

**Universidade Federal de Pernambuco
Centro de Ciências Biológicas
Programa de Pós-Graduação em Genética**

Sérgio de Sá Leitão Paiva Júnior

DNAstr.com: Ferramenta estatística aplicada à genética forense

**Recife
2016**

Sérgio de Sá Leitão Paiva Júnior

DNAstr.com: Ferramenta estatística aplicada à genética forense

Tese apresentada ao Programa de Pós-Graduação em Genética da Universidade Federal de Pernambuco como parte dos requisitos exigidos para obtenção do título de Doutor em Genética.

Orientador: Prof. Dr. Valdir de Queiroz Balbino

Coorientador : Dra. Sandra Maria dos Santos

Recife

2016

Catálogo na fonte
Elaine Barroso
CRB 1728

Paiva Júnior, Sérgio de Sá Leitão
DNAstr.com: ferramenta estatística aplicada à genética forense/ Sérgio de Sá Leitão Paiva Júnior– Recife: O Autor, 2016.

91 folhas : il., fig., tab.

Orientador: Valdir de Queiroz Albino

Coorientador: Sandra Maria dos Santos

Tese (doutorado) – Universidade Federal de Pernambuco.

Centro de Biociências. Genética, 2016.

Inclui referência e anexos

1. Genética forense 2. Biologia computacional 3. DNA I. Albino, Valdir de Queiroz (orientador) II. Santos, Sandra Maria dos (coorientadora) III. Título

576.5

CDD (22.ed.)

UFPE/CCB-2016-156

DNAstr.com: Ferramenta estatística aplicada à genética forense

Aprovado em 15 / 03 / 2016

Banca Examinadora:

Dr. Valdir de Queiroz Albino
Universidade Federal de Pernambuco

Dra. Sandra Maria dos Santos
Gerência geral de polícia científica - Pernambuco

Dr. Borko Stosic
Universidade Federal Rural de Pernambuco

Dra. Manuela Barbosa Rodrigues de Souza
Universidade Católica de Pernambuco

Dr. Carlos Alberto Santiago Figueiredo Júnior
Universidade Federal de Pernambuco

Recife
2016

**À minha esposa Walquíria e meus filhos Davi e
Marcos, minha mãe, irmãs, sogra e sogro.**

Agradecimentos

Agradeço à Deus, causa eficiente de todos os efeitos. À minha esposa e filhos que mais pagaram o ônus da minha dedicação ao doutorado. Ao meu orientador e amigo Valdir Balbino que sempre nos entusiasma com novos desafios, instigando-nos nos limites da nossa capacidade intelectual. Aos colegas do LABBE que, em todas as atividades, consegue tornar tudo bem mais fácil, num ambiente sadio e de cooperação. Ao PPGG, funcionários, professores e coordenadores. Ao laboratório de perícia em genética da polícia civil nas pessoas da Dra. Sanda Santos e do colega Carlos Souza que apoiaram este trabalho com valiosas contribuições. À Universidade Federal Rural de Pernambuco pelo apoio na etapa final deste projeto.

"Mesmo que já tenha feito uma longa caminhada, sempre haverá mais um caminho a percorrer." (Santo Agostinho)

Resumo

A introdução do uso de STRs (Short Tandem Repeat) como uma ferramenta para identificação humana, revolucionou a área forense nos últimos 20 anos. Muitos laboratórios de biologia molecular passaram a oferecer o serviço de genotipagem com o objetivo de identificação individual tanto no âmbito forense. O serviço de genotipagem é finalizado na emissão de um laudo que, em muitos casos, será utilizado em tribunais e serve como prova em conflitos na justiça, por isto o laudo deve ser isento de erros e para garantir a minimização de erros nas conclusões dos laudos é necessário um grande controle por parte do laboratório, desde a coleta das evidências passando pela genotipagem até os cálculos estatísticos.

O objetivo deste trabalho foi criar uma ferramenta destinada ao uso de laboratórios de genotipagem com foco forense. O DNAstr.com é um software construído na plataforma web que foca a automatização das etapas de cadastro até a emissão do laudo, permite a integração com outros aplicativos e equipamentos do laboratório, faz análises estatísticas do banco de dados e aplica alguns dos princípios estatísticos dos softwares utilizados na literatura para a resolução de casos de paternidade e análise de mistura.

Palavras-chave: DNA; STRs; Estatística Forense ; Software

Abstract

The introduction of the use of STR (Short Tandem Repeat) as a tool for human identification has revolutionized forensic science in the last 20 years. Many molecular biology laboratories began offering the genotyping service with the individual identification goal in civil and forensic context. The genotyping service terminates producing a report, which in most cases, will be used in courts and will serve as evidence of justice disputes. Therefore, the report should not contain mistakes, and to ensure the minimization of errors in the conclusions of the reports is needed a lot of control by the laboratory from genotyping to the statistical calculations.

The objective was to create a tool, oriented to the use of genotyping laboratories. The DNAstr.com is software that facilitates the creation of reports, automating the steps to register and report emission, allows integration with other applications and laboratory equipment, makes analyzes database statistics and apply some of the principles of statistical software used in the literature to resolve paternity cases, and mixture analysis.

Key words: DNA; STRs; Forensic Genetic; Software

Lista de Ilustrações

Figura 1- Haplótipos baseados em SNPs (Cox et al. 2012).	19
Figura 2-Diagrama Probabilidade de Exclusão	37
Figura 3- Fluxograma Calculo PI - Trio Simples	42
Figura 4- diagrama para casos Duo	43
Figura 5- Thresholds Analítico e Estocástico. 1-Pico considerado como ruído. 2- Pico considerado como alelo. 3- Alelo considerado homozigoto.	46
Figura 6- LR para o caso geral de misturas	50
Figura 7 : Tela inserir frequências alélicas via arquivo	57
Figura 8: cadastro manual de tabela de frequências	57
Figura 9: Inserir genotipagem formato plink	57
Figura 10: Tela de visualização das tabelas de frequências	59
Figura 11-Match/Partially dos Alelos	61
Figura 12-Match/Partially	61
Figura 13- Heatmap dos dados Match/Partially	61
Figura 14- Lista Match/Partially por locus	62
Figura 15- Pipeline Paternidade	62
Figura 16- Modulo de cadastro de pessoas	63
Figura 17- Tela de cadastro de casos	63
Figura 18- Cadastro das Pessoas nos Casos	64
Figura 19- Telas de Inserção de Genotipagem	65
Figura 20- Cadastro de Genotipagem	66
Figura 21- Tela Laudo	67
Figura 22-Heredograma	68
Figura 23- Tela Editor Visual Paternidade	71

Figura 24- Pipeline Forense	71
Figura 25- Tela Editor Forense	72

Lista de Tabelas

Tabela 1 - Demonstração do princípio de Hardy-Weinberg (Daniel e Clark 2007)	24
Tabela 2- Tabela de Expressão LR	39
Tabela 3- Tabelas do banco de dados	55
Tabela 4- Índices tabelas	58
Tabela 5- Comparação de Índices	59
Tabela 6- Comparação entre Frequências	60
Tabela 7- Estudo de casos de paternidade	68
Tabela 8- Resultados testes paternidade	68
Tabela 9-Caso Mistura 2 alelos	72
Tabela 10 - Caso Mistura 3 alelos	72
Tabela 11 - Caso Mistura 4 alelos	72
Tabela 12- Comparação PHR e Mx por Locus	73
Tabela 13- Cálculos LR Misturas ($\Theta=0$)	76
Tabela 14- Cálculos LR Misturas ($\Theta=0,01$)	76
Tabela 15 - Cálculos LR Misturas ($\Theta=0,03$)	76

Lista de Abreviaturas, Siglas e Símbolos

Item	Definição
CODIS	<i>Combined DNA Index System</i> (Sistema Combinado de Índices de DNA)
DNA	Ácido desoxirribonucleico
HE	Heterozigosidade esperada
HO	Heterozigosidade observada
HWE	<i>Hardy-Weinberg Equilibrium</i> (Equilíbrio de Hardy-Weinberg)
ISFG	<i>Internacional Society for Forensics Genetics</i> (Sociedade Internacional de Genética Forense)
PCR	<i>Polymerase Chain Reaction</i> (Reação em cadeia da polimerase)
PD	Poder de Discriminação
PE	Poder de Exclusão
PIC	Conteúdo de Informação Polimórfica
PM	Probabilidade de Match
SNP	<i>Single Nucleotide Polymorphism</i> (Polimorfismo de um Único Nucleotídeo)
STR	<i>Short Tandem Repeats</i> (Pequenas Repetições em Tandem)

Sumário

1. Introdução	15
2. Revisão da Literatura	16
2.1 Marcadores Moleculares	16
2.2 Nomenclatura de STRs com uso forense	20
2.3 Genética de Populações	21
2.4 Princípio de Hardy-Weinberg	22
2.5 Testes para desvios nos valores do equilíbrio de Hardy-Weinberg	25
2.6 Aplicação de genética de populações em genética forense	27
2.7 Softwares de uso em genética de populações	32
2.8 Visualização dos dados.	34
2.9 Estatística para Genética Forense	35
2.10 Aplicação do cálculo de LR para comparação de perfis genéticos em casos de paternidade.	40
2.11 Análise de Misturas.	45
2.12 Correção populacional	50
3. Objetivos	52
3.1 Objetivo geral	52
3.2 Objetivos específicos	52
4. Material e Métodos	52
4.1 Levantamento dos requisitos	52
4.2 Implementação	53
4.3 Amostras de genotipagens	54
5. Resultados	55
5.1 Base de dados do programa DNAAstr.com.	55
5.2 Modulo de Tabelas (genética de populações)	56
5.3 Módulo de Paternidade	62
5.4 Módulo de Análise de Misturas	71
6. Discussão	77
7. Conclusões	81
Referências	82
Anexo 1 – Diagrama do Banco de Dados	85

Anexo 2 – Formato dos Arquivos	87
Anexo 3 – Curriculum Vitae	89

1. Introdução

Os métodos em genética molecular aplicados à genética forense, têm se firmado ao longo dos últimos 20 anos com a chegada de novos equipamentos capazes realizar a genotipagem em um número cada vez maior de *loci* do genoma, em menor tempo e com uma maior sensibilidade. Estes avanços permitiram aumentar a capacidade de discriminação entre indivíduos, porém introduziram uma maior complexidade, tanto pelos cálculos envolvidos como pelas quantidades de informação armazenadas e processadas.

Os modelos estatísticos utilizados nos cálculos aplicados à genética forense têm como premissa a independência entre as frequências calculadas a partir dos *loci* da amostra populacional, sendo que quando não são verificadas estas condições iniciais, os valores calculados podem diferir do valor verdadeiro, trazendo um fator de erro. Esta tarefa de verificação fica a cargo do próprio experimentador. Ferramentas utilizadas para realizar os cálculos estatísticos (e.g. Familias; DNAmix) largamente utilizadas em laboratórios de genotipagem principalmente no Brasil, não possuem métodos para checar se marcadores moleculares, sumarizados em tabelas de frequências, estão ou não aptos para o uso, apesar de assumir estas premissas.

Este trabalho descreve o desenvolvimento da ferramenta DNAstr.com, um software que implementa, via web, os cálculos de genética forense, organiza dados dos casos estudados e das pessoas genotipadas, disponibiliza índices relacionados à genética de populações nas tabelas de frequências, implementa rotinas de segurança com criptografia SHA2, e possui uma interface amigável. Por isto esta ferramenta possibilita uma maior segurança na montagem dos laudos técnicos por parte de peritos que atuam nos laboratórios de genética forense. Pode

ser utilizada como uma ferramenta gerenciamento de laudos. Permite maior facilidade na capacitação dos usuários, pela facilidade de manuseio. Possui mecanismos de auditoria que permitem acompanhar as modificações importantes realizadas no banco de dados. Além disto a sua implementação é gratuita podendo ser instalada num computador desktop ou num servidor de rede ficando acessível apenas no ambiente interno do laboratório.

2. Revisão da Literatura

2.1 Marcadores Moleculares

O genoma humano consiste de 22 pares de cromossomos autossômicos e um par de cromossomos determinantes do sexo. Os cromossomos sexuais são representados pelas letras X e Y, sendo que o homem possui o par XY e a mulher o par XX. Os pares de cromossomos são descritos como homólogos por terem a mesma estrutura gênica. A posição ou localização de um gene ou de um marcador é comumente chamada de *locus* (*loci* no plural). Uma cópia de cada gene reside no mesmo *locus* em cada um dos cromossomos do par homólogo. Eventos mutacionais, aqui definidos como sendo mudanças na sequência de bases ao longo do DNA, podem ocorrer ao longo do tempo modificando uma das cópias de um locus em cada um dos cromossomos homólogos. Estas modificações geram diferenças denominadas alelos. Se dois alelos de um Locus forem idênticos, então este Locus é considerado homozigoto, e se forem diferentes, heterozigotos (Cox, Doudna, and O'Donnel 2012) .

Um marcador genético é uma sequência de DNA com localização cromossômica conhecida e que pode ser usado para identificação individual ou de espécies. Alguns tipos de marcadores usados:

- RFLP (*Restriction fragment length polymorphism* ou polimorfismo de comprimento de fragmentos de restrição).
- SSLP (*Simple sequence length polymorphism* ou polimorfismo de comprimento de sequência única)
- AFLP (*Amplified fragment length polymorphism* ou polimorfismo de comprimento de fragmento amplificado)
- RAPD (*Random amplification of polymorphic DNA* ou Polimorfismo de DNA Amplificado ao Acaso)
- VNTR (*Variable number of tandem repeat* ou repetições pequenas em tandem)
- SSR Microsatellite polymorphism, (*Simple sequence repeat* ou Sequências simples repetidas)
- SNP (*Single nucleotide polymorphism* ou Polimorfismo de base única)
- STR (*Short tandem repeat* ou Repetições curtas em Tandem)

Um genótipo (do grego “*genos*”, originar e “*typos*”, característica) é a caracterização dos alelos presentes num locus gênico. A genotipagem é a técnica pelo qual identificamos pequenas regiões do DNA denominadas MARCADORES, que variam de indivíduo para indivíduo de um ou mais loci do genoma, determinado assim seus alelos. A genotipagem do DNA para identificação humana foi descrita pela primeira vez pelo geneticista inglês Alec Jeffreys (Jeffreys, Wilson, and Thein 1985). Seu método consistia em localizar polimorfismos usando enzimas para marcar regiões curtas de DNA e cortar o genoma em pedaços para, então, serem identificados através da técnica de eletroforese em gel. Polimorfismos são diferenças encontradas em regiões do genoma de cada indivíduo provocadas por

mutações em um locus do cromossomo, sendo, em média, 1 par de base (pb) em cada 1000. Duas formas de variação são possíveis: Polimorfismos de sequência e Polimorfismos de tamanho. A partir do trabalho realizado por Alec Jeffreys, geneticistas passaram a trabalhar na possibilidade de identificação humana individual através da análise dos polimorfismos encontrados no genoma.

Um locus microssatélite ou STR (STR do inglês *Short Tandem Repeat* ou *repetições curtas em tandem*) é uma sequência curta de DNA repetida muitas vezes em bloco num local específico do cromossomo (e.g. CACG_n, onde o tema de repetição “CACG” é visto “n” vezes no locus). Um SNP é um tipo de polimorfismo de sequência. Consiste numa mutação de uma única base. Estima-se que existam ao menos 10 milhões de SNPs no genoma humano, ou um a cada 300 pb. Já foram identificados pelo menos quatro milhões, sua densidade é de um sítio a cada 1000 ou 3000 em DNA codificante e de um sítio a cada 500 a 1000 em regiões não codificantes. Seu mapeamento é uma importante ferramenta nos estudos de associação com doenças pois estes polimorfismos muitas vezes estão em desequilíbrio de ligação com regiões do genoma que desempenham papéis importantes no desenvolvimento de doenças. Tanto STRs como os SNPs, podem ser usados para identificação humana. Para alcançar a mesma precisão que STRs na identificação humana, são necessárias uma maior quantidade de marcadores do tipo SNPs (Kidd et al. 2006). Nesse sentido, pesquisadores vêm utilizando bancos de dados de SNPs para selecionar os marcadores mais informativos dentro de um conjunto de milhares de variantes (Kidd et al. 2012).

Regiões repetitivas do DNA com tamanho variando entre 5-500pb são designados de DNA satélite, e são exemplos de polimorfismos de tamanho. Estima-se que 3% do genoma humano seja composto por diferentes loci STR's de

tamanho diferentes. A maioria dos marcadores STRs são dinucleotídeos, seguidos em abundância pelos mono-, tri e tetranucleotídeos (Lander et al. 2001). Os STRs também podem ser divididos com relação ao padrão de repetição, podendo ser dos tipos repetição simples, repetições compostas (que consistem em duas ou mais repetições simples adjacentes) e repetições complexas que podem ter diversos blocos de repetições muito variáveis.

Grupos de STRs, ou SNPs, que estão fisicamente próximos em um mesmo cromossomo (ditos em desequilíbrio de ligação) raras vezes são afetados por eventos de recombinação durante a meiose e são, em geral, herdados em conjunto; estes agrupamentos são conhecidos como haplótipos (Figura 1). Marcadores herdados em conjunto são compilados em haplótipos, que passam a ter a sequência de marcadores posicionados na mesma ordem para cada indivíduo. Um subconjunto de marcadores é escolhido para definir cada haplótipo individualmente (Cox et al. 2012).

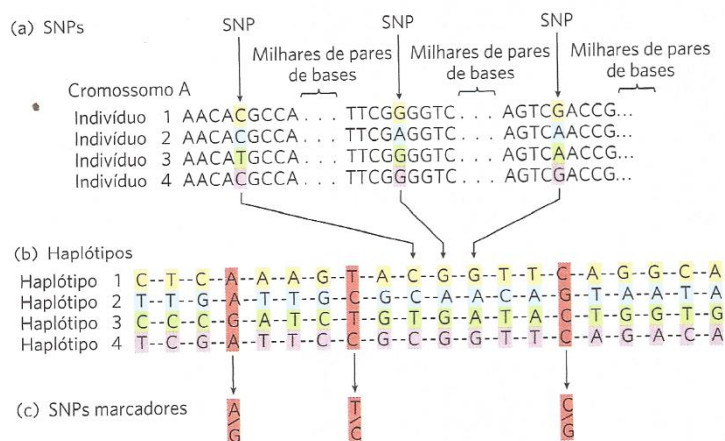


Figura 1- Haplotipos baseados em SNPs (Cox et al. 2012).

Para efeitos populacionais, o uso de polimorfismos é importante para a busca de padrões evolutivos relacionados a diferenças entre populações. As diferenças populacionais aumentam o poder estatístico dos marcadores (principalmente em alelos raros). Marcadores que apresentam frequências menores do que 0,95 para a maioria dos seus alelos são considerados polimórficos. Este corte em 0,95 serve para mostrar loci nos quais a variação alélica é mais comum. Um alelo é considerado raro quando sua frequência for menor que 0,005, muitos destes alelos são deletérios e são muitas vezes mantidos na população por mutações recorrentes. Com a definição de polimorfismos de 0,95 e se os alelos são combinados aleatoriamente em genótipos então estima-se que aproximadamente 10% da população será heterozigota para a maioria dos alelos comuns (Daniel and Clark 2007).

