

TICIANE HENRIQUES SANTA RITA

**FREQUÊNCIAS ALÉLICAS DE 15 STRs AUTOSSÔMICOS EM UMA AMOSTRA
POPULACIONAL DO DISTRITO FEDERAL DO BRASIL –
UM TERRITÓRIO QUE SURTIU DO NADA**

BRASÍLIA, 2013

UNIVERSIDADE DE BRASÍLIA
FACULDADE DE CIÊNCIAS DA SAÚDE
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS DA SAÚDE

TICIANE HENRIQUES SANTA RITA

**FREQUÊNCIAS ALÉLICAS DE 15 STRs AUTOSSÔMICOS EM UMA AMOSTRA
POPULACIONAL DO DISTRITO FEDERAL DO BRASIL –
UM TERRITÓRIO QUE SURTIU DO NADA**

Dissertação apresentada como requisito
parcial para a obtenção do título de Mestre
em Ciências da Saúde pelo Programa de
Pós-Graduação em Ciências da Saúde da
Universidade de Brasília.

Orientador: Prof. Dr. Gustavo Barcelos Barra

Brasília
2013

À Deus, pelo dom da vida e da sabedoria.

Aos meus pais Eloiza e Idelmo, exemplos de humildade, honestidade e caráter, pelo investimento em toda minha vida para que seja repleto de conhecimentos e por sonhar meus sonhos.

AGRADECIMENTOS

Ao meu irmão Igor “espírito de luz” mesmo não podendo te ver, sinto-me tocar nas vitórias e nos obstáculos. Obrigada por ser meu anjo protetor. Ao meu irmão companheiro Felipe Henriques, obrigada por acreditar em mim e fazer sentir orgulho de mim mesma. A Vilma Menezes por ser minha cúmplice em tudo e por cuidar da nossa família, enquanto estou longe de casa.

À família de Brasília, em especial Ademilton Félix, Renata Valadares e Sarah Félix. Obrigada pelo seio familiar que me proporcionaram, isto sem dúvidas, foi essencial.

Ao meu orientador, professor e amigo Gustavo Barra, com você não só aprendi como desenvolver uma pesquisa de qualidade, mas também, como ser um bom ser humano. Obrigada por todas as contribuições que são imensuráveis, por confiar, acreditar e investir seu tempo durante os últimos dois anos. Você é o responsável por esta conquista, tu és meu espelho e o exemplo a ser seguido.

Ao Laboratório Sabin, obrigada por participar das minhas conquistas nesse período, pela oportunidade e investir na ciência. Especialmente ao setor de Biologia Molecular (Dr^a Lara Velasco, Cláudia Sousa, Ceçília Caixeta, Ana Luisa Almeida, Patrícia Godoy, Déborah Costa, Aparecida Silva, Vanessa Marques e Camila Silva) pela convivência harmoniosa, pelo acolhimento e pela disposição em ajudar.

As amigas de mestrado (Camilla Chianca e Júlia Vasques), parceiras de estudo, de choros e altas risadas. Aos amigos que fiz nessa trajetória, Rodrigo Lourenço, Aline Cabral, Hermann Deny, Jonas Rodolfo, Marina Ponte, Eduardo Lourenço e Daniela Paniago, vocês são a diversão e o descanso.

As minhas amigas da época de faculdade e da vida (Renata Rego, Maysa Araújo e Nathalie Garcia) por todo incentivo em qualquer decisão e por ajudar a superar toda a saudade com essa distância.

Meus sinceros agradecimentos.

NOTAS DO ORIENTADOR

Agradecimentos:

Ao Laboratório Sabin,

Agradeço o investimento que a empresa faz em pesquisa científica e nos sonhos de seus colaboradores.

À Ticiane Henriques Santa Rita,

Alunos como você é que justificam a existência dos orientadores, pois fazem com que o tempo investido valha a pena.

RESUMO

O Distrito Federal do Brasil foi criado em 1960 na Região Centro-Oeste em uma área anteriormente despovoada. Em 2010, este território artificial estava povoado por 2.556.121 habitantes. Neste estudo, analisou-se na população do Distrito Federal as variações genéticas dos quinze *loci* STRs incluídos no kit AmpFISTR® NGM™ (Life Technologies) (D10S1248, D22S1045, D2S441, D1S1656, D12S391, D2S1338, D3S1358, D8S1179, D16S539, D18S51, D19S433, D21S11, FGA, TH01 e vWA). Os resultados indicam que o NGM™ é um sistema genético altamente informativo. Pois, a probabilidade de discriminar dois indivíduos selecionados ao acaso é de 99,99999999999999995% e a probabilidade de excluir um indivíduo falsamente acusado em um caso de paternidade é 99,99998%. Ademais, a análise das distâncias genéticas entre as populações das cinco macrorregiões brasileiras revelou que o Distrito Federal é mais semelhante ao Sudeste, ao Nordeste e a população brasileira como um todo do que as populações do Sul, Centro-Oeste e Norte. Esta conclusão está de acordo com a composição da população relatada no Censo 2010, em que a maioria dos imigrantes era do Nordeste e do Sudeste.

Palavras-chave: *Short tandem repeats*, Brasil, Brasília, Distrito Federal (Brasil), NGM™, Genética forense

ABSTRACT

The Federal District of Brazil was created in 1960 in the Central-West Region of Brazil in a previously unpopulated area. In 2010, this artificially founded district was populated by 2,562,963 inhabitants. In this study, we analyzed the genetic variations of the fifteen STRs *loci* (D10S1248, D22S1045, D2S441, D1S1656, D12S391, D2S1338, D3S1358, D8S1179, D16S539, D18S51, D19S433, D21S11, FGA, TH01 and vWA) included in the kit AmpFISTR® NGM™ (Life Technologies) in this population. The results indicate that the NGM™ is a highly informative genetic system. Therefore, the probability of discriminating two individuals selected at random is 99.9999999999999995% and the probability of excluding a falsely accused individuals in a paternity case is 99.99998%. Furthermore, analysis of genetic distances between populations of the five macroregions revealed that the Brazilian Federal District is more similar to the southeastern, northeastern and overall Brazil population compared to Southern, Central-Western and Northern populations. This conclusion agrees with the population composition reported in the 2010 National Survey Inquiries, in which most of the immigrants were from the northeast and the southeast.

Keywords: *Short* tandem repeats, Brazil, Brasília, Federal District (Brazil), NGM™ ,
Forensic genetics

LISTA DE FIGURAS

Figura 1 - Análise de coordenadas principais das distâncias genéticas de Nei (DA) entre as populações estudadas. Dados da matriz de distância genética (tabela 7)

Figura 2 - Dendrograma gerado pelo software Dispan (método *neighbor-joining*) a partir da matriz de distâncias genéticas de Nei (DA) das populações estudadas

LISTA DE TABELAS

Tabela 1 – Comparação entre as distribuições do local de nascimento da amostra populacional extraída do banco de dados e da população do Distrito Federal de acordo com o Censo 2010

Tabela 2 - Frequências alélicas de quinze STRs incluídos no kit NGM™ em uma amostra populacional do Distrito Federal (n = 462)

Tabela 3 – Heterozigose observada e heterozigose esperada dos quinze STRs incluídos no kit NGM™ em uma amostra populacional do Distrito Federal (n = 462)

Tabela 4 – Resultados do valor de P para o equilíbrio de Hardy-Weinberg dos quinze STRs incluídos no kit NGM™ em uma amostra populacional do Distrito Federal (n = 462)

Tabela 5 - Parâmetros forenses dos quinze STRs incluídos no kit NGM™ em uma amostra populacional do Distrito Federal (n = 462)

Tabela 6 – Parâmetros de paternidade para os quinze STRs incluídos no kit NGM™ em uma amostra populacional do Distrito Federal (n = 462)

Tabela 7 - Matriz da distância genética de Nei (DA) calculada pelo software DISPAN a partir das frequências alélicas deste estudo e de Raimann e colaboradores

LISTA DE SIGLAS

STR	<i>Short tandem repeats</i>
NGM	Next Generation Multiplex
DNA	Ácido desoxirribonucleico
PCR	Reação em cadeia da polimerase
ESSL	<i>Expanded European Standard Set of Loci</i>
CODIS	<i>Combined DNA Index System</i>
ESS	<i>European Standard Set</i>
CIP	Contéudo de informação polimórfica
PD	Poder de discriminação
PM	Probabilidade de coincidência (“ <i>matching</i> ”)
IPT	Índice de paternidade típico
IPC	Índice de paternidade combinado
PCoA	Análise de coordenadas principais

SUMÁRIO

1	INTRODUÇÃO	13
1.1	BRASÍLIA E O DISTRITO FEDERAL	13
1.2	OS “SHORT TANDEM REPEATS”	13
1.3	O KIT AMPFISTR® NGM™	14
1.4	A NECESSIDADE DE FREQUÊNCIAS ALÉLICAS LOCAIS PARA OS CÁLCULOS FORENSES E DE PATERNIDADE	14
1.5	PARÂMETROS FORENSES E DE PATERNIDADE	15
1.6	DESEQUILÍBRIO DE LIGAÇÃO ENTRE OS <i>LOCI</i>	16
1.7	DESCRIÇÕES DE FREQUÊNCIAS ALÉLICAS ENVOLVENDO O NGM™ NO BRASIL	16
1.8	DISTÂNCIAS GENÉTICAS ENTRE AS POPULAÇÕES	17
1.9	SÍNTESE E CONSIDERAÇÕES FINAIS	17
2	OBJETIVOS	19
3	MÉTODOS	20
3.1	POPULAÇÃO E BASE DE DADOS	20
3.2	EXTRAÇÃO DE DNA	20
3.3	AMPLIFICAÇÃO DE PCR E GENOTIPAGEM	21
3.4	ANÁLISE ESTATÍSTICA DOS DADOS	21
3.4.1	Comparação entre a amostra populacional e os dados do Censo 2010	21
3.4.2	Frequências alélicas	22
3.4.3	Heterozigose (ou heterozigosidade)	22
3.4.4	Equilíbrio de Hardy-Weinberg	23
3.4.5	Desequilíbrio de ligação	23
3.4.6	Conteúdo de informação polimórfica (CIP)	24
3.4.7	Probabilidade de coincidência (“ <i>matching</i> ”) (PM)	24
3.4.8	Poder de discriminação (PD)	24
3.4.9	Poder de exclusão (PE)	25
3.4.10	Índice de paternidade típico (IPT)	25
3.4.11	Distância genética	26
3.5	ASPECTOS ÉTICOS	27
4	RESULTADOS	28
4.1	A AMOSTRA POPULACIONAL DO ESTUDO REPRESENTA A POPULAÇÃO DO DISTRITO FEDERAL	28
4.2	FREQUÊNCIAS ALÉLICAS	28
4.3	HETEROZIGOSE	30
4.4	EQUILÍBRIO DE HARDY - WEINBERG	30
4.5	DESEQUILÍBRIO DE LIGAÇÃO	31
4.6	PARÂMETROS FORENSES	31
4.7	PARÂMETROS DE PATERNIDADE	32

4.8	DISTÂNCIA GENÉTICA	33
5	DISCUSSÃO	36
5.1	A AMOSTRA POPULACIONAL DO ESTUDO REPRESENTA A POPULAÇÃO DO DISTRITO FEDERAL	36
5.2	FREQUÊNCIAS ALÉLICAS.....	37
5.3	PARÂMETROS FORENSES.....	37
5.4	PARÂMETROS DE PATERNIDADE	39
5.5	DISTÂNCIA GENÉTICA	40
6	CONCLUSÃO	42
7	REFERÊNCIAS.....	43
	APÊNDICE I – ARTIGO CIENTÍFICO	48
	APÊNDICE II – PRODUÇÃO BIBLIOGRÁFICA	50
	ANEXO I – APROVAÇÃO PELO COMITÊ DE ÉTICA EM PESQUISA.....	59

PREFÁCIO

Este estudo descreve na população do Distrito Federal do Brasil, as frequências alélicas e os parâmetros forenses e de paternidade dos quinze marcadores de DNA do tipo “*Short Tandem Repeats*” (STRs) incluídos no ensaio comercial kit AmpFISTR® NGM™ (Life Technologies). A importância destas descrições reside no fato de subsidiar o uso do kit na população analisada. Pois, recomenda-se o uso de dados locais para os cálculos envolvidos na identificação humana e na investigação de vínculo genético.

Além disso, o Distrito Federal é um território artificial inaugurado em 1960 no centro do país com a finalidade de interiorizar a capital federal. O local escolhido era inicialmente despovoado e atualmente possui mais de dois milhões e meio de habitantes. Assim, trata-se de uma população nova e de rápido crescimento. Embora o NGM™ tenha sido lançado no Brasil em 2010 sua concepção baseou-se nas necessidades da comunidade forense europeia. Isto faz com que o conhecimento de suas características analíticas na população do Distrito Federal seja desejável antes do uso.

