



UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE BIOCIÊNCIAS
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS BIOLÓGICAS

ABIGAIL MARCELINO DOS SANTOS SILVA

**BIGASYSTEM: UMA FERRAMENTA BIOINFORMÁTICA PARA IDENTIFICAÇÃO
GENÉTICA HUMANA E BANCO DE DADOS**

Recife
2021

ABIGAIL MARCELINO DOS SANTOS SILVA

**BIGASYSTEM: UMA FERRAMENTA BIOINFORMÁTICA PARA IDENTIFICAÇÃO
GENÉTICA HUMANA E BANCO DE DADOS**

Tese apresentada ao Programa de Pós-Graduação em Ciências Biológicas da Universidade Federal de Pernambuco, como requisito parcial para obtenção do título de Doutora em Ciências Biológicas.

Área de Concentração Biotecnologia

Orientador: Prof. Dr. Valdir de Queiroz Balbino

Coorientador: Prof. Dr. Sérgio de Sá Leitão Paiva Junior

Recife

2021

Catálogo na Fonte:
Elaine C Barroso, CRB-4/1728

Silva, Abigail Marcelino dos Santos

BigASystem: uma ferramenta bioinformática para identificação genética humana e banco de dados / Abigail Marcelino dos Santos Silva – 2021.

157 f. : il., fig., tab.

Orientador: Valdir de Queiroz Balbino

Coorientador: Sérgio de Sá Leitão Paiva Junior

Tese (doutorado) – Universidade Federal de Pernambuco. Centro de Biociências. Programa de Pós-graduação em Ciências Biológicas, Recife, 2021.

Inclui referências e anexos.

1. Genética forense 2. Bioinformática 3. Banco de dados I. Balbino, Valdir Queiroz (orient.). II. Paiva Junior, Sérgio de Sá Leitão Paiva (coorient.) III. Título

576.5

CDD (22.ed)

UFPE/CB-2021-277

ABIGAIL MARCELINO DOS SANTOS SILVA

**BIGASYSTEM: FERRAMENTA BIOINFORMÁTICA PARA IDENTIFICAÇÃO
GENÉTICA HUMANA E BANCO DE DADOS**

Tese apresentada ao Programa de Pós-Graduação em Ciências Biológicas da Universidade Federal de Pernambuco, como requisito parcial para obtenção do título de doutora em Ciências Biológicas.

Aprovada em: 21/07/2021.

BANCA EXAMINADORA

Prof. Dr. Valdir de Queiroz Balbino
Universidade Federal de Pernambuco

Profa. Dr^a. Ana Maria Benko Iseppon
Universidade Federal de Pernambuco

Prof. Dr. Marco Jacinto Katzenberger Baptista Novo
Universidade Federal de Pernambuco

Prof. Dr. Martín Alejandro Montes
Universidade Federal Rural de Pernambuco

Prof. Dr. Rodrigo Soares de Moura Neto
Universidade Federal do Rio de Janeiro

Dedico este trabalho aos meus pais, Aluizio Francisco da Silva e Josefa Marcelino dos Santos Silva, que um dia sonharam e hoje compartilham este importante momento comigo. À minha irmã, Priscila Marcelino dos Santos Silva, que torce sempre pelo meu sucesso. A todos professores que tive ao longo da minha jornada de ensino médio, graduação e pós-graduação, e em especial aos Profs. Drs. Luiz Maurício-da-Silva (*in memoriam*) e Rosilda dos Santos Silva, com os quais dei os meus primeiros passos na minha carreira. A todos amigos e familiares que, direta ou indiretamente, sempre torceram por mim.

AGRADECIMENTOS

Em primeiro lugar, agradeço ao Dr. Valdir de Queiroz Albino, por ter sido orientador no meu doutorado. Sinto-me honrada por isto. Meu respeito e admiração pelo profissional e pessoa que é. Agradeço a orientação exemplar, baseada em alto e rigoroso nível científico, um empenho saudavelmente exigente, uma visão crítica e oportuna que têm contribuído para enriquecer todas as etapas do trabalho realizado. Agradeço também pela confiança, apoio, paciência e pela oportunidade de desenvolver este projeto. Foi uma oportunidade que precisava para me dedicar a algo que sempre desejei e me tornar a profissional que sou hoje. Considerarei o Dr. Valdir como o meu professor para sempre.

Um agradecimento em especial ao Dr. Sérgio de Sá Leitão Paiva Júnior, pela honra de tê-lo como coorientador uma pessoa e profissional extraordinários. Agradeço por sempre ter acreditado em mim, pela sua paciência, compreensão e apoio. Agradeço ainda a forma simples e cientificamente correta com que sempre abordou nossas reuniões de trabalho, sem nunca ter permitido que o desalento se instalasse. Achei que este trabalho se resumiria à construção de algoritmos. Mas o Dr. Sérgio me desafiou ainda mais, ao motivar a construção de um código a partir destes algoritmos. Como gosto de me aventurar, não seria novidade aceitar este projeto; confesso que me peguei, muitas vezes, pensando se, de fato, tinha atributos suficientes para isto. Também aproveito para dizer que sou e serei eternamente grata. Considerá-lo-ei, do mesmo modo, um professor para sempre.

À Dr^a. Ana Maria Benko Iseppon, pelos conselhos e valiosas sugestões, assim como também agradeço ao Drs. Marco Jacinto Katzenberger Baptista Novo e Drs. Martín Alejandro Montes pelas importantes contribuições para a melhoria deste documento.

Ao Dr. Rodrigo Soares de Moura Neto, assim como aos demais membros da banca examinadora citados no parágrafo anterior, pelas brilhantes considerações que auxiliaram na preparação final deste documento.

Aos Laboratório de Bioinformática e Biologia Evolutiva (LABBE) e Laboratório de Genética Molecular Humana (LGMH) da Universidade Federal de Pernambuco (UFPE), responsáveis pela prestação do Serviço de Investigação de Vínculo Genético

da UFPE, ambientes nos quais desenvolvi este trabalho, e por me fornecerem os perfis genotípicos e me facultarem o acesso ao banco de dados de frequências.

Ao Programa de Pós-Graduação em Ciência Biológicas (PPGCB) da UFPE, pela oportunidade de fazer meu doutorado. Agradeço aos funcionários, discentes, professores e coordenadores: muito obrigada pelo apoio, sugestões e disponibilidade.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela concessão da bolsa, que viabilizou a minha dedicação exclusiva para o desenvolvimento do meu projeto de doutorado.

À UFPE e a todos que a fazem, pelos excelentes professores e funcionários e por lutarem pela existência, manutenção e melhoria das pesquisas e dos programas de pós-graduação.

A todos os alunos e funcionários dos laboratórios LABBE/LGMH – UFPE, com os quais passei a maior parte do meu doutorado, pela amizade e por terem proporcionado um ambiente agradável sempre, o que é bastante importante para o bom desempenho de um trabalho. Não vou citar o nome de todos vocês para não correr o risco de esquecer o nome de alguém.

Não é possível agregar valor a Genética Forense sem a Biotecnologia
(SILVA, ABIGAIL, 2021).

RESUMO

Almejando praticidade e credibilidade em suas análises, as ferramentas bioinformáticas para análise estatística de identificação humana estão cada vez mais comuns na rotina de um geneticista forense. Porém, apesar dos notáveis avanços, ainda são insuficientes, já que não possibilitam em uma única ferramenta a identificação forense e o teste de paternidade com base nos tipos de marcadores genéticos (microssatélites autossômicos e do cromossomo Y) mais utilizados na rotina básica genética forense e a obtenção de valores de frequências alélicas e haplotípicas. Esses valores probabilísticos, alcançados a partir da validação de um banco de dados populacional, são indispensáveis na estatística forense. O *BigASystem (dataBase and human Identification Genetic Analysis System)* foi implementado com o intuito de atender estas necessidades. Para tanto, montou-se um banco de dados a partir de dados genéticos de locos microssatélites autossômicos e do cromossomo Y. Os algoritmos foram implementados a partir da linguagem de programação *Python* e posteriormente validados seguindo as normas da ISO17025 e da Sociedade Internacional de Genética Forense. O *software BigASystem* é constituído de dois módulos correspondentes a cada um dos tipos de dados genéticos. Cada módulo disponibiliza duas funções: uma voltada para validação de banco de dados populacional e outra voltada para a estatística de identificação humana. A validação de todos os módulos ocorreu conforme o esperado, tendo sido observado que os resultados gerados pelo *BigASystem* foram semelhantes àqueles obtidos com diferentes ferramentas de análise. O *BigASystem* garante as funções mais básicas para um laboratório forense de rotina e pesquisa, viabilizando o seu uso em análises de determinação de paternidade/maternidade, de identificação forense de indivíduos e de genética de populações.

Palavras-chave: Genética forense; STR; Banco de dados; *Software*.

ABSTRACT

Aiming for practicality and credibility in their analyses, bioinformatics tools for statistical analysis of human identification are increasingly common in the routine of a forensic geneticist. However, despite the remarkable advances, they are still insufficient, as they do not allow forensic identification and paternity testing based on the types of genetic markers (autosomal microsatellites and the Y chromosome) most used in basic routine forensic genetics and obtaining allelic and haplotype frequency values. These probabilistic values, reached from the validation of a population database, are indispensable in forensic statistics. The BigASystem (dataBase and human Identification Genetic Analysis System) was implemented in order to meet these needs. For this purpose, a database was set up based on genetic data from autosomal microsatellite loci and the Y chromosome. The algorithms were implemented using the Python programming language and later validated following the standards of ISO17025 and the International Society of Forensic Genetics. The BigASystem software consists of two modules corresponding to each type of genetic data. Each module provides two functions: one focused on population database validation and the other focused on human identification statistics. The validation of all modules took place as expected, and it was observed that the results generated by BigASystem were similar to those obtained with different analysis tools. BigASystem ensures the most basic functions for a routine forensic laboratory and research, enabling its use in paternity/maternity determination, forensic identification of individuals and population genetics analyses.

Keywords: Forensic genetics; STR; Database; Software

LISTA DE FIGURAS

Figura 1	Duas formas de variações existentes em DNA: (a) Polimorfismo de sequência e (b) Polimorfismo de comprimento.....	25
Figura 2	Processos envolvidos na geração de perfis de DNA após ocorrência de um caso forense.....	28
Figura 3	Fluxograma de funcionamento dos softwares de identificação humana.....	56
Figura 4	Fluxograma do funcionamento dos <i>softwares</i> utilizados na validação de banco de dados populacionais.....	58
Figura 5	Esboço de plano de validação de <i>software</i>	60

LISTA DE TABELAS

Tabela 1	Locos genéticos aut-STRs principais ou "Core STR loci" usados em teste de identificação humana.....	31
Tabela 2	Locos Y-STRs utilizados em exames genéticos forenses.....	34
Tabela 3	Escala verbal associada a valores Razão de Verossimilhança (RV) de testes de identificação forense por dados aut-STRs.....	46
Tabela 4	Tabela de predicado verbal de Hummel, assumindo-se probabilidades <i>a priori</i> iguais.....	49

LISTA DE ABREVIATURAS E SIGLAS

$A_{i_1}, \dots, A_{i_n}$	Alelo 1, ..., n do loco i
$\bar{A}$	Poder de exclusão Médio utilizada pela AABB do inglês, <i>average probability of exclusion</i> or <i>mean power of exclusion</i>
$p_{i_1}, \dots, p_{i_n}$	Frequencia dos alelos $A_{i_1}, \dots, A_{i_n}$ respectivamente do loco i .
p_{hap}	Frequência relativa de haplótipo em uma população
$p_{gen.multiloco}$	Probabilidade de genótipo multiloco em desequilíbrio de ligação em uma população
a	Alelo obrigatório da criança C
A, B, C, D, E...	Letras maiúsculas em dados genéticos representam alelos diferentes de um loco arbitrário, enquanto que $p_A, p_B, p_C, p_E...$ correspondem as suas frequências relativas
AABB	Associação Americana de Bancos de Sangue (do inglês, <i>American Association of Blood Banks</i>)
Alu	Trecho curto de DNA originalmente caracterizado pela ação da endonuclease de restrição <i>Arthrobacter luteus</i> (Alu)
$A_M A_P$	Forma alternativa de se referir ao genótipo G_c da criança C , sendo A_M e A_P os alelos de origem materna e paterna, respectivamente
aut-STR(s)	STR(s) autossômico(s)
bp	Um par de bases nitrogenada de nucleotídeo (do inglês, <i>base pair</i>). Na genética e biologia molecular, corresponde a dois nucleotídeos opostos complementares
b_x	Representa todos os n alelos para um marcador/loco i
BigASystem	<i>dataBase and human identification genetic Analysis Systems</i>
C	Cena do crime (e.g. teste de identificação humana forense) ou criança (em teste de paternidade)
CD	Poder ou capacidade discriminatória de haplótipo (CrY)

CODIS/CODIS-FBI/FBI CODIS	Sistema de Índice Combinado de DNA – Departamento Federal de Investigação dos EUA (do inglês, <i>Combined DNA Index System- Federal Bureau of Investigation/USA</i>)
Core STR loci	Locos STRs principais do CODIS-FBI dos EUA
COTS	<i>Softwares</i> comerciais (do inglês, <i>commercial off-the-shelf</i>)
CrAut(s)	Cromossomo(s) Autossômico(s)
CrX(s)	Cromossomo(s) X(s)
CrY(s)	cromossomo(s) Y(s)
D(a)	Diversidade alélica (CrY)
D(h)	Diversidade haplotípica(CrY)
DL	Desequilíbrio de ligação
DNA	Ácido desoxirribonucleico (do inglês, <i>deoxyribonucleic acid</i>)
DVI	Identificação de vítimas de desastres (do inglês, <i>Disaster Victim Identification</i>)
<i>E</i>	Evidências genéticas
EDNAP	Grupo Europeu de Perfil de DNA (do inglês, <i>European DNA Profiling Group</i>)
EHW	Equilíbrio de Hardy-Weinberg
EL	Equilíbrio de ligação
EM	Algoritmo de maximização esperada (do inglês, <i>Expectation Maximization</i>)
EUA	Estados Unidos da América
<i>F</i> ou <i>Fst</i>	Coeficiente de consanguinidade intra-pessoa (do inglês, <i>inbreeding coeficiente</i>). Pode se referir a um <i>pedigree</i> ou a uma população.
F(h)	Frequência haplotípica(CrY)
FBI	Departamento Federal de Investigação dos Estados Unidos (do inglês, <i>Federal Bureau of Investigation/USA</i>)
FSS	Serviço de Ciência Forense do Reino Unido (do inglês, <i>U.K. Forensic Science Service</i>)
G_c	Genótipo/haplótipo da amostra encontrada na cena do crime (teste de identificação forense) ou genótipo/haplótipo da criança C (teste de paternidade)

G_k	Genótipo tipados a partir da amostra referência K
G_M	Genótipo/haplótipo da mãe da criança C
G_s	Genótipo/haplótipo da amostra do suspeito
G_{SP}	Genótipo/haplótipo do suposto pai da criança C
H_d	Hipótese de defesa do inglês, <i>defense hypotheses</i>
HE_e	Heterozigosidade esperada
HE_o	Heterozigosidade observada
HO_e	Homozigosidade esperada
HO_o	Homozigosidade observada
H_p	Hipótese de acusação (do inglês, <i>prosecution hypotheses</i>)
HW	Hardy-Weinberg)
i	Nome aleatório para se referir a um único loco. Cada loco i dos n locos em um sistema multiloco é representado por $i = \{1, 2, \dots, n\}$
I	Informações/evidências não-genéticas
ID	Identificação
IM(s)	Inconsistência(s) Mendeliana(s)
Indel	Mutação do tipo inserção-deleção (do inglês, <i>Insertion–deletion mutation</i>)
Interpol	Organização Internacional de Polícia Criminal sediada na França (do inglês, <i>The International Criminal Police Organization</i>)
IP(s)	Índice(s) de paternidade ou parentesco)
IPA	Índice de paternidade ou parentesco acumulado
IPE	Probabilidade de exclusão individual ou caso-específica (do inglês - <i>individual probability of exclusion or case-specific power of exclusion</i>). Idem $PE_{\text{caso específico}}$
IPM	Probabilidade de inclusão média
INPI	Instituto Nacional da Propriedade Industrial (Brasil)
IRB	Conselho de Revisão Institucional (do inglês, <i>Institutional Review Board</i>)
ISFG	Sociedade Internacional de Genética Forense (do inglês, <i>International Society for Forensic Genetic</i>)

ISO 17025	Organização internacional de padronização 17025 (do inglês, <i>International Organization for Standardization 17025</i>)
IUPAC	União Internacional de Química Pura e Aplicada sediada nos EUA (do inglês, <i>International Union of Pure & Applied Chemistry</i>)
K	Amostra referência ou conhecida do inglês, <i>known sample</i>). Pode ser uma referência direta (amostra do próprio suspeito/desaparecido) ou indireta (amostra de parentes do suspeito/desaparecido/criança)
LABBE	Laboratório de Bioinformática e Biologia Evolutiva da UFPE
LGMH	Laboratório de Genética Molecular Humana da UFPE
LR	Razão de verossimilhança (do inglês, <i>likelihood ratio</i>)
M	Mãe da criança C
MAF	Probabilidade alélica mínima. Também pode ser chamada de P_{min}
Mb	Mega pares de bases ou Mega bases (do inglês, <i>mega base pairs</i>). É uma unidade medida em biologia molecular equivalente à 1.000.000 bp
MRY	Porção não recombinante do cromossomo Y (do inglês, <i>non-recombining portion of the Y-chromosome</i>)
mtDNA	DNA mitocondrial (do inglês, <i>mitochondrial DNA</i>)
NCBI	Centro Nacional de Informação Biotecnológica dos EUA (do inglês, <i>National Center for Biotechnology Information</i>)
NIST	Instituto Nacional de Padrões e Tecnologia dos EUA (do inglês, <i>National Institute of Standards and Technology</i>)
NRC(I ou II)	Conselho Nacional de Pesquisa do Canadá (do inglês, <i>National Research Council of Canada</i>)
P ou Probabilidade P	Probabilidade genotípica/haplotípica multiloco estimada para uma certa população por meio de um banco de dados
$P(aa)$ ou $P(ab)$	Correspondem à probabilidade P_i , especificando o genótipo formado pelos alelos a e b do loco i .
PAR (1 e 2)	Regiões pseudo autossômica do CrY (do inglês, <i>pseudoautosomal regions</i>)
PC	Probabilidade de coincidência de haplótipo (CrY)

PCR	Reação em cadeia de polimerase (do inglês, <i>polymerase chain reaction</i>)
PD	Poder de discriminação
PE	Poder de exclusão
PE _{caso específico}	Probabilidade de exclusão caso específico
P_i	Frequência genotípica de um único loco i
PIC	Conteúdo informativo de polimorfismo (do inglês, <i>polymorphism information content</i>)
PM	Probabilidade de coincidência (ou de <i>match</i>)
P_{min}	Probabilidade alélica mínima
PP	Probabilidade de paternidade ou parentesco ou ainda probabilidade <i>a posteriori</i> : $P(H_p)$
Q	Amostra questionada
rCRS	Sequência de referência revisada de Cambridge para o mtDNA humano (do inglês, <i>revised Cambridge Reference Sequence</i>)
RMNE	Frequência populacional de pais em potencial não excluídos (do inglês, <i>random man not excluded</i>)
RV	Razão de Verossimilhança
S	Suspeito
SINE(s)	Elemento(s) curto(s) intercalado(s) (do inglês, <i>short interspersed element[s]</i>)
SNP(s)	Polimorfismo(s) de nucleotídeo único (do inglês, <i>single nucleotide polymorphism[s]</i>)
SP	Suposto pai da criança C
STR(s)	Repetições curtas em tandem ou microssatélites (do inglês, <i>short tandem repeats</i>)
SWGDM	Grupo de Trabalho Científico dos EUA sobre Métodos de Análise de DNA (do inglês, <i>U.S. Scientific Working Group on DNA Analysis Methods</i>)
x	Cada um dos n alelos do marcador i . Ou seja, $x=\{1, 2, ..., n\}$
X-STR(s)	STR(s) do cromossomo X
YHRD	Banco de dados de referência de haplótipos do cromossomo Y (do inglês, <i>Y-Chromosome Haplotype Reference Database</i>)

Y-STR(s)	STR(s) do cromossomo Y
USA	Estados Unidos da América do (United States of America)
θ	Coeficiente de coancestralidade do inglês (do inglês <i>kinship coefficient or coancestry coefficient</i>). Pode se referir a um <i>pedigree</i> ou a uma população
θ_{XY}	Coeficiente de coancestralidade (<i>pedigree</i>) entre indivíduos X e Y
μ	Taxa de mutação para um determinado loco
μ_{IJ}	Taxa de mutação específica de alelo <i>I</i> para o <i>J</i> de um determinado loco
UFPE	Universidade Federal de Pernambuco
χ^2	Qui-quadrado

SUMÁRIO

1	INTRODUÇÃO.....	21
2	OBJETIVOS.....	22
2.1	OBJETIVO GERAL.....	22
2.2	OBJETIVOS ESPECÍFICOS.....	22
3	REVISÃO BIBLIOGRÁFICA.....	23
3.1	IDENTIFICAÇÃO HUMANA.....	23
3.2	MARCADORES MOLECULARES GENÉTICOS.....	24
3.2.1	Marcadores aut-STRs.....	30
3.2.2	Marcadores Y-STRs.....	32
3.3	GENÉTICA DE POPULAÇÕES.....	35
3.3.1	Banco de dados populacional.....	35
3.3.2	Modelos nível populacional.....	36
3.3.2.1	Estimativa de frequência alélica aut-STRs e haplotípica Y-STR	36
3.3.2.2	Estimativa de probabilidade de perfil genético aut-STR.....	37
3.4	ESTATÍSTICA FORENSE.....	40
3.4.1	Estatística Bayesiana.....	41
3.4.1.1	Estatística de teste de identificação forense.....	43
3.4.1.2	Estatística de teste de paternidade/parentesco.....	46
3.4.2	Poder de exclusão caso específico.....	51
3.4.3	Validação de banco de dados populacional.....	52
3.5	BIOINFORMÁTICA FORENSE.....	54
3.5.1	<i>Softwares</i> de identificação humana e análises complementares disponíveis na comunidade científica.....	55
3.5.2	Construção e validação de <i>software</i> de cálculos bioestatísticos.....	58
3.5.3	Bioinformática e ética.....	61
4	ARTIGO 1 - BIGASYSTEM: UM SOFTWARE PARA ESTATÍSTICA DE DNA FORENSE.....	63

5	REGISTRO DE PROGRAMA DE COMPUTADOR	
	BigASystem.....	76
6	CONCLUSÕES.....	79
	REFERÊNCIAS BIBLIOGRÁFICAS.....	83
	APÊNDICE A – REPRESENTAÇÃO GRÁFICA DOS RESULTADOS DA VALIDAÇÃO DO BIGASystem CITADOS NO ARTIGO 1.....	134
	A. MÓDULO 1 (AUT-STRs) - FUNÇÃO BANCO DE DADOS.....	134
	A.1. Valores dos parâmetros populacionais e porcentagem de erro.....	134
	A.2. Valores de <i>P-values</i> de EHW.....	136
	B. MÓDULO1 (AUT-STRs) - FUNÇÃO TESTE DE PATERNIDADE.....	137
	B.1. Valores de índice de paternidade acumulada e porcentagem de erro.....	137
	C. MÓDULO 1 (AUT- STRs) - FUNÇÃO IDENTIFICAÇÃO FORENSE.....	139
	C.1. Valores de razão de verossimilhança e porcentagem de erro.....	139
	D. MÓDULO 2 (Y-STRs) - FUNÇÃO BANCO DE DADOS.....	140
	D.1. Valores de parâmetros populacionais e porcentagem de erro.....	140
	E. MÓDULO 2 (Y-STRs) - FUNÇÃO IDENTIFICAÇÃO HUMANA.....	141
	E.1. Valores de razão de verossimilhança e porcentagem de erro	141
	APÊNDICE B – ARTIGO: ANÁLISE POPULACIONAL DE 20 LOCOS STR CODIS EXPANDIDO NA POPULAÇÃO DE PERNAMBUCO, REGIÃO NORDESTE DO BRASIL.....	143

APÊNDICE C - REGISTRO DE PROGRAMA DE
COMPUTADOR MutCodom (*Mutation Rate Estimate for
Codominant Markers*)

..... 155

APÊNDICE D - PEDIDO DE REGISTRO DE PROGRAMA DE
COMPUTADOR SSRgene (*Simulator of Simple Random
Sampling of Genetic data*)

..... 157

1. INTRODUÇÃO

Apesar da Genética Forense ser um ramo da ciência relativamente recente, dispõe-se de pelo menos 26 *softwares* aplicáveis à identificação humana disponíveis à comunidade científica. Com o intuito de trazer praticidade e aumentar a credibilidade das análises forenses, o uso destes *softwares* é cada vez comum no dia a dia do geneticista forense, principalmente no que diz respeito às análises estatísticas. Porém, ainda existem questões que precisam ser consideradas para que as ferramentas computacionais forneçam maior praticidade e eficiência aos usuários.

A validação de um banco de dados genéticos de dados, tais como os de locos microsatélites autossômicos e dos cromossomos X e Y (aut-STRs, X-STRs, Y-STRs), assim como àqueles derivados da análise de sequências do DNA mitocondrial (mtDNA), é uma etapa imprescindível na análise estatística forense. Tal procedimento não é possível com as ferramentas bioinformáticas atualmente disponíveis para identificação humana, implicando a necessidade de se realizar muitas manipulações exaustivas para a preparação dos arquivos de entrada dos vários *softwares* de genética de populações, decorrente da ausência de um formato padrão.

Também não se dispõe de um *software* que reúna a estatística dos dois tipos gerais de testes forenses (determinação de paternidade/maternidade [cível ou criminal] e identificação forense], utilizando um único formato de entrada e os dados dos principais tipos de marcadores genéticos (aut-STRs e Y-STRs). Diante desses pontos, há necessidade de desenvolvimento de um ambiente de análise que propicie maior praticidade para o cientista forense. O desenvolvimento desta solução viabilizaria a realização de cálculos estatísticos com diferentes classes de marcadores genéticos, contribuindo para melhorar a rotina de funcionamento de um laboratório de genética forense.

2. OBJETIVOS

2.1. OBJETIVO GERAL

- Desenvolver e validar um aplicativo (*BigASystem*) que integralize os cálculos estatísticos de identificação genética humana e de parâmetros populacionais de interesse forense a partir de dados STRs dos cromossomos autossômicos e Y.

2.2. OBJETIVOS ESPECÍFICOS

- Amostrar perfis genéticos aut-STRs e Y-STRs de casos forenses de paternidade e identificação humana, visando a construção de plataforma de banco de dados populacionais para obtenção e validação das frequências alélicas/haplotípicas;
- Criar módulos de estatísticas de determinação de paternidade/maternidade e de identificação forense, bem como de obtenção e validação de frequências alélicas/haplotípicas para implementação do *BigASystem*;
- Elaborar ensaios de estatística forense a partir dos dados amostrados para validação do *software BigASystem*.

3. REVISÃO BIBLIOGRÁFICA

3.1. IDENTIFICAÇÃO HUMANA

A identificação humana é o processo técnico científico pelo qual se determina a identidade de uma pessoa, ou um conjunto de diligências cuja finalidade é estabelecer uma identidade (Moreira, 1999; Vanrell, 2002; Khudhur & Croock, 2016). No processo de identificação, não se busca apenas estabelecer a semelhança de uma amostra a ser identificada com uma fonte potencial de evidência. Busca-se também a distinção de um certo objeto/indivíduo (potencial fonte de evidência) de todos os objetos considerados [demais possíveis fontes de evidências] (Champod, 2000; Khudhur & Croock, 2016).

Existem várias disciplinas nas Ciências Forenses voltadas à identificação humana, que vão desde àquelas mais convencionais até as baseadas na análise de dados moleculares. Análises da estrutura da arcada dentária, impressões digitais, grupos sanguíneos e das íris (Interpol, 2018 [Anexo 12]; Ballantyne, 2000; Tilstone et al., 2006) são exemplos de métodos convencionais. Os métodos moleculares, por sua vez, correspondem às análises de marcadores de DNA. Os métodos moleculares promoveram um grande desenvolvimento da identificação humana ao possibilitar, na maioria dos casos, a individualização inequívoca de uma amostra/indivíduo (Jeffreys, 1985a; 1985b; Wielbo, 2000; Margot, 2014).

A análise de marcadores de DNA é considerada atualmente como o padrão ouro na identificação humana (Ip et al., 2014; Thompson et al., 2013; Khudhur & Croock, 2016). A análise destes marcadores tem se tornado uma ferramenta indispensável na investigação criminal, na identificação de vítimas de desastres em massa e casos de pessoas desaparecidas, assim como nos testes de paternidade/maternidade cível e

criminal (Ip et al., 2014; Thompson et al., 2013; Khudhur & Croock, 2016). A identificação humana se divide em testes de identificação forense e de paternidade/parentesco por linhagem (Schanfield, 2000; Eldridge, 2019).

3.2. MARCADORES MOLECULARES GENÉTICOS

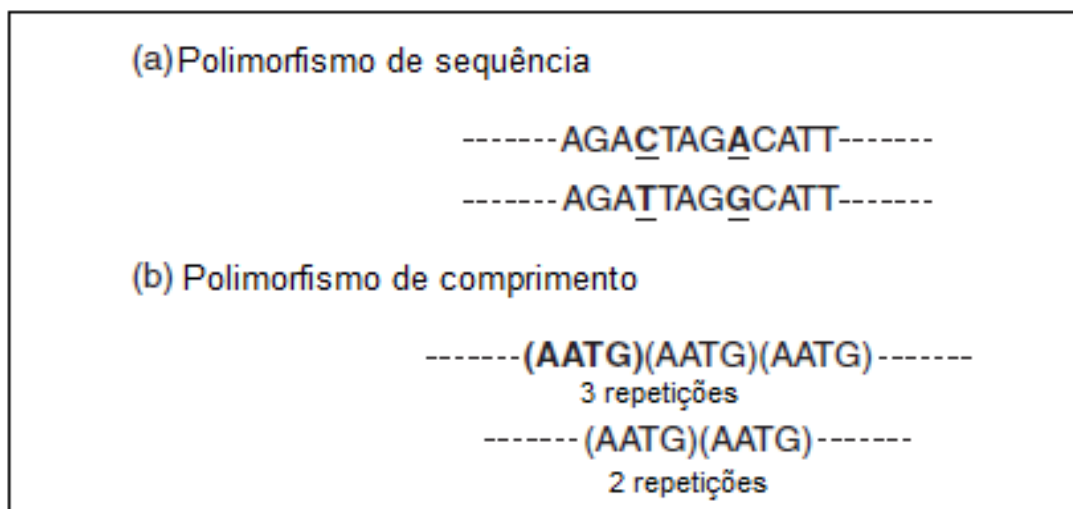
Do ponto de vista forense, não faz muito sentido analisar os 99,9% do genoma humano que são comuns entre os indivíduos [considerando apenas as mutações por substituição de nucleotídeos acumuladas desde o surgimento do humano moderno, a cerca de 150 000 anos] (Goodwin et al., 2007; Egeland et al., 2016; Niedzwiecki et al., 2017). O foco de interesse da genética forense é, assim, direcionado às regiões do genoma que são variáveis entre os indivíduos e que, graças ao advento das técnicas de biologia molecular, podem ser analisadas.

Marcadores genéticos podem ser definidos como caracteres que exibem diferenças entre dois ou mais indivíduos ou organismos, ou até mesmo entre grupos de organismos (Silva et al., 2012; Kanthaswamy, 2015; Kowalczyk et al., 2018). São também chamados de locos polimórficos, uma vez que apresentam duas ou mais formas variantes, e podem ser classificados em duas categorias: polimorfismos de sequência e de comprimento (Figura 1).

Os polimorfismos de sequência incluem os polimorfismos de nucleotídeo único – SNPs (do inglês, *single nucleotide polymorphism*), sequências altamente variáveis do DNA mitocondrial ou nuclear, elementos intercalados curtos – SINEs (do inglês, *short interspersed elements*), LINES (do inglês, *long interspersed elements*) e elementos SVAs (SINE/VNTR/Alu) [(Butler, 2005; Hameed et al., 2014; Sinha et al., 2015). Os polimorfismos de comprimento, tais como àqueles característicos dos locos minissatélites (ou VNTRs, do inglês, *variable number tandem repeats*), microsatélites

[STRs ou SSRs, do inglês, *short tandem repeats* e *simple sequence repeats*, respectivamente] e mini-STRs (Butler, 2005; Wyner et al., 2020; Short, 2016; Srivastava et al., 2019; Clark et al., 2019)

Figura 1- Duas formas de variações existentes em DNA: (a) Polimorfismo de sequência e (b) Polimorfismo de comprimento



Fonte: Butler (2005); Marwal & Gaur (2020)

Apesar de os polimorfismos de sequência terem aplicação na genética forense (Sheng et al., 2018; Parson, 2018; Gao et al., 2018; Parson et al., 2014; Kowalczyk et al., 2018; Butler, 2012), os polimorfismos de comprimento são os mais utilizados. A primeira aplicação do DNA na ciência forense ocorreu com o uso das VNTRs, em 1987, pelo geneticista britânico Alec Jeffreys. Isto ocorreu logo após a publicação da aplicação das VNTRs (apelidados à época de “*DNA fingerprinting*”) na identificação humana, descrita pelo próprio Jeffreys e colaboradores (Niedzwiecki et al., 2017).

Por questões de dificuldades em se analisar VNTRs em amostras criminais degradadas, os cientistas passaram a utilizar, a partir da década de 1990, a Reação em Cadeia da Polimerase (PCR, do inglês, *polymerase chain reaction*), que permitiu

a amplificação de regiões hipervariáveis mais curtas (STR), resultando na obtenção de milhares a milhões de cópias de uma sequência de DNA de interesse (Butler, 2012; Khudhur & Croock, 2016). Desde então, a análise de STRs passou a ser considerada como uma ferramenta biométrica confiável para fins de identificação humana (Rudin et al., 1996; AL-Zubaidi et al., 2014; Filoglu et al., 2014; Tripathi, 2015), sendo utilizada universalmente na resolução de casos forenses e em testes de paternidade/maternidade (Butler, 2006; El-Alfy et al. 2012; Tripathi, 2015, Jin et al, 2020). A combinação de amplificação por PCR com sequenciamento direto de DNA geralmente é a abordagem ideal para a identificação humana e provou ser um método confiável e reproduzível em casos forenses (Carracedo, 2013; Jin et al, 2020).

Os STRs são sequências polimórficas de um a seis nucleotídeos repetidos em *tandem*, encontradas nos genomas dos eucariontes (Litt & Luty, 1989; Fan & Chu, 2007; Virmond et al., 2016; Butler, 2012; Filoglu et al., 2014; Nurhantari & Suryadi, 2019; Srivastava et al., 2019), sendo os tetranucleotídeos mais utilizados na prática forense (Butler, 2012; Li et al., 2004; Vieira et al., 2016).