2.2 Nomenclatura de STRs com uso forense

A nomenclatura utilizada na genética forense para nomear os marcadores STRs, locus e alelos, foi recomendada pela Comissão Internacional de DNA da Sociedade Internacional de Genética Forense (ISFG do inglês *International Society to Forensics Genetics*). Um conjunto de três regras foram estabelecidas: A primeira regra estabelece que a leitura das sequências de DNA se faz da extremidade 5' para a 3' da fita de DNA (fita Watson); a segunda determina que quando um determinado STR está relacionado com genes codificantes de proteínas, essa será a cadeia utilizada como referência e o marcador terá a abreviatura da proteína que codifica (e.g. vWA, fator de von Willebrand), enquanto que em marcadores localizados fora da região codificante a nomenclatura será baseada na localização cromossômica da sequência descrita primariamente na literatura, iniciando com a

letra D (de DNA) seguido pelo número do cromossomo em que se localiza o STR, uma letra S (*single copy*, cópia simples) e um número que identifica a ordem no qual aquele marcador foi descrito no cromossomo (e.g. D12S324, que se situa no cromossomo 12 e no 324º locus descrito nesse cromossomo. A terceira regra diz respeito à classificação da sequência de repetição, em que o número de vezes que se repete uma determinada sequência é o número dado ao alelo em questão (e.g. na sequência: 5'- GG TCA TCA TCA TCA TCA GG-3' o nome do alelo será 5). No entanto, há unidades de repetição que estão incompletas, como é o caso do alelo 9.3 no marcador TH01 (alelo muito frequente na população portuguesa), o que significa que existem nove cópias da unidade de repetição tetranucleotídica e mais três nucleotídeos, estas microvariantes serão então classificadas com o número de repetições completas seguidas de um ponto decimal e do número de nucleotídeos existentes (Butler 2002)

2.3 Genética de Populações

O campo da genética de populações remonta ao início do século XX, sendo que um dos marcos iniciais foi a formulação do famoso Princípio de Hardy-Weinberg, em 1908. Seu início se deu com o estudo da proporção de genes através das características fenotípicas, pois inicialmente apenas estas características eram mensuráveis e suas diversas diferenças eram consideradas os alelos dos genes. A Genética de populações é um campo da genética que associa estatística e matemática para avaliar as variações hereditárias no tempo e no espaço através dos processos de mutação, migração e deriva genica avaliadas pela quantificação dos de diferentes alelos (Hartl and Clark 2007).

O modelo básico para genética de populações, chamado de modelo de gerações discretas, é dado por uma população ideal com as seguintes características: seu tamanho é infinito; a reprodução é aleatória entre os membros, ou seja, os cruzamentos entre os membros são aleatórios e possuem a mesma chance de ocorrer para qualquer membro da população (população panmítica); e as gerações são discretas, ou seja, não há sobreposição de gerações. Este modelo, apesar de simplista, pode ser utilizado como referência em muitas populações reais.

A partir deste modelo podemos imaginar “N” replicatas da população original, chamadas de populações locais ou demes, coexistindo ao mesmo tempo que são submetidas às mesmas forças evolucionárias, mas que diferem umas das outras pela aleatoriedade dos alelos transmitidos nas gerações passadas. A genética de populações tem focado seus esforços em compreender a variação existente entre os indivíduos das “N” subpopulações através da análise das frequências dos marcadores e seus alelos.

2.4 Princípio de Hardy-Weinberg

Em 1865, Gregor Mendel, um monge agostiniano e botânico, formulou as leis básicas da hereditariedade, hoje denominadas de Leis de Mendel. Seu trabalho de genética quantitativa foi fundamental para genética de populações por considerar as consequências da contínua autofecundação entre genótipos heterozigotos. Em 1908, Wilhelm Weinberg e G. H. Hardy publicaram, de forma independente, os mesmos resultados sobre a equação que envolve a frequência genotípica em uma população (Edwards 2008).

No modelo de Hardy-Weinberg, a relação matemática entre as frequências alélicas e as frequências genotípicas é dada por:

$$AA: p^2; Aa: 2pq; aa: q^2$$

Equação 1-Princípio de Hardy-Weinberg

Em que p^2 , $2pq$ e q^2 são as frequências dos genótipos **AA**, **Aa** e **aa** nos zigotos de qualquer geração, p e q são as frequências alélicas de **A** e **a** nos gametas da geração precedente e $p + q = 1$. As frequências mostradas na Equação 1 constituem a base do Princípio de Hardy-Weinberg ou equilíbrio de Hardy-Weinberg (EHW).

Um pré-requisito importante para que a Equação 1 seja válida diz respeito à população na qual as frequências são observadas, a saber: o organismo em estudo é diplóide; a reprodução é sexuada; as gerações são discretas; os genes em questão possuem dois alelos (esta equação pode ser expandida para mais de dois alelos); as frequências alélicas são idênticas em machos e fêmeas; o cruzamento entre os indivíduos é aleatório; o tamanho da população é infinito; mutações podem ser ignoradas e a seleção natural não afeta os alelos em consideração. Este conjunto de suposições resumem as premissas do modelo matemático de Hardy-Weinberg.

A Tabela 1 mostra que cruzamento aleatório dos genótipos equivale a união aleatória dos gametas no caso de dois alelos. As frequências genotípicas de AA, Aa e aa na geração parental são dadas por **P**, **Q** e **R**, onde $P + Q + R = 1$. Em termos de frequências genotípicas, as frequências alélicas p de **A** e q de **a** são as seguintes:

$$p = (2P + Q)/2 = P + \frac{Q}{2}$$

$$q = (2R + Q)/2 = R + \frac{Q}{2}$$

Equação 2- relação entre frequências alélicas e genotípicas.

Tabela 1 - Demonstração do princípio de Hardy-Weinberg (Daniel e Clark 2007)

Cruzamentos	Frequência de Cruzamentos	Frequências dos zigotos (descendentes)		
		AA	Aa	aa
AA x AA	P^2	1	0	0
AA x Aa	$2PQ$	$\frac{1}{2}$	$\frac{1}{2}$	0
AA x aa	$2PR$	0	1	0
Aa x Aa	Q^2	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$
Aa x aa	$2QR$	0	$\frac{1}{2}$	$\frac{1}{2}$
aa x aa	R^2	0	0	1
Totais (próximas gerações)		P'	Q'	R'

É possível deduzir então que as frequências genotípicas **P'**, **Q'** e **R'** da geração imediatamente seguinte podem ser matematicamente relacionadas às frequências alélicas da geração parental.

$$P' = P^2 + \frac{2PQ}{2} + \frac{Q^2}{4} = p^2$$

$$Q' = \frac{2PQ}{2} + 2PR + \frac{Q^2}{2} = 2pq$$

$$R' = \frac{Q^2}{4} + \frac{2QR}{2} + R^2 = q^2$$

Equação 3 - Relação frequência genotípica com frequência alélica da geração parental

Uma das mais importantes implicações do princípio de Hardy-Weinberg é que as frequências alélicas das próximas gerações são exatamente iguais às da geração anterior (Daniel and Clark 2007). Com o cruzamento aleatório as frequências se mantem constantes através das gerações. Estas frequências constantes implicam na ausência de forças evolutivas específicas, sendo que o mecanismo de herança mendeliana mantém a variação genética.

2.5 Testes para desvios nos valores do equilíbrio de Hardy-Weinberg

Para verificar se as frequências alélicas de uma população estão de acordo com a equação de Hardy-Weinberg, é necessário realizar algum teste estatístico, sendo que os mais comumente usados são o teste qui-quadrado e o teste exato de Fisher.

2.5.1 Teste qui-quadrado

O teste qui-quadrado (χ^2) é o teste mais comum para testar a independência entre linhas e colunas numa tabela de contingência (Conover 1980). A hipótese para o teste de qui-quadrado que define se uma população está em equilíbrio de Hardy-Weinberg (EHW) é dado por:

$$\begin{cases} H_0: A \text{ população está em equilíbrio de Hardy – Weinberg} \\ H_1: A \text{ população não está em equilíbrio de Hardy – Weinberg} \end{cases}$$

Para realizar este teste, devemos comparar as frequências dos genótipos na população contra as frequências esperadas quando a população está em EHW. Dado um locus denotado por **A1, A2,...,AK**, onde **K** é a quantidade de locus, definimos a quantidade de genótipos **AiAj** observados como: $n_{ij}, i \leq j, i, j = 1..k$ com **k = a** quantidade de alelos no locus. Então o número de alelos **ni** é dado por:

$$n_i = 2n_{ii} + \sum_{j=1}^{i-1} n_{ji} + \sum_{j=i+1}^k n_{ij},$$

Equação 4: Quantidade de alelos no locus

A probabilidade **Pij** do genótipo em equilíbrio de Hardy-Weinberg (EHW) é dado por:

$$p_{ij} \equiv P(A_i A_j) = \begin{cases} p_i^2, & i = j, \\ 2p_i p_j, & i \neq j, \end{cases}$$

Equação 5: Probabilidade do Genótipo $A_i A_j$ em EHW

Onde P_i é a probabilidade do Alelo A_i dada por: $p_i = n_i/(2n)$, $i=1..k$. Podemos estimar a frequência esperada E_{ij} para o locus em EHW como:

$$E_{ij} = \begin{cases} nP_i^2, & i = j \\ 2nP_i P_j, & i \neq j \end{cases}$$

Equação 6: Quantidade esperada de genótipos

O teste para a hipótese nula do equilíbrio de Hardy-Weinberg é construído como (Geisser and Johnson 1992) :

$$T = \sum_{i=1}^k \sum_{j=i}^k \frac{(n_{ij} - E_{ij})^2}{E_{ij}}.$$

Equação 7: teste qui-quadrado para EHW

2.5.2 Teste exato de Fisher

O teste exato de Fisher é um teste usado para analisar tabelas de contingência que calcula diretamente o valor exato do p-value, sem aproximação. Quando consideramos um locus dialélico, podemos calcular o valor exato do teste de Fisher pela fórmula:

$$P(n_{aa}n_{ab}n_{bb}|n_a n_b) = \frac{n! n_a! n_b! 2^{n_{ab}}}{(2n)! n_{aa}! n_{ab}! n_{bb}!}$$

Equação 8: Teste exato de Fisher para locus bialélico

Onde n_a e n_b são as quantidades dos alelos **a** e **b** e n_{aa} , n_{ab} e n_{bb} representa a quantidade observada dos genótipos **aa**, **ab** e **bb**. O teste exato de fisher também pode ser aplicado quando o locus possui mais de dois alelos. Neste caso, a avaliação de todas as possibilidades de alelos nos locus pode ser

computacionalmente muito custosa, especialmente quando o tamanho da amostra populacional é muito grande. Guo e Thonpson (1992) descreveram dois métodos aproximados para o cálculo do p-valor para o teste exato de fisher (Guo and a. 1992) baseados nas técnicas de reamostragem de Monte Carlo e no algoritmo Metropolis.

2.6 Aplicação de genética de populações em genética forense

A possibilidade da definição genotípica de indivíduos não relacionados para diversos loci presentes no genoma humano possibilitou a caracterização individual para os membros de uma população. Um conjunto de 12 STR loci pode dar um poder de discriminação entre os indivíduos da população de um em um trilhão (Fung and HU 2008). Estatísticas utilizadas para definir a identidade, ou *match* de perfis de um ou mais indivíduos, dadas as amostras, são baseadas na relação entre o genótipo dos indivíduos conhecidos e possivelmente envolvidos com a amostra e o genótipo de qualquer indivíduo que poderia estar aleatoriamente presente na população. Um conjunto de estatísticas relacionadas à genética de populações auxiliam na caracterização de uma população utilizada nos cálculos probabilísticos em genética forense.

2.6.1 Probabilidade de Match (p.M)

A probabilidade de match é a probabilidade de encontrar duas pessoas geneticamente não relacionadas e com o mesmo genótipo (Barrai and Larralde 1992). A probabilidade de Match pode ser considerada de duas formas: *A priori* é calculada quando conhecemos a frequência dos alelos e desejamos saber a probabilidade de dois genótipos iguais com estas frequências. *A posteriori*, quando

já temos a informação de um genótipo em particular e desejamos saber qual a probabilidade de encontrar outro genótipo igual. Em tabelas de frequências desejamos calcular o valor da probabilidade de Match a priori para cada locus em particular e para a tabela em geral.

Dados: **n** a quantidade de indivíduos genotipados; **i** e **j**, $i \leq j = 0, \dots, n$ as frequências de todos os alelos e **P_{ij}** a frequência de todos os genótipos o valor do Probabilidade de Match, **p.M**, a priori é dada por:

$$p.M = \sum_{i=0}^n \sum_{j=0}^n P_{ij}^2$$

Equação 9: Probabilidade de Match (p.M)

A probabilidade de Match acumulada informa a probabilidade de match quando consideramos todos os loci. Esta probabilidade é dada pelo produtório de p.M de todos os loci.

2.6.2 Poder de discriminação (PD)

O poder de discriminação (PD) (Jones 1972) é definido como a probabilidade de diferenciação entre dois indivíduos não relacionados. Esta medida está relacionada a probabilidade de match discutida anteriormente e é dada por:

$$PD = 1 - p.M$$

Equação 10: Poder de discriminação

Pela relação acima, percebe-se que quanto menor for a probabilidade de match, maior será o poder de discriminação que o locus ou a tabela possui.

2.6.3 Conteúdo Informativo de Polimorfismo (PIC)

O conceito de polimorfismo é usado para definir a variação genética numa população e tem sido aplicado em diversas áreas de estudo como ecologia, zoologia e microbiologia (Nagy et al. 2012). O conteúdo informativo de polimorfismo pode ser interpretado como a probabilidade que os alelos paternos e maternos podem ser deduzidos pelos alelos do filho, ou em outras palavras, a probabilidade dos alelos paternos terem sido transmitidos aos descendentes (Botstein et al. 1980). Dados: **n** o número de alelos; **pi** e **pj** as frequências dos alelos **i** e **j**, o valor de PIC é dado por:

$$PIC = 1 - \sum_{i=0}^n p_i^2 - \sum_{i=1}^{n-1} \sum_{j=i+1}^n 2p_i^2 p_j^2$$

Equação 11: PIC Conteúdo informativo de Polimorfismo

O valor PIC calculado por locus informa o quanto é informativo o locus para o polimorfismo. Valores de $PIC \geq 0,5$ indicam alto valor informativo; $0,5 > PIC \geq 0,25$ possui um conteúdo informativo médio e valores $PIC < 0.25$ possuem fraco conteúdo informativo.

2.6.4 Heterozigosidades Observada (**Ho**) e Esperada (**He**)

A Heterozigosidade Observada é definida como a razão entre a quantidade de indivíduos heterozigotos num locus pelo total de indivíduos. A Heterozigosidade Esperada é definida pela equação:

$$He = 1 - \sum_{i=0}^n P_i^2$$

Equação 12: Heterozigosidade Esperada HE

Onde **Pi**, **i = 0,...,n** são as frequências dos alelos no locus.

2.6.5 Poder de Exclusão (PE)

O poder de exclusão (PE) é definido como sendo a fração de indivíduos que têm um perfil de DNA que é diferente da de um indivíduo selecionado aleatoriamente, num caso típico de paternidade. O valor para cada caso individual pode variar. A média para um determinado locus é representado pela seguinte equação (Brenner and Morris 1992):

$$PE = H_o^2(1 - 2H_oH^2)$$

Equação 13: Poder de Exclusão (PE)

Onde **H_o** é a heterozigosidade observada e **H** é a frequência dos homozigotos. Podemos também calcular o PE para o conjunto de loci, CPE (poder de exclusão acumulado) dado por:

$$CPE = 1 - \prod_i^n (1 - PE_i)$$

Equação 14: Poder de exclusão acumulada (CPE)

2.6.6 Índice de paternidade típico (TPI)

O índice de paternidade típico reflete quantas vezes mais provável é que uma pessoa que está sendo testado seja o pai biológico, em vez de um indivíduo selecionado aleatoriamente. O índice de paternidade típico é atribuído a um locus, em vez de um caso individual. Geralmente, um TPI inferior a 1 é indicativo de não relacionamento. O TPI é representado pela seguinte equação:

$$TPI = \frac{1}{2H}$$

Equação 15: Índice de paternidade típico (TPI)

2.6.7 Índice de fixação de Wright (F)

De acordo com Falconer (FALCONER 1964), a endogamia ocorre quando se tem o acasalamento entre indivíduos que são relacionados por ascendência, tendo como primeiro efeito uma mudança nas frequências genotípicas de Hardy-Weinberg. Para o autor, o coeficiente de endogamia F mede a probabilidade de dois alelos, em qualquer locus de um indivíduo, serem originados a partir da cópia de apenas um alelo na geração anterior (autozigosidade). Sem considerar as forças evolucionárias que podem estar presentes numa população, o índice de fixação, ou estatística F , é desenvolvido em função das frequências alélicas e genotípicas (Nagylaki 1998), uma vez que os conceitos de autozigosidade (alelos iguais por descendência) e alozigosidade (diferentes por descendência) estão relacionados aos conceitos de homozigosidade e heterozigosidade em uma população de organismos endocruzados.

Wright (Wright 1965) discutiu o relacionamento entre gametas, mostrando o significado do coeficiente de endogamia e de parâmetros relacionados a pares de gametas em geral, realçando o fato de que eles podem ser interpretados tanto como coeficientes de correlação, quanto como a probabilidade dos genes idênticos serem de mesma origem. Wright sempre preferiu a interpretação de F como sendo uma correlação uma vez que F pode assumir valores negativos e neste caso o conceito de probabilidade falha.

Três parâmetros descritivos do índice de fixação foram propostos: Em termos de população total (T). Em termos de subpopulações (S) e em termos de indivíduos (I).

- F_{IT} - expressa a correlação entre os gametas que se unem para produzir os indivíduos em relação aos gametas da população total.

- F_{IS} - expressa a média das correlações, sendo cada uma delas proveniente dos gametas que se unem em cada subpopulação em relação aos gametas desta subpopulação.
- F_{ST} - expressa a correlação entre os gametas ao acaso dentro da subpopulação em relação aos gametas da população total.

O(WEIR 1996)(WEIR 1996)(WEIR 1996)(WEIR 1996) valor de F é definido em função da heterozigosidade através da seguinte expressão (WEIR 1996):

$$F = 1 - \frac{H_o}{H_e}$$

Equação 16: Índice de Fixação de Wright

Quando $F = 0$, assume-se que os acasalamentos são aleatórios, ou seja, não existe endogamia e as frequências genotípicas da população estarão de acordo com o equilíbrio de Hardy-Weinberg. Quando F assume qualquer outro valor entre 0 e 1, as frequências se alteram, sendo que para $F = 1$, tem-se endogamia total e a população só apresentará genótipos homozigóticos **AA** e **aa**, respectivamente nas frequências **p** e **q**.

2.7 Softwares de uso em genética de populações

Diversos métodos computacionais e softwares aplicados à genética de populações tem sido proposto nos últimos anos. No entanto, a maioria deles não possuem um claro direcionamento às aplicações forenses e muitas vezes é necessário o uso de diversos softwares para sintetizar todos os índices de interesse para a análise de tabelas de frequência.

2.7.1 Genepop

Este software foi originalmente escrito em 1995 e rescrito em 2008 (Rousset 2008). Entre os recursos disponíveis no software com aplicações em forense,

temos: Cálculo exato de fisher para o Equilíbrio de Hardy-Weinberg, Estatísticas F e cálculo de frequências alélicas. Este software possui versão Windows (DOS), Linux e uma interface WEB que roda em modo batch (em segundo plano) com os resultados sendo enviados pelo email. O formato de importação de dados é específico para o programa (formato NEXUS).

2.7.2 Alerquim

O Alerquim é um software de genética de populações bastante completo (Excoffier and Lischer 2010). Os algoritmos são implementados para análises intra- e inter-populacionais. Possui as seguintes funcionalidades utilizadas em genética forense: Estatística F; teste exato do HWE, cálculo de frequências alélicas. Possui interface para Windows, Linux e MAC. Aceita vários formatos para importação de dados e exporta em vários formatos sendo que na última versão foi implementado uma interface XML para exportação de dados.