Para a execução do trabalho, as frequências alélicas foram calculadas a partir do banco de dados de resultados de investigações de vínculo genético de um laboratório clínico do Distrito Federal (Laboratório Sabin). Trata-se assim de uma amostragem de conveniência. Entretanto, a distribuição do local de nascimento da amostra não foi significativamente diferente da população do território brasileiro conforme o Censo 2010, possibilitando a extrapolação dos dados. Ademais, é preciso ressaltar que o trabalho foi aprovado pelo comitê de ética da Universidade de Brasília, e houve dispensa de termo de consentimento devido a análise ser de banco de dados obtidos da rotina laboratorial sendo descritos de forma agregada impossibilitando a identificação de qualquer participante.

Por fim, a validação externa do estudo foi obtida pelo aceite do trabalho na revista forense “*International Journal of Legal Medicine*” de fator de impacto “*Journal Citation Reports*” de 2.8 e indexada no Pubmed (Medline). Na publicação (apêndice I), a autora desta dissertação divide primeira autoria com o seu orientador.

1 INTRODUÇÃO

1.1 BRASÍLIA E O DISTRITO FEDERAL

O Brasil é um país continental dividido em cinco macrorregiões: Norte, Nordeste, Centro-Oeste, Sudeste e Sul (1). O Distrito Federal e Brasília foram fundados em 21 de abril de 1960 baseado no desejo de transferir a capital federal do Rio de Janeiro para o centro do país. Ambos foram construídos em apenas 4 anos em um território anteriormente não povoado na região Centro-Oeste, como resultado dos esforços do presidente Juscelino Kubitschek, em conjunto com o arquiteto Oscar Niemeyer, o urbanista Lúcio Costa, e milhares de trabalhadores civis, conhecidos como "candangos" (2).

De acordo com o Censo 2010, a população do Distrito Federal é composta por 2.556.121 habitantes, 54% das pessoas nasceram localmente, e os 46% restantes eram imigrantes das outras macrorregiões brasileiras: Norte (2%), Nordeste (23,5%), Centro-Oeste (6,5%), Sudeste (12,5%) e Sul (1,5%) (3). Segundo estas frequências e outras descrições da literatura, as principais migrações aconteceram da região Nordeste e Sudeste (4). Além disso, o Brasil foi colonizado por Portugal, e o processo de mistura entre o nativo americano, africano e europeu ocorreu em diferentes proporções nas cinco macrorregiões, o que torna a população brasileira heterogênea (1, 5-7). Assim, a população do Distrito Federal é nova, mista, formada a partir de populações heterogêneas, de rápido crescimento e instalada em um território artificial.

1.2 OS “SHORT TANDEM REPEATS”

Os “*short tandem repeats*” (STRs) ou microssatélites são sequências de DNA de 2 a 7 nucleotídeos que se repetem consecutivamente. Tais sequências representam cerca de 3% do genoma humano total, são variáveis de pessoa para pessoas e transmitidas de forma mendeliana (8). Atualmente, são os marcadores de DNA mais utilizados na identificação humana e estudos de vínculo genético. Pois, são facilmente amplificados por PCR e apresentam ampla variabilidade genética, ou seja, são altamente polimórficos. Fato que diminui a probabilidade de que os

indivíduos testados apresentem os mesmos alelos (8, 9). Por causa da sua alta aplicabilidade existem empresas que produzem kits para detecção de STR que possibilitam a análise simultânea de 15 ou mais marcadores como especificado pelas sociedades forenses internacionais (8). Exemplos de kits comerciais para genotipagem de STRs são AmpFISTR® NGM™ (Life Technologies), PowerPlex® 16 (Promega) e AmpFISTR® Identifier™ (Life technologies).

1.3 O KIT AMPFLSTR® NGM™

O foco deste estudo são os STRs presentes no kit AmpFISTR® NGM™ (NGM™) lançado no Brasil em 2010 e desenvolvido pela Life Technologies de acordo com as recomendações da comunidade forense europeia. Os 15 *loci* STRs contidos neste ensaio comercial são: D10S1248, D22S1045, D2S441, D1S1656, D12S391, D2S1338, D3S1358, D8S1179, D16S539, D18S51, D19S433, D21S11, FGA, TH01 e vWa (10, 11). O conjunto compõe o “*Expanded European Standard Set of Loci*” (ESSL), cinco deles são novos marcadores (D10S1248, D22S1045, D2S441, D1S1656 e D12S391) e os demais fazem parte tanto do “*Combined DNA Index System*” (CODIS) americano quanto do ESSL europeu (12, 13). Portanto, a concepção do NGM™ baseou-se nas necessidades europeias e isto faz com que o conhecimento das características analíticas do kit na população do Brasil e, especificamente, do Distrito Federal seja desejável antes do uso.

1.4 A NECESSIDADE DE FREQUÊNCIAS ALÉLICAS LOCAIS PARA OS CÁLCULOS FORENSES E DE PATERNIDADE

Pelo fato de não haver a possibilidade de ter acesso aos perfis de DNA de todos os indivíduos existentes, são utilizados dados de amostras populacionais menores e extrapola-se a possibilidade de coincidência entre perfis genéticos. Dessa forma, as frequências alélicas de STRs são coletadas a partir de um grupo de pessoas e, em seguida, os postulados de genética populacional são aplicados para inferir se é razoável que um indivíduo tenha, ao acaso, o mesmo perfil genético que outro (8). Como a distribuição da frequência alélica de qualquer marcador de DNA é variável entre populações distintas, faz-se necessário estudos de frequências locais

para conhecer os alelos mais comuns numa determinada região, a fim de definir os marcadores mais informativos para as investigações forenses (14).

Os cálculos envolvidos nos estudos de vínculo genético também dependem das frequências alélicas da população em questão. Por exemplo, em um estudo de vínculo genético de paternidade ao se investigar o marcador D18S51 os seguintes genótipos são obtidos: mãe (17/20), criança (14/17) e o suposto pai (14/22). Este resultado caracteriza uma inclusão para o *locus* em questão. Diante de uma inclusão, três questionamentos são feitos: 1º qual é o alelo obrigatoriamente transmitido pelo pai? (o alelo 14); 2º qual é a probabilidade do suposto pai ter passado o alelo 14 para a criança? (50% de chance); 3º qual é a probabilidade de outro homem ter transmitido o alelo 14, ser o verdadeiro pai da criança e o suposto pai do caso ter sido incluído ao acaso? (frequência alélica do alelo 14 na população a qual os periciandos pertencem). Assim, a razão entre a probabilidade do suposto pai ter transmitido o alelo 14 (0,5) e a probabilidade de outro homem ter transmitido o alelo 14 para o filho (0,1201) corresponde ao índice de paternidade para o *locus* D18S51 ($0,5 / 0,1202 = 4,1631$). Isto é feito para o conjunto de quinze STRs, e o produto dos índices de paternidade individuais de cada *locus* é denominado índice de paternidade acumulado. Assim, a conclusão da investigação depende das frequências alélicas locais, sendo esta a importância de se conhecer as frequências alélicas para os STRs utilizados nos exames de paternidade.

1.5 PARÂMETROS FORENSES E DE PATERNIDADE

A partir dos genótipos dos indivíduos incluídos na amostra populacional calculam-se as frequências genotípicas e as frequências alélicas, e destas estimam-se os parâmetros forenses e de paternidade. Os parâmetros forenses incluem heterozigose esperada, heterozigose observada, equilíbrio de Hardy-Weinberg, conteúdo de informação polimórfica, probabilidade de coincidência (“*matching*”) e poder de discriminação. Os parâmetros de paternidade incluem poder de exclusão e índice de paternidade típico. A seção “Materiais e Métodos” desta dissertação contém uma descrição detalhada dos parâmetros supracitados e como foram calculados.

1.6 DESEQUILÍBRIO DE LIGAÇÃO ENTRE OS *LOCI*

Os cálculos forenses e de paternidade exigem que os STRs sejam estatisticamente independentes, ou seja, que estejam em equilíbrio de ligação. Normalmente, garante-se esta independência ao utilizar STRs localizados em cromossomos distintos, que é o caso da maioria dos *loci* incluídos no kit NGM™, exceto vWA e D12S391. Para STRs localizados no mesmo cromossomo (STRs sintênicos) é necessário comprovar que a segregação de seus alelos ocorre de forma independente (15). Desta forma, o desequilíbrio de ligação entre vWA e D12S391 foi calculado (vide materiais e métodos).

1.7 DESCRIÇÕES DE FREQUÊNCIAS ALÉICAS ENVOLVENDO O NGM™ NO BRASIL

Descrições de frequências alélicas envolvendo os *loci* do NGM™ em diferentes populações estão disponíveis na literatura (16-20). Entretanto, poucos estudos foram desenvolvidos para caracterizar a população brasileira, e nenhum caracterizou o Distrito Federal especificamente. Raimann e colaboradores, genotiparam 850 indivíduos das cinco macrorregiões brasileiras: Norte, Nordeste, Centro-Oeste, Sudeste e Sul com o kit NGM™. No entanto, destes 850 indivíduos apenas 91 eram da região Centro-Oeste. Além disso, quando as cinco macrorregiões foram comparadas entre si, três diferentes agrupamentos baseados em semelhança genética puderam ser identificados: um incluindo o Sul, Sudeste e Centro-oeste, outro o Norte, outro o Nordeste (21). Portanto, existem subpopulações dentro da população brasileira, o que justifica ainda mais o uso de frequências alélicas locais para que os cálculos de inclusão ao acaso sejam o mais realista possível.

As frequências alélicas dos cinco novos marcadores incluídos no ESS e presentes no kit NGM™ foram descritas também para população do Rio Grande do Sul (n=353). Neste trabalho, Rodenbusch e colaboradores descreveram os parâmetros forenses e de paternidade apenas para um conjunto limitado de *loci* (22). Ao comparar as frequências alélicas do Rio Grande do Sul (Sul do Brasil) com seis populações: Norte da Itália, Europa, África, Leste da Ásia, Oceania e México, observaram uma proximidade genética do Sul do Brasil com a Europa e uma menor similaridade genética com relação as demais. Além disso, os cinco novos *loci*

analisados são uma ferramenta poderosa para a identificação forense e testes de paternidade na população do Rio Grande do Sul.

Assim, os trabalhos de Raimann e Rodenbusch são duas descrições sobre o kit NGM anteriores a esta dissertação no Brasil, reforçando sua importância.

1.8 DISTÂNCIAS GENÉTICAS ENTRE AS POPULAÇÕES

Com base nas frequências alélicas de *locus* homólogos em duas ou mais populações é possível estimar a divergência genética entre elas. Isto é útil para esclarecer as relações evolutivas entre populações próximas (inter-relacionadas) e quanto menor a distância genética maior a similaridade entre as populações (23). A distância genética utilizada neste trabalho foi a distância de Nei (DA) que assume que as diferenças se originam a partir de mutações ou deriva genética (24). Trata-se de uma medida mais apropriada para marcadores do tipo STR e é possível ser mensurada a partir das frequências alélicas, dispensando a necessidade de se conhecer os genótipos dos indivíduos que dificilmente são publicados (23). O cálculo desta distância é feito com o auxílio de softwares específicos (vide Materiais e Métodos). Portanto, neste trabalho, para se conhecer a história da formação da população do Distrito Federal por meio de uma evidência genética comparou-se sua distância genética em relação às demais macrorregiões brasileiras.

1.9 SÍNTESE E CONSIDERAÇÕES FINAIS

Desta forma, as frequências alélicas dos STRs incluídos no kit NGM™ em uma amostra populacional do Distrito Federal são essenciais para estimar a probabilidade de ocorrência do perfil encontrado de forma aleatória, auxiliando as investigações forenses, bem como estudos de vínculo genético. A distribuição dos alelos é distinta entre as populações, portanto, recomenda-se que o próprio laboratório construa uma base de dados representativa da população que atende e use as frequências alélicas locais. É importante ressaltar que a depender do banco de dados (tamanho) e das frequências alélicas utilizadas (população) pode-se

encontrar evidências de DNA distintas, ou seja, uma frequência de perfil de DNA diferente (8, 25).

Assim, neste trabalho, investigou-se os parâmetros de eficiência forense e o polimorfismo genético dos quinze *loci* STRs presentes no kit NGM™ (D10S1248, D22S1045, D2S441, D1S1656, D12S391, D2S1338, D3S1358, D8S1179, D16S539, D18S51, D19S433, D21S11, FGA, TH01 e vWA) em uma amostra populacional do Distrito Federal com objetivo principal de descrever as frequências alélicas e estimar os parâmetros forenses e de paternidade. Além disso, analisou-se a distância genética da população do Distrito Federal em relação ao Brasil como um todo e as cinco macrorregiões brasileiras, com a finalidade de se produzir uma evidência genética sobre a formação da população deste território.

