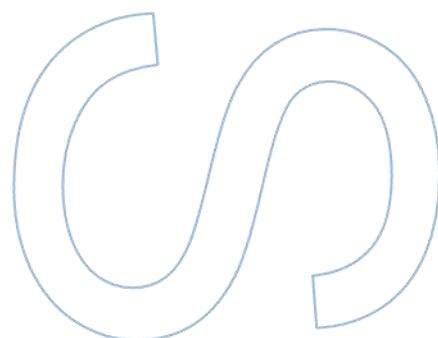
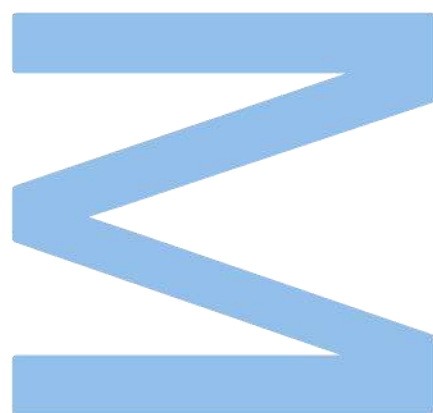


Marcadores genéticos STRs: Uma abordagem multipopulacional da diversidade e estrutura genética

Liana Patrícia Rocha Silva
Mestrado em Genética Forense
Departamento de Biologia
Faculdade de Ciências da Universidade do Porto
2025



Marcadores genéticos STRs: Uma abordagem multipopulacional da diversidade e estrutura genética

Liana Patrícia Rocha Silva

Dissertação realizada no âmbito do Mestrado em Genética Forense

Departamento de Biologia
2025

Orientadora

Luísa Azevedo, Investigadora Principal, Instituto de Ciências Biomédicas Abel Salazar (ICBAS-UP) e Unidade Multidisciplinar de Investigação Biomédica (UMIB)

Agradecimentos

À minha orientadora Luísa Azevedo, agradeço de coração por ter acreditado em mim em um momento tão incerto. Mesmo com todas as dificuldades e o fechamento do meu curso, aceitou ser minha orientadora e também agradeço pela paciência e dedicação. Sua confiança, compreensão e apoio fizeram toda a diferença.

Ao meu namorado, obrigada por estar sempre ao meu lado, por me incentivar a continuar mesmo nos dias mais difíceis e por celebrar comigo cada pequena conquista. À minha família, que foi meu porto seguro e não deixou que eu desistisse, devo uma gratidão imensa.

As minhas melhores amigas, que mesmo de longe não deixaram de me apoiar, agradeço pelo carinho e cuidado, pelas palavras de incentivo e por tornarem esse caminho mais leve, com risadas e companheirismo.

Agradeço a todos que mesmo sem perceber me encorajaram a persistir. Este trabalho não seria possível sem vocês.

Resumo

O uso de marcadores genéticos conhecidos como STRs (do inglês *Short Tandem Repeats*) transformou-se em uma ferramenta central na genética forense, permitindo a identificação individual a partir de amostras biológicas recolhidas de diferentes cenários. O sucesso na sua aplicabilidade resulta da heterogeneidade associada ao número de repetições entre indivíduos, que uma vez combinados para diferentes STRs, gera perfis genéticos distintivos e confiáveis.

Neste estudo, realizou-se uma análise multipopulacional baseada em uma revisão da literatura com o objetivo principal de descrever a diversidade e estrutura populacional de STRs autossómicos, STRs do cromossoma X (X-STRs) e STRs do cromossoma Y (Y-STRs). A análise focou em indicadores essenciais, como frequências alélicas, equilíbrio de Hardy-Weinberg (HWE), conteúdo de informação polimórfica (PIC), poder de discriminação (PD) e diversidade haplotípica (HD).