2.7.3 EasyDNA_PopuData

Software desenvolvido por (Fung and Hu ,2008) para implementação dos índices de genética de populações descritos no seu livro. Este software apresenta uma interface para Windows e inclui vários cálculos de índices de genética de população direcionados à forense, entre eles: Heterozigosidade Observada e Esperada com o erro padrão, Teste Qui-Quadrado, teste exato de fisher, Poder de discriminação e Poder de exclusão. O arquivo de importação é próprio e baseado em arquivo texto.

2.7.4 Plink

É uma ferramenta feita para estudos com dados de sequenciamento de genoma, especialmente na análise de SNPs e associação. Este software possui recursos para trabalhar com grande quantidade de marcadores genotipados de milhares de indivíduos (Purcell et al. 2007). Possui, entre outras, as seguintes funcionalidades: Teste Qui-Quadrado e teste exato de Fisher para o EHW, frequências alélicas e genotípicas. Sua interface é por linha de comando, trabalha com o formato PED e MAP que são arquivos textos que sumariza os alelos nos loci.

2.8 Visualização dos dados.

O perfil de DNA de dois indivíduos pode ter zero, um ou dois pares de alelos que são coincidentes em um determinado locus. Estes eventos podem ser chamados de *mismatches*, *partial matches* e *matches*, respectivamente, e tem probabilidades que dependem da proporção de alelos e de parâmetros populacionais como o F_{ST} que indicam a probabilidade de que um alelo compartilhado entre dois indivíduos seja idêntico por descendência. Visualizar estes dados pode ajudar a entender a distribuição dos alelos na população.

2.8.1 Match/Partial

A probabilidade de que uma pessoa, randomicamente escolhida na população, tenha um perfil genotípico específico diminui quando o número de loci genotipados cresce. Um elevado grau de partilha de alelos entre os pares de perfis sugere parentesco entre os indivíduos (Weir 2004). A tabela de match/partial mostra a quantidade de alelos compartilhados em todos os loci do banco de dados e entre todos os indivíduos do banco de dados comparados dois a dois. As linhas da tabela representam a quantidade de loci com dois alelos idênticos e as colunas representam a quantidade de loci com um alelo idêntico. Para a montagem da tabela cada indivíduo é comparado com todos os demais locus a locus e é contado

quantos loci compartilham dois alelos e quantos compartilham apenas um. O total é somado com os demais e colocado na célula correspondente da tabela.

2.9 Estatística para Genética Forense

A genética forense é o ramo da ciência forense dedicado à identificação biológica de indivíduos tanto no contexto do direito civil (determinação de parentesco) como no contexto do direito criminal (definição de participantes numa cena de crime).

Na ausência de erros, uma coincidência de perfis (*Match*) de DNA entre duas pessoas pode ser explicada de duas formas possíveis: ou o DNA pertence a mesma pessoa ou o indivíduo não é a fonte mas possui o mesmo perfil devido ao acaso. A estatística forense é a principal ferramenta utilizada para dar suporte a uma determinada hipótese de coincidência entre os perfis estudados e tem como premissas os conceitos de genética de populações nos cálculos dos resultados (Buckleton, Triggs Christopher & Walsh, 2005). Segundo estes autores, duas abordagens estatísticas podem ser empregadas nas análises da evidência: a abordagem Frequencista e a abordagem Bayesiana, sendo que a abordagem Bayesiana pode ser subdividida em Bayesiana lógica e Bayesiana completa, que leva em consideração mais que duas hipóteses a serem testadas.

Na abordagem frequencista é testada a frequência da hipótese da acusação ou hipótese nula (H_0) ocorrer ou não ocorrer. Sendo assim existem duas subdivisões nesta abordagem, a probabilidade de coincidência e a probabilidade de exclusão.

2.9.1 Probabilidade de Coincidência (match).

Na probabilidade de coincidência é calculada a probabilidade (Pr) da evidência (E), neste caso a evidência é o genótipo da amostra investigada ser originalmente pertencente a qualquer outra pessoa não relacionada ao suspeito $Pr(E|H_0)$. Se esta probabilidade é pequena (menor que 5%) dizemos então que H_0 não é suportada e a hipótese alternativa H_1 de que a evidência (E) vem do suspeito é aceita. Para calcular $Pr(E|H_0)$ devemos calcular a probabilidade do genótipo ter vindo de qualquer pessoa da população, que em geral é muito baixa, o que pode introduzir um viés na interpretação do resultado.

2.9.2 Probabilidade de Exclusão

A probabilidade de exclusão (EP) é a probabilidade de uma pessoa qualquer (randômica) ser excluída como contribuinte da amostra (no caso de misturas) ou de ser um parente relacionado no caso de testes de paternidade. A EP pode ser calculada sem a presença do genótipo do suspeito ou do pai em casos de paternidade pois seu valor indica a proporção de pessoas aleatórias na população que podem ser excluídas do relacionamento com o caso (Fung and Hu 2002). A EP é dada por:

$$EP_l = (1 - p_k)^2$$

Equação 17- Probabilidade de Exclusão

Onde l é o locus, p_k é a frequência populacional do(s) alelo(s) que não está presente nos perfis conhecidos envolvidos no estudo do caso (mistura ou

paternidade). É possível usar a regra do produto para calcular o EP acumulado nos loci estudados uma vez que os loci guardam independência entre si. Este valor é dado por:

$$EP_l = 1 - \prod_{l=1}^K (1 - p_k)^2$$

Equação 18- EP acumulado

O diagrama abaixo, baseado em (Fung and Hu ,2008), mostra o valor de EP para todas as possíveis variações nos genótipos do suposto pai, filho e mãe, dado P_i, P_j as frequências dos alelos i e j na população.

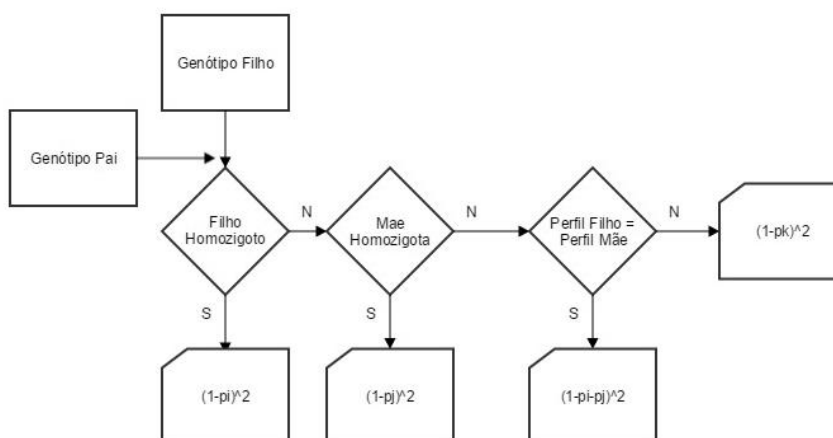


Figura 2-Diagrama Probabilidade de Exclusão

2.9.3 – Abordagem Bayesiana (Razão de Verossimilhança).

Na teoria das probabilidades, o teorema de Bayes mostra a relação entre probabilidades condicionais e incondicionais. Este teorema relaciona as probabilidades *a priori* tendo em conta novas evidências de forma a obter probabilidades *a posteriori*. O uso do teorema de Bayes passou a ser utilizado na

estatística forense pois a abordagem frequentista não suportava a possibilidade de acrescentar novas evidências, sendo já utilizada desde da década de mil novecentos e trinta em casos de paternidade.

Dados:

H_p : Hipótese de acusação

H_d : Hipótese de defesa

E : Representa a evidência

I : Evidência para o caso (genótipos)

O teorema de Bayes é dado por:

$$\Pr(H_p|E, I) = \Pr(E|H_p, I)\Pr(H_p|I)$$

$$\Pr(H_d|E, I) = \Pr(E|H_d, I) \Pr(H_d|I)$$

Equação 19- Teorema de Bayes

Através da razão de verossimilhança (*likelihood ratio* - LR) podemos relacionar as hipóteses de defesa e acusação antes da evidência (*prior odds*) dada pelo segundo termo da multiplicação, com as mesmas hipóteses depois da evidência (*posterior odds*). Desta forma podemos rescrever esta equação em função de LR :

$$Posterior Odds = Likelihood ratio \times Prior Odds$$

A valoração do termo *Prior Odds* não está relacionada a evidência, ou seja, em termos práticos para cálculos, este valor deverá ser retirado pois a genética forense utiliza-se apenas da evidência (que pode ser a mistura coletada pelo swab vaginal ou o genótipo de um filho num caso de paternidade).

A razão de verossimilhança (LR) avalia a força da evidência, que neste caso é a genotipagem da amostra biológica colhida para comparação, desta forma podemos rescrever a equação acima como:

$$LR = \frac{P(E|H_p)}{P(E|H_d)}$$

Equação 20: Razão de verossimilhança

Valores de LR igual a um indicam que a evidência não colabora nem com a hipótese de defesa nem com a da acusação. Valores de LR menores que um indicam que a evidência favorece a hipótese de defesa e para valores de LR maiores que 1 indicam que a evidência favorece a hipótese de acusação. A tabela 2 indica a valoração verbal para LR segundo os parâmetros da FSS (Forensic Science Service – UK). .

Tabela 2- Tabela de Expressão LR

LR	Expressão Oral	
1.000.000+	Extremamente Forte	 Suporta Hp
100.000	Muito Forte	
10.000	Forte	
1000	Moderadamente Forte	
100	Moderado	
10	Limitado	
1	Inconclusivo	

0,1	Limitado	
0,01	Moderado	
0,001	Moderadamente Forte	
0,0001	Forte	
0,00001	Muito Forte	
0,000001	Extremamente Forte	



Suporta H_d

2.10 Aplicação do cálculo de LR para comparação de perfis genéticos em casos de paternidade.

A razão de verossimilhança, LR (Likelihood Ratio), é utilizada para demonstrar a força da evidência em casos de paternidade. Nesta aplicação a LR é mais conhecida com Índice de Paternidade (PI). Sua estrutura permanece a mesma descrita no item anterior, no entanto as hipóteses são adaptadas para o contexto da paternidade. As hipóteses podem ser consideradas como:

H_p : O Suposto Pai (SP) é o verdadeiro Pai.

H_d : O Suposto Pai não é o verdadeiro Pai.

Colocando as hipóteses acima em função da razão de verossimilhança (LR), temos a seguinte equação:

$$LR = PI = \frac{\Pr(E|H_p)}{\Pr(E|H_d)}$$

Equação 21: Índice de Paternidade (PI)

2.10.1 Caso de Paternidade: Trio

Para casos onde estão envolvidos os genótipos da mãe, suposto pai e do filho, onde a paternidade está sendo questionada e não há dúvidas quanto a maternidade, as hipóteses são:

H_p : O Suposto Pai (SP) é o verdadeiro Pai (A mãe é verdadeira mãe).
 H_d : O suposto Pai não é o verdadeiro Pai (A mãe é a verdadeira mãe).

Obviamente que a maternidade pode ser questionada, neste caso muda-se a hipótese. Podemos agora relacionar as hipóteses dadas acima com o valor de PI através da equação:

$$PI = \frac{\Pr(G_C, G_M, G_{SP} | H_p)}{\Pr(G_C, G_M, G_{SP} | H_d)} = \frac{\Pr(G_C | G_M, G_{SP}, H_p)}{\Pr(G_C | G_M, G_{SP}, H_d)} \times \frac{\Pr(G_M, G_{SP} | H_p)}{\Pr(G_M, G_{SP} | H_d)}$$

Onde G_C = genótipo da criança, G_M = genótipo da mãe e G_{SP} = genótipo do suposto pai. Como os genótipos dos parentes (Mãe e Suposto Pai) não dependem da hipótese, podemos anular o segundo termo da multiplicação, ficando com:

$$PI = \frac{\Pr(G_C, G_M, G_{SP} | H_p)}{\Pr(G_C, G_M, G_{SP} | H_d)} = \frac{\Pr(G_C | G_M, G_{SP}, H_p)}{\Pr(G_C | G_M, G_{SP}, H_d)}$$

Equação 22- PI para caso simples de paternidade - Trio

Se não há incoerência nos loci com relação a maternidade e todos os alelos do suposto pai estão presentes então dizemos que $\Pr(G_C | G_M, G_{SP}, H_p) = 1$. O diagrama abaixo, baseado em (Fung and HU 2008), mostra o valor de PI para todas as possíveis variações nos genótipos do Suposto Pai, Filho e Mãe, dado P_i, P_j as frequências dos alelos i e j na população.

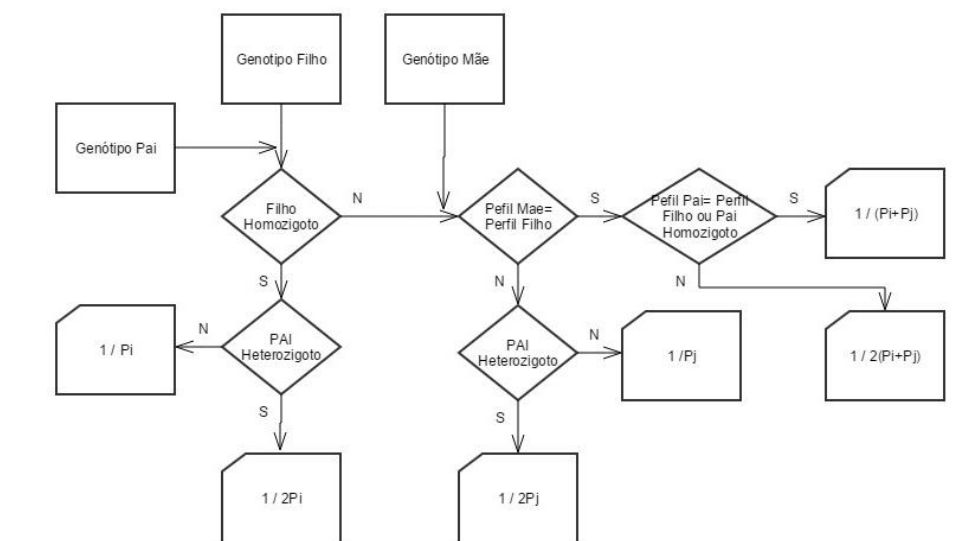


Figura 3- Fluxograma Calculo PI - Trio Simples

2.10.2 Caso de Paternidade: Duo

Nos casos onde não é possível determinar o genótipo da mãe, as hipóteses envolvidas mudam e deve considerar os alelos (genótipo) da mãe como de uma pessoa desconhecida.

H_p : O Suposto Pai (SP) é o verdadeiro Pai (O genótipo da mãe é desconhecido).

H_d : O suposto Pai não é o verdadeiro Pai (O genótipo da mãe é desconhecido).

O índice de paternidade (PI), será reescrito assim:

$$PI = \frac{\Pr(G_C, G_{SP} | H_p)}{\Pr(G_C, G_{SP} | H_d)} = \frac{\Pr(G_C | G_{SP}, H_p)}{\Pr(G_C | H_d)}$$

Equação 23-PI para o caso de Duo

Se não há incoerência nos loci e todos os alelos do suposto pai estão presentes então dizemos que $\Pr(G_c | G_{sp}, H_p) = 1$. O diagrama abaixo, baseado em (Fung and HU 2008), mostra o valor de PI para todas as possíveis variações nos genótipos do Suposto Pai e do Filho, dado P_i, P_j as frequências dos alelos i e j na população.

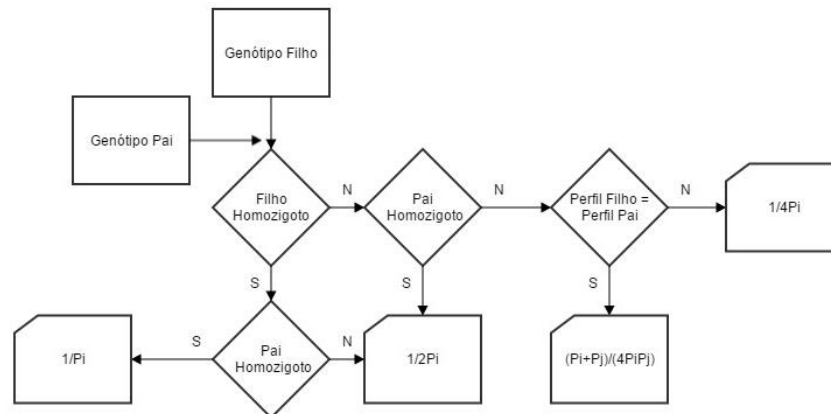


Figura 4- diagrama para casos Duo

2.10.3. Outros casos envolvendo parentesco.

Os possíveis cenários montados para caso de paternidade podem incluir as seguintes possibilidades: 1) O genótipo do suposto pai é desconhecido (e.g. por morte ou desaparecimento do genitor) mas o de um parente é conhecido; 2) A hipótese de acusação é que um parente do suposto pai é que é o verdadeiro pai. Podemos então definir as seguintes hipóteses:

H_p : Um parente próximo do Suposto Pai é o verdadeiro Pai .

H_d : O Verdadeiro Pai é desconhecido.

Ou

H_p : O Suposto Pai é o verdadeiro Pai .

H_d : O Verdadeiro Pai é um parente do Suposto Pai.

Este tipo de comparação de genótipos exige a definição do “alelo idêntico por descendência” que é o alelo comum que relaciona o genótipo do suposto pai com o do parente de acordo com o seu grau de parentesco. Podemos considerar as seguintes probabilidades de um dado alelo num genótipo do suposto pai ser idêntico por descendência com o de um parente (Balding and Nichols 1994).

$$K_0 = P(\text{Nenhum alelo é idêntico por descendência})$$

$$2k_1 = P(\text{Apenas um alelo é idêntico por descendência})$$

$$K_2 = P(\text{Ambos os alelos são idênticos por descendência})$$

Com as probabilidades acima, é possível definir a probabilidade condicional de duas pessoas relacionadas terem alelos idênticos por descendência.

$$P(A_i A_j | A_i A_j) = 2p_i p_j k_0 + \left[p_i \frac{1}{2} + p_j \frac{1}{2} \right] 2k_1 + k_2$$

Equação 24-Probabilidade condicional dos alelos por descendência

Podemos também definir o valor do coeficiente de parentesco ou consanguinidade, “F”, que é a probabilidade de dado dois alelos retirados aleatoriamente do suposto pai e do verdadeiro pai (no caso de paternidade) sejam idênticos por descendência. Sendo assim F é igual a K_1 . Podemos calcular os valores de “F” para os diversos graus de parentesco.

A razão de verossimilhança para casos que envolve familiares é definida como índice Avuncular (AI) e este índice se relaciona com o índice de paternidade (PI) através da formula:

$$AI = (1 - 2F) + 2FxPI$$

Equação 25- Índice Avuncular

2.11 Análise de Misturas.

Amostras biológicas como, fluidos corporais, encontradas em cenas de crime pode conter a presença de mais de dois perfis de DNA das pessoas possivelmente envolvidas no crime. Para identificação de perfis DNA numa mistura devemos aplicar um modelo matemático capaz de separar os alelos dos contribuintes na mistura para, assim, poder avalia-los. Estes modelos podem ser classificados em três categorias (Gill et al. 2015): 1) Modelos Binários, 2) Modelos Semi-Contínuos e 3) Modelos Contínuos. Esta classificação reflete como a informação sobre o tamanho ou área do pico no eletroferograma é usado nos cálculos da razão de verossimilhança. Para a avaliação estatística destes alelos na mistura, as metodologias podem ser agrupadas em métodos Bayesiano e em métodos frequentistas. Os métodos bayesianos diferem do frequentista pela possibilidade de utilizar os aspectos quantitativos e qualitativos dos picos/áreas dos alelos nos eletroferograma (resultado da genotipagem da mistura).

Para análise de Misturas utilizando o modelo binário, devem ser seguidos seis passos básicos (Clayton et al. 1998):

- Passo 1) Identificar a presença de misturas: A presença de misturas fica evidente quando mais de dois locus possuem no mínimo, três picos definidos como alelos.
- Passo 2) Designar os alelos da mistura: A definição de picos como alelos deve ser feita com critérios objetivos para evitar possíveis erros na genotipagem.
- Passo 3) Identificar o número de contribuintes na mistura.
- Passo 4) Estimar a proporção dos contribuintes da mistura.
- Passo 5) Fazer considerações sobre os genótipos presentes na mistura.

