18,333	17,690	41,140	49,517	23,268	28,198	69,947	27,516	23,031	25,659	137,001	28,642	67,404
Pb	0,102	0,000 *	0,917	0,934	0,256	0,000 *	0,720	0,135	0,731	0,844	0,954	0,900	0,460	0,804	0,998

DF: grau de liberdade; X^2 : qui quadrado; Pb: probabilidade - * significativo

FOB = amostra de 296 alunos de graduação da FOB/USP

A Tabela 18 representa a comparação da diversidade genética medida pelo coeficiente de consanguinidade nas três populações. (FOB, HILL, RIB).

Tabela 18: Análise da diversidade genética medida pelo coeficiente de consanguinidade das populações: FOB, HILL e RIB

	D8S1179	D21S11	D7S820	CSF1PO	D3S1358	TH01	D13S317	D16S539	D2S1338	D19S433	vWa	TPOX	D18S51	D5S818	FGA
Fis	0,034	-0,026	-0,002	-0,014	0,017	-0,086	0,014	0,022	0,013	-0,070	-0,002	-0,015	0,003	-0,006	-0,026
Fit	0,037	-0,025	0,000	-0,011	0,019	-0,078	0,016	0,026	0,017	-0,068	0,000	-0,014	0,004	-0,004	-0,024
Fst	0,003	0,001	0,002	0,003	0,001	0,007	0,002	0,004	0,004	0,001	0,002	0,002	0,001	0,001	0,002

Fis: Coeficiente de consanguinidade; Fit: Coeficiente de consanguinidade total; Fst: Índice de fixação total. FOB = amostra de 296 alunos de graduação da FOB/USP, HILL = amostra de 1036 pessoas de origem americana estudadas por HILL e colaboradores (HILL *et al.*, 2013) e RIB = amostra de 302 pessoas de origem portuguesa estudadas por Ribeiro e colaboradores (Ribeiro *et al.*, 2013).

A Tabela 19 representa a comparação da variância molecular entre as três populações e dentro de cada população estudada.

Tabela 19: Comparação da variância molecular entre as três populações e dentro de cada população: FOB, HILL e RIB.

	D8S1179	D21S11	D7S820	CSF1PO	D3S1358	TH01	D13S317	D16S539	D2S1338	D19S433	vWa	TPOX	D18S51	D5S818	FGA
Variação entre as populações	0,001	0,000	0,001	0,001	0,001	0,006	0,001	0,002	0,002	0,001	0,000	0,001	0,000	0,000	0,001
Variação na população FOB	0,409	0,422	0,400	0,370	0,388	0,393	0,394	0,390	0,444	0,402	0,406	0,339	0,441	0,364	0,434
Fator de fixação	0,032	-0,008	-0,003	-0,024	0,005	-0,146	-0,032	0,024	0,01	-0,103	0,03	-0,025	-0,005	-0,016	-0,02

FOB = amostra de 296 alunos de graduação da FOB/USP, HILL = amostra de 1036 pessoas de origem americana estudadas por HILL e colaboradores (HILL *et al.*, 2013) e RIB = amostra de 302 pessoas de origem portuguesa estudadas por Ribeiro e colaboradores (Ribeiro *et al.*, 2013).

Na Figura 6, está representada graficamente a heterozigosidade individual dos 296 alunos da FOB/USP, feitas pelo programa GenAlEx 6.5. O índice de heterozigosidade observada foi de 96,5%.

Figura 6: Heterozigosidade individual para 15 marcadores STR na amostra FOB.
FOB = amostra de 296 alunos de graduação da FOB/USP.

A Figura 7 representa a probabilidade de exclusão (PE) de amostras incompatíveis para os 15 marcadores STR utilizados, comparados nas três populações denominadas FOB, HILL e RIB, com seus valores reais médios e os valores máximos que podem atingir. A partir de 6 a 9 locos mais polimórficos analisados, é possível excluir 100% das falsas comparações. A ilustração abaixo foi feita e copiada na íntegra do programa GenAlEx 6.5.

Figura 7: Probabilidade de exclusão de acordo com as combinações de 15 locos em três populações comparadas (FOB, HILL, RIB).

P2Exc = probabilidade de exclusão, P2Exc FOB = probabilidade de exclusão na amostra de 296 alunos de graduação da FOB/USP, P2Exc HILL = probabilidade de exclusão na amostra de 1036 pessoas de origem americana estudadas por HILL e colaboradores (HILL *et al.*, 2013) e P2Exc RIB = probabilidade de exclusão na amostra de 302 pessoas de origem portuguesa estudadas por Ribeiro e colaboradores (Ribeiro *et al.*, 2013).

P2Exc1 Max FOB = probabilidade de exclusão máxima na amostra de 296 alunos de graduação da FOB/USP, P2Exc1 Max HILL = probabilidade de exclusão máxima na amostra de 1036 pessoas de origem americana estudadas por HILL e colaboradores (HILL *et al.*, 2013) e P2Exc1Max RIB = probabilidade de exclusão máxima na amostra de 302 pessoas de origem portuguesa estudadas por Ribeiro e colaboradores (Ribeiro *et al.*, 2013).

A Figura 8 representa a análise de variância molecular (AMOVA) entre as três populações denominadas FOB, HILL e RIB (0%), entre pessoas da mesma população (1%) e para cada pessoa de cada população (99%), feita e copiada na íntegra do programa Arlequin 3.5.