Os microsatélites constituem de 1 a 3% do genoma humano (Lander et al., 2001; Watson et al., 2004; Goodwin et al., 2007; Bagshaw, 2017). Estes marcadores podem ser encontrados nos cromossomos CrAut, CrX, CrY, sendo designados como aut-STRs, X-STRs e Y-STRs, respectivamente. As STRs possuem evolução supostamente neutra (Ellegren, 2004; Silva et al., 2012) e a taxa de mutação destes marcadores é cerca de 10^{-4} a 10^{-3} , ou quatro a seis ordens de magnitude maior do que a dos SNPs (Ellegren, 2000; Ellegren, 2004; Nachman & Crowell, 2000; Bagshaw, 2017)].

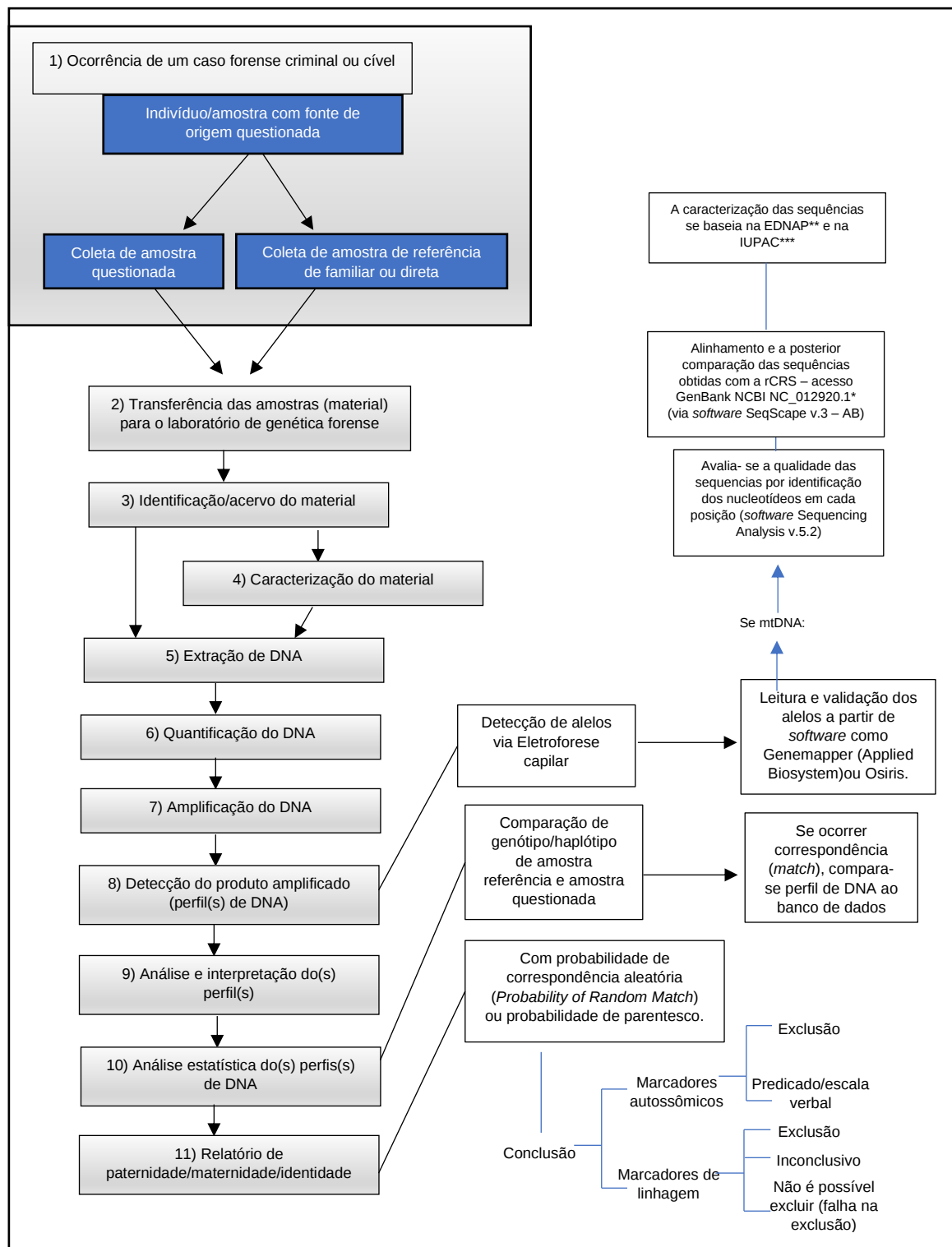
Independentemente da categoria do marcador, as análises de genética forense resultam na geração de perfis haplotípicos/genotípicos de DNA que envolvem os

seguintes processos (descritos por Goodwin et al. [2007] e Lee & Pagliaro [2013]): 1) ocorrência de um caso forense de interesse criminal ou cível; 2) Coleta de vestígios (material); 3) transferência do material para o laboratório; 4) identificação, catalogação e caracterização do material; 5) extração, quantificação e amplificação do material genético; 6) detecção do produto da amplificação (perfil de DNA); 7) análise e interpretação dos perfis genéticos obtidos; 8) avaliação estatística; e 9) emissão de relatório técnico (Figura 2).

O material em qualquer caso forense se refere a pelo menos duas amostras a serem avaliadas, excetuando-se àqueles casos que necessitam de verificação fenotípica (Virmond *et al.*, 2016; Kayser & de Knijff, 2011; Jobling & Gill, 2004): (1) evidência ou amostra questionada, Q (do inglês, *question sample*), que é a amostra/evidência com fonte de origem questionada (e.g. amostra oriunda da cena de crime; catástrofe ou acidente; criança com paternidade/maternidade questionada; indivíduo indocumentado); e (2) amostra de referência ou conhecida – K (do inglês, *known sample*), que corresponde a pessoas apontadas como potenciais fontes de evidência (e.g. suspeito de um crime; suposto pai/mãe ou indivíduos geneticamente vinculados a um indivíduo/amostra; parentes biológicos de Q; e objetos pessoais (Lee & Pagliaro, 2013; Espinheira, 2015; Butler, 2012; Butler, 2015aa). Um terceiro tipo de amostra seria àquelas de um banco de dados criminal ou de pessoas desaparecidas (Morrison et al., 2018), os quais não serão considerados neste trabalho.

Os STRs estão majoritariamente localizados em regiões não codificantes do genoma humano, estimando-se que não mais do que 8% estejam associados às regiões de codificação (Ellegren, 2000; Nurhantari & Suryadi, 2019). Os locos STRs são nomeados segundo regras já bem definidas. Considerando-se, por exemplo, o loco D12S391, temos: **D** representa DNA, **12** significa o cromossomo no qual o loco

Figura 2 - Processos envolvidos na geração de perfis de DNA após ocorrência de um caso forense. As amostras biológicas (material) são frequentemente caracterizadas mediante o DNA extraído



Fontes: Goodwin et al. (2007); Lee & Pagliaro (2013); Espinheira (2015); Butler (2005); Butler (2012); SWGDAM (2003); Allen et al. (2016); Roewer et al. (2020); Essam et al. (2020). * A numeração original (Anderson et al., 1981) foi mantida no rCRS com uma exclusão na sequência na posição 3107 para facilitar a correlação de trabalhos atuais baseados em rCRS com trabalhos anteriores baseados na numeração original; **do inglês, *European DNA Profiling Group* (Tully et al., 2001); *** do inglês, *International Union of Pure & Applied Chemistry* (Parson et al., 2014)

STR está localizado, **S** significa sequência de cópia única e **391** é o identificador único do marcador (Butler, 2005; Fan & Chu, 2007; Nurhantari & Suryadi, 2019).

Dos quatro tipos de STRs, os aut-STRs e Y-STRs são os mais utilizados na genética forense; o primeiro por permitir a individualização no geral e o segundo pelo fato de a maioria dos casos forenses criminais envolverem a identificação de amostras de indivíduos do sexo masculino (Jha et al., 2013; Niedzwiecki et al., 2017; Aceves et al., 2017; Dumache et al., 2018).

Em alguns casos de disputa familiar, os marcadores STR autossômicos não conseguirão produzir uma razão de verossimilhança conclusiva em um relacionamento questionado. Nesses casos, a análise de marcadores de DNA com padrões especiais de herança pode ser informativa, podendo auxiliar na confirmação de parentesco biológico entre perfis genéticos ligando famílias por meio de genealogia (linhagem) genética (Siegel et al., 2013; Niedzwiecki et al., 2017; Mahmood et al., 2020).

Para os marcadores autossômicos, na maioria dos casos, pode-se assumir que os genótipos fundadores de um heredograma são idênticos e independentemente distribuídos, dadas as suas frequências em uma população. A razão é que a dependência entre os genótipos autossômicos diminui muito rápido quando a relação se torna mais distante. Raramente haverá informações que possam ser utilizadas na distinção de um par de indivíduos quaisquer com relação de parentesco e com cinco a seis meioses de distância de um par não relacionado (Skare et al., 2009, Egeland et al., 2016).

Este contexto pode, no entanto, ser diferente para outros tipos de marcadores. Os marcadores do cromossomo Y, por exemplo, são transferidos de pai para filho quase inalterados. Então, uma subdivisão mais precisa do conceito de “não

relacionado” é necessária para se fazer inferências com base em tais dados: homens com ancestrais que vêm da mesma área geográfica têm muito mais probabilidade de apresentarem cromossomos Y semelhantes ou idênticos do que homens com ancestrais de áreas diferentes (Egeland et al., 2016).

3.2.1. Marcadores aut-STRs

Os marcadores mais utilizados na identificação humana são os locos principais do CODIS-FBI expandido [do inglês, *Expanded FBI CODIS Core STR Loci - Combined DNA Index System- Federal Bureau of Investigation/USA*] (Harris, 2015; STRbase-NIST - <https://strbase.nist.gov/coreSTRs.htm>; Tabela 1; Gettings et al., 2015). Para esses locos, já se dispõe de banco de dados populacionais, tais como o STRidER (Bodner et al., 2016)], além de publicações em periódicos especializados de *datasets* validados (Butler, 2006; Gouy & Zieger, 2017) e banco de dados criminais. Existem kits comerciais que abrangem esses locos (e.g. GlobalFiler™, PowerPlex® Fusion e PowerPlex® Fusion 6C System), permitindo a automatização da sua análise.

Os STR do CODIS têm modo de herança mendeliana, são evolutivamente neutros e não tem ligação gênica entre si. Isto resulta em equilíbrio de Hardy-Weinberg (EHW) e de ligação, de modo que um banco de dados populacional pode ser utilizado nos cálculos estatísticos forenses (Buckleton, 2005; Tsybovskii et al., 2017; Silva et al., 2012). Excetuando-se gêmeos univitelinos, este sistema leva a individualização humana [apesar de existirem questões envolvidas nesta premissa, mas que estão fora do escopo deste trabalho] (Evelt e Weir, 1998; Kaye, 2009a; Messina et al., 2018).

Dentre os desafios de interpretação mais comuns estão: trissomias como as do cromossomo 18 e 21; gêmeos univitelinos; tecidos transplantados ou outras situações de quimerismo; artefatos dos STRs (e.g. *stutters*, *pull-ups*, *spikes*, *strips*, *splits*);

desbalanço heterozigótico; alelos silenciosos; alelos *drop-out*; e alelos *off-ladder* (Bukcleton *et al.*, 2005; Weber-Lehmann *et al.*, 2014; Butler, 2015; Dumache *et al.*, 2018).

Tabela 1 - Locos genéticos aut-STRs principais ou "Core STR loci" usados em testes de identificação humana

Loco	Localização cromossômica	FBI-EUA	Europa*	Inglaterra	Alemanha	Interpol** standard set
D1S1656	1q42	X ^E	X			
D2S1338	2q35	X ^E	X	X		
D2S441	2p14	X ^E	X			
TPOX	2p25.3	X				
D3S1358	3p21.31	X	X	X	X	X
FGA	4q28	X	X	X	X	X
CSF1PO	5q33.1	X				
D5S818	5q23.2	X				
SE33	6q14		X		X	
D7S820	7q21.11	X				
D8S1179	8q24.13	X	X	X	X	X
D10S1248	10q26.3	X ^E	X			
TH01	11p15.5	X	X	X	X	X
vWA	12p13.31	X	X	X	X	X
D12S391	Chr 12	X ^E	X			
D13S317	13q31.1	X				
D16S539	16q24.1	X	X	X		
D18S51	18q21.33	X	X	X	X	X
D19S433	19q12	X ^E	X	X		
D21S11	21q21.1	X	X	X	X	X
D22S1045	22q12.3	X ^E	X			
Amelogenina	CrX e CrY	X	X	X	X	Opcional

Fontes: Butler (2006); Hares (2015); NIST STRBase (<https://strbase.nist.gov/coreSTRs.htm>). O "FBI- USA core loci" é um conjunto expandido que contém 13 locos originais (X) e 7 adicionais (X^E) (Hares, 2015). *(European Standard Set – ESS; **Organização Internacional de Polícia Criminal do inglês - The International Criminal Police Organization

3.2.2. Marcadores Y-STRs

Marcadores de DNA autossômico, tais como os locos STRs, são desvinculados, e “embaralhados” a cada geração; metade da informação genética de um indivíduo é de origem paterna e a outra, de origem materna. No entanto, os marcadores do CrY representam os chamados marcadores de linhagem paterna. As linhagens paternas podem ser seguidas com marcadores do cromossomo Y, uma vez que estes marcadores são transmitidos ao longo das gerações sem alteração [exceção feita àquelas decorrentes de eventos mutacionais] (Buckleton, 2005; Butler, 2012; Roewer et al., 2020). Seus locos são herdados como um "bloco", ao invés de sofrerem embaralhamento aleatório através de eventos de recombinação meiótica, de acordo com as expectativas da herança Mendeliana (Budowle et al., 2010; Butler, 2012; Mahmood et al., 2020). A maior parte do DNA no cromossomo Y é não recombinante, sendo que apenas as porções mais distais do cromossomo são capazes de se recombinar com o cromossomo X (Budowle et al., 2010; Colaco & Modi, 2018).

Por causa do padrão de herança masculino específico e da natureza haploide do cromossomo Y e do tamanho efetivo populacional menor, os marcadores do cromossomo Y são muito mais sensíveis à deriva genética e ao efeito fundador do que os marcadores autossômicos. As Y-STRs são amplamente utilizados para a avaliação da subestrutura da população ou diferenciação da população por meio do teste de distâncias genéticas pareadas (Egeland et al., 2016; Zhou et al., 2020).

Os cálculos estatísticos para uma probabilidade de correspondência aleatória não podem envolver a regra do produto (Roewer et al., 2020). Portanto, haplótipos obtidos de marcadores de linhagem nunca podem ser tão eficazes na diferenciação entre dois indivíduos quanto os genótipos de marcadores autossômicos, mas podem desempenhar um papel importante nas investigações forenses, bem como em outras

aplicações da identificação humana, tais como estudos históricos e arqueológicos a partir de restos mortais (Butler, 2012; Buckleton, 2005; Ambers et al., 2018). Dentre os desafios de interpretação estão: polissomias; microquimerismos (CrY); gêmeos univitelinos; artefatos de STRs; misturas de amostras; pseudogenes nucleares; e haplótipos comuns (Butler, 2012; Parsons & Coble, 2001; Coble et al., 2004; SWGDAM, 2014; Roewer et al., 2020).

O cromossomo Y corresponde a aproximadamente 0,82% (50Mb) do material genético nuclear dos homens (Butler, 2012; Colaco & Modi, 2018), sendo que cerca de 60% do cromossomo Y é constituído de heterocromatina, porção não codificante situada no braço longo (Butler, 2012; Colaco & Modi, 2018). O cromossomo Y é constituído de três partes: duas extremidades chamadas de PAR 1 e 2 (do inglês, *pseudoautosomal regions*) e MRY (do inglês, *non-recombining portion of the Y-chromosome* - 95% do CrY - mosaico de hetero e eucromatina) em cujas sequências eucromáticas (X-transposta, X-degenerada e amplicônica mantida por conversão genica) estão as 156 unidades de transcrição, que incluem 78 genes codificantes de proteína, assim como os marcadores de uso forense (Skaletsky et al., 2003; Butler, 2012). Enquanto os PARs e o braço curto são eucromáticos, uma grande porção do braço longo é heterocromática, com exceção da porção proximal justaposta ao centrômero, que é de natureza eucromática (Colaco & Modi, 2018).

Apesar de aproximadamente 400 locos Y-STRs terem sido descritos na literatura, os exames dos casos forenses incluem de 12 a 27 locos Y-STR com frequências de haplótipos disponíveis em bancos de dados populacionais (Hanson & Ballantyne, 2006; Willems et al., 2014; Butler, 2015b; Zhou et al., 2020). Eles constituem vários conjuntos de haplótipos, tais como o conjunto mínimo de haplótipos europeu, locos principais SWGDAM (do inglês, *U.S. Scientific Working Group on DNA*

Analysis Methods) e os dos kits comerciais *Yfiler®*, *Yfiler® Direct*, *PowerPlex® Y23*, *Yfiler® Plus* (Kayser *et al.*, 1997; SWGDAM, 2003; Gopinath *et al.*, 2016; Zhou *et al.*, 2020) descritos na Tabela 2.

Tabela 2 - Locos Y-STRs utilizados em exames genéticos forenses

Locos	Haplótipo mínimo europeu	SWGDAM	Yfiler® Yfiler® Direct	PowerPlex® Y23	Yfiler® Plus
DYS19	X	X	X	X	X
DYS385a	X	X	X	X	X
DYS385b	X	X	X	X	X
DYS389I	X	X	X	X	X
DYS389II	X	X	X	X	X
DYS390	X	X	X	X	X
DYS391	X	X	X	X	X
DYS392	X	X	X	X	X
DYS393	X	X	X	X	X
DYS438		X	X	X	X
(PENTA)					
DYS439		X	X	X	X
DYS437			X	X	X
DYS448			X	X	X
DYS456			X	X	X
DYS458			X	X	X
DYS635 (Y GATA C4)			X	X	X
Y GATA H4			X	X	X
DYS460*					X
DYS481*				X	X
DYS533*				X	X
DYF387S1a**					
DYF387S1b**					
DYS449**					
DYS518**					
DYS570**				X	
DYS576**				X	
DYS627**					
DYS549				X	
DYS643				X	
(PENTA)					

Fonte: Ballantyne et al., 2012; Engeland et al., 2016; Gopinath et al., 2016; Zhou et al., 2020.

*Y-STR altamente polimórfico; **Y-STR de mutação rápida (A taxa de mutação é comumente maior que 1%, e espera-se que parentes paternos distantes tenham haplótipos diferentes)

Bancos de dados Y-STRs globais como os YHRD (do inglês, *Y-Chromosome Haplotype Reference Database* - <http://www.yhrd.org>), Yfiler Haplotype Database (<http://www6.appliedbiosystems.com/yfilerdatabase/>), Ysearch (área de genealogia genética - <http://www.ysearch.org/>), STRBase. (http://www.cstl.nist.gov/biotech/strbase/y_strs.htm) e regionais como o US Y-STR Database (<http://www.usystrdatabase.org/>) estão disponíveis à comunidade científica.

3.3. GENÉTICA DE POPULAÇÕES

Componentes importantes para se avaliar o peso das evidências de DNA em casos de paternidade e de identificação forense a partir de dados aut-STR são as frequências alélicas e genotípicas. Para os casos de investigação de linhagem paterna Y, por sua vez, é de fundamental importância a atribuição das frequências haplotípicas (Evetts & Weir, 1998; Buckleton et al., 2005; Egeland et al., 2016; Roewer et al., 2020). São referidos como probabilidade P , por exemplo, do alelo paterno da criança ou de um perfil genotípico/haplotípico em uma população (Buckleton et al., 2005; Egeland et al., 2016). A área da genética que lida com estes parâmetro dentro e entre populações é a genética de populações (Lewontin & Hartl, 1991; Nature Research, 2020: <https://nature.com/subjects/population-genetics>), que faz uso de um banco de dados e de métodos baseados em modelos populacionais (Kaye e Sensabaugh, 2011; Asplen & Lane, 2003; Anslinger et al., 2001; Giannelli et al., 2011, NRCII, 1996; Egeland et al., 2016).

3.3.1. Banco de dados populacional

O banco de dados populacional é uma coleção de perfis de DNA que corresponde à um grupo de geralmente 100 ou mais indivíduos anônimos e

geneticamente não relacionados, selecionados da população contemporânea a qual o suspeito/suposto pai pertencem (Butler, 2010; Carracedo et al., 2013; Carracedo et al., 2014; Roewer et al., 2020). É uma amostra representativa da população, sendo geralmente obtida por meios convenientes, tais como doadores de sangue, banco de dados de perfis genotípicos de criminosos e indivíduos investigados em laboratório de paternidade (Evetts & Weir, 1998; Espinheira, 2015; Roewer et al., 2020). Estas informações necessitam ser validadas quanto à conformidade com o poder probatório a partir de parâmetros populacionais de interesse forense (Buckleton et al., 2005; Carracedo et al., 2013; Carracedo et al., 2014; Roewer et al., 2020).

3.3.2. Modelos nível populacional

A atribuição da probabilidade P é justa e razoavelmente baseada nas características evolutivas dos marcadores considerados e a forma pela qual são herdados ao longo das gerações (Neuhauser, 2001; Buckleton et al., 2005; Egeland et al., 2016). A seguir, são apresentados os principais modelos utilizados na estimativa da probabilidade P , que especificam a probabilidade de uma variante específica, alelo aut-STR ou haplótipo Y-STR, assim como de um genótipo específico aut-STR.

3.3.2.1. Estimativa de frequência alélica aut-STRs e haplotípica Y-STR

Considere que existe um conjunto de alelos k possíveis que são observados com dadas probabilidades (ou frequências alélicas) a partir de um banco de dados de perfis aut-STR e Y-STRs, obtêm-se as frequências alélicas e haplotípicas, que são calculadas por contagem direta (Holland & Parsons, 1999; Zhao et al., 2003; Butler, 2010; Steele & Balding, 2014; Egeland et al., 2016; Al-Azem et al., 2017; Andersen & Balding, 2019; Roewer et al., 2020).

Considere $C = (C_1, \dots, C_v)$ como um vetor de contagens de observações para v diferentes alelos/haplótipos encontrados no banco, e que seja N o número total de alelos contados (N corresponderá ao dobro do número de perfis contidos no banco de dados, quando o marcador for autossômico). A estimativa do parâmetro de frequência alélica/haplotípica é derivada da divisão das contagens C_i por N . Na forma vetorial, podemos escrever que o vetor P dos parâmetros de frequência alélica é calculado do seguinte modo (Egeland et al., 2016):

$$P = \frac{C}{N} \quad (1)$$

A contagem direta é um método conservador e pode, portanto, reduzir o poder de evidência, principalmente se o perfil Y-STR investigado tiver frequência abaixo de *1/tamanho do banco de dados*. A probabilidade de correspondência de um perfil haplotípico também pode ser estimada por métodos baseados em um modelo evolutivo de diversificação de haplótipos, tal como o Discreto de Laplace, entre outros (Krawczak, 2001; Roewer et al., 2000; Gusmão et al., 2006; Andersen et al., 2013; Andersen & Balding, 2017; Taylor et al., 2018; Andersen et al., 2019; Andersen et al., 2020; Roewer et al., 2020).

3.3.2.2. Estimativa de probabilidade de perfil genético aut-STR

A atribuição da probabilidade P neste caso é realizada não somente se utilizando as frequências alélicas do banco de dados, mas também modelos populacionais (NRCII, 1996; Evett & Weir, 1998; Song et al., 2010; Bradburd et al., 2016; Delgado & Neyra, 2019). Ela não é realizada a partir de contagem direta dos genótipos (do banco de dados) por que esta é mais propensa a erros de amostragem em relação a contagem direta dos alelos (NRCII, 1996). Esses modelos são: regra do produto (Teste de independência – o modelo mais utilizado); subestruturação populacional; e

população mista (Buckleton et al., 2005; Song et al., 2010; Egeland et al., 2016; Samantha et al., 2016; Bradburd et al., 2016).

O modelo do produto é baseado nas seguintes equações (Egeland et al., 2016):

$$P_i = \begin{cases} p_{i_1}^2, & A_{i_1} = A_{i_2} & \text{genótipo homozigoto} \\ 2p_{i_1}p_{i_2}, & A_{i_1} \neq A_{i_2} & \text{genótipo heterozigoto} \end{cases} \quad (2)$$

Considera-se que P_i é a frequência genotípica de um único loco i do genótipo $A_{i_1}A_{i_2}$ que pode ser obtida pelo produto da frequência alélica p_{i_1} e p_{i_2} (dos alelos A_{i_1} e A_{i_2} , respectivamente). Já a probabilidade do genótipo multiloco (P) pode ser atribuída a partir de probabilidades dos locos únicos P_i :

$$P = \prod_i P_i \quad (3)$$

As populações subestruturadas podem ser representadas por dois modelos (os aut-STR podem se apresentar genética e geograficamente estruturados), segundo Messina et al. (2018) e Aalbers et al. (2020):

- A. Modelo baseado na população sob efeito Wahlund (recomendação 4.1. da NRC II), que oferece uma correção para o desequilíbrio de HW decorrente do efeito Wahlund:

$$P_i = \begin{cases} p_{i_1}^2 + p_{i_1}(1 - p_{i_1})F, & A_{i_1} = A_{i_2} \\ 2p_{i_1}p_{i_2}, & A_{i_1} \neq A_{i_2} \end{cases} \quad (4)$$

F é o coeficiente de consanguinidade intra-pessoa da população (numericamente igual ao valor do coeficiente de coancestralidade θ (NRCII, 1996). Quanto maior a distância entre as subpopulações (que divergem por deriva genética), maior deve ser o tempo de separação entre elas e menor será o valor θ (Weir, 1996; Evett & Weir, 1998; NRCII, 1996; Crow e Kimura, 1970; Aalbers et al., 2020).

Quando não se tem o valor θ empírico de uma população, caso seja pequena (centenas de indivíduos) e isolada, esse parâmetro pode assumir um valor conservador de 0,03 [esse valor é intermediário entre os encontrados em populações com casamentos entre primos de primeiro e de segundo grau, além de ser um limite superior razoável para o que pode ser esperado (NRCII, 1996; Evett & Weir, 1998; SWGDAM, 2019; Aalbers et al., 2020)].

B. Modelo baseado em subpopulação centrada no suspeito de Balding e Nichols (Balding & Nichols, 1995; as equações 4.10 e recomendação 4.2 em NRC II, 1996; Evett & Weir, 1998; Aalbers et al., 2020)

Leva em consideração a análise de parentes próximos (irmãos, filhos, pais), e distantes (primos, tios). Depois, parentes de segundo grau e assim por diante, pessoas que nem conhece e nem tem sobrenomes semelhantes relacionadas mais remotamente (subpopulação) - Balding & Nichols, 1994; Buckleton et al., 2005; Malaspinas et al., 2011; Aalbers et al., 2020). Este modelo é representado pelas fórmulas:

$$P_i = \begin{cases} \frac{[3\theta + (1 - \theta)p_{i_1}][2\theta + (1 - \theta)p_{i_1}]}{(1 + \theta)(1 + 2\theta)}, & A_{i_1} = A_{i_2} \\ \frac{2[\theta + (1 - \theta)p_{i_1}][\theta + (1 - \theta)p_{i_2}]}{(1 + \theta)(1 + 2\theta)}, & A_{i_1} \neq A_{i_2} \end{cases} \quad (5)$$

$$P = \prod_i P_i \quad (6)$$

A multiplicação não corresponde à regra do produto, indicando eventos independentes. Faz parte da condicionalidade dos eventos (genótipos, cada um em um loco).

Já quando uma população com mistura recente (ou em transição) está em EHW e não está ainda em equilíbrio de ligação (uma demonstração deste estado pode ser vista na tabela 3.5 do NRCII, 1996), utiliza-se uma adaptação da fórmula de Balding e Nicols de toda a população misturada como apresentada por Law (2002) e Buckleton et al. (2005). Um exemplo extremo seria uma população na qual todos os indivíduos possuem um único tipo de perfil genotípico multiloco após uma geração. Outros modelos para populações mistas estão descritos em Pfaff et al. (2001), Thornton et al. (2012), Sousa et al. (2012), Bradburd et al. (2016), Zhou et al. (2017), Ragsdale & Gravel (2019), Boca et al. (2020) entre outras publicações.

3.4. ESTATÍSTICA FORENSE

A estatística forense é a aplicação de modelos de probabilidade e técnicas estatísticas a evidências científicas, como evidências de DNA e leis. Em contraste com as estatísticas cotidianas, os geneticistas forenses relatam as verossimilhanças como razões de verossimilhança – RV (Gill et al, 2018). Essa proporção de probabilidades é então usada por júris ou juízes para fazer inferências ou conclusões e decidir questões jurídicas (Ewett & Weir, 1998; Champod, 2000; Wielbo, 2000; Gill et al, 2018).

Para avaliar as evidências de forma adequada é necessário comparar as probabilidades das evidências sob duas proposições diferentes: H_p e H_d [hipótese de acusação e de defesa, respectivamente [do inglês, *prosecution and defense hypotheses*] (Egeland et al., 2016).

Na genética forense, ou mais especificamente na identificação humana, a comparação do perfil genético recuperado a partir da amostra questionada Q com o da amostra referência K (suspeito de fonte de evidência em potencial) é representada

pela hipótese H_p . A comparação daquela com a(s) amostra(s) de outros possível(is) suspeito(s) é representada por H_d . Neste trabalho, os possíveis suspeitos sempre se referirão a pessoas não relacionadas ao indivíduo da amostra questionada de uma específica população. Esta, por sua vez, será representada pelo banco de dados populacional (Champod, 2000; Wielbo, 2000; Gill et al, 2018;).

O suspeito da paternidade ou da execução de um crime/vítima são exemplos de indivíduos da amostra K , enquanto uma criança com paternidade questionada e materiais biológicos encontrados na cena do crime seriam amostras Q em um caso de paternidade e de identificação forense. Uma vez obtida a RV, o geneticista forense deve preparar um relatório baseado nas diretrizes da ISFG e em conformidade com os requisitos legais nacionais. O relatório deve conter a conclusão/interpretação da evidência na forma de uma declaração não somente quantitativa (como o valor RV) como também qualitativa [descrição verbal] (Roewer et al., 2020; Gill et al., 2020).

A conclusão é dada em referência ao suporte da hipótese de acusação H_p , já que foi ela que motivou a realização da análise forense. Se a hipótese H_p for “o indivíduo suspeito é a fonte do perfil genético questionado Q ” e os perfis genéticos das amostras Q e K compartilham todos os alelos para cada loco analisado, esta hipótese não é excluída ao considerar a hipótese alternativa (Champod, 2000; Buckleton, Clayton & Triggs, 2005; Roewer et al., 2020).

3.4.1. Estatística Bayesiana

Na genética forense a estatística mais utilizada é a Bayesiana (Evet & Weir, 1998; Gjertson et al., 2007; Roewer et al. 2020; Tillmar et al. 2017; Carracedo et al. 2000). A estatística Bayesiana (ou subjetivista) é representada pelo teorema de

Bayes, publicado por Thomas Bayes em 1764 (Stigler, 1982). O teorema de Bayes na genética forense é utilizado geralmente na forma de *odds* (Butler, 2015a).

Nos problemas forenses, haverá algum evento ou proposição H_p sobre o qual há incerteza, além das evidências genéticas (E) e não genéticas (I – informações condicionantes como declarações de testemunhas). O papel geral do geneticista forense é buscar como E afeta a incerteza de H_p . Se H_p e H_d são hipóteses exaustivas e mutualmente exclusivas, as *odds* do teorema de Bayes são (Evet & Weir, 1998):

$$\frac{P(H_p|E,I)}{P(H_d|E,I)} = \frac{P(E|H_p,I)}{P(E|H_d,I)} \times \frac{P(H_p|I)}{P(H_d|I)} \quad (7)$$

Para entender o funcionamento desta fórmula, é necessário entender o funcionamento do sinal de condicionamento. Isso geralmente é escrito como “|” e pode ser lido como "se" ou "dado" (Buckleton, 2005). A forma verbal desta fórmula é:

$$\text{Odds a posteriori} = \text{Razão de verossimilhança} \times \text{Odds a priori} \quad (8)$$

As *odds a priori* são as *odds* das hipóteses H_p antes da evidência de DNA. As *odds a posteriori* são as *odds* após a evidência do DNA. A razão de verossimilhança representa a relação entre as duas *odds*. Também diz como atualizar a opinião do geneticista forense de maneira lógica, depois de ouvir as evidências (Buckleton, 2005). Como cada probabilidade *a priori* é considerada igual a 0.5, ou seja, as *odds a priori* são 1:1* (Evet & Weir, 1998; Berry, 1991; Robertson e Vignaux, 1992; Essam et al., 2020; Egeland et al., 2016), as *odds a posteriori* correspondem à RV [por razão prática os termos I serão omitidos das fórmulas de RV daqui em diante (Evet & Weir, 1998; Essam et al., 2020):

$$RV = \frac{P(E|H_p)}{P(E|H_d)} \quad (9)$$

As *odds* de Bayes, e mais especificamente RV, assumem valores que vão de zero (quando H_p é falsa) até $+\infty$ (quando H_p é verdadeira) (Evet & Weir, 1998). As *odds* de Bayes podem estar a favor (>1) ou contra (quando é <1) a hipótese de acusação H_p , ou pode não dar suporte a nenhuma hipótese, quando é igual a 1 (Gill et al, 2018; Essam et al., 2020). Dessa maneira, se a probabilidade de ocorrência de um evento é de 80%, então as chances de ocorrência deste evento são de 4 para 1 (Evet & Weir, 1998; Gill et al, 2018; Essam et al., 2020).

3.4.1.1. Estatística de teste de identificação forense

Imagine que um crime foi cometido e um exame da cena do crime (C) revelou uma mancha de sangue, a qual foi coletada. Diante de informações recebidas, o investigador deteve um suspeito (S), que fornece uma amostra de material biológico. As hipóteses consideradas foram baseadas nas descrições de Ewett & Weir (1998), Buckleton, Walsh & Harbison (2005), Buckleton et al., 2011a, Milot et al., 2020 e Roewer et al., 2020):

- H_p : A pessoa $\underline{S}$ é a fonte do DNA;
- H_d : Um homem aleatório $\underline{A}$ da população de referência $\underline{R}$ é a fonte do DNA e $\underline{A}$ é um homem diferente de $\underline{S}$.