- Passo 6) Comparar os genótipos com os de referência.

Métodos para definição destes passos levam em consideração a informação sobre o tamanho do pico ou a área do pico no eletroferograma. Um pico no eletroferograma poderá ser considerado como um alelo quando o tamanho do seu pico for maior que o valor determinado como um limite chamado de *threshold* Analítico. A definição de um valor para o *threshold* analítico deve ser feita por meio de amostragens estatísticas realizadas pelo laboratório sob diferentes condições de concentração de amostras ou podem ser seguidas as recomendações definidas pelos fabricantes dos equipamentos. Para definir a condição de homozigidade de um alelo, ou seja, se o locus é homozigoto, um *Threshold* Estocástico deverá ser definido. O valor do *threshold* estocástico deverá ser definido por amostragem estatística realizada pelo laboratório, este valor não é indicado pelos fabricantes de equipamentos. Uma vez que um pico tem tamanho maior que o valor deste *threshold* então podemos afirmar que ele é um alelo homozigoto (Figura 5).

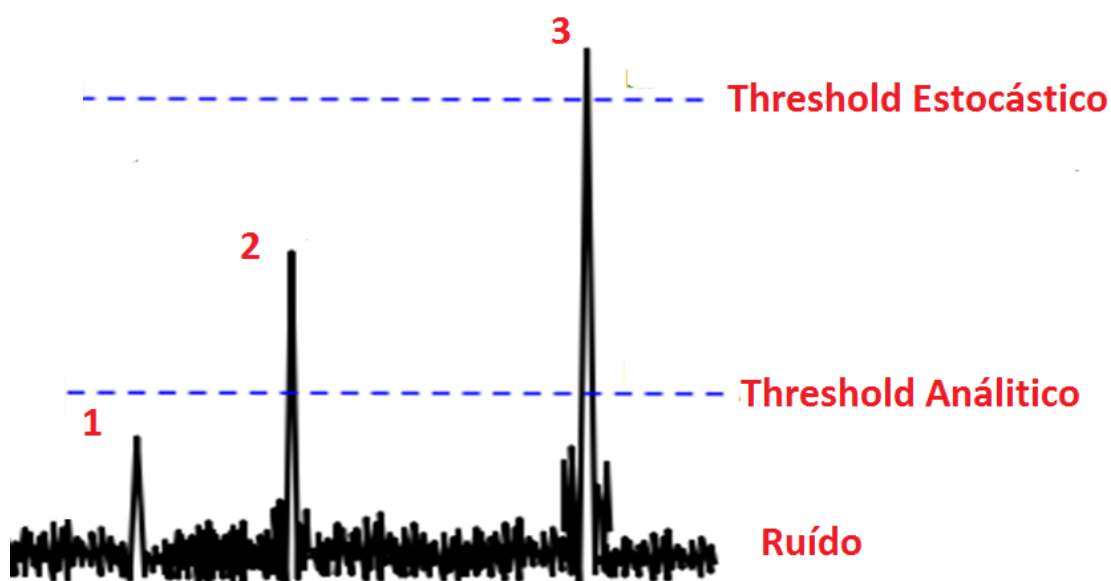


Figura 5- Thresholds Analítico e Estocástico. 1-Pico considerado como ruído. 2- Pico considerado como alelo. 3- Alelo considerado homozigoto.

Apesar dos limites definidos pelo Threshold analítico e estocástico, alguns picos devem ser descartados quando forem picos resultados de erros conhecidos durante o processo de PCR. A Taq Polimerase, enzima responsável pela amplificação das regiões de STR dos loci, pode, eventualmente, perder ou acrescentar blocos de repetição nos loci STR. Quando isto ocorre, é visto no eletroferograma um pico posicionado quatro blocos antes ou depois do pico verdadeiro, este pico é chamado de *Stutter*. Para um pico ser considerado um *Stutter* ele deverá ter até 15% do tamanho do pico originador e estar na posição correta. Outro erro causado pela Taq Polimerase é a inclusão ou retirada de adeninas na cauda poli-A (3'-end). Quando isto ocorre é possível verificar pequenas ondulações no pico.

Para a caracterização dos perfis na amostra, deve-se separar os contribuintes utilizando a proporção entre as alturas dos picos, em seguida deve-se determinar os alelos homozigotos e heterozigotos. Os picos com proporções iguais devem ser considerados como originários da mesma pessoa (Gill et al. 2006).

A proporção dos contribuintes na amostra (M_x) pode ser estimada pelo tamanho dos picos (ou áreas). Dados quatro alelos **A,B,C,D** com tamanho de picos dados por ϕA , ϕB , ϕC e ϕD . Os indivíduos X e Y possuem os genótipos **AB** e **CD** respectivamente. Se consideramos o indivíduo X como principal contribuinte na amostra, então podemos calcular M_x pela formula:

$$M_x = \frac{\phi A + \phi B}{\phi A + \phi B + \phi C + \phi D}$$

Equação 26- Proporção dos Contribuintes na Amostra

As proporções entre picos – PHR (Peak Height Ratio), é a razão entre os picos de um dado genótipo e é calculada pela formula:

$$PHR = \frac{\phi_{menor}}{\phi_{maior}}$$

Equação 27-Peak Height Ratio - PHR

Onde ϕ é o tamanho do pico. Para verificar qual é o genótipo de cada pessoa envolvida na amostra, deve-se considerar todos os picos presentes no locus dois a dois.

É importante salientar que estes métodos de determinação de picos não se aplicam quando não é observado o limite do threshold analítico. Isto pode acontecer quando a concentração de DNA nas amostras é muito baixa e o sequenciamento não dá conta de amplificar devidamente o DNA da mistura. Nestes caso outras abordagens que incluem a avaliação da probabilidade de um pico ser verdadeiro ou ser fruto de um artefato (e.g. falha na detecção do pico, deleção de picos que existem (drop-out), inserção de picos que não existem (drop-in) ou desbalanço entre os alelos heterozigotos) podem ser utilizados, porém estes métodos ainda não são aplicados em laboratórios de genética forense como rotina.

2.11.1 LR para Misturas

Dependendo do modelo utilizado para análise dos picos, devemos escolher o modelo para determinação da razão de verossimilhança que deverá ser utilizada para avaliar a força da evidência no caso de misturas. Quando os alelos estão bem definidos pela técnica de separação dos alelos, então podemos considerar o cálculo do LR a partir dos alelos da evidência (Mistura) e dos perfis dos envolvidos (Vitima

e suspeitos). Dadas as seguintes hipóteses quanto a origem dos contribuintes de uma mistura no caso criminal na presença da vítima e de um suspeito:

H_p : A Vítima e o Suspeito são os contribuintes da mistura (evidência)

H_d : A Vítima e um desconhecido são os contribuintes da mistura.

Dados os genótipos da mistura **M**, e **K** sendo o da Vítima e do(s)

Suspeito(s) assim podemos calcular LR através da fórmula:

$$LR = \frac{P(M, K|H_p)}{P(M, K|H_d)} = \frac{P(M|K, H_p) \times P(K|H_p)}{P(M|K, H_d) \times P(K|H_d)} = \frac{P(M|K, H_p)}{P(M|K, H_d)}$$

Equação 28-LR para misturas

Como os genótipos envolvidos **K**, não dependem das hipóteses, então podemos considerar $P(K|H_p) = P(K|H_d)$.

Alguns casos comuns podem envolver as hipóteses com um suspeito e dois desconhecidos, ou dois suspeitos e dois desconhecidos, sendo assim uma abordagem geral para o cálculo do LR pode ser resumido pelo cálculo de $P(M|K, H)$ onde **H** é a preposição especificada pelos contribuintes da mistura. Dados: a quantidade de alelos envolvidos na mistura; a quantidade de alelos que estão presentes na mistura mas ausentes nos envolvidos e a quantidade de contribuintes testados, o diagrama abaixo, baseado em (Fung and HU 2008), mostra o valor de **LR** para $P(M|K, H)$ nos casos onde a hipótese de acusação $P(M|K, H_p)=1$ que significa que todos os genótipos sob a hipótese da acusação são encontrados nas amostras.

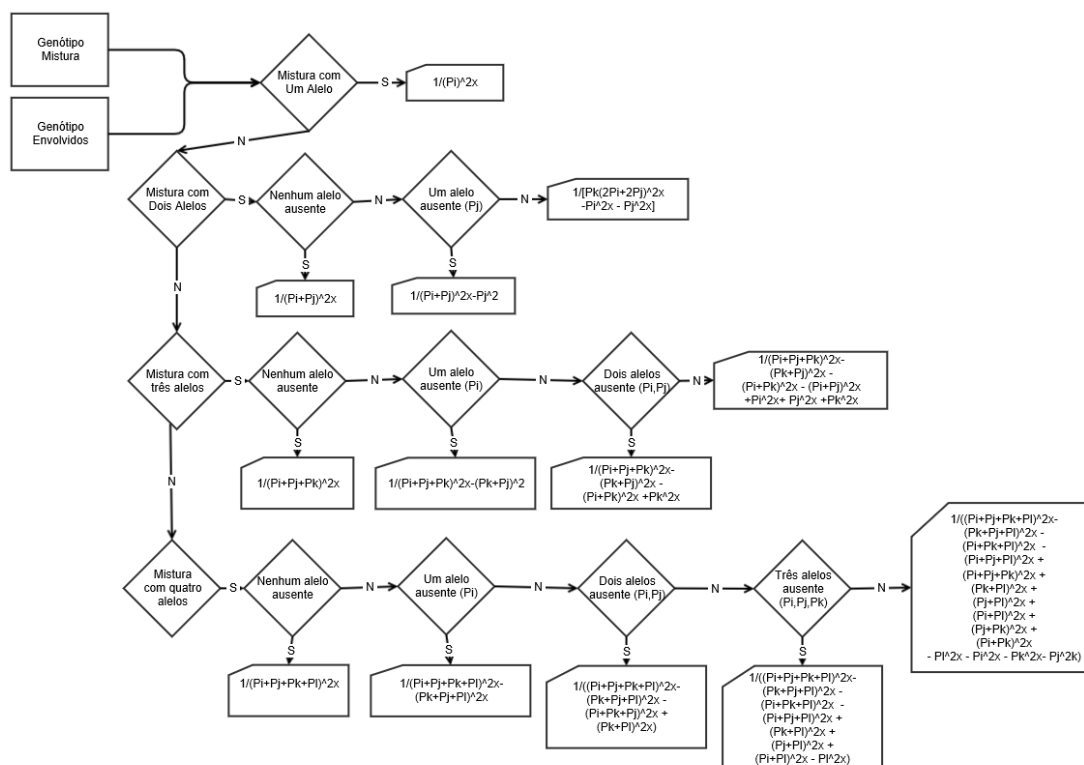


Figura 6- LR para o caso geral de misturas

2.12 Correção populacional

Nas populações a mistura dos alelos não é totalmente aleatória, pois seus indivíduos possuem alguma ancestralidade em comum. Para evitar, ou diminuir, os efeitos da subestruturação populacional nos cálculos do LR, é incorporado um valor para o índice de fixação de Wright (Fst) que em genética forense foi nomeado com o símbolo Θ (Theta). Este parâmetro é definido como a probabilidade de dois alelos serem idênticos por descendência (Buckleton, Triggs, and Walsh 2004). Quanto maior o valor deste índice, maior é a subestruturação populacional. O National Research Council (NRC-II 1996) recomenda a utilização de um valor $\Theta=0,01$ para populações não indígenas e $\Theta=0,03$ para populações indígenas. O NRC-II também definiu uma correção sobre o valor do LR quando se supõe que a vítima e o suspeito pertencem à mesma subpopulação e é dado pela seguinte formula:

$$P(A_i A_j | A_i A_j) = \begin{cases} \frac{(3\theta + (1 - \theta)P_i)(2\theta + (1 - \theta)P_j)}{(1 + \theta)(1 + 2\theta)}, & A_i = A_j \\ \frac{2(\theta + (1 - \theta)P_i)(\theta + (1 - \theta)P_j)}{(1 + \theta)(1 + 2\theta)}, & A_i \neq A_j \end{cases}$$

Equação 29- Correção populacional LR

3. Objetivos

3.1 Objetivo geral

Construir e validar, através de softwares de referência, um aplicativo (software) para uso em genética forense focado na rotina de um laboratório e que implemente os índices de genética de população em banco de dados forense.

3.2 Objetivos específicos

1. Construir uma interface Web, simples e objetiva que facilite o uso pelos técnicos e usuários envolvidos na rotina.
2. Disponibilizar rotinas para os cálculos estatísticos em genética forense nos seguintes casos: Trio Simples (mãe, suposto pai e filho) e Duo (suposto pai e filho) e nos casos de misturas com vítima e suspeito.
3. Cálculo dos parâmetros **Mx** e **PHR** para avaliação de picos em misturas.
4. Disponibilizar os seguintes cálculos aplicados em genética de populações nas tabelas de frequência forense: teste qui-quadrado para o equilíbrio de Hardy-Weinberg, poder de exclusão do locus, poder de discriminação, poder de match para o locus e para a tabela, probabilidade combinada de exclusão, conteúdo de informação polimórfica, índice de fixação de Wright, índice de paternidade típico e índice de paternidade típico acumulado, heterozigosidade observada e esperada e erro padrão, quantidade de alelos raros no locus.

4. Material e Métodos

4.1 Levantamento dos requisitos

Para levantamento dos requisitos funcionais do software foram entrevistados técnicos e responsáveis nos seguintes laboratórios: LABBE - laboratório de bioinformática e biologia evolutiva da UFPE, Laboratório de perícia e pesquisa em

genética forense da polícia civil de Pernambuco, Laboratório de genética forense da polícia civil da Paraíba. Um levantamento dos casos de uso foi realizado com base nas demandas encontradas nestes laboratórios.

Nesta etapa, também foi realizada uma avaliação dos pontos fortes e fracos dos softwares utilizados nas rotinas destes laboratórios, como o software Familias (Kling, Tillmar, and Egeland 2014), e o software DNAmix (Curran et al. 1999).

4.2 Implementação

Para a implementação dos requisitos levantados, foi criado o software DNAstr.com. Este software automatiza todas as etapas necessárias para a emissão de um laudo típico de genética forense. Além disto dispõe de rotinas para avaliação de tabelas de frequências que podem ser usadas para realização dos cálculos envolvidos nos laudos. Foram desenvolvidas rotinas para facilitar a inclusão no sistema de vários formatos de arquivos compatíveis com vários softwares de genética de populações.

A arquitetura de software utilizada foi a cliente-servidor com três camadas: interface, regras e dados. A implementação dos algoritmos do lado do servidor foi realizada utilizando a linguagem de programação PHP (*Hypertext Preprocessor* ou Processador de Hipertextos) que é uma linguagem de script open source de uso geral, muito utilizada, e especialmente adequada para o desenvolvimento web e que pode ser embutida dentro do HTML. Também foi utilizado, para dar suporte no lado do cliente, a linguagem de script JQuery, que é baseado em Javascript, e para a interface gráfica foi utilizado o framework Bootstrap que implementa, além de alguns códigos jquery, uma camada CSS (*Cascading Style Sheets* que em português seria algo como “folhas de estilo em cascata”). CSS é uma especificação que define como os elementos que compõem uma página, um documento ou

aplicação Web serão exibidos. Todos os algoritmos de cálculos foram implementados segundo a literatura corrente específica para cada caso. Também foi criado um domínio público específico para o software, o www.dnastr.com e implementado um certificado de segurança pública SHA2 (*Security Hash Algorithm* ou .Algoritmo HASH de segurança).

Os códigos em php foram escritos utilizando a IDE Aptana Studio V 3.0. O banco de dados MySQL foi desenvolvido utilizando o aplicativo MySQL Workbench 6.1 ce e o servidor MySQL 5.1 foi utilizado como gerenciador do banco de dados. O sistema foi hospedado no servidor apache v 2.4.10. Todos estes softwares são distribuídos sob a licença open-source.

4.3 Amostras de genotipagens

Para validar os módulos de tabelas, genotipagem, teste de paternidade e teste de mistura foram utilizadas amostras de arquivos de genotipagem da literatura e dados do laboratório LABBE/UFPE.

As amostras para validação das rotinas de tabela de frequência foram divididas em duas categorias: Tabelas com frequências dos alelos nos locus e Tabelas com genotipagem de indivíduos. Os cálculos de alguns índices descritos neste trabalho requerem dados de genotipagens individuais. A tabela apenas com frequências por locus utilizada como amostra para validação da importação e cálculos foi a descrita em Aguiar e colaboradores (Aguiar et al. 2014) que é uma das tabelas utilizadas como referência nos laboratórios pesquisados. Para a tabela com genotipagens foi utilizada a referida por C. Hill (Hill et al. 2013) com 1036 genótipos não relacionados.

Para validação dos resultados referentes aos testes de paternidade e análise de misturas foi utilizado a base de dados de genótipos criada para validação de softwares de genética forense (O'Connor et al. 2010) no formato CSV (separado por ponto-e-vírgula). Além destes dados, foram utilizados alguns arquivos de genotipagem no formato CMF e tabulado exportados pelos softwares GeneMapper e Osiris utilizado no LABBE sem qualquer identificação de pessoas.

5. Resultados

5.1 Base de dados do programa DNAstr.com.

Para dar suporte ao programa DNAstr.com e suas funcionalidades, foi projetada uma estrutura de banco de dados relacional utilizando o MySQL WorkBench 6.1 ce. A Tabela 3 descreve as entidades (tabelas do banco de dados) e os relacionamentos de cada entidade, no anexo 1 está o diagrama de dados completo.

Tabela 3- Tabelas do banco de dados

Nome	Descrição	Relacionamentos
Login	Dados dos login dos usuários	Laboratório, Perfil
Perfil	Perfil do sistema	Funcionalidades, Login
Funcionalidades	Funcionalidades disponíveis por perfil	Perfil
Laboratório	Armazena informações do laboratório	
Tabelas	Tabelas de frequência	Laboratório
Locus	Descrição dos Locus	
Alelos	Descrição dos alelos	
Frequências	Armazena as frequências alélicas das tabelas	Locus, Alelos, Tabelas
População	Dados das populações	Locus, Alelos, Tabela
Tipos Casos	Cadastro dos tipos de casos	
Sistemas	Cadastro dos sistemas de genotipagem	

Casos	Casos cadastrados	Tipo_caso, Laboratório	Sistemas,
Personagens	Personagens envolvidos nos casos		
Pessoas	Pessoas associadas aos casos		
Populacao_lab	Genotipagem das pessoas associadas aos casos.	Locus, Alelos, Pessoas, Laboratórios	
Auditoria	Tabela com os logs do sistema		
População_Forense	Tabela de genotipagem dos indivíduos associados aos casos forenses.	Caso,Locus,Laboratorio	
Modelo_mutacao	Cadastra os modelos de mutação associados às tabelas.	Tabelas, Locus, Laboratorio	
Cadastrar	Tabela com os dados de login a serem confirmados.		
Comandos	Tabela referente aos comandos executados sobre as tabelas.		

5.2 Modulo de Tabelas (genética de populações)

O modulo de tabelas possui as seguintes funcionalidades:

5.2.1 Inserir Tabela:

Neste modulo é possível cadastrar tabelas de frequência alélicas por locus ou tabelas de genótipos. As tabelas de genótipos são transformadas em tabelas de frequência pelo cálculo das frequências observadas nos indivíduos da tabela.

- I. Cadastro de tabela de frequências alélicas via arquivo: Para cadastrar uma tabela de frequências é necessário um arquivo do tipo TXT cujo formatação dos dados é especificado no anexo 2. A Figura 7 mostra a tela de cadastro.