Figura 8: Porcentagens de variação molecular entre as populações FOB, HILL e RIB (AMOVA).

FOB = amostra de 296 alunos de graduação da FOB/USP, HILL = amostra de 1036 pessoas de origem americana estudadas por HILL e colaboradores (HILL *et al.*, 2013) e RIB = amostra de 302 pessoas de origem portuguesa estudadas por Ribeiro e colaboradores (Ribeiro *et al.*, 2013)

A Figura 9 ilustra o eletroferograma obtido no programa de análise GeneMapper ID-X, mostrando os 2 picos alélicos (indicados pelas setas azul e preta) verificados no loco D18S51 da amostra de número 194.

Figura 9: Eletroferograma do loco D18S51 da amostra 194.

Interessantemente, esta amostra apresentou um pico (indicado pela seta preta) fora da escada e fora da escala dos tamanhos esperados para o referido loco, representadas pelas barras verticais cinzas e rosas do programa. A escada alélica do kit prevê alelos na faixa entre 7 (tamanho 264) até 27 (tamanho 344).

A Figura 10 ilustra o eletroferograma interpretado da amostra 194, do programa de análise GeneMapper ID-X. A interpretação é a seguinte: dentro da escada alélica, há um pico de tamanho 284 claramente identificado como alelo 12 (seta azul) com seu pico “stutter” 11 (seta amarela). O outro pico, colorido em preto, está claramente relacionado a um alelo de tamanho 354,28, de acordo com o programa (indicados pelas setas pretas). Por tratar-se de repetição de tetranucleotídeo, estima-se que este pico fora da escala seja referente a uma microvariante alélica, possivelmente a 29 (seta preta), corroborado pelo seu pico “stutter” 28 (seta vermelha).

A injeção capilar desta amostra foi repetida, obtendo-se o mesmo resultado. A análise no programa GeneMapper ID-X também foi refeita considerando-se e ajustando todos os parâmetros específicos do programa que permitem interpretação mais acurada e fidedigna, mantendo-se o resultado.

Figura 10: Eletroferograma da amostra 194, loco D18S51: alelo 12 e microvariante do alelo 29 com seus respectivos “stutters”, 11 e 28.

6 DISCUSSÃO

6 DISCUSSÃO

6.1 Por que formar e incrementar os Bancos de Dados?

Desde que o homem tomou conhecimento de si próprio, tem a necessidade de se identificar, de ser reconhecido como um ser único, por várias razões, desde psicológicas, emocionais, até racionais para fins de organização social. Neste último contexto, surgiram as necessidades de reconhecimentos em casos de medicina forense e para a aplicação da justiça quando houver demanda.

Houve toda uma evolução histórica da identificação da pessoa por meio de métodos antropométricos biológicos como as impressões digitais até que o desenvolvimento e revolução tecnológica permitiram a identificação laboratorial de uma pessoa, passando por testes de grupos sanguíneos, sistema HLA, odontologia forense e marcadores genéticos como os polimorfismos de DNA, muito estudados e utilizados nas últimas décadas.

Os polimorfismos genéticos se estabeleceram e se disseminaram por meio da migração e miscigenação de populações, podendo apresentar variações étnicas e regionais. Em decorrência da miscigenação da população brasileira, vários estudos têm mostrado a variabilidade genética dos marcadores STR e como estes são, atualmente, os mais elegíveis para a identificação humana, especialmente em situações forenses.

Por este motivo, esta pesquisa teve por objetivo estudar os referidos marcadores em alunos de graduação da FOB/USP com a finalidade de iniciar o protótipo de um banco de dados populacionais, visando a posterior criação de banco de dados de criminosos, pessoas desaparecidas, entre outros, para a prestação de serviço forense na região do interior do estado de São Paulo.

Os bancos de dados populacionais têm um papel fundamental por possibilitar a comparação genética de pessoas para as diversas finalidades a que se destinam, não somente casos forenses, mas também aos estudos de genética de populações, evolução, migração, miscigenação, análises de ligação, entre outros.

A análise das frequências alélicas e genótípicas permite inferir o grau de polimorfismo dos microssatélites para caracterizar a variabilidade genética da população. Neste sentido,

vários trabalhos brasileiros contendo este tipo de análise vêm sendo publicados e comparados, caracterizando de certa forma a população brasileira como um todo, conforme descrito na revisão de literatura (DE ASSIS POIARES et al., 2010; WOLFGRAMM EDE et al., 2011; VIEIRA et al., 2013).

No caso de Bauru, há uma população flutuante devido à migração de pessoas por motivos de trabalho e principalmente para estudo, haja vista o grande número de faculdades e universidades presentes na cidade.

Os estudos regionais englobam pessoas que estão em um determinado momento naquela região, mas que podem miscigenar com habitantes locais e, desta maneira, introduzir outros polimorfismos.

A ideia primordial do trabalho é de contribuir com o aumento de banco de dados da população brasileira em geral para as STR mais estudadas e preconizadas pelo CODIS do FBI, isto sem considerar quaisquer tipos de situações forenses, mas para se conhecer algumas características genéticas da população. Os 296 alunos estudados (amostra FOB/USP) representam pessoas residentes na região, ainda que algumas estejam residentes apenas por um período determinado de tempo em Bauru.