Os resultados da tipagem de DNA para a amostra de suspeitos e da cena do crime foram G_S e G_C , respectivamente. Sendo $E = (G_C, G_S)$, a probabilidade das evidências sob cada uma dessas hipóteses (Aitken & Taroni, 1996; Aitken, 1987; Balding & Donnelly, 1995a; Buckleton, Clayton e Triggs, 2005; Egeland et al., 2016; Roewer et al., 2020) é obtida do seguinte modo:

$$RV = \frac{P(E|H_p)}{P(E|H_d)} = \frac{P(G_S, G_C|H_p)}{P(G_S, G_C|H_d)} \quad (10)$$

Expandindo pela terceira lei da probabilidade [de modo suspeito ancorado] (Evet & Weir, 1998; Egeland et al., 2016) e considerando $P(G_S | H_p) = P(G_S | H_d)$, já que o fato de o suspeito ser ou não ser a fonte da amostra do crime não tem influência em G_S :

$$RV = \frac{P(G_C | G_S, H_p)}{P(G_C | G_S, H_d)} \times \frac{P(G_S | H_p)}{P(G_S | H_d)} = \frac{P(G_C | G_S, H_p)}{P(G_C | G_S, H_d)} \quad (11)$$

O numerador pode assumir valor 1 ou 0 quando $G_C = G_S$ e $G_C \neq G_S$, respectivamente, assumindo-se que os genótipos considerados podem ser determinados sem erros (Parson e Bandelt, 2007; Butler, 2012; Buckleton et al., 2005; Egeland et al., 2016; Roewer et al., 2020). Já o denominador corresponde a probabilidade P de genótipos (tópico 3.3. Genética de Populações). Considerando que houve correspondência (em inglês, *Match*) entre G_C e G_S , a RV fica:

$$RV = \frac{1}{P} \quad (12)$$

As três possíveis conclusões referentes a RVs dos testes por comparação de perfis Y-STRs são (DJ-USA, 2020; Roewer et al., 2020):

1. **Exclusão:** quando os perfis haplotípicos são diferentes, de modo que não poderiam ter sido originados da mesma fonte. Se houver duas ou mais diferenças de locos entre as amostras questionadas e conhecidas, as amostras podem ser excluídas como originárias da mesma pessoa ou linhagem;
2. **Inconclusivo:** quando não há dados suficientes que possibilitem uma interpretação ou os resultados haplotípicos obtidos foram ambíguos. Se houver uma diferença de locos entre as amostras questionadas e conhecidas, o resultado será inconclusivo. Uma única inconsistência pode ser uma mutação,

inclusive quando há somente uma meiose entre os indivíduos das amostras Q - criança - e K - genitor ou seu parente (Parsons et al., 1997).

- 3. Não é possível excluir (falha na exclusão):** quando os resultados haplotípicos são os mesmos em Q e em K e poderiam ter se originado da mesma pessoa ou linhagem. Um único loco não pode ser utilizado para apoiar uma interpretação de exclusão (Stewart et al., 2001; SWGDAM, 2003).

A escala verbal das RVs obtidas por testes de identificação por aut-STR (Aitken & Taroni, 1998; Broeders, 1999; Champod et al., 2000; Evett, 2000; Evett & Buckleton, 1989; Buckleton, 2005; Marquis et al., 2016; Eldridge, 2019; SWGDAM, 2017; SWGDA (Ad hoc), 2018; DJ-USA, 2018), conforme descrito na Tabela 3

A tipagem dos haplótipos só é útil para demonstrar exclusão. Como se espera que todos os parentes da mesma linhagem (salvo mutações) tenha o mesmo perfil haplotípico, é problemático estimar a força de uma hipótese quando o haplótipo de um suspeito não pode ser excluído, ou seja, K é indistinguível de Q (de Knijff 2003; Butler, 2012; Roewer et al., 2020).

Tabela 3 - Escala verbal associada a valores RV de teste de identificação forense por dados aut-STRs

RV	Forma verbal	Proposição suportada
≥ 1000000	Evidência extremamente forte	Suporte para H_p
> 10000	Muito forte	
$> 1000 - 10000$	Forte	
$> 100 - 1000$	Moderadamente forte	
$> 10 - 100$	Moderada	
$> 1 - 10$	Limitada	
1	Inconclusivo	Suporte para H_d
0,1 - <1	Limitada	
0,01 - <0,1	Moderada	
0,001 - <0,01	Moderadamente forte	
0,0001 - <0,001	Forte	
<0,0001	Muito forte	
0,000001	Extremamente forte	

Fonte: Buckleton (2005); Marquis et al. (2016); ENFSI (2016); Eldridge (2019); Gill (2020)

3.4.1.2. Estatística de teste de paternidade/parentesco

Considere que M seja a mãe do filho C e que SP seja o suposto pai. Seus genótipos serão designados G_M , G_C e G_{SP} , respectivamente. Considere disputas de paternidade (as quais normalmente resultam em processos civis), mantendo os mesmos sobescritos de hipóteses: H_p e H_d . Os termos “mãe” e “pai” em um caso de

paternidade (cível ou criminal) se referem aos pais biológicos. As duas proposições são:

- H_p : SP é o pai de C .
- H_d : Outro homem é o pai de C .

Considerando que E resume todas as evidências genéticas: G_M , G_C e G_{SP} , a RV fica assim (Evetts & Weir, 1998; Essam et al., 2020):

$$RV = \frac{P(E|H_p)}{P(E|H_d)} = \frac{P(G_C, G_M, G_{SP}|H_p)}{P(G_C, G_M, G_{SP}|H_d)} \quad (13)$$

A RV no teste de paternidade é chamada de índice de paternidade ou parentesco – IP (Berry, 1991; Robertson e Vignaux, 1992; Morling et al., 2002; Gjertson et al., 2007; Egeland et al., 2016; Essam et al., 2020). Aplicando várias vezes a terceira lei da probabilidade e considerando que se pode escrever o genótipo G_C da criança como $A_M A_P$, o IP pode ser apresentado na sua forma mais simplificada:

$$IP = RV = \frac{P(A_M|G_M)(A_P|G_{SP}, H_p)}{P(A_M|G_M)P(A_P|H_d)} \quad (14)$$

Os valores assumidos pelos termos do numerador ou denominador não condicionados a H_d correspondem às proporções de transmissão alélica mendeliana (0, $\frac{1}{2}$ e 1). Os termos que incluem H_d na condicionalidade corresponde a frequência alélica da população em questão (Egeland et al., 2016). A frequência alélica e não a probabilidade genotípica do banco de dados importa no teste de paternidade por dados aut-STRs. Os modelos populacionais regra do produto e população subestruturada devem ser abordados a nível de alelos (Egeland et al., 2016). Este último envolve frequência alélica e coeficiente θ populacional, enquanto o primeiro, apenas a frequência alélica (Buckleton, Clayton & Triggs, 2005; Amorim et al., 2016).

Uma vez obtido o IP para cada loco, os IPs são multiplicados, resultando no IPA (índice de paternidade acumulado). Na conclusão do teste de paternidade, normalmente uma probabilidade é usada no lugar das *odds*. Assumindo que as *odds a priori* são 1:1 (probabilidades *a priori* igual a $\frac{1}{2}$), elas podem ser convertidas em probabilidade de paternidade ($P(H_p)$ ou PP) ou probabilidade *a posteriori* assim (Egeland et al., 2016):

$$\text{Probabilidade a posteriori} = P(H_p) = \frac{IPA}{IPA + 1} \quad (15)$$

Essa é conhecida como equação Essen-Möller (Gjertson et al., 2007; Morimoto et al., 2016). As probabilidades de paternidade podem ser apresentadas na sua forma de expressão verbal, seguindo os predicados de Hummel (Hummel, Ihm e Schmidt, 1971; Caine et al., 2011; Geserick & Wirth, 2012; Morimoto et al., 2016; ENFSI, 2016) apresentados na Tabela 4. Probabilidades como o índice de Essen-Möller são restritas ao intervalo de 0 a 1 (ou de 0% a 100%) e são mais fáceis de interpretar do que RVs, mas isso vem com uma desvantagem: a necessidade de especificar probabilidades *a priori*. (Egeland et al., 2016).

O teste de paternidade lida principalmente com a investigação na sua forma de trio simples: mãe, filho e suposto pai. Mas também podem ser duos (suposto pai e filho – um tipo de teste de paternidade sem a mãe, meios-irmãos e teste complexo de paternidade [casos nos quais o suposto pai está ausente/falecido e um parente do mesmo é analisado no lugar- sua forma simplificada e no modelo populacional da regra do produto corresponde ao índice avuncular] (Evetts & Weir, 1998; Buckleton, Clayton & Triggs, 2005; Aceves et al., 2017; Eldridge, 2019).

Tabela 4 - Tabela de predicação verbal de Hummel. Assumindo probabilidades *a priori* iguais

PP	IP	Predicação verbal
$\geq 0,998$	≥ 399	Paternidade praticamente provada
$\geq 0,99$	≥ 99	Altamente provável
$\geq 0,95$	≥ 19	Muito provável
$\geq 0,90$	≥ 9	Provável
$\geq 0,10$	$\geq 1/9$	Nenhum predicação verbal (inconclusivo)
$\geq 0,05$	$\geq 1/19$	Improvável
$\geq 0,01$	$\geq 1/99$	Muito improvável
$\geq 0,02$	$\geq 1/399$	Altamente improvável
$< 0,02$	$< 1/399$	Praticamente excluída

Fonte: Hummel, Ihm e Schmidt (1971); Caine et al. (2011); Geserick & Wirth (2012); Morimoto et al. (2016); ENFSI, 2016; Eldridge, 2019

Na identificação de paternidade, a capacidade de excluir um indivíduo é determinada pelas leis de herança de Mendel e não por uma comparação com um conjunto de genótipos, tal como é realizado na identificação forense (Schanfield, 2000; Egeland et al., 2016).

Diante disso, quando há uma inconsistência Mendeliana (IM), tem-se uma exclusão de paternidade, havendo, porém, exceções. Eventos como padrão de três bandas causadas por trissomia (por mutação somática ou trissomia ou ainda translocações), presença de um alelo silencioso ou uma mutação podem dar origem a uma IM mesmo em casos com verdadeiros pais biológicos (Buckleton, Clayton & Triggs, 2005; Aceves et al., 2017; Rangel-Villalobos et al., 2019). Isto pode acontecer

porque o teste de paternidade inclui no mínimo duas gerações, ou seja, pelo menos uma meiose de distância (Nurhantari & Suryadi, 2019).

Uma abordagem do cálculo de IP considerando o IM é realizada a partir da extensão da fórmula de IP apresentada anteriormente considerando um modelo mutacional como descrito por Fimmers et al. [1992] (Gjertson e al., 2007; Coble et al., 2016):

$$IP = \frac{(\mu_{A,E} + \mu_{B,E})}{2p_E} \quad (16)$$

Considerando um caso no qual os genótipos do alegado pai, mãe e criança são (A, B) , (C, D) e (C, E) respectivamente, sendo μ_{IJ} igual a taxa de mutação específica para mudança de alelo I para o J . O I e J pertencem ao conjunto de alelos do marcador autossômico considerado e p_E é a frequência relativa do alelo E .

Ainda considerando o caso forense exemplificado no Tópico 3.4.1.1. para os dados de identificação humana, um caso forense frequentemente pode não ter uma amostra de S . No lugar desta amostra, utiliza-se como referência o material de um parente paterno K da mesma linhagem haplotípica de S , com o haplótipo G_K . Como se trata de um teste de parentesco, a RV passa a ser chamada de índice de parentesco (Buckleton, Clayton e Triggs, 2005; Egeland et al., 2016). Aplica-se a mesma lógica do teste de identificação, ou seja, RV é calculada por $1/N$, onde N é a frequência do haplótipo Y-STR obrigatório na população. Quando um filho é concebido, ele receberá o cromossomo Y de seu pai. Assim, a probabilidade de um homem transmitir seu cromossomo Y para seu filho é de 100% (ou 1,0), que se torna o numerador no cálculo da RV. A mãe não pode transmitir um cromossomo Y, uma vez que não tem nenhum, então ela é totalmente excluída do cálculo. H_p e H_d tornam-se assim (Allen & Pritchard, 2016):

- H_p : O homem testado transmitiu seu cromossomo Y a filho concebido por uma mulher.
- H_d : A mulher concebeu um filho ao qual um homem aleatório transmitiu seu cromossomo Y.

Uma alternativa de valor RV considerando populações subestruturada é dada por Balding & Nichols (1994), SWGDAM (2014) e Roewer et al. (2020):

$$RV \approx \frac{1}{\theta_Y + (1 - \theta_Y) \times p_{hap}} \quad (17)$$

na qual θ_Y o coeficiente de coancestralidade estimado para o cromossomo Y.

3.4.2. Poder de exclusão caso específico

Em alguns países, utiliza-se uma matemática diferente quando o caso de paternidade é criminal. No lugar de se utilizar a razão de verossimilhança, ou IP, e a probabilidade Bayesiana de paternidade, busca-se a frequência populacional de pais em potencial não excluídos [RMNE do inglês, *random man not excluded* (Buckleton, Clayton & Triggs, 2005; Bodner et al., 2016; Rasool & Hussain, 2020; Fung & Hu, 2016)] também chamada de $PE_{\text{Caso Específico}}$. Esta probabilidade é usada para dar uma luz se encontrar o pai verdadeiro pode ser fácil ou difícil (Schanfield, 2000; Lee et al., 2010; Rasool & Hussain, 2020).

O método RMNE é uma estimativa da fração de homens na população que carrega os alelos obrigatório da criança e poderia, portanto, ser o pai biológico. Diferentemente do índice de paternidade, a informação genética do suposto pai não é considerada, e sim o alelo (ou possíveis alelos) de origem paterna da criança (Schanfield, 2000; Rasool & Hussain, 2020; Fung & Hu, 2016). O $PE_{\text{Caso Específico}}$ não corresponde ao poder de exclusão PE ou ao poder de exclusão Médio $\bar{A}$ utilizados na validação de banco de dados populacionais (AABB, 2003 e 2013; Fung et al., 2002;

Fung & Hu, 2016), pois se referem à diversidade genética incluindo todos os alelos do loco. Considere um caso no qual a criança e a mãe têm genótipos $G_C=A_iA_i$ e $G_M=A_iA_j$ e um outro caso cujos genótipos são $G_C=A_iA_i$ e $G_M=A_iA_k$. Um homem aleatório sem o alelo A_i (ou seja, ambos seus alelos não são A_i) então será excluído. Assim, considerando que a frequência do alelo A_i é p_i , a probabilidade de RMNE é (Fung & Hu, 2016):

$$RMNE = 1 - (1 - p_i)^2 \quad (18)$$

E se $G_C=A_iA_j$ e $G_M=A_iA_j$, a probabilidade de RMNE é (Fung & Hu, 2016):

$$RMNE = 1 - (1 - p_i - p_j)^2 \quad (19)$$

E a probabilidade de exclusão para os dados haplotípicos (Buckleton et al., 2005):

$$PE = 1 - p_{hap} \quad (20)$$

Onde p_{hap} é a frequência de haplótipo da criança.

3.4.3. Validação de banco de dados populacional

É prática corrente de um laboratório validar o banco de dados da população que pretende usar antes de ir ao tribunal com estimativas de peso de evidencia. Esse é um recurso desejável, assim como a sua publicação ou deposição em bancos de domínio público (Buckleton et al., 2005; Coble et al., 2016; Tsybovskii et al., 2017; Hakim et al., 2020). Isso é útil para outros cientistas que desejam usar esses dados ou compará-los com suas próprias populações ou com outras populações (Zhivotovsky et al., 2009; Tsybovskii et al., 2017; Guevara et al., 2020; Ochoa & Storey, 2021).

Para validar o banco de dados, é necessário incluir algumas estatísticas para descrever os dados (Buckleton et al., 2005; Hakim et al., 2020). Estas estatísticas são

os parâmetros populacionais (baseados em Buckleton et al., 2005; Butler, 2012; Huston, 1998; Gouy & Zieger, 2017; Al-Azem et al., 2017):

- Para dados STRs autossômicos:** frequência alélica (F_a), número de tipos de alelos, frequência genotípica (zigótica) observada e esperada, número de tipos de genótipos esperado e observados, p -value da análise de EHW (a avaliação do EHW no banco aplicado na paternidade apenas para avaliar a tempo de validade do banco, um teste tipo “*Goodness of fit*” (em português Grau de ajuste) por teste exato de Fisher (Levene, 1949) por método de aleatorização de Monte Carlos (MMC) (Guo & Thompson, 1992; Engels, 2009), diversidade padrão (heterozigosidade esperada ou diversidade genética (HE_e) e observada (HE_o), homozigosidade esperada (HO_e) (Nei, 1987) e observada (HO_o) (Huston, 1998), probabilidade de coincidência ou *match* (PM) (Fisher, 1951), poder de discriminação (PD), conteúdo informativo de polimorfismo [em inglês – *polymorphism information content* – PIC] (Botstein et al., 1980; Guo & Elston, 1999), probabilidade ou poder de exclusão (PE) (Brenner & Morris, 1989; Huston, 1998; fórmula citada também em Butler, 2015a) - este não é o poder de exclusão de paternidade caso específico ou probabilidade de exclusão individual (IPE – do inglês- *individual probability of exclusion* or *case-specific power of exclusion* (Fung et al., 2002; ver também em Brenner, 2006 e Brenner, 1997a) nem o poder de exclusão de paternidade médio ($\bar{A}$ - em inglês - *average probability of exclusion* or *mean power of exclusion*) (Garber & Morris, 1983; Fung et al., 2002; ver também em Brenner, 2006), probabilidade de inclusão média (IPM), frequência alélica mínima [probabilidade alélica mínima (P_{min}) ou frequência alélica mínima - MAF do inglês *minimum allele frequency* - Buckleton e Curran, 2005; Budowle

et al., 1996; O'Connor et al., 2010; Butler, 2005; NRC II, 1996; Chakraborty, 1992; Weir, 1993a], índice de probabilidade típica [TPI, do inglês, *typical paternity index*] (Brenner & Morris, 1989);

- **Para dados de STRs do cromossomo Y:** frequência alélica, frequência haplotípica (Fh), número de tipos de haplótipos, haplótipos únicos (Schoske et al., 2004 Al-Azem et al., 2017), diversidade alélica ou genica (D(a) ou GD – equivale ao poder de exclusão dos aut-STR – Nei, 1973; Chakravarti, 1983; Kayser et al., 1997; Nei, 1987; Al-Azem et al., 2017; Peng et al., 2020), diversidade haplotípica - correspondente ao índice de discriminação dos aut-STR (D(h)) [Sensabaugh, 1982; Kayser et al., 1997; Al-Azem et al., 2017; Peng et al., 2020], probabilidade de coincidência (PC) e poder ou capacidade discriminatória [CD] (Kayser et al., 1997; Schoske et al., 2004; Al-Azem et al., 2017).

3.5. BIOINFORMÁTICA FORENSE

A bioinformática é uma disciplina que combina a matemática, estatística, ciência da informação, ciência da computação e biologia e ajuda a responder a questões biológicas (Zhang & Liu, 2013; Luscombe et al., 2001; Muntaha et al., 2018; Gauthier et al., 2019). Diferente da biologia computacional, que lida principalmente com modelagem de sistemas biológicos, a bioinformática é composta por dois componentes principais (Zhang & Liu, 2013; Muntaha et al., 2018):

- 1- O desenvolvimento de ferramentas de *software* e algoritmos;
- 2- A análise e interpretação de dados biológicos usando uma variedade de ferramentas de *software* e algoritmos particulares.

A bioinformática e o DNA forense são inerentemente interdisciplinares. Ambas extraem da estatística e da ciência de computação técnicas e métodos e as trazem para os problemas da biologia e para as leis (Butler, 2005; Bianchi & Liò, 2018). Na bioinformática forense além dos termos algoritmos e *software*, a palavra lei também é bem comum não somente em análises de amostras humanas como também não humanas (Kumar et al., 2012; Muntaha et al., 2018).

Ao realizar a identificação forense e para investigar parentesco, um dos principais problemas que um geneticista forense encontra é a análise estatística, que abrange a análise/validação do banco de dados populacional e o cálculo da RV com âmbito criminal ou cível. O uso de computadores nestes casos forenses economiza tempo e esforços e reduz a possibilidade de erros tanto de cálculo como de manipulação dos dados a serem analisados como os resultados [os dados de entrada e de saída- do inglês *input* e *output*, respectivamente (Coble et al., 2016).

3.5.1. Softwares de identificação humana e análises complementares disponíveis na comunidade científica

Embora o cálculo da razão de verossimilhança ou do índice de paternidade possa ser feito manualmente (Brenner, 2004; Egeland et al, 2016) é uma tarefa tediosa e propensa a erros humanos. Diante disto é necessário um *software* dedicado. E em casos mais complexos e simulação, é simplesmente impossível realizar cálculo estatísticos sem um *software* específico (Drábek, 2009; Egeland et al, 2016).

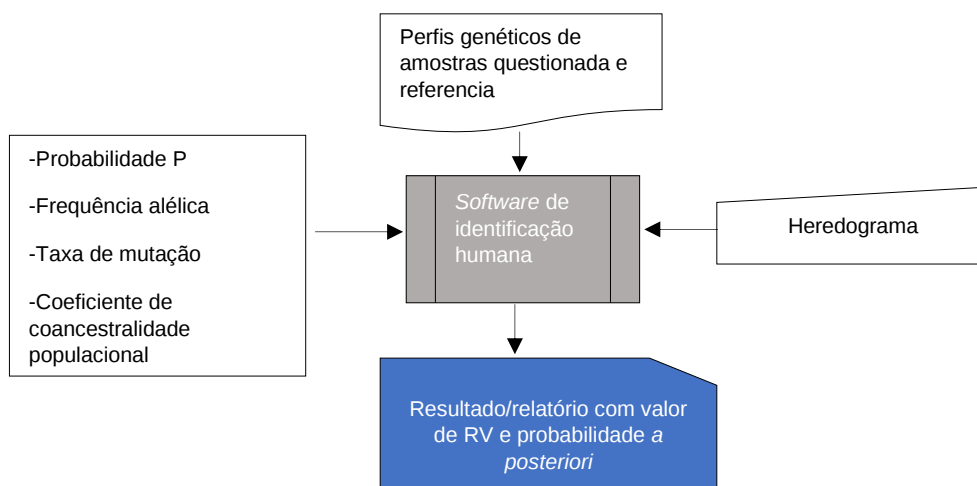
O uso de *software* vem desde as primeiras aplicações de DNA em casos de identificação de requisitados pelos tribunais na Europa e América do Norte, que concediam um prazo muito menor do que era necessário para execução da análise forense sem o uso de um *software* (Buckleton et al., 2005; Liu & Harbison, 2018). Nos

anos 90, época em que fazia manualmente os cálculos dos casos forenses, diante esse curto prazo recorriam a programadores de *software* para auxilia-los (Buckleton et al., 2005; Liu & Harbison, 2018). Atualmente há diversos *softwares* de cálculo estatístico forenses de identificação humana disponíveis na comunidade científica. E o papel desempenhado pelo *software* na genética forense vão além do cálculo estatístico propriamente dito: genotipagem, impressão de relatórios de especialistas, manipulação e validação de banco de dados populacional, comunicação com máquinas e redes (Drábek, 2009; Liu & Harbison, 2018).

Dentre os *softwares* disponíveis para cálculo estatístico de identificação humana (Planilha suplem. 1 e 2), o *software Familias* é o mais utilizado pela comunidade científica internacionalmente e, portanto, é um *software* de referência no momento (Drábek, 2009; Coble et al., 2016). A partir dele é possível fazer a análise estatística de paternidade simples e complexa e identificação forense (incluindo módulo DVI) por dados genotípicos (marcadores aut-STR). Ele permite importar as frequências alélicas e os perfis genotípicos a serem analisados em formato específico.

O *software Familias* assim como os demais funcionam basicamente como mostra a Figura 3:

Figura 3 – Fluxograma de funcionamento dos *softwares* de identificação humana



Fonte: a autora (2021)

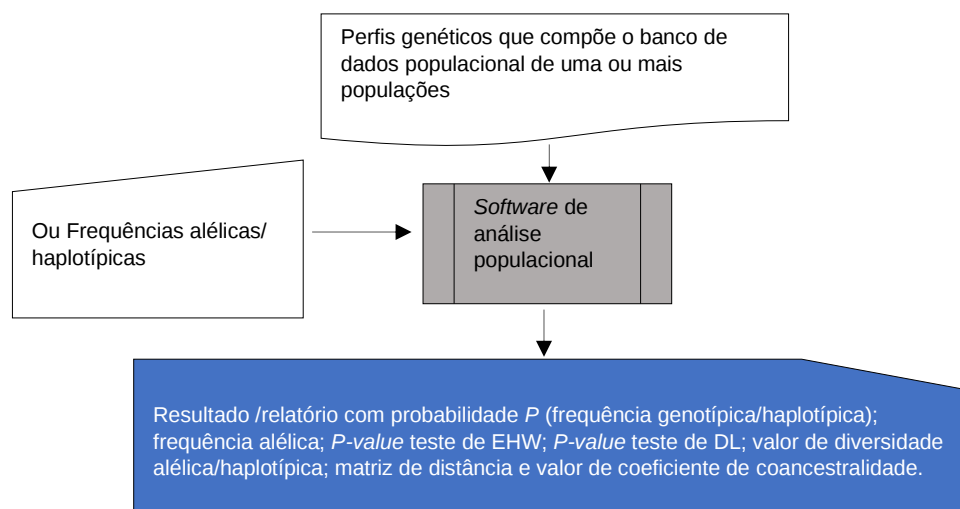
Como mostra a Figura 3 os *softwares* permitem a entrada (representados em branco no fluxograma) dos perfis de DNA, dos dados populacionais relevantes para o cálculo do estatístico da RV e do heredograma (apenas em casos de paternidade/parentesco). Uma vez com tais dados ele fará o gerenciamento de correspondência de perfil genético de amostra questionada com amostra de referência e o cálculo estatístico (representado em cinza). Posteriormente um arquivo de saída é gerado contendo os valores RVs e probabilidade *a posteriori* além de outras informações relevantes referentes aos cálculos estatísticos (representado em azul). Nenhum destes *softwares* tem prova formal de validação de código do *software* [como ITQS ou TickIt (Huber, 1998a)] durante o desenvolvimento. As publicações em periódicos revisados por pares foram consideradas provas suficientes de consistência científica e programática (Dajda & Jung, 2006; Phillips et al., 2008; Egeland et al., 2016).

Além dos *softwares* de identificação humana, há aqueles de análises complementares a partir dos quais é possível realizar a validação de banco de dados populacional. Estes *softwares* estão disponíveis tanto para dados genotípicos (Planilha suplem. 3) como haplotípicos (Planilha suplem. 4). Dentre eles o Arlequin é um dos mais utilizado na genética forense já que tem licença gratuita e disponibiliza análises para ambos tipos de dados. Ele permite a obtenção de frequência alélica/haplotípica e genotípica, diversidade de haplótipo e alélica e permite averiguar o DL, EHW. Este *software* também oferece uma possibilidade de comparação entre banco de dados diferentes gerando matriz de distância genética. Sendo assim é útil também para controle de qualidade, pois é possível uma indicação de falha do sequenciamento que geram alelo nulos, a partir da observação de excesso de homozigotos (isto para dados genotípicos). Isto é bastante útil para verificar se o

banco de dados se adapta ao modelo populacional de independência ou subestruturação populacional.

Os *softwares* utilizados na validação de banco de dados populacional no geral funcionam como na Figura 4. Eles importam perfis genéticos de um banco de dados ou de mais de um banco de dados (como o *Arlequin*) ou ainda dados de frequência alélica/haplotípica (em branco no fluxograma da Figura 4) e emite um relatório com as análises estatísticas solicitadas. Tanto o *input* como o *output* de cada *software* têm um formato de arquivo específico.

Figura 4 – Fluxograma do funcionamento dos *softwares* utilizados na validação de banco de dados populacionais



Fonte: a autora (2021)

3.5.2. Construção e validação de *software* de cálculos bioestatísticos

O *software* é considerado parte integrante do processo de prova e não deve ser tratado como um componente separado e isolado (Leclair et al., 2007; Coble et al., 2016). Embora a fórmula RV para testes de identificação humana seja uma técnica estatística bem conhecida, os produtos de *softwares* comerciais (COTS *softwares* -

do inglês *commercial off-the-shelf*) disponíveis para seu cálculo devem ser suficientemente validados (Drábek, 2009).

A validação do *software* de cálculo é feita para garantir que o mesmo resultado preciso seja obtido cada vez que o cálculo for executado. Essa validação de *software* leva a uma saída mais adequada e maior confiança nos resultados (Drábek, 2009; Coble et al., 2016). Ela é realizada seguindo diretrizes baseadas na ISO 17025 (EM ISO/IEC 17025, 1999; NATA, 2000; ABNT, 2005) disponíveis para o campo de perícia (Gjertson et al., 2007; SWGDAM, 2003; D.N.A. Advisory Board, 1998; Coble et al., 2016), biomédico (Green, 1996) e engenharia de *software* (Huber, 1998a, 1998b, 1998c; IEEE 2012 e 2008). O *software* deve atender aos pré-requisitos gerais de competência de laboratório de ensaio e calibração (EN/ISO/IEC 17025). Vários grupos descreveram exercícios de validação de pacotes de *softwares* internos, de código aberto ou comerciais (Drábek, 2009; Coble et al., 2016; Perlin et al., 2011; Brenner, 1997b; Riancho et al. 2003; Röck et al., 2011; Slooten, 2012; Mitchell et al., 2012; Gill & Haned, 2013; Kling et al., 2014; Puch-Solis & Clayton, 2014; Steele et al., 2014; Cowell et al., 2015; Dørum et al., 2015; Taylor et al., 2015; Bleka et al., 2016).

Apesar dessas publicações variarem em termos de abordagem de validação (Coble et al., 2016), o planejamento da validação pode ser apresentado em uma única figura (Figura 5).

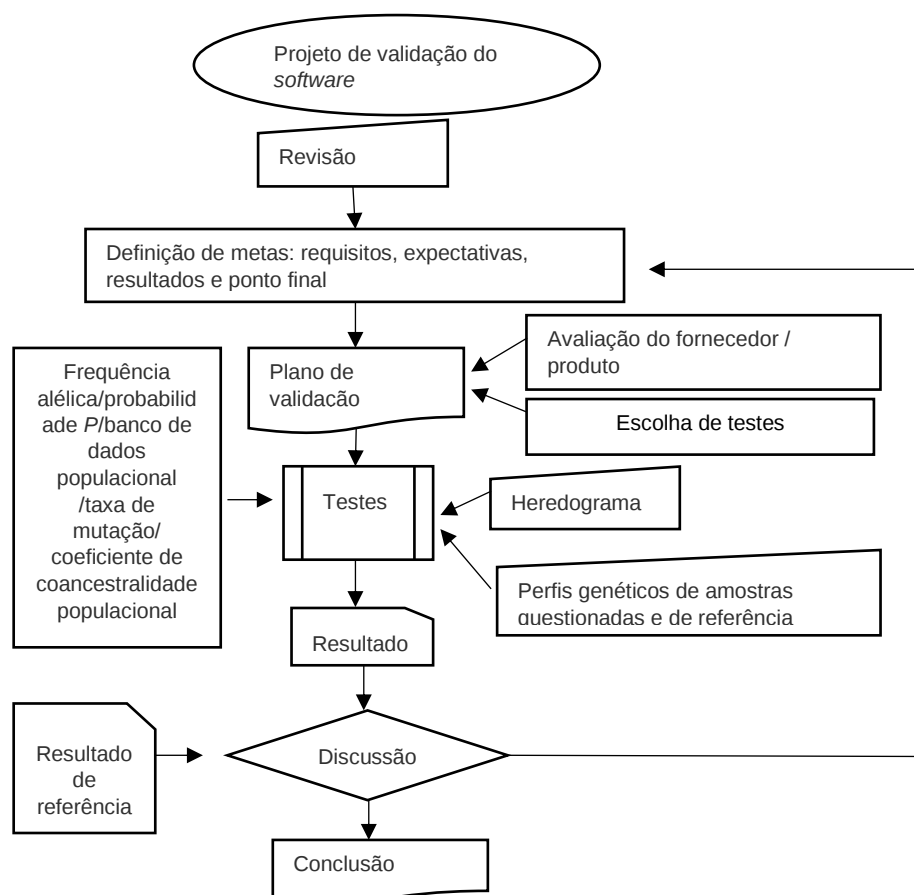
Assim a construção do *software* não se limita apenas a implementação de algoritmos pesquisados na literatura disponível na comunidade científica. Também significa levar em conta recomendações feitas pelos próprios usuários, que são padrões internacionais na genética forense (Drábek, 2009).

Pode-se argumentar que o uso impróprio de *software* é a causa mais frequente de erro do que o próprio *software*. No entanto, mesmo o pessoal mais qualificado pode

obter uma razão de probabilidade incorreta usando uma função de *software* não validada (Drábek, 2009). Os critérios de aceitação universal foram definidos como correspondência de razão de verossimilhança com referência (<http://dna-view.com/>; Brenner, 1993; Ayres, 2000) dentro de 1% de erro ($100 \times (\text{valor referência} - \text{valor do programa a ser validado}) / \text{valor referência}$) enquanto os *inputs* incorretos não danificarem ou interromperem o funcionamento do software e considerando algoritms baseados nos mesmos métodos estatísticos (Huber & Wiederoder, 1997; Coble et al., 2016).

O ponto final é definido como o cumprimento de todos os critérios de aceitação (re) definidos por um programa ou por uma combinação de programas (Drábek, 2009; Coble et al., 2016). Em cima disto se faz a discussão.

Figura 5 - Esboço de plano de validação de software



Fonte: adaptado de Drábek (2009)

3.5.3. Bioinformática e ética

A ética na genética forense também é aplicada na bioinformática e deve ser levado em conta tanto na construção e atualização do *software* como também no uso dele. O tipo de informação e a forma em que os dados são importados e reservados é um dos pontos que chama a atenção.

Normalmente as regiões de DNA utilizadas na genética forense são éxons ou outras não codificadoras. Mas isto não é suficiente quando se fala do impacto que a bioinformática pode trazer a sociedade incluindo direitos de privacidade e valores básicos da sociedade (Bianchi & Liò, 2018):

- A forte ligação da bioinformática com a tecnologia não a torna infalível;
- Cada tecnologia tem diferente sensibilidade e limites;
- Bioinformática não é necessariamente sinônimo de confidencialidade de informação genética. Isto está bastante ligado ao usuário da bioinformática.

Diferente de dados como impressão digital cuja aplicação é limitada apenas para identificação humana, o DNA forense tem muita informação que pode levar ao uso além da identificação, como características pessoais ligadas a aparência física, fisiologia e comportamento (Bianchi & Liò, 2018).