Os resultados indicam que os painéis de STRs dos *kits* Investigator *Argus X-12* (X-STRs) e do *kit PowerPlex Y23* (Y-STRs), apresentam alto grau de polimorfismo e alto poder discriminativo. A análise dos X-STRs apresentou um Poder Combinado de Discriminação (CPD) superior a 0,9999999983 enquanto os Y-STRs mostraram maior variação entre populações, mantendo altos níveis de poder de discriminação em grupos asiáticos, mas revelando valores mais reduzidos em comunidades geneticamente isoladas, como os Yakuts. Estes dados mostram que os X-STRs oferecem um desempenho uniformemente elevado, enquanto os Y-STRs são mais sensíveis à história demográfica e à subestrutura populacional.

Os dados analisados demonstram que a variabilidade genética é fortemente influenciada pela história demográfica e geográfica de cada população. Também ressaltam a importância de utilizar dados populacionais específicos, garantindo resultados estatísticos precisos e confiáveis em genética forense. Por fim, este estudo evidencia a relevância dos STRs como ferramentas essenciais, enquanto aponta a necessidade de expandir pesquisas para populações menos estudadas.

Palavras-chave: STRs, Y-STR, X-STR, STRs autossómicos, Marcadores Genéticos, Genética Forense, Genética Populacional

Abstract

The use of genetic markers known as Short Tandem Repeats (STRs) has become a central tool in forensic genetics, enabling individual identification from biological samples collected in different scenarios. The success of their applicability derives from the heterogeneity associated with the number of repeat units among individuals, which, when combined across different STR loci, generates distinct and reliable genetic profiles.

In this study, a multipopulation analysis was conducted based on a literature review, with the primary objective of describing the diversity and population structure of autosomal STRs, X chromosome STRs (X-STRs), and Y chromosome STRs (Y-STRs). The analysis focused on essential indicators such as allele frequencies, Hardy–Weinberg equilibrium (HWE), polymorphic information content (PIC), power of discrimination (PD), and haplotype diversity (HD).

The results indicate that the STR panels of the Investigator Argus X-12 kit (X-STRs) and the PowerPlex Y23 kit (Y-STRs) exhibit a high degree of polymorphism and strong discriminatory power. The X-STR analysis revealed a Combined Power of Discrimination (CPD) exceeding 0.9999999983, whereas Y-STRs showed greater variation among populations, maintaining high levels of discriminatory power in Asian groups but exhibiting reduced values in genetically isolated communities such as the Yakuts. These findings demonstrate that X-STRs provide consistently high performance, while Y-STRs are more sensitive to demographic history and population substructure.

The analyzed data demonstrate that genetic variability is strongly influenced by the demographic and geographic history of each population. It also highlights the importance of using population-specific data to ensure accurate and reliable statistical results in forensic genetics. Finally, this study underscores the relevance of STRs as essential tools while emphasizing the need to expand research on understudied populations.

Keywords: STRs, Y-STR, X-STR, Autosomal STRs, Genetic Markers, Forensic Genetics, Population Genetics.

1. Índice	
Lista de Tabelas	vi
Lista de Figuras	vii
Lista de Abreviaturas/ Siglas e Acrónimos	viii
Introdução.....	1
1. Marcadores genéticos.....	1
1.1. Critérios de escolha de marcadores para aplicações forenses	3
2. Repetições curtas em tandem- <i>Short Tandem Repeats</i> (STRs).....	3
2.1. Estrutura e Nomenclatura.....	4
2.2. Mutação	7
2.3. Tipos de STRs: características e localização	8
STRs autossómicos	8
STRs do Cromossoma X (X-STRs).....	9
STRs do Cromossoma Y (Y-STRs).....	10
2.4. Técnicas de deteção e análise	12
PCR multiplex	12
Eletroforese capilar	12
2.5. <i>Kits</i> comerciais	13
Investigator Argus X-12	14
<i>Kit</i> PowerPlex Y23	14
2.6. Banco de dados STRs.....	15
3. Genética populacional	15
3.1. Frequência alélica	15
3.2. Princípio de Hardy-Weinberg.....	16
3.3. Diversidade e estrutura populacional.....	17
4. Genética forense	18
4.1. Evolução metodológica e integração de marcadores.....	18
5. Objetivos.....	20
6. Metodologia	20