O impacto com relação ao objetivo de se formar um banco de dados é discreto devido ao tamanho reduzido da amostra, porém, o encontro de uma possível microvariante alélica nesta pequena amostra, o alelo 29 da D18S51, que deve ser confirmado, mostra de modo enfático a necessidade deste estudo e outros vindouros.

Neste sentido, serão feitos esforços para que estes números sejam aumentados em um futuro próximo e ainda corroborados por sequenciamento e estudos familiares, levando-se em conta toda a dificuldade inerente a este tipo de estudo.

As maiores dificuldades para a realização deste tipo de pesquisa dizem respeito à aprovação ética, ao número elevado de análises para que tenha significado populacional e à publicação que, atualmente, preconiza amostra de no mínimo 500 pessoas. Sem contar o alto custo técnico: kits, equipamentos, programas de análises e até a exigência de pesquisadores qualificados para tal propósito. Vale a pena lembrar que, para trabalhar na área, os profissionais são periodicamente submetidos a testes de proficiência para validade de resultados, devendo estar sempre atualizados e participar de cursos e reuniões científicas.

6.2 Características genéticas da população estudada

Foram analisados 15 locos autossômicos STR e a AMEL, como marcador de gênero, em amostra de 296 alunos de graduação da FOB/USP. Para atingir o objetivo proposto, foi comparada a distribuição alélica de 13 marcadores STR (D8S1179, D21S11, D7S820, CSF1PO, D3S1358, TH01, D13S317, D16S539, vWA, TPOX, D18S51, D5S818, FGA), em comum entre os alunos de graduação voluntários da FOB/USP (amostra FOB) e uma amostra de mais de 100 mil brasileiros previamente publicada (AGUIAR *et al.*, 2012) e ainda foram estudados os parâmetros estatísticos de genética de populações aplicáveis a finalidades forenses por meio da comparação com dados de amostras de população americana de referência em kits de identificação (HILL *et al.*, 2013) e com amostra de população portuguesa (RIBEIRO *et al.*, 2013). Estas populações foram escolhidas por apresentarem os mesmos marcadores e os parâmetros estatísticos que foram propostos.

A garantia de que a distribuição alélica encontrada dentre todos os voluntários deste estudo pudesse ser comparada a outros estudos, independentemente da maioria ser do gênero feminino (79,73%), foi alcançada pelo teste goodness of fit χ^2 index (GFI) entre 61 mulheres selecionadas aleatoriamente com todos os 61 homens da amostra, que provou que não houve nenhuma diferença significativa (GFI = 0,99) que impedisse a continuidade das análises.

A mesma análise foi feita com estas 61 mulheres comparadas as outras 174 restantes da amostra para todos os locos estudados. Como esperado, os resultados revelaram alta similaridade na distribuição alélica destes dois grupos avaliados (GFI = 0,99).

Para finalizar a normalização das amostras, foram agrupados os 61 homens e as 61 mulheres selecionadas aleatoriamente e comparado este novo grupo com toda a amostra (296 pessoas). Os resultados revelaram novamente nenhuma diferença significativa dentre os locos avaliados (GFI = 0,99). Logo, estes resultados garantiram a possibilidade de comparação da distribuição alélica do presente estudo com qualquer outro estudo.

Então, comparando-se as frequências alélicas dos 13 marcadores STR encontradas na amostra FOB/USP com as da população representativa brasileira não se observaram diferenças significativas, indicando que são praticamente iguais uma vez que o valor de p foi de 0,99 para todas as análises.

Uma vez normalizadas as amostras e sendo garantido estatisticamente que a amostra FOB/USP, embora pequena, poderia ser comparada a outras, foram feitos os demais testes estatísticos propostos, pela comparação com amostras de populações americana e portuguesa. Em relação às frequências alélicas dos Estados Unidos e de Portugal, também não houve diferença significativa nas distribuições alélicas. Embora exista uma vasta miscigenação no território brasileiro como um todo, a distribuição de frequências alélicas não variou significativamente entre as populações comparadas.

Quanto ao equilíbrio de Hardy-Weinberg, retratando as avaliações de heterozigose, observou-se que, para quase todos os marcadores, os índices de heterozigosidade esperados situaram-se muito próximos àqueles efetivamente observados, com diferenças não significativas com nível de significância de 5%, com convergência plena entre si, demonstrando uma situação de normalidade de distribuição populacional. Os locos D21S11 e TH01 apresentaram diferenças significativas, que podem ser explicadas pelo fato de alguns alelos menos frequentes destes locos não terem sido encontrados na amostra estudada. Embora estes marcadores tenham mostrado diferença significativa por este teste, com resultado inferior ao nível de significância admitido de 0,05, o que poderia expressar a não confirmação do equilíbrio de Hardy-Weinberg na população, o fato é que os métodos estatísticos (AMOVA e AMOVA com pareamento) foram desenvolvidos para o conjunto dos marcadores, de acordo com os tutoriais dos programas GenAlEx 6.5 e Arlequin 3.5 utilizados. Assim, o nível de significância não deve ser comparado diretamente com os valores p de cada marcador, e sim com o conjunto de marcadores. Logo, a população FOB/USP encontra-se em equilíbrio segundo Hardy-Weinberg.