Uma forma de gerenciar a manipulação dos dados seria é a criação de um banco de dados nacional armazenando os perfis de toda a população (Jeffreys et al., 1985a, 1985b, 1991; Jeffreys, 2004; Hakim et al., 2020) cujo órgão levasse em conta declarações de direitos humanos a nível nacional e das nações a nível internacional. Enquanto isto não acontece os requisitos legais de recolhimentos de amostras para teste de DNA. Uma das exigências é que as amostras de DNA sejam tornadas completamente anônimas para o laboratório de teste, de forma que a informação genética não possa ser reconectada com o indivíduo testado e a aprovação do

conselho de revisão institucional [IRB do inglês *Institutional Review Board*] (Butler, 2005; Hakim et al., 2020).

4. ARTIGO 1 - BIGASYSTEM: UM SOFTWARE PARA ESTATÍSTICA DE DNA FORENSE
(A SER SUBMETIDO AO PERIÓDICO FORENSIC SCIENCE INTERNATIONAL: GENETICS)

Abigail Marcelino dos Santos Silva¹

Sérgio de Sá Leitão Paiva Junior^{1,2}

Valdir de Queiroz Balbino¹ (Corresponding author)

¹ Departamento de genética, Universidade Federal de Pernambuco, Prédio do CCS - Centro de Ciências da Saúde, Av. Prof. Moraes Rego, 1235, Cidade Universitária, Recife, Pernambuco, Brasil

² Unidade Acadêmica de Serra Talhada, Universidade Federal Rural de Pernambuco, Av. Gregório Ferraz Nogueira, s/n, Serra Talhada, Pernambuco, Brasil

Destaques:

- BigASystem (dataBase and human Identification Genetic Analysis System) calcula todas as estatísticas-padrão necessárias para dados de populações e evidências forenses de STR autossômicos e do cromossomo Y na identificação forense e teste de paternidade/parentesco.
- Está disponível gratuitamente, requer um arquivo de entrada muito simples e os resultados podem ser baixados como tabelas convenientes.

- Exporta dados genotípicos/haplotípicos e de frequência alélica/haplotípica para o formato específico de um dos softwares de genética de populações utilizados na genética forense.

Abstract

A análise estatística de dados populacionais e de evidências forenses de marcadores microssatélites autossômicos e do cromossomo Y na identificação humana forenses e no teste de paternidade/parentesco é indispensável para a aplicação da genética forense como uma ciência. Por ser uma etapa muito exaustiva e, conseqüentemente, propensa a erros quando realizada manualmente, o uso de softwares já é bastante comum. Apesar de várias alternativas disponíveis, a maioria dos softwares de identificação humana não oferece estatísticas de análises populacionais na perspectiva forense. Caso possuam um banco de dados, os softwares disponíveis são voltados apenas para a análise de dados autossômicos e não são gratuitos, além de serem focados em análises de genética e evolução populacionais. E, até onde é do nosso conhecimento, nenhum dos programas calcula todos os parâmetros de interesse forense para dados STR autossômicos e do cromossomo Y. Apresentamos aqui uma ferramenta BigASystem (dataBase and human Identification Genetic Analysis Systems) que preenche estas lacunas. Esse programa permite a utilização de um arquivo de entrada intuitivo e calcula todos os parâmetros populacionais forenses para dados aut-STR e Y-STRs, além de estimar a razão de verossimilhança e a probabilidade *a posteriori* de testes de identificação forense humana e de paternidade/parentesco mais comuns em laboratórios genéticos forenses.

Palavras chaves: estatística, análise populacional, identificação forense humana, teste de paternidade, STR.

1 INTRODUÇÃO

As análises estatísticas de identificação humana baseadas em locos microsatélites autossômicos (aut-STRs) em equilíbrio de ligação e de Y-STRs (cromossomo Y) necessitam das frequências alélicas e haplótípicas, obtidas e validadas por meio da análise estatística populacional forense, que dependem da obtenção de parâmetros populacionais de interesse forense. Os aut-STR e Y-STR são os marcadores com maior demanda a nível cível e criminal [1-6].

Vários softwares estão disponíveis para a condução destas análises, porém nenhum deles fornece em uma única plataforma a possibilidade de realização de análises estatísticas populacionais e de evidências forenses dos marcadores aut-STR e Y-STRs. Assim os geneticistas forenses lidam com a exaustiva construção de *inputs* específicos para diferentes softwares e com a indisponibilidade de softwares para análises específicas (e.g. análise populacional de dados Y-STR). Softwares como o Forstatistics [7] e Genostat (acessível em <http://www.bioforensics.com/genostat>) estão disponíveis para identificação humana.

Uma alternativa que demandaria um menor número de ferramentas seria através do uso do Genoproof (disponível em <http://www.qualitytype.de/en/qualitytype/genoproof>) e do YHRD [8]. O Genoproof não tem licença gratuita e, apesar de realizar análises de evidências de marcadores aut-STR e Y-STR, este software só inclui as estatísticas para os primeiros e nenhuma para os Y-STRs (limitando-se ao fornecimento de informações de *match* dos perfis evidencial e de referência. O YHRD, por sua vez,

permite a análise estatística de identificação humana para dados Y-STRs, mas baseada no uso de frequências haplotípicas de um banco de dados global, reduzindo, assim, a qualidade das inferências que dependem do conhecimento da estruturação genética de populações locais. Adicionalmente, não disponibiliza ferramenta estatística de parâmetros populacionais forenses.

Recentemente foi publicado o STRAFF [9], ferramenta que supostamente realiza tais análises, mas que efetivamente não calcula os parâmetros ao nível de haplótipo, inclusive as frequências haplotípicas, que são valores de maior interesse na genética forense. O BigASystem, que é aqui apresentado, fornece à comunidade uma ferramenta gratuita e que realiza em uma única plataforma análises populacionais e de evidências genéticas aut-STR e Y-STR de identificação humana forenses e teste de paternidade/parentesco. O analista pode, por exemplo, utilizar *data sets* internos do seu laboratório, sendo mais específico para aplicação em um específico caso forense. É validado por comparação de resultados estatísticos obtidos por outro software ou cálculo manual. O BigASystem pode ser encontrado online em ***http://www.****

*Ainda em construção. Justificativa: aguardando registro do manual do software no INPI. Informações a respeito do software e sua validação com o professor Valdir de Queiroz Balbino chefe do Laboratório de Bioinformática e Biologia Evolutiva do Departamento de Genética da UFPE.

2.RECURSO DO SOFTWARE

O BigASytem é um software baseado em Windows® e desenvolvido em linguagem Python. O software usa um banco de dados implementado em formato CSV para armazenar dados genotípicos/haplotípicos para análise estatística populacional. Também gera relatórios para as estatísticas populacionais e de evidências genéticas, em formato CSV (*comma-separated values*) e Excel®, para análises populacionais, e

no formato WORD® para análise de evidências genéticas. Adicionalmente, o software gera um arquivo no formato específico do Arlequin [10] a partir do banco de dados de genótipos de marcadores aut-STR e Y-STRs, assim como também de dados de frequências alélicas ou haplotípicas em formato CSV. O Arlequin [10] é um dos softwares de genética de populações mais utilizados para análises mais aprofundadas dos dados populacionais de interesse forense. O BigASystem pode ser integrado ao WORD® e EXCEL® (independentemente da versão). São programas bastantes conhecidos e que fornecem uma interface editável do relatório para que vários detalhes (e.g. número do caso; identificação do laboratório no qual a análise está sendo realizada e o nome dos analistas envolvidos) possam ser fornecidos no relatório consolidado.

Arquivos *inputs* e banco de dados modelos estão disponíveis na página *web* já apresentada. O BigASystem é compatível com alelos com repetições parciais, ou seja, microvariantes. No banco de dados os dados ausentes devem ser indicados com um espaço vazio.

A) PARÂMETROS POPULACIONAIS FORENSES

O BigASystem estima a homozigotidade esperada e observada (H_e e H_o , respectivamente), além do PIC e de outros parâmetros forenses (e.g. PE, PM, PD e TPI) para dados populacionais de marcadores aut-STRs. Os algoritmos para estes cálculos foram baseados nas fórmulas de Huston [11] e Butler [12]. A estatística para dados Y-STR permite a obtenção do valor de diversidade haplotípica (equação análoga à da diversidade alélica), probabilidade de correspondência aleatória (RMP) e da capacidade de discriminação (DC) [13-14]. A fórmula de diversidade alélica [15]

implementada no BigASystem foi $D(A) = \left(\frac{n}{n-1}\right) \times (1 - \sum_{i=1}^n p_i^2)$, considerando um loco com n alelos, representados por $i=\{1, 2, \dots, n\}$ e com frequências alélicas p_i .

O BigASystem calcula as frequências alélicas e haplotípica (Y-STR) por método de contagem [16] para todos os locos do arquivo de entrada e retorna uma tabela completa contendo também os parâmetros populacionais forenses em um formato CSV e MS Excel.

B) RAZÃO DE VEROSSIMILHANÇA DE TESTE DE IDENTIFICAÇÃO POR DADOS AUT-STR

O BigASystem permite a obtenção do índice de paternidade (IP), índice acumulado de paternidade (IPA) e probabilidade de paternidade (PP) para os tipos de casos mais comum de rotina forense: trio simples (suposto pai, mãe e filho alegado), duo (suposto pai e filho alegado), meios-irmãos e pai falecido por índice avuncular com ou sem a mãe. O software calcula estes parâmetros mediante as fórmulas de Fimmers et al. [17-19] e Essen-Möller [18,20-22] considerando o modelo mutacional *uniform* [23] e frequência alélica mínima (MAF) [24]. O BigASystem também considera um algoritmo adicional conversão de heredograma em rede imaginária [25] e as descrição verbal de IPA segundo predicado verbal de Hummel [26].

São necessários dois inputs: um contendo os perfis genéticos com suas distintas identificações e outro com cada membro do caso com seus ascendentes imediatos em formato CSV. Já para obtenção dos valores de Razão de verossimilhança de identificação humana, o BigASystem se baseia em Balding & Nichols [27] com interpretação verbal das conclusões [28-30]. É necessária a importação dos perfis de evidencial e de referência considerado em formato CSV. Um relatório é emitido com os valores estatísticos e perfis genéticos investigados, nos formatos CSV, MS Excel e

Word. Os valores de frequência podem ser obtidos a partir da análise do próprio banco de dados realizado BigASystem ou pode ser de forma direta nos formatos CSV.

C) RAZÃO DE VEROSSIMILHANÇA DE TESTE DE PARENTESCO POR LINHAGEM Y-STR

A fórmula implementada no BigASystem para obtenção do IP é baseada em modelo de população subestruturada [31-33] com representação verbal [33-34]. É necessária a importação dos perfis genéticos de amostra evidencial e de referência. As frequências haplotípicas podem ser calculadas pelo próprio BigASystem a partir de banco de dados construído previamente ou importação direta no formato CSV. O output é análogo ao falado anteriormente.

3 TESTE DE VALIDAÇÃO

Os resultados gerados das análises populacionais de dados de 20 marcadores aut-STR (N=540) e 17 Y-STRs (N=202) com o BigASystem utilizando banco de dados construídos a partir de dados obtidos no Laboratório de Bioinformática e Biologia Evolutiva da Universidade Federal de Pernambuco (LABBE-UFPE) foram comparados com resultados obtidos pelo software STRAF (v.1.0.5. on-line) e cálculos manuais e foram satisfatórios com porcentagem de erro de 0 a 8% para todos os parâmetros. Os P-values do Equilíbrio de Hardy-Weinberg foram concordantes, já que todas as comparações assumiram valores acima do nível de significância de $\alpha=0,05$. Os valores de RV e PP tanto para os testes de paternidade/parentesco por evidência aut-STRs (incluindo modelo mutacional e MAF) e Y-STRs e para os testes de identificação (erro percentual de 0 a 4.9%) foram concordantes com resultados obtidos a partir de do uso do software Familias [35], cálculo manual e Genostat, respectivamente.

4 LIMITAÇÕES

O BigASystem foi desenvolvido para verificar conjuntos de dados em genética forense antes da publicação ou uso em estudos de caso. Não é seu propósito fornecer análises aprofundadas de dados genéticos para projetos de pesquisa genética, para os quais softwares mais robustos, tais como Arlequin [10] ou Genepop [36], devem ser considerados. Em testes de reconhecimento de paternidade complexo por evidências aut-STRs, o BigASystem só realiza testes envolvendo um parente paterno por vez. O BigASystem no momento não é aplicado para análise de outros marcadores de interesse forenses, tais como X-STRs e sequências de mtDNA.

5 FUTUROS DESENVOLVIMENTOS

São bem-vindas sugestões de desenvolvimento e relatório de Bugs, que podem ser dirigidos aos autores. Na segunda versão (em desenvolvimento), implementaremos uma interface de usuário, uma rede Bayesiana para casos de paternidade forense, inclusão de mais modelos mutacionais nas funções de teste de paternidade/parentesco, além de modelo de alelo nulos. No teste de parentesco por linhagem Y-STR, pretendemos implementar o algoritmo de Laplace, incluindo modelo mutacional. Além disso, pretende-se implementar modelo de alelo nulo para identificação forense por evidências para aut-STRs. Pretendemos incluir outros tipos de marcadores genéticos de uso forense, tais como X-STR, sequências mt-DNA, SNPs e INDELS, além de uma estatística para obtenção de RVs a partir de dois ou mais tipos de marcadores associados.

6 REFERÊNCIAS

- [1] D. JHA, J. RIJAL, B. TULADHAR, B. POKHAREL, Mutations or Exclusion: An Uncommon Parentage Assessment Case, *Scientific World*, 11(11) (2013) 74-76.
- [2] E. Niedzwiecki, S. Debus-Sherrill, M.B. Field, Understanding Familial DNA Searching: Coming to a Consensus on Terminology. National Institute of Justice, Office of Justice Programs, U.S. Department of Justice. Fairfax, VA: ICF International, 2017.
- [3] M.E.G. Aceves, O.R. Rentería, X.X. Díaz-Navarro, H. Rangel-Villalobos, Paternity tests in Mexico: Results obtained in 3005 cases, *Journal of forensic and legal medicine*, 55 (2018) 1–7.
- [4] R. Dumache, A. Enache, L. Barbarii, C. Constantinescu, A. Pascalau, C. Jinca, S. Arghirescu, Chimerism Monitoring by Short Tandem Repeat (STR) Markers in Allogeneic Stem Cell Transplantation, *Clinical laboratory*, 64(9) (2018) 1535–1543.
- [5] J.A. Siegel, P.J. Saukko, M.M. Houck, *Encyclopedia of Forensic Sciences*, 2ed, Academic Press, 2013.
- [6] H. Mahmood, N. Salman, K. Salih, D. Hasan, M. Al-Zubaidi, Frequency of Y-chromosome STRs using PowerPlex® Y23 System in Iraqi population, *Egyptian Journal of Forensic Sciences* 10 (2020) 12.

- [7] N. Rasool, W. Hussain, ForeStatistics: A windows-based feature-rich software program for performing statistics in forensic DNA analysis, paternity and relationship testing, *Forensic Science International* 307 (2020) 110142.
- [8] S. Willuweit, L. Roewer, International Forensic Y Chromosome User Group (2007). Y chromosome haplotype reference database (YHRD): update, *Forensic science international: Genetics* 1(2) (2007) 83–87.
- [9] A. Gouy, M. Zieger, STRAF-A convenient online tool for STR data evaluation in forensic genetics, *Forensic science international: Genetics* (30) (2017) 148–151.
- [10] L. Excoffier, H.E.L. Lischer, Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows, *Molecular ecology resources* 10(3) (2010) 564-7.
- [11] K.A. Huston, Statistical analysis of STR data, *Profiles DNA*, (1998), pp. 14–15.
- [12] J.M. Butler, Forensic DNA testing, *Cold Spring Harb. Protoc.* 6 (2011) 1438–1450.
- [13] M. Kayser, A. Caglià, D. Corach, et al., Evaluation of Y-chromosomal STRs: a multicenter study, *Int J Legal Med.* 110(3) (1997) 125–133.
- [14] R. Schoske, P.M. Vallone, M.C. Kline, et al., High throughput Y-STR typing of U.S. populations with 27 regions of the Y chromosome using two multiplex PCR assays, *Forensic Sci Int.* 139(2–3) (2004) 107–121.
- [15] M. Nei, *Molecular Evolutionary Genetics*. 1st ed. Columbia University Press, USA. 1987.
- [16] M.M. Holland, T.J. Parsons, Mitochondrial DNA sequence analysis - validation and use for forensic casework, *Forensic Sci Rev.* 11(1) (1999) 21–50.

- [17] R. Fimmers, L. Henke, J. Henke, M. Baur, How to deal with mutations in DNA-testing, in: C. Rittner, P.M. Schneider (Eds.), *Advances in Forensic Haemogenetics*, vol. 4, Springer-Verlag, Berlin, 1992, pp. 285–287.
- [18] D.W. Gjertson, C.H. Brenner, M.P. Baur, A. Carracedo, F. Guidet, J.A. Luque, R. Lessig, W.R. Mayr, V.L. Pascali, M. Prinz, P.M. Schneider, N. Morling, ISFG: Recommendations on biostatistics in paternity testing, *Forensic Sci. Int. Genetics* 1(3) (2007) 223-231.
- [19] M.D. Coble, J. Buckleton, J.M. Butler, T. Egeland, R. Fimmers, P. Gill, L. Gusmão, B. Guttman, M. Krawczak, N. Morling, W. Parson, N. Pinto, P.M. Schneider, S.T. Sherry, S. Willuweit, M. Prinz, DNA Commission of the International Society for Forensic Genetics: Recommendations on the validation of software programs performing biostatistical calculations for forensic genetics applications, *Forensic Science International: Genetics*, 25 (2016) 191-197.
- [20] E. Essen-Möller, Die Beweiskraft der Ähnlichkeit im Vaterschaftsnachweis - theoretische Grundlagen, *Mitt. Anthropol. Ges. (Wien)* 68 (1938) 9–53.
- [21] E. Essen-Möller, C. Quensel, Zur Theorie des Vaterschaftsnachweises aufgrund von Ähnlichkeitsbefunden, *Z. Ges. Gerichtl. Med.* 31 (1939) 70–96.
- [22] C. MORIMOTO, S. MANABE, T. KAWAGUCHI, C. KAWAI, S. FUJIMOTO, Y. HAMANO, R. YAMADA, F. MATSUDA, K. TAMAKI, Pairwise Kinship Analysis by the Index of Chromosome Sharing Using High-Density Single Nucleotide Polymorphisms, *PloS one*, 11(7) (2016) e0160287.
- [23] DAWID, A. P.; MORTERA, J.; PASCALI, V. L. Non-fatherhood or mutation? A probabilistic approach to parental exclusion in paternity testing. **Forensic Science International**, v.124, n.1, p. 55-61, 2001.

- [24] K.L. O'Connor, S.P. Myers, E.L.R. Butts, C.R. Hill, J.M. Butler, P.M. Vallone, Candidate Reference Family Data: A Tool for Validating Kinship Analysis Software. Poster #35, 21th International Symposium on Human Identification San Antonio, TX, 11–14 (2010). Available in: http://www.cstl.nist.gov/biotech/strbase/pub_pres/Promega2010_OConnor.pdf
- [25] N. Tinker, D. Mather, KIN: Software for Computing Kinship Coefficients. The Journal of Heredity, 84(3) (1993).
- [26] K. Hummel, P. Ihm, V. Schmidt, Biostatistical opinion of parentage—Biostatistische Abstammungsbegutachtung. Table v. 1, part III—Tabellenband 1, 3. Stuttgart: Teil. Fischer, 1971.
- [27] D.J. Balding, R.A. Nichols, A method for quantifying differentiation between populations at multi-allelic loci and its implications for investigating identity and paternity, Genetica, 96 (1995) 3–12.
- [28] SWGDAM - Scientific Working Group on DNA Analysis Methods. Recommendations of the SWGDAM Ad Hoc Working Group on Genotyping Results Reported as Likelihood Ratios, 2017. Available in: <https://www.swgdam.org/publications> and notified in <https://www.swgdam.org/faq> em 2018.
- [29] C.G.G. AITKEN, F. TARONI, A verbal scale for the interpretation of evidence, Science Justice, 38 (1998) 279–281.
- [30] DJ-USA- Department of Justice. Uniform Language for Testimony And Reports For Forensic Autosomal DNA Examinations Using Probabilistic Genotyping Systems - Autosomal Dna Examinations Using Probabilistic Genotyping Systems Ultr, U.S.

Department of Justice, 2018. Available in: <https://www.justice.gov/olp/page/file/1095976/>.

[31] D.J. Balding, R.A. Nichols, DNA profile match probability calculation: how to allow for population stratification, relatedness, database selection and single bands, *Forensic Science International*, 64 (1994) 125–140.

[32] SWGDAM. Interpretation Guidelines for Y-Chromosome STR Typing. 2014. Available in: http://swgdam.org/SWGDAM_YSTR_Guidelines_APPROVED_01092014_v_02112014_FINAL.pdf

[33] L. Roewer, M.M. Andersen, J. Ballantyne, J.M. Butler, A. Caliebe, D. Corach, M.E. D'Amato, L. Gusmão, Y. Hou, P. Knijff, W. Parson, M. Prinz, P.M. Schneider, D. Taylor, M. Vennemann, S. Willuweit, DNA commission of the International Society of Forensic Genetics (ISFG): Recommendations on the interpretation of Y-STR results in forensic analysis, *Forensic Science International: Genetics*, 48 (2020) 102308.

[34] DJ-USA- Department of Justice. Uniform Language for Testimony and Reports for Forensic YSTR Data Examinations, U.S. Department of Justice, 2020. URL: <https://www.justice.gov/olp/page/file/1095976/>.

[35] D. Kling, A.O. Tillmar, T. Egeland, Familias 3-Extensions and new functionality, *Forensic Science International: Genetics*, 13 (2014) 121-127.

[36] F. Rousset, genepop'007: a complete re-implementation of the genepop software for Windows and Linux, *Molecular ecology resources* 8(1) (2008) 103-106.

5. REGISTRO DE PROGRAMA DE COMPUTADOR BigASystem

Quando: 14/04/2021.

Como: Via e-mail.

Com quem: Diretoria de Inovação e Empreendedorismo da Universidade Federal de Pernambuco – POSITIVA -UFPE, intermediária do pedido de registro no INPI (Instituto Nacional da Propriedade Industrial), conforme as instruções em <https://www.ufpe.br/positiva>.

Autores: Abigail Marcelino dos Santos Silva, Sérgio de Sá Leitão Paiva Junior e Valdir de Queiroz Balbino.

Função do programa: Banco de dados de genótipos e haplótipos bem como sua validação, e análises estatísticas de identificação forense e de parentesco humanos, incluindo linhagem.

Tipo de dados que analisa: genótipos de microssatélites autossômicos e haplótipos de microssatélites do cromossomo Y e X e sequências mitocondriais.



REPÚBLICA FEDERATIVA DO BRASIL

MINISTÉRIO DA ECONOMIA

INSTITUTO NACIONAL DA PROPRIEDADE INDUSTRIAL

DIRETORIA DE PATENTES, PROGRAMAS DE COMPUTADOR E TOPOGRAFIAS DE CIRCUITOS INTEGRADOS

Certificado de Registro de Programa de Computador

Processo Nº: **BR512021001362-0**

O Instituto Nacional da Propriedade Industrial expede o presente certificado de registro de programa de computador, válido por 50 anos a partir de 1º de janeiro subsequente à data de 07/03/2021, em conformidade com o §2º, art. 2º da Lei 9.609, de 19 de Fevereiro de 1998.

Título: BigASystem

Data de criação: 07/03/2021

Titular(es): UNIVERSIDADE FEDERAL DE PERNAMBUCO

Autor(es): ABIGAIL MARCELINO DOS SANTOS SILVA; SÉRGIO DE SÁ LEITÃO PAIVA JÚNIOR; VALDIR DE QUEIROZ BALBINO

Linguagem: PYTHON

Campo de aplicação: BL-02

Tipo de programa: AP-01; AT-01; TC-01

Algoritmo hash: SHA-512

Resumo digital hash:

05d394915d4805b3d805b1bd63f346ac1c69cf570ca9c2953744555da0a9f1b29c3c6e9a0d0c7ebe78509d468ad3b2994a411edefb1bc73105d0b2af8c2ad5ab

Expedido em: 29/06/2021

Aprovado por:

Carlos Alexandre Fernandes Silva
Chefe da DIPTO

6. CONCLUSÕES

A importância da identificação pessoal e da verificação de vínculo genético entre indivíduos é inquestionável nas áreas social, jurídica e de segurança pública. A biologia molecular trouxe grandes contribuições para isto. A identificação de bebês, de pessoas desaparecidas, de restos mortais de indivíduos mortos em desastres naturais, atentados terroristas ou guerras, além do reconhecimento da paternidade e maternidade, a possibilidade de criação e utilização de banco de dados para estudos populacionais, comprovam a importância e a variedade de situações nas quais a identificação genética humana é utilizada.

A análise destes dados é um processo lento e exaustivo, não apenas pela execução da estatística em si como também pelos aparecimentos de algumas situações como ausência dos genitores ou parentes diretos da criança cuja paternidade está sendo questionada, ou ainda presença de inconsistências. Ao analisar estas inconsistências deve ser levada em consideração a possibilidade da ocorrência de mutação genética. Além das questões da rotina, ainda tem a análises da pesquisa, as quais lidam com grande quantidade de dados, inviabilizando serem feitas por métodos não automatizados.

A solução apresentada nesse trabalho mostrou que é possível em uma única plataforma (BigASystem) fornecer ferramenta para armazenamento e manutenção de banco de dados aut-STR e Y-STRs compatível com aplicativos de suporte como o Excel, MySQL e Postgres. Nesta mesma plataforma são fornecidas ferramentas para validação destes bancos e consequentemente frequências alélicas/haplotípicas/genotípicas indispensáveis para a estatística Bayesiana dos testes de identificação humana, também disponibilizados.

O BigASystem mostra que é possível realizar a análise dos casos de rotina com praticidade e credibilidade. Como os códigos são baseados em funções redundantes, traz confiabilidade ao processo de cálculo das razões de verossimilhança, além de trabalhar com formatos de arquivos compatíveis com os que normalmente são utilizados na rotina, e são compatíveis com qualquer sistema operacional. Além de não exigir conhecimentos profundos em probabilidade, apesar de serem necessários para que suas funções do aplicativo sejam utilizadas da melhor forma possível.

Nesta tese foram considerados algoritmos que formalizam todo o processo necessário para se realizar a análise probabilística na rotina de teste de paternidade mais comuns e teste de identificação forense. Os resultados obtidos foram satisfatórios sob vários pontos de vista. O aplicativo atingiu o objetivo de automatizar as análises estatísticas a partir de sua representação através de heredogramas e arquivos de importação compatíveis com os normalmente utilizados no laboratório além de permitir facilidade de manipulação de dados populacionais. Também estão dentro das normas de construção de *software* apoiadas pela ISFG e ISO17025 tanto em relação a construção dos algoritmos em si como também em relação a realização de validações.

Já quanto às funções de validação de banco de dados populacional, o *software BigASystem* mostrou que possível a realização de análises tanto intra como interpopulacionais com maior praticidade. O *software* de fato automatizou muitos procedimentos que eram feitos manualmente, sujeitos a erros humanos.

Em um só *software* se têm a possibilidade fazer as análises de paternidade, identificação e de genética de populações. A perspectiva é melhorar, colocando

uma interface do usuário estruturada da forma democrática possível, aumentar as ferramentas de análises e ainda ampliar o número de análises populacionais.

A validação do BigASystem foi realizada com êxito de forma que pode ser utilizado na pesquisa e na rotina. Porém, é importante validações adicionais, assim como críticas e opiniões para serem consideradas nas próximas versões do software.

O BigASystem será disponibilizado em uma página web acompanhado de um tutorial e uma apostila (estão sendo providenciados) e email para contato.

Pretende-se futuramente também permitir a interação dos diferentes tipos de dados de motivos de STRs tanto nas análises de rotina como nas de pesquisas forenses e até mesmo pesquisas mais profundas nos aspectos de genética de populações, estender a aplicação do software para outros tipos de dado, como X-STRs, sequencias mt-DNA, INDELS, SNPs e outros dados genéticos a nível de sequenciamento de última geração, tanto nas análises de rotina como nas de pesquisa.

No Brasil vê-se que ainda há uma grande deficiência de um sistema automatizado inclusive de pesquisa populacional na genética forense. Os valores das frequências alélicas, por exemplo, por si só têm um grau de incerteza já que são estimados a partir de uma amostra da população. Imagine se, além disto, ainda houver erros na estimação destes valores. Pode levar a sobre- ou subestimação das frequências alélicas. E isto pode trazer um impacto significativo em análises que dependem destes valores como estudos na área de genética de populações propriamente ditas ou análises de áreas estritamente dependentes de tais valores populacionais como a genética forense.

A genética forense utiliza o conceito da ciência como argumento da credibilidade de suas análises. A crença não faz parte da ciência e não se encaixa na metodologia científica. A qualidade nas análises quer seja na biologia molecular ou estatística devem existir de modo que todas elas sejam reprodutíveis.

REFERÊNCIAS BIBLIOGRÁFICAS

AABB - AMERICAN ASSOCIATION OF BLOOD BANKS, EUA. **Annual Report Summary for Testing in 2013**. 11 ed. Bethesda, Maryland: Prepared by the Relationship Testing Program Unit. Standards for Relationship Testing Laboratories, 2013. Disponível em: <<http://www.aabb.org/sa/facilities/Pages/relationshipreports.aspx>>. Acesso em: 28 ago. 2020.

AABB - AMERICAN ASSOCIATION OF BLOOD BANKS, EUA. **Annual Report Summary for Testing in 2003**. Prepared by the Relationship Testing Program Unit, 2003. Disponível em: <<http://www.aabb.org/sa/facilities/Pages/relationshipreports.aspx>>. Acesso em: 30 mar. 2020.

AALBERS, S. E.; HIPPI, M. J.; KENNEDY, S. R.; WEIR, B. S. Analyzing population structure for forensic STR markers in next generation sequencing data. **Forensic science international. Genetics**, v.49, p.102364, 2020. doi:10.1016/j.fsigen.2020.102364

ABNT - ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **ABNT NBR ISO/IEC 17025: Requisitos gerais para a competência de laboratórios de ensaio e calibração**. Rio de Janeiro, 2005.

ACEVES, M. E. G.; ROMERO RENTERÍA, O.; DÍAZ-NAVARRO, X. X.; RANGEL-VILLALOBOS, H. Paternity tests in Mexico: Results obtained in 3005 cases.

Journal of forensic and legal medicine, v.55, p.1–7, 2018.

<https://doi.org/10.1016/j.jflm.2018.02.003>

AITKEN, C.; TARONI, F. Interpretation of scientific evidence. **Science Justice**, v.36, p.290–292, 1996.

AITKEN, C. G. G. The use of statistics in forensic science. **Journal of the Forensic Science Society**, v.27, p.113–115, 1987.

AITKEN, C. G. G.; TARONI, F. A verbal scale for the interpretation of evidence. **Science Justice**, v.38, p.279–281, 1998.

AL-AZEM, M.; EL ANDARI, A.; MANSOUR, I. (2017) Estimation of Allele and Haplotype Frequencies for 23 YSTR Markers in the Lebanese Population. **Forensic Research & Criminology International Journal**, v.5, n.2, p.00150. DOI: 10.15406/frcij.2017.05.00150

ALLEN, R. W. & PRITCHARD, J. Parentage Testing and Kinship Analysis, in: Payne-James, J. & Byard, R. W. (eds.), **Encyclopedia of Forensic and Legal Medicine**, Elsevier: 2ed, 2016.

AMBERS, A.; VOTRUBOVA, J.; VANEK, D.; SAJANTILA, A.; BUDOWLE, B. Improved Y-STR typing for disaster victim identification, missing persons investigations, and historical human skeletal remains. **International journal of legal medicine**, v.132, n.6, p.1545–1553, 2018. doi: 10.1007/s00414-018-1794-8

AMORIM, A.; CRESPILO, M.; LUQUE, J. A.; PRIETO, L.; GARCIA, O.; GUSMÃO, L.; ALER, M.; BARRIO, P. A.; SARAGONI, V. G.; PINTO, N.

Formulation and communication of evaluative forensic science expert opinion-A GHEP-ISFG contribution to the establishment of standards. **Forensic science international. Genetics**, v.25, p.210–213, 2016. doi:10.1016/j.fsigen.2016.09.003

ANDERSEN, M. M. & BALDING, D. J. Y-profile evidence: Close paternal relatives and mixtures. **Forensic science international. Genetics**, v.38, p.48–53, 2019. doi:10.1016/j.fsigen.2018.10.004

ANDERSEN, M. M. & D.J. BALDING. How convincing is a matching Y-chromosome profile? **PLoS Genetic**, v.13, n.11, p.1007028, 2017.

ANDERSEN, M. M.; CALIEBE, A.; KIRKEBY, K.; KNUDSEN, M.; VIHRS, N.; CURRAN, J. Estimation of Y haplotype frequencies with lower order dependencies **Forensic Science International: Genetics**, v.46, p.102214, 2020.

ANDERSEN, M. M.; CURRAN, J.; DE ZOETE, J.; TAYLOR, D.; BUCKLETON, J. Modelling the dependence structure of Y-STR haplotypes using graphical models. **Forensic Science International: Genetics**, v.37, p.29–36, 2018.

ANDERSEN, M. M.; ERIKSEN, P.S.; MORLING, N. The discrete Laplace exponential family and estimation of Y-STR haplotype frequencies. **Journal of Theoretical Biology**, v.329, p.39–51, 2013.

ANSLINGER, K.; WEICHHOLD, G.; KEIL, W.; BAYER, B.; EISENMENGER, W. Identification of the skeletal remains of Martin Bormann by mtDNA analysis.

International Journal of Legal Medicine, v.114, n.3, p.194-196, 2001.

doi:10.1007/s004140000176

ASPLEN, C.; LANE, S. A. Considerations in the development of DNA databases.

Interfaces, v.36: p.1–2, 2003.

AYRES, K. L. Relatedness testing in subdivided populations. **Forensic Science**

International, v.114, p.107–115, 2000.

BALDING, D. J.; DONNELLY, P. Inference in forensic identification. **Journal of**

the Royal Statistical Society, v.158, p.21–53, 1995a.

DOI:10.2307/2983402.

BALDING, D. J.; DONNELLY, P. Inferring identity from DNA profile evidence.

Proceedings of the National Academy of Sciences, USA, v.92, p.11741–

11745, 1995b. DOI:10.1073/pnas.92.25.11741.

BALDING, D. J.; DONNELLY, P. **The prosecutor's fallacy and DNA evidence.**

Criminal Law Review, GenEthx: Genetics and Ethics Database, Bioethics

Research Library of the Kennedy Institute of Ethics, 1994b. p.711–721.