The screenshot shows a web interface titled 'Formato 1' with a red header. In the top right corner, there is a button labeled 'Ajuda'. Below the header, there is a section for file upload with the text 'Arquivo Para Upload' and a button 'Escolher arquivo'. To the right of the button, it says 'Nenhum arquivo selecionado'. Below this, there are two text input fields: 'Nome (Tabela):' and 'Descrição da Tabela:'. At the bottom left, there is a blue button labeled 'Enviar'.

Figura 7 : Tela inserir frequências alélicas via arquivo

- II. Cadastro de tabela de frequências alélicas manualmente: Nesta opção o usuário deverá inserir manualmente o nome da tabela, a descrição e as frequências alélicas por locus. A Figura 8 mostra a tela de cadastro.

The screenshot shows a web interface titled 'Inserir Tabela Manualmente' with a red header. In the top right corner, there is a button labeled 'Ajuda'. Below the header, there are two text input fields: 'Nome (Tabela):' and 'Descrição da Tabela:'. Below these, there is a row of controls: a dropdown menu for 'Locus' with 'D3S1358' selected, a dropdown menu for 'Alelo' with '12' selected, a text input field for 'Frequencia', and a button labeled 'Inserir Locus/Alelo'.

Figura 8: cadastro manual de tabela de frequências

- III. Cadastro de tabela de genotipagem (população): existem três possibilidades de inserção de genotipagem, são elas: Formato do plink (PED e MAP), formato do Genpop e um formato próprio, as descrições estão no anexo 2. A Figura 9 mostra a tela de cadastro do formato plink.

The screenshot shows a web interface titled 'Formato 3 plink (PED, MAP)' with a red header. In the top right corner, there is a button labeled 'Ajuda'. Below the header, there are two sections for file upload. The first section is for 'Arquivo Format PED para upload:' with a button 'Escolher arquivo' and the text 'Nenhum arquivo selecionado'. The second section is for 'Arquivo Format MAP para upload:' with a button 'Escolher arquivo' and the text 'Nenhum arquivo selecionado'. Below these, there are two text input fields: 'Nome (Tabela):' and 'Descrição da Tabela:'. At the bottom left, there is a blue button labeled 'Enviar'.

Figura 9: Inserir genotipagem formato plink

5.2.2 Visualizar Tabela:

Neste modulo é possível verificar as frequências alélicas em cada locus e todos os índices de genética de população associados as tabelas de frequências cadastradas. Os algoritmos implementados neste módulo são responsáveis pelos seguintes índices:

Tabela 4- Índices tabelas

NI	Número de indivíduos no locus
NL	Número de alelos no locus
PAR	Quantidade de alelos raros no locus
Ho	Heterozigosidade Observada no locus. Ho media
He	Heterozigosidade Esperada no locus
Se	Erro padrão.
TPI	Índice de paternidade típico e TPI acumulado
F	Índice de fixação WRIGHT
PM	Poder de Match do Locus e PM combinado
PD	Poder de Discriminação
PE	Poder de Exclusão do Locus - PE typical
HWE	Teste qui-quadrado para o equilíbrio de Hardy-Weinberg
PIC	Conteúdo de Informação Polimórfica

A Figura 10 mostra a tela onde é possível visualizar a tabela de frequência.

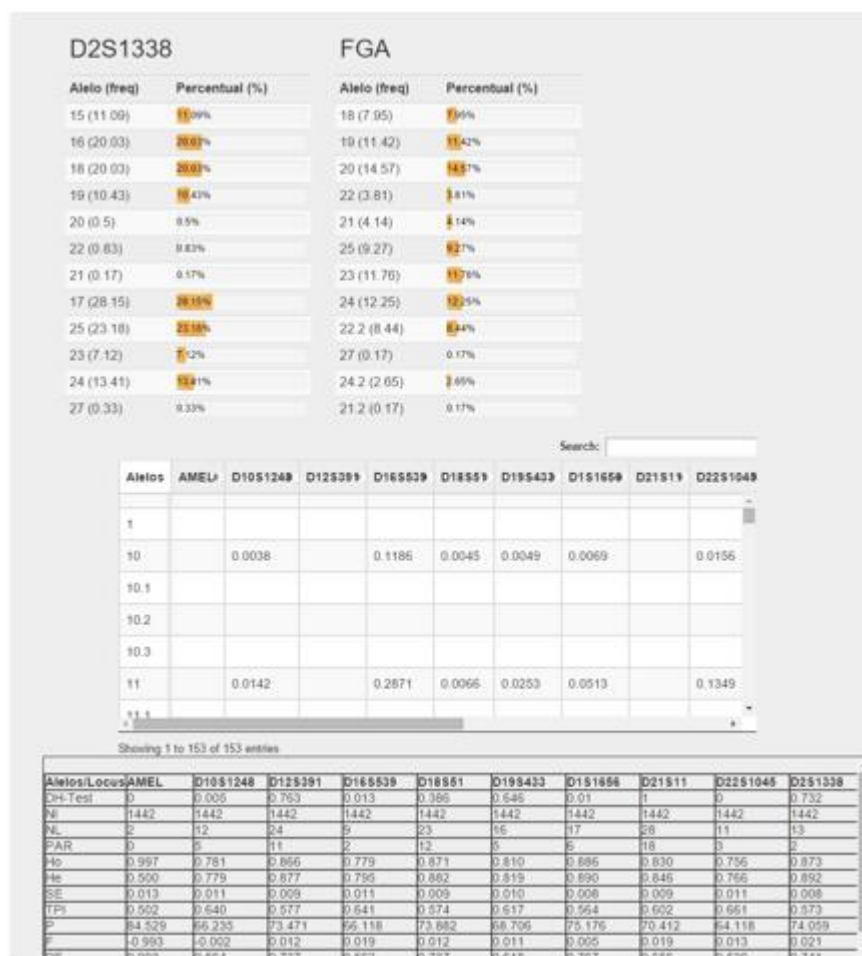


Figura 10: Tela de visualização das tabelas de frequências

Validamos os resultados com o software EasyDNA_PopuData e a planilha PowerStatsV12.xlsx. Para esta validação consideramos um locus da base de dados (Hill et al. 2013). Duas limitações foram encontradas nos softwares: O software EasyDNA_PopuData não permite trabalhar com um “n” amostral maior que 300 genótipos e a planilha PowerStats não trabalha com mais de 600. A Tabela 5 está dividida em duas colunas com os resultados dos índices avaliados.

Tabela 5- Comparação de Índices

	N=300			N=600	
Índice	DNAstr.com	DNA_easy	PowerStats	DNAstr.com	PowerStats
DH-Test	0,021	X	X	0	X
PAR	3	X	X	5	X
Ho	0,799	0,797	0,798	0,802	0,801
He	0,781	0,78	X	0,779	X

SE	0,024	0,024	X	0,0017	X
TPI	2,492	X	2,48	2,521	2,51
F	-0,024	X	X	-0,030	X
PE	0,598	0,575	0,595	0,602	0,600
PD	0,912	0,918	0,913	0,913	0,913
PM	0,088	X	0,087	0,087	0,087
PIC	0,747	X	0,75	0,745	0,74

As frequências dos alelos nos locus calculadas pelos softwares estão na Tabela 6.

Tabela 6- Comparação entre Frequências

	N=300			N=600	
Alelo	DNAstr.com	DNA_easy	PowerStats	DNAstr.com	PowerStats
8	0,0033	0,003	0,003	0,0017	0,002
9	X	X	X	0,0017	0,002
10	0,0017	0,002	0,002	0,0017	0,002
11	0,0151	0,015	0,015	0,0158	0,016
12	0,0803	0,08	0,081	0,0775	0,077
13	0,2625	0,262	0,261	0,2817	0,281
14	0,2826	0,282	0,283	0,2867	0,288
15	0,2341	0,237	0,234	0,2075	0,207
16	0,0920	0,092	0,093	0,10	0,10
17	0,0268	0,027	0,027	0,0225	0,023
18	X	X	X	0,0017	0,002
19	0,0017	0,002	0,002	0,0017	0,002

A planilha PowerStats permite a visualização no formato Match/Partially, onde as linhas e colunas representam os alelos e as células representam a quantidade de “match” encontrados nos perfis para o mesmo locus. A ferramenta DNAstr.com também possui esta funcionalidade, mostrada na Figura 11 abaixo para o locus utilizado nos cálculos anteriores.

Match/Partial	8	9	10	11	12	13	14	15	16
10									
11									
12	1		1	1	5				
13		1	1	5	29	37			
14				10	36	108	41		
15	1	1		2	10	75	65	30	
16					4	40	29	30	6

Figura 11-Match/Partially dos Alelos

A ferramenta DNAAstr.com permite também mostrar o gráfico Match/Partially (Figura 12-Match/Partially), com todos os loci, neste caso as linhas e colunas representam a quantidade de loci, as células representam a quantidade de pares de genótipos com Match/Partially. Além deste gráfico, um Heatmap dos dados (Figura 13) e uma lista com os totais de match/partially e total por locus (Figura 14).

Match/Partial	0	1	2	3	4	5	6	7	8	9	10
0	1	46	342	1301	3738	8150	12642	15050	14226	10677	586
1	10	83	458	1818	4821	9296	13032	13807	11720	7208	356
2	7	66	356	1160	2809	4806	5966	5508	3983	2100	804
3	5	35	139	422	1000	1508	1564	1267	749	326	95
4	0	10	46	138	233	298	263	181	99	31	10
5	0	2	9	20	41	46	35	14	8	2	0
6	0	0	0	2	4	3	5	1	0	0	0

Figura 12-Match/Partially

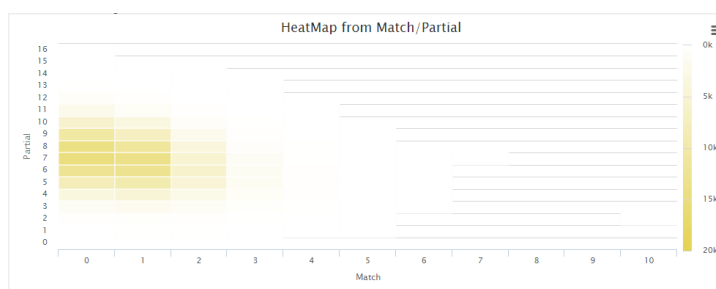


Figura 13- Heatmap dos dados Match/Partially

Matches	Partially	Total
D1051248 10.228%	D1051248 7.811%	D1051248 8.074%
D125391 7.29%	D125391 4.566%	D125391 4.239%
D165539 9.414%	D165539 7.281%	D165539 7.495%
D18551 8.28%	D18551 5.67%	D18551 4.41%
D195433 6.24%	D195433 6.688%	D195433 6.692%
D151656 4.465%	D151656 4.766%	D151656 4.516%
D21511 4.713%	D21511 6.053%	D21511 5.38%

Figura 14- Lista Match/Partially por locus

5.3 Módulo de Paternidade

O modulo de paternidade foi desenvolvido com dois objetivos: o primeiro é servir como base para emissão de laudos automaticamente em laboratórios de teste de paternidade e o segundo foi elaborado para facilitar a visualização e montagem de casos.

Para atender o primeiro objeto foi desenvolvido uma funcionalidade que agrupa todos os passos necessários para a conclusão de um caso de paternidade numa única tela. Este recuso chama-se “Pipeline”. A Figura 15 abaixo mostra esta tela com as suas opções.

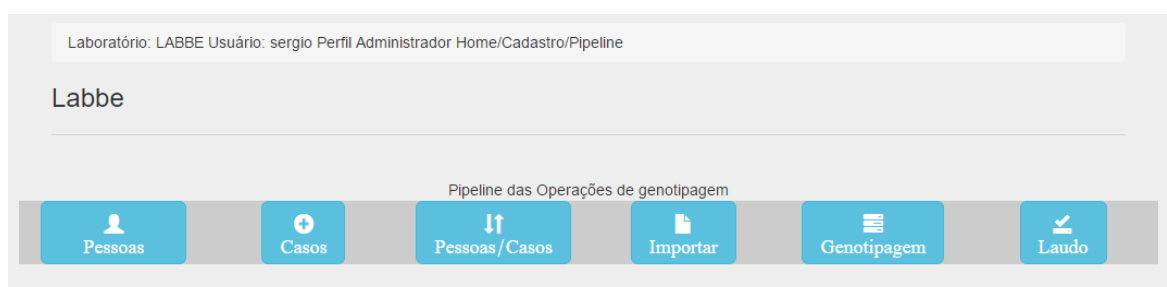


Figura 15- Pipeline Paternidade

5.3.1 Cadastro de Pessoas

No modulo de cadastro de Pessoas (Figura 16), deve-se informar um identificador único por pessoa ou amostra. Este identificador poderá ser o CPF ou outro conjunto qualquer de caracteres até 14 posições. O sexo é optativo, pois este

modulo também é usado para cadastrar amostras utilizadas em análise de misturas.

A interface do formulário 'Cadastro Individual de Pessoas' possui um cabeçalho vermelho com o título 'Cadastro Individual de Pessoas'. Abaixo dele, uma barra cinza contém o texto 'Informar dados sobre as pessoas que serão genotipadas' e um botão 'Ajuda' à direita. O corpo do formulário é branco e contém três campos obrigatórios marcados com um ponto: 'Cpf/Doc:' com um campo de texto, 'Nome :' com um campo de texto, e 'Sexo : Feminino' com uma lista suspensa. No canto inferior esquerdo, há um botão azul com o texto 'Enviar'.

Figura 16- Modulo de cadastro de pessoas

5.3.2 Cadastro de Casos

O módulo de cadastro de casos é utilizado para criar um novo caso. Nesta tela, Figura 17, é registrado o código do caso, que é um identificador único para cada laboratório; a data da inclusão do caso; o sistema utilizado na genotipagem e o tipo de caso que pode ser: Trio, Duo, Mistura ou reconstrução na presença dos avós.

A interface do formulário 'Cadastro de Casos' possui um cabeçalho vermelho com o título 'Cadastro de Casos'. Abaixo dele, uma barra cinza contém o texto 'Informar dados sobre as pessoas que serão genotipadas' e um botão 'Ajuda' à direita. O corpo do formulário é branco e contém quatro campos obrigatórios marcados com um ponto: 'Código Caso:' com um campo de texto, 'Data : dd/mm/aaaa' com um campo de texto, 'Sistema utilizado na genotipagem: Cofiler' com uma lista suspensa, e 'Tipo : Mãe, Suposto Pai e Filho' com uma lista suspensa. No canto inferior esquerdo, há um botão azul com o texto 'Enviar'.

Figura 17- Tela de cadastro de casos

5.3.3 Cadastro das pessoas nos casos

Neste módulo é possível associar as pessoas cadastradas no sistema à um determinado caso também cadastrado no sistema (Figura 18). Nesta

funcionalidade o sistema verifica qual o tipo de caso para adaptar o número de pessoas e a função de cada um dentro do caso (ex. mãe, pai, filho).

Figura 18- Cadastro das Pessoas nos Casos

5.3.4 Importar Genotipagem

A importação dos genótipos pode ser feita utilizando os formatos CMF versão 3.2 que foi criado pelo CODIS (Combined DNA Index System) para padronizar a troca de dados entre os aplicativos de genética forense ou pelo formato tabular com separação de campos por ponto-e-vírgula, que é o padrão utilizado por alguns softwares de genotipagem. Estes formatos permitem que vários genótipos sejam exportados juntos, permitindo o uso de apenas um arquivo por caso. A , Figura 19 mostra a tela de cadastro de genotipagem.

Associar arquivo de genotipagem aos casos

Formato de interface entre o CODIS e sistemas externos.
Para maiores informações sobre este formato, baixe o arquivo [aqui](#)

Caso: caso_trio_simples

Mãe : CPF: 3 Nome: 3
Escolher a amostra 20-2SP_D11_04 ▼

Suposto Pai : CPF: 2 Nome: 2
Escolher a amostra 20-2SP_D11_04 ▼

Filho : CPF: 11 Nome: 11
Escolher a amostra 20-2SP_D11_04 ▼

Enviar

Importar Arquivos de Genotipagem

Codigo do Caso Labbe_02

DADOS DA GENOTIPAGEM Arquivo CMF

ID de submissão : 20141021_131906
 Data de submissão : 2014-10-21T16:19:36
 Responsável submissão : Abigail
 Kit Usado na Genotipagem : PowerPlex Fusion
 Responsável recebimento sistema :sergio

Locus/Alelos	Jose Ferreira da Silva CPF = 1	Carlos Figueredo CPF = 2	Moises Freitas CPF = 3	Responsável
AMEL	X	X/Y	X/Y	Abigail
D3S1358	OL15.3/16	15/17	14/16	Abigail
D13S317	12/14	8/12	11/12	Abigail
Penta E	12	8/14	12/13	Abigail
D16S539	10/13	11	10/12	Abigail
D18S51	13/16	12/19	14/16	Abigail
D2S1338	17/19	16/20	17/23	Abigail
CSF1PO	7/10	12	7/12	Abigail
Penta D	7/13	8/9	13/15	Abigail

Figura 19- Telas de Inserção de Genotipagem

5.3.5 Cadastro Genotipagem

Correções na genotipagem podem ser feitas nesta opção (Figura 20). O campo Código do caso ou CPF/ID é utilizado para localizar a genotipagem ou montar uma tela para inserção da genotipagem. Quando um locus é modificado, o campo ficará verde se o alelo for válido ou vermelho caso o alelo ou formato do campo esteja inválido ou não cadastrado.

Cadastro de Genotipagem

Cadastrar e Corrigir Genótipos. Ajuda

Código Caso ou CPF/ID: labbe_01 Localizar Caso

Tipo de Caso: Mãe, Suposto Pai e Filho
 Suposto Pai : 1 Jose Antonio da Silva
 Mãe : 2 Maria dos Santos
 Filho : 3 Tonho dos Santos

Escolher Alelos por locus

Locus/Alelos	Maria dos Santos	Jose Antonio da Silva	Tonho dos Santos
AMEL	X/X	X/Y	X/Y
CSF1PO	7/10	12/12	7/12
D16S539	10/13	11/11	10/12
D3S1358	14/18	15/17	14/16
D7S820	8/10	10/11	10/11
TH01	8/9.3	9/9	7/8.7
TPOX	8/11	8/8	8/8

Inserir

Figura 20- Cadastro de Genotipagem

5.3.5 Laudo

O Laudo é o último passo do processo de análise dos casos de paternidade. Para emissão do laudo é necessário indicar a tabela de frequência que será utilizada para o cálculo e o número do caso. É possível emitir um resumo com os cálculos de índice de paternidade (IP) por locus, índice de paternidade acumulada (IPC) e a probabilidade a posteriori (W) por locus e total. É possível salvar nos formatos PDF, XLS excel, CSV e copiar para a área de transferência. O laudo ainda permite verificar a genotipagem por locus; alelo compartilhado; frequência do alelo compartilhado; uma mensagem de observação por locus que indica exclusão, inclusão, inconsistências nos locus (alelos não coincidentes) ou se o locus não está cadastrado na tabela de frequência utilizada; quantidade total de locus e a quantidade de locus que indicam exclusão.