2 OBJETIVOS

- Descrever as frequências alélicas e os parâmetros forenses dos quinze STRs incluídos no kit AmpFISTR® NGM™ (Life Technologies) em uma amostra populacional do Distrito Federal;
- Analisar a distância genética da população do Distrito Federal em relação ao Brasil como um todo e suas cinco macrorregiões (Norte, Nordeste, Sul, Sudeste e Centro-Oeste).

3 MÉTODOS

3.1 POPULAÇÃO E BASE DE DADOS

O banco de dados de resultados das investigações de vínculo genético realizadas pelo laboratório Sabin de Análises Clínicas foi utilizado como fonte dos perfis genéticos deste estudo. Durante o período de maio de 2011 a outubro de 2012 o laboratório realizou 1167 investigações.

Os critérios de inclusão foram ser periciando em estudo de vínculo genético, ter perfil genético determinado pelo kit AmpFISTR® NGM™, residir no Distrito Federal e, por fim, ter apresentado informação sobre a sua naturalidade. Foram excluídos os indivíduos em duplicidade, ou seja, que tiveram perfil genético presente mais de uma vez na base de dados e também caso tivesse algum grau de parentesco com os participantes já incluídos (exemplo: os filhos nos casos de paternidade por serem relacionados aos pais recrutados).

Assim, 462 adultos não relacionados (230 mulheres e 232 homens) residentes no Distrito Federal foram selecionados e os perfis de STRs anonimizados foram recuperados e os parâmetros forenses e de paternidade calculados e descritos em um contexto populacional (de forma agregada). Por se tratar de uma amostragem de conveniência, avaliou-se sua representatividade em relação a população do Distrito Federal (vide item 3.4.1). Os aspectos éticos deste trabalho estão descritos no item 3.5.

3.2 EXTRAÇÃO DE DNA

Para a análise de parentesco de rotina, amostras de sangue são coletadas a partir da punção da ponta de dedo anelar em um cartão de papel Whatman® FTA™. O DNA é extraído a partir de manchas de sangue presente neste cartão. Perfura-se um disco de 1,2 mm do cartão FTA com o auxílio de um “Harris Micro Punch” que em seguida é colocado em tubo de amplificação de PCR e lavado com 200 µl de FTA Purification Reagent (Life Technologies).

Após 10 minutos de agitação horizontal (700 rpm), à temperatura ambiente, o reagente FTA é removido, e repete-se esse passo de purificação por três vezes. Em

seguida, faz-se a lavagem do disco com 200 µl de água livre de DNase/RNase por 05 minutos sob agitação horizontal (700 rpm) e após a retirada de todo o líquido a reação de PCR é executada no próprio tubo utilizado na extração.

3.3 AMPLIFICAÇÃO DE PCR E GENOTIPAGEM

Para a rotina de investigação de vínculo genético, o DNA contido em um disco de 1,2 mm (0,5 - 1,0 ng) é amplificado pela reação de PCR “multiplex” do kit AmpFISTR® NGM™ em um termociclador GeneAmp PCR System 9700. A reação é executada de acordo com o protocolo do fabricante (26) . Em seguida, os produtos da PCR são detectados por eletroforese capilar no Analisador Genético 3500. Após amplificação, 1 µl do produto de PCR é adicionado a 8,5 µl de formamida Hi-Di e 0,5µl de 600 LIZ GeneScan. Os alelos são identificados utilizando software GeneMapper®-X v.1.2. O conjunto de fluorescência utilizado é o G5 (6FAM, VIC, NED, PET e LIZ) (todos equipamentos e reagentes deste item são fabricados pela Life Technologies, Foster City, CA, EUA).

3.4 ANÁLISE ESTATÍSTICA DOS DADOS

3.4.1 Comparação entre a amostra populacional e os dados do Censo 2010

Como as frequências genotípicas e alélicas foram calculadas a partir do banco de dados de um laboratório clínico, do Distrito Federal (Laboratório Sabin), refere-se à uma amostragem de conveniência. Neste sentido, comparou-se a distribuição do local de nascimento dos periciandos com a distribuição da naturalidade da população do Distrito Federal, de acordo com o Censo 2010. O teste chi-quadrado do software Graphpad Prism 6.0 (27) foi utilizado para a execução desse cálculo e o nível de significância considerado foi 0,05.

3.4.2 Frequências alélicas

As frequências alélicas dos 15 microssatélites do kit AmpFISTR® NGM™ nos 462 indivíduos foram obtidas por meio do programa Arlequin v.3.1 (28), pelo método de contagem direta, ou seja, número de vezes em que o alelo é observado, dividido pelo número total de alelos averiguados para o marcador.

3.4.3 Heterozigose (ou heterozigosidade)

Significa a proporção de indivíduos com dois alelos diferentes em determinado *locus*, ou seja, corresponde a quantidade de indivíduos heterozigóticos. Quanto maior a heterozigose, maior é a variabilidade genética do *locus* na população em questão, assim, a probabilidade de que dois indivíduos tenham o mesmo perfil torna-se menor. O ideal é utilizar marcadores genéticos que tenham heterozigose superior a 70%, pois são mais informativos (8).

A heterozigose pode ser esperada ou observada. A heterozigose observada determina a proporção real de heterozigotos para um determinado *locus* na amostra estudada. É calculada diretamente dos genótipos encontrados na população. Trata-se do número de heterozigotos em cada *locus* dividido pelo número total de indivíduos pesquisados. A heterozigose esperada é a fração estimada de heterozigotos para um *locus*, assumindo a condição de equilíbrio de Hardy-Weinberg na população. Difere-se da observada por ser calculada a partir das frequências alélicas segundo a fórmula (29):

$$\text{Heterozigotos} + \text{Homozigotos} = 1$$

$$\text{Heterozigotos} = 1 - \text{Homozigotos}$$

$$\text{Homozigotos} = \text{Soma das frequências alélicas ao quadrado (Freq. Alelo } 1^2 + \text{ Freq. Alelo } 2^2 + \text{ Freq. Alelo } 3^2 + \text{ Freq. Alelo } 4^2 + \dots + \text{ Freq. Alelo } N^2)$$

As heterozigosidades esperada e observada foram obtidas pelo software Arlequin v. 3.1 (28). É importante calculá-las individualmente para constatar se existe redução na heterozigose observada o que pode ser consequência do excesso de homozigotos (devido a um problema na técnica de PCR, pois, o sistema amplifica

preferencialmente um alelo em detrimento de outro) ou ocorrência de endogamia na população (8). Portanto, o ideal é que apresentem valores similares.

3.4.4 Equilíbrio de Hardy-Weinberg

A partir das frequências alélicas observadas obtém-se as frequências genotípicas esperadas considerando que as frequências alélicas não devem modificar ao longo das gerações se o *locus* é estável. Caso a frequência genotípica esperada seja semelhante à frequência genotípica observada pressupõe-se que o *locus* na população analisada está em equilíbrio de Hardy-Weinberg (8). A questão básica desse pressuposto é definir se os alelos encontrados dentro do *locus* são independentes, caso contrário, a população está em desequilíbrio, o qual é calculado pelo teste chi-quadrado pelo software Arlequin v. 3.1 (28).

A população pode estar em desequilíbrio quando se analisa subpopulações ou se tem excesso de homozigose devido a amplificação preferencial um alelo em detrimento de outro (8). Os valores do equilíbrio de Hardy-Weinberg foram corrigidos pelo cálculo de Bonferroni para neutralizar a análise de comparações múltiplas, considerando resultados estatisticamente significantes se $p < 0.0033$ (valor de alfa 0.05 foi dividido pelo número de STRs investigados) (8, 30).

3.4.5 Desequilíbrio de ligação

Denomina-se STRs sintênicos os localizados no mesmo cromossomo. E, caso se utilize STRs sintênicos, o ideal é que estejam a uma distância física de pelo menos 50 megabases entre si. Esta regra é utilizada para garantir a independência entre os *loci*. Para STRs localizados à uma distância menor é necessário calcular o desequilíbrio de ligação, ou seja, a independência entre eles (8, 15). No kit AmpFISTR® NGM™, os STRs vWA e D12S391 são sintênicos (cromossomo 12) e a distância entre eles é de 6.3 megabases. Neste caso, o desequilíbrio de ligação foi calculado para se conhecer a forma com que o vWA e o D12S391 são transmitidos,

já que apenas *loci* independentes devem ser incluídos nos cálculos (15). Para isto o software Arlequin v. 3.1 foi utilizado (28).

3.4.6 Conteúdo de informação polimórfica (CIP)

É utilizado para aferir o quanto um marcador genético é informativo, sendo uma medida capaz de estimar o quão polimórfico é o STR. O grau de polimorfismo é determinado pelo número de alelos e suas frequências (31). Recomenda-se utilizar marcadores com valor de CIP maior que 0,5 (50%) (32), pois possuem maior eficácia em diferenciar indivíduos (33). Esse parâmetro foi mensurado pelo programa PowerStats v. 12 (34), trata-se de um cálculo complexo, e sua fórmula foi descrita por Botstein e colaboradores (32).

3.4.7 Probabilidade de coincidência (“*matching*”) (PM)

É a probabilidade de que dois indivíduos selecionados ao acaso tenham genótipo idêntico para um determinado *locus* (8). É calculado pela soma dos quadrados das frequências genotípicas observadas. Pois, a chance de encontrar um determinado genótipo = sua frequência observada; a chance de encontrar um determinado genótipo e, em seguida, o mesmo genótipo = sua frequência observada ao quadrado. Assim, a soma das frequências observadas ao quadrado é igual a probabilidade de coincidência para o *locus* (35-37). Neste trabalho este parâmetro foi calculado pelo PowerStats v.12 (34). Os valores individuais de cada *locus* podem ser multiplicados para obtenção da probabilidade de coincidência combinada (33), que representa a probabilidade de dois indivíduos não relacionados selecionados randomicamente terem perfis genéticos idênticos para todos os STRs testados.

3.4.8 Poder de discriminação (PD)

É a probabilidade de encontrar dois indivíduos ao acaso na população e eles possuírem perfis genéticos diferentes. Refere-se da probabilidade contrária de se

encontrar dois perfis iguais (probabilidade de coincidência). Ou seja, é calculado pela fórmula $PD = 1 - PM$ (8, 33). Nesta investigação, foi calculado através do programa PowerStats v.12 (34). Os valores individuais de cada *locus* podem ser agrupados no **poder de discriminação combinado** de acordo com as fórmulas: $1 - [(PM_1) \times (PM_2) \times (PM_3) \times \dots \times (PM_N)]$ ou $1 - (PM \text{ combinado})$, o qual foi executado no Microsoft Excel.

3.4.9 Poder de exclusão (PE)

Corresponde à probabilidade de se excluir qualquer indivíduo que tenha sido falsamente acusado em teste de paternidade (33). É equivalente a frequência de todos os homens, em uma determinada população, que não tem os alelos concordantes com o alelo paterno obrigatório (alelo herdado do pai biológico) da criança em um teste de paternidade (8, 33). O valor é determinado pela seguinte fórmula: $PE = h^2(1-2hH^2)$, onde h significa a frequência observada de heterozigotos e H é a frequência observada de homozigotos (33). Este parâmetro foi calculado pelo programa PowerStats v. 12 (34). Os valores individuais de cada *locus* podem ser agrupados no **poder de exclusão combinado** pelo Microsoft Excel seguindo a fórmula: $1 - [(1 - PE_1) \times (1 - PE_2) \times (1 - PE_3) \times \dots \times (1 - PE_N)]$ (8, 38).

3.4.10 Índice de paternidade típico (IPT)

Esse parâmetro estatístico de estudo de vínculo genético determina a probabilidade de que o indivíduo que está sendo testado seja realmente o pai biológico da criança, do que um homem selecionado aleatoriamente. Os índices de paternidade (IP) de cada *locus* representam a razão entre a probabilidade do suposto pai e a probabilidade de qualquer outro indivíduo da população, transmitir o alelo em questão para o filho. A probabilidade de qualquer outro indivíduo transmitir o alelo paterno obrigatório corresponde a frequência do alelo na população avaliada. O **índice de paternidade combinado (IPC)** é o produto dos índices de paternidade individuais para cada *locus* investigado e determina a força da evidência genética do suposto pai ser o pai biológico. Dessa forma, quanto maior o índice maior é a evidência genética (8, 39).

A partir das frequências alélicas é possível calcular o **índice de paternidade típico** para cada *locus* na população avaliada, e corresponde ao valor para o índice de paternidade mais comum a ser observado na rotina (33). Este parâmetro foi calculado pelo programa PowerStats v. 12, através desta fórmula: $IP = (H + h)/2H$, onde *h* significa a frequência de observada de heterozigotos e *H* é a frequência observada de homozigotos (34).

