



PONTÍFICA UNIVERSIDADE CATÓLICA DE GOIÁS
PRÓ-REITORIA DE PÓS-GRADUAÇÃO E PESQUISA
PROGRAMA DE MESTRADO EM GENÉTICA
NÚCLEO DE PESQUISA REPLICON

Jonas Garcia de Almeida

ANÁLISES COMPARATIVAS DAS PROBABILIDADES DE PATERNIDADE OBTIDAS
A PARTIR DE DIFERENTES BANCOS DE DADOS POPULACIONAIS DO BRASIL

Goiânia-Goiás
2015

Jonas Garcia de Almeida

ANÁLISES COMPARATIVAS DAS PROBABILIDADES DE PATERNIDADE OBTIDAS
A PARTIR DE DIFERENTES BANCOS DE DADOS POPULACIONAIS DO BRASIL

Dissertação Apresentada ao Programa de Pós-Graduação Stricto Sensu em Genética da Pontifícia Universidade Católica de Goiás para Obtenção do Título de Mestrado em Genética.

Área de Concentração: Genética

Orientador: Prof. Dr. Aparecido Divino da Cruz

Co-orientadora: Prof^a Dr^a Thaís Cidália Vieira

Goiânia-Goiás
2015

Dados Internacionais de Catalogação da Publicação (CIP)
(Sistema de Bibliotecas PUC Goiás)

A447a Almeida, Jonas Garcia de.
Análises comparativas das probabilidades de paternidade obtidas a partir de diferentes bancos de dados populacionais do Brasil [manuscrito] / Jonas Garcia de Almeida – Goiânia, 2015. xiv, 81 p. : il. ; 30 cm.

Dissertação (mestrado) – Pontifícia Universidade Católica de Goiás, Programa de Pós-Graduação *Stricto Sensu* em Genética.

“Orientador: Prof. Dr. Aparecido Divino da Cruz”.
Bibliografia.

1. Testes de paternidade. 2. Genética. I. Título.

CDU 575(043)

Ata de Aprovação



PONTIFÍCIA UNIVERSIDADE CATÓLICA DE GOIÁS
PRÓ-REITORIA DE PÓS-GRADUAÇÃO E PESQUISA

Av. Universitária, 1069 • Setor Universitário
Caixa Postal 86 • CEP 74605-010
Goiânia • Goiás • Brasil
Fone: (52) 3946.1070 • Fax: (62) 3946.1070
www.pucgoias.edu.br • prope@pucgoias.edu.br

ATA COMPLEMENTAR Nº 102/2015

MESTRADO EM GENÉTICA DA PONTIFÍCIA UNIVERSIDADE CATÓLICA DE GOIÁS

DISCENTE: JONAS GARCIA DE ALMEIDA

DEFENDIDA EM 26 DE MARÇO DE 2015 E Aprovado COM CONCEITO... A

O título foi alterado (não () sim) _____

BANCA EXAMINADORA

fovere
Prof. Dr. Aparecido Divino da Cruz
(presidente-orientador)

[Assinatura]
Profa. Dra. Thais Cidália Vieira Gigonzac / UEG
Co-Orientadora

Lysa Bernardes Minasi
Profa. Dra. Lysa Bernardes Minasi / PUC Goiás
(membro interno)

Jaqueline Gleice A. Freitas
Profa. Dra. Jaqueline Gleice Aparecida Freitas / UEG
(membro externo)

Dedicatória

Dedico esse Trabalho ...

Aos meus queridos pais e irmãos pelo apoio alicerçados na doutrina familiar que me deram forças todos os dias para que eu possa alcançar o melhor em minha vida.

Agradecimentos

A minha família em especial meu pai Joaquim Geraldo de Almeida e Helena Garcia de Almeida, mesmo não tendo formação escolar sempre enfatizaram a mim e meus irmãos a importância e a carga que os estudos possuem.

Ao Prof. Dr. Aparecido Divino da Cruz, "Peixoto", por acreditar em minha formação de geneticista, por ter participado na concretização do sonho. Agradeço pela atenção, amizade, cumplicidade, pela lealdade, competência e principalmente por me nortear na vida profissional e pessoal compartilhando sua experiência e ensinamento paterno.

À Profa Dra. Thais Cidália Vieira, por ter sido a providência que Deus colocou no meio da minha jornada de desenvolvimento. Agradeço por ter se disposta a enfrentar o desafio. Muito obrigado pela atenção, disponibilidade e principalmente pelo exemplo profissional.

À Profa Dra. Lysa Bernardes Minasi, por ter sido mais uma providência que Deus colocou no final de minha jornada de formação a mestre, aceitando em ser participante da banca. Agradeço por ter se disposta em passar o crivo rigoroso em minha dissertação.

Aos Colegas do Replicon, em especial, Cristiano Luiz Ribeiro e Raimundo da Silva Lima pelo auxílio na elaboração dos dados estatísticos desta pesquisa, e incentivos intelectuais prestados.

Aos professores do Programa de Mestrado e Pós-graduação em Genética da Universidade Católica de Goiás, por compartilhar seus conhecimentos e pelos incentivos.

A minha querida noiva, Naiane Iara de França Pinheiro por ter me apoiado e me amparado nas horas de aflição, e por estar sempre ao meu lado.

Epígrafe

“Gostaria de dizer para você que viva como quem sabe que vai morrer um dia, e que morra como quem soube viver direito. Amizade só faz sentido se traz o céu para mais perto da gente, e se inaugura aqui mesmo o seu começo. Mas, se eu morrer antes de você, acho que não vou estranhar o céu. Ser seu amigo, já é um pedaço dele.”

Chico Xavier

RESUMO

Para fins de identificação humana na área forense os Estados Unidos criou o Banco de Dados (BD) do *Sistema de Índice de DNA Combinado* (CODIS), contendo 13 *loci*: CSF1PO, FGA, TH01, TPOX, VWA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51 e D21S11. Foi realizado no presente estudo, a comparação dos dados obtidos das frequências alélicas (FA), índices de paternidade (IP) por marcador e probabilidades de paternidade (PPs) obtidas a partir de 9 BD populacionais, sendo 1 nacional (BR), utilizando 241 casos contendo a $PP \geq 99,99$ com o uso de 13 *loci* STRs (CODIS) e mais 2 marcadores STRs do Penta-D e E, cedidos pelos BDs do Núcleo de Pesquisa Replicon, da Pontifícia Universidade Católica de Goiás (PUC/GO) e do LaGene, Laboratório Central de Saúde Pública (SES/GO). Os 241 casos foram analisados nas situações de Trio (suposto pai, filho e genitora) e Duo (suposto pai e filho) usando 13 marcadores STRs (CODIS) através do teorema de *Bayes* e pela Razão de Verossimilhança (RV). Os resultados foram submetidos aos testes de Kruskal-Wallis, Qui-quadrado, Correlação de *Spearman* e Exato de Fisher usando o *software* SPSS (2010). As análises estáticas da RV dos casos contendo o Trio resultaram em 1.148 resultados com $PPs \geq 99,99\%$ e 21 resultados com $PPs < 99,99\%$, e em situação de Duo 1.686 resultados com $PPs \geq 99,99\%$ e 483 resultados com $PPs < 99,99\%$. As análises das médias dos IPs de cada marcador demonstraram os marcadores STRs do D21S11 e FGA com os maiores poder de inclusão e TH01 juntamente com o D3S1358 os menores poderes de inclusão. As PPs não apresentaram diferenças estatísticas significativas contendo em sua maioria correlações positivas de moderadas a fortes entre os 8 BDs comparados ao BDs populacional do BR. Este estudo demonstrou as interferências estatísticas que cada BDs de frequências alélicas pode exercer nos cálculos de PP quando se utiliza apenas 13 *loci* para confirmação de vínculo genético, tornando-se assim mais significativas nos casos em situação de Duo. De acordo as informações disponíveis nos BDs de frequências alélicas das diferentes regiões geográficas do Brasil, tornou-se possível concluir, que as frequências alélicas obtidas, bem como os IPs por marcador e as PPs obtidas, sugerem fortes similaridades às encontradas no banco de dados nacional.

Palavras Chaves: *loci*, Bandos de dados populacionais, Frequências alélicas, Índice de Paternidade, Probabilidade Paternidade.

ABSTRACT

For human identification purposes in Forence area, the United States created the database (DB) of *DNA Index System Combined (CODIS)*, containing 13 loci: CSF1PO, FGA, TH01, TPOX, VWA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51 and D21S11. Then the DBs of the markers of the Y-STR X-STRs haplotypes were created. Comparison of the obtained data was conducted from allele frequencies (AF), paternity index (PI) by marker and paternity probabilities (PPs) obtained from nine BD population, one of those national (BR), using 241 cases containing the $PP \geq 99.99$ using 13 loci STRs (CODIS) and 2 more STRs markers of Penta D- and E, granted by the BDs of Núcleo de Pesquisas Replicon, of the Pontifícia Universidade Católica de Goiás (PUC / GO) and LaGene, Laboratório Central de Saúde Pública (SES / GO). 241 cases were analyzed in Trio (alleged father, son and progenitor) and Duo (alleged father and son) situations, using 13 STRs markers (CODIS) by Bayes' theorem and the Likelihood Ratio (LR). The results were submitted to the Kruskal-Wallis, chi-square, Spearman correlation and Fisher's exact tests and SPSS software (2010). Static analysis of RV cases containing the Trio resulted in 1,148 results $PPs \geq 99.99\%$ and 21 results $PPs < 99.99\%$, and in a position to Duo 1686 results with $PPs \geq 99.99\%$ and 483 results $PPs < 99.99\%$. The analysis of the IP average of each marker showed the D21S11 of STRs markers and FGA with the highest power of inclusion and TH01 and D3S1358 with the lowest power of inclusion. The PPs did not show significantly different, containing mostly positive correlation, of moderate to strong, between 8 BDs compared to the BR population databases. This study demonstrated the statistics interference that each allele frequencies DB can have in PP calculations using only 13 loci of genetic, thus making it more significant in cases Duo situation. According to the information available in databases of gene frequencies of the different geographical regions of Brazil, it became possible to conclude that the allele frequencies obtained and the IPs per marker and PPs obtained, suggest strong similarities to those found in the national database.

Keywords: *loci*, population data base, allelic frequencies, Paternity Index, Paternity Probability

SUMÁRIO

RESUMO	viii
ABSTRACT	ix
FIGURAS	xi
TABELAS	xii
QUADROS	xiii
ANEXOS	xiv
LISTA DE ABREVIATURAS.....	xv
1. INTRODUÇÃO	16
1.1 A formação da população brasileira.....	16
1.2 O uso do DNA na identificação humana	18
1.3 Marcadores STRs (do inglês, <i>Short Tandem Repeats</i>).....	19
1.4 Nomenclatura dos <i>Loci</i> STRs	21
1.5 Bancos de frequências alélicas.....	23
1.6 Cálculos realizados na comprovação de vínculo genético.....	24
1.7 Mutações em STRs	28
2. OBJETIVOS	30
2.1 Objetivo Geral.....	30
2.2 Objetivos Específicos.....	30
3. METODOLOGIA	31
3.1 Grupo amostral.....	31
3.2 Obtenção e seleção dos Bancos de dados (BDs).....	31
3.3 Análises estatísticas.....	33
4. RESULTADOS	34
4.1 Análises comparativas das frequências alélicas dos bancos de dados	34
4.3 Probabilidades de Paternidade obtidas nos casos em situações de Trio e Duo	43
5. DISCUSSÃO	51
5.1 O potencial informativo das frequências alélicas encontradas nos 9 BDs populacionais, sendo 1 nacional (BR) usados na pesquisa.....	51
5.2 As PP obtidas com o uso dos 13 marcadores STRs (CODIS) nos casos em situação de Trio e Duo.....	53
5.3 Avaliação dos dados obtidos pelo nos testes estatísticos Correlação de <i>Spearman</i> e Exato de Fisher comparando as frequências alélicas, IPS e PPs.....	56
6. CONCLUSÃO.....	59
7. REFERÊNCIAS BIBLIOGRÁFICAS	60
ANEXOS	66

FIGURAS

Figura 1. Pintura de Modesto Brocos retratando a miscigenação brasileira em um modelo empírico de branqueamento- Rio de Janeiro, RJ 1936 (Cã, 1895).....	17
Figura 2. Representação da nomenclatura do locus D16S539 (adaptado Butler 2010).	22
Figura 3. Mapa do Brasil, contendo regiões geográficas de onde foram selecionados os nove bancos de frequências alélicas contendo os 13 loci STRs do CODIS selecionados para o presente estudo (adaptado de IBGE, 2015).....	33
Figura 4. Distribuição das frequências alélicas do marcador D18S51.....	34
Figura 5. Distribuição das frequências alélicas do marcador D21S11.....	35
Figura 6. Distribuição das frequências alélicas do marcador D3S1358.....	35
Figura 7. Distribuição das frequências alélicas do marcador TH01.....	36
Figura 8. Distribuição das frequências alélicas do marcador CSF1PO.....	36
Figura 9. Distribuição das frequências alélicas do marcador D13S317.....	37
Figura 10. Distribuição das frequências alélicas do marcador D16S539.....	38
Figura 11. Distribuição das frequências alélicas do marcador D5S818.....	38
Figura 12. Distribuição das frequências alélicas do marcador D7S820.....	39
Figura 13. Distribuição das frequências alélicas do marcador D8S1179.....	40
Figura 14. Distribuição das frequências alélicas do marcador FGA.....	40
Figura 15. Distribuição das frequências alélicas do marcador TPOX.....	41
Figura 16. Distribuição das frequências alélicas do marcador vWA.....	41
Figura 17. Relação entre o número de inclusões obtidas nas Análises da RV em 13 marcadores STRs (CODIS).	44
Figura 18. Relação entre o número de exclusões obtidas nas Análises da RV em 13 marcadores STRs (CODIS), ($p < 0,001$).	44
Figura 19. Distribuição espacial das PPs obtidas a partir de 241 indivíduos analisados entre os BDs das regiões avaliadas.	50

TABELAS

Tabela 1. Informação dos 13 STRs CODIS.	24
Tabela 2. Fórmulas utilizadas nos cálculos de confirmação de paternidade em casos de Trio (SP, F e Mãe biológica).	26
Tabela 3. Fórmulas usadas nos cálculos de confirmação de paternidade em casos de Duo (SP e SF).	26
Tabela 4. Representação das frequências das Mutações e Média do Poder de Exclusão dos locus STRs (AABB, 2004).	29
Tabela 5. Bancos de dados de frequências alélicas obtidos pela pesquisa.	32
Tabela 6. Bancos de frequências alélicas selecionados com 13 <i>loci</i> genéticos recomendados pela AABB 2010.	32
Tabela 7. Número de casos com suas frequências alélicas não apresentadas nos Bancos de dados utilizados na pesquisa.	42
Tabela 8. Número de probabilidades de paternidade obtidas nas análises realizadas em casos em situação de Trio (suposto pai, suposto filho e mãe biológica).	43
Tabela 9. Número de probabilidades de paternidade obtidas nas análises realizadas em casos em situação de Duo (Suposto pai e Suposto filho).	43
Tabela 10. Probabilidades Paternas em um caso de mutação no marcador FGA, nas situações de Trio e Duo.	45

QUADROS

Quadro 1. Análise de correlação de <i>Spearman</i> entre os IPs de cada marcador STR entre as regiões dos BDs populacionais, comparados com o banco de dados de frequências alélicas nacional (BR).....	47
Quadro 2. Análise de correlação de <i>Spearman</i> entre as probabilidades de paternidade (PP) dos marcadores STR e as regiões dos bancos de dados populacionais, comparados com o banco nacional (BR),	49

ANEXOS

Tabela 1. Frequências alélicas do <i>locus</i> D18S51.	66
Tabela 2. Frequências alélicas do <i>locus</i> D21S11,	67
Tabela 3. Frequências alélicas do <i>locus</i> D3S1358.	68
Tabela 4. Frequências alélicas do <i>locus</i> TH01.	69
Tabela 5. Frequências alélicas do <i>locus</i> CSF1PO	70
Tabela 6. Frequências alélicas do loci D13S317	71
Tabela 7. Frequências alélicas do loci D16S359,	72
Tabela 8. Frequências alélicas do <i>locus</i> D5S818	73
Tabela 9. Frequências alélicas do <i>locus</i> D7S820	74
Tabela 10. Frequências alélicas do <i>locus</i> D8S1179	75
Tabela 11. Frequências alélicas do <i>locus</i> FGA.....	76
Tabela 12. Frequências alélicas do <i>locus</i> TPOX	77
Tabela 13. Frequências alélicas do <i>locus</i> vWA.....	78
Tabela 14. Frequências alélicas dos 13 marcadores STRs (CODIS) em 102.807 indivíduos da população do Brasil	79

LISTA DE ABREVIATURAS

- STR:** Repetição curta em Tandem
- PCR:** Reação em Cadeia da Polimerase
- SNP:** polimorfismo de nucleotídeo simples
- DNA:** Ácido Desoxirribonucleico
- CODIS:** Sistema Combinado de índice de DNA
- BDS:** Bancos de Dados
- RV:** Razão de Verossimilhança
- IPC:** Índice de Paternidade Combinado
- PP:** Probabilidade Paterna
- IP:** Índice de Paternidade
- SP:** Suposto Pai
- F:** Filho
- M:** Mãe
- PE:** Poder de Exclusão
- AABB:** Associação Americana de Banco de Sangue
- APO:** Alelo Paterno Obrigatório
- FA:** Frequência Alélica
- CAPs:** Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
- BR:** Brasil
- GO:** Goiás
- BA:** Bahia
- SC:** Santa Catarina
- SP:** São Paulo
- PA:** Pará
- MS:** Mato Grosso do Sul
- RJ:** Rio de Janeiro
-
-

1. INTRODUÇÃO

1.1 A formação da população brasileira

O Brasil desde a sua descoberta no ano de 1.500 até 1.800, foi o berço migratório intenso de várias civilizações. Uma das primeiras migrações foi iniciada com cerca de 500.000 portugueses e 4 milhões de Africanos, povoando os Estados da Bahia, Espírito Santo, Rio de Janeiro, Pernambuco, São Paulo, Santa Catarina e Rio Grande do Sul. (Cordeiro, 2009).

A região Sul do Brasil caracteriza-se pela elevada taxa de miscigenação, estabelecida pelo cruzamento de imigrantes europeus, escravos africanos, e ameríndios durante os últimos cinco séculos de ocupação territorial do Brasil (IBGE, 2010).

A população do Sudeste brasileiro, além de possuir origem Européia e Oriental, também é bastante miscigenada devido às migrações dos povos vindos de outras regiões brasileiras, tais como Nordeste, Sul e Sudeste durante o período de industrialização regional observado nos séculos XVII e XVIII (IBGE, 2010).

A região Centro-Oeste do Brasil tem uma peculiar caracterização fenotípica, fruto da miscigenação de populações de diversas origens, das quais incluem latino-americanos, Japoneses, Portugueses, Espanhóis, Árabes, além de famílias migrantes de outros Estados brasileiros (IBGE, 2010).

A região Nordeste principalmente as populações dos Estados do Piauí, Rio Grande do Norte e Pernambuco podem ser caracterizadas como uma população exogâmica, composta por indivíduos de ascendências de nativos Americanos, Africanos e Europeus (Silva, 2004).

Uma evidência histórica sobre os primeiros estudos empíricos sobre miscigenação humana foi apresentada no Congresso Universal das Raças, realizado em Londres no início do século XX, com o estudo publicação e demonstração pelo médico brasileiro João Batista Lacerda, o qual apresentou a obra de arte de Modesto Brocos y Gómez de 1895, como exemplo de miscigenação da população brasileira. João Batista demonstrou uma evidência científica com sua tese de branqueamento (Figura 1). Ele defendeu a tese que brancos e negros não eram espécies distintas, e sim raças cuja miscigenação produzia uma prole fecunda. Ainda segundo a tese do médico, a imigração e a ausência de preconceito racial

cooperariam para extinguir os mestiços, pois em um século a população brasileira seria majoritariamente branca, e os negros e índios provavelmente desapareceriam (seyferth, 2011).



Figura 1. Pintura de Modesto Brocos retratando a miscigenação brasileira em um modelo empírico de branqueamento- Rio de Janeiro, RJ 1936 (Cã, 1895).

Hoje com os estudos avançados em genética de população, hipóteses históricas podem ser constatadas e transformadas em dados mais precisos em relação à etnia e miscigenação de uma população. Uma demonstração desses dados foi apresentada pelo estudo realizado por Pena (2011), o qual contou com a participação voluntária de 934 indivíduos de diferentes regiões brasileiras para um estudo de ancestralidade. Pena (2011) usou um painel de 40 polimorfismos de DNA de inserção-deleção para ancestralidade. O painel foi validado para ascendência em Europeus, Africanos e Ameríndios. O estudo demonstrou que no Brasil, a ascendência europeia foi predominante em todo o território nacional. Porém, as diferenças populacionais no padrão de ocupação regional resultaram em uma proporção Africana maior no Nordeste (30,3%), seguido em ordem decrescente pelo Sudeste (18,9%), Sul (12,7%) e Norte (10,9%). Por outro lado, a proporção Ameríndia foi maior na população do Norte (19,4%), que também se apresentou relativamente uniforme nas três outras regiões.