BALDING, D. J.; DONNELLY, P.; NICHOLS, R. A. [DNA Fingerprinting: A Review

of the Controversy]: Comment: Some Causes for Concern about DNA

Profiles. **Statistical Science**, v.9, n.2, p.248-251, 1994.

doi:10.1214/ss/1177010489.

BALDING, D. J.; DORMELLY, P. How convincing is DNA evidence? **Nature**,

v.368, p.285-286, 1994a. DOI:10.1038/368285a0.

- BALDING, D. J.; NICHOLS, R. A. A method for quantifying differentiation between populations at multi-allelic loci and its implications for investigating identity and paternity. **Genetica**, v.96, p.3–12, 1995. DOI:10.1007/BF01441146.
- BALDING, D. J.; NICHOLS, R. A. DNA profile match probability calculation: how to allow for population stratification, relatedness, database selection and single bands. **Forensic Science International**, v.64, p.125–140, 1994. DOI: 10.1016/0379-0738(94)90222-4.
- BALLANTYNE, J. Mass disaster genetics. **Nature Genetics**, v.15, p.329–331, 1997. DOI: 10.1038/ng0497-329.
- BALLANTYNE, J. Serology: overview, in: Siegel, J.A.; Saukko, P.J.(Eds.), **Encyclopedia of Forensic Sciences**. London: Academic Press, 2000. p.1322–1331.
- BALLANTYNE, K.N.; KEERL, V.; WOLLSTEIN, A.; CHOI, Y.; ZUNIGA, S.B.; RALF, A.; VERMEULEN, M.; DE KNIJFF, P.; KAYSER, M. A new future of forensic Y-chromosome analysis: rapidly mutating Y-STRs for differentiating male relatives and paternal lineages. **Forensic science international. Genetics**, v.6, n.2, p.208-218, 2012.
- BERRY, D. A. Probability of Paternity, in: Aitken, C. G. G.; Stoney, D. A. (eds), **The Use of Statistics in Forensic Science** New York: Ellis Horwood, 1991. p.150-156.
- BERRY, D. A.; GEISSER, S. Inference in cases of disputed paternity, in: DeGroot, M. H. et al., **Statistics and the Law**. New York: John Wiley and Sons, 1986. p.353–382.

BINI, C.; CECCARDI, S.; LUISELLI, D.; FERRI, G.; PELOTTI, S.; COLALONGO, C.; FALCONI, M.; PAPPALARDO, G. Different informativeness of the three hypervariable mitochondrial DNA regions in the population of Bologna (Italy). **Forensic Science International**, v.135, n.1, p.48–52, 2003. DOI:10.1016/s0379-0738(03)00167-1.

BLEKA, Ø.; STORVIK, G.; GILL, P. EuroForMix: an open source software based on a continuous model to evaluate STR DNA profiles from a mixture of contributors with artefacts. **Forensic Science International: Genetics**, v.21, p.35–44, 2016.

BOCA, S. M.; HUANG, L.; ROSENBERG, N. A. On the heterozygosity of an admixed population.

BODNER, M.; BASTISCH, I.; BUTLER, J. M.; FIMMERS, R.; GILL, P.; GUSMÃO, L.; MORLING, N.; PHILLIPS, C.; PRINZ, M.; SCHNEIDER, P. M.; PARSON, W. Recommendations of the DNA Commission of the International Society for Forensic Genetics (ISFG) on quality control of autosomal Short Tandem Repeat allele frequency databasing (STRidER). **Forensic science international: Genetics**, v.24, p.97–102, 2016. DOI:10.1016/j.fsigen.2016.06.008.

BODNER, M.; BASTISCH, I.; BUTLER, J. M.; FIMMERS, R.; GILL, P.; GUSMÃO, L.; MORLING, N.; PHILLIPS, C.; PRINZ, M.; SCHNEIDER, P. M.; PARSON, W. Recommendations of the DNA Commission of the International Society for Forensic Genetics (ISFG) on quality control of autosomal Short Tandem Repeat allele frequency databasing (STRidER). **Forensic science**

international. **Genetics**, v.24, p.97–102, 2016.
doi:10.1016/j.fsigen.2016.06.008

BOTSTEIN, D.; WHITE, R.L.; SKOLNICK, M.; DAVIS, R.W. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. **The American Journal of Human Genetics**, v.32, n.3, p.314-31, 1980.

BRADBURD, G. S.; RALPH, P. L.; COOP, G. M. A Spatial Framework for Understanding Population Structure and Admixture. **PLoS Genetic**, v.12, n.1, p.e1005703, 2016. doi:10.1371/journal.pgen.1005703

BRADBURD, G.S.; RALPH, P.L.; COOP, G. M. A Spatial Framework for Understanding Population Structure and Admixture. **PLoS Genetic**, v.12, n.1, p. e1005703, 2016. doi:10.1371/journal.pgen.1005703

BRENNER, C. H. A note on paternity computation in cases lacking a mother. **Transfusion**, v.33, p.51–54, 1993. DOI:10.1046/j.1537-2995.1993.33193142310.x

BRENNER, C. H. **How to Solve Any Kinship Problem by Hand [Online]**, 2004, p. 1–4. Disponível em: <<http://dna-view.com/index.html>>. Acesso em: 10 jan. 2020.

BRENNER, C. H. Symbolic kinship program. **Genetics**, v.145, p.535–542, 1997b. PMID: 9071605.

BRENNER, C. **Mutations in paternity**, 2006. Disponível em: <<http://dna-view.com/mudisc.htm>>. Acesso em 12 mar. 2020.

BRENNER, C. **What's wrong with the "exclusion probability"**. Disponível em: <http://dna-view.com/exclusn.htm#definition>, 1997a. Acesso em: 26 dez. 2019.

BRENNER, C.; STAUB, R. W. **Can DNA solve this?** Phoenix, AZ: Poster presented at the 14th International Symposium on Human Identification, September 30–October 2, 2003. p. 1–2.

BRENNER, C.H. & MORRIS, J.W. Paternity Index Calculations in Single Locus Hypervariable DNA Probes: Validation and Other Studies. **The International Symposium on Human Identification**, Promega Corporation, Madison, WI, USA, p. 21-53, 1989.

BROEDERS, A. P. A. Some observations on the use of probability scales in forensic identification. **Forensic Linguistics**, v.6, p.228–241, 1999. DOI: 10.1558/sll.1999.6.2.228

BUCKLETON, J. A Framework for Interpreting Evidence, in: Buckleton, J. S.; Triggs, C. M.; Walsh, S. J (eds.), **Forensic DNA Evidence Interpretation**, CRC Press, 2005.

BUCKLETON, J. S.; KRAWCZAK, M.; WEIR, B. S. The interpretation of lineage markers in forensic DNA testing. **Forensic science international. Genetics**, v.5, p.78–83, 2011a.

BUCKLETON, J. S.; TRIGGS, C. M.; CURRAN, J. M. Detection of Deviations From Genetic Equilibrium – A Commentary on Budowle B, Moretti TR, Baumstark AL, Defenbaugh DA, Keys KM. Population Data on the Thirteen CODIS Core Short Tandem Repeat Loci in African Americans, Caucasians,

Hispanics, Bahamians, Jamaicans And Trinidadians in (1999) Journal of Forensic Sciences, 44, 1277-1286. **Journal of Forensic Sciences**, v.46, n.1, p.198-202, 2001.

BUCKLETON, J., TRIGGS, C., CURRAN, J., BUDOWLE, B., AND CHAKRABORTY, R., "Detection of Deviation from Genetic Equilibrium—A Commentary on Budowle B, Moretti TR, Baumstark AL, Defenbaugh DA, Keys KM. Population Data on the Thirteen CODIS Core Short Tandem Repeat Loci in African Americans, US Caucasians, Hispanics, Jamaicans, and Trinidadians. J Forensic Sci 1999; 44: 1277–86". **Journal of Forensic Sciences**, v. 46, N. 1, p.198-202. DOI:10.1520/JFS14945J.

BUCKLETON, J.; CLAYTON, T. & TRIGGS, C. Parentage Testing, in: Buckleton, J. S.; Triggs, C. M.; Walsh, S. J (eds.), **Forensic DNA Evidence Interpretation**, CRC Press, 2005.

BUCKLETON, J.; CURRAN, J. Sampling Effects, in: Buckleton, J. S.; Triggs, C. M.; Walsh, S. J (eds.), **Forensic DNA Evidence Interpretation**, CRC Press, 2005.

BUCKLETON, J.; CURRAN, J.; GOUDET, J.; TAYLOR, D.; THIERY, A.; & WEIR, B. S. Population-specific FST values for forensic STR markers: A worldwide survey. **Forensic science international. Genetics**, v.23, p.91–100, 2016. <https://doi.org/10.1016/j.fsigen.2016.03.004>

BUCKLETON, J.; TRIGGS, C. M.; WALSH, S. J. **Forensic DNA Evidence Interpretation**. Boca Raton, Florida: CRC Press, 2005.

BUCKLETON, J.; WALSH, S.; HARBISON, S. Cap. 9: Nonautosomal Forensic Markers, in: Buckleton, J. S., Triggs, C. M., Walsh, S. J. (eds.), **Forensic DNA Evidence Interpretation**. CRC Press 2005.

BUDOWLE, B. SNP Typing Strategies. **Forensic Science International**, v.146, p.139-142, 2004.

BUDOWLE, B., MONSON, K. L.; CHAKRABORTY, R. Estimating minimum allele frequencies for DNA profile frequency estimates for PCR-based loci. **International Journal of Legal Medicine**, v.108, p.173–176, 1996. DOI:10.1007/BF01369786

BUDOWLE, B.; ALLARD, MW.; WILSON, M. R.; CHAKRABORTY, R. Forensics and mitochondrial DNA: applications, debates, and foundations. **Annual Review of Genomics and Human Genetics**, v.4, p.119-141, 2003. doi:10.1146/annurev.genom.4.070802.110352.

BUDOWLE, B.; MONSON, K. L., GIUSTI, A. M., BROWN, B. L. The assessment of frequency estimates of Hinf I-generated VNTR profiles in various ethnic databases. **Journal of forensic sciences**, v.39, p.988–1008, 1994b. PMID: 7910844.

BUDOWLE, B.; MONSON, K.; GIUSTI, A.; BROWN, B. The assessment of frequency estimates of Hae III-generated VNTR profiles in various reference databases. **Journal of forensic sciences**, v.39, p.319–352, 1994a. DOI:10.1520/JFS13605J.

BUDOWLE, B.; MORETTI, T. R.; BAUMSTARK, A. L.; DEFENBAUGH, D. A.; KEYS, K. M. Population data on the thirteen CODIS core short tandem repeat loci in African Americans, U.S. Caucasians, Hispanics, Bahamians,

Jamaicans, and Trinidadians. **Journal of forensic sciences**, v.44, n.6, p.1277–1286, 1999a. PMID: 10582369.

BUDOWLE, B.; PLANZ, J. V.; CAMPBELL, R.; EISENBERG, A. J. Chapter 26 - Molecular Diagnostic Applications, in: Patrinos, G. P. & Ansorge, W. J. (eds.), **Forensic Science Molecular Diagnostics**, 2ed, Academic Press, 2010, Pages 393-406. doi:10.1016/B978-0-12-374537-8.00026-2.

BUDOWLE, B.; PLANZ, J. V.; CAMPBELL, R.; EISENBERG, A. J. Molecular Diagnostic Applications in Forensic Science. **Molecular Diagnostics**, Elsevier, p.393–406, 2010. doi:10.1016/b978-0-12-374537-8.00026-2

BUDOWLE, B.; WILSON, M. R.; DIZINNO, J. A.; STAUFFER, C.; FASANO, M. A.; HOLLAND, M. M.; MONSON, K. L. Mitochondrial DNA regions HVI and HVII population data. **Forensic Science International**, v.103, p.23–35, 1999b. DOI:10.1016/s0379-0738(99)00042-0

BUTLER, J. M. The future of forensic DNA analysis. ***Philosophical transactions of the Royal Society of London. Series B, Biological sciences***, v.370, n.1674, p.20140252, 2015aa. doi: 10.1098/rstb.2014.0252

BUTLER, J. M. **Advanced topics in Forensic DNA Typing – interpretation**. Ed.1, Academic Press, 2015a.

BUTLER, J. M. **Advanced Topics in Forensic DNA Typing: Methodology**. Elsevier Academic Press, San Diego, 2012.

BUTLER, J. M. **Forensic DNA Typing: Biology, Technology, and Genetics of STR Markers**. Ed. 2, Elsevier Academic Press, New York, 2005.

BUTLER, J. M. **Fundamentals of Forensic DNA Typing**. Elsevier Academic Press, San Diego, 2010.

BUTLER, J. M. Review: Genetics and Genomics of Core STR Loci Used in Human Identity Testing. **Journal of Forensic Sciences**, in press, v.51, n.2, p.253-65, 2006.

BUTLER, J. M. STR Alleles and Amplification Artifacts. **Advanced Topics in Forensic DNA Typing: Interpretation**, p.47–86, 2015. doi:10.1016/b978-0-12-405213-0.00003-8

BUTLER, J. M. The future of forensic DNA analysis. **Philos Trans R Soc Lond B Biol Sci.**, v.370, n.1674:20140252, 2015b. DOI:10.1098/rstb.2014.0252

BUTLER, J.M. Genetics and genomics of core STR loci used in human identity testing. **Journal of forensic sciences**, v.51, n.2, p.253-265, 2006.

BUTLER, J.M. Short tandem repeat typing technologies used in human identity testing. **BioTechniques**, v.43, p.2-5, 2007.

CARRACEDO, A. Biology/DNA, in: Siegel, J. A.; Saukko, P. J.; Houck, M. M. (eds)., **Encyclopedia of Forensic Sciences**, 2ed, Academic Press, 2013.

CARRACEDO, A.; BÄR, W.; LINCOLN, P.; MAYR, W., MORLING, N.; OLAISEN, B.; SCHNEIDER, P.; BUDOWLE, B.; BRINKMANN, B.; GILL, P.; HOLLAND, M.; TULLY, G.; WILSON, M. DNA Commission of the International Society for Forensic Genetics: guidelines for mitochondrial DNA typing. **Forensic Science International**, v.110, n.2, p.79-85, 2000. DOI: 10.1016/j.fsigen.2014.07.010

CARRACEDO, Á.; BUTLER, J. M.; GUSMÃO, L.; LINACRE, A.; PARSON, W.; ROEWER, L.; SCHNEIDER, P. M. New guidelines for the publication of genetic population data. **Forensic Science International: Genetics**, v.7, p.217–220, 2013.

CARRACEDO, Á.; BUTLER, J. M.; GUSMÃO, L.; LINACRE, A.; PARSON, W.; ROEWER, L.; SCHNEIDER, P. M. Update of the guidelines for the publication of genetic population data. **Forensic Science International: Genetics**, v.10, p. A1–A2, 2014.

CHAKRABORTY, R. Sample size requirements for addressing the population genetic issues of forensic use of DNA typing. **Human Biology**, v.64, p.141–160, 1992.

CHAKRABORTY, R.; DANKER-HOPFE, H. Analysis of population structure: a comparative study of different estimators of Wright's fixation indices, in: Rao, C. R.; Chakraborty, R. (ed.), **Handbook of statistics 8: Statistical methods in biological and medical sciences**. Amsterdam: North-Holland, 1991, p 203-254.

CHAKRABORTY, R.; KIDD, K. K. The utility of DNA typing in forensic work. **Science**, v.254, p.1735–1739, 1991.

CHAKRABORTY, R.; STIVERS, D. N. Paternity exclusion by DNA markers: effects of paternal mutations. **Journal Forensic Science**, v.41, p.671–677, 1996.

CHAKRAVARTI, A. & LI, C. C. The effect of linkage on paternity calculations. In: Walker, R. H. (ed.), **Inclusion probabilities in parentage testing**. American Association of Blood Banks, Arlington, VA, pp 411–420, 1983.

CHAMPOD, C. Identification/individualization: Overview and Meaning of ID, in: Siegel, J. A.; Saukko, P. J. (Eds.), **Encyclopedia of Forensic Sciences**, London: Academic Press, 2000.

CHAMPOD, C.; EVETT, I.; JACKSON, G.; BIRKETT, J. Comments on the scale of conclusions proposed by the ad hoc committee of the ENFSI marks working group. **Inf. Bull. Shoeprint/Toolmark Examiners**, v.6, p.11–18, 2000.

COBLE, M. D.; BUCKLETON, J.; BUTLER, J.M.; EGELAND, T.; FIMMERS, R.; GILL, P.; GUSMÃO, L.; GUTTMAN, B.; KRAWCZAK, M.; MORLING, N.; PARSON, W.; PINTO, N.; SCHNEIDER, P.M.; SHERRY, S.T.; WILLUWEIT, S.; PRINZ, M. DNA Commission of the International Society for Forensic Genetics: Recommendations on the validation of software programs performing biostatistical calculations for forensic genetics applications. **Forensic Science International: Genetics**, v.25, p.191–197, 2016.

COBLE, M. D.; JUST, R. S.; O'CALLAGHAN, J. E.; LETMANYI, I. H.; PETERSON, C. T.; IRWIN, J. A.; PARSONS, T. J. Single nucleotide polymorphisms over the entire mtDNA genome that increase the power of forensic testing in Caucasians. **International Journal of Legal Medicine**, v.118, p.137–146, 2004. DOI:10.1007/s00414-004-0427-6.

COBLE, M. D.; LOREILLE, O.; WADHAMS, M. J.; EDSON, S. M.; MAYNARD, K.; MEYER, C. E.; NIEDERSTÄTTER, H.; BERGER, C.; BERGER, B.; FALSETTI, A. B.; GILL, P.; PARSON, W.; FINELLI, L. N. Mystery solved:

the identification of the two missing Romanov children using DNA analysis.

PloS one, v.4, n.3, e4838, 2009. DOI:10.1371/journal.pone.0004838

COLACO, S. & MODI, D. Genetics of the human Y chromosome and its association with male infertility. **Reproductive Biology and Endocrinology**, v.16, p.14, 2018. doi: 10.1186/s12958-018-0330-5

COLE, J. B. PyPedal: a computer program for pedigree analysis. **Computers and Electronics in Agriculture**, v.57, p.107–113, 2007.

CROW, J. F.; KIMURA, M. **An Introduction to Population Genetics Theory**. London: Harper and Row, 1970.

CURRAN, J. M.; BUCKLETON, J. S.; TRIGGS, C. M. What is the Magnitude of the Subpopulation Effect? **Forensic Science International**, v.135, n.1, p.1-8, 2003. DOI:10.1016/s0379-0738(03)00171-3.

D.N.A. ADVISORY BOARD. Quality assurance standards for forensic DNA testing laboratories. **Forensic Science Communication**, v.2, p.1–15, 1998.

DAJDA, T.; JUNG, M. LR-calculation of any kinship situation using a new graphical interface: generate two or more hypotheses, draw the family trees and assign the DNA-profiles to person symbols. **International Congress Series**, v.1288, p.474–476, 2006.

DAWID, A. P.; MORTERA, J.; PASCALI, V. L. Non-fatherhood or mutation? A probabilistic approach to parental exclusion in paternity testing. **Forensic Science International**, v.124, n.1, p. 55-61, 2001.

DAWID, A. P.; MORTERA, J.; PASCALI, V.; VAN BOXEL, D. Probabilistic expert systems from forensic inference from genetic markers. **Scandinavian Journal of Statistics**, v.29, p.577–595, 2002.

DAWID, A.P. Which likelihood ratio? **Law, Probability & Risk**, v.3, n.1, p.65–71, 2004.

DE KNIJFF, P. Son, give up your gun: Presenting Y-STR results in court. **Profiles in DNA**, v.6, n.2, p.3–5, 2003. Disponível em: <http://www.promega.com/profiles/602/ProfilesinDNA_602_03.pdf>.

DELGADO, E. & NEYRA, C. D. Allele frequencies of 21 autosomal STR markers in a mixed race Peruvian population applied to forensic practice. **Spanish Journal of Legal Medicine**, v.45, n.3, p.92-97, 2019. doi:10.1016/j.remle.2018.09.001.

DJ-USA- Department of Justice. **Uniform Language for Testimony and Reports for Forensic YSTR Data Examinations, U.S.** Department of Justice, 2020. URL: <https://www.justice.gov/olp/page/file/1095976/>. Acesso em: 17 fev 2021.

DJ-USA- Department of Justice. **Uniform Language for Testimony And Reports For Forensic Autosomal DNA Examinations Using Probabilistic**

Genotyping Systems - Autosomal Dna Examinations Using Probabilistic Genotyping Systems Ultr, U.S. Department of Justice, 2018

Disponível em: <https://www.justice.gov/olp/page/file/1095976/>.

DØRUM, G.; KLING, D.; BAEZA-RICHER, C.; GARCÍA-MAGARIÑOS, M.; SÆBØ, S.; DESMYTER, S.; EGELAND, T. Models and implementation for relationship problems with dropout. **International Journal of Legal Medicine**, v.129; p.411–423, 2015.

DRÁBEK, J. Validation of software for calculating the likelihood ratio for parentage and kinship. **Forensic Science International: Genetics**, v.3, p.112–118, 2009.

DUMACHE, R.; ENACHE, A.; BARBARII, L.; CONSTANTINESCU, C.; PASCALAU, A.; JINCA, C.; ARGHIRESCU, S. Chimerism Monitoring by Short Tandem Repeat (STR) Markers in Allogeneic Stem Cell Transplantation. **Clinical laboratory**, v.64, n.9, p.1535–1543, 2018. doi: 10.7754/Clin.Lab.2018.180409

EGELAND, T.; KLING, D.; MOSTAD, P. Chapter 2 - Basics and Chapter 6 - Models for pedigree inference, in: Egeland, T.; Kling, D.; Mostad, P. (eds.), **Relationship Inference with Familias and R**, Elsevier Academic Press, 2016. Pages 11-50 and 147-187. doi: 10.1016/B978-0-12-802402-7.00006-0.

EGELAND, T.; KLING, D.; MOSTAD, P. Chapter 2 - Basics. **Relationship Inference with Familias and R**, Elsevier Academic Press, 2016.
doi:10.1016/b978-0-12-802402-7.00002-3

EGELAND, T.; MOSTAD, P.F.; MEVAG, B.; STENERSEN, M. Beyond traditional paternity and identification cases. Selecting the most probable pedigree. **Forensic Science International**, v.110, p.47–59, 2000.

ELDRIDGE, H. Juror comprehension of forensic expert testimony: A literature review and gap analysis. **Forensic Science International: Synergy**, v.1, p.24-34, 2019. doi: 10.1016/j.fsisyn.2019.03.001.

ELLEGREN, H. Heterogeneous mutation processes in human microsatellite DNA sequences. **Nature Genetics**, v.24, p.400–402, 2000.
<https://www.ncbi.nlm.nih.gov/pubmed/10742106>

ELLEGREN, H. Microsatellite mutations in the germline: implications for evolutionary inference. **Trends in Genetics**, v.16, p.551–558, 2000.

ELLEGREN, H. Microsatellites: simple sequences with complex evolution. **Nature Reviews Genetics**, v.5, p.435–445, 2004.

EN ISO/IEC 17025: 1999. **Standard General requirements for the Competence of Testing and Calibration Laboratories**. Geneva, Switzerland: International Standardization Organization, 1999.

ENFSI - European Network of Forensic Science Institutes. **Guideline for Evaluative Reporting in Forensic Science: Strengthening the Evaluation of Forensic Results across Europe (STEOFRAE)**. Approved version 3.0, 2016. Disponível em: http://enfsi.eu/wpcontent/uploads/2016/09/m1_guideline.pdf Acesso em: 18 fev 2021.

ENGELS, W. R. Exact Tests for Hardy–Weinberg Proportions. **Genetics**, v.183, n.4: 1431-1441, 2009. DOI:10.1534/genetics.109.108977.

ESPINHEIRA, R. M. Identificação genética de desconhecidos, in: Corte-Real, F.; Vieira, D. N. (ed.), **Princípios de genética forense**. Coimbra: Imprensa da Universidade de Coimbra, 2015. DOI:10.14195/978-989-26-0957-7_3.

ESSAM, K. & DIAB, A. Role of DNA in Paternity Testing. **Journal of Forensic Sciences & Criminal Investigation**, v.14, n.2, p.555-582, 2020. doi:10.19080/JFSCI.2020.14.555882.

EVETT, I. W. A Discussion of Recommendation 5.1 of the Second NRC Report on DNA Profiling. **FSS Report**, N.: TN 821, 1996.

EVETT, I. W. Evaluating DNA profiles in the case where the defense is it was my brother. **Journal of the Forensic Science Society**, v.32, p.5-14, 1992.

EVETT, I. W. Verbal conventions for handwriting opinions. **Journal Forensic Science**, v.45, p.508–509, 2000.

EVETT, I. W.; BUCKLETON, J. S., Some aspects of the Bayesian approach to evidence evaluation. **Journal of the Forensic Science Society**, v.29, p.317–324, 1989.

EVETT, I. W.; WEIR, B. S. **Interpreting DNA evidence: Statistical Genetics for Forensic Scientists**. Sunderland: Sinauer, 1998.

FAN, H.; CHU, J. Y. A Brief review of short tandem repeat mutation. **Genomics, Proteomics & Bioinformatics**, v.5, p.1, 2007.
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5054066/pdf/main.pdf>

FAN, H.; CHU, J.Y. A brief review of short tandem repeat mutation. **Genomics Proteomics Bioinformatics**, v.5, n.1, p.7-14, 2007.

FILOGLU, G.; BULBUL, O.; RAYIMOGLU, G.; YEDIAY, F. E.; ZORLU, T.; ONGOREN, S.; & ALTUNCUL, H. Evaluation of reliability on STR typing at leukemic patients used for forensic purposes. **Molecular biology reports**, v.41, n.6, p.3961–3972, 2014. doi:10.1007/s11033-014-3264-9

FIMMERS, R.; HENKE, L.; HENKE, J.; BAUR, M. How to deal with mutations in DNA-testing, in: Rittner, C.; Schneider, P. M. (eds.), **Advances in Forensic Haemogenetics**, v.4. Berlin: Springer-Verlag, 1992. p.285–287.

FISHER, R. A. Standard calculations for evaluating a blood-group system. **Heredity (Edinb)**, v.5, n.1, p. 95-102, 1951

FUNG, W. K.; CHUNG, Y.; WONG, D. Power of exclusion revisited: probability of excluding relatives of the true father from paternity. **International Journal of Legal Medicine**, v.116, n.2, p.64-67, 2002.

FUNG, W.K. & HU, Y.-Q. DNA: Statistical Analysis, in: Payne-James, J. & Byard, R. W. (eds.) **Encyclopedia of Forensic and Legal Medicine**, Elsevier: Second Edition, 2016.

GAO, Z.; CHEN, X.; ZHAO, Y; ZHAO, X.; ZHANG, S.; YANG, Y.; WANG, Y.; ZHANG, J. Forensic genetic informativeness of an SNP panel consisting of 19 multi-allelic SNPs. **Forensic Science International: Genetics**, v.34, Pages 49-56, 2018. doi:10.1016/j.fsigen.2018.01.006.

GARBER, R. A.; MORRIS, J. W. General Equations for the Average Power of Exclusion for Genetic Systems of n Codominant Alleles in One-Parent and No-Parent Cases of Disputed Parentage, in Walker, R. H. (ed.), **Inclusion Probabilities in Parentage Testing**. USA: AABB, 1983. p.277-280.

GAUTHIER, J.; VINCENT, A. T.; CHARETTE, S. J.; DEROME, N. A brief history of bioinformatics. **Briefings in bioinformatics**, v.20, n.6, p.1981–1996, 2019. DOI:10.1093/bib/bby063.

GESERICK, G. & WIRTH, I. Genetic Kinship Investigation from Blood Groups to DNA Markers. Transfusion medicine and hemotherapy : offizielles Organ der Deutschen Gesellschaft fur. **Transfusionsmedizin und Immunhamatologie**, v.39, n.3, p.163–175, 2012. doi:10.1159/000338850

GETTINGS, K. B.; APONTE, R. A.; VALLONE, P. M.; BUTLER, J. M. STR allele sequence variation: Current knowledge and future issues. **Forensic Science International: Genetics**, v.18, p.118–130, 2015. doi:10.1016/j.fsigen.2015.06.005

GIANNELLI, P. C.; IMWINKELRIED, E. J.; PETERSON, J. L. Reference Guide on Forensic Identification Expertise. De National Academy of Sciences. **Reference Manual on Scientific Evidence**. 3 ed. Washington, DC: The National Academies Press, 2011.

GILL, P. An assessment of the utility of single nucleotide polymorphisms (SNPs) for forensic purposes. **International Journal of Legal Medicine**, v.114, p.204–210, 2001.

GILL, P.; BRENNER, C.; BRINKMANN, B.; BUDOWLE, B.; CARRACEDO, A.; JOBLING, M. A.; DE, K.; KAYSER, M.; KRAWCZAK, M.; MAYR, W. R.; MORLING, N.; OLAISEN, B.; PASCALI, V.; PRINZ, M.; ROEWER, L.; SCHNEIDER, P. M.; SAJANTILA, A.; TYLER-SMITH, C. DNA Commission of the International Society of Forensic Genetics: Recommendations on forensic analysis using Y-chromosome STRs. **Forensic Science International**, v.124, p.5-10, 2001

GILL, P.; FOREMAN, L.; BUCKLETON, J. S.; TRIGGS, C. M.; ALLEN, H. A comparison of adjustment methods to test the robustness of an STR DNA database comprised of 24 European populations. **Forensic Science International**, v.131, p.184–196, 2003.

GILL, P.; HANED, H. A new methodological framework to interpret complex DNA profiles using likelihood ratios. **Forensic Science International: Genetics**, v.7, p.251–263, 2013.

GILL, P.; HICKS, T.; BUTLER, J. M.; CONNOLLY, E.; GUSMÃO, L.; KOKSHOORN, B.; MORLING, N.; VAN, O.; PARSON, W.; PRINZ, M.; SCHNEIDER, P. M.; SIJEN, T.; TAYLOR, D. DNA commission of the ISFG: Assessing the value of forensic biological evidence - Guidelines highlighting the importance of propositions: Part I: evaluation of DNA profiling comparisons given (sub-) source propositions. **Forensic Science International: Genetics**, v.36, p.189-202, 2018.

GILL, P.; HICKS, T.; BUTLER, J. M.; CONNOLLY, E.; GUSMÃO, L.; KOKSHOORN, B.; MORLING, N.; VAN, O.; PARSON, W.; PRINZ, M.; SCHNEIDER, P. M.; SIJEN, T.; TAYLOR, D. DNA Commission of the International Society for Forensic Genetics: Assessing the value of forensic biological evidence - Guidelines highlighting the importance of propositions. Part II: Evaluation of biological traces considering activity level propositions. **Forensic Science International: Genetics**, v.44, p.102186, 2020.

GILL, P.; WERRETT, D. J.; BUDOWLE, B.; GUERRIERI, R. An assessment of whether SNPs will replace STRs in national DNA databases. **Science and Justice**, v.44, p.51–53, 2004.

GJERTSON, D. W.; BRENNER, C. H.; BAUR, M. P.; CARRACEDO, A.; GUIDET, F.; LUQUE, J. A.; LESSIG, R.; MAYR, W. R.; PASCALI, V.L.; PRINZ, M.; SCHNEIDER, P. M.; MORLING, N. ISFG: Recommendations on

biostatistics in paternity testing. **Forensic Science International: Genetics**, v.1, n.3, p.223-231, 2007.

GOODWIN, W.; LINACRE, A.; HADI, S. **An Introduction to Forensic Genetics**. 1 ed. Chichester: John Wiley & Sons Ltd, 2007. ISBN (HB) 9780470010259.

GOODWIN, W.; LINACRE, A.; VANEZIS, P. The use of mitochondrial DNA and short tandem repeat typing in the identification of air crash victims. **Electrophoresis**, v.20, p. 1707–1711, 1999.

GOPINATH, S.; ZHONG, C.; NGUYEN, V.; GE, J.; LAGACÉ, R. E.; SHORT, M. L.; MULERO, J. J. Developmental validation of the Yfiler(®) Plus PCR Amplification Kit: An enhanced Y-STR multiplex for casework and database applications. **Forensic science international. Genetics**, v.24, p.164–175, 2016. <https://doi.org/10.1016/j.fsigen.2016.07.006>

GOUY, A. & ZIEGER, M. STRAF-A convenient online tool for STR data evaluation in forensic genetics. *Forensic science international. Genetics*, v.30, p.148–151, 2017. doi:10.1016/j.fsigen.2017.07.007

GREEN, J. M. A practical guide to analytical method validation. **Analytical Chemistry**, v.68, p.305–309, 1996.

GUEVARA, E. K.; PALO, J. U. Ö.; VERSTI. S.; KING, J. L.; SEIDEL, M.; STOLJAROVA, M.; WENDT, F. R.; BUS, M. M.; GUENGERICH, A.; CHURCH, W. B.; GUILLÉN, S.; ROEWER, L.; BUDOWLE, B.; SAJANTILA, A. Genetic assessment reveals no population substructure and divergent

regional and sex-specific histories in the Chachapoyas from northeast Peru.

PLoS ONE, v.15, n.12, p.e0244497, 2020. doi: 10.1371/journal.pone.0244497

GUO, S. W.; THOMPSON, E. A. Performing the Exact Test of Hardy-Weinberg Proportion for Multiple Alleles. **Biometrics**, V.48, N.2, p.361-372, 1992.

GUO, X & ELSTON, R. C. Linkage informative content of polymorphic genetic markers. **Human Heredity**, v.49, p.112–118, 1999.

GUSMÃO, L.; BUTLER, J. M.; CARRACEDO, A.; GILL, P.; KAYSER, M.; MAYR, W. R.; MORLING, N.; PRINZ, M.; ROEWER, L.; TYLER-SMITH, C.; SCHNEIDER, P. M. DNA Commission of the International Society of Forensic Genetics. DNA Commission of the International Society of Forensic Genetics (ISFG): an update of the recommendations on the use of Y-STRs in forensic analysis. **Forensic Science International**, v.157, p.187-97, 2006.

GUSMÃO, L.; SÁNCHEZ-DIZ, P.; ALVES, C.; BELEZA, S.; LOPES, A.; CARRACEDO, A.; AMORIM, A. Grouping of Y-STR haplotypes discloses European geographic clines. **Forensic science international**, v.134, n.2-3, p.172–179, 2003.

HAKIM, H. M.; KHAN, H. O.; LALUNG, J.; NELSON, B. R.; CHAMBERS, G. K.; & EDINUR, H. A. Autosomal STR Profiling and Databanking in Malaysia: Current Status and Future Prospects. **Genes**, v.11, n.10, p.1112, 2020. doi: 10.3390/genes11101112

HAMEED, I.; JEBOR, M.; AAMERA, J.; CHEAH, Y. K.; HAIDER, K.; ALI, H.; MUATAZ, A. Genetic variation and DNA markers in forensic analysis. **African Journal of Biotechnology**, v.13, p.3122-3136, 2014. doi:10.5897/AJB2013.13160.