Além disso, calcula-se a probabilidade de paternidade. No cálculo estatístico é obrigatório o uso de uma probabilidade “*a priori*”. Como os peritos desconhecem estas evidências não-genéticas, e devem permanecer neutros na disputa, utilizam o valor de 0,5 (50%) em seus cálculos para que não exista tendência matemática prévia a favor ou contra a paternidade. A probabilidade de paternidade é calculado da seguinte forma (39):

$$\begin{aligned} \text{Probabilidade de paternidade} &= 0,5 \times IPC / [0,5 \times (IPC) + (1-0,5)] \\ &= 0,5 \times 10.000 / [0,5 \times (10.000) + (1-0,5)] \\ &= 5.000 / 5.000 + 0,5 \\ &= 0,9999 \times 100\% \\ &= 99,99\% \end{aligned}$$

A fórmula pode ser simplificada para: $IPC / (IPC + 1)$ (39). No geral, o valor obtido é dado em percentagem (multiplicado por 100) e os testes de vínculo genético tem como regra para aceitação de paternidade uma probabilidade acima de 99,99% que implica em um IPC maior que 10.000 (8, 39).

3.4.11 Distância genética

Com base nas frequências alélicas foram calculadas as distâncias genéticas da população estudada (Distrito Federal) em relação às cinco macrorregiões brasileiras (Norte, Nordeste, Sul, Sudeste e Centro-Oeste) e o Brasil como um todo. A finalidade desta análise foi investigar a semelhança genética entre as populações analisadas. As frequências alélicas para cinco macrorregiões brasileiras, e do Brasil como um todo foram obtidas do trabalho de Raimann e colaboradores (21). Em seguida, as distâncias genéticas segundo a metodologia de Nei (1983) foram

mensuradas pelo software Dispan (40), devido ao fato desse programa utilizar com frequências alélicas, diferentemente dos demais softwares para mesma finalidade que exigem genótipos. Normalmente, os genótipos não são publicados.

A matriz de distância genética gerada pelo Dispan correlaciona as diferenças entre as populações analisadas e para facilitar a sua interpretação, os dados foram representados graficamente pela análise de coordenadas principais (PCoA) do software GenAlEx v.6.5 (41). Nesta análise os quinze STRs foram considerados. Resumidamente, este tipo de gráfico dispõe os valores das distâncias em um espaço dimensional dividido em 04 quadrantes. As populações que concentrarem-se no mesmo quadrante são mais semelhantes. A distância entre os pontos representativos de cada população também indicam proximidade genética, quanto mais próximos mais semelhantes.

Além disso, a segunda forma para se visualizar os dados da matriz de distâncias genéticas utilizada no estudo foi o dendrograma elaborado pelo software Dispan (40). Através do método chamado de *neighbor-joining* o programa faz uma reconstrução na forma de árvore filogenética com base nos valores das distâncias. O principal objetivo é encontrar pares, no caso, regiões que diminuam o comprimento do ramo que é formado a depender da similaridade construindo, assim, um gráfico de ramificações (42).

3.5 ASPECTOS ÉTICOS

O projeto de pesquisa intitulado como “Análise genética populacional de 22 microssatélites no Distrito Federal”, que também contempla a análise de STRs do Cromossomo Y não incluídos nesta dissertação, foi encaminhado ao Comitê de Ética da Faculdade de Ciências da Saúde da Universidade de Brasília (UnB), e aprovado em (13/12/2012), através do protocolo 165/12 (Anexo I), de acordo com o estabelecido pela Resolução nº 196/96 (Brasil, 1996). Por se tratar de análise de banco de dados, houve a dispensa do termo de consentimento informado específico.

4 RESULTADOS

4.1 A AMOSTRA POPULACIONAL DO ESTUDO REPRESENTA A POPULAÇÃO DO DISTRITO FEDERAL

Inicialmente, foi questionado se a amostra populacional originada de uma base de dados representaria a composição da população do Distrito Federal. Neste sentido, comparou-se a distribuição da naturalidade dos 462 indivíduos incluídos no presente estudo com a do último Censo (2010) (3) (Tabela 1). A comparação revelou que não existia diferença estatística significativa entre as proporções ($p = 0,26$).

Tabela 1 – Comparação entre as distribuições do local de nascimento da amostra populacional extraída do banco de dados e da população do Distrito Federal de acordo com o Censo 2010

	Este trabalho	%	Censo 2010	%
Centro-Oeste	33	7,14	164.215	6,50
Nordeste	116	25,11	602.104	23,50
Norte	13	2,81	47.783	2,00
Sudeste	66	14,29	321.368	12,50
Sul	5	1,08	38.634	1,50
Distrito Federal	229	49,57	1.382.017	54,00
Total	462	100	2.556.121	100

chi-quadrado = 6,419, graus de liberdade = 5, $p = 0,26$

4.2 FREQUÊNCIAS ALÉLICAS

As frequências alélicas dos quinze *loci* STRs avaliados na amostra populacional estão apresentadas na Tabela 2.

4.3 HETEROZIGOSE

A heterozigose observada variou entre 0,753 (D22S1045) a 0,876 (D2S1338 e FGA), a heterozigose esperada variou entre 0,745 (D22S1045) a 0,892 (D1S1656) e houve uma considerável similaridade entre os dois parâmetros em todos os *loci* (Tabela 3).

Tabela 3 – Heterozigose observada e heterozigose esperada dos quinze STRs incluídos no kit NGM™ em uma amostra populacional do Distrito Federal (n = 462)

<i>Locus</i>	Heterozigose observada	Heterozigose esperada
D10S1248	0,7810	0,7830
vWA	0,8030	0,8050
D16S539	0,7960	0,7860
D2S1338	0,8760	0,8790
D8S1179	0,8050	0,8080
D21S11	0,8240	0,8540
D18S51	0,8700	0,8840
D22S1045	0,7530	0,7450
D19S433	0,8110	0,8150
TH01	0,7980	0,7940
FGA	0,8760	0,8740
D2S441	0,7660	0,7670
D3S1358	0,7790	0,7800
D1S1656	0,8650	0,8920
D12S391	0,8700	0,8870

4.4 EQUILÍBRIO DE HARDY - WEINBERG

As frequências genótípicas demonstraram que os STRs avaliados estão em equilíbrio de Hardy-Weinberg ($p > 0,05$), exceto o *locus* D2S441 ($p = 0,011$). No entanto, este desvio não permaneceu significativo após aplicação da correção de Bonferroni ($p = 0,05 / 15 = 0,0033$) (Tabela 4).

Tabela 4 – Resultados do valor de P para o equilíbrio de Hardy-Weinberg dos quinze STRs incluídos no kit NGM™ em uma amostra populacional do Distrito Federal (n = 462)

Locus	Valor de P
D10S1248	0,2970
vWA	0,5380
D16S539	0,8520
D2S1338	0,7960
D8S1179	0,9060
D21S11	0,2680
D18S51	0,4510
D22S1045	0,9950
D19S433	0,7400
TH01	0,8160
FGA	0,0680
D2S441	0,0110*
D3S1358	0,4620
D1S1656	0,3080
D12S391	0,9650

* valor de $p < 0,05$

4.5 DESEQUILÍBRIO DE LIGAÇÃO

Não foi detectado evidência de desequilíbrio de ligação entre os *loci* vWA e D12S391 ($p = 0,18$) refletindo uma independência entre os seus respectivos alelos.

4.6 PARÂMETROS FORENSES

O **conteúdo de informação polimórfica** variou entre 0,7056 e 0,8813. Os cinco *loci* mais polimórficos foram D1S1656, D12S391, D18S51, D2S1338, e FGA. Os cinco *loci* menos polimórficos foram D22S1045, D2S441, D3S1358, D10S1248, D16S539 (Tabela 5).

A **probabilidade de coincidência (“*matching*”)** combinada foi $4,85 \times 10^{-20}$, sendo esta a probabilidade de selecionar dois indivíduos aleatoriamente com perfis iguais. As probabilidades de coincidência para cada *locus* podem ser evidenciadas na tabela 5.

Tabela 6 – Parâmetros de paternidade para os quinze STRs incluídos no kit NGM™ em uma amostra populacional do Distrito Federal (n = 462)

<i>Locus</i>	Poder de exclusão	Índice de paternidade típico
D10S1248	0,5650	2,2871
vWA	0,6047	2,5385
D16S539	0,5926	2,4574
D2S1338	0,7480	4,0526
D8S1179	0,6087	2,5667
D21S11	0,6456	2,8519
D18S51	0,7349	3,8500
D22S1045	0,5153	2,0263
D19S433	0,6209	2,6552
TH01	0,5966	2,4839
FGA	0,7480	4,0526
D2S441	0,5379	2,1389
D3S1358	0,5611	2,2647
D1S1656	0,7262	3,7258
D12S391	0,7349	3,8500

4.8 DISTÂNCIA GENÉTICA

A tabela 7 mostra as distâncias genéticas entre as populações estudadas.

Tabela 7 - Matriz da distância genética de Nei (DA) calculada pelo software DISPAN a partir das frequências alélicas deste estudo e de Raimann e colaboradores (21) (n = 462)

	Brasil	CO	ND	N	SD	S	DF
Brasil	0						
CO	0,0114	0					
ND	0,0071	0,0206	0				
N	0,0166	0,0265	0,0266	0			
SD	0,0045	0,0188	0,0123	0,0232	0		
S	0,0043	0,0141	0,0152	0,0198	0,0122	0	
DF	0,0043	0,0146	0,0098	0,0204	0,0086	0,0078	0

CO - Centro-oeste, ND - Nordeste, N - Norte, SD - Sudeste, S – Sul, DF – Distrito Federal

De acordo com a análise de coordenadas principais destas distâncias genéticas o Distrito Federal, a região Sudeste, Nordeste e o território nacional foram agrupados próximos, e em um mesmo quadrante, sugerindo uma maior semelhança genética (Figura 1). A região Sul se posicionou à uma distância intermediária e em um outro quadrante. Por fim, as regiões Norte e Centro-Oeste se posicionaram mais distantes, e em quadrantes distintos do agrupamento que inclui o Distrito Federal, o que sugere uma menor semelhança genética.

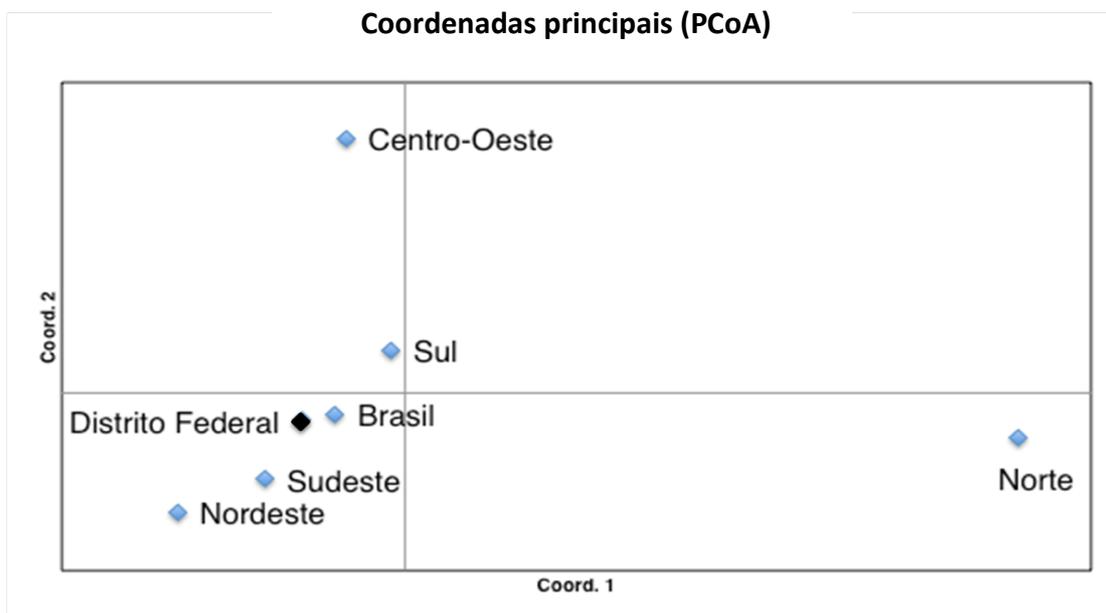


Figura 1 - Análise de coordenadas principais das distâncias genéticas de Nei (DA) entre as populações estudadas. Dados da matriz de distância genética (tabela 7)

O dendrograma gerado pelo software Dispan a partir das distâncias genéticas (Figura 2) corrobora com o gráfico de coordenadas principais (Figura 1). O Distrito Federal, Sudeste, Nordeste e Brasil como um todo situam-se na mesma ramificação. O Sul e Centro-Oeste na mesma ramificação, mas distantes entre si, sendo que o Sul está posicionado próximo ao Brasil como um todo. Por fim, o Norte em uma ramificação distinta e distante das demais.