Dos alelos encontrados, 9.307 estão em heterozigose e 164 são brancos, implicando em homozigose ou alelo não determinado. Os 164 alelos dentro dos padrões de homozigose, suportam a tese de amostra essencialmente heterozigota para o conjunto dos 15 locos autossômicos STR, sendo que algumas pessoas atingiram 100% de heterozigosidade e a maioria apresentou entre 60% a 90%. O índice de heterozigosidade observada foi de 96,5%, estando de acordo com o esperado e refletindo, assim, o equilíbrio de Hardy-Weinberg em que se encontrou a amostra FOB/USP.

Os alelos mais frequentes, portanto menos informativos, foram o alelo 9 do loco TH01 (0,412) e o alelo 8 do loco TPOX (0,517). Vale ressaltar que o loco TH01 e o D16S539, ambos com 7 alelos, foram os menos polimórficos

encontrados na amostra FOB/USP e o D18S51 foi o de maior polimorfismo (17 alelos), seguido do loco FGA (15 alelos).

Os locos D2S1388 (PE = 0,61 e TPI = 0,025986) e D18S51 (PE = 0,591 e TPI = 0,028853) foram os mais discriminatórios e apresentaram a maior probabilidade de exclusão e consequentemente o menor índice de coincidência aleatória, indicando serem estes importantes marcadores genéticos para a identificação humana.

O índice de Shannon, que permite prever sem dúvidas o próximo achado quando próximo de zero, com valores superiores a 1 na amostra FOB/USP, evidenciou que o alto polimorfismo não permite estimar ao acaso a identificação de uma pessoa utilizando estes marcadores.

A alta probabilidade de exclusão (PE) revelou que a partir de 6 a 9 locos mais polimórficos analisados, é possível excluir 99,999999 % das falsas comparações para todas as populações.

A análise de variância molecular (AMOVA) entre as três populações (0%), entre pessoas da mesma população (1%) e para cada pessoa de cada população (99%), mostrou que as pessoas têm uma alta variância individual e esta variância é mantida de forma homogênea entre as pessoas da mesma amostra (região) e entre as populações estudadas.

O presente estudo reforçou a conclusão de outros trabalhos populacionais (GOES *et al.*, 2004; DEL CASTILLO *et al.*, 2009; DE ASSIS POIARES *et al.*, 2010; WOLFGRAMM EDE *et al.*, 2011; AGUIAR *et al.*, 2012; VIEIRA *et al.*, 2013; ALVES *et al.*, 2014; SANTOS *et al.*, 2015) de frequências alélicas dos 15 marcadores STR autossômicos, confirmou o elevado grau polimórfico, a alta heterozigosidade, o elevado grau de discriminação dos marcadores utilizados, a comparabilidade das amostras e pequenas diferenças individuais que podem ser de interesse em casos forenses.

Chamou-nos a atenção a presença de um pico fora da escada padrão e fora da escala de tamanhos inferidos pelo do kit. Na amostra de número 194, loco D18S51, a escada alélica do kit prevê variação alélica entre 7 a 27, porém foi verificado um pico claramente relacionado a um alelo com cerca de 10 pares de bases a mais do que o alelo 27. Por tratar-se de repetição de tetranucleotídeo, estima-se que seja referente a uma microvariante do alelo 29, ainda não bem definido nas populações, não contemplado nos kits e nem na base de dados STRbase da Applied Genetics Group of the National Institute of Standards and Technology, atualizada em

agosto de 2015 e acessada via <http://www.cstl.nist.gov/biotech/strbase>. Para a comparação, algumas microvariantes já registradas que ladeiam esta possível microvariante encontrada neste trabalho são apresentadas em forma de anexo no final da tese.

Para a confirmação desta hipótese, são indicados o sequenciamento e as genotipagens de pais, irmãos ou filhos (se houver) para se verificar o comportamento padrão familiar. Caso na família não haja este padrão, pode ser postulada a tese de duplicação de tetranucleotídeo. Confirmada esta microvariante, este trabalho trará uma contribuição bastante significativa para a atualização da STRBase.

Embora a amostra seja limitada, vale ressaltar que 91,35% dos alunos matriculados na FOB/USP/Bauru participou do estudo, indicando um alto índice de adesão e colaboração com a pesquisa científica por parte desta comunidade.

7 CONCLUSÃO

7 CONCLUSÃO

Existe semelhança na distribuição de frequências alélicas dos 15 marcadores STR autossômicos entre as populações comparadas. Além disso, foram confirmados o elevado grau de polimorfismo, a alta heterozigosidade, o equilíbrio de Hardy-Weinberg e o padrão de variância molecular da população estudada.

Foi encontrada uma microvariante alélica pertencente ao loco D18S51, possivelmente referente ao alelo 29, fora da escada e da escala padrão. Portanto, análises familiares e sequenciamento do DNA devem ser feitos para se verificar a possibilidade de ocorrência familiar ou duplicação de nucleotídeos. Após a confirmação, recomenda-se a inclusão na base de dados internacional STRBase.

Desta forma, inclusive pelas pequenas diferenças que apresentam, pode-se concluir que estes valores podem e devem ser incorporados ao banco de dados populacional brasileiro e representam o protótipo para a criação do primeiro banco regional de dados genéticos da população local. Mais amostras devem ser analisadas com a finalidade de aplicações em análises forenses, estudos de vínculo genético e estudos genéticos populacionais, incrementando a proposta desta tese.