HANSON, E.; BALLANTYNE, J. Comprehensive annotated STR physical map of the human Y chromosome: Forensic implications. **Legal medicine (Tokyo, Japan)**, v.8, p.110-120, 2006. DOI:10.1016/j.legalmed.2005.10.001.

HARDY, G. H. Mendelian proportions in a mixed population. **Science**, v.28, p.49–50, 1908.

HARES, D. R. Expanding the CODIS core loci in the United States. **Forensic Science International: Genetics**, v.6, n.1, e52-4, 2012. DOI: 10.1016/j.fsigen.2011.04.012.

HARES, D.R. Selection and implementation of expanded CODIS core loci in the United States. **Forensic Science International: Genetics**, v.17, p.33–34, 2015.

HOLLAND, M. M. & PARSONS, T. J. Mitochondrial DNA sequence analysis - validation and use for forensic casework. **Forensic Science Review**, v.11, n.1, p.21-50, 1999.

HOLLAND, M. M.; PARSONS, T.J. Mitochondrial DNA sequence analysis—validation and use for forensic casework. **Forensic Science Review**, v.11, p.21–50, 1999.

HUBER, L. Qualification and validation of software and computer systems in laboratories—Part 2: qualification of vendors. **Accreditation and Quality Assurance**, v.3, p.2–5, 1998a.

HUBER, L. Qualification and validation of software and computer systems in laboratories—Part 3: installation and operational qualification. **Accreditation and Quality Assurance**, v.3, p.140–144, 1998b.

HUBER, L. Qualification and validation of software and computer systems in laboratories—Part 4: evaluation and validation of existing systems. **Accreditation and Quality Assurance**, v.3, p.317–321, 1998c.

HUBER, L.; WIEDERODER, H. Qualification and validation of software and computer systems in laboratories—Part 1: validation during development. **Accreditation and Quality Assurance**, v.2, p.360–366, 1997.

HUMMEL, K.; CONRADT, J. Calculation of Percentage of Cases on File with an Unnamed Father in 100 One-man and 100 Two-man Cases (Filiation Cases) from South-West Germany in 1976-1981. **Zeitschrift für Rechtsmedizin**, v.88, p.277-284, 1982.

HUMMEL, K.; IHM, P.; SCHMIDT, V. **Biostatistical opinion of parentage—Biostatistische Abstammungsbegutachtung**. Table v. 1, part III—Tabellenband 1, 3. Stuttgart: Teil. Fischer, 1971. (Ver também Hummel, K. Hummel's Verbal Predicates. Disponível em: <<http://dna-view.com/hummel.htm>>).

HUSTON, K. A. Statistical Analysis of STR Data. Technical Tips, Profiles in DNA, **GenePrint™**, p. 14-15, 1998.

IEEE. Standard for Software and System Test Documentation, in: **IEEE Std 829-2008**, p. 1–150, 2008. DOI:10.1109/IEEESTD.2008.4578383.

IEEE. Standard for System and Software Verification and Validation, in: **IEEE Std 1012-2012 (Revision of IEEE Std 1012-2004)**, p.1–223, 2012. DOI:10.1109/ IEEESTD.2012.6204026.

INTERPOL - INTERNATIONAL CRIMINAL POLICE ORGANIZATION. **Disaster victim identification guide**, 2018. Disponível em: <[https://www.interpol.int file › 18Y1344 E DVI_Guide](https://www.interpol.int/file/18Y1344/E/DVI_Guide)>.

IP, S. C.; LIN, S. W.; LI, C.; LAI, K. M. Forensic DNA typing strategy of degraded DNA on discarded cigarette ends using the AmpF ℓ STR® Identifiler®, Identifiler® Plus and MiniFiler™ PCR amplification kits. **Science & justice: journal of the Forensic Science Society**, v.54, n.4, p.311–315, 2014. DOI:10.1016/j.scijus.2014.04.007.

ISENBERG, A. R. Forensic mitochondrial DNA analysis. In Saferstein, R. (Ed.), **Forensic science handbook**. Vol. II. Upper Saddle River, New Jersey: Pearson Prentice Hall, 2004. p. 297–327.

JEFFREYS, A. J. Two decades of DNA fingerprints. **BBC News**, 2004. Disponível em: <<http://news.bbc.co.uk/2/hi/science/nature/3636050.stm>>. Acesso em: 20 fev. 2020.

JEFFREYS, A. J.; ALLEN, M. J.; ARMOUR, J. A.; COLLICK, A.; DUBROVA, Y.; FRETWELL, N.; GURAM, T.; JOBLING, M.; MAY, C. A.; NEIL, D. L. Mutation processes at human minisatellites. **Electrophoresis**, v.16, p.1577–1585, 1995.

JEFFREYS, A. J.; BROOKFIELD, J. F.; SEMEONOFF, R. Positive identification of an immigration test-case using human DNA fingerprints. **Nature**, v.317, p.818–819, 1985b.

JEFFREYS, A. J.; MACLEOD, A.; TAMAKI, K.; NEIL, D. L.; MONCKTON, D. G. Minisatellite repeat coding as a digital approach to DNA typing. **Nature**, v.354, p.204–209, 1991.

JEFFREYS, A. J.; WILSON, V.; THEIN, S. L. Hypervariable ‘minisatellite’ regions in human DNA. **Nature**, v.314, p.67–73, 1985a.

JHA, D.; RIJAL, J.; TULADHAR, B.; POKHAREL, B. Mutations or Exclusion: An Uncommon Parentage Assessment Case. **Scientific World**, v.11, n.11, p.74-76, 2013. doi:10.3126/sw.v11i11.8556

JIN, X.; ZHANG, X.; SHEN, C.; LIU, Y.; CUI, W.; CHEN, C.; GUO, Y.; & ZHU, B. A Highly Polymorphic Panel Consisting of Microhaplotypes and Compound Markers with the NGS and Its Forensic Efficiency Evaluations in Chinese Two Groups. **Genes**, v.11, n.9, p.1027, 2020. doi: 10.3390/genes11091027

JOBLING, M. A.; TYLER-SMITH, C. The human Y chromosome: An evolutionary marker comes of age. **Nature Reviews Genetics**, v.4, p.598–612, 2003.

JOBLING, M.A.; GILL, P. Encoded evidence, DNA in forensic analysis. **Nature Reviews Genetics**, v.5, p.739-751, 2004.

KANTHASWAMY, S. Review: domestic animal forensic genetics – biological evidence, genetic markers, analytical approaches and challenges. **Animal Genetics**, v. 46, n. 5, p. 473-484, 2015.

KAYE, D. H.; SENSABAUGH, G. Reference Guide on DNA Identification Evidence, in: **National Academy of Sciences. Reference Manual on Scientific Evidence**. 3 ed. Washington: The National Academies Press (Federal Judicial Center), 2011.

KAYE, D. Rounding Up the Usual Suspects: A Logical and Legal Analysis of DNA Trawling Cases. **North Carolina Law Review**, v.87, p.425-503, 2009b.

KAYE, D. H. Identification, individualization and uniqueness: What's the difference? **Law, Probability and Risk**, v.8, p.85–94, 2009a.

KAYSER, M.; CAGLIÀ, A.; CORACH, D.; FRETWELL, N.; GEHRIG, C.; GRAZIOSI, G.; HEIDORN, F.; HERRMANN, S.; HERZOG, B.; HIDDING, M.; HONDA, K.; JOBLING, M.; KRAWCZAK, M.; LEIM, K.; MEUSER, S.; MEYER, E.; OESTERREICH, W.; PANDYA, A.; PARSON, W.; PENACINO, G.; ... ROEWER, L. Evaluation of Y-chromosomal STRs: a multicenter study. **International journal of legal medicine**, v.110, n.3, p.125–149, 1997. DOI:10.1007/s004140050051

KAYSER, M.; DE KNIJFF, P. Improving human forensics through advances in genetics, genomics and molecular biology. **Nature Reviews Genetics**, v.12, p.179-192, 2011.

KHUDHUR, S. D. & CROOCK, M. S. Biometrics System based Human Identification using STR DNA Marker. **International Journal of Computer Applications (0975 – 8887)**, v.138, n.7, 2016.

KLING, D.; TILLMAR, A. O.; EGELAND, T. Familias 3—extensions and new functionality. **Forensic Science International: Genetics**, v.13, p.121–127, 2014. DOI:10.1016/j.fsigen.2014.07.004

KLING, R. **Computerization and Controversy: Value Conflicts and Social Choices**. 2 ed. London: Academic Press, 1991.

KOWALCZYK, M.; ZAWADZKA, E.; SZEWCZUK, D.; GRYZIŃSKA, M.; JAKUBCZAK, A. Molecular markers used in forensic genetics. **Medicine, science, and the law**, v.58, n.4, p.201–209, 2018. doi: 10.1177/0025802418803852

KRAWCZAK, M. Forensic evaluation of Y-STR haplotype matches: a comment. **Forensic Science International**, v.118, p.114–115, 2001.

KRAWCZAK, M. Informativity assessment for biallelic single nucleotide polymorphisms. **Electrophoresis**, v.20, p.1676–1681, 1999.

KRAWCZAK, M. Kinship testing with X-chromosomal markers: Mathematical and statistical issues. **Forensic Science International: Genetics**, v.1, p.111–114, 2007.

KUMAR, A.; SINGH, N.; GAURAV, S. S. Forensic Bioinformatics: An innovative technological advancement in the field of Forensic Medicine and Diagnosis. **International Journal of Medicine and Biomedical Research**, v.1, p.35–38, 2012. doi:10.14194/ijmbr.116.

LANDER, E. S.; LINTON, L. M.; BIRREN, B.; NUSBAUM, C.; ZODY, M. C.; BALDWIN, J.; DEVON, K.; DEWAR, K.; DOYLE, M.; FITZHUGH, W.; FUNKE, R.; GAGE, D.; HARRIS, K.; HEAFORD, A.; HOWLAND, J.; KANN, L.; LEHOCZKY, J.; LEVINE, R.; MCEWAN, P.; MCKERNAN, K.; ... INTERNATIONAL HUMAN GENOME SEQUENCING CONSORTIUM. Initial sequencing and analysis of the human genome. **Nature**, v.409, n.6822, p.860–921, 2001. DOI:10.1038/35057062.

LAW, B. Statistical Interpretation of DNA Evidence. Ph.D. thesis, Department of Statistics. Auckland, N. Z: University of Auckland, 2002.

LECLAIR, B.; SHALER, R.; CARMODY, G. R.; ELIASON, K.; HENDRICKSON, B. C.; JUDKINS, T.; NORTON, M. J.; SEARS, C.; SCHOLL, T. Bioinformatics and human identification in mass fatality incidents: the world trade center disaster. **Journal of forensic sciences**, v.52, n.4, p.806–819, 2007. DOI:10.1111/j.1556-4029.2007.00456.x

LEE, H. C.; PAGLIARO, E. M. Forensic Evidence and Crime Scene Investigation. **Journal Forensic Investigation**, v.1, n.2, p. 5, 2013.

- LEE, J. C.; TSAI, L.; LINACRE, A.; HSIEH, H. Distribution of paternity index and random man not excluded values from a simulated parentage testing study. **Forensic Science Journal**, v.9, n.1, p.9-18, 2010.
- LEE, Y. S.; KIM, W. Y.; JI, M.; KIM, J. H.; BHAK, J. MitoVariome: a variome database of human mitochondrial DNA. **BMC Genomics**, v. 10, p. S12, 2009. DOI: 10.1186/1471-2164-10-S3-S12.
- LEVENE, H. On a matching problem arising in genetics. **Annals of Mathematical Statistics**, v.20, p.91-94, 1949.
- LEWONTIN, R. C.; HARTL, D. L., Population genetics in forensic DNA typing. **Science**, v.254, p.1745–1750, 1991.
- LI, C.; ZHAO, S.; ZHANG, N.; ZHANG, S.; HOU, Y. Differences of DNA methylation profiles between monozygotic twins' blood samples. **Molecular biology reports**, v.40, n.9, p.5275-5280, 2013.
- Li, Y. C.; Korol, A. B.; Fahima, T.; & Nevo, E. Microsatellites within genes: structure, function, and evolution. **Molecular biology and evolution**, v.21, n.6, p.991–1007, 2004. <https://doi.org/10.1093/molbev/msh073>
- LITT, M.; LUTY, J. A. A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. **American journal of human genetics**, v.44, n.3, p.397–401, 1989.
- LIU, L.; LI, Y.; LI, S.; HU, N.; HE, Y.; PONG, R.; LIN, D.; LU, L.; LAW, M. Comparison of Next-Generation Sequencing Systems. **Journal of biomedicine & biotechnology**, v.2012, Article ID:251364, 2012. DOI:10.1155/2012/251364.

- LIU, Y. Y. & HARBISON, S. A review of bioinformatic methods for forensic DNA analyses. *Forensic science international. Genetics*, v.33, p.117–128, 2018. doi:10.1016/j.fsigen.2017.12.005
- LUSCOMBE, N. M.; GREENBAUM, D.; GERSTEIN, M. What is bioinformatics? A proposed definition and overview of the field. **Methods of Information in Medicine**, v.40, n.4, p.346-358, 2001.
- MAHMOOD, H.; SALMAN, N.; SALIH, K.; HASAN, D.; AL-ZUBAIDI, M. (2020). Frequency of Y-chromosome STRs using PowerPlex® Y23 System in Iraqi population. **Egyptian Journal of Forensic Sciences**, v.10, p.12, 2020. doi:10.1186/s41935-020-00186-3.
- MALASPINAS, A. S.; SLATKIN, M.; SONG, Y. S. Match probabilities in a finite, subdivided population. **Theoretical population biology**, v.79, n.3, p.55–63, 2011. doi:10.1016/j.tpb.2011.01.003
- MARGOT, P. Identification and the Development of Forensic Science, in: Bruinsma, G.; Weisburd, D. (eds.), **Encyclopedia of Criminology and Criminal Justice**. New York: Springer, 2014.
- MARQUIS, R.; BIEDERMANN, A.; CADOLA, L.; CHAMPOD, C.; GUEISSAZ, L.; MASSONNET, G.; MAZZELLA, W. D.; TARONI, F.; HICKS, T. Discussion on how to implement a verbal scale in a forensic laboratory: Benefits, pitfalls and suggestions to avoid misunderstandings. **Science & Justice**, v.56, n.5, p.364-370, 2016. doi: 10.1016/j.scijus.2016.05.009.

- MARWAL, A. & GAUR, R. K. Molecular markers: tool for genetic analysis, in: Verma, A. S. & Singh, A. (eds.), **Animal Biotechnology**, Elsevier Academic Press: Second Edition, 2020. <https://doi.org/10.1016/C2016-0-01620-0>
- MEESTER, R.; SJERPS, M. The evidential value in the DNA database search controversy and the two-stain problem. **Biometrics**, v.59, p.727–732, 2003.
- MENDEL, G. Experiments in plant hybridization, in: Fisher, R.A.; Bennett, J.H. (eds.), **English Translation and Commentary 1866**. Edinburgh: Oliver and Boyd, 1965.
- MESSINA, F.; DI CORCIA, T.; RAGAZZO, M.; SANCHEZ MELLADO, C.; CONTINI, I.; MALASPINA, P.; CIMINELLI, B. M.; RICKARDS, O.; JODICE, C. (2018). Signs of continental ancestry in urban populations of Peru through autosomal STR loci and mitochondrial DNA typing. **PloS one** , v.13, n.7, p.e0200796. doi: 10.1371/journal.pone.0200796
- MILOT, E.; BAECHLER, S.; CRISPINO, F. Must the random man be unrelated? A lingering misconception in forensic genetics. **Forensic Science International: Synergy**, v.2, p.35–40, 2020.
- MITCHELL, A. A.; TAMARIZ, J.; O'CONNELL, K.; DUCASSE, N.; BUDIMLIJA, Z.; PRINZ, M.; CARAGINE, T. Validation of a DNA mixture statistics tool incorporating allelic dropout and drop-in. **Forensic Science International: Genetics**, v.6, p.749–761, 2012.
- MOREIRA, R. P.; FREITAS, A. Z. V. M. **Dicionário de Odontologia legal**. Rio de Janeiro: Guanabara Koogan, 1999.

MORIMOTO, C.; MANABE, S.; KAWAGUCHI, T.; KAWAI, C.; FUJIMOTO, S.; HAMANO, Y.; YAMADA, R.; MATSUDA, F.; & TAMAKI, K. Pairwise Kinship Analysis by the Index of Chromosome Sharing Using High-Density Single Nucleotide Polymorphisms. **PloS one**, v.11, n.7, p.e0160287, 2016. doi:10.1371/journal.pone.0160287

MORLING, N.; ALLEN, R. W.; CARRACEDO, A.; GEADA, H.; GUIDET, F.; HALLENBERG, C.; MARTIN, W.; MAYR, W. R.; OLAISEN, B.; PASCALI, V. L.; SCHNEIDER, P. M; PATERNITY TESTING COMMISSION OF THE INTERNATIONAL SOCIETY OF FORENSIC GENETICS. Paternity Testing Commission of the International Society of Forensic Genetics: recommendations on genetic investigations in paternity cases. **Forensic science international**, v.129, n.3, p.148–157, 2002. DOI: 10.1016/s0379-0738(02)00289-x

MORRIS, J. W.; GARBER, R. A.; D'AUTREMONT, J.; BRENNER, C. H. The avuncular index and the incest case. **Advances in Forensic Haemogenetics**. Berlin: Springer-Verlag, 1988. p.607-611.

MORRISON, J.; WATTS, G.; HOBBS, G.; DAWNAY, N. Field-based detection of biological samples for forensic analysis: Established techniques, novel tools, and future innovations. **Forensic science international**, v.285, p.147–160, 2018. doi: 10.1016/j.forsciint.2018.02.002

MUNTAHA, S. T.; Hasnain, M. J. U.; KHAN, M. W. A.; RAFIQ, F.; PERVEZ, M. T.; HUSSAIN, T. Role of Bioinformatics in Forensic Science. **Fuuast Journal of Biology**, v.8, n.1, p.133-138, 2018.

MUNTAHA, S. T.; HASNAIN, M. J. U.; KHAN, W. A.; RAFIQ, F.; PERVEZ, M. T.;
 HUSSAIN, T. Role of Bioinformatics in Forensic Science. **Fuuast Journal
 of Biology**, v.8, n.1, p.133 – 138, 2018.

NACHMAN, M. W.; CROWELL, S. L. Estimate of the mutation rate per nucleotide
 in humans. **Genetics**, v.156, p.297–304, 2000.

NATA - NATIONAL ASSOCIATION OF TESTING AUTHORITIES AUSTRALIA.
**ISO/IEC 17025 Application Document: Supplementary Requirements
 for Accreditation in the Field of Forensic Science**. 2000.

NATURE RESEARCH. **Population Genetics. Latest research and news /
 Nature**. Disponível em: <[https://www.nature.com/subjects/population-
 genetics](https://www.nature.com/subjects/population-genetics)>. Acesso em: 12 jun. 2020.

NEI, M. Analysis of gene diversity in subdivided populations. **Proceedings of the
 National Academy of Sciences USA**, v.70, p.3321-3323, 1973.

NEI, M. F-statistics and analysis of gene diversity in subdivided populations.
Annals of Human Genetics, v.41, p.225-233, 1977.

NEI, M. **Molecular Evolutionary Genetics**. Columbia University Press, New York,
 1987.

NEUHAUSER, C. Mathematical models in population genetics, in: Cannings, C.
 (Ed.) **Handbook of Statistical Genetics**. Chichester: John Wiley and Sons,
 2001.

NIEDZWIECKI, E.; DEBUS-SHERRILL, S.; FIELD, M. B. **Understanding Familial
 DNA Searching: Coming to a Consensus on Terminology**. National

Institute of Justice, Office of Justice Programs, U.S. Department of Justice.
Fairfax, VA: ICF International, 2017.

NRC I — NATIONAL RESEARCH COUNCIL. **Committee on DNA Technology in Forensic Science, Board on Biology, Commission on Life Sciences, DNA Technology in Forensic Science.** Washington, DC: National Academy Press, 1992.

NRC II — NATIONAL RESEARCH COUNCIL. **Committee on DNA Forensic Science, The Evaluation of Forensic DNA Evidence.** Washington, DC: National Academy Press, 1996.

NURHANTARI, Y. & SURYADI, H. Genetic Inconsistency in Paternity Investigation. **KnE Life Sciences**, v.4, n.11, p.47–55, 2019. doi: 10.18502/kls.v4i11.3851

O'CONNOR, K. L.; MYERS, S. P.; BUTTS, E. L. R.; HILL, C. R.; BUTLER, J. M.; VALLONE, P. M. Candidate Reference Family Data: A Tool for Validating Kinship Analysis Software. Poster #35, **21th International Symposium on Human Identification San Antonio**, TX, p.11–14, 2010 Disponível em: <http://www.cstl.nist.gov/biotech/strbase/pub_pres/Promega2010_OConnor.pdf>. Acesso em:06 mar. 2020.

OCHOA, A. & STOREY, J. D. Estimating FST and kinship for arbitrary population structures. **PLoS Genetic**, v.17, p.1, p.e1009241, 2021. doi: 10.1371/journal.pgen.1009241

PARSON, W. Age Estimation with DNA: From Forensic DNA Fingerprinting to Forensic (Epi)Genomics: A Mini-Review. **Gerontology**, v.64, n.4, p.326-332, 2018. doi: 10.1159/000486239.

PARSON, W.; BANDELT, H. J. Extended guidelines for mtDNA typing of population data in forensic science. **Forensic Science International: Genetics**, v.1, p.13–19, 2007.

PARSON, W.; GUSMÃO, L.; HARES, D. R.; IRWIN, J. A.; MAYR, W. R.; MORLING, N.; POKORAK, E.; PRINZ, M.; SALAS, A.; SCHNEIDER, P. M.; PARSONS, T. J.; DNA COMMISSION OF THE INTERNATIONAL SOCIETY FOR FORENSIC GENETICS. DNA Commission of the International Society for Forensic Genetics: revised and extended guidelines for mitochondrial DNA typing. **Forensic science international: Genetics**, v.13, p.134–142, 2014. DOI:10.1016/j.fsigen.2014.07.010.

PARSONS, T. J.; COBLE, M. D. Increasing the forensic discrimination of mitochondrial DNA testing through analysis of the entire mitochondrial DNA genome. *Croatian Medical Journal*, v.42, p.304–309, 2001.

PARSONS, T. J.; MUNIEC, D. S.; SULLIVAN, K.; WOODYATT, N.; GREINER, R. A.; WILSON, M. R.; BERRY, D. L.; HOLLAND, K. A.; WEEDN, V. W.; GILL, P.; HOLLAND, M. M. A high observed substitution rate in the human mitochondrial DNA control region. **Nature Genetics**, v.15, p.363–368, 1997.

PENG, D., ZHANG, Y., REN, H. et al. Identification of sequence polymorphisms at 58 STRs and 94 iiSNPs in a Tibetan population using massively parallel

sequencing. **Scientific Reports – Nature**, v.10, p.12225, 2020.

<https://doi.org/10.1038/s41598-020-69137-1>

PERLIN, M. W.; LEGLER, M. M.; SPENCER, C. E.; SMITH, J. L.; ALLAN, W. P.; BELROSE, J. L.; DUCEMAN, B. W. Validating TrueAllele® DNA mixture interpretation. **Journal of forensic sciences**, v.56, n.6, p.1430–1447, 2011. DOI:10.1111/j.1556-4029.2011.01859.x.

PFAFF, C. L.; PARRA, E. J.; BONILLA, C.; HIESTER, K.; MCKEIGUE, P. M.; KAMBOH, M. I.; HUTCHINSON, R. G.; FERRELL, R. E.; BOERWINKLE, E.; SHRIVER, M. D. Population structure in admixed populations: effect of admixture dynamics on the pattern of linkage disequilibrium. **American journal of human genetics**, v.68, n.1, p.198–207, 2001. doi:10.1086/316935

PHILLIPS, C.; BALLARD, D.; GILL, P.; COURT, D. S.; CARRACEDO, A.; LAREU, M. V. The recombination landscape around forensic STRs: Accurate measurement of genetic distances between syntenic STR pairs using HapMap high density SNP data. **Forensic Science International: Genetics**, v.6, p.354–65, 2012.

PHILLIPS, C.; FONDEVILA, M.; GARCÍA-MAGARIÑOS, M.; RODRIGUEZ, A.; SALAS, A.; CARRACEDO, A.; LAREU, M. V. Resolving relationship tests that show ambiguous STR results using autosomal SNPs as supplementary markers. **Forensic Science International**, v.2, p.198– 204, 2008. DOI: 10.1016/j.fsigen.2008.02.002.

PUCH-SOLIS, R.; CLAYTON, T. Evidential evaluation of DNA profiles using a discrete statistical model implemented in the DNA LiRa software. **Forensic Science International: Genetics**, v.11, p.220–228, 2014.

RAGSDALE, A. P. & GRAVEL, S. Models of archaic admixture and recent history from two locus statistics. **PLoS Genetic**, v.15, n.6, p.e1008204, 2019. doi:10.1371/journal.pgen.1008204

RANGEL-VILLALOBOS, H.; AGUILAR-VELÁZQUEZ, J.; GARCÍA-ACEVES, M.; GUTIÉRREZ-GONZÁLEZ, C.; MARTÍNEZ-CORTÉS, G. Frequency of null alleles in paternity cases from Mexico for the locus Penta E. **Forensic Science International: Genetics Supplement Series**, v.7, n.1, p.617-619, 2019. doi: 10.1016/j.fsigss.2019.10.112.

RASOOL, N. & HUSSAIN, W. ForeStatistics: A windows-based feature-rich software program for performing statistics in Forensic DNA analysis, Paternity and relationship testing. **Forensic Science International**, v.307, p.110142, 2020. doi:10.1016/j.forsciint.2020.110142.

RIANCHO, J. A. & ZARRABEITIA, M. T. A Windows-based software for common paternity and sibling analyses. **Forensic Science International**, v.135, p.232–234, 2003.

ROBERTSON, B.; VIGNAUX, G. A. Unhelpful Evidence in Paternity Cases. **New Zealand Law Journal**, v. 1992, n. 9, p.315-317, 1992.

- RÖCK, A.; IRWIN, J.; DÜR, A.; PARSONS, T.; PARSON, W. SAM: String-based sequence search algorithm for mitochondrial DNA database queries, **Forensic Science International: Genetics**, v.5, p.126–132, 2011.
- ROEDER, K. DNA fingerprinting: a review of the controversy. **Statistical Science**, v.9, p.222–278, 1994.
- ROEWER, L.; ANDERSEN, M. M.; BALLANTYNE, J.; BUTLER, J. M.; CALIEBE, A.; CORACH, D.; D'AMATO, M. E.; GUSMÃO, L.; HOU, Y.; KNIJFF, P.; PARSON, W.; PRINZ, M.; SCHNEIDER, P. M.; TAYLOR, D.; VENNEMANN, M.; WILLUWEIT, S. DNA commission of the International Society of Forensic Genetics (ISFG): Recommendations on the interpretation of Y-STR results in forensic analysis. **Forensic Science International: Genetics**, v.48, p.102308, 2020. doi: 10.1016/j.fsigen.2020.102308.
- ROEWER, L.; KAYSER, M.; DE KNIJFF, P.; ANSLINGER, K.; BETZ, A.; CAGLIA, A.; CORACH, D.; FUREDI, S.; HENKE, L.; HIDDING, M.; KARGEL, H.J.; LESSIG, R.; NAGY, M.; PASCALI, V.L.; PARSON, W.; ROLF, B.; SCHMITT, C.; SZIBOR, R.; TEIFEL-GREDING, J.; KRAWCZAK, M. A new method for the evaluation of matches in non-recombining genomes: application to Y-chromosomal short tandem repeat (STR) haplotypes in European males. **Forensic Science International**, v.114, p.31–43, 2000.
- RUDIN, N.; INMAN, K.; STOLOVITZKY, G.; RIGOUTSOS, I. DNA Based Identification, in: Jain, A. K.; Bolle, R.; Pankanti, S.; (eds.), **Biometrics**, Springer US, 1996; pp. 287-309.

SCHANFIELD, M. S. Parentage Testing, in: Siegel, J.; Saukko, P. (eds.),
Encyclopedia of Forensic Sciences. London: Academic Press, 2000.

SCHNEIDER, P. M. Scientific standards for studies in forensic genetics. **Forensic science international**, v.165, n.2-3, p.238–243, 2007. DOI: 10.1016/j.forsciint.2006.06.067

SCHOSKE, R.; VALLONE, P. M.; KLINE, M. C.; REDMAN, J. W. et al. High throughput Y-STR typing of U.S. populations with 27 regions of the Y chromosome using two multiplex PCR assays. **Forensic Science International**, v.139, n.2-3, p.107-121, 2004.

SENSABAUGH, G. F. Biochemical markers of individuality. In: Saferstein, R, (ed.),
Forensic science handbook. Prentice-Hall, Englewood Cliffs, pp 340-415, 1982.

SHENG, X.; BAO, Y.; ZHANG, J. S.; LI, M.; LI, Y. N.; XU, Q. N.; ZHANG, S. H.; LI, C. T. Research Progress on InDel Genetic Marker in Forensic Science. **Fa yi xue za zhi**, v.34, n.4, p.420–427, 2018. doi:10.12116/j.issn.1004-5619.2018.04.016

SHORT, S. DNA: Basic Principles. **Encyclopedia of Forensic and Legal Medicine**. Elsevier, 2016, p324–329. doi:10.1016/b978-0-12-800034-2.00151-8

SIEGEL, J. A.; SAUKKO, P. J.; HOUCK, M. M. **Encyclopedia of Forensic Sciences**, 2ed, Academic Press, 2013.

SILVA, N. M.; PEREIRA, L.; POLONI, E. S.; CURRAT, M. Human Neutral Genetic Variation and Forensic STR Data. **PLoS one**, v.7, n.11, e49666, 2012. DOI:10.1371/journal.pone.0049666.

SILVA, N. M.; PEREIRA, L.; POLONI, E. S.; CURRAT, M. Human neutral genetic variation and forensic STR data. **PloS one**, v.7, n.11, p.e49666, 2012. doi:10.1371/journal.pone.0049666

SINHA, S.; MURPHY, G.; BROWN, H.; GREEN, A.; MONTGOMERY, A.; CARROLL, M.; TABAK, J. Retrotransposable elements: Novel and sensitive DNA markers and their application in human identity. **Forensic Science International: Genetics Supplement Series**, v.5, p.e627 - e629, 2015.

SKALETSKY, H.; KURODA-KAWAGUCHI, T.; MINX, P. J.; CORDUM, H. S.; HILLIER, L.; BROWN, L. G.; REPPING, S.; PYNTIKOVA, T.; ALI, J.; BIERI, T.; CHINWALLA, A.; DELEHAUNTY, A.; DELEHAUNTY, K.; DU, H.; FEWELL, G.; FULTON, L.; FULTON, R.; GRAVES, T.; HOU, S. F.; LATRIELLE, P.; ... PAGE, D. C. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. **Nature**, v.423, n.6942, p.825–837, 2003. DOI:10.1038/nature01722.

SKARE, Ø.; SHEEHAN, N.; EGELAND, T. Identification of distant family relationships. **Bioinformatics**, v.25, p.2376-2382, 2009.

SLOOTEN, K. Match probabilities for multiple siblings. **Forensic Science International: Genetics**, v.6, p.466–468, 2012.

SONG, X.; ZHOU, Y.; YING, B.; WANG, L.; LI, Y.; LIU, J.; BAI, X.; LU, X.; WANG, J.; YE, Y. Short-tandem repeat analysis in seven Chinese regional populations. **Genetics and Molecular Biology**, v.33, n.4, p.605-609, 2010.

SOUSA, V. C.; BEAUMONT, M. A.; FERNANDES, P.; COELHO, M. M.; CHIKHI, L. Population divergence with or without admixture: selecting models using an ABC approach. **Heredity**, v.108, n.5, p.521–530, 2012. doi: 10.1038/hdy.2011.116

SRIVASTAVA, D.; ..., SIDDIQUI, M. H. Modulation of Gene Expression by Microsatellites in Microbes in New and Future Developments, in: Singh, H. B.; Gupta, V. K.; Jogaiah, S. (eds.), **Microbial Biotechnology and Bioengineering**, Elsevier, 2019. doi: 10.1016/C2014-0-00307-3

STEELE, C. D. & BALDING, D. J. Choice of population database for forensic DNA profile analysis. **Science & Justice: Journal of the Forensic Science Society**, v.54, n.6, p.487-493, 2014. <https://doi.org/10.1016/j.scijus.2014.10.004>

STEELE, C. D.; GREENHALGH, M.; BALDING, D. J. Verifying likelihoods for low template DNA profiles using multiple replicates. **Forensic Science International: Genetics**, v.13, p.82–89, 2014. DOI:10.1016/j.fsigen.2014.06.018

STEWART, J. E.; FISHER, C. L.; AAGAARD, P. J.; WILSON, M. R.; ISENBERG, A. R.; POLANSKEY, D.; POKORAK, E.; DIZINNO, J. A.; BUDOWLE, B.

Length variation in HV2 of the human mitochondrial DNA control region.

Journal of Forensic Sciences, v.46, p.862–870, 2001.

STIGLER, S. M. Thomas Bayes's Bayesian Inference. **Journal of the Royal Statistical Society. Series A (General)**, v.145, n.2, p. 250-258, 1982. DOI: 10.2307/2981538.

SWGDM - Scientific Working Group on DNA Analysis Methods. **Addendum to SWGDAM interpretation guidelines for autosomal STR typing by forensic DNA testing laboratories to address next generation sequencing**. Technical report, 2019.

SWGDM - Scientific Working Group on DNA Analysis Methods. **Recommendations of the SWGDAM Ad Hoc Working Group on Genotyping Results Reported as Likelihood Ratios, 2017**. Disponível em: <https://www.swgdam.org/publications> e notificado em <https://www.swgdam.org/faq> em 2018. Acesso em :18 fev 2021.

SWGDM. Guidelines for mitochondrial DNA (mtDNA) nucleotide sequence interpretation. **Forensic Science Communications**, v.5, n.2, 2003. Disponível em: <http://www2.fbi.gov/hq/lab/fsc/backissu/april2003/swgdammitodna.htm>. Acesso em: 08 jan. 2020.

SWGDM. **Interpretation Guidelines for Y-Chromosome STR Typing**. 2014. Disponível em: http://swgdam.org/SWGDM_YSTR_Guidelines_APPROVED_01092014_v_02112014_FINAL.pdf. Acesso em: 13 jan. 2020.