7.	Caracterização dos estudos incluídos.....	22
7.1.	STRs autossómicos.....	22
7.2.	X-STRs.....	22
7.3.	Y-STRs.....	22
	Limitações metodológicas.....	24
8.	Marcadores X-STR.....	24
8.1.	Populações e Amostragem.....	25
8.2.	Metodologias usadas pelos artigos.....	26
8.3.	Análise estatística	27
8.3.1.	Equilíbrio de Hardy Weinberg e desequilíbrio de ligação	27
8.4.	Parâmetros forenses	27
8.4.1.	Conteúdo de Informação de Polimorfismo (PIC).....	27
8.4.2.	Poder de Discriminação (PD) e Poder Combinado de Discriminação (CPD)	28
8.4.3.	Diversidade Haplotípicas (HD).....	30
8.4.4.	Análise de Estrutura Populacional e diferenciação populacional.....	31
8.5.	Conclusão	31
9.	Marcadores Y-STR.....	32
9.1.	População e amostragem.....	32
9.2.	Metodologias usada nos artigos	34
9.3.	Análise estatística dos Y-STRs	34
9.3.1.	Diversidade genética (GD).....	34
9.3.2.	Diversidade de Haplótipos (HD) e Capacidade Discriminatória (DC) ...	36
9.4.	Conclusão	37
10.	Conclusões	38
11.	Referências Bibliográficas	39
12.	Anexos	47

Lista de Tabelas

Tabela 1. Tipos de marcadores genéticos	2
Tabela 2. Caracterização dos estudos e a sua aplicação	23
Tabela 3. Amostras das populações estudadas; NR: não relacionados; AN: anónimos.	26
Tabela 4. Metodologia utilizada nos estudos analisados.....	26
Tabela 5. Valores de PIC das populações analisadas com 12 X-STRs.	28
Tabela 6. Amostras das populações analisadas com 23 Y-STRs	33
Tabela 7. Metodologia usada para amplificação e genotipagem dos Y-STRs	34
Tabela 8. <i>Locus</i> mais e menos polimórfico das populações analisadas.....	35
Tabela 9. Diversidade haplotípica (HD), capacidade de discriminação (DC) e número de haplótipos observados em populações analisadas por Y-STR.....	36

Lista de Figuras

Figura 1. Imagem ilustrativa de STRs com unidade repetitiva GATA	6
Figura 2. Ilustração esquemática de deslizamento da fita durante a replicação	8
Figura 3. Esquema ilustrativo da metodologia utilizada neste trabalho	21
Figura 4. Ilustração geográfica das populações estudadas na análise dos marcadores X-STRs.....	25
Figura 5. Poder de discriminação dos 12X-STRs populações analisadas. A) Poder de discriminação dos 12X-STRs em amostras de mulheres das populações analisadas; B) Poder de discriminação de LG1-LG4 em mulheres da Eritreia e Noruega; C) Poder de discriminação dos 12X-STRs em homens das populações analisadas; D) Poder de discriminação de LG1-LG4 em homens da Eritreia e Noruega.	29
Figura 6. Diversidade haplotípica dos 12 X-STRs das populações analisadas	30
Figura 7. Ilustração geográfica das populações estudadas na análise dos marcadores do Y-STRs	33
Figura 8. Diversidade genética das populações analisadas no estudo de marcadores do Y-STRs.....	35