1.2 O uso do DNA na identificação humana

O DNA tem auxiliado geneticistas e especialistas forenses na confirmação e identidade de suspeitos e vítimas em assassinatos, como também casos de abuso sexual e nos estudos de vínculo genético. A determinação da identidade genética pelo exame em DNA pode ser considerada um dos produtos mais revolucionários da genética molecular humana moderna (Pena, 2011).

O surgimento da identificação humana iniciou em 1900 com a descoberta do sistema ABO para realização da tipagem sanguínea por Karl Landsteiner, este sistema é encontrado em todas as células do corpo humano. Esta descoberta foi aplicada na identificação humana em 1910, por Von Dungern e Hirsfeld, ficando conhecido como herança mendeliana do grupo sanguíneo sendo o primeiro marcador sorológico utilizado nas pesquisas e perícias forense. Porém este sistema apresentou uma variabilidade muito limitada, pois classificam as pessoas em apenas quatro tipos (grupos A, B, AB, O). O próximo evento significativo ocorreu em 1927, com o sistema MN descrito por Landsteiner e Levine, seguido pela descoberta do sistema Rh, em 1940, por Levine (Butler, 2011; Monte, 2004).

Marcando a segunda fase da evolução dos testes de identificação humana, em 1954, foi apresentado o sistema de histocompatibilidade, com a descoberta de antígenos nas superfícies de leucócitos, conhecido como complexo Antígenos Leucocitários Humanos (HLA). Composto por um conjunto de genes que codificam moléculas de histocompatibilidade o complexo encontra-se localizado no braço curto do cromossomo 6, onde estão agrupados três sub-regiões: HLA classe I (A, B e C), HLA classe II (DR, DQ e DP) e HLA classe III. Devido seu alto poder discriminatório forneceu aos cientistas um meio prático para identificar exclusões de pelo menos 95% dos supostos pais investigados. O uso de metodologia para identificar HLA foi muito utilizada para complementar as metodologias anteriores, dando mais eficácia ao processo de identificação de indivíduos. No entanto estas metodologias, não foram totalmente adequadas para este fim, devido ao seu alto custo operacional (Jeffreys, 2005; Alves, 2005).

Em 1953 com a descoberta do DNA como molécula do patrimônio genético dos seres vivos, Watson e Crick revolucionaram a Biologia molecular provocando mudanças científicas e tecnológicas capazes de identificar particularidades no DNA de cada indivíduo (Buckleton, 2005). Posteriormente, Alec Jeffreys em 1985 marcou a era das pesquisas

forenses de identificação humana, com a descoberta de regiões de DNA que continham sequências repetitivas consecutivas com tamanhos variando entre 7 a 100 bases nitrogenadas com números de repetições de 100 vezes ou mais. Estas regiões contendo sequências repetitivas no DNA com tamanhos variados poderiam diferir de indivíduo para indivíduo (Jeffreys 1985). Estes polimorfismos ficaram conhecidos posteriormente como VNTRs (do inglês: *Variable Number Tandem Repeats*), ou regiões minissatélites. Devido a sua variabilidade na população, comparados aos sistemas antes utilizados, a Inglaterra começou a empregar os VNTRs em seus estudos, como também foi utilizada para resolver um caso de imigração e, logo em seguida, de estupro e homicídio (Butler, 2012).

Nos Estados Unidos da América os VNTRs passaram a ser utilizados a partir de 1986 (Butler 2005). Alec Jeffreys nesse mesmo período utilizou a técnica do RFLP (do inglês, *Restriction fragment length polymorphism*) com emprego de enzimas de restrição para separar regiões específicas dos VNTRs, a metodologia gerou resultados conhecidos como “impressões digitais” (Jeffreys 1985, Butler, 2012).

Em 1983, Kary Mullis desenvolveu a técnica de PCR (Polymerase Chain Reaction). Essa técnica permite que pequenas quantidades de DNA sejam amplificadas exponencialmente. Assim, os fragmentos curtos de DNA, como os STRs (do inglês, *Short Tandem Repeat*) podem ser utilizados na técnica (Souza, 2010). Hoje essa metodologia usa a enzima – *Taq DNA polimerase* – que sintetiza uma sequência complementar do DNA selecionado, desde que um pequeno fragmento (o iniciador ou *primer* marcado com fluorescência) já esteja pareado em uma das cadeias do DNA no ponto escolhido para o início da síntese (Figura 5) (Rumjanek, 2001).

1.3 Marcadores STRs (do inglês, *Short Tandem Repeats*)

A seleção dos *loci* STR na identificação humana e em estudos populacionais está relacionada ao grau de polimorfismo de cada *locus*, e tal medida é feita com base na ocorrência de um maior ou menor número de alelos heterozigotos para um dado *locus* em uma determinada população. Para a avaliação do grau de polimorfismo, é necessária a análise de parâmetros estatísticos objetivando a determinação do padrão de diversidade genética destes marcadores na população (Aquino, 2011).

Os *loci* STRs apresentam sequências curtas de repetições de pares de bases (Jeffreys, 2005 & Ferreira; Grattapalha, 1995). A diversidade nos *loci* STRs está no número de repetições com sequência variando de 1 a 5 de bases nitrogenadas. Devido a essas repetições cada indivíduo possui um tamanho diferente na região do seu DNA (Jeffreys, 2005 & Ferreira e Grattapalha, 1995). Essas regiões são analisadas e encontradas unicamente, podendo assim identificar o perfil genético de um indivíduo (Carracedo, 2007).

Os STRs são classificados como simples, quando se apresentam em repetições com apenas unidades idênticas quanto à sequência e comprimento (Ex. GAGTn); compostos, quando apresentam repetições contendo 2 ou mais repetições simples adjacentes (GAGTnTATAn); e complexos, quando apresentam repetições com blocos de unidades repetidas espaçados por sequências intervenientes que variam mais ou menos em comprimento (Ex. GAGTnAAGAGTnTAGAGTn) (Urquhart, 1994).

Considera-se polimórfico o *loci* existente em uma população que possui pelo menos dois alelos, em que um deles contenha sua frequência igual ou superior a 1% na população. Os polimorfismos correspondem a variações ao nível de sequência e ao nível do tamanho da molécula de DNA (Butler, 2012).

Uma das características importantes dos marcadores é o seu elevado polimorfismo. Esses polimorfismos podem estar no número de repetições, no tamanho da unidade de repetição ou ainda na composição da unidade de repetição, sendo que esta última é dividida em repetições simples, compostas, complexas e complexas hipervariáveis. As repetições simples contêm unidades de repetição com tamanho idêntico que se repetem ao longo de toda a sequência, as repetições compostas compreendem 2 ou mais unidades de repetição que se repetem ao longo da sequência, as repetições complexas contêm vários blocos de unidades de repetição variáveis com tamanho variável, e por fim as repetições complexas hipervariáveis contêm vários alelos não consensuais que diferem em tamanho e sequência. Nem todos os alelos contêm unidades de repetição completas, podendo ter somente 1, 2 ou 3 nucleotídeos no final da repetição, pelo que adquirem o nome de microvariantes (Butler, 2012).

Para validar o uso dos marcadores microssatélites em casos de paternidade, o pesquisador se atenta nas seguintes observações: Eles devem possuir herança mendeliana, taxas de mutação baixa, o estudo da população testada deverá apresentar frequências alélicas consistentes e de acordo com o princípio de Hardy-Weinberg para que possam ser incorporadas nos cálculos das frequências genotípicas e de probabilidade paterna, verificar se

os *locus* não estão ligados, para que a regra do produto possa ser utilizado nos cálculos estatísticos e que o poder de exclusão seja suficientemente apresentado (Hammond, 1994).

Apesar dos *loci* microssatélites serem eficientes para distinção genética individual, a maioria oferece pouco poder para permitir a distinção de etnia com informações precisar sobre as características de um indivíduo (Shiver, 1997). Isto se deve ao fato que a maior parte (80-90%) da variação genética humana e intra-populacional, sendo que somente 10-20% da variação total e formada com a variação entre populações (Rosenberg, 2002).

Apesar da limitação em se obter distinções étnicas, Bamshad (2003) e Pedrosa (2006), utilizaram *loci* STRs para estimar o grau de mistura populacional e avaliar a ancestralidade genômica ao nível individual. Com os resultados do estudo os pesquisadores concluíram que os melhores marcadores para este tipo de investigação seriam aqueles que apresentassem os maiores diferenciais de frequências alélicas entre os grupos definidos. Pois marcadores com estas características são muito uteis em estudos envolvendo genética humana, mapeamento genético, e genética forense. Assim, para que esses marcadores sejam utilizados de maneira correta o pesquisador deverá realizar estudos populacionais prévios, avaliando as diferenças existentes entre frequências alélicas de cada população.

1.4 Nomenclatura dos *Loci* STRs

Para facilitar a comunicação entre a comunidade Forense por todo o mundo existem regras de nomenclatura dos marcadores genéticos. Essas regras foram recomendadas em 1997, pela Comissão Internacional de DNA da Sociedade Internacional de Genética Forense (ISFG) (Butler, 2012).

A primeira regra estabelece que a leitura das sequências de DNA seja feita da extremidade 5' → 3'. A segunda determina que quando um *locus* STR está relacionado com genes codificantes de proteínas, essa será a cadeia utilizada como referência e o *locus* terá a abreviatura da proteína que codifica, como é o exemplo do *locus* vWA (fator de von Willebrand), enquanto que em outros *loci* o critério é o da sequência descrita primariamente na literatura, em que o número seguinte à letra D que significa DNA é o número do cromossomo em que se localiza o STR e os números seguintes à letra S que refere-se ao *loci* de DNA com sequência de cópia única que em seguida são numerados com identificadores,

como é exemplo o marcadores D12S324 que se situa no cromossoma 12 sendo o 324º *locus* descrito neste cromossomo (Figura 8).

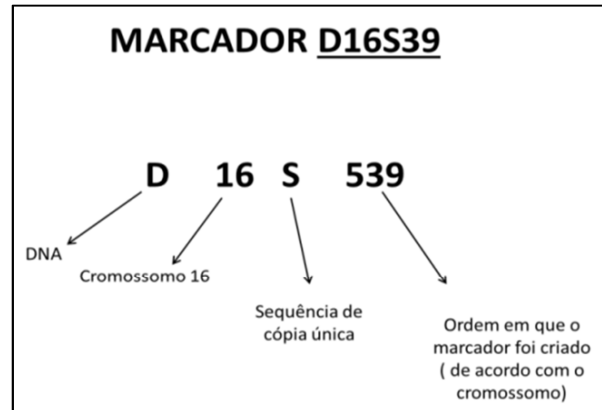


Figura 2. Representação da nomenclatura do locus D16S539 (adaptado Butler 2010).

A terceira regra diz respeito à sequência de repetição, que está no enumera de vezes que se repete uma determinada sequência no qual se numera o alelo em questão, por exemplo, na sequência: 5'- GG TCA TCA TCA TCA TCA GG-3' o nome do alelo será 5. Já nas unidades de repetição que se encontram incompletas, como é o caso do alelo 9.3 do *locus* TH01, significa que existem 9 cópias da unidade de repetição tetranucleotídica e mais três nucleotídeos, estas microvariantes serão então classificadas com o número de repetições completas seguidas de um ponto decimal e do número de nucleotídeos existentes de seguida (Butler, 2012). Como também existem dois alelos para cada *locus* representado em cada indivíduo do sexo feminino, com exceção dos cromossomos sexuais, no sexo masculino apresentam apenas um alelo para cada *locus* do cromossomo X e do Y (Otto, 2004).

Para além das regras de nomenclatura referidas, os marcadores devem, ainda, cumprir determinados requisitos, tal como possui elevado poder discriminatório, possuem valores de heterozigose superiores a 70%, facilidade de reprodutibilidade da técnica, estar localizado em cromossomos distintos ou distantes um do outro quando presentes no mesmo cromossomo, conter baixas taxas de mutações, promover produtos de PCR com tamanho reduzido e nível de artefatos ou bandas inespecíficas baixas (Butler, 2012).

1.5 Bancos de frequências alélicas

De acordo com as Diretrizes para Operações de Identificação de DNA de Fatalidade em Massa, de 2010, o laboratório responsável pelas identificações dos indivíduos deverá selecionar um Banco de Dados (BD) de frequência alélica adequado à população em que pertence o indivíduo testado. Um BD incorreto poderá distorcer o compartilhamento dos alelos, pois alelos de uma população podem ser raros, mas em outra população não. Os BD de frequências alélicas de muitos grupos étnicos são publicados em revistas e estão disponíveis para uso público. A criação, validação e uso de um banco de frequência alélica de uma nova população, é criado quando o foco está em uma pequena ou única população que não é bem representada por bancos de dados já publicados.

Nos Estados Unidos, em 1998, houve a criação do mais conhecido e conceituado BD do CODIS, que combina a área da biologia molecular com a tecnologia eletrônica, formado a partir de perfis genéticos de DNA extraído de evidências biológicas coletados em cenas de crimes (Silva, 2006). Atualmente o CODIS é constituído por 13 loci: CSF1PO, FGA, TH01, TPOX, VWA, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51 e D21S11 (Butler, 2010) (Tabela 1).

A distribuição da frequência alélica de qualquer *locus* de DNA é variável entre populações distintas. Por isso é de extrema importância estabelecer as frequências populacionais identificando-se os alelos mais comuns e mais informativos para as investigações forenses de identificação humana para aquela população (Butler, 2010).

Para que a identificação humana mediante o uso de STRs seja incorporada com efetiva confiabilidade, é crucial a realização de estudos populacionais abrangentes que contemple toda a diversidade representativa de uma população. A criação desses bancos de dados e suas frequências devem ser divulgadas na literatura científica (Egeland, 2006).

Tabela 1. Informação dos 13 STRs CODIS.

Marcador STR	Localização	Sequência repetitiva	Número de repetição
D18S51	18q21.3	[GAAA]	08-10, 10.2, 11-13, 13.2, 14-27
D21S11	21q21.1	[TCTA], [TCTG]	24, 24.2, 25, 25.2, 26-28, 28.2, 29, 29.2, 30, 30.2, 31, 31.2, 32, 32.2, 33, 33.2, 34, 34.2, 35, 35.2, 36-38
D3S1358	3p21.31	[AGAT], [TCTA]	02-20
TH01	11p15.5	[AATG]	04-09, 9.3, 10, 11, 13.3
CSF1PO	5q33.1	[AGAT]	06-15
D13S317	13q31.1	[GATA]	07-15
D16S539	16q24.1	[GATA]	05, 08-15
D5S818	5q23.2	[AGAT]	07-16
D7S820	7q21.11	[GATA]	06-14
D8S1179	8q.24.13	[TATC]	07-18
FGA	4q28	[TTTC] ₃ TTTTTCT [CTTT] _n CTCC [TTCC] ₂	16-18, 18.2, 19, 19.2, 20, 20.2, 21, 21.2, 22, 22.2, 23, 23.2, 24, 24.2, 25, 25.2, 26-30, 31.2, 43.2, 44.2, 45.2, 46.2
TPOX	2p25.3	[AATG]	06-13
vWA	12p13.31	[AGAT]	10-22
AMELOGENINA	Xp22.1.22.3 e Y	NA*	X, Y

Fonte: http://www.cstl.nist.gov/strbase/str_fact.htm e Power Plex[®] 16 System, Manual Técnico 2013 (adaptado).
* Não se aplica; **X**: Cromossomos X; **Y**: Cromossomo Y; **p**: braço curto do cromossomo; **q**: braço longo do cromossomo.

1.6 Cálculos realizados na comprovação de vínculo genético

O estudo das probabilidades de parentesco implica o conhecimento populacional das características alélicas e das suas frequências relativas, presumindo a existência de equilíbrio de Hardy Weinberg na população de referência e a transmissão genética mendeliana com o sistema codominante. Este estudo tem de ser realizado antes da sua aplicação na rotina dos casos (Drabel, 2009)

O conjunto de alelos de todos os loci constitui o perfil alélico de um indivíduo para realizar o teste de vínculo genético, esse perfil alélico (investigado) é usado para identificar as possíveis coincidências existentes no perfil alélico de outro indivíduo (investigante). As coincidências alélicas entre os perfis dos envolvidos permitem que o pesquisador estabeleça inferências a partir de cálculos estatísticos. Os resultados combinados das estimativas estatísticas permitem concluir probabilisticamente o vínculo genético existente entre os indivíduos envolvidos no teste (Guerra, 2007).

A probabilidade combinada de um perfil alélico em estudo é estimada a partir das frequências alélicas de bancos de dados da população da qual fazem parte os investigadores e investigados. É crucial que as frequências dos alelos dos marcadores genéticos usados na investigação seja representativo na população e contemple toda a diversidade individual e de

suas etnias fundadoras. Neste contexto quanto mais raro um perfil alélico, mais forte é a evidência de que as duas amostras de DNA usadas em comparação correspondem ao genoma de uma mesma pessoa (Reno, 2000; Butler, 2010).

As frequências alélicas depositadas em um determinado BD são usadas para realizar as análises de Razão de Verossimilhança (RV), formado com a probabilidade e a hipótese. O grau da RV é medido pela frequência da incidência representada pelo número de vezes em que determinado perfil genético ocorre na população, como, por exemplo, 1 em 5 bilhões ou 1 em 20 bilhões de pessoas (Jobim, 1999).

Assim os resultados das análises obtidas nos testes de vínculos genéticos são realizados estatisticamente a partir de cálculos baseados nas frequências alélicas da população. Os índices empregados nos cálculos estatísticos são: Índices de Paternidade (IP), Índice de Paternidade Combinado (IPC), Probabilidade de Paternidade (PP) e Probabilidade de Exclusão (PE) (Primorac, 2000).

O IP é obtido através da razão existente entre a probabilidade de que um homem investigado como suposto pai seja o pai biológico do filho investigante e a probabilidade de que outro homem, ao acaso, possa ser o verdadeiro pai biológico (Jobim, 2012). Calcula-se o IP de acordo com a frequência do APO (Alelos Paternos Obrigatórios) presentes no filho investigante. Em seguida um IP é multiplicado pelo outro consecutivamente, chegando-se ao valor dos produtos, chegando-se ao IPC. Quanto maior o valor do IPC, maior será a probabilidade do suposto pai ser o pai biológico do filho investigante (Jobim, 2012).

O IP de cada *locus* representa a razão entre a probabilidade do suposto pai e a probabilidade de qualquer outro indivíduo da população transmitir o APO para o filho investigante. A probabilidade de qualquer outro indivíduo transmitir o alelo paterno obrigatório corresponde à frequência do alelo na população avaliada (Butler, 2005).

De acordo com Jacewicz (2004), para usar as frequências alélicas geradas a partir da genotipagem de uma determinada população o pesquisador deverá levar em conta se a origem e a etnia dos envolvidos analisados.

De acordo com o teorema de *Bayes*, cada IP é calculado de acordo com a conformação alélica apresentada pelo conjunto em situação de Trio (composto pelo suposto pai, filho investigante e mãe biológica ou SP,F,M), ou Duo (composto pelo suposto pai e filho ou SP,F), aplicando-se assim uma fórmula para cada conjunto alélico apresentado (Tabelas 2 e 3) (Tillmar, 2010).

Tabela 2. Fórmulas utilizadas nos cálculos de confirmação de paternidade em casos de Trio (SP, F e Mãe biológica).

SF ¹	Mãe	SP ²	Numerador	Denominador	IP ³
A _i A _i	A _i A _i	A _i A _i	1	p _i	1/p _i
		A _i A _j ou j ≠ i	1/2	p _i	1/2p _i
		A _i A _k ou j ≠ i, j	0	p _i	0
A _i A _i	A _i A _j ou i ≠ j	A _i A _i	1/2	p _i /2	1/p _i
		A _i A _j ou j ≠ i	1/4	p _i /2	1/2p _i
		A _i A _k ou k ≠ i, j	0	p _i /2	0
A _i A _j ou i ≠ j	A _i A _i	A _j A _j	1	p _j	1/p _j
		A _j A _k ou k ≠ j	1/2	p _j	1/2p _j
		A _k A _i ou k, l ≠ j	0	p _i	0
A _i A _j ou i ≠ j	A _i A _j ou i ≠ j	A _i A _i	1/2	(p _i + p _j)/2	1/(p _i + p _j)
		A _i A _j	1/2	(p _i + p _j)/2	1/(p _i + p _j)
		A _j A _k , k ≠ j	1/4	(p _i + p _j)/2	1/2((p _i + p _j))
		A _k A _l , k, l ≠ j	0	(p _i + p _j)/2	0
A _i A _j ou i ≠ j	A _i A _k	A _j A _j	1/2	p _j /2	1/p _j
		A _j A _i ou i ≠ j	1/4	p _j /2	1/2p _j
		A _k A _l ou k, l ≠ j	0	p _j /2	0

Fonte: Adaptado de *Annual Report Summary for testing in DNA* of 2004 (AABB, 2004).