PDF	Copy	Excel	CSV	Search:			
Locus/Alelos	Mãe: 3	Suposto Pai: 2	Filho: 11	Alelo Ref.	Freq	IP	IPC
AMEL	X;X	X;Y	X;X	X	----	----	
CSF1PO	9;11	10;12	9;12	12	0.31529	1.5858	1
D13S317	12;12	11;13	11;12	11	0.29570	1.6909	2
D16S539	9;12	11;11	11;12	11	0.29015	3.4465	5
D18S51	14;16	12;14	12;14	12	0.11949	4.1845	3
D21S11	28;29	28;31.2	29;31.2	31.2	0.10419	4.7989	1
D3S1358	15;17	15;16	15;16	16	0.27286	1.8324	3

Quantidade Total de locus: 14

Quantidade de locus considerados: 13

Quantidade de locus que indicam exclusão: 0

IPC: 254360.82230881

W : 0.99999606859241 ou 99.999606859241%

Figura 21- Tela Laudo

Para validação dos casos de paternidade, utilizamos o software famílias versão 3.0. com o banco de dados de genótipos aparentados do NIST (Connor et al. 2010). Para os cálculos foi utilizada a tabela de frequência Brasil 2014 (Aguilar et al. 2014). Os seguintes locus foram utilizados para os cálculos: D8S1179, D21S11, D7S820, CSF1PO, D3S1358, TH01, D13S317, D16S539, vWA, TPOX, D18S51, D5S818, FGA.

A Figura 22 mostra o heredograma (Connor et al. 2010) utilizado como referência para a montagem dos casos analisados. Os números abaixo de cada símbolo representam os identificadores dos indivíduos no banco de dados.

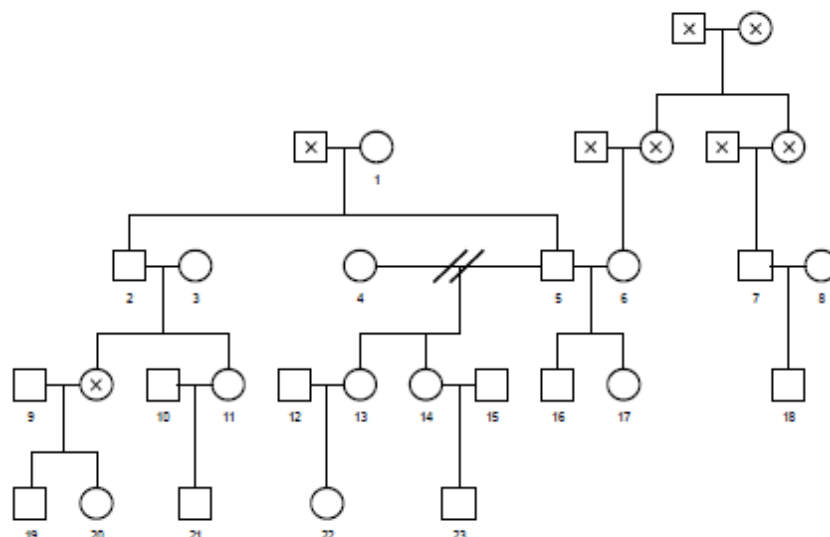


Figura 22-Heredograma usado na validação

Foram testadas as seguintes combinações:

Tabela 7- Estudo de casos de paternidade

Tipo	Identificação Caso	Descrição	ID Mãe	ID Suposto Pai	ID Filho
Trio Simples	caso_trio_simples	Inclusão	3	2	11
Trio Simples	caso_trio_simples1	Inclusão	8	7	18
Trio Simples	caso_trio_simples2	Exclusão	6	19	17
Trio Simples	caso_trio_simples3	Exclusão	11	7	21
Duo Simples	caso_duo_simples	Inclusão		9	20
Duo Simples	caso_duo_simples1	Inclusão		5	16
Duo Simples	caso_duo_simples2	Exclusão		2	18
Duo Simples	caso_duo_simples3	Exclusão		7	20

Tabela 8- Resultados testes paternidade

Caso_trio_simples			Caso_trio_simples1		
Locus	IP (LR)		Locus	IP (LR)	
	Familias	DNAstr		Familias	DNAstr
CSF1PO	1,5858	1,5858	CSF1PO	3,1717	3,1717
D13S317	1,6909	1,6909	D13S317	5,1225	5,1224
D16S539	3,4465	3,4465	D16S539	2500	2500
D18S51	4,1845	4,1845	D18S51	12,8817	12,8816
D21S11	4,7990	4,7989	D21S11	2,2416	2,2416
D3S1358	1,8325	1,8324	D3S1358	5,5316	5,5316
D5S818	1,5262	1,5262	D5S818	2,8871	2,8871

D7S820	3,0954	3,0954	D7S820	0,9734	0,9734
D8S1179	1,9819	1,9819	D8S1179	1,7669	1,7669
FGA	3,4294	3,4294	FGA	---	---
TH01	3,3623	3,3622	TH01	1,0748	1,0748
TPOX	0,8660	0,8661	TPOX	2,1982	2,1982
vWA	8,0000	8,0000	vWA	3,8932	3,8932
IPC	254.376,0843	254.366,4647	IPC	296312653,2	296.300.801
W	0,999996	0,999996	W	>0,999999	0,999995

Caso_trio_simples2			Caso_trio_simples3		
Locus	IP (LR)		Locus	IP (LR)	
	Familias	DNAstr		Familias	DNAstr
CSF1PO	----	1,8831	CSF1PO	2,9649	2,9649
D13S317	1,271274567	1,2713	D13S317	----	----
D16S539	----	1,7232	D16S539	----	----
D18S51	----	----	D18S51	----	----
D21S11	----	----	D21S11	2,1190	2,119
D3S1358	1,678657759	1,6786	D3S1358	----	----
D5S818	1,526158354	1,5262	D5S818	2,887086064	2,8871
D7S820	----	----	D7S820	4,2269	4,2269
D8S1179	----	----	D8S1179	0,9341	0,9341
FGA	4,453241895	4,4532	FGA	----	----
TH01	4,547108039	4,5471	TH01	2,2736	2,2736
TPOX	2,198166711	2,1982	TPOX	----	----
vWA	5,891710364	5,8917	vWA	3,8932	3,8932
IPC	854,1117	854,0861	IPC		162,8296
W	0,998831	0,998830	W		0,996639

Caso_duo_simples			Caso_duo_simples1		
Locus	IP (LR)		Locus	IP (LR)	
	Familias	DNAstr		Familias	DNAstr
CSF1PO	0,792920803	0,79292	CSF1PO	0,792920803	0,79292
D13S317	3,405955794	3,40592	D13S317	1,690919851	1,6909
D16S539	0,980776775	0,98078	D16S539	1,723246597	1,72325
D18S51	1,735887377	1,73587	D18S51	2,318702467	2,31868
D21S11	7,575833333	7,57576	D21S11	1,059512205	1,0595
D3S1358	5,531640668	5,53159	D3S1358	1,755558814	1,75554
D5S818	0,721771516	0,72177	D5S818	1,443543032	1,44354
D7S820	2,485031792	2,48503	D7S820	2,115327664	2,11533
D8S1179	2,66741169	2,66739	D8S1179	3,918227412	3,91819
FGA	1,6368011	1,63677	FGA	1,714711934	1,71468
TH01	2,679225184	2,67923	TH01	2,27355402	2,27355
TPOX	1,439434178	1,43945	TPOX	1,099083355	1,09909
vWA	2,41621703	2,41622	vWA	1,472927591	1,47293
IPC	14060,459850	14059,714390	IPC	752,44040	752,39080

W	0,999929	0,999929	W	0,99867	0,99867
---	----------	----------	---	---------	---------

Caso_duo_simples2			Caso_duo_simples3		
Locus	IP (LR)		Locus	IP (LR)	
	Familias	DNAstr		Familias	DNAstr
CSF1PO	1,585841606	1,58584	CSF1PO	1,585841606	1,58584
D13S317	0,845459926	0,84545	D13S317	2,561238603	2,56121
D16S539	1,723246597	1,72325	D16S539	0,861623298	0,86162
D18S51	---	----	D18S51	----	----
D21S11	---	----	D21S11	----	----
D3S1358	----	----	D3S1358	2,765820334	2,76579
D5S818	0,763079177	0,76308	D5S818	1,443543032	1,44354
D7S820	1,057663832	1,05766	D7S820	-----	----
D8S1179	0,883463496	0,88345	D8S1179	0,883463496	0,88345
FGA	1,714711934	1,71468	FGA	----	----
TH01	1,13677701	1,13678	TH01	1,13677701	1,13678
TPOX	1,099083355	1,09909	TPOX	1,099083355	1,09909
vWA	0,943289439	0,94329	vWA	-----	----
IPC	3,3293	3,3291	IPC	15,4232	15,4226
W	0,769010	0,769007	W	0,939110	0,939108

5.3.6 Editor Visual – Casos de Paternidade

Também é possível montar casos utilizando o editor visual. Com esta ferramenta o usuário poderá testar de forma simples várias configurações de heredogramas, sem a necessidade de criar um caso, apenas com o cadastro das pessoas envolvidas e suas genotipagens. Para a validação dos resultados foram realizados os mesmos testes feitos na validação do laudo, os valores de mostraram idênticos ao calculado pela opção laudo.

Figura 23- Tela Editor Visual Paternidade

5.4 Módulo de Análise de Misturas

O módulo de análise de misturas possui basicamente a mesma estrutura do módulo de paternidade. A funcionalidade “Pipeline Forense” desenvolvida para facilitar o cadastro e análise dos dados, também foi customizado para a análise de misturas. A Figura 24 abaixo mostra a sequência de operação para a análise de misturas forense.

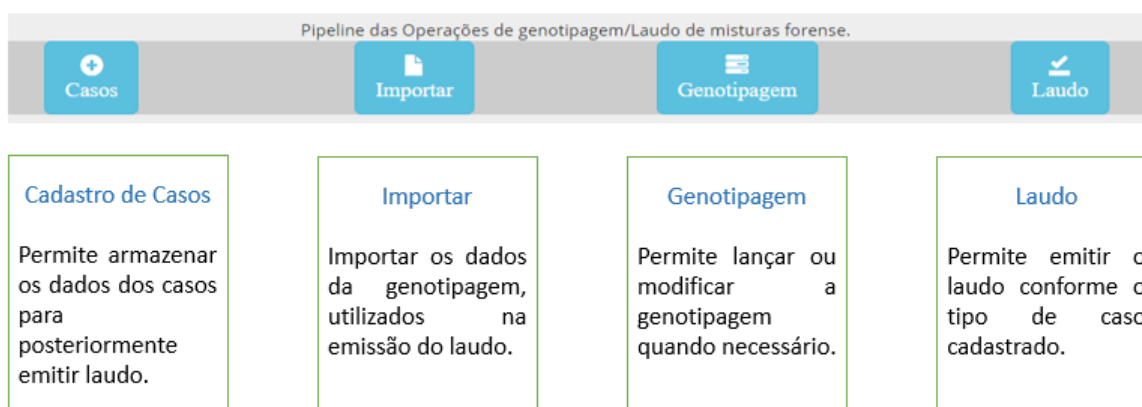


Figura 24- Pipeline Forense

5.4.1 Editor Visual – Misturas

Também foi desenvolvido um editor visual para construção de casos Forenses. Nele é possível, além de mostrar o laudo com os cálculos da razão de

verossimilhança, avaliar as relações Mx e PHR entre os picos da amostra de mistura.

Figura 25- Tela Editor Forense

Para validação dos resultados foi montado um caso com a genotipagem dos suspeito, vítima e mistura. Na genotipagem da mistura foram incluídos locus com 1,2,3 e 4 alelos, variando o tamanho dos picos entre eles.

Tabela 9-Caso Mistura 2 alelos

	Vitima				Suspeito				Mistura			
Locus	A1	ϕ	A2	Φ	A1	ϕ	A2	ϕ	A1	ϕ	A2	Φ
D8S1179	12	2720	13	2685	12	3117	13	2720	12	2514	13	2658
D22S1045	15	7127			16	6689			15	6706	16	4694
D10S1248	13	4508	16	4503	16	4901	16	4716	13	4544	16	9088

Tabela 10 - Caso Mistura 3 alelos

	Vitima				Suspeito				Mistura					
Locus	A1	Φ	A2	ϕ	A1	Φ	A2	ϕ	A1	Φ	A2	ϕ	A3	ϕ
D16S539	12	7743			13	3816	14	3303	12	7820	14	6256	13	4379
TPOX	9	2973	11	2768	8	5943			9	2522	11	2643	8	1513
D2S441	11.3	2090	14	1952	13	1992	14	1957	11.3	1860	13	2114	14	3171
TH01	8	1754	9.3	1743	8	1902	9	1491	8	1673	9	1232	9.3	1325
D5S818	11	9532			12	5287	13	5229	11	5148	12	4808	13	4500
D7S820	12	6175	13	4786	9	5587	12	5249	9	4954	12	5228	13	4823
D12S391	18	6221	16	5593	22	6268	22	5461	16	5136	18	5139	22	7708

Tabela 11 - Caso Mistura 4 alelos

	Vitima				Suspeito			
Locus	A1	ϕ	A2	ϕ	A1	Φ	A2	Φ

D3S1358	16	3552	17	3232	15	2997	18	2725
CSF1PO	10	4437	12	4028	11	4079	15	3779
D21S11	30.2	5642	32.2	5841	29	5597	30	6108
D19S433	13	1890	15	1548	11	1211	14	1435
FGA	20	2266	23	2202	21	2162	22	1541
D13S317	12	6425	13	4788	11	5377	14	5942
SE33	19	5011	28.2	4266	28.2	5047	29.2	4319
Mistura								
Locus	A1	ϕ	A2	ϕ	A3	Φ	A4	Φ
D3S1358	15	2564	17	2515	16	1538	18	1257
CSF1PO	10	4090	12	3815	11	2454	15	1907
D21S11	29	6306	32.2	5223	30.2	3783	30	2611
D19S433	11	1565	13	1244	14	939	15	622
FGA	20	2220	21	2137	22	1332	23	1068
D13S317	11	4630	13	4636	12	2778	14	2318
SE33	19	4874	26.2	4312	28.2	2924	29.2	2156

Onde ϕ é o tamanho do pico no eletroferograma. Os cálculos dos valores de PHR e Mx para todas as combinações dos alelos em dois genótipos possíveis foi realizado através da planilha MixtureAnalysisMaster (Gill et al. 1998) para comparação com o software DNASTR.com. Foram escolhidos para comparação seis loci sendo dois loci com quatro alelos, dois com três alelos e dois com dois alelos.

Tabela 12- Comparação PHR e Mx por Locus

D3S1358							
Cálculos							
Alelo:15 Height: 2564 Alelo:17 Height: 2515 Alelo:16 Height: 1538 Alelo:18 Height: 1257							
Genótipos		DNASTR.com			MixtureAnalysisMaster		
Major	Minor	Mx	PHR1	PHR2	Mx	PHR1	PHR2
15/17	16/18	0,645	0,98	0,82	0,645	0,98	0,82
15/16	17/18	0,521	0,6	0,5	0,521	0,6	0,5
15/18	17/16	0,485	0,49	0,61	0,485	0,49	0,61
17/18	15/18	0,515	0,61	0,49	0,515	0,61	0,49
17/16	15/18	0,479	0,5	0,6	0,479	0,5	0,6
16/18	15/17	0,355	0,82	0,98	0,355	0,82	0,98

D21S11							
Cálculos							
Alelo:29 Height: 6306							
Alelo:32.2 Height: 5223							
Alelo:30.2 Height: 3783							
Alelo:30 Height: 2611							
Genótipos		DNAstr.com			MixtureAnalysisMaster		
Major	Minor	Mx	PHR1	PHR2	Mx	PHR1	PHR2
29/32.2	30.2/30	0,643	0,83	0,69	0,643	0,83	0,69
29/30.2	32.2/30	0,563	0,6	0,5	0,563	0,60	0,50
29/30	32.2/30.2	0,498	0,41	0,72	0,498	0,41	0,72
32.2/30	29/30	0,502	0,72	0,41	0,502	0,72	0,41
32.2/30.2	29/30	0,437	0,5	0,6	0,437	0,50	0,60
30.2/30	29/32.2	0,357	0,69	0,83	0,357	0,69	0,83

D16S539							
Cálculos							
Alelo:12 Height: 7820							
Alelo:14 Height: 6256							
Alelo:13 Height: 4379							
Genótipos		DNAstr.com			MixtureAnalysisMaster		
Major	Minor	Mx	PHR1	PHR2	Mx	PHR1	PHR2
12/12	14/13	0,424	NA	0,7	0,424	NA	0,70
14/14	12/13	0,339	NA	0,56	0,339	NA	0,56
13/13	12/14	0,237	NA	0,8	0,237	NA	0,80
12/14	12/13	0,588	0,74	NA	0,588	0,74	NA
14/13	12/13	0,444	0,31	NA	0,444	0,31	NA
12/14	14/13	0,641	0,51	NA	0,641	0,51	NA
14/13	12/12	0,576	0,7	NA	0,576	0,70	NA
12/13	14/14	0,661	0,56	NA	0,661	0,56	NA
12/14	13/13	0,763	0,8	NA	0,763	0,80	NA
12/13	12/14	0,359	0,74	NA	0,359	0,74	NA
12/13	14/13	0,556	0,31	NA	0,556	0,31	NA
12/13	14/13	0,359	0,51	NA	0,359	0,51	NA

TPOX	
Cálculos	
Alelo:9 Height: 2522	
Alelo:11 Height: 2643	

Alelo:8 Height: 1513							
Genótipos		DNAstr.com			MixtureAnalysisMaster		
Major	Minor	Mx	PHR1	PHR2	Mx	PHR1	PHR2
9/9	11/8	0.378	NA	0.57	0,378	NA	0,57
11/11	9/8	0.396	NA	0.60	0,396	NA	0,60
8/8	9/11	0.227	NA	0.95	0,227	NA	0,95
9/11	9/8	0.636	0.61	NA	0,636	0,61	NA
11/8	9/8	0.512	0.29	NA	0,512	0,29	NA
9/11	11/8	0.625	0.66	NA	0,625	0,66	NA
11/8	9/9	0.622	0.57	NA	0,622	0,57	NA
9/8	11/11	0.604	0.60	NA	0,604	0,60	NA
9/11	8/8	0.773	0.95	NA	0,773	0,95	NA
9/8	9/11	0.375	0.61	NA	0,375	0,61	NA
9/8	11/8	0.488	0.29	NA	0,488	0,29	NA
9/8	11/8	0.375	0.66	NA	0,375	0,66	NA

D8S1179							
Cálculos							
Alelo:12 Height: 2514							
Alelo:13 Height: 2658							
Genótipos		DNAstr.com			MixtureAnalysisMaster		
Major	Minor	Mx	PHR1	PHR2	Mx	PHR1	PHR2
12/12	13/13	0.486	NA	0.95	0,486	NA	0,95
12/13	13/13	0.972	NA	NA	0,972	NA	NA
12/12	12/13	-0.028	NA	0.95	-0,028	NA	0,95
12/13	12/13	NA	0.95	NA	NA	0,95	NA
12/13	12/12	1.028	NA	NA	1,028	NA	NA
13/13	12/12	0.514	NA	NA	0,514	NA	NA
13/13	12/13	0.028	NA	NA	0,028	NA	NA

D22S1045							
Cálculos							
Alelo:15 Height: 6706							
Alelo:16 Height: 4694							
Genotipos		DNAstr.com			MixtureAnalysisMaster		
Major	Minor	Mx	PHR1	PHR2	Mx	PHR1	PHR2
15/15	16/16	0.588	NA	0.70	0,588	NA	0,70
15/16	16/16	1.176	NA	NA	1176,000	NA	NA
15/15	15/16	0.176	NA	0.70	0,176	NA	0,70
15/16	15/16	NA	0.70	NA	NA	0,70	NA
15/16	15/15	0.824	NA	NA	0,824	NA	NA

16/16	15/15	0.412	NA	NA	0,412	NA	NA
16/16	15/16	-0.176	NA	NA	-0,176	NA	NA

Para a comparação dos valores da razão de verossimilhança foi utilizado o software DNAmix (Curran et al. 1999). Foram utilizados três variações no valor do Theta populacional (ou Fst) : $\Theta=0$; $\Theta=0,01$; $\Theta=0,03$.