Lista de Abreviaturas/ Siglas e Acrónimos

AMOVA	Análise De Variância Molecular (<i>Analysis Of Molecular Variance</i>)
CPD	Poder Combinado de Discriminação (<i>Combined Power Of Discrimination</i>)
CPE	Poder Combinado de Exclusão (<i>Combined Power Of Exclusion</i>)
CPI	Probabilidade Combinada de Inclusão (<i>Combined Probability Of Inclusion</i>)
CODIS	Sistema de Índice Combinado de DNA (<i>Combined DNA Index System</i>)
DC	Capacidade de Discriminação (<i>Discrimination Capacity</i>)
DG	Diversidade Genética (<i>Genetic Diversity</i>)
DH	Diversidade Haplotípica (<i>Haplotype Diversity</i>)
DNA	Ácido Desoxirribonucleico (<i>Deoxyribonucleic Acid</i>)
ESS	Conjunto Padrão Europeu (<i>European Standard Set</i>)
FCUP	Faculdade de Ciências da Universidade Do Porto
FIT	Coeficiente de Endogamia Total (<i>Overall Inbreeding Coefficient</i>)
FIS	Coeficiente de Endogamia Dentro das Populações (<i>Inbreeding Coefficient Within Populations</i>)
FST	Coeficiente de Diferenciação/Fixação (<i>Fixation Index</i>)
HET	Heterozigotia
HOM	Homozigotia
HWE	Equilíbrio de Hardy-Weinberg (<i>Hardy–Weinberg Equilibrium</i>)
INDEL	Inserção ou Deleção (<i>Insertion–Deletion</i>)
IP	Índice de Paternidade (<i>Paternity Index</i>)
ISFG	Sociedade Internacional de Genética Forense (<i>International Society For Forensic Genetics</i>)
MPS	Sequenciação Massiva em Paralelo (<i>Massive Parallel Sequencing</i>)
PE	Poder de Exclusão (<i>Power Of Exclusion</i>)

PCR	Reação em Cadeia da Polimerase (<i>Polymerase Chain Reaction</i>)
PDF	Poder de Discriminação Para Mulheres
PDM	Poder de Discriminação Para Homens
PIC	Conteúdo de Informação de Polimorfismos (<i>Polymorphic Information Content</i>)
RFLP	Polimorfismo de Comprimento de Fragmentos de Restrição (<i>Restriction Fragment Length Polymorphism</i>)
RM	Mutação Rápida
RST	Índice de Diferenciação Genética (<i>Genetic Differentiation Index</i>)
SM	Mutação Lenta
SNP	Polimorfismo de Nucleótido Único (<i>Single Nucleotide Polymorphism</i>)
STR	Repetição Curta em Tandem (<i>Short Tandem Repeat</i>)
UP	Universidade do Porto
VNTR	Repetição Em Tandem de Número Variável (<i>Variable Number Tandem Repeat</i>)
X-STR	STR do Cromossoma X
Y-STR	STR do Cromossoma Y

Introdução

1. Marcadores genéticos

O genoma humano é constituído por aproximadamente $3,08 \times 10^9$ de pares de bases, distribuídos em autossomas, cromossomas sexuais e em DNA mitocondrial, sendo que a maior parte não é codificante, e difere menos de 0.1% entre indivíduos [1]. É precisamente nesta pequena parte variável que se encontra a chave para a individualidade biológica e a compreensão da variabilidade populacional. No sentido da identificação dessa variação inter-individual são usados marcadores genéticos, que são sequências de DNA que variam em sequência ou número de repetições, ou seja, podem ocorrer por substituições de bases, ou por inserções ou deleções de conjunto de bases que levam a alterações no número de repetições de uma sequência específica de DNA. Os marcadores genéticos são amplamente utilizados na identificação biológica, definição de relações de parentesco, reconstrução de relações evolutivas e mapeamento de genes associadas a doenças [2].

Entre os marcadores genéticos mais utilizados estão os *Short Tandem Repeats* (STRs) descritas na tabela 1. Estes marcadores são repetições de pequenas sequências de DNA cujo número de repetições varia entre indivíduos e pode ser determinado a nível molecular via Reação em Cadeia da Polimerase (PCR) e eletroforese capilar [3]. A elevada heterozigotia faz com que valores do conteúdo informativo polimórfico (PIC) frequentemente excedam 0.5, caracterizando-os como altamente informativos dado o elevado poder discriminatório [4].