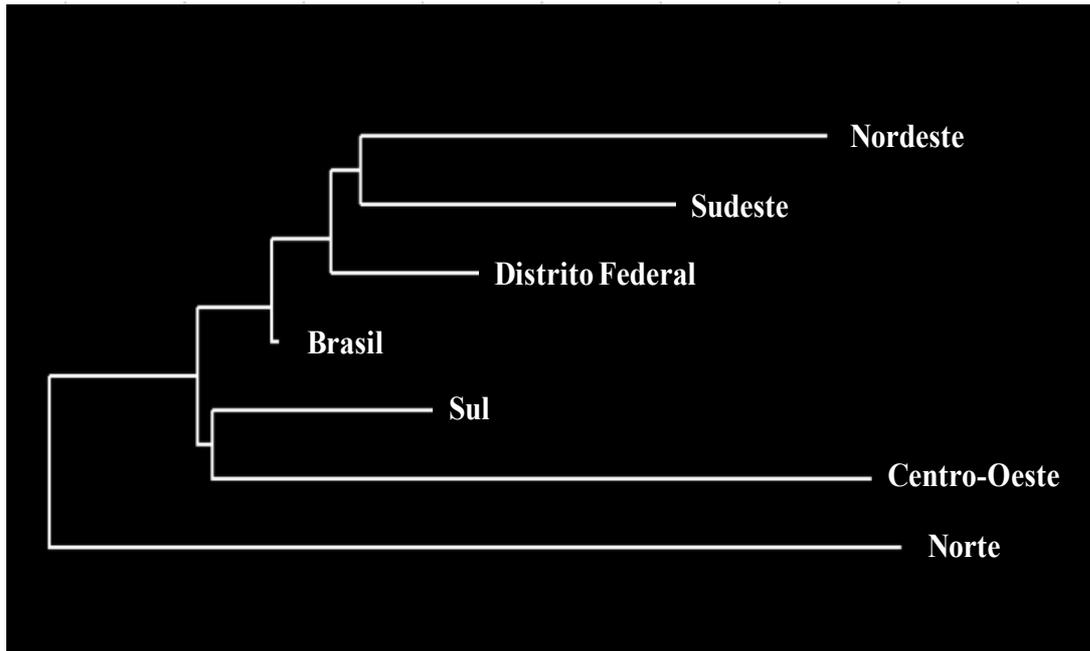


Figura 2 - Dendrograma gerado pelo software Dispan (método *neighbor-joining*) a partir da matriz de distâncias genéticas de Nei (DA) das populações estudadas

5 DISCUSSÃO

Este trabalho visa descrever as frequências alélicas e parâmetros forenses de quinze marcadores genéticos do tipo STR, para uma amostra populacional do Distrito Federal e, além disso, investigar a composição desta população utilizando como ferramenta o DNA de seus indivíduos. Os STRs são marcadores genéticos extremamente polimórficos e utilizados para identificação humana e investigação de vínculo genético. Recomenda-se o uso de frequências alélicas locais para os cálculos envolvidos nestas análises, como probabilidade de coincidência e índice de paternidade.

Entretanto, o Distrito Federal é um território artificial inaugurado à 53 anos em um local até então despovoado. Sua população fundadora era composta apenas por imigrantes e, historicamente, as principais migrações aconteceram da região Nordeste e Sudeste (4). Atualmente, o Distrito Federal possui mais de 2.500.000 de habitantes, e metade destas pessoas nasceram localmente. Este fato reforça ainda mais a necessidade de se conhecer as frequências alélicas da região.

Ademais, a composição da população brasileira é complexa (mista) e com contribuições de povos distintos (europeu, nativo americano e africano) e em diferentes proporções. Estudos recentes revelam uma maior contribuição dos nativos americanos para as frequências alélicas da região Norte, africana para a região Nordeste e europeia para as demais (1, 5-7). Assim, o Distrito Federal é uma população mista formada a partir de populações heterogêneas.

5.1 A AMOSTRA POPULACIONAL DO ESTUDO REPRESENTA A POPULAÇÃO DO DISTRITO FEDERAL

Neste estudo, a amostra populacional do Distrito Federal foi originada a partir dos indivíduos que procuraram um laboratório clínico local, para realização de exame de vínculo genético de paternidade. Assim, trata-se de uma amostragem de conveniência. O banco de dados de STRs do laboratório foi analisado e as frequências alélicas e parâmetros forenses foram descritos em um contexto populacional e de forma agregada conforme aprovado pelo comitê de ética. Por isto,

foi necessário comparar a distribuição do local de nascimentos (macrorregiões) dos indivíduos incluídos no estudo com a distribuição de naturalidade da população relatada no Censo 2010. Não foi observado diferença significativa nesta análise e conclui-se que a amostra populacional utilizada nessa investigação representa a população do Distrito Federal. Em ambas populações a maior parte dos indivíduos era nascida no próprio Distrito Federal, seguidos das regiões Nordeste e Sudeste.

5.2 FREQUÊNCIAS ALÉLICAS

Os quinze marcadores genéticos avaliados foram: D10S1248, D22S1045, D2S441, D1S1656, D12S391, D2S1338, D3S1358, D8S1179, D16S539, D18S51, D19S433, D21S11, FGA, TH01 e vWA. Suas frequências alélicas se encontram na tabela 2 e, atualmente, são as utilizadas na rotina do laboratório clínico que as originou, em substituição às do território nacional, outrora utilizadas (21). Portanto, o presente estudo é de grande relevância por caracterizar uma amostragem significativa (462 indivíduos) e específica do território. Assim, as frequências alélicas locais descritas neste trabalho agregam maior confiança estatística aos cálculos forenses e de vínculo genético.

5.3 PARÂMETROS FORENSES

A diversidade genética dos marcadores foi primeiramente investigada por meio da heterozigose esperada e observada. Estes parâmetros foram maiores que 70% para todos os *loci*, o que confere a população genótipos diversificados e, conseqüentemente, maior capacidade do conjunto de STRs em diferenciar indivíduos. Ademais, os valores esperados e observados foram próximos sugerindo ausência de problemas inerentes à técnica de PCR e endogamia na população.

Em seguida, a análise do equilíbrio de Hardy-Weinberg revelou que todos os *loci* analisados apresentavam-se estáveis ao longo das gerações, exceto o D2S441 ($p = 0,011$). Entretanto, a significância foi perdida após correção de Bonferroni para análises múltiplas, já que o valor de P para todos os *loci* são analisados simultaneamente pelo software Arlequin. Nestas circunstâncias, o conjunto de STRs analisados está em equilíbrio de Hardy-Weinberg tornando possível seu uso em

teste de paternidade e identificação humana. Ademais, Rodenbusch e colaboradores ao investigarem a população do Rio Grande do Sul ($n = 353$) também notaram que o marcador D2S441 estava em desequilíbrio ($p = 0,022$) (22). Pontes e colaboradores ao estudarem a população do Norte de Portugal ($n = 213$) constataram esse mesmo dado ($p = 0,008$) (16). No entanto, ao analisar a população brasileira total ($n = 835$), Raimann e colaboradores evidenciaram equilíbrio ($p = 0,68$) (21), talvez em decorrência do maior tamanho amostral. Em todos os casos supracitados o kit NGM™ utilizado foi o mesmo deste trabalho.

O teste para desequilíbrio de ligação entre vWA e D12S391 mostrou independência entre os dois *loci*. Os achados corroboram com outros quatro estudos (11, 15, 21, 43), e supõe-se que essa independência seja em consequência das elevadas taxas de mutação destes STRs ou presença de um sítio de recombinação entre os *loci* apesar da proximidade (11).

O conteúdo de informação polimórfica, probabilidade de coincidência e poder de discriminação são parâmetros forenses relacionados. O *locus* que apresentar maior conteúdo de informação polimórfica, ou seja maior número de alelos encontrados, também será o com menor probabilidade de coincidência e maior poder de discriminação. Dois dos novos *loci* recomendados pela comunidade forense europeia (D1S1656 e D12S391) encontram-se entre os cinco com maior conteúdo de informação polimórfica no kit NGM™, juntamente com (D18S51, D2S1338, e FGA). Portanto, são *loci* altamente polimórficos e recomendados também para a população do Distrito Federal.

Por motivo de comparação, os parâmetros forenses dos conjuntos de STRs (kits) são normalmente descritos e comparados de forma combinada. Os índices de probabilidade de coincidência combinada e poder de discriminação combinado, juntamente com os parâmetros de paternidade, esclarecem as características analíticas do kit. Assim, na população do Distrito Federal, a probabilidade de coincidência combinada (probabilidade de ocorrência de 02 perfis de DNA idênticos ao acaso) é $4,85 \times 10^{-20}$. Para uma melhor contextualização deste valor, a probabilidade de ser sorteado na megasena com uma única tentativa é de $1,1 \times 10^{-8}$ (44). Assim, é $2,2 \times 10^{11}$ (220 bilhões) vezes mais provável ganhar na megasena do que encontrar dois indivíduos com o mesmo genótipo ao acaso utilizando o conjunto de quinze STRs deste trabalho.

Raimann e colaboradores demonstraram que a probabilidade de coincidência do kit NGM™ na população brasileira é de $4,0 \times 10^{-20}$, assemelhando-se ao dado obtido neste trabalho, pois estão dentro da mesma ordem de grandeza (21). Os outros dois kits comerciais mais utilizados na população brasileira descreveram que a probabilidade de coincidência combinada é de $1,15 \times 10^{-18}$ (Identifiler™ – Life Technologies) e $2,08 \times 10^{-19}$ (PowerPlex® 16 - Promega), respectivamente (45, 46). Apesar de pouca disparidade no desempenho desses ensaios comerciais, nota-se que o NGM™ apresenta menor valor de probabilidade de coincidência que os demais kits.

Neste estudo, o poder de discriminação combinado, ou seja, probabilidade de diferenciar indivíduos escolhidos aleatoriamente, foi de 0,9999999999999999995 ou 99,9999999999999995% (18 casas após a vírgula). Para facilitar a compreensão, quanto mais próximo de 100%, melhor a capacidade de diferenciação. Raimann e colaboradores ao analisar a população brasileira total obteve 99,9999999999999996% (17 casas após a vírgula) (21). Este parâmetro para o kit PowerPlex® 16 também na população brasileira total é 99,9999999999999997% (17 casas após a vírgula) (46) e para o Identifiler™ é 99,9999999999999998% (16 casas após a vírgula) (45). O kit NGM™ na população do Distrito Federal apresenta parâmetros forenses melhores que o Identifiler™, o PowerPlex® 16, e inclusive o próprio kit na população brasileira como um todo.

5.4 PARÂMETROS DE PATERNIDADE

O poder de exclusão combinado obtido neste trabalho é de 0,9999998 ou 99,99998%, valor idêntico ao obtido por Raimann e colaboradores (99,99998%) (21). Para os kits Identifiler™ e o PowerPlex® 16 na população brasileira o valor do poder de exclusão acumulado é 99,99992% (45) e 99,99995%(46), respectivamente.

Faz-se necessário esclarecer as diferenças entre o poder de exclusão e o poder de discriminação. Pois, são os dois parâmetros que resumem as características analíticas dos kits para identificação humana. O poder de exclusão é a capacidade do conjunto de STRs excluir um indivíduo falsamente acusado em um teste de paternidade. O poder de discriminação é a capacidade do conjunto de STRs diferenciar um indivíduo do outro. Para discriminar um indivíduo de um outro

compara-se o genoma diplóide, ou seja, os genótipos completos (os dois alelos para o *locus*). Por outro lado, o poder de exclusão considera o genoma haplóide, ou seja, a presença ou ausência do alelo paterno obrigatório e diferentes genótipos podem conter o alelo paterno obrigatório. Por isto, o poder de exclusão normalmente é menor que o poder de discriminação.

Extrapolando o conceito de índice de paternidade combinado para os índices de paternidade típicos obtêm-se o valor de 6131861,1, o qual corresponde a uma probabilidade de paternidade acumulada de 99,999984%. Sendo esses os valores típicos que serão obtidos na rotina diária. Lembrando que o índice mínimo aceitável é de 10.000 que corresponde a 99,99% (39). A mesma extrapolação para o trabalho de Raimann e colaboradores resulta em 6157696,78, o qual corresponde à 99,999983% (21). Já para o trabalho de Aguiar e colaboradores (PowerPlex® 16) foi obtido o valor de 2221267,4, o qual corresponde à 99,999955% (46). Portanto, parâmetros praticamente idênticos nestas descrições.

Conclui-se que o kit NGM™ na população do Distrito Federal apresenta parâmetros de paternidade os quais atendem as exigências das sociedades forenses.

5.5 DISTÂNCIA GENÉTICA

A formação recente e não-natural do território faz do Distrito Federal um importante objeto de estudo antropológico e genético, por se tratar de uma população jovem e originada de migrações. Historicamente, sabe-se que as principais migrações ocorreram do Nordeste e do Sudeste. Entretanto, este estudo de frequências alélicas permite a comprovação deste dado a partir de uma evidência científica, no caso, genética. As distâncias genéticas de Nei (DA) foram calculadas a partir das frequências alélicas descritas por Raimann (21) e as originadas neste trabalho. Como os valores das distâncias correspondem a comparação das populações aos pares, o formato da matriz dificulta a visualização e a interpretação dos resultados. Por isto, usa-se o gráfico de coordenadas principais e/ou dendrograma. A comparação entre as duas representações gráficas revela uma concordância completa.