Legenda: ¹ - Combinação alélica do Suposto Filho. ² - Combinação alélica do Suposto Pai. ³ - Fórmula para calcular o índice de paternidade. p_{i,j,k,l} - Frequência do alelo.

Tabela 3. Fórmulas usadas nos cálculos de confirmação de paternidade em casos de Duo (SP e SF).

F ¹	SP ²	Numerador	Denominador	IP ³
A _i A _i	A _i A _i	p _i ³	p _i ⁴	1/p _i
	A _i A _j ou j ≠ i	p _i ² p _j	2p _i ³ p _j	1/2p _i
A _i A _j , i ≠ j	A _i A _i	p _i ² p _j	2p _i ³ p _j	1/2p _i
	A _i A _j ou j ≠ i	p _i p _j (p _i ² p _j)	(2p _i p _j) ²	(p _i +p _j)/(4p _i p _j)
	A _i A _k ou k ≠ j	p _i p _j p _k	4p _i ² p _j p _k	1/4p _i

Fonte: Adaptado de *Annual Report Summary for testing in DNA* of 2004 (AABB, 2004).

Legenda: ¹ - Combinação alélica do Suposto Filho. ² - Combinação alélica do Suposto Pai. ³ - Fórmula para calcular o índice de paternidade. p_{i,j,k} - Frequência do alelo.

A PP é uma medida que define o quanto as evidências indicam que a pessoa testada pode ser o pai biológico do filho investigante. A PP baseia-se no Teorema de *Bayes*, um método que determina uma probabilidade *a posteriori* a partir da genotipagem da mãe, do investigante e do suposto pai. Para estimar a PP, deve-se assumir uma probabilidade *a priori* de que o suposto pai testado seja o verdadeiro pai biológico. Na maioria dos países, a probabilidade *a priori* empregada em todas as análises estatísticas é de 1/2 (Primorac, 2000). O IPC é o produto dos índices de paternidade individuais para cada *locus* investigado, ele determina o poder da evidência genética do suposto pai ser o pai biológico do filho investigante. Assim, quanto maior o valor do IPC, maior será a evidência genética (Butler 2005). Assim é usado para obter a PP *a posteriori* de 1/2 a seguinte fórmula: Prob = IP/(IP+1)

x 100%. Esta probabilidade posteriori chamada de “PP”. Por exemplo, em uma análise onde o $IP = 10.000$, calculando a PP: $10.000 / (10.000 + 1) = 0,99$, multiplicando por 100, chega-se PP de 99,9900%.

O PE corresponde à probabilidade de se excluir qualquer indivíduo que tenha sido falsamente acusado em teste de paternidade. Isto equivale à frequência de todos os homens, em uma determinada população, que não possuem os alelos coincidentes ao APO da criança em um teste de paternidade (Butler, 2005).

Para calcular o PE usa-se a seguinte fórmula: $PE = h^2(1-2hH^2)$, onde h significa a frequência observada de heterozigotos e H é a frequência observada de homozigotos. Os valores individuais de cada locus podem ser agrupados no poder de exclusão combinado pelo seguinte fórmula: $1 - [(1 - PE1) \times (1 - PE2) \times (1 - PE3) \times \dots \times (1 - PEN)]$ (Tillmar, 2010).

Internacionalmente é recomendado e consensuado no mínimo 13 marcadores STRs para a investigação de vínculo genético entre pessoas derivadas de uma mesma população. Porém, quanto maior o número de marcadores analisados, menor a incerteza sobre a inferência estatística e maior a probabilidade de acerto. Isto é, a probabilidade a posteriori se aproxima da verdade biológica dos dados em análise. De acordo com a Sociedade Brasileira de Medicina Legal, a exclusão de paternidade é declarada quando o suposto pai apresenta exclusão em no mínimo de três Alelos Paterno Obrigatório (APO) no perfil alélico do investigante (Silva, 2006 & Dolinsky, 2007).

Nas investigações de parentesco, os casos mais comuns são os de investigação de paternidade simples. Nestes casos participam da investigação o suposto pai, o investigante e mãe caracterizando um Trio investigatório. Em alguns casos participam somente o suposto pai e o investigante caracterizando um Duo investigatório. Os cálculos estatísticos para a determinação de um vínculo genético são cruciais e de extrema importância para conclusão de exame de vínculo genético (Dawid, 2002).

Seguindo os parâmetros internacionais de controle de qualidade e recomendações de órgãos especializados em perícias de vínculo genético, como por exemplo, o Grupo de Línguas Espanholas e Portuguesa da Sociedade Internacional de Genética Forense e a *Associação Americana de Banco de Sangue (AABB)*, para que um laudo de investigação de paternidade possa ser considerado conclusivo, os resultados obtidos mediante a análise dos *loci* dos cromossomos autossômicos devem fornecer um IPC igual ou superior a 10.000. Isto indica que o indivíduo testado tem dez mil vezes mais chance de ser o pai biológico em

relação a qualquer outro indivíduo da mesma população. Nesta participativa, a Probabilidade de Paternidade (PP) é de 99,99% (AABB, 2008).

1.7 Mutações em STRs

Mutações podem ser encontradas tanto em células reprodutoras quanto em células somáticas. Pesquisadores suspeitam que as mutações nos STR resultem no deslizamento de cadeias de DNA durante o processo de replicação (*replication slippage*) ou de uma deficiência na capacidade de reparação da replicação. Esse evento mutacional pode ser na forma de uma única mudança de base ou em toda a repetição do STR. Atualmente, estudos sugerem que a taxa de mutação é proporcional ao grau de polimorfismo (número de alelos) de cada *locus* STR, sendo maior nos alelos mais polimórficos (Butler, 2010).

Quando, por ocasião da investigação individual usando DNA, se observa uma mutação em um dado locus, o IP é calculado de acordo com a taxa de mutação do locus no qual foi encontrada a incompatibilidade (Brenner, 2004). As taxas de mutação em marcadores STRs variam entre 10^{-3} e 10^{-4} . Em geral, a inclusão da taxa de mutação no cálculo diminui-se o valor do IP. No entanto, o índice de paternidade combinado, permanece elevado e com a PP esperada de dos 99,99% (Brinkmann, 2001 & Di Lonardo, 2004). Quando a PP se encontra abaixo dos 99,99% devido à inserção do cálculo estatístico de uma mutação encontrada em um dos marcadores avaliados recomenda-se a inserção de mais marcador no cálculo para se alcançar uma PP confiável (Penacino, 2003). A estrutura repetitiva do *locus*, número de cernes mutacionais e o tipo de mutação dos STRs devem ser observados antes de se realizar a análise estatística em um teste de vínculo genético (Brinkmann, 2001).

A AABB (2004) recomenda que nos casos de inconsistências únicas, não deve ser ignorado, devendo-se realizar cálculos incluindo a taxa de mutação para o *locus* mutado. O índice de paternidade de um *locus* com uma inconsistência é pequeno: maior do que zero, mas substancialmente menor do que um. Isso faz com que a mutação leve o IPC para baixo, diminuindo a probabilidade de paternidade para aquele conjunto de *loci*.

Nos casos em que se encontra mutação em um dos *locus* dos participantes do teste de vínculo genético, o IP é calculado de acordo com a fórmula μ/\hat{A} proposta por Fimmers R, 1992. Onde, μ : frequência de mutação do marcador e \hat{A} : média do poder de exclusão daquele do marcador, utilizando os dados da Tabela 4. Tendo como exemplo o marcador vWA, no

qual, a mãe apresenta os alelos 18/19 o filho investigante 16/19 e o suposto pai os alelos 15/17, resultando do IP será = $0,00170/0,6196 = 0,0027$.

Tabela 4. Representação das frequências das Mutações e Média do Poder de Exclusão dos locus STRs (AABB, 2004).

Marcador	Frequência de Mutações	Média do poder de Exclusão
CSF1PO	0,00153	0,5043
TPOX	0,00012	0,3716
THO1	0,00009	0,5457
vWA	0,00170	0,6196
D16S539	0,00109	0,5396
D7S820	0,00116	0,5950
D13S317	0,00142	0,5286
D5S818	0,00116	0,4618
FGA	0,00319	0,7281
D8S1179	0,00159	0,6305
D18S51	0,00320	0,7400
D21S11	0,00147	0,6825
D3S1358	0,00128	0,5850

De acordo com Panacino (2003), taxas de mutações moderadas não interferem significativamente no uso de marcadores STRs nas análises de comprovação de paternidade, desde que as taxas de mutação sejam conhecidas e possam ser incorporadas em razão da probabilidade estatística do índice de paternidade confirmada. No entanto, o uso dos 15 marcadores STR apontam os resultados dos exames de vínculo genético de populações constituindo, portanto, os bancos de dados (BD) genéticos baseados em valores das frequências alélicas que podem ser utilizados em testes de vínculo genético, assim como em estudos de identificação humana, devido às semelhanças verificadas com outras populações do Brasil e do mundo.

A comparação das frequências alélicas entre os BDs podem permitir avaliar a probabilidade de vínculo genético entre indivíduos, com um valor mais fidedigno, respeitando as variações mutacionais de uma determinada população. Porventura, a utilização de frequências alélicas de bancos de dados de outras regiões pode não ser representativas para a população de uma determinada cidade, ou mesmo da região em que ela pertença.

2. OBJETIVOS

2.1 Objetivo Geral

Comparar os dados obtidos das frequências alélicas (FA), índices de paternidades individuais (IP) e as probabilidades de paternidade (PPs) obtidas a partir de diferentes bancos de dados (BD) populacionais, utilizando as frequências alélicas de nove diferentes Estados do Brasil.

2.2 Objetivos Específicos

1- Analisar o potencial informativo das frequências alélicas dos 9 BDs populacionais, sendo 1 BDs nacional (BR);

2- Analisar as estimativas da RV para inferir as PP usando 13 marcadores STRs (CODIS) em situação de Trio e Duo (sem o perfil alélico da genitora), usando as frequências alélicas dos 9 BDs populacionais, sendo 1 nacional (BR) a partir de 241 casos de vínculo genético de inclusão, em situação de Trio e Duo;

3- Analisar as possíveis divergências existentes entre os resultados obtidos com as análises usando os 9 BDs populacionais, comparando as FA, IPS, PPs e os números de inclusão e exclusão dos perfis alélicos individuais;

4- Propor um novo banco de dados de frequências alélicas nacional com a união de nove bancos de dados regionais de frequências alélicas representados nas cinco regiões brasileiras;

3. METODOLOGIA

3.1 Grupo amostral

Para realização do estudo proposto, foram utilizados 241 casos de vínculo genético em situação de Trio, realizados entre indivíduos da população de Goiás com 13 marcadores STRs (CODIS) e mais dois marcadores STRs Pentas (Penta-E, Penta-D) contendo as PPs \geq 99,99%. Os perfis alélicos dos 241 casos foram cedidos pelos BDs do Núcleo de Pesquisa Replicon, da Pontifícia Universidade Católica de Goiás (PUC/GO) e do LaGene, Laboratório Central de Saúde Pública Dr. Giovanni Cysneiros, da Secretaria de Saúde do Estado de Goiás (SES/GO).

3.2 Obtenção e seleção dos Bancos de dados (BDs)

Para realizar o levantamento dos BDs considerou-se previamente os bancos publicados em periódicos com endosso da AABB, Scielo, CAPS e PubMed, sendo alguns estudos obtidos mediante solicitação direta aos autores. Com as pesquisas feitas pela internet, com as palavras chave STRs paternity foram obtidos 272.000 artigos relacionados teste de paternidade utilizando STRs. A pesquisa foi afinada acrescentando a palavra autossomal, na qual foram obtidos 126.00 resultados e logo em seguida acrescentadas as palavras CODIS, resultando assim em 842 resultados, desses 166 estavam relacionados a pesquisa contendo o assunto objetivado neste trabalho, porém somente 15 artigos (Tabela 05) continhas os dados necessários para a pesquisa, dos quais 09 artigos serviram para compor os dados estatísticos deste estudo (Tabela 06) contendo assim eles os 13 STRs. Após a seleção dos artigos contendo as publicações utilizando as populações das 05 regiões brasileiras, sendo 06 artigos obtidos por solicitações feitas aos autores. Após a obtenção BDs, os trabalhos cujas análises realizadas contendo 13 marcadores STRs (CODIS) foram utilizadas na pesquisa.

Tabela 5. Bancos de dados de frequências alélicas obtidos pela pesquisa.

BDs populacional	Número de indivíduos	Número de marcadores STRs (CODIS)	Autor
BR	100.000	13	Aguiar et al. 2011
AL	598	07	Silvia LAF, 2002
AM	307	12	Francez, 2011
SP	55	13	Martins, 2006
BA	150	13	Santos, 2004
DF	462	08	Rita, 2013
GO	986	13	Vieira, 2013
MS	247	13	Figueiredo, 2009
PA	325	13	Rodrigues, 2005
PE	546	11	Dellalibera, 2004
RJ	300	13	Góes, 2004
RS	316	13	Leite, 2006a
SC	452	13	Moret, 2009
MA	177	00	Ferreira, 2004
SE	259	08	Pimentel, 2004

Foram selecionados 9 BDs populacionais, sendo 1 BDs Nacional (BR) (Tabela 6), todos contendo os 13 marcadores STRs a serem utilizados na pesquisa. Os BDs populacionais obtidos encontram-se distribuídos nas cinco regiões brasileiras (Figura 3).

Tabela 6. Bancos de frequências alélicas selecionados com 13 *loci* genéticos recomendados pela AABB 2010.

BDs populacional	Número de indivíduos	Número de marcadores STRs (CODIS)	Autor
BR	100.000	13	Aguiar <i>et al.</i> 2011
SP	55	13	Martins, 2006
BA	150	13	Santos, 2004
GO	986	13	Vieira, 2013
MS	247	13	Figueiredo, 2009
PA	325	13	Rodrigues, 2006
RJ	300	13	Góes, 2004
RS	316	13	Leite, 2006
SC	452	13	Moret, 2009



Figura 3. Mapa do Brasil, contendo regiões geográficas de onde foram selecionados os nove bancos de frequências alélicas contendo os 13 loci STRs do CODIS selecionados para o presente estudo (adaptado de IBGE, 2015).

3.3 Análises estatísticas

Os dados obtidos foram tabulados em planilhas do *software Microsoft Office Excel*[®] Professional Plus 2010, bem como a confecção de tabelas, quadros e figuras. Os IPs e PPs, dos 241 casos de vínculo genéticos em situação de Duo e Trio contendo os 13 marcadores STRs utilizados pelo CODIS foram obtidos através do teorema de *Bayes* e pela Razão de Verossimilhança (RV).

Os testes de Kruskal-Wallis, Qui-quadrado, Correlação de *Spearman* e Exato de Fisher foram utilizados para comparar as frequências alélicas, IPs, PPs e os números de inclusão e exclusão dos perfis alélicos individuais foram realizados com o auxílio do *software SPSS* (2010). Foi considerado um nível de significância de 0,05, com o intervalo de confiança de 95%. A efetivação deste trabalho foi fundamentada com a observância das Normas de Pesquisa de Saúde. Por tratar-se de um estudo baseado em um levantamento de dados secundários, não foi necessária a aplicação de um Termo de Consentimento Livre e Esclarecido (TCLE).

4. RESULTADOS

4.1 Análises comparativas das frequências alélicas dos bancos de dados

O presente estudo comparou as frequências alélicas de 13 *loci* STRs (CODIS) em 9 BDs populacionais (contendo oito BDs populacionais regionais diferentes e um BD populacional nacional (BR) totalizando 102.806 indivíduos analisados, a fim de verificar a distribuição das frequências alélicas e identificar a variabilidade alélica de cada marcador.

As FAs do marcador D18S51 (figura 1), apresentou uma distribuição alélica geral homogênea entre os BDs populacionais, pelo teste de Kruskal-Wallis, indicando não haver diferença estatisticamente significativa comparando todos os BDs ($p=0,09$). Porém, a análise individual demonstrou haver diferença estatisticamente significativa confrontando as FAs dos BDs de SP ($p<0,05$) e MS ($p<0,05$) com o BD nacional. Os alelos 10, 13, 16, 20, 22 e 25, foram considerados os mais frequentes ($f \geq 0,10$). Adicionalmente, pode-se inferir que o marcador D18S51 possui alelos de elevada diversidade populacional, pelo número de alelos (n: 21 alelos), e uma homogeneidade na distribuição das frequências alélicas entre os BDs.

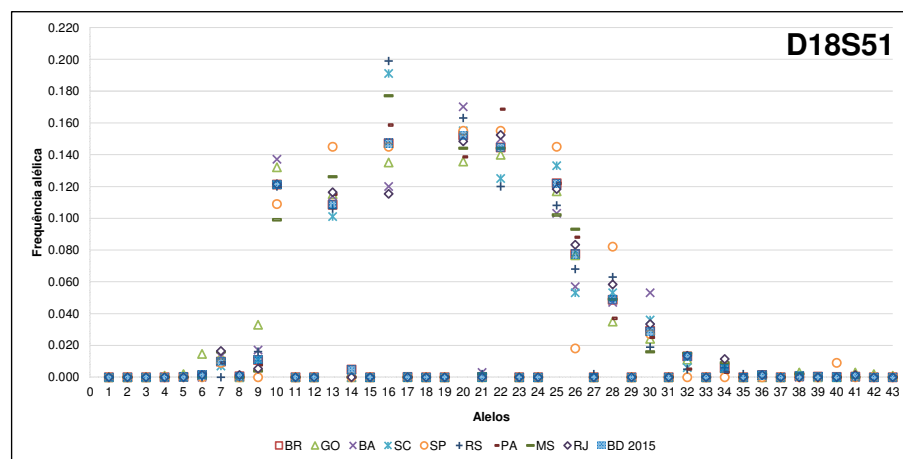


Figura 4. Distribuição das frequências alélicas do marcador D18S51.

No marcador D21S11 (figura 5), apresentou uma homogeneidade nas FAs entre os BDs populacionais ($p= 0,99$), segundo o teste de Kruskal-Wallis, indicando que houve diferença estatisticamente significativa comparando todos os BDs. Também não houve diferença estatisticamente significativa ($p>0,05$) quando comparadas as FAs dos BDs das oito

regiões do país com o BD nacional. Dentre os trinta alelos presentes neste marcador (alta diversidade populacional), os alelos 28, 29 e 30 foram considerados os mais frequentes ($f \geq 0,12$).

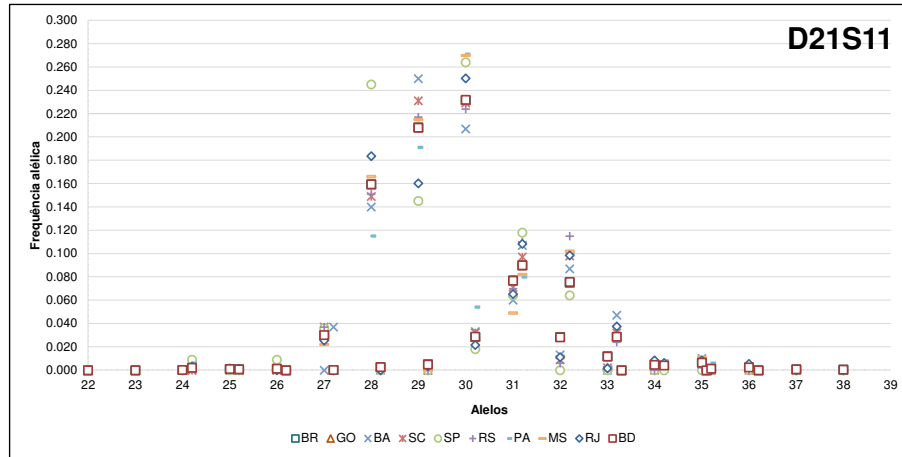


Figura 5. Distribuição das frequências alélicas do marcador D21S11.

Pode-se observar que no marcador STR do D3S1358 (figura 6) possui uma distribuição alélica homogênea entre os BDs populacionais ($p=1,00$). Contendo 13 alelos representativos para a população brasileira, sendo os alelos 15, 16 e 17 considerados os mais frequentes, por apresentarem suas frequências acima de 0,16. Adicionalmente, pode-se inferir que o marcador D3S1358 apresentou em sua maioria, uma variação das FAs para os alelos mais frequentes.