- TAYLOR, D.; BUCKLETON, J.; EVETT, I. Testing likelihood ratios produced from complex DNA profiles. **Forensic Science International: Genetics**, v.16, p.165–171, 2015.
- TAYLOR, D.; CURRAN, J.; BUCKLETON, J. Likelihood ratio development for mixed Y-STR profiles. **Forensic Science International: Genetics**, v.35, p.82–96, 2018.
- THOMPSON, R.; ZOPPIS, S.; MCCORD, B., IN: CAPUTO, M.; IRISARRI, M.; ALECHINE, E. CORACH, D. A DNA extraction method of small quantities of bone for high-quality genotyping. **Forensic science international: Genetics**, v.7, n.5, p.488–493, 2013. doi:10.1016/j.fsigen.2013.05.002.
- THORNTON, T.; TANG, H.; HOFFMANN, T. J.; OCHS-BALCOM, H. M.; CAAN, B. J.; & RISCH, N. Estimating kinship in admixed populations. **American journal of human genetics**, v.91, n.1, p.122–138, 2012. doi: 10.1016/j.ajhg.2012.05.024
- TILLMAR, A. O.; KLING, D.; BUTLER, J. M.; PARSON, W.; PRINZ, M.; SCHNEIDER, P. M.; EGELAND, T.; GUSMÃO, L. DNA Commission of the International Society for Forensic Genetics (ISFG): Guidelines on the use of X-STRs in kinship analysis. **Forensic science international: Genetics**, v.29, p.269-275, 2017.
- TRIGGS, C.; WALSH, K. A. J.; BUCKLETON, J.S. Assessing probabilities considering eyewitness evidence. **Science & Justice**, v.35, p.263–266, 1995.

TRIPATHI, V. A. The Effects of Forensic DNA Typing on the FRR and FAR of a Biometric System. **SCJAS**, 2015.

TSYBOVSKII, I. S.; VEREMEICHIK, V. M.; KOTOVA, S. A.; KRITSKAYA, S. V.; EVMENENKO, S. A.; & UDINA, I. G. Developing forensic reference database by 18 autosomal STR for DNA identification in Republic of Belarus. **Russian Journal of Genetics**, v.53, p.275–284, 2017. doi:10.1134/S1022795417020132

TULLY, G.; BÄR, W.; BRINKMANN, B.; CARRACEDO, A.; GILL, P.; MORLING, N.; PARSON, W.; SCHNEIDER, P. Considerations by the European DNA profiling (EDNAP) group on the working practices, nomenclature and interpretation of mitochondrial DNA profiles. **Forensic Science International**, v.124, p.83–91, 2001.

VANRELL, J. P. **Odontologia legal e antropologia forense**. 1 ed. Rio de Janeiro: Guanabara Koogan, 2002.

VIEIRA, M. L.; SANTINI, L.; DINIZ, A. L.; & MUNHOZ, C. Microsatellite markers: what they mean and why they are so useful. **Genetics and molecular biology**, v.39, n.3, p.312–328, 2016. doi:10.1590/1678-4685-GMB-2016-0027

VIRMOND, M. B.; ROBERT, A. W.; BRITO, P. B.; MASSUDA, T. Y. C. Fenotipagem forense pelo DNA através de SNPs. **Revista Brasileira de Ciências Criminais**, v.5, n.2, p.37-47, 2016.

WAHLUND, S. Zusammensetzung von Population und Korrelationserscheinung vom Standpunkt der Vererbungslehre aus betrachtet. **Hereditas**, v.11, p.65-106, 1928.

WALSH, K. A. J.; BUCKLETON, J. S. Calculating the frequency of occurrence of a blood type for a "random man." **Journal of the Forensic Science Society**, v.31, p.49–58, 1991.

WALSH, K. A. J.; BUCKLETON, J. S.; TRIGGS, C. M. Assessing prior probabilities considering geography. **Journal of the Forensic Science Society**, v.34, p.47–51, 1994.

WALSH, K.; BUCKLETON, J. S. A discussion of the law of mutual independence and its application to blood-group frequency data. **Journal of the Forensic Science Society**, v.28, p.95–98, 1988.

WATSON, J. D.; BAKER, T. A.; BELL, S. P.; GANN, A.; LEVINE, M.; LOSICK, R. **Biologia Molecular do Gene**. 5 ed. São Paulo: Artmed editora, 2004.

WEBER-LEHMANN, J.; SCHILLING, E.; GRADL, G.; RICHTER, D. C.; WIEHLER, J.; ROLF, B. Finding the needle in the haystack: differentiating "identical" twins in paternity testing and forensics by ultra-deep next generation sequencing. **Forensic science international: Genetics**, v.9, p.42–46, 2014. DOI:10.1016/j.fsigen.2013.10.015.

WEINBERG, W. Über den Nachweis der Vererbung beim Menschen. **Jahresh. Verein f. Varerl Naturk.Wurtemb.**, v.64, p.368–382, 1908.

WEIR, B. S. Forensic Population Genetics and the National Research Council (NRC). **American Journal of Human Genetics**, v.52, p.437, 1993a.

WEIR, B. S. **Genetic Data Analysis, Vol. II.** Sinauer, MA: Sunderland, 1996.

WEIR, B. S. **Human identification: The use of DNA markers.** Dordrecht: Kluwer Academic Publishers, 1995.

WEIR, B. S. Independence tests for VNTR alleles defined as quantile bins. **American Journal of Human Genetics**, v.53, p.1107–1113, 1993b.

WIELBO, D. Fingerprints, in: Siegel, J.; Saukko, P. (eds.), **Forensic Science in encyclopedia of forensic sciences.** London: Academic Press, 2000.

WRIGHT, S. The genetical structure of populations. *Annals of Eugenics*, v.15, p.159-171, 1951.

WYNER, N.; BARASH, M.; MCNEVIN, D. Forensic Autosomal Short Tandem Repeats and Their Potential Association with Phenotype. **Frontiers in Genetics**, v.11, p.884, 2020. doi: 10.3389/fgene.2020.00884

ZHANG, Y. & LIU, S. L. Bioinformatics, in Maloy, S.; Hughes, K (eds.), **Brenner's Encyclopedia of Genetics.** 2 ed. Academic Press, 2013. p.338-340. ISBN 9780080961569.

ZHAO, H.; PFEIFFER, R.; GAIL, M. H. Haplotype analysis in population genetics and association studies. **Pharmacogenomics**, v.4, n.2, p.171–178, 2003.

ZHIVOTOVSKY, L. A., AHMED, S., WANG, W., & BITTLES, A. H. The forensic DNA implications of genetic differentiation between endogamous communities. **Forensic science international**, v.119, n.3, p.269–272, 2001.

ZHIVOTOVSKY, L. A.; MALYARCHUK, B. A.; DERENKO, M. V.; WOZNIAK, M.;
& GRZYBOWSKI, T. Developing STR databases on structured populations:
the native South Siberian population versus the Russian population.
Forensic science international. Genetics, v.3, n.4, p.e111–e116, 2009.
doi: 10.1016/j.fsigen.2008.08.001

ZHOU, Y.; QIU, H.; XU, S. Modeling Continuous Admixture Using Admixture-
Induced Linkage Disequilibrium. **Scientific Reports – nature**, v.7, p.43054,
2017. doi: 10.1038/srep43054

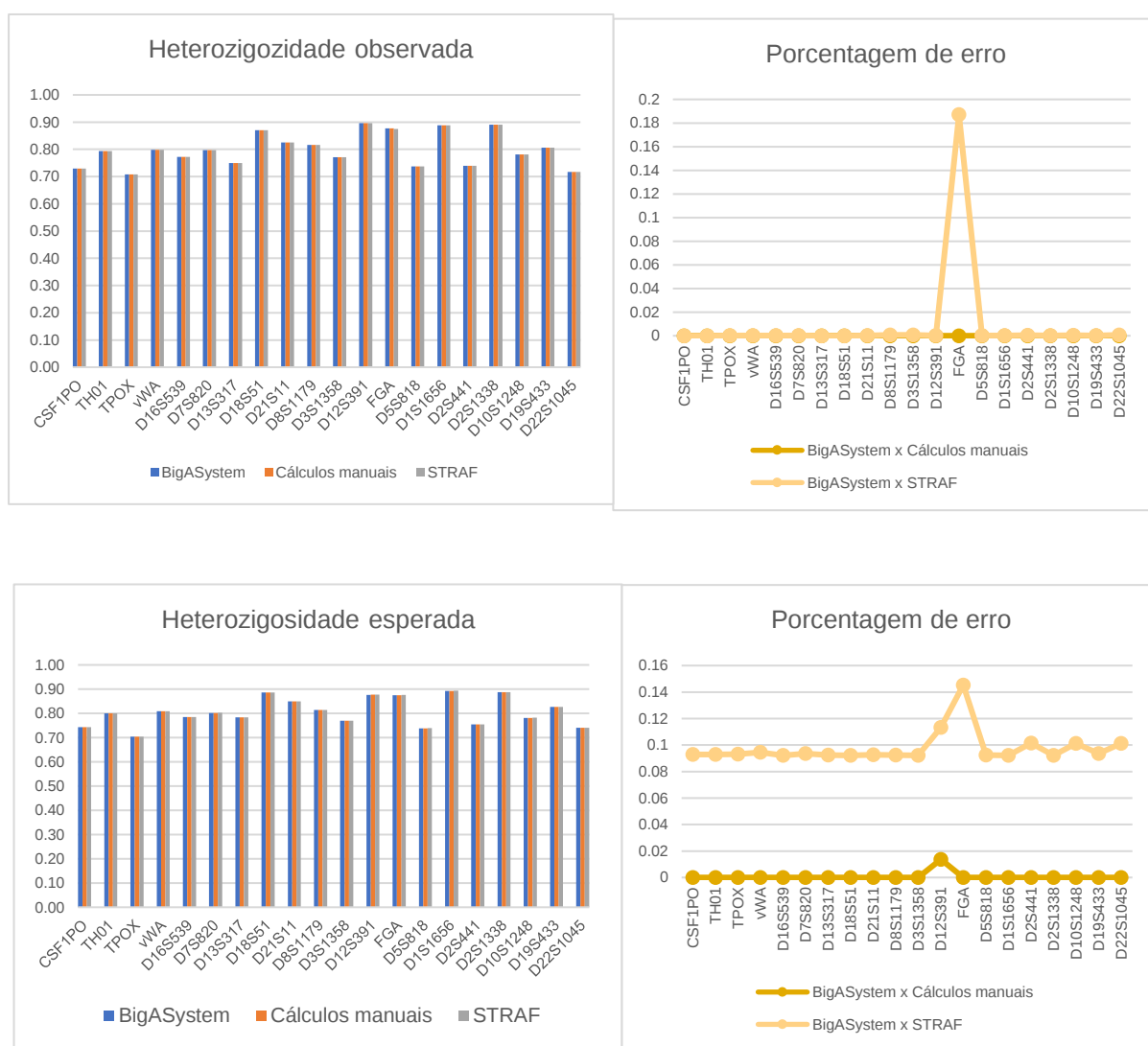
ZHOU, Y.; YAO, Y.; LIU, B.; YANG, Q.; ZHOU, Z.; SHAO, C.; LI, S.; TANG, Q.;
XIE, J. Characterizing Y-STRs in the Evaluation of Population Differentiation
Using the Mean of Allele Frequency Difference between Populations.
Genes, v.11, n.5, p.566, 2020. doi: 10.3390/genes11050566

APÊNDICE A – REPRESENTAÇÃO GRÁFICA DOS RESULTADOS DA VALIDAÇÃO DO BIGASYSTEM CITADOS NO ARTIGO 1

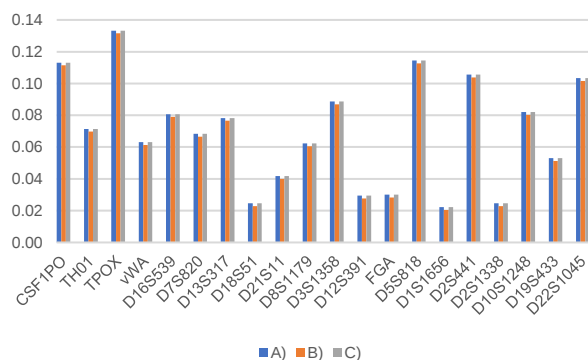
A. MÓDULO 1 (AUT-STRs) - FUNÇÃO BANCO DE DADOS

A.1. Valores dos parâmetros populacionais e porcentagem de erro

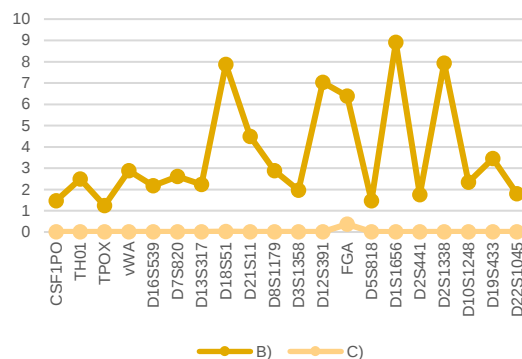
Foram analisados 540 perfis genéticos aut-STRs (20 locos do CODIS expandido) de indivíduos não relacionados de Pernambuco obtidos no Laboratório de Bioinformática e Biologia Evolutiva (LABBE-UFPE).



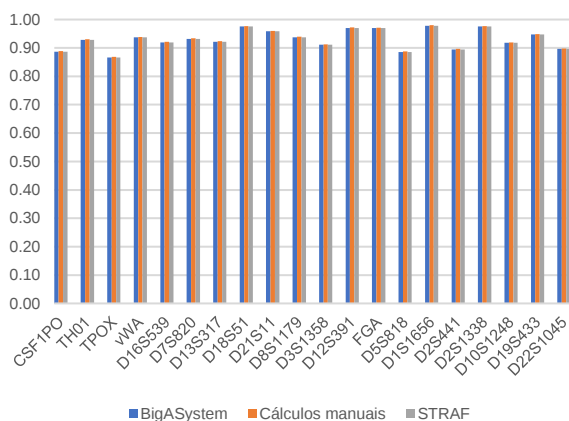
Probabilidade de coincidência



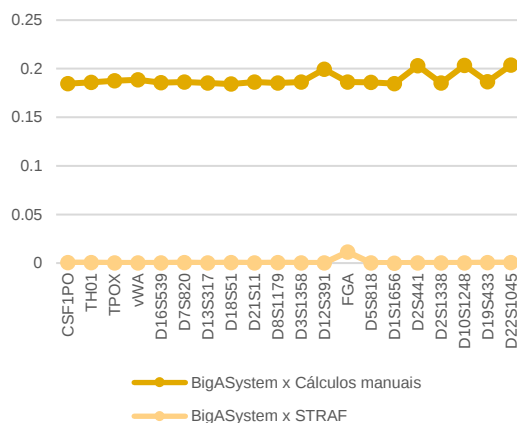
Porcentagem de erro



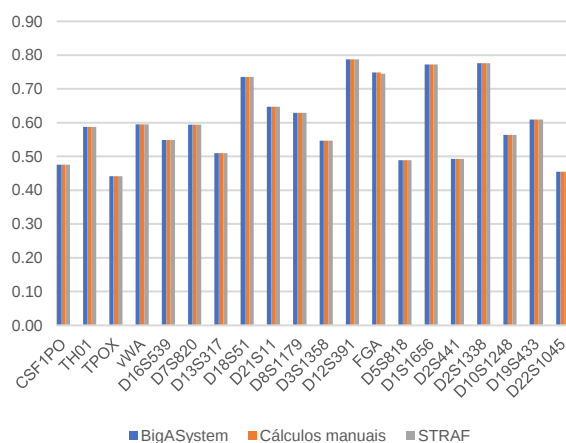
Probabilidade de discriminação



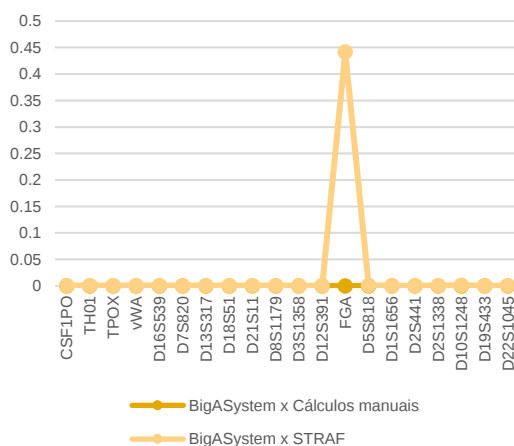
Porcentagem de erro

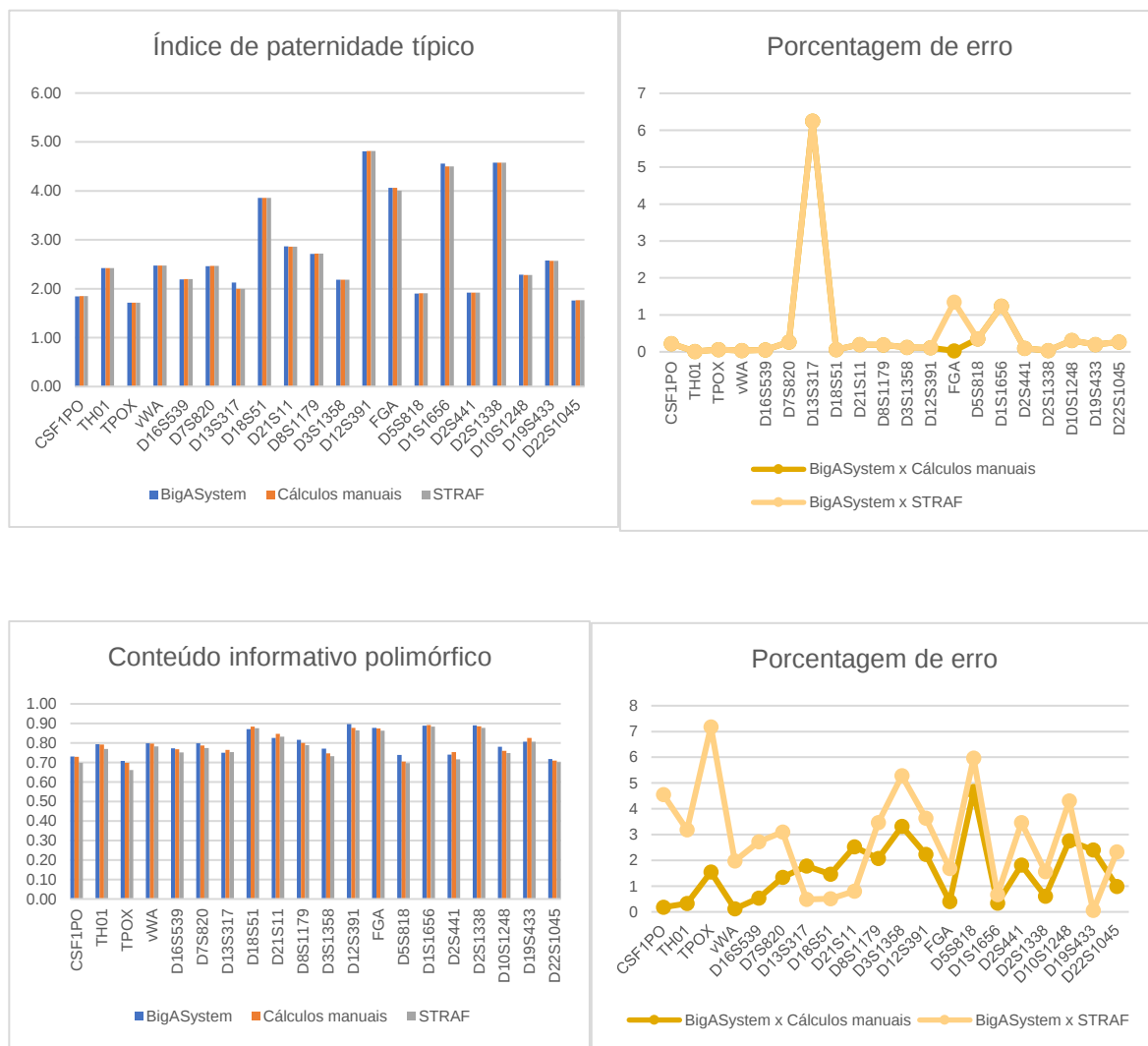


Poder de exclusão



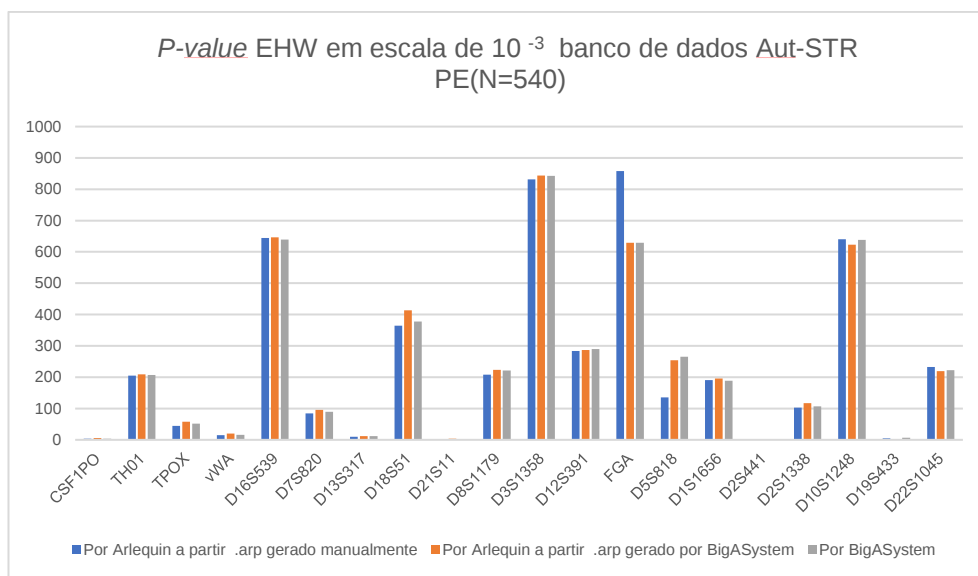
Porcentagem de erro





A.2.Valores de *P-values* de EHW

Foram analisados 540 perfis genéticos aut-STRs (20 locos do CODIS expandido) de indivíduos não relacionados de Pernambuco obtidos no Laboratório de Bioinformática e Biologia Evolutiva (LABBE-UFPE).

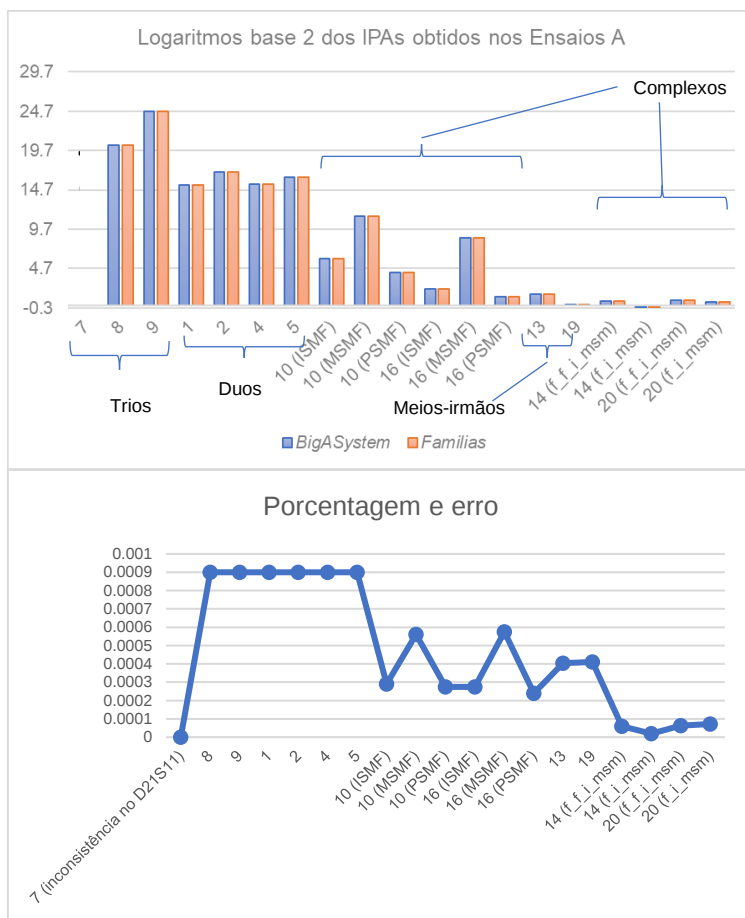


B. MÓDULO1 (AUT-STRs) - FUNÇÃO TESTE DE PATERNIDADE

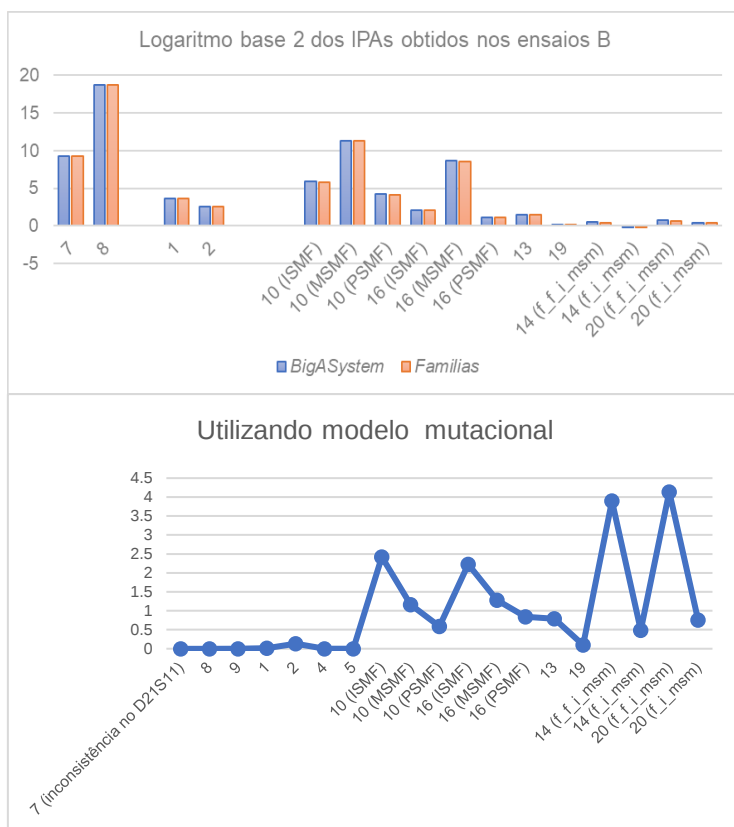
B.1.Valores de índice de paternidade acumulada e porcentagem de erro

Foram analisados treze casos forenses baseados no NIST (*O'Connor et al., 2010 - considerando os 15 locos correspondentes aos do multiplex *Identyfiler®* da *Applied Biosystems*) cujos números foram dados aleatoriamente visando apenas a discriminação: três trios -(casos 7, 8 e 9), quatro duos (1, 2, 4 e 5), dois meios-irmãos e quatro complexos - ausência do genitor questionado que pode ser o suposto pai (SP) ou suposta mãe SM) em casos de paternidade e maternidade respectivamente (10, 16, 14 e 20). Nos casos complexos foram utilizadas siglas que representam o parente do genitor alegado: ISMF – irmão(ã) de SM falecida (SMF), MSMF – mãe de SMF, PSMF – pai de SMF, f_f_i_msm – filho(a) do(a) filho(a) de ISMF, f_i_msm - filho(a) de ISMF.

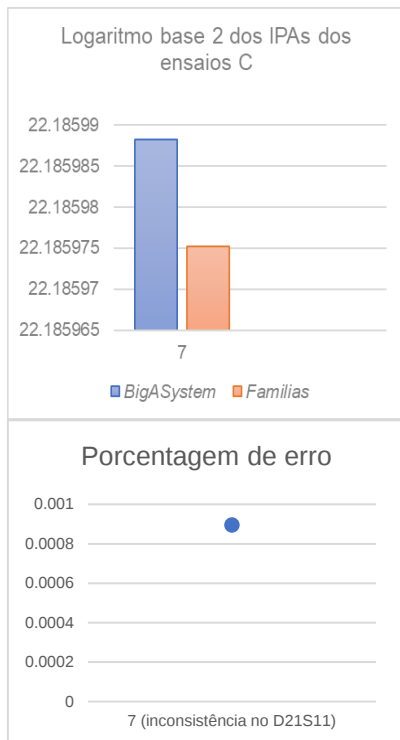
*K.L. O'Connor, S.P. Myers, E.L.R. Butts, C.R. Hill, J.M. Butler, P.M. Vallone, Candidate Reference Family Data: A Tool for Validating Kinship Analysis Software. Poster #35, 21th International Symposium on Human Identification San Antonio, TX, 11–14 (2010). Available in: http://www.cstl.nist.gov/biotech/strbase/pub_pres/Promega2010_OConnor.pdf



Sem considerar
modelo
mutacional e
MAF no cálculo
do IP



Considerando
modelo mutacional
no cálculo do IP

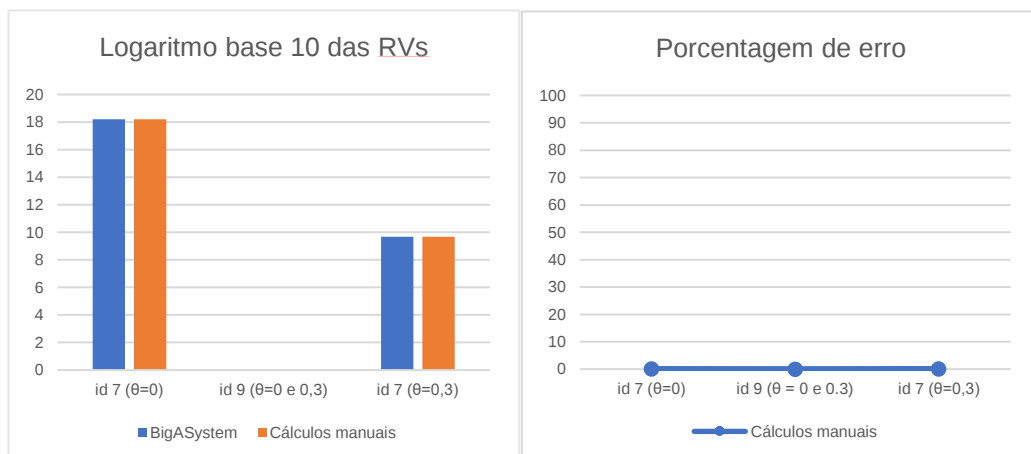


Considerando MAF
no cálculo do IP

C. MÓDULO 1 (AUT- STRs) - FUNÇÃO IDENTIFICAÇÃO FORENSE

C.1. Valores de razão de verossimilhança e porcentagem de erro

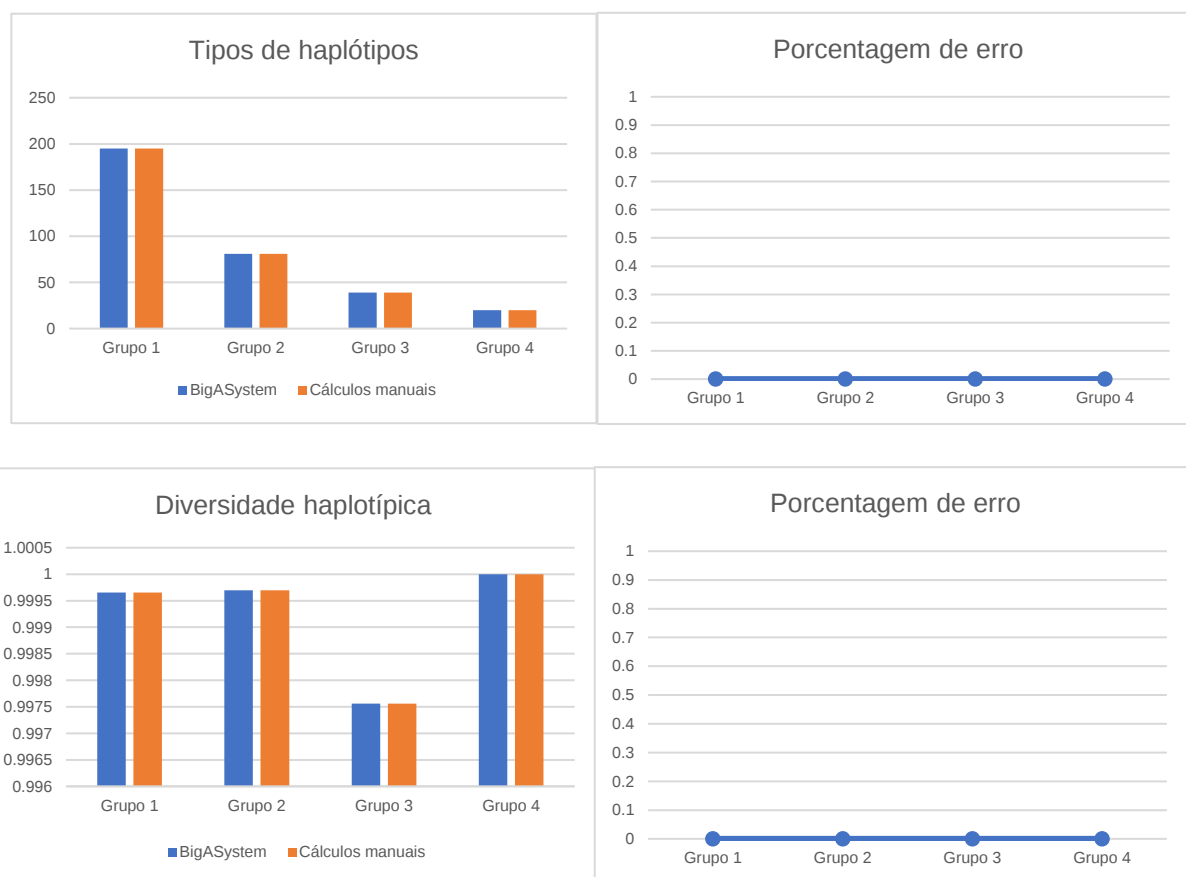
Foram analisados dois casos de identificação forense fictícios considerando 15 locos correspondentes aos do multiplex *Identyfiler*® (da *Applied Biosystems*) e coeficientes de coancestralidade populacionais iguais a zero e a 0,3 (valores escolhidos aleatoriamente – veja que este último tem um valor elevado para uma população humana).