Os resultados indicam que a população do Distrito Federal é mais próxima geneticamente do Brasil, Sudeste e Nordeste. Resultado que corrobora com as descrições históricas e também com os locais de nascimento conforme o Censo populacional (3). A região Sul apresenta-se com uma semelhança menor quando comparada com Nordeste e Sudeste. Interessantemente, o Centro-Oeste local onde o território foi construído posicionou-se relativamente distante, o que sugere uma menor semelhança genética com o Distrito Federal quando comparado com Sudeste, Nordeste e o Sul. Por fim, a região Norte se posicionou distante de todas as demais regiões. Este resultado corrobora com a descrição de que a mistura entre o nativo americano, africano e europeu ocorreu em diferentes proporções nas cinco macrorregiões (1, 5-7).

6 CONCLUSÃO

- As frequências alélicas, os parâmetros forenses e de paternidade revelam que AmpFISTR® NGM™ é um sistema genético altamente informativo para a população do Distrito Federal. O poder de discriminação combinado e o poder de exclusão combinado foram 99,99999999999999995% e 99,99998%, respectivamente.
- O Distrito Federal é geneticamente mais próximo das populações do Sudeste, Nordeste e Brasil como um todo do que das populações do Sul, Centro-Oeste e Norte.

7 REFERÊNCIAS

1. Instituto Brasileiro de Geografia e Estatística (IBGE): 2010;[cited 26 de Janeiro de 2013]. Available from: <http://www.ibge.gov.br/estadosat/>.
2. Habermalz U. Brasília - eine Stadt aus dem Nichts: auf den Spuren Oscar Niemeyers. Berner Fachhochschule 2006.
3. Censo Demográfico (2010) Instituto Brasileiro de Geografia e Estatística (IBGE) 2010 [cited 26 de Janeiro de 2013]. Available from: <http://www.ibge.gov.br/estadosat/perfil.php?sigla=df>.
4. Codeplan. Evolução dos movimentos migratórios para o Distrito Federal: 1959- 2010. Demografia Em Foco. Mauro Guimarães Moncaio ed. Brasília 2013. p. 79.
5. Godinho NMO, Gontijo CC, Diniz MCEG, Falcão-Alencar G, Dalton GC, Amorim CEG, et al. Regional patterns of genetic admixture in South America. Forensic Science International: Genetics 2008;1(1):329–30.
6. Cordeiro Q, Souza BR, Correa H, Guindalini C, Hutz MH, Vallada H, et al. A review of psychiatric genetics research in the Brazilian population. Rev Bras Psiquiatr. 2009;31(2):154-62.
7. Pena SD, Bastos-Rodrigues L, Pimenta JR, Bydlowski SP. DNA tests probe the genomic ancestry of Brazilians. Braz J Med Biol Res. 2009;42(10):870-6.
8. Butler JM. Forensic DNA Typing: Biology, Technology, and Genetics of STR markers. Second Edition ed. Press EA, editor 2005.
9. Butler JM. Forensic DNA testing. Cold Spring Harb Protoc. 2011;(12):1438-50.
10. Schneider PM. Expansion of the European standard set of DNA database loci—the current situation. Institute of Legal Medicine, University Hospital of Cologne, Germany: Profiles in DNA; 2009. p. 6-7.

11. Budowle B, Ge J, Chakraborty R, Eisenberg AJ, Green R, Mulero J, et al. Population genetic analyses of the NGM STR loci. *Int J Legal Med*. 2011;125(1):101-9.
12. Butler JM. Biology and Genetics of New Autosomal STR Loci Useful for Forensic DNA Analysis. *Forensic Sci Rev* [Internet]. 2012. Available from: http://www.cstl.nist.gov/strbase/pub_pres/Butler-Hill-FSR2012-newSTRloci.pdf.
13. Oldroyd N. Implementation of Next-Generation STR Kits in Europe: Challenges and Solutions In: *Green R, Wang D, Lagace R, Norona W, Mulero J, Hennessy L, et al., editors.*: Life Technologies Corporation.; 2011.
14. Butler JM. Fundamentals of forensic DNA typing. Elsevier Academic Press; 2010.
15. O'Connor KL, Tillmar AO. Effect of linkage between vWA and D12S391 in kinship analysis. *Forensic Sci Int Genet*. 2012;6(6):840-4.
16. Pontes ML, Pinheiro MF. Population data of the AmpFISTR® NGM™ STR loci in a North of Portugal sample. *Forensic Sci Int Genet*. 2012;6(5):e127-8.
17. Ribeiro T, Dario P, Vital N, Sanches S, Espinheira R, Geada H, et al. Population data of the AmpFISTR® NGM™ loci in South Portuguese population. *Forensic Sci Int Genet*. 2013;7(2):e37-9.
18. Parys-Proszek A, Kupiec T, Wolańska-Nowak P, Branicki W. Genetic variation of 15 autosomal STR loci in a population sample from Poland. *Leg Med (Tokyo)*. 2010;12(5):246-8.
19. Westen AA, Haned H, Grol LJ, Harteveld J, van der Gaag KJ, de Knijff P, et al. Combining results of forensic STR kits: HDplex validation including allelic association and linkage testing with NGM and Identifiler loci. *Int J Legal Med*. 2012;126(5):781-9.
20. Jedrzejczyk M, Jacewicz R, Szram S, Berent J. Genetic population studies on 15 NGM™ STR loci in central Poland population. *Forensic Sci Int Genet*. 2012; 6(4):e119-20.

21. Raimann PE, Hirschfeld-Campolongo G, Schumacher S, de Mattos Almeida SE, da Silva CM, Rodenbusch R, et al. Population genetic analyses of the AmpFISTR® NGM™ in Brazil. *Int J Legal Med.* 2012;126(2):337-41.
22. Rodenbusch R, Schumacher S, Mardini AC, Chula FG, Michelon CT, Gastaldo AZ, et al. Allele frequencies of the five new generation forensic STR (D1S1656, D2S441, D10S1248, D12S391 and D22S1045) in the population from Rio Grande do Sul, Southern Brazil. *Forensic Sci Int Genet.* 2012;6(1):e55-7.
23. Takezaki N, Nei M. Genetic distances and reconstruction of phylogenetic trees from microsatellite DNA. *Genetics.* 1996;144(1):389-99.
24. Takezaki N, Nei M. Empirical tests of the reliability of phylogenetic trees constructed with microsatellite DNA. *Genetics.* 2008;178(1):385-92.
25. Butler JM. *Advanced topics in forensic DNA typing: methodology 2012.*
26. AmpFISTR® NGM™ PCR Amplification Kit User Guide. Life Technologies Corporation - Applied Biosystems; 2012.
27. Software Graphpad Prism 6.0 2013. Available from: <http://graphpad.com/quickcalcs/chisquared1.cfm>.
28. Excoffier L, Lischer HE. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Mol Ecol Resour.* 2010;10(3):564-7.
29. McMANUS C, PAIVA S, CORRÊA PS, SEIXAS L, MELO CB. *Estatísticas para descrever genética de populações. INCT: Informação genético-sanitária da pecuária brasileira;* 2011.
30. Rice WR. Analyzing tables of statistical tests. *Evolution.* 43. 1989. p. 223-5.
31. Guo X, Elston RC. Linkage information content of polymorphic genetic markers. *Hum Hered.* 1999;49(2):112-8.

32. Botstein D, White RL, Skolnick M, Davis RW. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am J Hum Genet.* 1980;32(3):314-31.
33. Huston KA. Statistical Analysis of STR Data. *Profiles in DNA.* 1998:14-5.
34. Tereba A. Tools for analysis of population statistics. 1999. p. 14-6.
35. Fisher RA. Standard calculations for evaluating a blood-group system. *Heredity (Edinb).* 1951;5(1):95-102.
36. Jones BJ, Brewer JK. An analysis of the power of statistical tests reported in the *Research Quarterly.* *Res Q.* 1972;43(1):23-30.
37. Tillmar A. *Populations and Statistics in Forensic Genetics.* Linköping University, Faculty of Health Sciences 2010.
38. Scientific Working Group on DNA Analysis Methods (SWGDM): *Interpretation Guidelines for Autosomal STR Typing by Forensic DNA Testing Laboratories.* United States 2010.
39. AABB. *Parentage Testing Accreditation Requirements Manual.* 3 ed. Bethesda, Md.: Walker R.H; 1998.
40. Ota T. *DISPAN: Genetic distance and phylogenetic analysis* Pennsylvania 1993. p. Institute of Molecular Evolutionary Genetics, Pennsylvania State University.
41. Peakall R, Smouse PE. GenAlEx 6.5: genetic analysis in Excel. *Population genetic software for teaching and research--an update.* *Bioinformatics.* 2012; 28(19):2537-9.
42. Saitou N, Nei M. The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol.* 1987;4(4):406-25.

43. Phillips C, Fernandez-Formoso L, Garcia-Magariños M, Porras L, Tvedebrink T, Amigo J, 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.* 2011;5(3):155-69.
44. Caixa Econômica Federal 2013 [cited 2013]. Available from: <http://www1.caixa.gov.br/loterias/loterias/megasena>.
45. de Assis Poiares L, de Sá Osorio P, Spanhol FA, Coltre SC, Rodenbusch R, Gusmão L, et al. Allele frequencies of 15 STRs in a representative sample of the Brazilian population. *Forensic Sci Int Genet.* 2010;4(2):e61-3.
46. Aguiar VR, Wolfgramm EeV, Malta FS, Bosque AG, Mafia AeC, Almeida VC, 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.* 2012;6(4):504-9.

APÊNDICE I – ARTIGO CIENTÍFICO

(Publicado no periódico científico *International Journal of Legal Medicine*)

Int J Legal Med
DOI 10.1007/s00414-013-0866-z

POPULATION DATA

Allele frequencies for 15 autosomal STRs in a population sample from the Federal District (Brazil)—a territory that arose from nothing

Gustavo Barcelos Barra · Ticiane Henriques Santa Rita · Camilla Figueiredo Chianca · Lara Franciele Ribeiro Velasco · Claudia Ferreira de Souza

Received: 8 February 2013 / Accepted: 24 April 2013
© Springer-Verlag Berlin Heidelberg 2013

Abstract The Federal District (Brazil) was created in 1960 in the Central-West Region of Brazil in a previously unpopulated area. In 2010, this artificially founded district was populated by 2,562,963 inhabitants. In this study, the genetic variations of the 15 Next Generation Multiplex (NGMTM) short tandem repeat loci were analyzed. The results indicate that the NGMTM is a highly informative genetic system in this population, which is more similar to the southeastern, northeastern, and overall Brazil populations. This conclusion agrees with the population composition reported in the 2010 National Survey Inquiries, in which most of the immigrants were from the northeast and the southeast.

Keywords Short tandem repeats · Brazil · Brasília · Federal District (Brazil) · NGMTM · Forensic genetics

Brazil is a continental country divided into five macroregions: North, Northeast, Central-West, Southeast, and South [1]. Brazil's Federal District and Brasília, the actual federal capital, were founded on April 21, 1960 based on a desire to transfer the federal capital from Rio de Janeiro to the center of the country. The Federal District and Brasília were built in only

4 years in previously unpopulated territory in Brazil's Central-West Region as a result of the efforts of President Juscelino Kubitschek, together with the architect Oscar Niemeyer, the urbanist Lúcio Costa, and thousands of civil workers, known as “candangos” [2]. According to the most recent National Survey Inquiries in 2010, the Federal District population was comprised of 2,562,963 inhabitants; 54 % of people were born locally, and the remaining 46 % were immigrants from the other Brazilian macroregions: North (2 %), Northeast (23 %), Central-West (6 %), Southeast (13 %), and South (2 %) [3]. Brazil was colonized by Portugal, and the admixture of Native Americans, Africans, and Europeans occurred in different proportions among the five macroregions [4]. In this work, we investigated the forensic efficiency parameters and the genetic polymorphism of 15 short tandem repeat (STR) loci (D10S1248, D22S1045, D2S441, D1S1656, D12S391, D2S1338, D3S1358, D8S1179, D16S539, D18S51, D19S433, D21S11, FGA, TH01, and vWA) in individuals from Brazil's Federal District. To this end, anonymized STR profiles of 462 unrelated adults (230 females and 232 males) residing in the Federal District were retrieved from the Sabin Laboratory paternity testing results database; 49.57 % of these people were born locally, and 50.43 % immigrated from other Brazilian macroregions. The complete distribution of birth-place for all participants, given in Table S1, was similar to the distribution in the 2010 National Survey Inquiries ($\chi^2=7.641$, $df=5$, $p=0.17$). The allele frequencies and forensic parameters were analyzed and described in a population context (aggregated form). The University of Brasília Health Science Faculty Ethical Committee approved this study and waived the need for specific informed consent (protocol 165/12). This paper follows the guidelines for the publication of population data requested by the journal [5], and Sabin Laboratory is ISO9001/2008-certified and participates in GHEP/ISFG proficiency testing. For routine kinship analysis, blood samples were collected from the ring finger tip on a

Gustavo Barcelos Barra and Ticiane Henriques Santa Rita contributed equally to this work.