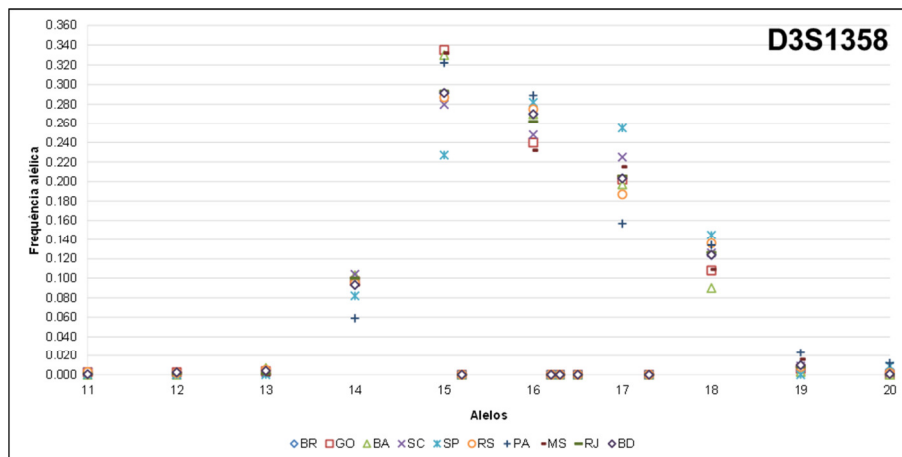


Figura 6. Distribuição das frequências alélicas do marcador D3S1358.

O marcador TH01 (figura 7), apresentou uma distribuição alélica e de FAs, homogênea entre os BDs populacionais ($p = 0,99$). Contendo 8 alelos representativos, sendo os alelos 6, 7, 8 e 9.3 considerados os mais frequentes ($f \geq 0,12$), o alelo 7 obteve a frequência com maior variação entre BDs populacionais de SP (0,150) e BA (0,183) com a diferença entre as frequências de 0,033. Adicionalmente, pode-se inferir que o marcador TH01, possui aproximadamente 30% dos *locus* com frequências discriminativas.

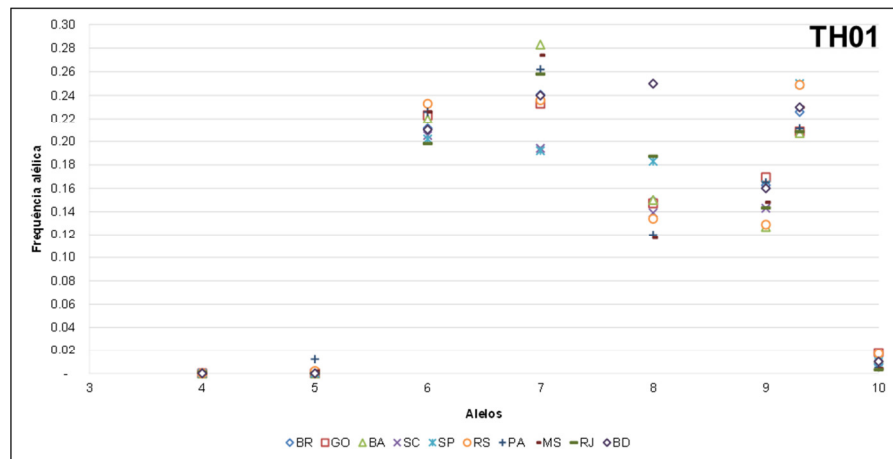


Figura 7. Distribuição das frequências alélicas do marcador TH01.

O marcador STR do CSF1PO (figura 8) apresentou uma distribuição alélica homogênea entre os BDs populacionais. Contendo 17 alelos, os alelos 10, 11 e 12 apresentaram-se como os mais frequentes, representando 17,64% de todos alelos, sendo o alelo 12 o mais variável entre os BDs populacionais de SP (0,327), e o menos frequente o BD da BA (0,277), com uma diferença entre as frequências de 0,05.

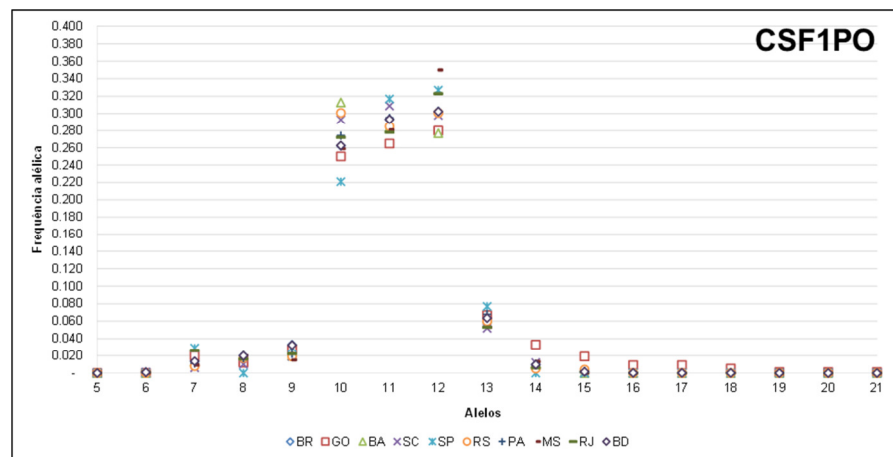


Figura 8. Distribuição das frequências alélicas do marcador CSF1PO.

O marcador STR do D13S317 (figura 9), apresentou uma distribuição alélica homogênea entre os BDs populacionais ($p=0,98$, segundo o teste de Kruskal-Wallis. Com 13 alelos representados, este marcador apresentou 11 alelos com frequências variadas. Apenas 15% dos alelos, os alelos 8 e 11, apresentaram-se como os mais frequentes, sendo o alelo 11 o mais variável entre os BDs populacionais do PA (0,222) e BA (0,336), com a diferença entre as frequências de 0,114. Por apresentar essas informações pode-se considerar que o marcador STR do D13S317 possui um alelo bastante variante entre as populações do PA e BA. Adicionalmente, pode-se inferir que o marcador D13S317, possui em sua maioria *locus* com frequências discriminativas por região.

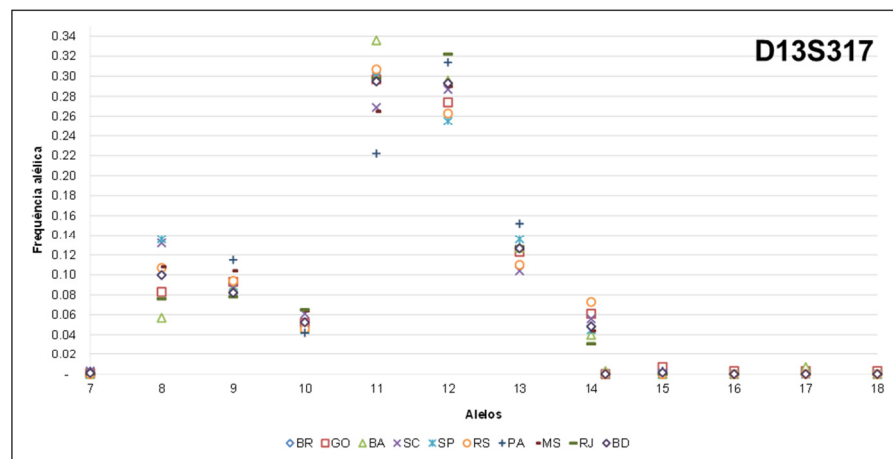


Figura 9. Distribuição das frequências alélicas do marcador D13S317.

O marcador STR do D16S539 (figura 10), apresentou uma distribuição alélica homogênea ($p=1,00$), para o teste de Kruskal-Wallis comparando os BDs populacionais. Contendo 16 alelos, quatorze destes apresentaram alelos com as frequências variadas. Com apenas 12,5% das FAs de todos os alelos, os alelos 11 e 12 apresentaram-se como os mais frequentes ($f \geq 0,24$). O alelo 13 apresentou-se como o mais variável entre os BDs populacionais do SP (0,096) e RJ (0,180), com a diferença entre as frequências de 0,084. Adicionalmente, pode-se inferir que o marcador D16539, possui em sua maioria alelos com frequências discriminativas.

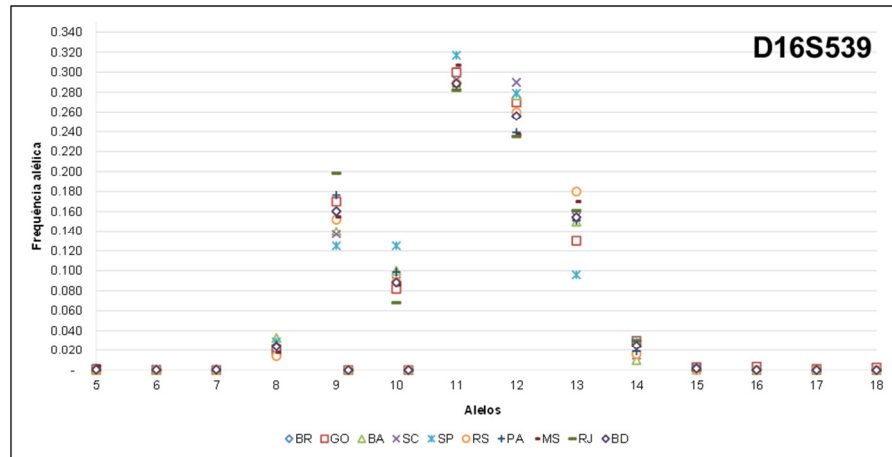


Figura 10. Distribuição das frequências alélicas do marcador D16S539.

O marcador STR D5S818 (figura 11), apresentou uma distribuição alélica homogênea ($p=0,99$), segundo o teste de Kruskal-Wallis, comparando os BD populacionais. Contendo 14 alelos representados, o marcador D5S818 apresentou 12 alelos com as frequências variadas. Pertencendo apenas 14% dos alelos, os alelos 11 e 12 apresentaram-se como os mais frequentes, sendo ambos, mais variáveis entre os BDs populacionais do PA (11: 0,285; 12: 0,400) e SP (11: 0,427; 12: 0,264), com a diferença entre a frequência do alelo 11 de 0,142 e o alelo 12 de 0,136. Por apresentar essas informações pode considerar que o marcador STR do D5S818 possui dois alelos bastante variantes entre as populações do PA e SP. Adicionalmente, pode-se inferir que o marcador D5S818, possui em sua maioria *locus* com baixo poder discriminativo entre as regiões analisadas.

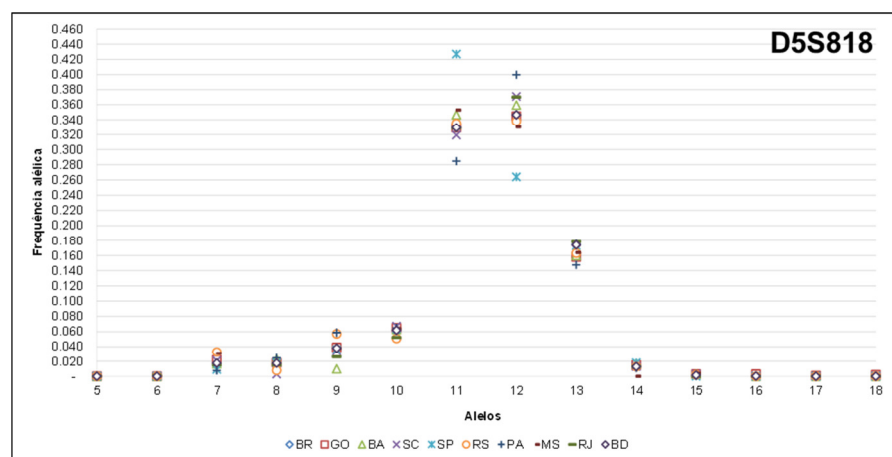


Figura 11. Distribuição das frequências alélicas do marcador D5S818.

Podemos observar que marcador STR do D7S820 (figura 11), apresentou uma distribuição alélica homogênea entre os BDs populacionais ($p=0,97$, pelo teste de Kruskal-Wallis). O marcador STR do D7S820 contendo 16 alelos representados para a população brasileira, apresentou 15 alelos com frequências variadas. Pertencendo apenas 6,66% dos alelos, o alelo 11 apresentou-se como o mais frequente, sendo ele, o mais variável entre os BDs populacionais de SP (0,218) e BA (0,320), com a diferença entre as frequências de 0,112. Por apresentar essas informações pode considerar que o marcador STR do D7S820 possui um alelo significativamente variante entre as populações do SP e BA. Adicionalmente, pode-se inferir que o marcador D7S820, possui em sua maioria *locus* com frequências de poder discriminativo.

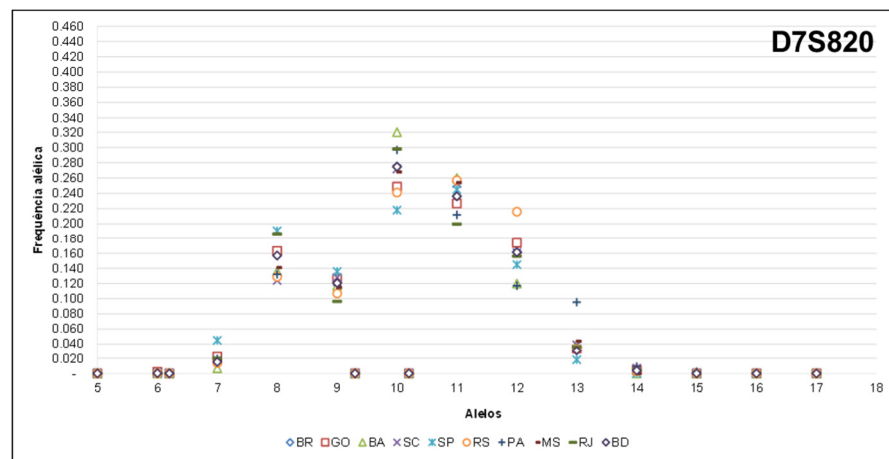


Figura 12. Distribuição das frequências alélicas do marcador D7S820.

O marcador STR do D8S1179 (figura 13), apresentou uma distribuição alélica homogênea ($p=1,00$), segundo o teste de Kruskal-Wallis, entre os BDs populacionais. Contendo 14 alelos, apresentou 12 alelos com as frequências variadas. Pertencendo apenas 14,3% dos alelos, os alelos 13 e 14 apresentaram-se como os mais frequentes. O alelo 15 obteve a maior variação alélica entre os BDs populacionais de SP (0,100) e BA (0,173), com a diferença as frequências de 0,073. Por apresentar essas informações pode considerar que o marcador STR do D81179 possui um alelo com pouca variância entre as populações do SP e BA (figura 19). Adicionalmente, pode-se inferir que o marcador D8S1179, possui em sua maioria alelos com frequências pouco discriminativas.

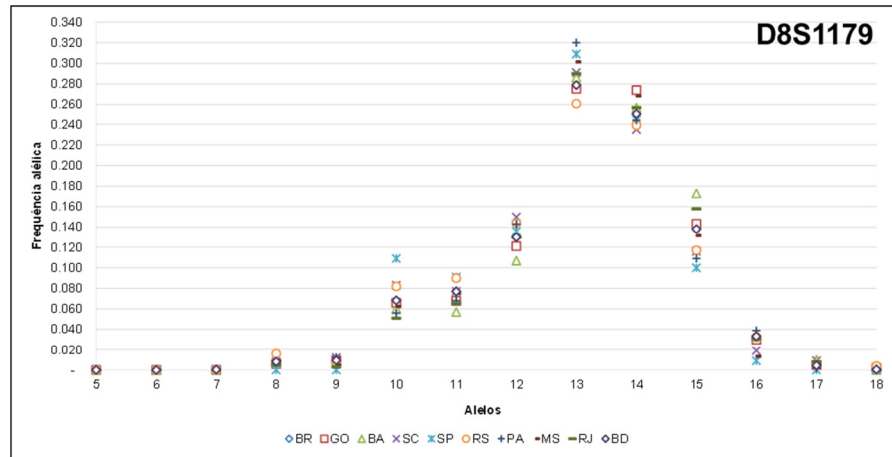


Figura 13. Distribuição das frequências alélicas do marcador D8S1179.

O marcador STR do FGA (figura 14), apresentou uma distribuição alélica homogênea entre os BDs populacionais ($p=0,99$), segundo o teste de Kruskal-Wallis. Contendo 20 alelos, com suas frequências variadas, o alelo 21 apresentou-se como o mais frequente, tendo ele a maior variação alélica entre os BDs populacionais de MS (0,091) e SP (0,132), com a diferença entre as frequências de 0,043. Por apresentar essas informações pode considerar que o marcador STR do D81179 possui um alelo com uma baixa variação entre as populações do MS e SP (figura 20). Adicionalmente, pode-se inferir que o marcador FGA, possui em sua maioria alelos com frequências alélicas com um baixo poder discriminativo.

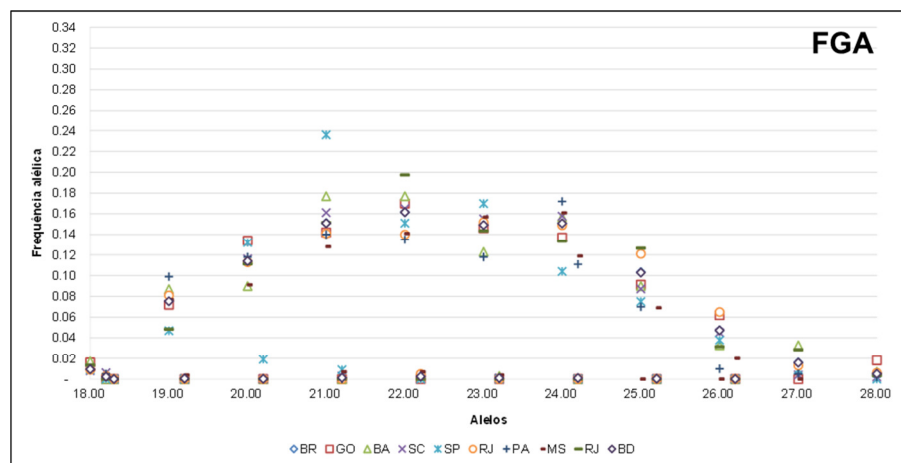


Figura 14. Distribuição das frequências alélicas do marcador FGA.

O marcador TPOX (figura 15), apresentou uma distribuição alélica homogênea entre os BDs populacionais para o marcador STR do TPOX. Todos os 10 alelos avaliados para este marcador, apresentam frequências variadas. O alelo 11 foi o mais frequente, apresentando

uma variação de 0,22 (SP) a 0,316 (MS). O alelo 12, para o BD de SP apresentou uma FA de 0,23, superior 4,6x a média (0,05) das FAs obtidas nas demais regiões.

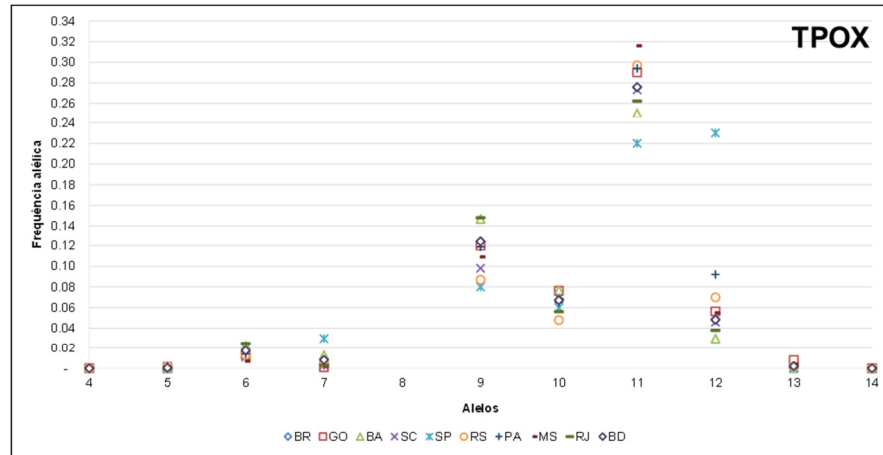


Figura 15. Distribuição das frequências alélicas do marcador TPOX.

O marcador STR do vWA (figura 16), apresentou uma distribuição alélica homogênea entre os BDs populacionais ($p=0,97$), segundo o teste de Kruskal-Wallis. Com 18 alelos representados, este marcador apresentou 15 alelos com FA avaliadas. Os 16 e 17 foram considerados os alelos mais frequentes ($f \geq 0,21$), por apresentarem suas frequências alélicas acima da média. Pode-se observar também que as frequências alélicas desses alelos apresentaram uma maior variação entre os BDs populacionais, sendo o alelo 16, variando de SP (0,218) a MS (0,301), e o alelo 17, de 0,229 (SC) a 0,288 (PA). Adicionalmente, pode-se inferir que o marcador vWA, possui em sua maioria alelos com frequências discriminativas entre as regiões.