Um outro tipo de marcador usado em genética forense são os *Single Nucleotide Polymorphisms* (SNPs) que são marcadores bialélicos na maioria dos casos, que apresentam apenas duas variantes possíveis em uma dada posição cromossómica, tendo por isso um menor poder discriminatório (tabela 1) [5]. Ainda assim, a sua utilização é importante em estudos de genética forense e populacional, pois, aplicam-se na análise de DNA degradado ou altamente fragmentado. Em contextos forenses, esses dados podem orientar investigações iniciais quando não há correspondência direta nas bases de dados, fornecendo pistas sobre a provável origem populacional do doador. Além disso, em análises populacionais, permitem estimativas robustas da composição ancestral do indivíduo examinado [4].

Embora os dois tipos de marcadores referidos acima sejam os mais usados, existem outros marcadores que são importantes nos estudos moleculares em genética forense e populacional que também se encontram discriminados na tabela 1.

Tabela 1. Tipos de marcadores genéticos

Marcadores	Princípio	Vantagens	Desvantagens
Repetições em Tandem de Número Variável (VNTRs) [6]	Unidades de repetição que variam tipicamente de 10 a 100 pares de bases	Elevada diversidade alélica, elevada heterozigosidade observada em muitos <i>loci</i> e elevado poder de discriminação	Relativamente escassos no genoma e exigem grandes quantidades de DNA intacto
Repetições Curtas em Tandem (STRs) [7], [8]	Unidades de repetição tipicamente de 2 a 6 pares de bases	Elevadas taxas de mutação e facilidade de deteção por meio de amplificação baseada em PCR, Análise simultânea de muitos <i>loci</i> informativos a partir de quantidades mínimas de material biológico	Elevadas taxas de mutação por geração, o que resulta no surgimento de novas variantes alélicas
Polimorfismo de Nucleotídeo Único (SNP) [4]	Variação em uma única posição nucleotídica numa sequência de DNA	Facilidade na amplificação usando PCR multiplex, Baixa taxa de mutação (10^{-8})	Carácter bialélico sendo menos informativos É necessário um número elevado para atingir probabilidades combinadas equivalentes às obtidas com STRs [4]
Inserção ou Deleção (INDELs) [9]	Inserção ou deleção de uma pequena sequência de DNA	Amplificados em multiplex com fragmentos de PCR muito curto, úteis para amostras de DNA altamente degradado	Poder de discriminação relativamente baixo em comparação com os STRs

1.1. Critérios de escolha de marcadores para aplicações forenses

A seleção dos marcadores genéticos para aplicação forense depende de um conjunto complexo de fatores técnicos, estatísticos e populacionais, de forma a garantir que o perfil obtido seja altamente discriminatório e robusto no contexto específico da investigação.

- **Grau de polimorfismo**

Medido pelo conteúdo informativo polimórfico (PIC) e pela heterozigosidade observada (H_o). Marcadores com valores elevados nestes índices oferecem maior capacidade de distinguir indivíduos não aparentados, reduzindo a probabilidade de coincidências [10]. Essa propriedade é essencial quando se pretende alcançar probabilidades combinadas de discriminação (CPD) ou exclusão (CPE) próximas a 1, que é a condição fundamental para fortalecer a evidência em tribunal [10].

- **Representatividade populacional das frequências alélicas**

A estimativa das frequências deve fundamentar-se em amostras populacionais de dimensão adequada, preferencialmente com pelo menos 1000 indivíduos, a fim de assegurar robustez estatística, minimizando erros amostrais e identificar eventuais desequilíbrios de ligação (LD) entre *loci* altamente polimórficos [11].

- **Condições da amostra**

Muitos casos forenses envolvem material degradado ou em quantidade reduzida. Nesse cenário, marcadores com regiões alvo curtas têm vantagem técnica devido à maior eficiência na amplificação por PCR, o que explica a preferência por STRs curtos ou SNPs quando o DNA está degradado [12].

2. Repetições curtas em tandem-*Short Tandem Repeats* (STRs)

As repetições curtas em tandem (STR) ou microssatélites, têm geralmente 2 a 6 pares de bases que se repetem em tandem em locais específicos do genoma, ocupando cerca de 3% de todo o genoma [13,14,15].