Electronic supplementary material The online version of this article (doi:10.1007/s00414-013-0866-z) contains supplementary material, which is available to authorized users.

G. B. Barra · L. F. R. Velasco · C. F. de Souza
Sabin Laboratory and Institute, SCLRN 710/711 bloco A,
Brasília, DF 70.750-531, Brazil

G. B. Barra (✉) · T. H. Santa Rita · C. F. Chianca
Programa de Pós-Graduação em Ciências da Saúde,
Universidade de Brasília, Campus Darcy Ribeiro,
Brasília, DF 70.910-900, Brazil
e-mail: gbbarra@gmail.com

Published online: 04 May 2013

 Springer

APÊNDICE II – PRODUÇÃO BIBLIOGRÁFICA

Totais de produção

Artigos completos publicados em periódico -> 1

Trabalhos publicados em anais de eventos -> 12

Apresentações de trabalhos (Congresso) -> 3

Demais produções bibliográficas ->1

PRINCIPAIS RESUMOS ENVIADOS PARA CONGRESSO

Frequências alélicas de 15 *short tandem repeats* autossômicos em uma amostra populacional do Distrito Federal (Brasil) - um território que surgiu do nada para a realidade. (Trabalho selecionado para apresentação oral no 40º Congresso Brasileiro de Análises Clínicas em 2013).

T. Santa Rita¹, C. Chianca¹, L. Abdalla², S. Costa², C. Sousa², G. Barra².

1- Universidade de Brasília, Brasília, Brazil, 2- Laboratório Sabin, Brasília, Brazil.

Introdução:

O Distrito Federal do Brasil e de Brasília, a atual capital federal, foi fundado em 21 de abril de 1960 a partir do desejo de transferir a capital federal do Rio de Janeiro para o centro do país. Em 2010, a população do Distrito Federal é composta de 2.562.963 habitantes, 54% das pessoas nasceram localmente e os 46% restantes eram imigrantes de outras macrorregiões brasileiras. O teste de paternidade está progressivamente a tornar-se um teste de laboratório clínico. Esta análise é feita com base na existência ou não de alelos *short tandem repeats* (STR) que são compartilhados entre a criança e o suposto pai. Se tiver inclusão de paternidade há duas possibilidades: o suposto pai é o pai biológico ou, devido ao acaso, ele tem um perfil idêntico de STR do verdadeiro pai biológico. A possibilidade de dois indivíduos não aparentados possuírem perfis de STR idênticos pode ser determinada estatisticamente (Random match probability). Neste trabalho, investigamos as frequências alélicas e os parâmetros forenses de 15 loci STR autossômicos na população do Distrito Federal.

Material e Métodos:

Através da análise retrospectiva de nosso banco de dados de resultados do laboratório, recuperamos os perfis de STR de 462 adultos não relacionados (230 mulheres e 232 homens) envolvidos em testes de paternidade realizados entre Maio de 2011 e Outubro de 2012. Todos os perfis foram de indivíduos que residem no

Distrito Federal. As frequências alélicas, desvio de equilíbrio de Hardy-Weinberg, conteúdo de informação polimórfica, poder de discriminação, poder de exclusão, random match probability, índice típico de paternidade, e as distâncias genéticas de Nei para outras populações das macrorregiões brasileiras foram calculados para os 15 STRs incluídos no AmpFISTR® NGM™ (Applied Biosystems). O comitê de ética da Universidade de Brasília aprovou este estudo.

Resultados:

Nenhum desvio significativo do equilíbrio de Hardy-Weinberg foi encontrado. Para todos os STRs, a random match probability combinado foi de $4,85 \times 10^{-20}$, o poder de discriminação combinado foi 0,99999999999999999951, e o poder de exclusão combinado foi 0,9999998. O maior índice de paternidade foi 4,0526 para D2S1338 e FGA e o menor foi de 2,0263 para D22S1045. As distâncias genéticas de Nei mostraram que a população do Distrito Federal foi agrupada em um cluster composto por Sudeste, Nordeste e populações do Brasil. Assim, o Distrito Federal assume uma posição intermediária em relação ao Sul, e é muito distante do Norte e até do Centro-Oeste, onde o território está localizado.

Conclusão:

Os 15 loci STR estudados são um sistema genético altamente informativo para a população do Distrito Federal. Além disso, as distâncias genéticas encontradas foram adequadas para a composição da população de acordo com o Inquérito de Pesquisa Nacional em 2010, em que a principal proporção de imigrantes no Distrito Federal era do Nordeste e Sudeste. As frequências alélicas descritas aqui serão usadas nos casos de rotina.

EDTA-mediated inhibition of nuclease protects cell-free fetal DNA from ex-vivo degradation in blood samples. (Trabalho premiado pela Research in Molecular Pathology Division no congresso *American Association for Clinical Chemistry* em 2013).

J. Vasques¹, T. Santa Rita², C. Chianca², L. Velasco¹, L. Abdalla¹, S. Costa¹, G. Barra¹. Laboratório Sabin, Brasília, Brazil¹; Universidade de Brasília, Brasília, Brazil²

Background:

The cell-free DNA (cfDNA) in bloodstream is becoming an important analyte. Several studies have shown that its measurement is useful for monitoring patients with cancer and other health conditions. Moreover, the fetal cfDNA in the maternal circulation is currently the method of choice for non-invasive determination of fetal genetic traits. However, cfDNA is unprotected and it may be susceptible to known blood high nuclease activity. In this study, we use Y chromosome-specific sequence (DYS-14) levels in maternal serum and EDTA-plasma to investigate the ex-vivo impact of this nuclease activity on cfDNA quantity.

Methods:

This study used banked (-20°C) EDTA-plasma and serum paired samples (n=34) from women with a male fetus (gestational age, 12 ± 4 weeks). Informed consent was obtained from all participants and institutional review board approved the study. Experiments were done with sets of 9-10 paired samples, except the -20°C/4°C/24°C assay (EDTA- plasma and serum sample were not paired). Nucleic acid extraction was performed by using Nuclisens easyMAG (bioMerieux). DYS-14 assays were performed on a StepOne Real-time PCR Systems (Applied Biosystems) by using of hydrolysis probe chemistry and absolute quantification. The results were shown as median (Genomic Equivalents/mL). Statistical analyses were Wilcoxon's test or Friedman's test. For exogenous nuclease assay samples were treated or not with 25U of DNase I (Fermentas) for 1 hour at 37°C before DYS-14 assay. For endogenous nuclease assay a hydrolysis probe and a passive reference (ROX) were added to the crude serum or EDTA-plasma, fluorescence increase were measured for 10 hours at 37°C in the Real-time PCR System. For serum nuclease inhibition assay a serial dilution of EDTA from 0.000005 to 50 mM was used.

Results

DYS-14 quantity was higher in EDTA-plasma than in serum (24.77 versus 18.13 GE/ml, P=0.0137). This difference increased after specimen exposure to 37°C for 24 hours, 22.22 GE/ml for EDTA-plasma versus 5.18 GE/ml for serum, P=0.002. Next, the samples were subject to -20°C, 4°C or 24°C for 24 hours and no difference was observed on DYS-14 concentration in EDTA-plasma, 36.24, 38.02 and 37.31 GE/ml (p=0.328), respectively. In serum, DYS-14 concentration was reduced at 24°C (11.60 GE/ml) compared to -20°C (18.65 GE/ml) and 4°C (17.90 GE/ml), p = 0.0002. Furthermore, DNase-I treatment did not alter the DYS-14 amount in EDTA-plasma (untreated 16.07 versus treated 17.57 GE/ml, p=0.42) but completely eliminates it in serum, (untreated 9.21 versus treated 0.0003 GE/ml, p=0.0039). Additionally, hydrolysis probe degradation could be detected in serum (12.7-fold) but not in EDTA-plasma (0.28-fold). Finally, hydrolysis probe and DYS-14 degradation in serum was inhibited from 0.5 mM of EDTA.

Conclusions

We found that cfDNA is subject to a temperature-triggered degradation in serum but not in EDTA-plasma. Moreover, exogenous and endogenous nucleases were active only in serum. EDTA probably by chelation of divalent ions inhibits blood nucleases conferring an ex-vivo protection to cfDNA. Ultimately, we developed a real-time fluorescence method to detect nuclease activity in clinical samples.

Comparison of four commercial assays for human papillomavirus detection in consecutive clinical laboratory samples including men and women. (Trabalho selecionado para apresentação oral no congresso *American Association for Clinical Chemistry* em 2013).

T. Santa Rita¹, M. Caixeta², C. Chianca¹, L. Velasco², L. Abdalla², S. Costa², G. Barra². ¹*Universidade de Brasília, Brasília, Brazil*, ²*Laboratorio Sabin, Brasília, Brazil*,

Background:

Currently, there are several commercial assays for Human *Papillomavirus* (HPV) detection. However, there are substantial differences between their proposals, such as genotypes detected, methodology, and viral genomic target region that confer unique analytical performances for each one. Here, we compare four commercial HPV assays in consecutive samples from our HPV detection routine that's include 1/3 of men and 2/3 women.

Methods:

The assays were Hybrid Capture (HC2-Qiagen), Papillocheck (Greiner-bio-one), Clart-HPV2 (Biomerieux), and Real-Time High-Risk HPV (Abbott-PCR). Ninety two (61 women and 31 men) consecutive genital samples were included. Requesting physician performed the samples collections in Specimen Transport Medium (Qiagen). Results were shown as HPV positivity (any HPV detected) and High-Risk HPV positivity (at least one HR-HPV detected). Statistical analysis were chi-square test and Cohen' Kappa agreement coefficient. In divergence investigation, fail was characterized by a negative result in a sample classified positive in any other method, except if the genotype was not detected by the method.

Results:

The overall concordance was 78.7% for women and 45.2% for men. HPV positivity was 19.7, 24.6, 27.9 and 27.9% (P=0.69) in women and 29.0, 58.1, 67.7 and 32.3% in men (P=0.0035) for HC2, Papillocheck, Clart-HPV2 and Abbott-PCR, respectively. HR-HPV positivity was 18.0, 19.7, 24.6 and 27.9% (P=0.58) in women and 22.6, 38.7, 48.4 and 32.3% in men (P=0.18) for HC2, Papillocheck, Clart-HPV2 and Abbott-PCR, respectively. The concordance and kappa between each method are shown in Table 1. HC2, Papillocheck, Clart-HPV2 and Abbott-PCR fail in 13.1, 9.8, 8.2 and 0% of the samples in women and in 35.5, 12.9, 3.2, and 12.9% in men, respectively. Not detected genotype account for 8.2% and 19.3% of the divergence in women and men, respectively.

Conclusion:

For women, HC2, Papillocheck, Clart-HPV2 and Abbott-PCR shown similar analytic performances. For men, Papillocheck and Clart-HPV2 shows better analytic performance than HC2 and Abbott-PCR.

Table 1. The concordance (%) and kappa between each HPV assay

	Method 1	Method 2	HPV Detection			HR HPV Detection		
			Agreement (%)	Kappa	Interpretation	Concordance (%)	Kappa	Conclusion
Women	HC2	Papillocheck	91.80	0.763	Good	93.44	0.778	Good
	HC2	Clart-HPV2	91.80	0.776	Good	93.44	0.806	Very Good
	HC2	Abbott-PCR	85.25	0.597	Moderate	90.16	0.726	Good
	Papillocheck	Clart-HPV2	86.89	0.662	Good	90.16	0.709	Good
	Papillocheck	Abbott-PCR	83.61	0.577	Moderate	88.52	0.686	Good
	Clart-HPV2	Abbott-PCR	90.16	0.755	Good	93.44	0.831	Very Good
Men	Method 1	Method 2	Concordance (%)	Kappa	Agreement	Concordance (%)	Kappa	Agreement
	HC2	Papillocheck	70.97	0.456	Moderate	83,87	0.632	Good
	HC2	Clart-HPV2	48.39	0.101	Poor	67,74	0.343	Fair
	HC2	Abbott-PCR	77.42	0.469	Moderate	83,87	0.599	Moderate
	Papillocheck	Clart-HPV2	77.42	0.521	Moderate	83,87	0.668	Good
	Papillocheck	Abbott-PCR	74.19	0.512	Moderate	90,32	0.786	Good
	Clart-HPV2	Abbott-PCR	64.52	0.37	Fair	83,87	0.674	Good

HCV genotype distribution and determination of others variables related to the virus by assessing a clinical laboratory results database. (Congresso American Association for Clinical Chemistry em 2012).

G. B. Barra^{1,2}, C. F. Chianca², T. H. Santa Rita², P.G.G. Costa, W.L.P. Vivas³, P.O.S. Almeida³, L.F.R. Velasco¹, L.F. Abdala¹, S.S.S. Costa¹.