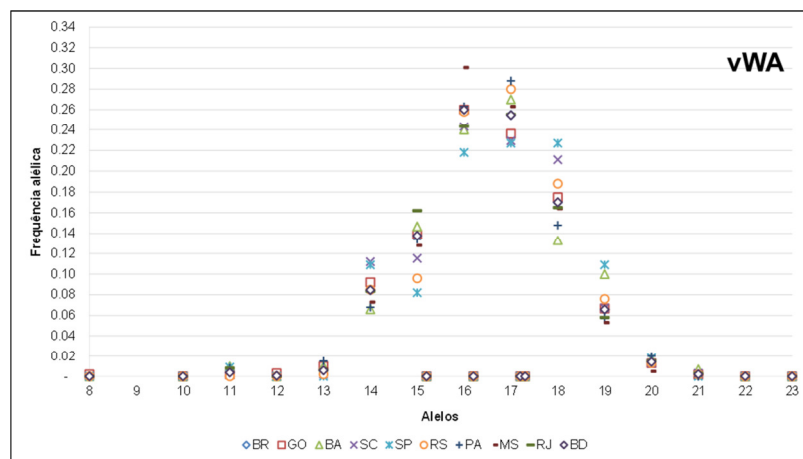


Figura 16. Distribuição das frequências alélicas do marcador vWA.

Nos 241 casos de paternidade comprovadas dos casos testados, 169 alelos não exibiram suas frequências nos bancos de dados de frequências alélicas. A Tabela 7 apresenta o número de alelos não representados nos 13 marcadores STRs (CODIS) de acordo com os Bancos de Dados populacionais utilizados na pesquisa.

Observa-se que o alelo 32.2 do marcador STR do D21S11 (tabela 7), exclusivamente no BDs de GO, não esteve presentes em 18 indivíduos testados. Observa-se também que no BDs nacional (BR) não obteve ausência de alelos em relação aos demais indivíduos e BDs testados.

Tabela 7. Número de casos com suas frequências alélicas não apresentadas nos Bancos de dados utilizados na pesquisa.

LOCI	ALELOS	BR	GO	BA	SC	SP	RS	PA	MS	RJ	TOTAL
D21S11	25		1	1	1	1					4
	25.2		2	2	2	2	2	2			12
	30		13								13
	32		2			2					4
	32.2		18								18
	33		2	2		2			2		8
	34.2		1	1		1					3
	35		1	1		1					3
	36		2			2	2	2	2		12
	37		1			1	1	1	1	1	7
38		1	1			1	1	1	1	7	
D18S51	10										2
	11					2					2
	22					2	2				4
	24			1		1	1	1		1	5
	27			1	1	1	1	1	1		6
TH01	13*		1	1	1	1	1	1	1	1	8
D3S1358	19		1			1					2
FGA	44.2*		1	1	1	1	1	1	1	1	8
	46.2		1	1	1	1	1	1	1	1	8
CSF	08					1					1
	14					1					1
	15			1	1	1			1		4
D5S818	14							5			5
TPOX	13			1	1	1					2
	14		1			1	1	1	1	1	7
D7S820	14			1		1			1		4
D8S1179	17			1		1					2
	18			1	1	1			1	1	5
vWA	21					1		1			2
Total			49	18	14	32	14	13	19	8	169

Legenda: * Alelo sem representação de frequência em todos os bancos de dados utilizados na pesquisa.

4.3 Probabilidades de Paternidade obtidas nos casos em situações de Trio e Duo

As análises estáticas da RV realizadas nos 241 casos de paternidade com 13 marcadores STRs usando as frequências alélicas disponíveis nos BDs populacionais avaliados, resultaram 2.168 análises, com resultados contendo 2.148 casos com PPs $\geq 99,99\%$ e 21 casos com PPs $< 99,99\%$ (Tabela 8).

Tabela 8. Número de probabilidades de paternidade obtidas nas análises realizadas em casos em situação de Trio (suposto pai, suposto filho e mãe biológica).

Referencial	Número de análises	(%)
PPs ¹ \geq do 99,99%	2.148	99
PPs $<$ dos 99,99%	21	01
Total	2.169	100

Legenda: ¹ Probabilidade Paterna.

As análises estáticas realizadas nos 241 casos de paternidade confirmada utilizando as frequências alélicas disponíveis nos BDs populacionais avaliados, resultaram 2.169 análises, com resultados contendo 1.686 casos com PPs $\geq 99,99\%$ e 483 casos com PPs $< 99,99\%$ (Tabela 9).

Tabela 9. Número de probabilidades de paternidade obtidas nas análises realizadas em casos em situação de Duo (Suposto pai e Suposto filho).

Referencial	Número de análises	(%)
PPs ¹ \geq do 99,99%	1.686	78
PPs $<$ dos 99,99%	483	22
Total	2.169	100

Legenda: ¹ Probabilidade Paterna.

A relação entre os números de inclusões obtidas com as frequências alélicas disponíveis nos BDs populacionais selecionados para realização das análises da RV está apresentada na figura 17.

O número de inclusões obtidas nos casos em situação de Duo e Trio pela análise da RV usando as frequências alélicas dos BDs populacionais (figura 17), selecionados para esta pesquisa apresentaram diferenças estatisticamente significativas ($p < 0,01$), podendo assim evidenciar que nos casos de investigação paterna em que a mãe biológica é também analisada para identificar categoricamente os alelos de origem paterna, que possui uma grande importância na avaliação probabilística para confirmação de uma paternidade.

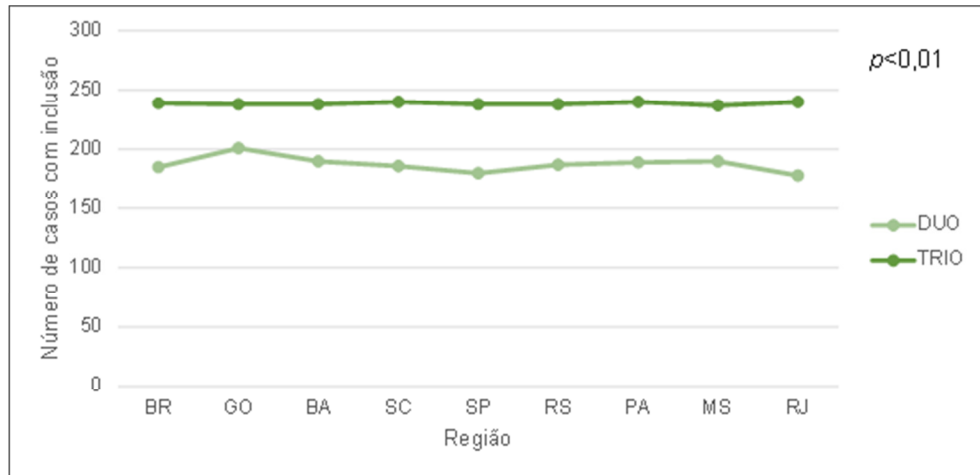


Figura 17. Relação entre o número de inclusões obtidas nas Análises da RV em 13 marcadores STRs (CODIS).

A relação entre os números de exclusões obtidas utilizando as frequências alélicas disponíveis nos BDs populacionais (figura 18), selecionados para realização das análises da RV contendo 13 marcadores STRs (CODIS).

De acordo com a figura 18 o número de exclusões obtidas nos casos em situação de Duo e Trio pela análise da RV nos BDs populacionais, selecionados para esta pesquisa, apresentam diferenças altamente significativa ($p < 0,001$), podendo assim apontar que nos casos de investigação paterna em que a mãe biológica é também analisada para identificar categoricamente os alelos de origem paterna, possui um papel fundamental na avaliação probabilística de exclusão no estudo de vínculo genético.

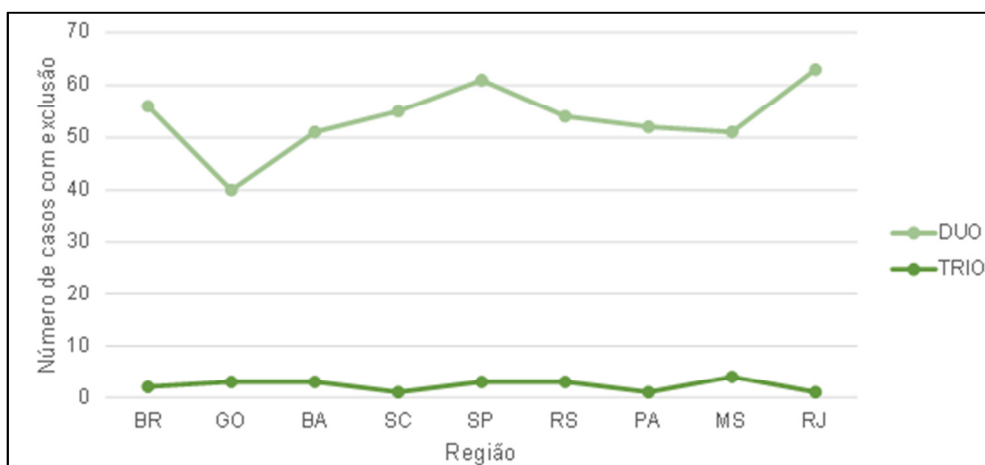


Figura 18. Relação entre o número de exclusões obtidas nas Análises da RV em 13 marcadores STRs (CODIS), ($p < 0,001$).

Foi identificado uma mutação de um cerne de repetição no marcador do FGA em um dos filhos investigantes do estudo realizado, cuja mãe biológica apresentando os alelos 24,24 o filho investigante 23,24 e suposto pai 21,22. Todas as PPs das análises realizadas para este contendo Trio (com o perfil alélico da genitora) e Duo (sem o perfil alélico da genitora) geraram resultados com valores < 99,99%.

As PPs obtidas a partir dos BDs populacionais, através das análises realizadas em situação de Trio e Duo contendo uma mutação no marcador FGA em uma dos casos testados encontra-se na Tabela 10.

Tabela 10. Probabilidades Paternas em um caso de mutação no marcador FGA, nas situações de Trio e Duo.

Banco de Dados populacionais	PP ¹ – Trio ²	PP – Duo ³
BR	99,2960%	96,6362%
GO	99,9489%	99,0280%
BA	99,3617%	95,7413%
SC	99,9554%	97,4118%
SP	99,8424%	99,0011%
RS	99,3383%	95,5392%
PA	99,6004%	98,0211%
MS	99,9197%	98,3595%
RJ	99,4117%	97,6553%
≠ entre a maior e menor PP	00,6594%	03,4888%

Legenda: ¹ Probabilidade de Paternidade. ² Conjunto de participantes contendo o suposto pai, filho e mãe. ³ Conjunto de participantes contendo o suposto pai e filho.

De acordo com a tabela 10 os casos em situação de Duo em relação aos conjuntos de Trio obtiveram as menores PPs. As PPs obtidas nas análises da RV usando as frequências alélicas dos BDs populacionais do BR, BA e RS obtiveram as menores PPs entre todos os BDs usados nas análises. Observa-se também que houve uma diferença significativa entre a menor e maior PP obtida. Diante aos resultados obtidos nos conjuntos em situação de Trio e Duo, pode-se dizer que as análises realizadas usando os 13 marcadores STRs (CODIS) contendo mutação alélica, não são suficientes para análises e inferir que o indivíduo testado é mesmo o pai verdadeiro do filho investigante.

A fim de observar a similaridade nos valores de IPs distribuídos entre os BDs regionais, comparados ao BD nacional, o Quadro 01 apresenta a análise da correlação de *Spearman* entre estes IPs obtidos pelas análises da RV em 13 marcadores STR (CODIS).

Quadro 1. Análise de correlação de *Spearman* entre os IPs de cada marcador STR entre as regiões dos BDs populacionais, comparados com o banco de dados de frequências alélicas nacional (BR).

Marcador	GO		BA		SC		SP		RS		PA		MS		RJ	
	r	Valor p	r	Valor p	r	Valor p	r	Valor p	r	Valor p	r	Valor p	r	Valor p	r	Valor p
D18S51	0,69	< 0,0001	0,62	< 0,0001	0,62	< 0,0001	0,24	0,002	0,61	< 0,0001	0,62	< 0,0001	0,27	< 0,0001	0,47	0,64
D21S11	0,81	< 0,0001	0,25	0,002	0,10	0,11	0,001	0,95	0,10	0,14	0,09	0,14	0,41	< 0,0001	-0,01	0,86
D3S1358	0,42	< 0,0001	0,83	< 0,0001	0,28	< 0,0001	0,32	< 0,0001	0,63	< 0,0001	0,26	< 0,0001	0,86	< 0,0001	0,15	0,02
TH01	0,21	0,001	0,69	< 0,0001	0,68	< 0,0001	0,71	< 0,0001	0,66	< 0,0001	0,67	< 0,0001	0,66	< 0,0001	0,64	< 0,0001
CSF1PO	0,38	< 0,0001	0,82	< 0,0001	0,80	< 0,0001	0,70	< 0,0001	0,86	< 0,0001	0,85	< 0,0001	0,84	< 0,0001	0,80	< 0,0001
D13S317	0,02	0,71	0,53	< 0,0001	0,04	0,49	0,54	< 0,0001	0,52	< 0,0001	0,52	< 0,0001	0,55	< 0,0001	0,54	< 0,0001
D16S539	0,85	< 0,0001	0,70	< 0,0001	0,90	< 0,0001	0,83	< 0,0001	0,90	< 0,0001	0,89	< 0,0001	0,91	< 0,0001	0,90	< 0,0001
D5S818	0,66	< 0,0001	0,51	< 0,0001	0,54	< 0,0001	0,46	< 0,0001	0,69	< 0,0001	0,45	< 0,0001	0,46	< 0,0001	0,72	< 0,0001
D7S820	0,62	< 0,0001	0,48	< 0,0001	0,64	< 0,0001	0,31	< 0,0001	0,66	< 0,0001	0,50	< 0,0001	0,46	< 0,0001	0,66	< 0,0001
D8S1179	0,88	< 0,0001	0,43	< 0,0001	0,69	< 0,0001	0,28	< 0,0001	0,58	< 0,0001	0,40	< 0,0001	0,65	< 0,0001	0,65	< 0,0001
FGA	0,83	< 0,0001	0,55	< 0,0001	0,31	< 0,0001	0,53	< 0,0001	0,29	< 0,0001	0,33	< 0,0001	0,63	< 0,0001	0,46	< 0,0001
TPOX	0,17	0,01	0,26	< 0,0001	0,86	< 0,0001	0,26	< 0,0001	0,85	< 0,0001	0,71	< 0,0001	0,54	< 0,0001	0,84	< 0,0001
vWA	0,34	< 0,0001	0,35	< 0,0001	0,19	0,001	0,35	< 0,0001	0,13	0,05	0,52	< 0,0001	0,25	0,001	0,19	0,001
Mediana	0,53		0,54		0,51		0,42		0,58		0,52		0,58		0,54	
DP	0,29		0,19		0,29		0,23		0,26		0,23		0,21		0,28	
IC	0,16		0,10		0,16		0,13		0,14		0,13		0,11		0,15	

Legenda: DP- Desvio Padrão; IC- Intervalo de Confiança; r – valor de rho,

Observou-se que os IPs obtidos através das frequências alélicas cada BDs populacional selecionados para esta pesquisa, apresentaram índices de correlação variados (r: -0,01 a 0,91), pelo teste de correlação de *Spearman*. Foi considerado um valor de $r \geq 0,8$, como sendo de alta correlação ou alta similaridade entre os IPs estudados.

A comparação dos BDs de GO com o BD nacional apresentou correlação positiva para os marcadores D21S11 (r: 0,81), D16S539 (0,85), D8S1179 (r: 0,88) e FGA (r: 0,83). O BD da BA apresentou o marcador D3S1358 (r: 0,83) e o marcador CSF1PO (r: 0,82). O Estado de Santa Catarina apontou o marcador CSF1PO (r: 0,82) e o D16S539 (0,90) como os marcadores com alta similaridade comparado ao banco de dados nacional. O marcador D16S539 foi o único marcador indicado com alta correlação entre o BD de SP e o BD nacional (r: 0,83). O BD do RS apontou o marcador CSF1PO (r: 0,83) e o TPOX (r: 0,85) com os dois marcadores de alta correlação com o BD nacional. Para o BD do PA foram identificados os marcadores CFF1PO (r: 0,85) e D16S539 (r: 0,89) com alta similaridade com o BD nacional. O MS apresentou três marcadores STRs com alta correlação com o BD nacional: D3S1358 (r: 0,86), CSF1PO (r: 0,84) e o marcador D16S539 (r: 0,91). Por último, o BD do RJ indicou o marcador CSF1PO (r: 0,80), D16S539 (r: 0,90) e TPOX (r: 0,84).

O BD de SP apresentou o menor número de marcadores indicado com alta correlação, 7,7% (1/13). Por outro lado, o BD do Estado de Goiás apresentou o maior número de marcadores com alta similaridade com o banco de dados nacional, quatro marcadores (30,8%). Neste sentido, pode-se inferir que os IPs obtidos a partir dos 241 indivíduos do presente estudo quando utilizado o BD de GO, obteve maior similaridade com o BD nacional, indicando uma possível homogeneidade entre as FAs dos 13 marcadores utilizados. Em contrapartida, utilizando o BDs de SP, apenas um marcador foi apresentado como marcador homogêneo com o BD nacional.

Avaliando as frequências dos marcadores que apresentaram alta correlação entre os BDs comparados, os marcadores D18S51, TH01, D13S317, D5S818 e D7S820 não foram apontados como alta correlação entre os bancos de dados regionais e o BD nacional. Os marcadores D21S11, D8S1179 e FGA foram indicados apenas uma vez (12,5%) em cada BD regional. Os marcadores D3S1358 e TPOX surgiram em 25% (2/8) das comparações realizadas entre os BDs regionais e o BD nacional, para cada marcador. Por último, os marcadores CSF1PO e D16S539 foram indicados, individualmente, em 75% (6/8) das comparações realizadas entre os BDs regionais e o BD nacional, com alta similaridade.

Avaliando as médias obtidas dos IPS de cada marcador pode-se identificar o marcador D21S11 (IP médio: 6,71) e o FGA (IP médio: 5,55) com o maior poder de inclusão entre os 13 marcadores estudados. Por outro lado, os marcadores TH01 (IP médio: 2,75) e D3S1358 (IP médio: 2,93) apresentaram o menor poder de inclusão entre os marcadores STR avaliados. Tais resultados permitem inferir que a distribuição alélica de D21S11 e FGA podem ser mais discriminativos entre os marcadores, considerando que o valor de IP elevado está relacionado ao poder de inclusão de um determinado alelo. Avaliando individualmente estes marcadores, podemos observar que os alelos 11, 12 e 15 do marcador FGA apresentam baixa frequência na população do Estado de Goiás, porém, não observada em nenhum dos outros sete BDs avaliados. Possivelmente, a baixa frequência de um alelo na população é um fator decisivo para o aumento do IP individual deste marcador e, conseqüentemente, o aumento do poder de inclusão deste marcador no perfil alélico do indivíduo.

Desta forma, a utilização de um BDs mais representativo favorece o poder de inclusão de marcadores discriminativos, com alelos que apresentam baixa frequência na população de uma determinada região. O Quadro 2 apresenta a análise de correlação de *Spearman* entre as PPs obtidas para cada BD comparando ao banco de dados nacional.

Quadro 2. Análise de correlação de *Spearman* entre as probabilidades de paternidade (PP) dos marcadores STR e as regiões dos bancos de dados populacionais, comparados com o banco nacional (BR),

Marcador	GO	BA	SC	SP	RS	PA	MS	RJ	Valor p
D18S51	0,99	0,98	0,98	0,96	0,98	0,99	0,98	0,99	<0,001
D21S11	0,99	0,76	0,92	0,78	0,89	0,90	0,86	0,93	<0,001
D3S1358	0,86	0,84	0,88	0,75	0,92	0,90	0,83	0,89	<0,001
TH01	0,81	0,88	0,81	0,84	0,85	0,79	0,79	0,87	<0,001
CSF1PO	0,87	0,95	0,96	0,87	0,97	0,98	0,95	0,99	<0,001
D13S317	0,97	0,98	0,99	0,99	0,99	0,99	0,99	0,99	<0,001
D16S539	0,93	0,99	0,97	0,94	0,94	0,99	0,96	0,95	<0,001
D5S818	0,95	0,98	0,99	0,97	0,99	0,99	0,96	0,95	<0,001
D7S820	0,99	0,97	0,99	0,96	0,99	0,97	0,98	0,99	<0,001
D8S1179	1,00	0,97	0,98	0,92	0,98	0,97	0,98	0,95	<0,001
FGA	0,96	0,99	0,90	0,89	0,96	0,85	0,56	0,97	<0,001
TPOX	0,96	1,00	0,99	0,83	0,97	0,97	0,99	0,99	<0,001
vWA	1,00	0,94	0,96	0,90	0,96	0,97	1,00	0,97	<0,001
Média	0,94	0,94	0,95	0,89	0,95	0,94	0,91	0,96	
DP	0,06	0,07	0,06	0,07	0,04	0,07	0,13	0,04	
IC	0,03	0,04	0,03	0,04	0,02	0,04	0,07	0,02	

Legenda: DP, desvio padrão. IC, intervalo de confiança.