São altamente polimórficos, pois há uma variação no número de repetições entre os indivíduos, tornando-se uma ferramenta valiosa para estudos forenses, como testes de paternidade e identificação humana [16,14]. Essa elevada variabilidade decorre, principalmente do fenómeno de deslizamento durante a replicação. Um mecanismo no qual a DNA polimerase pode adicionar ou remover repetições completas de forma relativamente frequente, aumentando a diversidade alélica na população [17].

As atuais aplicações práticas exploram esse elevado grau de variabilidade pois do ponto de vista estatístico, muitos *loci* STR frequentemente exibem valores elevados de PIC, frequentemente superiores a 0.6, o que indica elevado poder discriminatório [18].

Loci STR estão incorporados em sistemas *multiplex* comerciais que tipicamente analisam de 15 a 23 *loci* autossómicos em simultâneo e permitem cálculos robustos do CPD e do CPE. Em muitos *kits* validados esses parâmetros atingem valores próximos de 1, indicando uma baixíssima probabilidade teórica de perfis idênticos entre indivíduos não aparentados [17].

Além dos marcadores autossómicos clássicos presentes nas bases de dados como o CODIS (*Combined DNA Index System*) e reconhecidos internacionalmente, esses painéis podem incluir STRs localizados nos cromossomas sexuais. Os X-STR (STR associadas ao cromossoma X) oferecem vantagens específicas para a resolução de relações biológicas envolvendo hereditariedade paterna em filhas [18], enquanto os Y-STR (STR associadas ao cromossoma Y) são usados para rastrear linhagens masculinas diretas e resolver casos com predominância ou exclusividade de DNA masculino, como o caso de abuso sexual em que o agressor é do sexo masculino [19].

2.1. Estrutura e Nomenclatura

A estrutura molecular dos STRs consiste em unidades curtas de sequência nucleotídica repetidas em tandem. O motivo básico, como por exemplo a sequência GATA, define cada repetição, e o número de repetições determina o alelo observado num indivíduo, refletindo a contribuição haplóide herdada de cada progenitor [2].

A *International Society for Forensic Genetics* (ISFG) recomenda que o motivo repetitivo seja definido a partir dos primeiros nucleotídeos na direção 5' da cadeia direta no *GenBank*. Esta padronização evita ambiguidades relacionadas à orientação da sequência e assegura comparabilidade internacional dos resultados entre laboratórios [20].

A nomenclatura dos STRs segue convenções internacionais, como os sistemas CODIS e ESS (*European Standard Set*), nos quais os *loci* STR são organizados em painéis *multiplex* validados, contendo um conjunto fixo de marcadores centrais. Cada *locus* possui um identificador alfanumérico sistemático que reflete sua localização cromossômica. Por exemplo, no *locus* DXS10135:

- **D** refere-se ao DNA
- **X** representa o cromossoma **X**
- **S** representa STR
- **10135°** marcador identificado no cromossoma **X**

Analiticamente, cada alelo é numerado conforme o número integral de repetições detetadas no produto amplificado por PCR. O tamanho do fragmento é inferido por eletroforese capilar e calibrado com padrões internos. É importante notar que o tamanho aparente obtido pode diferir ligeiramente do real devido a propriedades físico-químicas do marcador e do corante fluorescente [20], exigindo atenção especial em comparações internacionais ou interlaboratoriais.

A designação alélica é tradicionalmente baseada na contagem do número total de repetições completas presentes em cada alelo [5]. Por exemplo, um alelo designado “4” em um *locus* com motivo tetranucleotídico GATA, indica que o motivo se encontra repetido 4 vezes consecutivas. O número de repetições desse motivo pode variar de pessoa para pessoa, como ilustrado na Figura 1 [21]. Em alguns casos, podem existir repetições parciais adicionais, indicadas por sufixos decimais (por exemplo, 12.2 = doze repetições completas + duas bases adicionais) [20]. Essa nomenclatura uniforme é essencial para assegurar consistência em bases de dados forenses e populacionais.