¹Laboratório Sabin de Análises Clínicas, Brasília, Brasil; ²Universidade de Brasília, Brasília, Brasil; ³Universidade Tiradentes, Aracaju, Brasil

Introduction

In 2011, FDA announced a recall of an in vitro nucleic acid amplification test for the quantitation of HCV RNA that has been shown to under-quantitate a subset of genotype 4 in patient specimens by approximately 1.0-1.5 log₁₀ in the absence of any sequence mismatches. The distribution of these test include your country. As the HCV genotype has significant geographic variation, this recall encourage us to determine the prevalence of HCV genotypes and others variables related to the virus by assessing our clinical laboratory results database, and evaluate the possible impact of this under-quantitation in our region.

Methods

Through retrospective analysis of our HCV genotyping database, we assessed the samples results between January 2005 and September 2011. 480 samples were

analyzed, 394 (82%) genotypes were determined and 86 (17.9%) had negative results (no genotype detected). Of these, 280 (58,2%) perform quantitation of HCV RNA in the same day. The genotype prevalences and RNA quantification were identified and presented by gender and age group <50 years (n=162) and ≥50 years (n=118). HCV genotyping was performed by three different techniques in the analyzed period, PCR-RFLP, reverse hibridization, and RT-qPCR, they were also compared.

Results

The observed genotype distribution was HCV-1 288 (73%), HCV-3 91 (23%), HCV-2 13 (3.2%), and HCV-4 2 (0.5%). The genotypes HCV-6 and HCV-5 were not found. The three techniques show related capabilities, since they result in similar genotypic distributions frequency (p=0.98). The average viral load was 6.11 log₁₀ copies/ml and did not differ with genotype (p=0.48), or gender of the patient (p=0.54). By separating the sample into age groups <50 and ≥ 50 years the average viral load was 6.13 and 6.41 log₁₀ copies/ml (P=0.032) respectively.

Conclusion

In this study we noted that the distribution of HCV in our samples is HCV-1, HCV-3, HCV-2 and HCV-4. The genotype 4 is rare in our region, as was found in only two individuals, or 0.5% of samples in six and half years. Thus, the impact of under-quantitation in our region would be minimal. When comparing the different techniques used for molecular genotyping showed that all three techniques (RT-qPCR real-time Reverse Hybridization and PCR-RFLP) have similar capabilities to determine the genotypes. With respect to viral load, no difference was observed between genotypes, or in genotypes according to gender, nor in the genotypes according to age. However, the group aged ≥ 50 had a mean viral load significantly greater than <50.

Paternity inclusion and exclusion in different types of genetic kinship investigations conducted in a clinical laboratory of the Federal District (Brazil).

(Congresso American Association for Clinical Chemistry em 2012).

C. Chianca¹, T. Santa Rita¹, J. Vasques², L. Abdalla², S. Costa², C. Sousa², G. Barra².

¹Universidade de Brasília, Brasilia, Brazil; ²Laboratório Sabin de Análises Clínicas, Brasilia, Brazil.

Background:

The paternity test is progressively becoming a clinical laboratory test. This analysis is based on comparing at least fifteen short tandem repeat (STR) DNA markers between the child and alleged father, in the presence or absence of the biological mother, which classifies the exam in trio or duo, respectively. The incompatibility in three or more STRs characterizes a paternity exclusion. The compatibility between all regions characterizes a paternity inclusion. The description of the types of cases (duo/trio) and results (inclusion/exclusion) in clinical laboratories are scarce. In this

work, these parameters were evaluated retrospectively after eighteen months of implementation of this test in our laboratory.

Methods:

Through the retrospective analysis of our database, we assessed the genetic kinship investigations performed between May 2011 and October 2012. Nine hundred and fourteen investigations were conducted, all involving individuals residing in Brazil's Federal District. The type of case, the conclusion and their distribution over the months were recorded and presented as absolute and/or relative frequency and mean \pm standard deviation, when appropriate. The chi-square test was used to compare the obtained ratios. University of Brasília ethical committee approved this study.

Results:

Out of a total of 914 paternity cases, the trios occurred in a higher prevalence compared to duos, 596 (65.2%) versus 318 (34.8%). Moreover, the inclusions were more prevalent than exclusions, 610 (66.7%) versus 304 (33.3%). Besides that, the proportion of inclusion/exclusion were similar between trios and duos, 201 (33.72%) exclusions for trios and 103 (32.39%) exclusions for duos ($p=0.68$). Furthermore, the inclusion/exclusion and trio/duo proportions remained homogeneous in the eighteen-month studied period, $34.78 \pm 6.78\%$ for exclusions ($p = 0.45$) and $34.72 \pm 5.26\%$ for duos ($p = 0.97$).

Conclusion:

In paternity testing, the trios were more frequent than duo. This result can be explained by the fact that trio has lower complexity analysis than duo, because the alleles not transmitted from mother are determined with precision, making it a test cheaper than duo. Furthermore, the paternity inclusion is the most prevalent result type and the inclusion/exclusion proportion observed in this study is similar to that reported by forensic laboratories (32%). Moreover, the homogeneity in inclusion/exclusion and trio/duo proportions suggest that these parameters associated with paternity testing remain similar over time.

Distribution of genotypes/species of hpv in penile lesions biopsies (Congresso American Association for Clinical Chemistry em 2012).

Gustavo Barcelos Barra¹, Ticiane Henriques Santa Rita², Maria Cecilia Sant' Anna Soares Borges Caixeta¹, Camilla Figueiredo Chianca², Wanessa Lordêlo Pedreira Vivas², Patrícia de Oliveira Santos Almeida², Lara Franciele Ribeiro Velasco¹, Lidia Freira Abdlla¹, Sandra Santana Soares Costa¹.

Laboratório Sabin de Análises Clínicas¹, Brasília-DF; Universidade Tiradentes², Aracaju-SE

Introduction

Investigations on the human papillomavirus (HPV) are mostly focused on women. However, HPV infection in men and its consequences are not completely

understood. The limited studies available suggest that the virus prevalence in men is high, and that they are the spreader of the infection. Thus, the aim of this study was to describe the HPV infection in penile lesions by comparing the results of biopsies with viral genotyping.

Methods

Through retrospective analysis of our laboratory results database, we assessed patients who had undergone biopsies and HPV genotyping in penile injury on the same day, between February 2009 and May 2011. Forty-three samples met this condition. Next, we evaluated the prevalence of HPV genotypes/species, and concordance between the two methodologies. In addition, we correlated the different cytological abnormalities evidenced by biopsy (exophytic lesion, hyperkeratosis, parakeratosis, irregular acanthosis, papillomatosis, hypergranulosis, koilocytotic atypia, keratinization and hyperplasia and loss of basal polarity) with the presence of the virus, the type of infection (simple and multiple) and its oncogenicity (high and low risk). The average age of patients was 30.3 ± 7.9 years, and genotyping was performed by PapilloCheck (Greiner Bio-One), which evaluates 24 different HPV types (18 high risk and 6 low risk).

Results

Genotyping demonstrated that 72.1% of penile lesions were positive for HPV. Among these, 38.7% were single infections and 61.3% multiple infections. Of 68 found viruses, 57.3% were low risk, 42.6% high risk, and the five most prevalent genotypes were 6 (25%), 11 (13.2%), 42 (10.3%), 39 (5.9%) and 44/55 (5.9%). When we grouped the genotypes in species, A10 was the most prevalent (44.1%), consisting of low-risk genotypes, followed by A7 (14.7%) consisting of high-risk genotypes. However the biopsy demonstrated that 65.1% of the penile lesions showed cellular changes suggestive of viral infection. Thus, the agreement among the two methods was 69.6% ($P = 0.078$). The cellular changes associated with HPV were hyperkeratosis ($P = 0.003$), irregular acanthosis ($P = 0.003$) and papillomatosis ($P = 0.037$). The Koilocitotic atypia was close to significance ($P = 0.09$). In addition, we observed that parakeratosis ($P = 0.046$), koilocitotic atypia ($P < 0.0001$), hyperplasia ($P = 0.024$) and loss of basal polarity ($P = 0.024$) are more prevalent in multiple infections than in simple. However, we found no association between virus oncogenicity and cytological alterations.

Conclusion

In penile lesions for HPV positivity is high, multiple infections surpass the simple, and the genotypes/species of low risk are the most prevalent. The correlation between biopsy and genotyping is not complete, suggesting the presence of HPV types not covered by the genotyping test run, and the presence of the virus in the form of subclinical lesions. On the other hand, the association among the multiple infections and some types of cellular changes indicate the need for the presence of more than one type of viruses for its appearance in men. Thus, it is noted that important information about HPV infection in men could be extracted by analyzing a database of a clinical laboratory.

HPV infection pattern and genotype distribution in cervical cytology and in penile lesion biopsies. (Congresso American Association for Clinical Chemistry em 2012).

Gustavo Barcelos Barra¹, Ticiane Henriques Santa Rita², Maria Cecília Sant' Anna Soares Borges Caixeta¹, Camilla Figueiredo Chianca², Wanessa Lordêlo Pedreira Vivas², Patrícia de Oliveira Santos Almeida², Lara Franciele Ribeiro Velasco¹, Lidia Freira Abdlla¹, Sandra Santana Soares Costa¹.

Laboratório Sabin de Análises Clínicas¹, Brasília-DF
Universidade Tiradentes², Aracaju-SE

Background

Investigations on the human HPV are mostly focused on women. However, HPV infection in men and its consequences are not completely understood. The aim of this study was to describe the HPV infection pattern and genotypes distribution in cervical cytology and in penile lesions biopsies by assessing our clinical laboratory results database.

Methods

We assessed patients who had undergone cervical cytology/penile lesion biopsies and HPV genotyping on the same day, between February 2009 and May 2011. 1731 women and 43 men samples met this condition. Positivity, type of infection, oncogenicity, and genotypes were identified and presented. In addition, we correlated these parameters with different cytological abnormalities evidenced by cytology (Bethesda system), or by biopsy. Genotyping was performed by PapilloCheck (GreinerBio-One).

Results

Positivity were 38,37% and 72.1% for cervical cytology and penile lesions, respectively. Multiple infections account for 41,5% in cervical cytology, and 61.3% in penile lesions. Of 1311 found viruses in cervical cytology, 73,7% were high risk, and of 68 found viruses in penile lesion 42.6% were high risk. In cervical cytology, positivity ($P < 0,0001$) and high risk genotypes ($P = 0,0087$) increased with the degree of cell injury. Moreover, the seven most common genotypes in normal cervical cytology were 16/56/44/53/6/39/42, in ASC-US 16/52/56/31/53/39/73, in LSIL 16/56/66/68/53/31/51, in HSIL 16/58/73/52/56/39/68, and in penile biopsies, 6/11/42/59/44/31/51. Additionally, hyperkeratosis ($P = 0.003$), irregular acanthosis ($P = 0.003$), and papillomatosis ($P = 0.037$) were associated with HPV in penile biopsies. Parakeratosis ($P = 0.046$), koilocitotic atypia ($P < 0.0001$), hyperplasia ($P = 0.024$) and loss of basal polarity ($P = 0.024$) were more prevalent in multiple infections than in simple.

Conclusion

There are differences in positivity, multiple infections, and high risk genotypes prevalence between cervical cytology and penile lesion. Also, proportion of positivity and high risk genotypes increase with degree of cell injury in cervical cytology. Some cellular/hystological changes are significantly associated with HPV and multiple infections in penile lesions.

ANEXO I – APROVAÇÃO PELO COMITÊ DE ÉTICA EM PESQUISA



Universidade de Brasília
Faculdade de Ciências da Saúde
Comitê de Ética em Pesquisa – CEP/FS

PROCESSO DE ANÁLISE DE PROJETO DE PESQUISA

Registro do Projeto no CEP: **165/12**

Título do Projeto: “Análise populacional de 22 loci microssatélites no Distrito Federal”.

Pesquisadora Responsável: Gustavo Barcelos Barra

Data de Entrada: 17/10/12

Com base na Resolução 196/96, do CNS/MS, que regulamenta a ética em pesquisa com seres humanos, o Comitê de Ética em Pesquisa com Seres Humanos da Faculdade de Ciências da Saúde da Universidade de Brasília, após análise dos aspectos éticos e do contexto técnico-científico, resolveu **APROVAR** o projeto **165/12** com o título: “Análise populacional de 22 loci microssatélites no Distrito Federal”, analisado na 11ª Reunião Ordinária, realizada no dia 11 de dezembro de 2012.

O pesquisador responsável fica, desde já, notificado da obrigatoriedade da apresentação de um relatório semestral e relatório final sucinto e objetivo sobre o desenvolvimento do Projeto, no prazo de 1 (um) ano a contar da presente data (item VII.13 da Resolução 196/96).

Brasília, 13 de dezembro de 2012.

Thiago Rocha da Cunha
Vice-Coordenador do CEP-FS/UnB