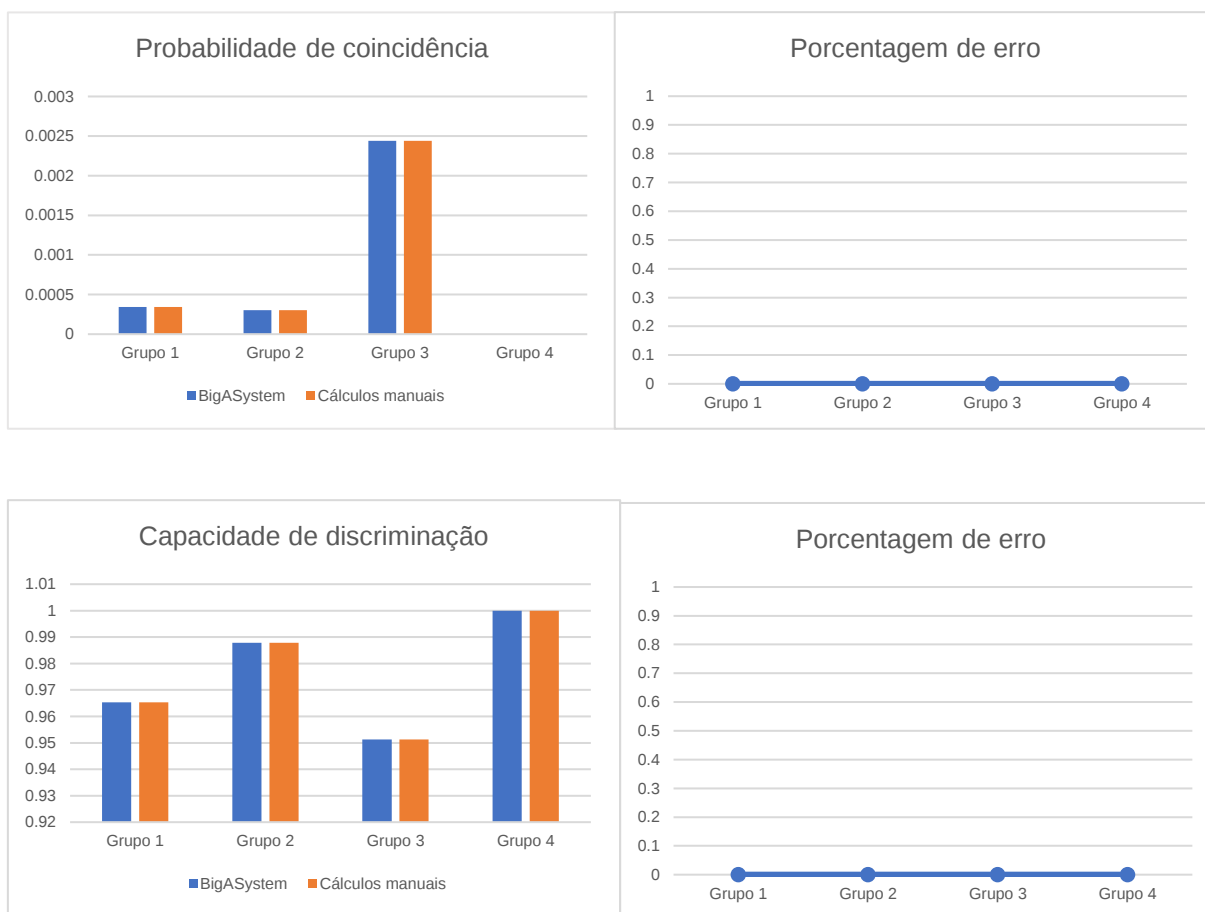


D. MÓDULO 2 (Y-STRs) - FUNÇÃO BANCO DE DADOS

D.1. Valores de parâmetros populacionais e porcentagem de erro

Foram analisados 202 perfis haplotípicos 17 marcadores Y-STRs (DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS385a/b, DYS438, DYS439, DYS437, DYS448, DYS456, DYS458, DYS635 e Y-GATA H4) obtidos de indivíduos não relacionados. Esses perfis foram disponibilizados pelo LABBE -UFPE.

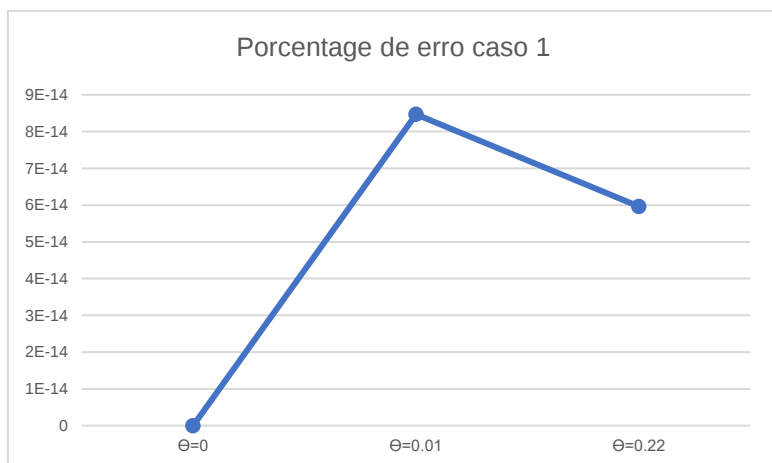
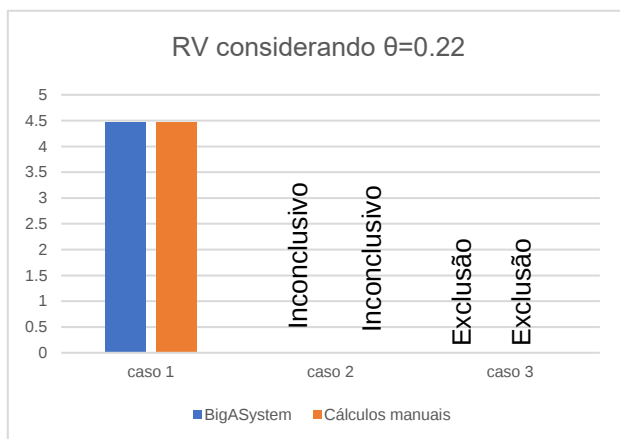
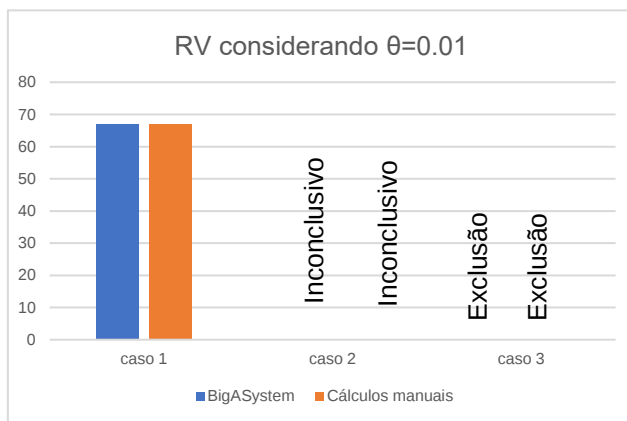
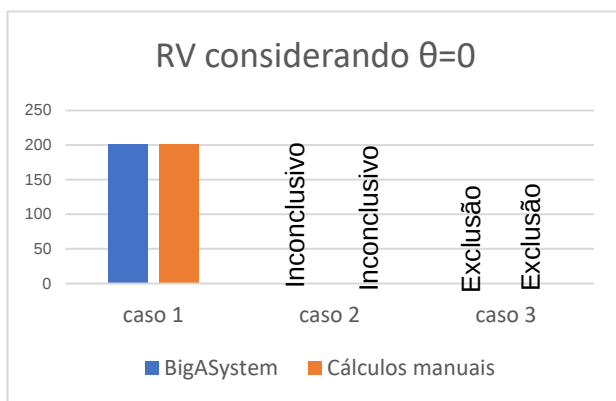




E. MÓDULO 2 (Y-STRs) - FUNÇÃO IDENTIFICAÇÃO HUMANA

E.1. Valores de razão de verossimilhança e porcentagem de erro

Foram analisados três casos de identificação forense fictícios (um de não exclusão, outro de exclusão e um inconclusivo por 17 locos Y-STRs) por teste de linhagem com amostra de referência direta considerando valor de coancestralidade populacional igual a 0, 0.01 e 0.22 (escolhidos aleatoriamente).



**APÊNDICE B – ARTIGO: ANÁLISE POPULACIONAL DE 20 LOCOS STR CODIS
EXPANDIDO NA POPULAÇÃO DE PERNAMBUCO, REGIÃO NORDESTE DO
BRASIL**

**(A SER SUBMETIDO COMO SHORT COMMUNICATIONS AO PERIÓDICO
GENETICS AND MOLECULAR BIOLOGY)**

Abigail Marcelino dos Santos Silva¹, Sérgio de Sá Leitão Paiva Junior^{1,2}, Valdir de
Queiroz Balbino¹

¹ Departamento de genética – Universidade Federal de Pernambuco

² Unidade Acadêmica de Serra Talhada – Universidade Federal Rural e Pernambuco

Resumo

Foram obtidos dados genéticos populacionais a partir da análise perfis genotípicos dos 20 short tandem repeats (STR) CODIS EXTENDIDO presentes em até 540 indivíduos não aparentados do estado de Pernambuco. Os STR são locos utilizados na área de genética forense. Atualmente um dos conjuntos locos mais utilizados para estas finalidades são os 20 STR do CODIS expandido, um programa do Federal Bureau of Investigation (FBI) para bancos de dados de DNA. Corresponde aos primeiros passos de um trabalho que abrange toda a região Nordeste do Brasil. Dentre os dados genéticos populacionais estão as frequências alélicas, parâmetros de genética forenses e comparação com populações de outros estados do Brasil. O loco D1S1656 apresentou-se com maior polimorfismo em relação aos demais e D1S1656 apresentou maior poder de discriminação. Todos os locos estão em equilíbrio de Hardy-Weinberg, alguns após a correção de Bonferroni.

Palavras chaves: frequências alélicas, dados populacionais, STRs autossômico, CODIS Expandido, genética forense, Pernambuco, Nordeste do Brasil

O estado de Pernambuco contém aproximadamente 8.800.000 habitantes (ver em <http://www.ibge.gov.br/estadosat/perfil.php?sigla=pe#>) e está localizado na região Nordeste do Brasil. Esta região é caracterizada como exogâmica e mista (originada de indivíduos ancestrais Africanos, Europeus e norte-americanos nativos). Essa característica revigora ainda mais a importância de estudo populacional humano, o qual ainda é muito reduzido na região Nordeste inclusive no estado de Pernambuco.

Um dos dados moleculares para este estudo são as regiões polimórficas do DNA. Os polimorfismos, quando analisados por marcadores genéticos, tornam-se elementos importantes para os estudos intra e inter-populacional, por distinguirem as diferentes versões alélicas de um loco a vários locos, herdadas ou não, de um gene ou segmentos do genoma (sequência de DNA), fornecendo ferramentas para uma ampla gama de aplicações. Essas aplicações podem ser em estudos populacionais propriamente ditos como também em áreas que dependem diretamente desses estudos populacionais e marcadores polimórficos, como a genética forense.

Os marcadores genéticos mais utilizados nas investigações forenses cíveis ou criminais, como investigação de parentesco e identificação humana, são os *shorts tandem repeats* (STRs). As sequências apresentam vários tamanhos e são classificadas pelo tamanho da unidade de repetição e pelo número de vezes que se repetem. Além dos próprios perfis genéticos STRs dos indivíduos investigados, as frequências alélicas dos STRs são dados utilizados na investigação e são utilizados como bancos de dados para uso na Genética Forense e de populações.

Este trabalho teve como objetivo a obtenção da frequência alélica estimada dos 20 marcadores autossômicos do CODIS expandido a partir de dados da população Pernambucana, além de parâmetros populacionais de interesse forense e comparação com dados da população de outros estados brasileiros.

Para isso, mostras de sangue de quinhentos e quarenta indivíduos (317 machos e 223 fêmeas) não relacionados da população do estado de Pernambuco foram obtidas de casos de paternidade do Laboratório de Bioinformática e Biologia Evolutiva da Universidade Federal de Pernambuco. O tamanho da amostra varia de loco a loco devido a utilização de diferentes kits multiplex para amplificá-la, além do insucesso da amplificação em alguns locos. O DNA foi extraído de células do sangue periférico a partir do método do cartão FTA (Whatman Bioscience, Cambridge, UK) seguindo as instruções do fabricante. A amplificação dos STRs alvos DNA foi realizada de acordo com as recomendações do fabricante dos kits multiplex FUSION (Promega Corporation, Madison, WI) e GLOBAL FILE (Thermo-Fisher, Waltham, MA USA). Os 20 locos foram tipados por eletroforese capilar via sequenciador ABI 3.500 Genetic Analyzer (Applied Biosystems, USA). A identificação dos alelos foi feita por comparação dos fragmentos amplificados com os allelic ladders do kit de amplificação multiplex e segundo diretrizes do ISFG (Bodnera *et al.*, 2016). A amplificação e a tipagem foram realizadas com controles positivos e negativos segundo instruções do kit multiplex Global Filer e Fusion (controles positivos – DNA CONTROL 007 e 2800M, respectivamente) (Applied Biosystems). A identificação dos alelos foi feita por comparação dos fragmentos amplificados com os *allelic ladders* e foram visualizados pelo *software* OSIRIS v.2.9 (Goor *et al.*, 2011).

As frequências dos alelos, o polimorfismo conteúdo de informação (PIC), o poder de discriminação (PD), poder de exclusão (PE), probabilidade de coincidência ou

match (MP) e típica índice de paternidade (TPI) foram calculados para cada loco usando o software GenAEx v. 6.5 (Peakall and Smouse, 2012) e planilhas eletrônicas. O Arlequin v. 3. 5 (Excoffier and Lischer, 2010) foi usado para calcular o valor P do teste exato de Equilíbrio de Hardy- Weinberg (EHW) e a comparação entre a população pernambucana com outras populações brasileiras com dados conjugados [20, 21] (Aguiar *et al.*, 2012; Aguiar *et al.*, 2013) ou não (Francez *et al.*, 2011; Nascimento *et al.*, 2011; Farias *et al.*, 2013; Dalton *et al.*, 2009; Wolfgramm *et al.*, 2011; Vieira *et al.*, 2013; Alves *et al.*, 2014; Rodenbusch *et al.*, 2012; Chula *et al.*, 2009; Góes *et al.*, 2004; Ocampos *et al.*, 2009; Sousa *et al.*, 2014; Fridman *et al.*, 2008; Fridman, 2008) utilizando dados de frequência alélicas dos 13 locos do CODIS (somente os 13 devido ao número limitado de locos STRs autossômicos estudados no Brasil) por Análise distância aos pares (FST) e diferenciação da população. Também foi realizado pelo Arlequin com estes dados o teste de variação genética (AMOVA). A construção dendrograma baseada na distância genética de Nei foi realizada no DISPAN (Ota, 1993), utilizando dados de frequência alélica de 13 locos (nos estudos dos quais foram obtidas as frequências alélicas de locos STR autossômicos de vários locais no Brasil, excetuando Amapá já que só coincide em 12 locos, citados no próximo parágrafo só coincidem em 13 locos, os quais estão no CODIS - CSF1PO, TH01, TPOX, vWA, D16S539, D7S820, D13S317, D18S51, D21S11, D8S1179, D3S1358, FGA, D5S818). A visualização do dendrograma foi feita do MEGA v6 (Tamura *et al.*, 2013).

As frequências alélicas e os valores dos parâmetros populacionais obtidos para os 20 locos STRs autossômicos do CODIS EXPANDIDO na população de Pernambuco estão na tabela S1. Os valores referentes a comparações de distância genética F_{st} (DF_{st}), AMOVA e diferenciação populacional dos dados da população

pernambucana deste artigo com dados de 11 outras populações brasileiras (dos estados Amapá-AP, Bahia-BA, Ceará-CE, Espírito Santo-ES, Goiás-GO, Paraná-PR, Rio Grande do Sul-RS, Rio de Janeiro -RJ, Santa Catarina -SC, São Paulo-SP, mais Distrito Federal (ver mapa na figura 1) e ainda dados conjugados de vários estados brasileiros (Brasil –BR), totalizando 12 dados populacionais, estão apresentados nas tabelas S2 e S3, respectivamente.

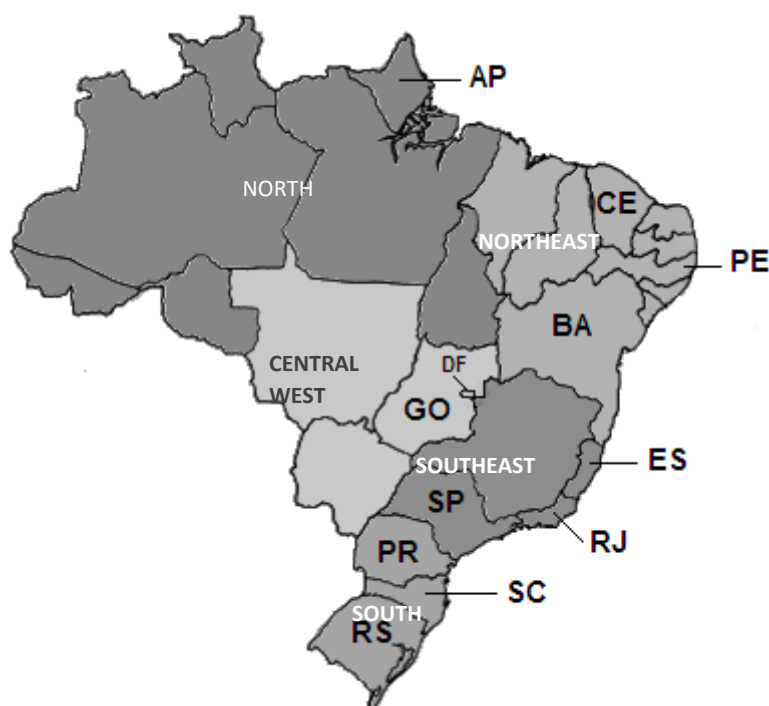


Figura 1 - Mapa do Brasil-BR (Baseado no IBGE - <http://mapas.ibge.gov.br/>). Estado(s) localizado(s) na região Norte: Amapá-AP; Nordeste: Bahia-BA, Ceará-CE, Pernambuco-PE; Centro-Oeste: Goiás- GO, Distrito Federal-DF (apesar de ser unidade federativa brasileira, não é considerado como um estado); Sudeste: Espírito Santo-ES, Rio de Janeiro-RJ, São Paulo-SP; Sul: Paraná-PR, Rio Grande do Sul-RS e Santa Catarina-SC

Todos os 20 locos do STR em Pernambuco avaliados nesta análise atingiu o Equilíbrio de Hardy-Weinberg (EHW) com $P > 0,05$, exceto os locos CSF1PO ($P = 0,0037$), TPOX ($P = 0,04481$), vWA ($P = 0,01502$), D13S317 ($P = 0,00986$), D2S441 ($P = 0,00104$), D19S433 ($P = 0,00508$) os quais após a Correção de Bonferroni ($P > 0,0025$) passaram a não ter mais diferenças estatisticamente significativas dos valores esperados no EHW. Comparando a população pernambucana com outras populações do nordeste e de outras regiões brasileiras, ela é geneticamente mais distante de Amapá para os marcadores vWA, D13S317, FGA, D5S818 (com DF_{st} de 0.00272, 0.00195, 0.00254, 0.00542, respectivamente), de Goiás em CSF1PO, TH01, D7S820, D21S11, D3S1358 (com DF_{st} de 0,00101, 0,03111, 0,00045, 0,18918, 0,00033), de São Paulo apenas para o loco D8S1179 ($DF_{st} = 0.06727$) e Rio Grande do Sul para o loco D16S539 ($DF_{st} = 0.00125$), Santa Catarina para TPOX (DF_{st} de 0.00145). Já comparando nossos resultados com a população brasileira mostrou um alto grau de concordância para todos marcadores, com DF_{st} próximos de 0.

O loco D21S11 apresentou-se com maior índice de variação genética entre indivíduos de diferentes populações, 1.99% e o D3S1358 o menor valor, com praticamente nenhuma variação, isto é, 0%.

Considerando que na comparação por diferenciação populacional duas populações são consideradas como significativamente diferentes quando o valor P apresenta-se menor do que o nível de significância para cada loco observa-se que Pernambuco está mais distante de Goiás, Santa Catarina, Amapá, São Paulo e Ceará ($P < 0,05/12$ em 3, 3, 1, 1 e 1, respectivamente, dos 13 locos CODIS original). Possivelmente isto pode ser justificado pela diferença da estrutura étnica além de Pernambuco estar geograficamente distante deles. Já na comparação por diferenciação de dados de PE com os dados conjugados brasileiros, observou-se

diferença significativa em 1 dos 13 locos. No dendograma apresentado na figura 2 observa-se que pela distância de Nei obtida a partir de um conjunto de 12 locos STRs autossômicos de Pernambuco não se agrupou com nenhum dos estados do Nordeste considerado. Vê-se claramente menor distância genética entre Bahia, Rio de Janeiro e Distrito Federal e entre São Paulo, Paraná, Santa Catarina e Rio Grande do Sul, que podem ser explicados pela aproximação geográfica e razões históricas coincidentes. É necessário um estudo mais detalhado para maior suporte desta análise. Inclusive uma análise de subestruturação da população brasileira.

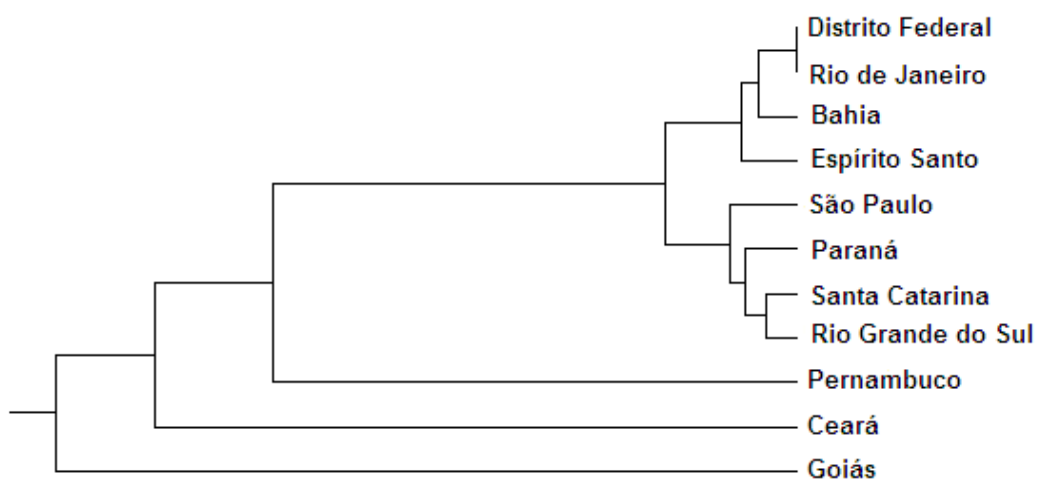


Figura 2 - Dendograma construído pelo método UPGMA a partir de dados de distância de Nei estimada para o conjunto de 12 locos STRs autossômicos.

Isso faz-nos perceber que nem sempre dados conjugados são representativos para todos os estados em todos os locos, ainda mais quando há prevalências de indivíduos de um dos estados brasileiros. A presença de diferenças significativas entre populações diferentes é importante para escolha de banco de dados na investigação

forense, pois na análise estatística pode levar a valores significativamente mais altos ou mais baixos do que o que se esperava.

No estado de Pernambuco a taxa de heterozigosidade observada (H_o) varia de 0,7084 (TPOX) a 0,8962 (D12S391), o poder de discriminação (PD) varia entre 0,8668 (TPOX) a 0,9779 (D1S1656). Já a probabilidade de exclusão (PE) varia de 0,4414 (TPOX) a 0,7877 (D12S391). Os marcadores genéticos mais polimórficos são D1S1656, D2S1338 e D18S51 ($PIC= 0,8919, 0,8854$ e $0,8833$, respectivamente) e os menos polimórficos são TPOX, D5S818 e D22S1045 ($PIC= 0,6976, 0,7050$ e $0,7099$, respectivamente). Os 20 locos juntos apresentaram: PD acumulado =1,0, PE acumulado=0,999999995, PM acumulado= $4,39363 \times 10^{-25}$ e TPI acumulado= $2,2 \times 10^{10}$.

Os parâmetros forenses relacionados aos 20 locos STR são úteis em termos de variabilidade e para fins forenses cíveis e criminais envolvendo a população Pernambuco. A perspectiva é fazer um estudo ainda mais amplo e a partir de dados buscados em um banco de dados único Brasileiro com regras padrão relacionadas tanto aos procedimentos de biologia molecular como também à genética de populações, geografia e antropologia. Isso garantiria além da uniformização dos dados para melhor interpretação, a qualidade dos dados. E esses dois pontos são imprescindíveis para maior veracidade dos estudos populacionais.

Agradecimentos

Este trabalho foi apoiado pelo Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

Referências

Aguiar VR, Wolfgramm EV, Malta FS, Bosque AG, Mafia AC, Almeida VC, Caxito FA, Pardini VC, Ferreira AC, Louro ID (2012) Updated Brazilian STR allele frequency data using over 100,000 individuals: an analysis of CSF1PO, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D21S11, FGA, Penta D, Penta E, TH01, TPOX and vWA locos, *Forensic Sci Int Genet.* 6(4): 504-9. DOI: 10.1016/j.fsigen.2011.07.005. Epub 2011 Aug 2.

Aguiar VRC, Wolfgramm EV, Malta FSV, Bosque A G, Mafia AC, Almeida VCO, Caxito FA, Pardini VC, Ferreira ACS, Louro ID (2013) Corrigendum to “Updated Brazilian STR allele frequency data using over 100,000 individuals: An analysis of CSF1PO, D3S1358, D5S818, D7S820, D8S1179, D13S317, D16S539, D18S51, D21S11, FGA, Penta D, Penta E, TH01, TPOX and vWA locos” [*Forensic Sci. Int.: Genet.* 6 (4) (2012) 504–509], *Forensic Science International: Genetics* 7: 321–325. DOI: <http://dx.doi.org/10.1016/j.fsigen.2011.07.005>

Alves HB, Leite FPN, Sotomaior VS, Rueda FF, Silva R, Moura-Neto RS (2014) STR data for 15 autosomal STR markers from Paraná (Southern Brazil), *Int J Legal Med* 128: 269–270. DOI: 10.1007/s00414-013-0859-y

Bodnera M, Bastisch I, Butler JM, Fimmers R, Gill P, Gusmão L, Morling N, Phillips C, Prinz M, Schneider PM, Parson W (2016) Recommendations of the DNA Commission of the International Society for Forensic Genetics (ISFG) on quality control of autosomal Short Tandem Repeat allele frequency databasing (STRidER), *Forensic Science International: Genetics* 24: 97–102. DOI: <http://dx.doi.org/10.1016/j.fsigen.2016.06.008>

Chula FGL, Rodenbusch R, Schumacher S, Grandi T, Michelon CT, Gastaldo AZ, Costi C, Carvalho B, da Silva CMD (2009) 15 STR locos frequencies with mutation rates in the population from Rio Grande do Sul, Southern Brazil, *Forensic Science International: Genetics* 3: e35–e38. DOI:10.1016/j.fsigen.2008.05.006

Dalton GC, Maia FAS, Mendes CRBO, Pak HI, de Paula KAA, Trindade-Filho A, Oliveira SF (2009) Genetic profile of Federal District of Brazil based on 18 STR autosomal locos. *Forensic Science International: Genetics Supplement Series* 2: 373–375. DOI:10.1016/j.fsigss.2009.08.131

Excoffier L and Lischer HEL (2010) Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows, *Molecular Ecology Resources* 10: 564-567. DOI: 10.1111/j.1755-0998.2010.02847.x

Farias DAB, Málaga SMR, Vasconcelos FR, Sales DC, Araújo RWB, Lopes MEM, Sá RCA, Guedes MIF, Builes JJ, Gaviria A (2013) Population genetics data on 15 STR markers in the population of Ceará State, Northeast Region of Brazil, *Forensic Science International: Genetics Supplement Series* 4: e250–e251. DOI: <http://dx.doi.org/10.1016/j.fsigss.2013.10.128>

Francez PAC, Rodrigues EMR, Frazão GF, Borges NDR, Santos SEB (2011) Allelic frequencies and statistical data obtained from 12 codis STR locos in an admixed population of the Brazilian Amazon, *Genetics and Molecular Biology*, 34 1: 35-39. DOI: <http://dx.doi.org/10.1590/S1415-47572011000100007>

Fridman C (2008) Erratum to “Brazilian population profile of 15 STR markers”. [*Forensic Science International: Genetics* 2/2(2008): e1–e4], *Forensic Science International: Genetics* 2: 382. DOI:10.1016/j.fsigen.2008.06.001

Fridman C, dos Santos PCC, Kohler P, Garcia CF, Lopez LF, Massad E, Gattás GJF (2008) Brazilian population profile of 15 STR markers. *Forensic Science International: Genetics* 2: e1–e4. DOI:10.1016/j.fsigen.2007.09.003

Góes ACS, da Silva DA, Gil ÉHF, da Silva MTD, Pereira RW, de Carvalho EF (2004) Allele frequencies data and statistic parameters for 16 STR locos—D19S433, D2S1338, CSF1PO, D16S539, D7S820, D21S11, D18S51, D13S317, D5S818, FGA, Penta E, TH01, vWA, D8S1179, TPOX, D3S1358—in the Rio de Janeiro population, Brazil, *Forensic Science International* 140: 131–132. DOI:10.1016/j.forsciint.2003.11.023

Goor RM, Neall LF, Hoffman D, Sherry ST (2011) A Mathematical Approach to the Analysis of Multiplex DNA Profiles, *Bull Math Biol.*73(8): 1909-1931. DOI: 10.1007/s11538-010-9598-0

Nascimento E, Cerqueira E, Gusmão L (2011) Population database defined by 13 autosomal STR locos in a representative sample from Bahia, Northeast Brazil, *Forensic Science International: Genetics* 5: e38–e40. DOI: <http://dx.doi.org/10.1016/j.fsigen.2009.11.004>

Ocampos M, Fernandes RC, Latorre AFS, da Silva CMD, Korndorfer FP, Giamarusti AC, Menezes ME (2009) 15 STR locos frequencies in the population from Santa Catarina, Southern Brazil. *Forensic Science International: Genetics* 3: e129–e131. DOI:10.1016/j.fsigen.2008.10.010

Ota T (1993) DISPAN: Computer Program for Genetic Distance and Phylogenetic Analysis. Institute of Molecular Evolutionary Genetics, The Pennsylvania State University, Mueller Laboratory, University Park, USA,.

Peakall R and Smouse PE (2012) GenAEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update, *Bioinformatics* 28: 2537-2539.

Rodenbusch R, Schumacher S, Mardini AC, Chula FGL (2012) Allele frequencies of the five new generation forensic STR (D1S1656, D2S441, D10S1248, D12S391 and D22S1045) in the population from Rio Grande do Sul, Southern Brazil, *Forensic Science International: Genetics* 6: e55–e57. DOI: 10.1016/j.fsigen.2011.04.016

Sousa MLAPO, de Oliveira MAT, Auler-Bittencout EA, Soares-Vieira JA, Munoz DR, Iwamura ESM (2014) Population data of 16 autosomal STR locos of the Powerplex ESX 17 System in a Brazilian Population from the State of São Paulo, *Forensic Science International: Genetics* 11: e15–e17. DOI: <http://dx.doi.org/10.1016/j.fsigen.2014.03.018>

Tamura k, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular Biology and Evolution* 30: 2725-2729. DOI: 10.1093/molbev/mst197

Vieira TC, Silva DM, Gigonzac MAD, Ferreira VL, Gonçalves MW, da Cruz AD (2013) Allelic frequencies and statistical data obtained from 15 STR locos in a population of the Goiás State, *Genetics and Molecular Research* 12, 1: 23-27. DOI <http://dx.doi.org/10.4238/2013.January.16.5>

Wolfgramm EV, Silva BC, Aguiar VRC, Malta FSV, Castro AM, Souza Ferreira AC, Prezoti ANL, Paula F, Louro ID (2011) Genetic analysis of 15 autosomal and 12 Y-STR locos in the Espírito Santo State population, Brazil, *Forensic Science International: Genetics* 5: e41–e43. DOI: 10.1016/j.fsigen.2010.05.001

APÊNDICE C - REGISTRO DE PROGRAMA DE COMPUTADOR MutCodom
(Mutation Rate estimate for Codominant markers)

Quando: 14/04/2021.

Como: Via e-mail.

Com quem: Diretoria de Inovação e Empreendedorismo da Universidade Federal de Pernambuco – POSITIVA -UFPE, intermediária do pedido de registro no INPI (Instituto Nacional da Propriedade Industrial), conforme as instruções em <https://www.ufpe.br/positiva>.

Autores: Abigail Marcelino dos Santos Silva, Sérgio de Sá Leitão Paiva Junior e Valdir de Queiroz Balbino.

Função do programa: Calcular taxa de mutação.

Tipo de dados que analisa: genótipos de marcadores codominantes como os microssatélites autossômicos.



REPÚBLICA FEDERATIVA DO BRASIL

MINISTÉRIO DA ECONOMIA

INSTITUTO NACIONAL DA PROPRIEDADE INDUSTRIAL

DIRETORIA DE PATENTES, PROGRAMAS DE COMPUTADOR E TOPOGRAFIAS DE CIRCUITOS INTEGRADOS

Certificado de Registro de Programa de Computador

Processo Nº: **BR512021001365-4**

O Instituto Nacional da Propriedade Industrial expede o presente certificado de registro de programa de computador, válido por 50 anos a partir de 1º de janeiro subsequente à data de 30/10/2019, em conformidade com o §2º, art. 2º da Lei 9.609, de 19 de Fevereiro de 1998.

Título: MutCodom

Data de criação: 30/10/2019

Titular(es): UNIVERSIDADE FEDERAL DE PERNAMBUCO

Autor(es): ABIGAIL MARCELINO DOS SANTOS SILVA; SÉRGIO DE SÁ LEITÃO PAIVA JÚNIOR; VALDIR DE QUEIROZ BALBINO

Linguagem: PYTHON

Campo de aplicação: BL-02

Tipo de programa: AP-01; AT-01; TC-01

Algoritmo hash: SHA-512

Resumo digital hash:

5f004c4d4dd74beab4838cdd05ee5e1b05957e87aa006310f4f787d6eadd05b972ef9c24a11aa1ff2a7671fe2c2201a525d36e16d6613c616571f94d4469accb

Expedido em: 29/06/2021

Aprovado por:

Carlos Alexandre Fernandes Silva
Chefe da DIPTO

APÊNDICE D - PEDIDO DE REGISTRO DE PROGRAMA DE COMPUTADOR
SSRgene (Simulator of Simple Random Sampling of Genetic data)

Quando: 07/06/2021

Como: Via e-mail

Com quem: Diretoria de Inovação e Empreendedorismo da Universidade Federal de Pernambuco – POSITIVA -UFPE, intermediária do pedido de registro no INPI (Instituto Nacional da Propriedade Industrial), conforme as instruções em <https://www.ufpe.br/positiva>.

Autores: Abigail Marcelino dos Santos Silva, Sérgio de Sá Leitão Paiva Junior e Valdir de Queiroz Balbino

Função do programa: disponibiliza ferramenta de manipulação de amostras

Tipo de dados que analisa: dados genotípicos de marcadores codominantes diplóides.

Situação atual: em andamento.