REFERÊNCIAS

REFERÊNCIAS

- AGUIAR, V. R. et al. Updated Brazilian STR allele frequency data using over 100,000 individuals: an analysis of CSF1PO, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D21S11, FGA, Penta D, Penta E, TH01, TPOX and vWA loci. **Forensic Sci Int Genet**, v. 6, n. 4, p. 504-9, Jul 2012.
- ALGEE-HEWITT, B. F. et al. Individual Identifiability Predicts Population Identifiability in Forensic Microsatellite Markers. **Curr Biol**, Mar 15 2016.
- ALLOR, C.; EINUM, D. D.; SCARPETTA, M. Identification and characterization of variant alleles at CODIS STR loci. **J Forensic Sci**, v. 50, n. 5, p. 1128-33, Sep 2005.
- ALVES, H. B. et al. STR data for 15 autosomal STR markers from Parana (Southern Brazil). **Int J Legal Med**, v. 128, n. 2, p. 269-70, Mar 2014.
- BALDING, D. J.; NICHOLS, R. A. A method for quantifying differentiation between populations at multi-allelic loci and its implications for investigating identity and paternity. **Genetica**, v. 96, n. 1-2, p. 3-12, 1995.
- BAR, W. et al. DNA recommendations. Further report of the DNA Commission of the ISFH regarding the use of short tandem repeat systems. International Society for Forensic Haemogenetics. **Int J Legal Med**, v. 110, n. 4, p. 175-6, 1997.
- BARBER, M. D.; PARKIN, B. H. Sequence analysis and allelic designation of the two short tandem repeat loci D18S51 and D8S1179. **Int J Legal Med**, v. 109, n. 2, p. 62-5, 1996.
- BECKER, D. et al. New alleles and mutational events at 14 STR loci from different German populations. **Forensic Sci Int Genet**, v. 1, n. 3-4, p. 232-7, Dec 2007.
- BERGLUND, E. C.; KIIALAINEN, A.; SYVANEN, A. C. Next-generation sequencing technologies and applications for human genetic history and forensics. **Investig Genet**, v. 2, p. 23, 2011.
- BONACCORSO, N. S. **Aspectos técnicos, éticos e jurídicos relacionados com a criação de banco de dados criminais de DNA no Brasil**. 2010. Universidade de São Paulo, São Paulo.
- BUTLER, J. M. **Forensic DNA typing: biology & technology behind STR markers**. 2nd edition. Waltham MA: Elsevier Academic Press, 2005.
-

BUTLER, J. M. **Advanced Topics in Forensic DNA Typing: Methodology**. 1st edition. Walthman, MA: Elsevier Academic Press, 2012. 704.

BUTLER, J. M. **Advanced Topics in Forensic DNA Typing: Interpretation**. 1st edition. Walthman, MA: Elsevier Academic Press, 2014. 608.

BUTLER, J. M. Why DNA Interpretation Has Become More Challenging in Recent Years. **American Academy of Forensic Sciences**, Orlando, FL, p. 20, 2015. Disponível em: <http://www.cstl.nist.gov/strbase/pub_pres/Butler-DNA-interpretation-AAFS2015.pdf>. Acessado em 03/24/2016.

BUTLER, J. M.; HILL, C. R. Biology and Genetics of New Autosomal STR Loci Useful for Forensic DNA Analysis. **Forensic Sci Rev**, v. 24, n. 1, p. 15-26, Jan 2012.

CARRACEDO, A. et al. New guidelines for the publication of genetic population data. **Forensic Sci Int Genet**, v. 7, n. 2, p. 217-20, Feb 2013.

CARRACEDO, A. et al. Update of the guidelines for the publication of genetic population data. **Forensic Sci Int Genet**, v. 10, p. A1-2, May 2014.

CARRACEDO, A. et al. Publication of population data for forensic purposes. **Forensic Sci Int Genet**, v. 4, n. 3, p. 145-7, Apr 2010.

CHULA, F. G. et al. 15 STR loci frequencies with mutation rates in the population from Rio Grande do Sul, Southern Brazil. **Forensic Sci Int Genet**, v. 3, n. 2, p. e35-8, Mar 2009.

COLLINS, P. J. et al. Developmental validation of a single-tube amplification of the 13 CODIS STR loci, D2S1338, D19S433, and amelogenin: the AmpFI STR Identifiler PCR Amplification Kit. **J Forensic Sci**, v. 49, n. 6, p. 1265-77, Nov 2004.

CONSORTIUM, I. H. G. S. Initial sequencing and analysis of the human genome. **Nature**, v. 409, n. 6822, p. 860-921, Feb 15 2001.

CONSORTIUM, I. H. G. S. Finishing the euchromatic sequence of the human genome. **Nature**, v. 431, n. 7011, p. 931-45, Oct 21 2004.

DALTON, G. C. et al. Genetic profile of Federal District of Brazil based on 18 STR autosomal loci. **Forensic Science International: Genetics Supplement Series**, v. 2, n. 2, p. 373-375, 2009.

DE ASSIS POIARES, L. et al. Allele frequencies of 15 STRs in a representative sample of the Brazilian population. **Forensic Sci Int Genet**, v. 4, n. 2, p. e61-3, Feb 2010.

DEL CASTILLO, D. M. et al. Populational genetic data for 15 STR markers in the Brazilian population of Minas Gerais. **Leg Med (Tokyo)**, v. 11, n. 1, p. 45-7, Jan 2009.

DELLALIBERA, E. et al. Genetic analysis of 13 STR loci in the population from the State of Pernambuco, northeast Brazil. **Forensic Sci Int**, v. 146, n. 1, p. 57-9, Nov 10 2004.