De acordo com o Quadro 2, observa-se que as PPs obtidas nas análises da RV dos BDs populacionais selecionados para esta pesquisa não apresentaram diferenças estatísticas significativas pelo teste de correlação de *Spearman* ($p < 0,001$), bem como apresentando em sua maior parte valores de $r \geq 0,80$ com correlação positivas fortes entre os BDs populacionais, comparados ao BDs populacional do BR. Porém, os marcadores D21S11 do BD populacional da BA e D3S1358 do BD do PA e MS, as PPs obtidas em relações as PPs obtidas no BD nacional (BR) obtiveram valores de $r < 0,8$, assumindo assim uma correlação positiva moderada entre eles.

A Figura 25 indica a distribuição espacial dos valores de PPs obtidos para cada BD estudado. É possível observar um único indivíduo que alterando o banco de dados populacional, pode-se observar a variação de PPs de 0,229, utilizando o BD do Estado da Bahia, para 0,764 do BD de SP. A média dos PPs médios foi de 0,372. Tal dado permite indicar que os BDs de SC (PP médio: 0,396), MS (IP médio: 0,460) e SP (0,764) apresentaram um maior poder de inclusão da PP, utilizando os 13 marcadores STR utilizados (figura 25). Tal observação, permite inferir que segundo o BD utilizado a PP pode alterar consideravelmente, variando em 53,5% o poder de inclusão de vínculo genético.

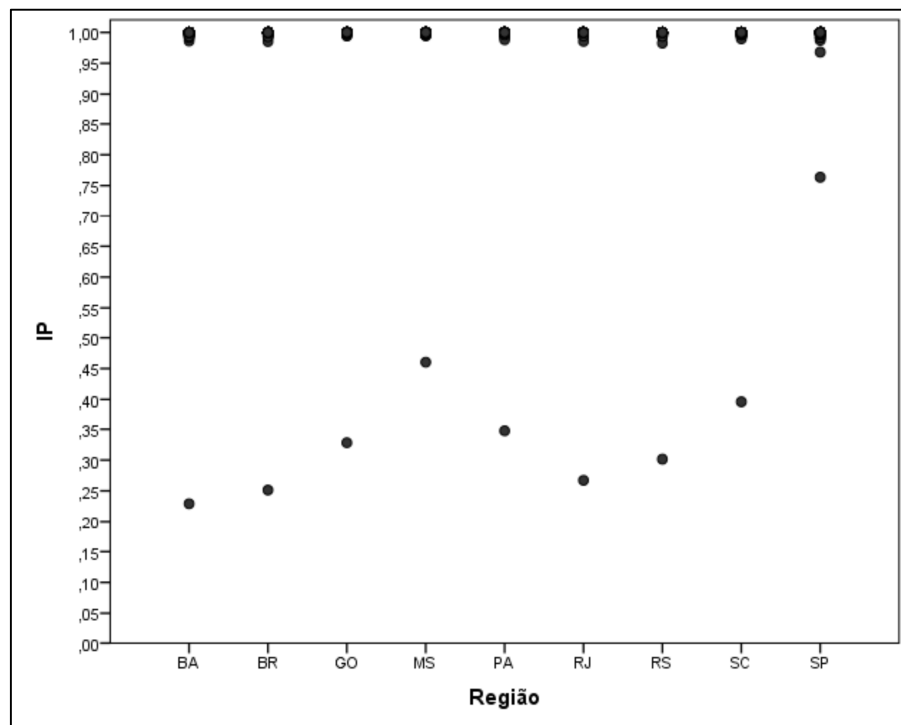


Figura 19. Distribuição espacial das PPs obtidas a partir de 241 indivíduos analisados entre os BDs das regiões avaliadas.

5. DISCUSSÃO

5.1 O potencial informativo das frequências alélicas encontradas nos 9 BDs populacionais, sendo 1 nacional (BR) usados na pesquisa

O presente estudo buscou comparar as frequências alélicas entre os bancos de dados a fim de permitir avaliar a probabilidade de vínculo genético entre indivíduos, com um valor mais fidedigno, respeitando as variações mutacionais de uma determinada população. Neste sentido, Francez (2012) afirma que a existência de várias populações sugere que os indivíduos de uma espécie não se distribuem aleatoriamente, mas sim de acordo com diversos fatores, como pela segregação espacial ou pela estruturação social. Sendo assim, estudos filogenéticos de populações humanas têm sido desenvolvidos visando a compreensão da história, demografia e evolução da espécie humana, além de estudos de ancestralidade. Por outro lado, este mapeamento tem sido utilizado no estudo de vínculo genético e identificação humana em casos criminais.

Porventura, a utilização de frequências alélicas de bancos de dados de outras regiões pode não ser representativa para a população de uma determinada cidade, ou mesmo da região em que ela pertença. Dessa forma, o presente estudo permitiu uma melhor compreensão das diferenças alélicas e do poder de inclusão dos marcadores STR utilizados rotineiramente, na área civil ou criminal na genética forense, com um banco de dados mais representativo. Infelizmente, poucos estudos foram realizados no Brasil comparando os bancos de dados populacionais existentes.

Os resultados de PPs $<$ ou $\geq 99,99\%$ observados no presente estudo, utilizando os 9 BDs populacionais, sendo 1 nacional (BR), apresentando diversos alelos variantes entre eles. Tais resultados corroboram com as observações feitas por Butler (2012b) que demonstra a importância da valorização do perfil genético de um indivíduo através de um marcador de DNA, tendo como base as frequências alélicas de cada *locus* na população de referência,

A pesquisa realizada por Cordeiro (2009), sobre a imigração de africanos para os Estados da BA, RJ, SP, SC e RS entre os anos de 1500 e 1800, torna-se mais evidente com o estudo realizado por Aguiar (2011), o qual demonstra que a população brasileira possui fortes semelhanças com povo Africano com relação ao marcador STRs do D21S11. Estes estudos reforçam esta pesquisa realizada onde se utilizou o marcador STR do D21S11, podendo assim

inferir que os alelos mais frequentes (28, 29 e 30) disponíveis nos 5 BDs populacionais da BA, RJ, SP, SC e RS podem estar relacionados com a forte presença destes povos nestes Estados, e também dizer que estes tais alelos podem ser fortes candidatos para auxiliar na identificação e origem dos indivíduos presentes nestas regiões. Adicionalmente, segundo a avaliação das FAs dos 13 marcadores CODIS foi possível observar uma distribuição alélica geral homogênea entre todos os BDs populacionais. Porém, a análise individual demonstrou haver diferença estatisticamente significativa confrontando as FAs dos BDs de SP ($p < 0,05$) e MS ($p < 0,05$) com o BD nacional, para o marcador D18S51.

Segundo o IFSG (2007), as frequências alélicas podem variar entre as populações, isso se deve possivelmente pela composição étnica da população estudada na geração de um BDs. Esses resultados podem alterar a força exercida sobre a frequência de tal alelo na população estudada em comparação a outra população em que pertencem os indivíduos testados. Essa informação se confirma com os resultados das análises obtidas nos marcadores STRs D3S135, D13S317 e D5S818. O marcador STRs do D3S135 em relação à variação das frequências alélicas do alelo 15 apresentou variação das frequências entre os BDs de Goiás e São Paulo, o marcador STR D13S317, e o marcador D5S818 apresentou a mesma informação nas análises comparativas das frequências dos alelos 11 e 12 nos BDs de SP e PA (com o alelo 11 mais frequente no PA e com o alelo 12 mais frequente em SP). A utilização de um desses BDs populacionais sem o prévio estudo sobre as origens dos indivíduos testados poderão gerar resultados equivocados, podendo incluir ou excluir indivíduos de uma paternidade que contenham em seu perfil alélico esses alelos variantes, ou até mesmo gerar resultados para inocentar ou acusar tais indivíduos.

Particularmente, a partir dos resultados obtidos entre a comparação entre as frequências alélicas de todos os BDs populacionais e 1 BDs populacional (BR), pode-se inferir que o marcador STR do TH01 apresentou-se com marcador de baixo poder de inclusão para realização de testes de vínculo genético, por apresentar-se um baixo número de alelos relativamente variáveis entre os BDs e por ter poucos alelos discriminativos (42%). Além dessas informações, o marcador STR do TH01 quando submetido às análises da RV, mantém estável por conter em sua maioria alelos frequentes.

Os marcadores STR do CSF1PO, D16S539, D7S820, D8S1179 E FGA apresentaram na avaliação comparativa um alto poder discriminativo por apresentarem na maioria de seus alelos frequências baixas, como também entre os BDs populacionais utilizados na pesquisa.

Desta forma, pode-se inferir que estes marcadores apresentaram-se um alto poder de inclusão populacional, segundo o alelo estudado.

Corroborando com o estudo populacional feito por Moret (2009) o qual apresenta que entre os alelos 10 e 11, do marcador TPOX, os alelos de maior frequência populacional. Dependendo do BD populacional, esse alelo irá representar 30% das frequências alélicas em relação aos outros alelos disponíveis no marcador TPOX, como é o caso do BD do MS que apresentou a frequência de 0,316. Assim, esta frequência faz com que esse alelo se torne menos discriminativo para um estudo de inclusão de paternidade, quando apresentado em um APO de um indivíduo do MS, e mais discriminativo quando apresentado em um APO de outro indivíduo de SP. Tal informação reforça a inferência feita por Corte-Real (2004): *“Independente de todo este desenvolvimento técnico, não é possível afirmar-se uma probabilidade de paternidade ou a probabilidade que numa mancha pertença a determinado suspeito, sem o estudo da população de referência na qual decorre esta investigação”*.

Adicionalmente, os dados das FAs e comparando os BDs avaliados, corroboram com a orientação das Diretrizes para Operações de Identificação de DNA de Fatalidade em Massa de 2010 para operações de identificação humana em condições de fatalidade em massa, no qual o pesquisador deve observar a presença e ausência das frequências dos *loci* antes de realizar os cálculos de probabilidade de vínculo genético em BD nacionais e regionais. Portanto, um dos bancos de dados avaliados por este estudo não constava a frequência em um dos alelos apresentado no indivíduo testado, fazendo com que a análises da RV se torne ineficiente devido a correção de tal dado.

5.2 As PP obtidas com o uso dos 13 marcadores STRs (CODIS) nos casos em situação de Trio e Duo

Os dados obtidos a partir das PP pelo presente estudo corrobora com Elegante (2006), que relatou que diferentes populações apresentam alelos em frequências diferentes e algumas vezes únicos. Sendo assim a maioria dos bancos de frequências alélicas selecionados para este estudo não possuíam algumas frequências alélicas em alguns alelos presentes nos indivíduos submetidos aos cálculos estatísticos, Além de apresentarem frequências diferentes entre os *loci* de alguns alelos, geraram valores abaixo e acima dos 99,99% de PP, ou seja, estatisticamente aceitando a paternidade em alguns casos, e em outros, ou até mesmo

excluindo a paternidade por não alcançar a PP recomendada pela AABB (2010) e IFSG (2010).

De acordo com o resumo anual do relatório para teste de DNA, publicado pela AABB de 2010, os testes de paternidade sem o perfil alélico da mãe biológica pode apresentar uma série de problemas, pois o índice de paternidade poderá ser cortado ao meio, principalmente em casos contendo alelos homocigotos, que podem chegar a um IPC contendo um décimo do que quando o perfil alélico da genitora é utilizado, cgerando assim PPs com valores < 99,99%, Situação que justifica a quantidade de resultados obtidos com as análises da RV realizadas nos casos em que o perfil alélico materno não foi incluso.

Nos casos em situação de Duo quando há uma aparente inconsistência (mutação) o resultado poderá não ser conclusivo devido à ausência do perfil alélico materno. Talvez por esse motivo, o teste de paternidade em que se utilizam os alelos da mãe, mesmo que a maternidade não está sendo contestada, torna-se importante, pois na avaliação da relação questionada, os alelos da mãe biológica melhora a possibilidade de se obter resultados mais precisos, como também servindo como um controle de qualidade (AABB, 2010). Com essas observações e diante dos resultados deste estudo pode-se considerar que a ausência do perfil elélico da mãe influenciou consideravelmente nos resultados dos cálculos de vínculo genético, fazendo com que as PP sejam reduzidas ao ponto de que os resultados das análises não sejam conclusivos para afirmar a paternidade. Tal fato pode ocasionar análises incompletas ou até mesmo equivocadas.

Penacino e colaboradores (2003) demonstram que em certos casos são acrescentados mais *locus* no cálculo estatístico para se chegar a uma PP confiável, desde que as taxas de mutação sejam conhecidas, podendo assim ser incorporadas nos cálculos estatísticos para obtenção de um resultado confiável na contestação de uma paternidade. Isto reforça a importância em se ter um número significativo ou adequado de *loci* a serem usados nas análises estatísticas para confirmação de paternidade, principalmente nos casos em que tenham apresentado algum tipo de mutação. Esta observação também pode ser identificada no presente estudo, pois em uma mutação alélica encontrada em um dos casos analisados, não alcançou o a PP desejada, pois os 13 *loci* contendo uma mutação alélica não foram suficientes para concluir as análises estatísticas, deixando assim o resultado do exame inconclusivo, variando a PP de 0,229 a 0,764.

Os valores baixos das PPs obtidas com as frequências dos bancos de dados do Estado de Goiás e Mato Grosso do Sul, demonstraram a importância de se ter um BD contendo uma quantidade significativa de alelos variáveis e uma quantidade significativa de *loci* que possam garantir resultados confiáveis em uma análise de vínculo genético.

Babol-Pokora e colaboradores (2005) descreveram uma análise comparativa semelhante a este estudo, os quais utilizaram 10 alelos do kit AmpFISTR® SGM Plus® contendo 08 alelos utilizados pelo CODIS. Os autores avaliaram 150 casos em situação de Trio e Duo. Os resultados das Probabilidades de Paternidade obtidas demonstraram valores inferiores a 99,99% em 60% dos casos analisados em situação de Trio, como também todos os valores em situação de Duo obtidos com a PP foram menores que 99,94%. Segundo a *Comissão Polonesa de Genética Forense*, esses valores são considerados insuficientes para julgar casos de paternidade, Para os resultados obtidos neste estudo nos bancos de frequência alélica utilizados, somente 1% dos casos em situação de Trio e 23% dos casos em situação de Duo obtiveram valores abaixo do recomendado. Ou seja, os valores obtidos pelas análises em situação de Duo não foram consideravelmente significativos para realização da conclusão das análises desses casos de investigação vínculo genéticos.

Por outro lado, os resultados utilizando o banco de frequência brasileira, somente um caso em situação de Trio contendo uma mutação, obteve uma PP abaixo de 99,99%, enquanto 55 casos em situação de Duo obteve índice abaixo de 99,99% de PP. Portanto, os resultados obtidos nos casos em situação de Duo corroboram com os resultados do estudo realizado por Babol-Pokora, e colaboradores (2005). Adicionalmente, os resultados suportam a insuficiência estatística encontrada em casos em que a situação de Duo nas análises de RV em que se utilizam os 13 *loci* STRs do CODIS para realização do estudo de vínculo Genético,

Pode-se avaliar também os resultados obtidos com a utilização das frequências alélicas do Brasil realizado por Aguiar e colaboradores (2011), no qual gerou a maior quantidade de PPs acima da PP de 99,99%, podendo assim, ser justificado pela quantidade de indivíduos testados e a variedade étnica e da mistura genômicas inseridas nos 100.000 indivíduos obtidos na população brasileira. Essas variações também fizeram com que a distribuição das frequências alélicas alcançasse valores mais observáveis.

Além do caso relatado com a PP abaixo dos 99,99% de vínculo de parentesco utilizando os 13 marcadores STRs, existem casos mais complexos em que a utilização de 15 STRs não garante o poder de discriminação desejado, resultando assim em um caso sem

resultado conclusivo. Testes de vínculo genéticos chamados 'casos atípicos', em que o suposto pai é falecido e não dispendo de parentes suficientes para a reconstrução do perfil alélico do mesmo, ou em casos em que os pretensos genitores são relacionados e possuem semelhança genética, estão entre as situações em que a utilização de 15 *loci* genéticos pode não apresentar resultados incontestáveis. Até mesmo 'casos típicos' de investigação de paternidade podem ser dificultados, quando são encontradas mutações (Betz, 2007).

Outro dado relevante sobre o número de *loci* a ser utilizado em uma análise de investigação de paternidade, em que o perfil alélico materno não foi inserido nas análises de RV, está no fato que os 13 *loci* poderão influenciar nos resultados gerando falsas inclusões, principalmente em situações nos casos em que apresentam mutações, como também nos casos em que um parente próximo do suposto pai será testado (Poetsch, 2006; Betz, 2007).

Portanto, os casos que não são confirmados estatisticamente com 13 ou até mesmo com 15 STR *loci* (por exemplo, a paternidade deficiente, pares de parentes, ou possíveis eventos de mutação), sugere-se que o pesquisador utilize *loci* STR adicionais ou marcadores alternativos de DNA tais como os *loci* dos cromossomos X ou Y (Poetsch, 2006, Betz, 2007, ISFG, 2007, AABB 2009).

5.3 Avaliação dos dados obtidos pelo nos testes estatísticos Correlação de Spearman e Exato de Fisher comparando as frequências alélicas, IPS e PPs

As distribuições gerais das frequências alélicas, dos *loci* STR, em várias populações brasileiras são bastante semelhantes, assim como demonstrados nos estudos comparativos das frequências alélicas realizados por Leite (2006b) e Rodrigues (2007), Apesar dessas semelhanças, alguns BDs populacionais utilizados não apresentaram frequências alélicas em alguns de seus alelos, limitando ainda mais as análises estatísticas para se confirmar o vínculo genético entre os envolvidos testados, usando os 13 marcadores STRs (CODIS). Estes dados confirmados com o teste estatístico de correlação de Spearman entre os IPs gerados nas análises da RV nos 241 casos, demonstrando que há diferenças estatísticas significativas entre os BDs analisados e o BDs populacional nacional.

Diante das informações acerca da ausência de frequências alélicas entre os BDs utilizados nesta pesquisa, confrontou-se todas as frequências alélicas de todos os 9 BDs populacionais avaliados neste estudo com as frequências alélicas fornecidas pelo BDs

populacional do BR, propondo-se assim um novo BDs populacional, composto pelas frequências alélicas de 100.000 indivíduos do Brasil mais 102.807 indivíduos das 6 regiões do Brasil (Anexo II).

De acordo com os dados obtidos pelo teste de correlação de *Spearman* foi possível identificar uma alta similaridade entre o BD de GO e o BD nacional para os marcadores D21S11 (r : 0,81), D16S539 (0,85), D8S1179 (r : 0,88) e FGA (r : 0,83). Segundo Góes (2006) e Pena (2004) no Brasil devido à miscigenação verifica-se que a distribuição de frequências alélicas não apresenta variação significativamente entre os diferentes grupos étnicos, o que pode ser observado no presente estudo. Adicionalmente, o BD do Estado de Goiás apresentou o maior número de marcadores com alta similaridade com o banco de dados nacional, quatro marcadores (30,8%). Neste sentido, pode-se inferir que os IPs obtidos a partir dos 241 indivíduos do presente estudo quando utilizado o BD de GO, obteve maior similaridade com o BD nacional, indicando uma possível homogeneidade entre as FAs dos 13 marcadores utilizados. Em contrapartida, utilizando o BDs de SP, apenas um marcador foi apresentado como marcador homogêneo com o BD nacional.

A avaliação dos IPs por marcador tornou-se uma ferramenta adicional útil na avaliação de divergências populacionais. Avaliando as médias obtidas dos IPS de cada marcador pode-se identificar o marcador D21S11 (IP médio: 6,71) e o FGA (IP médio: 5,55) apresentaram o maior poder de inclusão entre os 13 marcadores estudados. Por outro lado, os marcadores TH01 (IP médio: 2,75) e D3S1358 (IP médio: 2,93) apresentaram o menor poder de inclusão entre os marcadores STR avaliados. Desta forma, é possível inferir que a distribuição alélica de D21S11 e FGA podem ser mais discriminativos entre os marcadores, considerando que o valor de IP elevado está relacionado ao poder de inclusão de um determinado alelo. Não foi encontrando nenhum estudo, utilizando marcadores STRs que avaliasse o IP destes marcadores para comparar diferenças entre regiões diferentes.