Tabela 13- Cálculos LR Misturas ($\Theta=0$)

Locus	Vitima Alelo(Frequência)	Suspeito Alelo(Frequência)	Mistura Alelo(Frequência)	DNAstr.com	DNAmix
D3S1358	16(0.27)/17(0.20)	15(0.30)/18(0.12)	15(0.30)/17(0.20)/16(0.27)/18(0.12)	13,98	14.0
D16S539	12(0.25)	13(0.15)/14(0.02)	12(0.25)/14(0.02)/13(0.15)	137,82	137.8
CSF1PO	10(0.27)/12(0.32)	11(0.29)/15(0.01)	10(0.27)/12(0.32)/11(0.29)/15(0.01)	169,98	170.0
TPOX	9(0.12)/11(0.28)	8(0.45)	9(0.12)/11(0.28)/8(0.45)	1,74	1.7
D8S1179	12(0.13)/13(0.28)	12(0.13)/13(0.28)	12(0.13)/13(0.28)	5,93	5.9
D21S11	30.2(0.03)/32.2(0.09)	29(0.21)/30(0.24)	29(0.21)/32.2(0.09)/30.2(0.03)/30(0.24)	10,08	10.1
D2S441	11.3(0.06)/14(0.26)	13(0.03)/14(0.26)	11.3(0.06)/13(0.03)/14(0.26)	49,66	49.7
D19S433	13(0.25)/15(0.14)	11(0.02)/14(0.28)	11(0.02)/13(0.25)/14(0.28)/15(0.14)	74,12	74.1
TH01	8(0.15)/9.3(0.22)	8(0.15)/9(0.16)	8(0.15)/9(0.16)/9.3(0.22)	6,86	6.9
FGA	20(0.11)/23(0.15)	21(0.15)/22(0.16)	20(0.11)/21(0.15)/22(0.16)/23(0.15)	21,43	21.4
D22S1045	15(0.35)	16(0.34)	15(0.35)/16(0.34)	2,83	2.9
D5S818	11(0.33)	12(0.35)/13(0.18)	11(0.33)/12(0.35)/13(0.18)	8,14	8.1
D13S317	12(0.30)/13(0.13)	11(0.30)/14(0.05)	11(0.30)/13(0.13)/12(0.30)/14(0.05)	34,49	34.5
D7S820	12(0.16)/13(0.03)	9(0.12)/12(0.16)	9(0.12)/12(0.16)/13(0.03)	16,91	16.9
D10S1248	13(0.27)/16(0.11)	16(0.11)/16(0.11)	13(0.27)/16(0.11)	6,74	6.7
D12S391	18(0.20)/16(0.04)	22(0.09)/22(0.09)	16(0.04)/18(0.20)/22(0.09)	20,50	20.5

Tabela 14- Cálculos LR Misturas ($\Theta=0,01$)

Locus	Vitima Alelo(Frequência)	Suspeito Alelo(Frequência)	Mistura Alelo(Frequência)	DNAstr.com	DNAmix
D3S1358	16(0.27)/17(0.20)	15(0.30)/18(0.12)	15(0.30)/17(0.20)/16(0.27)/18(0.12)	13,63	13,6
D16S539	12(0.25)	13(0.15)/14(0.02)	12(0.25)/14(0.02)/13(0.15)	98,96	99
CSF1PO	10(0.27)/12(0.32)	11(0.29)/15(0.01)	10(0.27)/12(0.32)/11(0.29)/15(0.01)	89,35	89,4
TPOX	9(0.12)/11(0.28)	8(0.45)	9(0.12)/11(0.28)/8(0.45)	1,72	1,7
D8S1179	12(0.13)/13(0.28)	12(0.13)/13(0.28)	12(0.13)/13(0.28)	5,25	5,3
D21S11	30.2(0.03)/32.2(0.09)	29(0.21)/30(0.24)	29(0.21)/32.2(0.09)/30.2(0.03)/30(0.24)	10,08	10,1
D2S441	11.3(0.06)/14(0.26)	13(0.03)/14(0.26)	11.3(0.06)/13(0.03)/14(0.26)	36,27	36,3
D19S433	13(0.25)/15(0.14)	11(0.02)/14(0.28)	11(0.02)/13(0.25)/14(0.28)/15(0.14)	55,02	55,0
TH01	8(0.15)/9.3(0.22)	8(0.15)/9(0.16)	8(0.15)/9(0.16)/9.3(0.22)	6,47	6,5
FGA	20(0.11)/23(0.15)	21(0.15)/22(0.16)	20(0.11)/21(0.15)/22(0.16)/23(0.15)	20,61	20,6
D22S1045	15(0.35)	16(0.34)	15(0.35)/16(0.34)	2,79	2,8
D5S818	11(0.33)	12(0.35)/13(0.18)	11(0.33)/12(0.35)/13(0.18)	8,18	8,2
D13S317	12(0.30)/13(0.13)	11(0.30)/14(0.05)	11(0.30)/13(0.13)/12(0.30)/14(0.05)	30,22	30,2
D7S820	12(0.16)/13(0.03)	9(0.12)/12(0.16)	9(0.12)/12(0.16)/13(0.03)	14,66	14,7
D10S1248	13(0.27)/16(0.11)	16(0.11)/16(0.11)	13(0.27)/16(0.11)	5,89	5,9
D12S391	18(0.20)/16(0.04)	22(0.09)/22(0.09)	16(0.04)/18(0.20)/22(0.09)	16,15	16,2

Tabela 15 - Cálculos LR Misturas ($\Theta=0,03$)

Locus	Vitima Alelo(Frequência)	Suspeito Alelo(Frequência)	Mistura Alelo(Frequência)	DNAstr.com	DNAmix
D3S1358	16(0.27)/17(0.20)	15(0.30)/18(0.12)	15(0.30)/17(0.20)/16(0.27)/18(0.12)	13,07	13,1
D16S539	12(0.25)	13(0.15)/14(0.02)	12(0.25)/14(0.02)/13(0.15)	64,4	64,4
CSF1PO	10(0.27)/12(0.32)	11(0.29)/15(0.01)	10(0.27)/12(0.32)/11(0.29)/15(0.01)	48,76	48,8
TPOX	9(0.12)/11(0.28)	8(0.45)	9(0.12)/11(0.28)/8(0.45)	1,69	1,7
D8S1179	12(0.13)/13(0.28)	12(0.13)/13(0.28)	12(0.13)/13(0.28)	4,29	4,3
D21S11	30.2(0.03)/32.2(0.09)	29(0.21)/30(0.24)	29(0.21)/32.2(0.09)/30.2(0.03)/30(0.24)	10,08	10,1
D2S441	11.3(0.06)/14(0.26)	13(0.03)/14(0.26)	11.3(0.06)/13(0.03)/14(0.26)	23,22	23,2

D19S433	13(0.25)/15(0.14)	11(0.02)/14(0.28)	11(0.02)/13(0.25)/14(0.28)/15(0.14)	37,84	37,8
TH01	8(0.15)/9.3(0.22)	8(0.15)/9(0.16)	8(0.15)/9(0.16)/9.3(0.22)	5,86	5,9
FGA	20(0.11)/23(0.15)	21(0.15)/22(0.16)	20(0.11)/21(0.15)/22(0.16)/23(0.15)	19,22	19,2
D22S1045	15(0.35)	16(0.34)	15(0.35)/16(0.34)	2,62	2,6
D5S818	11(0.33)	12(0.35)/13(0.18)	11(0.33)/12(0.35)/13(0.18)	8,26	8,3
D13S317	12(0.30)/13(0.13)	11(0.30)/14(0.05)	11(0.30)/13(0.13)/12(0.30)/14(0.05)	24,84	24,8
D7S820	12(0.16)/13(0.03)	9(0.12)/12(0.16)	9(0.12)/12(0.16)/13(0.03)	11,63	11,6
D10S1248	13(0.27)/16(0.11)	16(0.11)/16(0.11)	13(0.27)/16(0.11)	4,72	4,7
D12S391	18(0.20)/16(0.04)	22(0.09)/22(0.09)	16(0.04)/18(0.20)/22(0.09)	11,22	11,2

6. Discussão

A probabilidade combinada utilizada em laudos de paternidade e misturas requer que as frequências dos alelos nos loci sejam independentes, está independência é verificada através da análise da tabela de frequência utilizada. O trabalho realizado por Budowle e colaboradores (Budowle et al. 1994) demonstrou que ao ignorar o equilíbrio populacional em tabelas de frequências de marcadores de STRs, a probabilidade condicionada combinada nos locus fica afetada por um fator de até 10. A variação de outros fatores populacionais também podem implicar em mudanças na distribuição da razão de verossimilhança (LR), isto ficou demonstrado no trabalho realizado por Taylor e colaboradores (Taylor et al. 2014). Vários índices relacionados à genética de populações como: o PIC (Conteúdo Informativo de Polimorfismo); Heterozigosidade observada e esperada; Poder de Exclusão (PE); Índice de Paternidade Típico (TPI) e estatística F, podem ser utilizados para validar um banco de dados de frequências alélicas. Um laboratório que utiliza seus próprios dados deve verificar estes cálculos para garantir que seus dados não estão enviesados por desequilíbrios e aumentar a confiabilidade necessária no uso das frequências alélicas.

Nos laboratórios forenses é comum o uso da tabela Brasil de frequências alélicas (Aguilar et al. 2014) o que pode ser considerado uma boa prática, uma vez que esta tabela representa uma amostra significativa da população Brasileira. Para

o uso tabelas de frequências retirada de populações locais é importante realizar os cálculos dos índices estatísticos que podem ser obtidos através de diferentes softwares específicos de genética de populações ou através de planilhas do Excel criadas para este fim. O grande problema está na exportação/manipulação dos dados entre estes programas. Cada programa de genética de população requer diferentes formatos de arquivos para entrada/saída de dados, o que na prática dificulta a realização dos cálculos. Softwares utilizados pelos laboratórios e descritos na literatura, não possuem interfaces de exportação de dados nos diversos formatos adequados.

O DNAsr.com disponibiliza uma interface no qual o usuário pode avaliar todos estes índices, como demonstrado. A ferramenta permite a exportação das tabelas nos formatos genepop e alerquim. Isto possibilita ao usuário a verificação de outros cálculos não disponibilizados no software (e.g. o cálculo do pvalue exato para o EHW, que não é disponibilizado no software devido ao tempo de execução deste algoritmo na plataforma web).

O trabalho realizado por Clayton e colaboradores (Clayton et al. 1998) definiu seis regras que são utilizadas para confirmar possíveis alelos no eletroferograma baseado no tamanho do pico ou na sua área. Esta metodologia dá conta de alelos muito bem definidos, onde seus picos passam nos critérios de thresholds analíticos e estocásticos ou quando não se trata de artefatos como *Stutters*. A proporção dos contribuintes na amostra pode definir a quantidade de contribuintes da mistura. Os cálculos do PHR e Mx são cruciais nesta metodologia para avaliação dos verdadeiros genótipos presentes na amostra. Esta metodologia pode falhar quando a concentração de DNA é muito baixa (<100 pg) que é comum em cenas de crime. Nestas concentrações foram verificados o aparecimento em maior quantidade de

artefatos como drop-out ou drop-in. Para a solução deste problema novos modelos estatísticos estão sendo propostos como o modelo contínuo que implementa a distribuição normal para estimar a probabilidade dos verdadeiros alelos e que utiliza informações do tamanho ou área do alelo no eletroferograma, ou o uso do modelo semicontínuo, para definir a probabilidade de drop-in e drop-out e assim definir o cálculo do LR (Buckleton et al. 2004). Esta abordagem ainda não está presente na prática dos laboratórios forenses no Brasil.

A ferramenta DNAstr.com incorporou o modelo demonstrado por Cleyton e colaboradores. Esta ferramenta ajuda os peritos a definir qual a melhor combinação de genótipos para a mistura e assim comparar com os genótipos da vítima e suspeito. A ferramenta permite exportar casos forenses para dois softwares: DNAmix (Curran et al. 1999) e software R pelo pacote Forensim (Haned and Gill 2011). O DNAmix é um software bastante utilizado nos laboratórios de genética forense no Brasil. Ele implementa o modelo binário para o cálculo do valor de LR. O pacote Forensim implementa o modelo semicontínuo, no software estatístico R. Este pacote permite calcular o valor do LR levando em conta as probabilidades de drop-in e drop-out.

O teste de paternidade tem como objetivo revelar relações de parentesco na presença de duas teses conflituosas (chamadas de Tese de acusação e Tese de defesa). Na tese de acusação busca-se provar através dos genótipos dos indivíduos que existe uma relação parental. Na tese de defesa tenta-se provar que não existe tal relação. A maioria dos casos envolve a presença de três pessoas: Um filho, sua mãe e um suposto pai. Outras configurações são possíveis como por exemplo quando não se tem o genótipo da mãe, neste caso apenas o filho e suposto pai estão envolvidos.

O DNSstr.com disponibiliza duas formas de testar as hipóteses de paternidade nos casos de trio (mãe, filho e suposto pai) e duo (filho e suposto pai). Na primeira forma um *pipeline* de cinco passos: 1) cadastrar pessoas: neste passo são cadastradas as pessoas envolvidas no caso que se deseja resolver. 2) Cadastrar o caso: neste passo é armazenado o caso, data e tipo do caso. 3) Cadastrar as pessoas nos casos: Aqui você relaciona os papéis de cada pessoa no caso. 4) Cadastro da genotipagem: neste passo você pode inserir a genotipagem de cada pessoa, pode ser manualmente ou através de arquivos no formato tabular do GeneMapper ou XML do CODIS. 5) Gerar o laudo: a última etapa permite gerar uma tabelas com todos os alelos dos envolvidos, o valor do IP (índice de paternidade) por locus, o IPC (índice de paternidade acumulada), o valor da probabilidade de paternidade W e quantos locus indicam exclusão. O software DNAstr.com ainda disponibiliza uma interface mais simples com um editor gráfico para edição das relações, neste editor são necessários apenas três passos: 1) Definir os envolvidos. 2) Montar a relação. 3) Colocar os parâmetros do teste. Os resultado pode ser exportado para os formatos PDF, Excel e CSV.

Este formato de software tem o potencial de ser implementado na rotina de laboratórios de genética tanto forense. A possibilidade de armazenar e recuperar os dados de casos e a genotipagem dos envolvidos, permite simular facilmente novos cálculos envolvendo novas hipóteses, facilita o acompanhamento de todos os casos resolvidos e a serem resolvidos. Esta ferramenta adaptada para gestão permite à gestores e técnicos se concentrarem em outros aspectos importantes e complexos na análise de um caso como todo.

7. Conclusões

Vinte anos após a introdução de STRs (*Short Tandem Repeat*) para análise de sequencias de DNA encontradas em cenas de crimes e também para resolver casos de paternidade (Gill et al. 2015), vivemos a era da informação em massa com o rápido desenvolvimento de equipamentos de genotipagem. Este desenvolvimento traz consigo novos desafios como o uso de interfaces mais produtivas, o armazenamento dos dados mais confiável, interação com outras ferramentas e confiabilidade nos cálculos.

O software DNAstr.com é o primeiro software disponível gratuitamente na plataforma WEB e que é capaz de fazer cálculos na área civil (paternidade), criminal (misturas) e também na área de genética de populações aplicada aos bancos de dados forense. Esta ferramenta pode ser muito útil em laboratório de genética onde o uso de softwares é restrito a poucos devido à dificuldade de utilização.

Referências

- Aguiar, Vitor R. C., Amanda M. de Castro, Vanessa C. O. Almeida, Frederico S. V. Malta, Alessandro C. S. Ferreira, and Iúri D. Louro. (2014) *New CODIS Core Loci Allele Frequencies for 96,400 Brazilian Individuals*. *Forensic Science International: Genetics* 13:e6–12. Retrieved (<http://linkinghub.elsevier.com/retrieve/pii/S1872497314000921>).
- Balding, D. J. and R. A. Nichols. (1994) *DNA Profile Match Probability Calculation: How to Allow for Population Stratification, Relatedness, Database Selection and Single Bands*. *Forensic science international* 64:125–40.
- Barrai, I. and a. Rodriguez Larralde. (1992) *The Probability of Matching Genetic Markers in Different Samples*. *International Journal of Anthropology* 7:71–78.
- Botstein, D., R. L. White, M. Skolnick, and R. W. Davis. (1980) *Construction of a Genetic Linkage Map in Man Using Restriction Fragment Length Polymorphisms*. *American journal of human genetics* 32:314–31.
- Brenner, Charles and Jeffrey Morris. (1992) *Paternity Index Calculations in Single Locus Hypervariable Dna Probes : Validation and Other Studies*. Pp. 21–53 in *International Symposium on Human Identification*.
- Buckleton, John, Christopher Triggs, and Simon Walsh. (2004) *Forensic DNA Evidence Interpretation*. CRC Press.
- Buckleton, John, TriggsChristopher, and Simon Walsh. (2005) *Forensic DNA Evidence Interpretation*.
- Budowle, B., K. L. Monson, A. M. Giusti, and B. L. Brown. (1994) *The Assessment of Frequency Estimates of Hae III-Generated VNTR Profiles in Various Reference Databases*. *Journal of forensic sciences* 39:319–52.
- Butler, John M. (2002) *Forensic DNA Typing*. 2nd ed. Elsevier Academic Press.
- Clayton, T. M., J. P. Whitaker, R. Sparkes, and Peter. Gill. (1998) *Analysis and Interpretation of Mixed Forensic Stains Using DNA STR Profiling*. *Forensic Science International* 91:55–70.
- Connor, Kristen L. O., Steven P. Myers, Erica L. R. Butts, Carolyn R. Hill, John M. Butler, and Peter M. Vallone. (2010) *Candidate Reference Family Data : A Tool for Validating Kinship Analysis Software*. in *21st International Symposium on Human Identification Poster Abstracts*.
- Conover, W. (1980) *Pratical Nompametric Statistics*. 2nd ed. John Wiley.
- Cox, Michael, Jennifer Doudna, and Michael O´Donnel. (2012) *Biologia Molecular - Princípios E Técnicas*. edited by Artmed.
- Curran, J. M., C. M. Triggs, J. Buckleton, and B. S. Weir. (1999) *Interpreting DNA Mixtures in Structured Populations*. *Journal of forensic sciences* 44:987–95.
- Daniel, Hartil and Andrew Clark. (2007) *Pincipio de Genética de Populações*. artmed.
- Edwards, a. W. F. (2008) *G. H. Hardy (1908) and Hardy-Weinberg Equilibrium*.

Genetics 179:1143–50.

Excoffier, Laurent and Heidi E. L. Lischer. (2010) *Arlequin Suite Ver 3.5: A New Series of Programs to Perform Population Genetics Analyses under Linux and Windows*. *Molecular Ecology Resources* 10:564–67.

FALCONER, D. S. (1964) *Introduction of Quantitative Genetics*.

Fung, Wing K. and Yue-Qing Hu. (2002) *Evaluating Mixed Stains with Contributors of Different Ethnic Groups under the NRC-II Recommendation 4.1*. *Statistics in medicine* 21:3583–93.

Fung, Wing Kam and Yue-Qing HU. (2008) *Statistical DNA Forensics*. edited by C. John Wiley & Sons Ltd, The Atrium, Southern Gate.

Geisser, S. and W. Johnson. (1992) *Testing Hardy-Weinberg Equilibrium on Allelic Data from VNTR Loci*. *American journal of human genetics* 51:1084–88.

Gill, Peter, C. H. Brenner, J. S. Buckleton, A. Carracedo, M. Krawczak, W. R. Mayr, N. Morling, M. Prinz, P. M. Schneider, and B. S. Weir. (2006) *DNA Commission of the International Society of Forensic Genetics: Recommendations on the Interpretation of Mixtures*. *Forensic Science International* 160:90–101. Retrieved (<http://linkinghub.elsevier.com/retrieve/pii/S0379073806002659>).