DIVNE, A. M.; EDLUND, H.; ALLEN, M. Forensic analysis of autosomal STR markers using Pyrosequencing. **Forensic Sci Int Genet**, v. 4, n. 2, p. 122-9, Feb 2010.

EGYED, B. et al. Analysis of eight STR loci in two Hungarian populations. **Int J Legal Med**, v. 113, n. 5, p. 272-5, 2000.

ELLEGREN, H. Microsatellites: simple sequences with complex evolution. **Nat Rev Genet**, v. 5, n. 6, p. 435-45, Jun 2004.

EXCOFFIER, L.; LISCHER, H. E. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. **Mol Ecol Resour**, v. 10, n. 3, p. 564-7, May 2010.

FERREIRA DA SILVA, L. A. et al. Allele frequencies of nine STR loci--D16S539, D7S820, D13S317, CSF1PO, TPOX, TH01, F13A01, FESFPS and vWA--in the population from Alagoas, northeastern Brazil. **Forensic Sci Int**, v. 130, n. 2-3, p. 187-8, Dec 4 2002.

FIGUEIREDO, A. L. S.; PARADELA, E. R. **Bancos de dados de DNA: Uma ferramenta investigativa útil.** Âmbito Jurídico. Rio Grande. 32 2006.

FIGUEIREDO, M. S. et al. Brazilian Caucasian population data for 15 STR loci (PowerPlex 16 kit). **J Forensic Sci**, v. 49, n. 1, p. 167-9, Jan 2004.

GASTALDO, A. Z. et al. Allele frequencies of 20 autosomal STR in the population from Rio Grande do Sul Southern Brazil. **Forensic Sci Int Genet**, v. 19, p. 15-7, Nov 2015.

GILL, P. et al. Considerations from the European DNA profiling group (EDNAP) concerning STR nomenclature. **Forensic Sci Int**, v. 87, n. 3, p. 185-92, Jun 23 1997.

GILL, P. et al. An assessment of whether SNPs will replace STRs in national DNA databases--joint considerations of the DNA working group of the European Network of Forensic Science Institutes (ENFSI) and the Scientific Working Group on DNA Analysis Methods (SWGDM). **Sci Justice**, v. 44, n. 1, p. 51-3, Jan-Mar 2004.

GOES, A. C. et al. Allele frequencies data and statistic parameters for 16 STR loci-D19S433, D2S1338, CSF1PO, D16S539, D7S820, D21S11, D18S51, D13S317, D5S818, FGA, Penta E, TH01, vWA, D8S1179, TPOX, D3S1358-in the Rio de Janeiro population, Brazil. **Forensic Sci Int**, v. 140, n. 1, p. 131-2, Feb 10 2004.

GOMES, A. V. et al. 13 STR loci frequencies in the population from Paraiba, Northeast Brazil. **Forensic Sci Int**, v. 173, n. 2-3, p. 231-4, Dec 20 2007.

GRUBWIESER, P. et al. Unusual variant alleles in commonly used short tandem repeat loci. **Int J Legal Med**, v. 119, n. 3, p. 164-6, May 2005.

HAUGE, X. Y.; LITT, M. A study of the origin of 'shadow bands' seen when typing dinucleotide repeat polymorphisms by the PCR. **Hum Mol Genet**, v. 2, n. 4, p. 411-5, Apr 1993.

HEINRICH, M. et al. Characterisation of variant alleles in the STR systems D2S1338, D3S1358 and D19S433. **Int J Legal Med**, v. 119, n. 5, p. 310-3, Sep 2005.

HILL, C. R. et al. U.S. population data for 29 autosomal STR loci. **Forensic Sci Int Genet**, v. 7, n. 3, p. e82-3, May 2013.

HUEL, R. L. et al. Variant alleles, triallelic patterns, and point mutations observed in nuclear short tandem repeat typing of populations in Bosnia and Serbia. **Croat Med J**, v. 48, n. 4, p. 494-502, Aug 2007.

JEFFREYS, A. J.; BROOKFIELD, J. F.; SEMEONOFF, R. Positive identification of an immigration test-case using human DNA fingerprints. **Nature**, v. 317, n. 6040, p. 818-9, Oct 31-Nov 6 1985.

JEFFREYS, A. J.; WILSON, V.; THEIN, S. L. Hypervariable 'minisatellite' regions in human DNA. **Nature**, v. 314, n. 6006, p. 67-73, Mar 7-13 1985a.

JEFFREYS, A. J.; WILSON, V.; THEIN, S. L. Individual-specific 'fingerprints' of human DNA. **Nature**, v. 316, n. 6023, p. 76-9, Jul 4-10 1985b.

JOVANOVIĆ, S. et al. Developmental validation of a fully integrated sample-to-profile rapid human identification system for processing single-source reference buccal samples. **Forensic Sci Int Genet**, v. 16, p. 181-94, May 2015.

KLINE, M. C. et al. STR sequence analysis for characterizing normal, variant, and null alleles. **Forensic Sci Int Genet**, v. 5, n. 4, p. 329-32, Aug 2011.

KLINTSCHAR, M.; WIEGAND, P. Polymerase slippage in relation to the uniformity of tetrameric repeat stretches. **Forensic Sci Int**, v. 135, n. 2, p. 163-6, Aug 12 2003.