Com estes resultados, a análise das FAs, IPs, PPs e números de inclusão dos 13 marcadores STR avaliados pelo presente estudo, comparando-os entre os diferentes BDs de oito regiões do país e um banco de dados nacional confirmaram o elevado grau polimórfico e alta heterozigose dos mesmos. Ao final, a proposta de um novo banco de dados (Tabela 24 em Anexo) permite incorporar estes valores para o ajuste do banco de dados nacional (BR), assim como novas análises populacionais devem ser feitas para melhor elucidação da variabilidade dos *locus* STR, para aplicações em análises forenses, estudos de vínculo genético e estudos

genéticos populacionais. Consequentemente, a obtenção da PP a partir de dados mais representativos permite confirmar ou excluir uma relação de vínculo genético de parentesco ou identificação criminal com finalidade jurídica tornando-se mais confiável, evitando consequências negativas na área de criminal, de família ou de direitos humanos.

6. CONCLUSÃO

As análises comparativas dos resultados obtidos neste estudo demonstraram a interferência estatística que cada banco de dados de frequências alélicas pode exercer nos cálculos de probabilidade de paternidade quando se utiliza apenas 13 *loci* na análise e comprovação um vínculo genético. Essa interferência estatística se torna mais significativa nos casos em que o perfil alélico materno não é inserido na análise, podendo assim impossibilitar uma análise precisa. Demonstra também a existência de alelos com frequências divergentes entre cada região do país. Isso reforça a importância em se ter um banco regional contendo características genotípicas da população, principalmente aquelas populações isoladas.

De acordo as informações disponíveis nos bancos de dados de frequências alélicas de cada região, foi possível concluir, que as frequências alélicas obtidas, bem como os índices de paternidade por marcador, e as probabilidades de paternidade, possuem fortes similaridades com o banco de dados nacional.

Os resultados gerados neste estudo (Tabelas em anexo) complementam e corroboram para um banco de dados representativo, em que os pesquisadores poderão utilizar nas análises de vínculo genético, nos indivíduos da população do Estado de Goiás, e demais Estados brasileiros.

Estudos que preconizam a caracterização da distribuição alélica de diferentes regiões do Brasil utilizando os marcadores microssatélites tornam-se imprescindíveis para a identificação de uma baixa diferenciação genética entre elas, demonstrando a capacidade dos marcadores microssatélites em detectar diferenciações genéticas entre populações.

Os resultados obtidos nesta pesquisa criam uma avaliação crítica e analítica no que se trata em distribuição populacional e étnica que se apresenta no Brasil, e a importância que se tem em se criar um BDs contendo todas as informações necessárias para realização de estudos de vínculo genético e de identificação humana.

7. REFERÊNCIAS BIBLIOGRÁFICAS

AGUIAR VRC, WALFGRAMM EV, MALTA FSV, *et al.* Corrigendum to “Updated Brazilian STR allele frequency data using over 100,000 individuals. An analysis of. CSF1PO, D3S1358, D5S818, D7S820, D13S317, D13S317, D16S539, D18S51, D21S11, FGA, Penta D, Penta E, TH01, TPOX e *Vwa loci*, Forensic Science International: Genetics; 2011, 771:1-6.

ALVES C, SOUZA T, VEIGA S, *et al.*. Importância do sistema de histocompatibilidade humano (HLA) em Pediatria. Revista Pediatria São Paulo. 2005; 27 (4): 274-86.

ANNUAL REPORT SUMMARY FOR TESTING IN 2004. AABB [Capturado 03 nov, 2014] Disponível em: <https://www.aabb.org/sa/facilities/Documents/>

ANNUAL REPORT SUMMARY FOR TESTING IN 2008. ABB [Capturado 03 nov, 2014] Disponível em: <https://www.aabb.org/sa/facilities/Documents/>

ANNUAL REPORT SUMMARY FOR TESTING IN 2010. AABB [Capturado 03 nov, 2014] Disponível em: <https://www.aabb.org/sa/facilities/Documents/>

AQUINO JG. Desenvolvimento e validação de sistema multiplex X-STR para identificação humana por DNA. [Dissertação] Rio de Janeiro: Universidade Estadual do Rio de Janeiro; 2011.

BABOL-POKORA K, JACEWICZ R, PEPIINSKI, *et al.*. Is SGM Plusk the sufficient system for paternity testing? International Congress Series, 2006; 1288: 450–452.

Bamshad MJ, Wooding S, Waltskins WS, Ostler CT, Batzer MA, Jorde LP. Human population genetic structure and inference of group membership. Am J Hum Gent; 2003, 72: 578-589.

BETZ, THOMAS. Paterniplex, a highly discriminative decaplex STR multiplex tailored for investigating special problems in paternity testing. Electrophoresis, Goettingen, 2007; 28 (21): 3868-3874.

BRENNER C H. Multiple mutações, covert mutations and false exclusions in paternity case Work. Int Cong Series; 2004, 1261:112-114.

BROOKES, ANTHONY J. The essence of SNPs Gene, July 1999; 234 (2):177-18.

BUTLER JM. Short Tandem Repeat Markers, in Fundamentals of Forensic DNA Typing. 2 ed. 147-173, Estados Unidos América, Elsevier; 2012a.

BUTLER JM. Commonly used Short Tandem Repeat Markers and Commercial Kits in Forensic DNA Typing. 2 ed. 85-121, Estados Unidos América, Elsevier; 2012b.

- BUTLER JM. *Advanced Topics in Forensic DNA Typing: Methodology USA*. Academic Press. 2012c; 699.
- BUTLER JM. Short Tandem Repeat (STR) Loci and Kits, in *Advanced Topics in Forensic DNA Typing: Methodology*. 2 ed. 99-139. Estados Unidos América, Elsevier; 2012d.
- BUTLER JM. *Fundamentals of forensic DNA typing*. Elsevier Academic Press; 2010.
- BUTLER JM. Short tandem repeat typing technologies used in human identity testing, *Biotechniques*; 2007, 43 (4) ii – v.
- BUTLER JM. Genetics And Genomics of Core Short Tandem Repeat Loci Used in Human Identity testing. *Bio Techniques*; 2006, 43 (4).
- BUTLER JM. *Forensic DNA Typing: Biology, Technology, and Genetics of STR markers*. 2 ed. Press EA; 2005.
- BUTLER JM. Recent Developments in Y-Short Tandem Repeat and Y-Single Nucleotide Polymorphism Analysis. *Forensic Sci Ver*; 2003, 15: 91-111.
- CARRACEDO A, SOBRINO B, LAREU MV. *Forensic DNA typing technologies: a review*, U K: Elsevier; 2007.
- CARVALHO SPM. *Avaliação da qualidade do DNA obtido de saliva humana armazenada e sua aplicabilidade na identificação forense em odontologia legam*. [Dissertação] Bauru: Faculdade de Odontologia de Bauru da Universidade de São Paulo; 2009.
- CORDEIRO O, SOUZA BR, CORREA H, *et al*. A review of psychiatric genetics research in the Brazilian population, *Ver, Bras, Psiquiatr*, 2009; 31 (2): 154-62.
- DAWID AP, MORTERA J, PASCALI V L. *et al*. Probabilistic Expert Systems for Forensic Inference from Genetic Markers, *Scandinavian Journal of Statistics*, 2002; 29 (4): 577–595.
- DELLALIBERA E, HAVRO MLB, SOUZA M, *et al*. Genetic analysis of 13 STR loci in the population from the State of Pernambuco, northeast Brazil, *Forensic Science International*, 2004; 146: 57–59.
- DOLINSKY LC, PEREIRA LMCV, *et al*. DNA Forense Artigo de Revisão, *Saúde e Meio Ambiente em Revista*, Universidade Unigranrio, 2007; 2 (2): 11-22.
- EGELAND T, KULLE B, ANDREASSEN R. Essen-Moller and Identification based on DNA, *Chance*, 2006, 19 (02): 27-31.
- FERREIRA FL, LEAL-MESQUITA ERL, SANTOS SEB, *et al*. Genetic characterization of the population of São Luís, MA, Brazil, *Genetics and Molecular Biology*, 2005; 28 (1): 22-31.

- FERREIRA ME, GRATTAPAGLIA D. Introdução ao uso de marcadores RAPD e RFLP em análise genética, Brasília: Embrapa; 1995.
- FIGUEIREDO HP. Avaliação das frequências alélicas de 15 loci STRS na população de pessoas nascidas no Estado do Mato Grosso do Sul. [Dissertação] Campo Grande: Universidade Católica de Don Bosco; 2009.
- FIMMERS R, HENKE L, HENKE J, *et al.* How to deal with mutations in DNA testing, In: Rittner C, Schneider PM, eds, *Advances in Forensic Haemogenetics 4*, Berlin: Springer Verlag, 1992; 285-7.
- FRANCEZ ABC, RODRIGUES EMR, FRAZÃO GF, *et al.* Allelic frequencies and statistical data obtained from 12 CODIS STR loci in an admixed population of the Brazilian Amazon, *Genetics and Molecular Biology*, 2011; 34(1): 35-39.
- GÓES AC, DA SILVA DA, FONSECA GIL EH, *et al.* Allelic frequencies data and statistic parameters for 16 STR marcadores-D19S433, D2S1338, CSF1PO, D16S539, D7S820, D21S11, D18S51, D13S317, D5S818, FGA, Penta E, TH01, vWA, D8S1179, TPOX, D3S1358-in the Rio de Janeiro population, Brazil, *Forensic Sci, Int*, 2004; 140: 131-132.
- GUERRA R. Crime sem corpo, 8 ed, São Paulo: Revista Ciência Criminal; 2007.
- HAMMOND HÁ, LI JIN, ZHONG Y, THOMAS CASKEY C, *et al.* Evaluation of 13 short tandem repeat loci for use in personal identification applications, *Am J Hum Genet*, 1994; 55: 175-189.
- IBGE – Instituto Brasileiro de Geografia e Estatística: Censo Demográfico 2010. Características Gerais da População e Instrução, Rio Grande do Sul, Rio de Janeiro, 2010.
- Jacewicz R, Berent J, Prośniak A, Dobosz T, Kowalczyk E, Szram S, Paternity determination of the deceased defendant in STR against RFLP analysis, *International Congress Series*, 2004; 1261: 523-525.
- JEFFREYS A. Genetic fingerprinting, *Nature Medicine*, out 2005; 11 (10): 14-18.
- JEFFREYS AJ, WILSON V, THEIN SL. Hypervariable 'minisatellite' regions in human DNA *Nature*, 1985a; 314 6006: 67-73.
- JEFFREYS AJ, WILSON V. Thein SL, Individual-specific 'fingerprints' of human DNA, *Nature*, 1985b, 316: 6023-76.
- JOBIM LF, COSTA LR, DA SILVA M. Identificação Humana- Identificação Médico Legal Perícias Odontológicas Identificação Humana pelo DNA. Campinas, SP: Millennium; 2012.

- JOBIM LF, JOBIM MR, BRENNER C. Identificação humana pelo DNA. Identificação Humana, Sagra Luzzatto. Porto Alegre; 1999; 239-303.
- LEITE FPN. Análise da estrutura genética da população do Rio Grande do Sul através de microssatélites autossômicos e de cromossomos sexuais [Dissertação]. Campo Grande: Universidade Federal do Rio Grande do Sul; 2006a.
- LEITE FPN. *The genetic structure of southern Brazil revealed by autosomal markers* Porto Alegre, RS, Brazil, 2006b.
- MARTINS JA. Estudo de frequências alélicas de STRs do cromossomo X na população brasileira de Araraquara – SP [dissertação]. Universidade Estadual Paulista, Araraquara; 2008.
- MARTINS JA, PANETO GG, PEREIRA GA. Alvarenga VLS, Cicarelli RMB, Genetic population data from Araraquara region (SP State, Brazil) using Power Plex® 16 Systems, International Congress Series, 2006; 1288: 424–426.
- MONTE SJH, MOITA N, JOSÉ M. *et al.* Polimorfismo do sistema HLA em uma amostra de mestiços da população de Teresina, Piauí, Revista da Associação Médica Brasileira, October 2004, 50(4): 422-426.
- MORET T. Identificação Humana: Uma proposta metodológica para obtenção de DNA de ossos e implementação de bancos de dados de frequência alélica de STRs autossômicos na população de Santa Catarina [Dissertação]. Universidade Federal de Santa Catarina, Florianópolis; 2009.
- MULLIS KB, FALOONA FA. Specific synthesis of DNA in vitro via a polymerasecatalyzed chain reaction. *Methods Enzymol.* 1987; 155: 335-350.
- NEI M. *Molecular Evolutionary Genetics*, Universidade de Columbia; New York; 1973.
- OTTO PG, OTTO PA, FROTA-PESSOA O. *Genética Humana e Clínica* . 2. ed. São Paulo, ROCA; 2004.
- PEDROSA MAF. Composição genética de quatro populações remanescentes de quilombos do Brasil com base em microssatélites e marcadores ancestrais [Dissertação]. Universidade de Brasília, Brasília; 2006.
- PENA SDJ, PIETRO GD, FUCHSHUBER-MORAES M, *et al.* The Genomic Ancestry of Individuals from Different Geographical Regions of Brazil Is More Uniform Than Expected, 2011; PloS ONE: 6 (2).

- PENACINO G A, SALA D, CORACH. Are DNA tests infallible? Published by Elsevier Science B,V. S0531-5131, 2003; 02: 00558-7.
- PIMENTEL BJ, AZEVEDO DA, SILVA LAF. Population Genetics of Eleven STR Loci in the State of Sergipe, Northeastern Brazil, Journal of Forensic Sciences, 2004; 49: 1-2.
- POETSCH M, LÜDCKE C, REPENNING A, *et al.* The problem of single parent/child paternity analysis-practical results involving 336 children and 348 unrelated men; Forensic Sci Int; 2006, 2,159(2-3): 98-103.
- POETSCH M, PETERSMANN H, REPENNING A. Development of two pentaplex systems with X-chromosomal STR loci and their allele frequencies in a northeast German population, Forensic Sci Int, 2005; 155: 71-76.
- PRIMORAC D, SCHARNFELD MS. Application of forensisc DNA testing in the legal system, Croatian Med J, 2000; 41 (1): 32-46.
- RENO J, MARCUS D, LEARY ML. Samuels JE, THE FUTURE OF FORENSIC DNA TESTING: Predictions of the research and development working group, Washington: National Commission of the Future of DNA Evidence, Nov 2000; 183697. Disponível em: <https://www.ncjrs.gov/pdffiles1/nij/183697.pdf>
- RITA THS. Frequências Alélicas de 15 STRS autossômicos em uma amostra populacional do Distrito Federal do Brasil – Um território que surgiu do nada. [Dissertação] Brasília: Universidade de Brasília; 2013.
- RODRIGUES EMR, PALHA TJ, SANTOS SEB. Allele frequencies data and statistic parameters for 13 STR loci in a population of the Brazilian Amazon Region, Forensic Science International, 2007; 168: 244-247.
- RODRIGUES EMR, PALHA TJBF, SANTOS SEB. Allele frequências data and statistic parameters of 13 STRs loci in a population of the Brazilian Amazon Region, Genetics molecular Biology, 2006; 28 (1): 22-21.
- ROSENBERG NA, PRITCHARD JK, WEBER JL, *et al.* Genetic struture of human population, 2002; 298:2381-2385.
- RUMJANEK FD, RINZLER CMC. Os exames de DNA nos tribunais, Ciências Hoje, 2001; 29 (169): 24 – 30.
- SANTOS MV, ANJOS MJ, ANDRADE L. Population genetic data for the STR marcadores using the AmpFISTR® Identifiler™ Kit in Bahia, Brazil, Elsevier B,V, 2004; 03: 01628-5.

SEYFERTH G. O futuro era branco, Revista de História da Biblioteca Nacional [Internet]. jun 2011 6 [acesso em 2 , dez 2014]. Disponível em:

<http://www.revistadehistoria.com.br/secao/perspectiva/o-futuro-era-branco>

SILVA LAF, PASSOS NS. DNA Forense – Coleta de Amostras Biológicas em Locais de Crimes para Estudo de DNA, Maceió: UFAL, 2006, 84.

SILVA LAF, PIMENTEL BJ, AZEVEDO DA, *et al.*. Allele frequencies of nine STR loci—D16S539, D7S820, D13S317, CSF1PO, TPOX, TH01, F13A01, FESFPS and vWA—in the population from Alagoas, Northeastern Brazil, 2002, Forensic Science International, 2002; 130:187–188.

SILVA RHA, MUSSE JO, MELANI RFH, *et al.* Human bite mark identification and DNA technology in forensic dentistry, Braz J Oral Sci, Oct-Dec, 2006; 5:1193-7.

SOUZA JC, QUEIROZ PR. Aplicações dos marcadores moleculares baseados em DNA nas questões de identificação humana no âmbito cível e forense. Ver Mostra de Produção Científica da Pós-Graduação Lato Sensu da PUC Goiás, 2010; 1: 26.

THOMPSON & THOMPSON, Genética Médica, 6 ed. Rio de Janeiro, RJ Guanabara Koogan S.A; 2002.

THOMPSON & THOMPSON. Genética Médica. 7 ed. Rio de Janeiro, RJ Guanabara Koogan S.A; 2007.

TILLMAR A. Populations and Statistics in Forensic Genetics, Linköping University, Faculty of Health Sciences; 2010.

URQUHART A, KIMPTON C P, DOWNES TJ, *et al.* Variation in short tandem repeat sequences--a survey of twelve microsatellite loci for use as forensic identification markers, Int J Legal Med, 1994; 107 (1): 13-20.

VIEIRA TC, SILVA DM, GIGONZAC MAD, *et al.* Allelic frequencies and statistical data obtained from 15 STR marcadores in a population of the Goiás State, Genet, Genet Mol, Res, 2013; 12 (1): 23-27.

WATSON JDE, CRICK FHC. Molecular Structure of Nucleic Acids, Nature, 1953; 171:737–738.

ANEXOS

Tabela 11. Frequências alélicas do *locus* D18S51.

D18S51	BR	GO	BA	SC	SP	RS	PA	MS	RJ
04	0,00001								
05	0,00001								
06	0,00001								
07	0,00002								
08	0,00009	0,002							
09	0,00120	0,015		0,001			0,002		0,001
10	0,00988	0,013	0,013	0,007	0,009		0,009	0,016	0,016
10,2	0,00056		0,017						0,001
11	0,01077	0,033	0,137	0,011		0,016	0,008	0,004	0,005
12	0,12112	0,132	0,137	0,121	0,109	0,120	0,120	0,099	0,121
12,2	0,00004								
12,3	0,00000								
13	0,10853	0,115	0,114	0,101	0,145	0,106	0,115	0,126	0,116
13,2	0,00476								
13,3	0,000004								
14	0,14722	0,135	0,120	0,191	0,145	0,199	0,159	0,177	0,115
14,2	0,00011								
14,3	0,00001								
14,5	0,00000								
15	0,15221	0,136	0,170	0,155	0,155	0,163	0,139	0,144	0,148
15,2	0,00012		0,003			0,002		0,002	
16	0,14449	0,140	0,150	0,125	0,155	0,120	0,169	0,144	0,152
16,2	0,00002								
16,5	0,00000								
17	0,12192	0,117	0,103	0,133	0,145	0,108	0,122	0,102	0,118
18	0,07756	0,077	0,057	0,053	0,018	0,068	0,088	0,093	0,083
18,2	0,00002					0,002			
19	0,04862	0,035	0,047	0,053	0,082	0,063	0,037	0,049	0,058
19,2	0,00002								
20	0,02909	0,024	0,053	0,036	0,027	0,019	0,025	0,016	0,033
20,2	0,00002								
21	0,01304	0,011	0,013	0,006		0,005	0,005	0,015	0,013
21,2	0,000004								
22	0,00553	0,002	0,003	0,006		0,006	0,003	0,009	0,011
22,2						0,002			
23	0,00138			0,001		0,001		0,002	0,001
23,2	0,00000								
24	0,00075			0,001			0,002	0,002	
25	0,00037								
26	0,00019				0,009				
27	0,00029								0,001
28	0,00003								

BR: Brasil GO: Goiás BA: Bahia SC: Santa Catarina SP: São Paulo RS: Rio Grande do Sul PA: Pará MS: Mato Grosso do Sul RJ: Rio de Janeiro,

Tabela 12. Frequências alélicas do *locus* D21S11.