Gill, Peter, Hinda Haned, Oyvind Bleka, Oskar Hansson, Guro Dørum, and Thore Egeland. (2015) *Genotyping and Interpretation of STR-DNA: Low-Template, Mixtures and Database matches—Twenty Years of Research and Development*. *Forensic Science International: Genetics* 18:100–117. Retrieved October 2, 2015 (<http://www.sciencedirect.com/science/article/pii/S1872497315000654>).

Gill, Peter, R. Sparkes, R. Pinchin, T. Clayton, J. Whitaker, and J. Buckleton. (1998) *Interpreting Simple STR Mixtures Using Allele Peak Areas*. *Forensic Science International* 91:41–53. Retrieved ([http://dx.doi.org/10.1016/S0379-0738\(97\)00174-6](http://dx.doi.org/10.1016/S0379-0738(97)00174-6)).

Guo, Sun Wei and Thompson Elizabeth a. (1992) *Guo.Thom1992.Biometrics.HWTest.pdf*. *Biometrics*.

Haned, Hinda and Peter Gill. (2011) *Analysis of Complex DNA Mixtures Using the Forensim Package*. *Forensic Science International: Genetics Supplement Series* 3:e79–80. Retrieved January 11, 2016 (<http://www.sciencedirect.com/science/article/pii/S1875176811000400>).

Hartl, Daniel L. and Andrew G. Clark. (2007) *Principles of Population Genetics*. 4th ed. artmed.

Hill, Carolyn R., David L. Duewer, Margaret C. Kline, Michael D. Coble, and John M. Butler. (2013) *U.S. Population Data for 29 Autosomal STR Loci*. *Forensic science international. Genetics* 7:e82–83. Retrieved (<http://www.sciencedirect.com/science/article/pii/S1872497312002712>).

Jeffreys, A., V. Wilson, and S. Thein. (1985) *Individual-Specific “fingerprints” of Human DNA*. *Nature* 316:76–79.

Jones, D. a. (1972) *Blood Samples: Probability of Discrimination*. *Journal - Forensic Science Society* 12:355–59.

Kidd, Kenneth K., Judith R. Kidd, William C. Speed, Rixun Fang, Manohar R. Furtado, F. C. L. Hyland, and Andrew J. Pakstis. (2012) *Expanding Data and*

Resources for Forensic Use of SNPs in Individual Identification. Forensic Science International: Genetics 6:646–52. Retrieved (<http://dx.doi.org/10.1016/j.fsigen.2012.02.012>).

Kidd, Kenneth K., Andrew J. Pakstis, William C. Speed, Elena L. Grigorenko, Sylvester L. B. Kajuna, Nganyirwa J. Karoma, Selemanni Kungulilo, Jong Jin Kim, Ru Band Lu, Adekunle Odunsi, et al. (2006) *Developing a SNP Panel for Forensic Identification of Individuals. Forensic Science International* 164:20–32.

Kling, Daniel, Andreas O. Tillmar, and Thore Egeland. (2014) *Familias 3 - Extensions and New Functionality. Forensic science international. Genetics* 13:121–27. Retrieved March 6, 2015 (<http://www.scopus.com/inward/record.url?eid=2-s2.0-84907334835&partnerID=tZOtx3y1>).

Lander, E. S., L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, et al. (2001) *Initial Sequencing and Analysis of the Human Genome. Nature* 409:860–921.

Nagy, Sándor, Péter Poczai, Istvan Cernák, Ahmad Gorji Mousapour, Geza Hegedus, and János Taller. (2012) *PICcalc : An Online Program to Calculate Polymorphic Information Content for Molecular Genetic Studies. 670–72.*

Nagylaki, Thomas. (1998) *Fixation Indices in Subdivided Populations. Genetics* 148:1325–32.

NRC-II. (1996) *The Evaluation of Forensic DNA Evidence. NATIONAL ACADEMY PRESS.*

O'Connor, Kristen L., Steven P. Myers, Erica L. R. Butts, Carolyn R. Hill, John M. Butler, and Peter M. Vallone. (2010) *Candidate Reference Family Data: A Tool for Validating Kinship Analysis Software. in 21st International Symposium on Human Identification Poster Abstracts.*

Purcell, Shaun, Benjamin Neale, Kathe Todd-Brown, Lori Thomas, Manuel a R. Ferreira, David Bender, Julian Maller, Pamela Sklar, Paul I. W. de Bakker, Mark J. Daly, et al. (2007) *PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. American journal of human genetics* 81:559–75.

Rousset, François. (2008) *GENEPOP'007: A Complete Re-Implementation of the GENEPOP Software for Windows and Linux. Molecular Ecology Resources* 8:103–6.

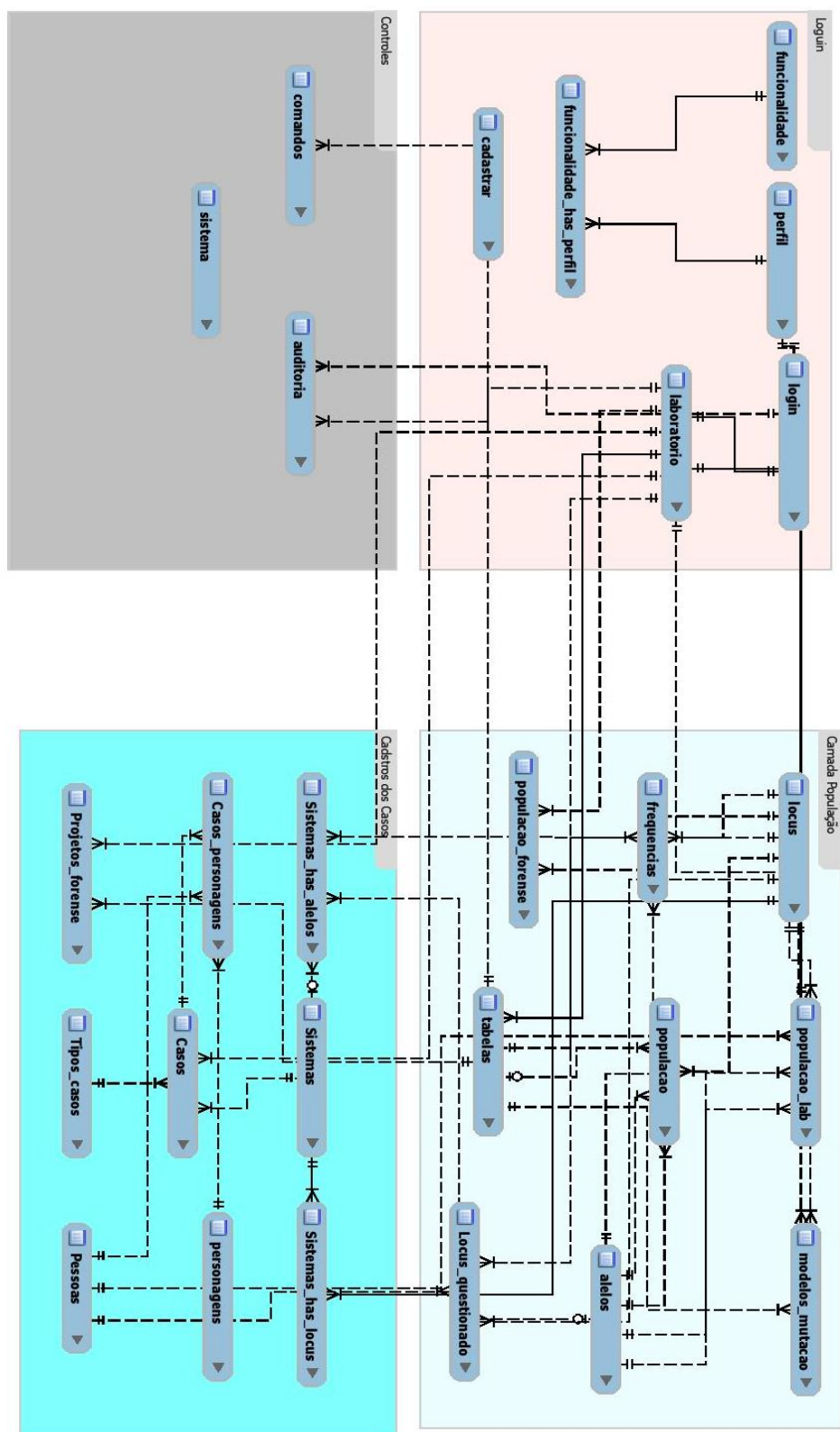
Taylor, D., J. a. Bright, J. Buckleton, and J. Curran. (2014) *An Illustration of the Effect of Various Sources of Uncertainty on DNA Likelihood Ratio Calculations. Forensic Science International: Genetics* 11:56–63. Retrieved (<http://dx.doi.org/10.1016/j.fsigen.2014.02.003>).

WEIR, B. S. (1996) *Genetic Data Analysis II: Methods for Discrete Poopulation Genetics. 2.ed.*

Weir, Bruce S. (2004) *Matching and Partially-Matching DNA Profiles. Journal of forensic sciences* 49:1009–14.

Wright, Sewall. (1965) *The Interpretation of Population Structure by F-Statistics with Special Regard to Systems of Mating. Evolution* 19:395–420. Retrieved (<http://www.jstor.org/stable/2406450>).

Anexo 1 – Diagrama do Banco de Dados



Anexo 2 – Formato dos Arquivos

Inserção de Tabelas

Formato 1: Tabela de Frequências

Inserção de Tabela de Frequências. As frequências alélicas são dados por locus onde cada linha equivale a um alelo, cada coluna equivale a um locus. As frequências são colocadas nas células. Primeira Linha: Nomes dos Locus. Demais Linhas : Primeira coluna o nome do alelo, demais colunas a frequência por locus. Separador: Ponto e vírgula.

1	Alleles	D10S1248	D12S391	D16S539	D18S51
2	4				
3	5			0.000498	
4	6			0.000202	
5	6.3				
6	7			0.000088	
7	7.3				

Formato 2: Tabela de Genótipos

Primeira linha : Locus ; Nome dos loci (um por coluna). Segunda linha : Alelos ; Todos os alelos presentes no locus (por coluna). Demais Linhas: ID do indivíduo; Alelos por indivíduos (cada coluna com os alelos do locus correspondente). Separador de colunas: Ponto e Vírgula

1	Locus	D8S1179	D21S11	D7S820
2	Alelos	14/13/15/12/10/16	30.2/28/31.2/29/30/33.2/31	12/11/8/10/9
3	1	10/14	28/30.2	11/12
4	2	13/14	28/31.2	11/12
5	3	13/13	28/29	8/10
6	4	14/15	29/30	8/8
7	5	12/14	30/30.2	11/11

Formato 3: Tabela de Genótipos

Primeira linha : SampleName ; Nome dos loci (um por coluna). Repetir cada loci nas colunas (um para cada alelo). Demais linha : Id do individuo ; Um alelo por coluna para cada locus. Separador de colunas: Ponto e Virgula.

SampleCode	AMEL	AMEL	D10S1248	D10S1248	D12S391	D12S391
BC11352	X	Y	13	15	17	18
GA05070	X	Y	12	14	17,3	23
GA05071	X	Y	14	16	19	20
GC03394	X	Y	15	16	19	19
GT36864	X	Y	15	17	17,3	20
GT36866	X	Y	13	16	18	20

Formato Inserção de genotipagem

A primeira linha deverá conter os seguintes labels : id;nome;sexo;locus1;locus2...,locusN. As outras linhas deverão conter um individuo cada com as seguintes colunas: id; nome; sexo; alelo_locus1_1/alelo2_locus1_2; alelo_locusN_1/alelo2_locusN_2 Onde: id: identificação do individuo (cpf ou outra) nome: nome da pessoa. Sexo: M=masculina e F=Feminino. Locus1..LocusN = Nome de cada locus. alelo1_locus1/alelo2_locus1 = Os alelos de cada locus

id	nome	sexo	D8S1179	D21S11	D7S820	CSF1PO	
1	1	F	10/14	28/30.2	11/12	10/11	:
2	2	M	13/14	28/31.2	11/12	10/12	:
3	3	F	13/13	28/29	8/10	9/11	:
4	4	F	14/15	29/30	8/8	10/10	:

Anexo 3 – Curriculum Vitae

Sérgio de Sá Leitão Paiva Júnior

Curriculum Vitae

Dados pessoais

Nome Sérgio de Sá Leitão Paiva Júnior
Filiação Maria Auxiliadora Brito
Nascimento 13/01/1974 - Joao Pessoa/PB - Brasil
Carteira de Identidade 4255683 ssp - PE - 06/02/1998
CPF 908.213.194-34

Formação acadêmica/titulação

2004 - 2006 Mestrado em Biometria e Estatística Aplicada.
 Universidade Federal Rural de Pernambuco, UFRPE, Recife, Brasil
 Título: Detecção de vasos sanguíneos em imagens de fundo de olho, Ano de obtenção: 2006
 Orientador: Borko Stosic
 Bolsista do(a): Coordenação de Aperfeiçoamento de Pessoal de Nível Superior

2000 - 2004 Graduação em Lic. Computação.
 Universidade Federal Rural de Pernambuco, UFRPE, Recife, Brasil

1990 - 1994 Ensino Médio (2o grau) .
 Instituto Federal de Pernambuco, IFPE, Recife, Brasil, Ano de obtenção: 1994

Formação complementar

2015 - 2015 Curso de curta duração em Validação Interna para Interpretação de Perfis Genéticos de Misturas. (Carga horária: 40h).
 Academia Integrada de Segurança Pública, ACADPOL, Brasil

Produção

Produção bibliográfica

Artigos completos publicados em periódicos

1. LIMA, RICARDO ALEXANDRE C., ALMEIDA, LUCIANO F., LYRA, WELLINGTON S., SIQUEIRA, LUCAS A., GAIÃO, EDVALDO N., **PAIVA JUNIOR, SÉRGIO S.L.**, LIMA, RAFAELA L.F.C. Digital movie-based on automatic titrations. Talanta (Oxford). , v.147, p.226 - 232, 2016.
2. DE LUCENA, RODRIGO MENDONÇA, ELSZTEIN, CAROLINA, BARROS DE SOUZA, RAFAEL,

DE BARROS PITA, WILL, **PAIVA JR., S'RGIO DE S' LEIT'O**, DE MORAIS JR., MARCOS ANTONIO Genetic Interaction between *HOG1* and *SLT2* Genes in Signalling the Cellular Stress Caused by Sulphuric Acid in *Saccharomyces cerevisiae*. Journal of Molecular Microbiology and Biotechnology. , v.25, p.423 - 427, 2016.

3. CHAGAS, BÁRBARA SIMAS, COMAR, MANOLA, GURGEL, ANA PAVLA ALMEIDA DINIZ, **PAIVA, SÉRGIO**, SERACENI, SILVA, DE FREITAS, ANTONIO CARLOS, CROVELLA, SERGIO Association Study between Cervical Lesions and Single or Multiple Vaccine-Target and Non-Vaccine Target Human Papillomavirus (HPV) Types in Women from Northeastern Brazil. Plos One. , v.10, p.e0132570 - , 2015.

4. OLIVEIRA, Wagner Luís Mendes de, Mangueira EVC, VILELA, Marinalda Anselmo, **PAIVA, SÉRGIO**, LEAL, N.C., ALMEIDA, Alzira de Diversity of SCCmec Types in Methicillin Resistant Staphylococcus spp. Causing Hospital-Associated Infections.. Jacobs Journal of Microbiology and Pathology. , v.2, p.20 - , 2015.

5. DE LUCENA, RODRIGO MENDONÇA, ELSZTEIN, CAROLINA, DE BARROS PITA, WILL, DE SOUZA, RAFAEL BARROS, **DE SÁ LEITÃO PAIVA JÚNIOR, SÉRGIO**, DE MORAIS JUNIOR, MARCOS ANTONIO Transcriptomic response of *Saccharomyces cerevisiae* for its adaptation to sulphuric acid-induced stress. Antonie van Leeuwenhoek (Gedrukt). , v.11, p.1 - 14, 2015.

6. AMARAL, CAROLINA MM, CETKOVSKÁ, KATERINA, GURGEL, ANA PAD, CARDOSO, MARCUS V, CHAGAS, BÁRBARA S, **PAIVA JÚNIOR, SÉRGIO SL**, DE LIMA, RITA, SILVA-NETO, JACINTO C, SILVA, LUIZ AF, MUNIZ, MARIA TC, BALBINO, VALDIR Q, FREITAS, ANTONIO C MDM2 polymorphism associated with the development of cervical lesions in women infected with Human papillomavirus and using of oral contraceptives. Infectious Agents and Cancer. , v.9, p.24 - , 2014.

Trabalhos publicados em anais de eventos (completo)

1. Ronnie E.S. Santos, Cleyton V. C. Magalhaes, Jorge Correia Neto, **PAIVA JÚNIOR, S. S. L.** PROGLIB: uma linguagem de programação baseada na escrita de LIBRAS In: 22 simpósio brasileiro de informática na educação, 2011, aracaju.
22 simpósio brasileiro de informática na educação. , 2011.

Trabalhos publicados em anais de eventos (resumo)

1. **PAIVA JÚNIOR, S. S. L.**, BALBINO, V. Q. Caracterização in silico da proteína YLR194CP durante a fermentação alcoólica por *Saccharomyces cerevisiae* In: 2 jornada de pós-graduação em genética, 2012, recife.
2 jornada de pós-graduação em genética. , 2012.

2. Pedro Lemos A. Junior, Jose N. Cordeiro, **PAIVA JÚNIOR, S. S. L.**, gaiao Desenvolvimento de um calorímetro microcontrolado, portátil e de baixo custo In: 16 encontronacional de química analítica, 2011, campos do Jordão.
16 encontro nacional de química analítica. , 2011.

3. Ricardo Lima, **PAIVA JÚNIOR, S. S. L.**, Rafaela L. F. C. Lima Determinação da Acidez de Óleos Vegetais usando um Analisador Fluxo-Batelada com Detecção por Filmagem Digital. In: 34 reunião nacional de química analítica, 2011, Florianópolis.
anais da 34 reunião nacional de química analítica. , 2011.

4. gaiao, **PAIVA JÚNIOR, S. S. L.**, Joseane Jales da Nobrega Uso de imagem digital como ferramenta de auxílio na aprendizagem de química por alunos com deficiência visual. In: 34 reunião nacional de química analítica, 2011, florianópolis.
Anais do 34 reunião nacional de química analítica. , 2011.

Apresentação de trabalho e palestra

1. PAIVA JÚNIOR, SÉRGIO SL

Considerações para o uso de softwares em Laboratório de Genética Forense, 2015. (Conferência ou palestra, Apresentação de Trabalho)

2. PAIVA JÚNIOR, S. S. L.

Tópicos de Computação com Aplicação em Química, 2011. (Conferência ou palestra, Apresentação de Trabalho)

Patentes e registros

Patente

A Confirmação do status de um pedido de patentes poderá ser solicitada à Diretoria de Patentes (DIRPA) por meio de uma Certidão de atos relativos aos processos1. gaiao, **PAIVA JÚNIOR, S. S. L.**

Analizador portátil para determinação de umidade em gás através de imagens digitais, 2013. Categoria: Produto. Instituição onde foi depositada: INPI - Instituto Nacional da Propriedade Industrial. País: Brasil. Natureza: Patente de Invenção. Número do registro: BR1020130324124. Data de depósito: 19/12/2013. Depositante/Titular: Edvaldo da Nobrega Gaião, Carlos Zurita. Depositante/Titular: Universidade Federal Rural de Pernambuco.

Orientações e Supervisões

Orientações e supervisões

Orientações e supervisões concluídas

Trabalhos de conclusão de curso de graduação

1. Yulianne Maria Siqueira Bezerra. **Análise da Interações no processo de aprendizagem colaborativa mediada por redes sociais educacionais com alunos do ensino técnico e superior**. 2012. Curso (Sistema de Informação) - Universidade Federal Rural de Pernambuco