LEWIS, P. R.; REYNOLDS, K.; GAGG, C. **Forensic Materials Engineering: Case Studies** Boca Raton, FL: CRC Press, 2004.

LINCOLN, P. J. DNA recommendations--further report of the DNA Commission of the ISFH regarding the use of short tandem repeat systems. **Forensic Sci Int**, v. 87, n. 3, p. 181-4, Jun 23 1997.

LYNCH, M.; RITLAND, K. Estimation of pairwise relatedness with molecular markers. **Genetics**, v. 152, n. 4, p. 1753-66, Aug 1999.

MORALES-VALVERDE, A. et al. Characterisation of 12 new alleles in the STR system D18S51. **Forensic Science International: Genetics Supplement Series**, v. 2, n. 2, p. 43-44, 2009.

NASCIMENTO, E.; CERQUEIRA, E.; GUSMAO, L. Population database defined by 13 autosomal STR loci in a representative sample from Bahia, Northeast Brazil. **Forensic Sci Int Genet**, v. 5, n. 1, p. e38-40, Jan 2011.

NIELSEN, R. et al. Genotype and SNP calling from next-generation sequencing data. **Nat Rev Genet**, v. 12, n. 6, p. 443-51, Jun 2011.

PEAKALL, R.; SMOUSE, P. E. GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. **Molecular Ecology Notes**, v. 6, p. 288-295, 2006.

PEAKALL, R.; SMOUSE, P. E. GenAEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research--an update. **Bioinformatics**, v. 28, n. 19, p. 2537-9, Oct 1 2012.

PHILLIPS, C. et al. Analysis of global variability in 15 established and 5 new European Standard Set (ESS) STRs using the CEPH human genome diversity panel. **Forensic Sci Int Genet**, v. 5, n. 3, p. 155-69, Jun 2011.

PHILLIPS, C.; FONDEVILA, M.; LAREAU, M. V. A 34-plex autosomal SNP single base extension assay for ancestry investigations. **Methods Mol Biol**, v. 830, p. 109-26, 2012.

PUERS, C. et al. Identification of repeat sequence heterogeneity at the polymorphic short tandem repeat locus HUMTH01[AATG]_n and reassignment of alleles in population analysis by using a locus-specific allelic ladder. **Am J Hum Genet**, v. 53, n. 4, p. 953-8, Oct 1993.

REGATEIRO, F. J. **Manual de Genética Médica**. 2nd. Coimbra: Imprensa da Universidade de Coimbra, 2007.

RIBEIRO, T. et al. Population data of the AmpFISTR(R) NGM loci in South Portuguese population. **Forensic Sci Int Genet**, v. 7, n. 2, p. e37-9, Feb 2013.

RODRIGUES, E. M.; PALHA TDE, J.; DOS SANTOS, S. E. Allele frequencies data and statistic parameters for 13 STR loci in a population of the Brazilian Amazon Region. **Forensic Sci Int**, v. 168, n. 2-3, p. 244-7, May 24 2007.

ROEWER, L. DNA fingerprinting in forensics: past, present, future. **Investig Genet**, v. 4, n. 1, p. 22, 2013.

ROSENBERG, N. A. A population-genetic perspective on the similarities and differences among worldwide human populations. **Hum Biol**, v. 83, n. 6, p. 659-84, Dec 2011.

RUITBERG, C. M.; REEDER, D. J.; BUTLER, J. M. STRBase: a short tandem repeat DNA database for the human identity testing community. **Nucleic Acids Res**, v. 29, n. 1, p. 320-2, Jan 1 2001.

SANTOS, S. M. et al. Distribution of forensic marker allelic frequencies in Pernambuco, Northeastern Brazil. **Genet Mol Res**, v. 14, n. 2, p. 4303-10, 2015.

SILVA, D. A. et al. Statistical analyses of 14 short tandem repeat loci in Brazilian populations from Rio de Janeiro and Mato Grosso do Sul states for forensic and identity testing purposes. **Forensic Sci Int**, v. 139, n. 2-3, p. 173-6, Jan 28 2004.

SILVA, M. B. et al. Allele frequencies of fifteen STR loci in a population from Central Brazil. **Forensic Sci Int Genet**, v. 4, n. 5, p. e151-2, Oct 2010.

SWGDAM. Interpretation Guidelines for Autosomal STR Typing by Forensic DNA Testing Laboratories. APPROVED 1/14/10. Disponível em: < <http://www.fbi.gov/about-us/lab/codis/swgdam-interpretation-guidelines> >. Acessado em 10/12/2015.

SONG, S.; DEY, D. K.; HOLSINGER, K. E. Genetic diversity of microsatellite loci in hierarchically structured populations. **Theor Popul Biol**, v. 80, n. 1, p. 29-37, Aug 2011.

TAUTZ, D. **Notes on the defunction and nomenclature dimly repetitive DNA sequences**. Basel: 1993.

URQUHART, A. et al. Variation in short tandem repeat sequences--a survey of twelve microsatellite loci for use as forensic identification markers. **Int J Legal Med**, v. 107, n. 1, p. 13-20, 1994.

VIEIRA, T. C. et al. Allelic frequencies and statistical data obtained from 15 STR loci in a population of the Goiás State. **Genet Mol Res**, v. 12, n. 1, p. 23-7, 2013.