D21S11	BR	GO	BA	SC	SP	RS	PA	MS	RJ
12	0,000004	0,000004							
22	0,000004	0,000004							
23	0,000004	0,000004							
24	0,0001	0,0001							
24,2	0,002	0,002			0,009	0,002	0,006	0,004	0,003
25	0,001	0,001				0,002	0,003	0,002	0,001
25,2	0,001	0,001						0,002	0,001
26	0,001	0,001			0,009		0,003		
26,2	0,00003	0,00003							
27	0,030	0,030		0,037	0,036	0,037	0,031	0,022	0,025
27,2	0,0001	0,0001	0,037						
28	0,159	0,159	0,140	0,149	0,245	0,151	0,115	0,166	0,183
28,2	0,003	0,003		0,001					
29	0,208	0,208	0,250	0,231	0,145	0,217	0,191	0,215	0,160
29,2	0,005	0,005					0,001		0,003
30	0,232	0,232	0,207	0,229	0,264	0,224	0,271	0,270	0,250
30,2	0,028	0,028	0,033	0,032	0,018	0,031	0,054	0,026	0,021
31	0,077	0,077	0,060	0,071	0,064	0,070	0,078	0,049	0,065
31,2	0,090	0,090	0,107	0,097	0,118	0,110	0,080	0,082	0,108
32	0,028	0,028	0,013	0,009		0,006	0,012	0,011	0,011
32,2	0,075	0,075	0,087	0,098	0,064	0,115	0,100	0,102	0,098
33	0,012	0,012		0,002		0,003	0,003		0,001
33,2	0,028	0,028	0,047	0,032	0,027	0,024	0,029	0,031	0,037
33,3	0,00001	0,00001							
34	0,005	0,005	0,003				0,006	0,005	0,008
34,2	0,004	0,004	0,003	0,003		0,006	0,008	0,002	0,006
35	0,006	0,006	0,010	0,007		0,002	0,003	0,011	0,008
35,1	0,000004	0,000004							
35,2	0,001	0,001					0,006		
36	0,002	0,002	0,003						0,005
36,2	0,00003	0,00003							
37	0,001	0,001							
38	0,0003	0,0003		0,001					
39	0,00003	0,00003							

BR: Brasil **GO:** Goiás **BA:** Bahia **SC:** Santa Catarina **SP:** São Paulo **RS:** Rio Grande do Sul **PA:** Pará **MS:** Mato Grosso do Sul **RJ:** Rio de Janeiro,

Tabela 13. Frequências alélicas do *locus* D3S1358.

D3S1358	BR	GO	BA	SC	SP	RS	PA	MS	RJ
05		0,001							
06	0,00002								
07	0,00003	0,001							
08	0,00002								
09	0,0001								
10	0,0001	0,001							
11	0,001	0,003		0,003		0,003			0,001
12	0,002	0,003				0,002	0,002	0,002	0,001
13	0,004	0,004	0,007	0,002		0,005	0,003		0,001
14	0,093	0,097	0,103	0,104	0,082	0,095	0,059	0,093	0,100
15	0,291	0,335	0,330	0,279	0,227	0,287	0,322	0,332	0,293
15,2	0,0001								
16	0,270	0,240	0,270	0,248	0,282	0,275	0,263	0,232	0,262
16,2	0,00001								
16,3	0,000004								
16,5	0,000004								
17	0,203	0,202	0,197	0,225	0,255	0,187	0,157	0,215	0,205
17,3	0,00002								
18	0,124	0,108	0,108	0,126	0,145	0,136	0,134	0,109	0,126
19	0,010	0,006	0,003	0,009		0,008	0,023	0,016	0,008
20	0,001			0,002		0,002	0,012		
21	0,00003								

BR: Brasil **GO:** Goiás **BA:** Bahia **SC:** Santa Catarina **SP:** São Paulo **RS:** Rio Grande do Sul **PA:** Pará **MS:** Mato Grosso do Sul **RJ:** Rio de Janeiro,

Tabela 14. Frequências alélicas do *locus* TH01.

TH01	BR	GO	BA	SC	SP	RS	PA	MS	RJ
04	0,0001								0,001
05	0,002	0,001				0,002	0,012	0,002	
06	0,211	0,223	0,220	0,205	0,202	0,233	0,226	0,226	0,198
07	0,240	0,233	0,283	0,194	0,192	0,236	0,226	0,274	0,258
08	0,149	0,147	0,150	0,139	0,183	0,134	0,120	0,117	0,187
8,2	0,00001								
09	0,163	0,169	0,127	0,143	0,163	0,129	0,165	0,148	0,143
9,2	0,000004								
9,3	0,226	0,208	0,207	0,305	0,250	0,249	0,211	0,230	0,208
10	0,008	0,018	0,013	0,014	0,010	0,017	0,005	0,004	0,003
11	0,0005								
12	0,00001								
13	0,00002								
13,3	0,00000								
14	0,00001								
15	0,000004								
16	0,000004								
18	0,00001								

BR: Brasil **GO:** Goiás **BA:** Bahia **SC:** Santa Catarina **SP:** São Paulo **RS:** Rio Grande do Sul **PA:** Pará **MS:** Mato Grosso do Sul **RJ:** Rio de Janeiro.

Tabela 15. Frequências alélicas do *locus* CSF1PO.

CSF1PO	BR	GO	BA	SC	SP	RS	PA	MS	RJ
05	0,00002								
06	0,001			0,001					0,001
07	0,013	0,020	0,013	0,006	0,029	0,008	0,011	0,009	0,026
08	0,020	0,012	0,020	0,010		0,017	0,017	0,020	0,016
09	0,033	0,028	0,020	0,019	0,029	0,019	0,022	0,015	0,020
10	0,262	0,250	0,313	0,293	0,221	0,301	0,274	0,259	0,272
11	0,294	0,265	0,284	0,309	0,317	0,285	0,295	0,281	0,278
12	0,302	0,280	0,277	0,298	0,327	0,301	0,303	0,350	0,323
13	0,064	0,067	0,063	0,052	0,077	0,060	0,068	0,053	0,053
14	0,010	0,033	0,010	0,012		0,005	0,009	0,013	0,006
15	0,001	0,019				0,004	0,002		0,001
16	0,00004	0,009							
17		0,009							
18		0,005							
19		0,001							
20		0,001							
21		0,001							
22	0,000004								
23	0,000004								

BR: Brasil **GO:** Goiás **BA:** Bahia **SC:** Santa Catarina **SP:** São Paulo **RS:** Rio Grande do Sul **PA:** Pará **MS:** Mato Grosso do Sul **RJ:** Rio de Janeiro,

Tabela 16. Frequências alélicas do loci D13S317.

D13S317	BR	GO	BA	SC	SP	RS	PA	MS	RJ
04	0,000								
05	0,000								
06	0,000								
07	0,001	0,001		0,003			0,003		
08	0,100	0,083	0,057	0,132	0,136	0,107	0,100	0,108	0,076
09	0,082	0,093	0,087	0,087	0,082	0,094	0,115	0,104	0,078
10	0,053	0,052	0,047	0,061	0,045	0,046	0,042	0,064	0,065
11	0,295	0,297	0,336	0,269	0,300	0,307	0,222	0,265	0,298
12	0,293	0,274	0,296	0,287	0,255	0,263	0,314	0,290	0,322
13	0,126	0,123	0,127	0,104	0,136	0,110	0,152	0,126	0,128
14	0,048	0,061	0,040	0,056	0,045	0,073	0,049	0,044	0,031
14,2			0,003						
15	0,001	0,007		0,001			0,003		
16	0,000	0,003							
17		0,003	0,007						
18		0,003							

BR: Brasil **GO:** Goiás **BA:** Bahia **SC:** Santa Catarina **SP:** São Paulo **RS:** Rio Grande do Sul **PA:** Pará **MS:** Mato Grosso do Sul **RJ:** Rio de Janeiro.

Tabela 17. Frequências alélicas do loci D16S359.

D16S539	BR	GO	BA	SC	SP	RS	PA	MS	RJ
05	0,001	0,001		0,001			0,002	0,004	
06	0,0003			0,001					
07	0,0004								
08	0,024	0,019	0,033	0,016	0,029	0,014	0,026	0,018	0,025
09	0,160	0,170	0,140	0,137	0,125	0,152	0,177	0,155	0,198
9,2	0,00001								
10	0,088	0,082	0,100	0,090	0,125	0,089	0,099	0,086	0,068
10,2	0,000004								
11	0,289	0,300	0,287	0,292	0,317	0,289	0,286	0,307	0,282
12	0,256	0,270	0,277	0,290	0,279	0,260	0,239	0,237	0,235
13	0,155	0,130	0,150	0,155	0,096	0,180	0,151	0,170	0,161
14	0,025	0,030	0,010	0,015	0,029	0,016	0,019	0,024	0,030
15	0,002	0,003	0,003	0,002			0,003		
16	0,0001	0,003							
17	0,00002	0,001							
18		0,002							

BR: Brasil **GO:** Goiás **BA:** Bahia **SC:** Santa Catarina **SP:** São Paulo **RS:** Rio Grande do Sul **PA:** Pará **MS:** Mato Grosso do Sul **RJ:** Rio de Janeiro.

Tabela 18. Frequências alélicas do *locus* D5S818.

D5S818	BR	GO	BA	SC	SP	RS	PA	MS	RJ
05	0,000004								
06	0,000006								0,001
07	0,017	0,022	0,017	0,021	0,009	0,033	0,008	0,031	0,015
08	0,018	0,018	0,023	0,003	0,018	0,008	0,025	0,016	0,015
09	0,038	0,039	0,020	0,034	0,036	0,057	0,059	0,038	0,026
10	0,062	0,065	0,063	0,067	0,055	0,051	0,065	0,064	0,052
11	0,329	0,329	0,347	0,319	0,427	0,335	0,285	0,353	0,328
12	0,346	0,345	0,360	0,371	0,264	0,338	0,400	0,330	0,370
13	0,174	0,158	0,160	0,168	0,173	0,163	0,148	0,164	0,178
14	0,013	0,015	0,017	0,013	0,018	0,013	0,011		0,013
15	0,002	0,003	0,003	0,003		0,002	0,002	0,004	
16	0,0002	0,003							
17	0,0001	0,001							
18	0,00002	0,002							
19	0,000004								

BR: Brasil **GO:** Goiás **BA:** Bahia **SC:** Santa Catarina **SP:** São Paulo **RS:** Rio Grande do Sul **PA:** Pará **MS:** Mato Grosso do Sul **RJ:** Rio de Janeiro.

Tabela 19. Frequências alélicas do *locus* D7S820.

D7S820	BR	GO	BA	SC	SP	RS	PA	MS	RJ
05	0,00003								
06	0,0003	0,002							
6,2	0,00004								
07	0,015	0,022	0,007	0,017	0,045	0,014	0,002	0,018	0,020
08	0,157	0,163	0,136	0,124	0,191	0,129	0,132	0,141	0,187
09	0,121	0,126	0,117	0,128	0,136	0,107	0,117	0,115	0,096
9,3	0,00001								
10	0,275	0,249	0,320	0,272	0,218	0,241	0,297	0,268	0,298
10,2	0,000004								
11	0,236	0,227	0,260	0,253	0,245	0,257	0,212	0,254	0,200
12	0,161	0,174	0,120	0,161	0,145	0,216	0,117	0,161	0,156
13	0,030	0,033	0,040	0,039	0,018	0,033	0,095	0,044	0,036
14	0,004	0,005		0,007		0,003	0,009		0,005
15	0,0001						0,002		
16	0,0002								
17	0,000004								

BR: Brasil **GO:** Goiás **BA:** Bahia **SC:** Santa Catarina **SP:** São Paulo **RS:** Rio Grande do Sul **PA:** Pará **MS:** Mato Grosso do Sul **RJ:** Rio de Janeiro.

Tabela 10. Frequências alélicas do *locus* D8S1179.

D8S1179	BR	GO	BA	SC	SP	RS	PA	MS	RJ
03	0,000004								
04	0,000001								
05	0,00001								
06	0,0001								
07	0,0003								
08	0,008	0,005	0,007	0,008		0,016	0,007	0,005	0,005
09	0,009	0,009	0,010	0,012		0,008	0,012	0,000	0,003
10	0,068	0,066	0,060	0,083	0,109	0,082	0,056	0,062	0,051
11	0,077	0,068	0,057	0,077	0,091	0,090	0,068	0,075	0,065
12	0,130	0,121	0,107	0,150	0,136	0,144	0,143	0,130	0,130
13	0,279	0,275	0,286	0,291	0,309	0,261	0,320	0,301	0,290
14	0,251	0,274	0,257	0,235	0,245	0,239	0,244	0,268	0,257
15	0,138	0,143	0,173	0,116	0,100	0,117	0,109	0,131	0,158
16	0,034	0,030	0,033	0,019	0,009	0,033	0,039	0,013	0,031
17	0,004	0,005	0,010	0,009		0,006	0,002	0,005	0,008
18	0,001	0,002				0,004	0,001		
19	0,0001								
20	0,00001								
21	0,000004								
22	0,00001								

BR: Brasil **GO:** Goiás **BA:** Bahia **SC:** Santa Catarina **SP:** São Paulo **RS:** Rio Grande do Sul **PA:** Pará **MS:** Mato Grosso do Sul **RJ:** Rio de Janeiro.

Tabela 11. Frequências alélicas do *locus* FGA.

FGA	BR	GO	BA	SC	SP	RS	PA	MS	RJ
07	0,00004								
08	0,000001								
09	0,00002								
10	0,000002								
11	0,00002	0,003							
12	0,00002	0,001							
13	0,00002								
14	0,00002								
15	0,00003	0,002							
16	0,0005	0,003							0,001
17	0,0015	0,002		0,003					0,001
17,2	0,00003								
18	0,0093	0,016	0,017	0,008	0,009	0,009	0,010	0,009	0,013
18,2	0,0024			0,006		0,003	0,005		
18,3	0,000004								
19	0,0754	0,072	0,087	0,079	0,047	0,081	0,099	0,077	0,048
19,2	0,0006							0,004	
20	0,1139	0,134	0,090	0,116	0,132	0,113	0,118	0,091	0,111
20,2	0,0002				0,019				
21	0,1511	0,142	0,177	0,161	0,236	0,141	0,140	0,128	0,151
21,2	0,0013			0,003	0,009			0,007	
22	0,1617	0,170	0,177	0,165	0,151	0,140	0,135	0,141	0,197
22,2	0,0024		0,003	0,003		0,005	0,002	0,007	
23	0,1493	0,146	0,123	0,155	0,170	0,152	0,118	0,157	0,143
23,2	0,0011		0,003	0,001				0,004	
24	0,1509	0,137	0,157	0,158	0,104	0,149	0,172	0,161	0,133
24,2	0,0003						0,111	0,119	
25	0,1034	0,089	0,090	0,087	0,075	0,121	0,070		0,126
25,2	0,0002							0,069	
26	0,0474	0,062	0,033	0,045	0,038	0,065	0,010		0,033
26,2	0,00003							0,020	
27	0,0160	0,018	0,033	0,004	0,009	0,013	0,005		0,028
28	0,0051	0,004	0,007	0,002		0,006	0,002	0,005	0,003
28,2	0,00001								
29	0,0018	0,001		0,001			0,002		0,003
29,2	0,00002								
30	0,0007							0,002	0,001
30,2	0,000173								
31	0,0001								
31,2	0,0011			0,001		0,002			
32	0,0001								
32,2	0,0002								
33	0,00004								
33,2	0,00002								
39,2	0,000004								
40	0,00001								
41,2	0,000004								
42	0,00001								
42,2	0,00005								
43	0,000004								
43,2	0,0006								
44,2	0,0003								
45,2	0,0003								
46,2	0,0003								
47	0,000004								
47,2	0,00004								
48,2	0,00002								

BR: Brasil **GO:** Goiás **BA:** Bahia **SC:** Santa Catarina **SP:** São Paulo **RS:** Rio Grande do Sul **PA:** Pará **MS:** Mato Grosso do Sul **RJ:** Rio de Janeiro.

Tabela 12. Frequências alélicas do *locus* TPOX.

TPOX	BR	GO	BA	SC	SP	RS	PA	MS	RJ
04	0,000004								
05	0,001			0,002		0,002	0,002		
06	0,018	0,013	0,023	0,011	0,020	0,013	0,014	0,007	0,025
07	0,008	0,001	0,013	0,003	0,030	0,005	0,003	0,002	0,003
08	0,456	0,430	0,460	0,496	0,530	0,476	0,410	0,440	0,465
09	0,124	0,120	0,147	0,098	0,080	0,087	0,120	0,109	0,148
10	0,067	0,076	0,077	0,069	0,060	0,048	0,066	0,068	0,056
11	0,276	0,290	0,250	0,273	0,220	0,297	0,294	0,316	0,262
12	0,048	0,056	0,030	0,046	0,230	0,070	0,092	0,055	0,038
13	0,002	0,008		0,001		0,002	0,003	0,004	0,001
14	0,0001								
15	0,0001								
16	0,00003								
17	0,00001								
18	0,00001								
21	0,0001	0,001							
22		0,001							
24		0,002							

BR: Brasil **GO:** Goiás **BA:** Bahia **SC:** Santa Catarina **SP:** São Paulo **RS:** Rio Grande do Sul **PA:** Pará **MS:** Mato Grosso do Sul **RJ:** Rio de Janeiro.

Tabela 13. Frequências alélicas do *locus* vWA.

vWA	BR	GO	BA	SC	SP	RS	PA	MS	RJ
08	0,00001	0,002							
10	0,00004								
11	0,004	0,002	0,010	0,003	0,009		0,008	0,004	0,008
12	0,001	0,003							
13	0,006	0,010	0,010	0,002		0,002	0,015	0,007	0,006
14	0,084	0,092	0,066	0,112	0,109	0,085	0,068	0,073	0,083
15	0,138	0,139	0,147	0,115	0,082	0,096	0,134	0,128	0,162
15,2	0,000004								
16	0,260	0,260	0,240	0,243	0,218	0,258	0,263	0,301	0,243
16,2	0,000004								
17	0,255	0,236	0,270	0,229	0,227	0,280	0,288	0,263	0,255
17,2	0,000								
17,3	0,00001								
18	0,170	0,175	0,133	0,211	0,227	0,188	0,148	0,164	0,165
19	0,066	0,670	0,100	0,067	0,109	0,076	0,057	0,053	0,058
20	0,014	0,013	0,017	0,013	0,018	0,013	0,020	0,005	0,015
21	0,002	0,002	0,007	0,003		0,002		0,002	0,003
22	0,0002								
23	0,000004								

BR: Brasil **GO:** Goiás **BA:** Bahia **SC:** Santa Catarina **SP:** São Paulo **RS:** Rio Grande do Sul **PA:** Pará **MS:** Mato Grosso do Sul **RJ:** Rio de Janeiro.

Tabela 20. Frequências alélicas de 13 marcadores STRs (CODIS) em 102.807 indivíduos da população Brasileira

Alelos	D18S51 (N 102.807)	D21S11 (N 102.807)	D3S1358 (N 102.807)	TH01 (N 102.807)	CSF1PO (N 102.807)	D13S317 (N 102.807)	D16S539 (N 102.807)	D5S818 (N 102.807)	D7S820 (N 102.807)	D8S1179 (N 102.807)	FGA (N 102.807)	TPOX (N 102.807)	Vwa (N 102.807)
3,0													
4,0													
5,0							0,001					0,001	
6,0				0,210								0,018	
6,2													
7,0				0,240	0,013	0,001		0,018	0,015			0,008	
8,0				0,250	0,020	0,100	0,024	0,018	0,157	0,008		0,456	
9,0	0,001			0,160	0,033	0,082	0,160	0,038	0,120	0,009		0,124	
9,2													
9,3				0,230									
10,0	0,010			0,010	0,263	0,053	0,088	0,062	0,275	0,068		0,067	
10,2	0,001												
11,0	0,011		0,001		0,293	0,295	0,289	0,329	0,236	0,077		0,276	0,004
12,0	0,121		0,002		0,302	0,293	0,256	0,347	0,161	0,130		0,048	0,001
12,2													
12,3													
13,0	0,109		0,004		0,064	0,126	0,154	0,174	0,031	0,279		0,002	0,006
13,2	0,005												
13,3													
14,0	0,147		0,093		0,010	0,048	0,025	0,013	0,004	0,251			0,085
14,2													
14,3													
14,5													
15,0	0,152		0,292		0,001	0,002	0,002	0,002		0,138			0,138

Alelos	D18S51 (N 102.807)	D21S11 (N 102.807)	D3S1358 (N 102.807)	TH01 (N 102.807)	CSF1PO (N 102.807)	D13S317 (N 102.807)	D16S539 (N 102.807)	D5S818 (N 102.807)	D7S820 (N 102.807)	D8S1179 (N 102.807)	FGA (N 102.807)	TPOX (N 102.807)	Vwa (N 102.807)
15,2													
16,0	0,144		0,270							0,033	0,001		0,260
16,2													
16,3													
17,0	0,122		0,203							0,004	0,001		0,255
17,2													
17,3	0,077									0,001	0,009		0,170
18,0			0,124										
18,2	0,000										0,002		
18,3											0,075		
19,0	0,049		0,010								0,001		0,066
19,2											0,114		
20,0	0,029		0,001								0,000		0,014
20,2											0,151		
21,0	0,013										0,001		0,002
21,2											0,162		
22,0	0,006										0,002		
22,2											0,149		
23,0	0,001										0,001		
23,2											0,151		
24,0	0,001	0,000									0,001		
24,2		0,002									0,103		
25,0		0,001											
25,2		0,001									0,047		
26,0		0,001											
26,2											0,016		