WHITTLE, M. R.; ROMANO, N. L.; NEGREIROS, V. A. Updated Brazilian genetic data, together with mutation rates, on 19 STR loci, including D10S1237. **Forensic Sci Int**, v. 139, n. 2-3, p. 207-10, Jan 28 2004.

WOLFGRAMM EDE, V. et al. Genetic analysis of 15 autosomal and 12 Y-STR loci in the Espírito Santo State population, Brazil. **Forensic Sci Int Genet**, v. 5, n. 3, p. e41-3, Jun 2011.

ANEXO

ANEXO 1- Microvariantes alélicas da STR D18S51. Trechos específicos, de interesse ilustrativo neste estudo, copiados de: http://www.cstl.nist.gov/biotech/strbase/var_D18S51.htm.

Alelo	Tamanho do alelo	Equipamento	Kit	Contribuidor	Verificação/Método(s) de confirmação	Notas	Frequência
6	257.45	ABI 3130XL	ID	Genetica, DNA Laboratories, Inc., Cincinnati OH	Amostra do pai		
6 [2]	256.65	ABI 3130XL	ID	Genetica, DNA Laboratories, Inc., Cincinnati OH	Amostra do pai		
7	265	ABI 310	PS	Carmen Tomas	Reextração e reamplificação	Puna(Salta, Argentina)	2 em 108
7 [3]	265.81	ABI 377	PS	Igor Kornienko, Tatjana Tarasova	Reamplificação	Amostras russas somente	1 em 614 até o momento
8	269.67	ABI 377	PS	Igor Kornienko, Tatjana Tarasova DoD, Russia	Reamplificação	Amostras russas somente	1 em 658
8 [2]	266.38	ABI 3100	PS	Jan Nicklas, VT Forensic Laboratory	Reamplificação		
8 [3]	265.85	ABI 3100	ID	Marsha Garcia, Orchid Cellmark, Nashville TN	Reextração		
9.2	271.92	ABI 3100	ID	Marsha Garcia, Orchid Cellmark, Nashville	Reamplificação		
9.2 [2]	272.49 to 272.53	ABI 3130	PS	Laboratoire de sciences judiciaires et de médecine légale, Montréal, Canada		Sangue	2 vezes no mesmo caso
10.2 [2]	276.48	ABI 3130	PS	Laboratoire de sciences judiciaires et de médecine légale, Montréal, Canada			
24.2 [3]	349.02	ABI 3500xl	ESX 16	Malin Sanga, Swedish National Laboratory of Forensic Science (SKL)	Reamplificação e reeletroforese	Amostra de referência	1 – primeira observação
27	347.36	ABI 377	PS	Chris PiwonkaTexas DPS	Multiple samples extracted	Infrator condenado	1 em 49.000
28 (>27)	349.14	ABI 310	ID	Oscar Garcia, Forensic Genetics Department, Autonomous Police of the Basque Country	Reamplificação e reanálise	Amostra de infrator condenado	
28 (>27) [2]	351.85	ABI 3730	ID	DNA Forensic Division, Supreme Prosecutor's Office, South Korea	Reanálises com ABI 3130		1 em 31.330
28.1	351.52	ABI 377	PS	Catherine Allor	Reamplificação e reanálise	Amostra de paternidade	1 em 33.300
28.3	352	ABI 3100	SGM+	RivkaAdelman, IDFS, Israel Police, Jerusalem, Israel	Múltiplas amostras extraídas e reamplificação	3 amostras de estudo de caso e duplicata de amostra de referência	1
28.3 [2]	352.36	ABI 3100	PS	Angela Withrow - M U F S C	Reamplificação x 1		1
28.3 [3]	352.46	ABI 3100	PS	Nicole Hoffman M U F S C	Sem verificação prévia		2
29.3	355.88	ABI 3100	ID	Marsha Garcia Orchid Cellmark, Nashville TN	Reextração		

29.3 [2] (>27)	356.15, 355.94	ABI 3130	PS	Laboratoire de sciences judiciaires et de médecine légale, Montréal, Canada	Reamplificação e reanálise	Amostra de referência	Duplicata
>27 (?)	359.64	ABI 3730	ID	DNA Forensic Division, Supreme Prosecutor's Office, South Korea	Reamplificação e reanálise com ABI 3130		1 em 31.330
>27 (33 ?)	367.53, 367.6	ABI 3130	ID	Laboratoire de sciences judiciaires et de médecine légale, Montréal, Canada	Reamplificação	Elevado stutter e n+4 franqueadora >27; possivelmente Latino	2 vezes no mesmo caso
30	356.43 (ID) and 356.85 (NGM)	ABI 310	ID & NGM SElect	Laboratory Identification and Population Genetics, Institute of Genetics of National University of Colombia	Reextração e reamplificação	Amostra de referência	1 em 9.904
34	375.80	ABI 3100	PS	Julia Harnett, National Forensic Services, RCMP	Múltiplas amostras, sequenciamento de amostra representativa por NIST		4
40	396.37	ABI 310	PS	Christine Moraczewski NE State Patrol Crime Lab	Reextração e reamplificação	Infrator condenado	1

*53 variantes relatadas em 07/08/2015

Abreviações dos nomes dos kits de amplificação:

CO: Applied Biosystems COfiler

PR: Applied Biosystems Profiler

PS: Applied Biosystems Profiler Plus

ID: Applied Biosystems Identifier

PP: Promega PowerPlex 16

CV: Promega CTTV

CA: Promega CTTA

MP: Promega MonoPlex