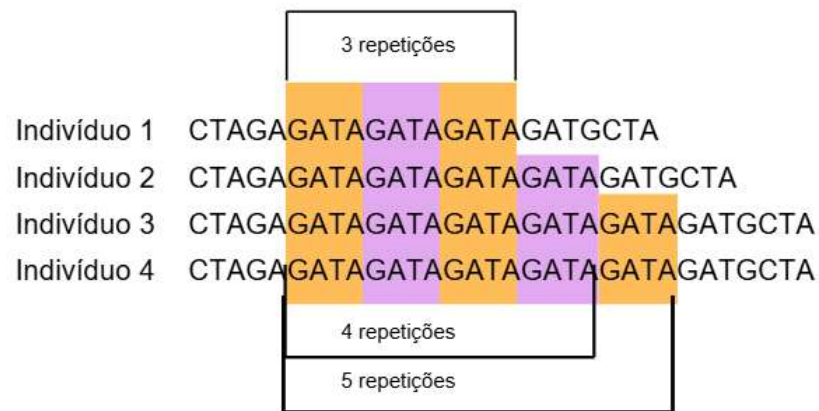


Figura 1. Imagem ilustrativa de STRs com unidade repetitiva GATA

As regiões flangeadoras também influenciam a nomenclatura, pois mutações nestas regiões podem alterar o tamanho aparente dos fragmentos obtidos por eletroforese capilar. Em certos sistemas comerciais, utilizam-se modificadores de mobilidade para ajustar a migração dos fragmentos, garantindo correspondência com as tabelas alélicas oficiais [20]. É importante, ainda, referir que mutações nas sequências onde hibridam os oligonucleotídeos iniciadores (*primers*) durante as reações de PCR podem gerar alelos nulos por falha de amplificação, fenómeno que deve ser considerado em análises forenses [20].

A estrutura interna dos STRs pode ser simples ou complexa.

- **STRs simples:** apresentam uma única unidade repetitiva contínua, tornando a contagem previsível e a interpretação eletroforética mais direta.
- **STRs complexos:** possuem interrupções, motivos imperfeitos ou blocos adjacentes distintos. Essa diversidade aumenta a heterozigotia e requer maior cuidado no desenho de *primers* para evitar amplificação inespecífica [22].

A variabilidade entre indivíduos decorre, principalmente, do número distinto de repetições do motivo central, mas também pode resultar de mutações intra repetitivas ou em regiões adjacentes, detetáveis por métodos de elevada resolução, como a

sequenciação massiva em paralelo (do inglês Massive Parallel Sequencing, MPS) [5]. Além disso, a estrutura repetitiva pode originar variantes atípicas, como os alelos trialélicos, decorrentes de duplicações intracromossômicas ou mosaicismos germinativos [20].

Do ponto de vista funcional, os STRs classificam-se segundo o tipo e tamanho dos motivos repetitivos: dinucleotídeos (2 nucleotídeos), trinucleotídeos (3 nucleotídeos), tetranucleotídeos (4 nucleotídeos), pentanucleotídeos (5 nucleotídeos) ou hexanucleotídeos (6 nucleotídeos). Em contexto forense, predominam os tetranucleotídeos, devido ao equilíbrio entre elevada variabilidade alélica e menor incidência de artefactos de amplificação, como os picos *stutter* [5].

A interpretação estatística baseia-se em bases de dados alélicos de referência e na independência entre os *loci* analisados, validada por estudos de desequilíbrio de ligação, permitindo estimativas robustas de coincidência [23, 24].

2.2. Mutação

Os STRs possuem elevadas taxa de mutação em relação a outros marcadores genéticos, situando-se na ordem de 10^{-3} , 10^{-4} por geração. Essa taxa é influenciada pela idade e sexo dos indivíduos, pelo aumento do número de repetições e pela unidade de repetição e sua composição. Apesar das elevadas taxas de mutação, estas não são uniformes entre os *loci*. Os STRs tetranucleotídeos apresentam uma taxa de mutação cerca de 50% inferior em relação aos dinucleotídeos em *loci* STR humanos não patogénicos, sendo que, em STR associados a doença, a taxa de mutação dos tetranucleotídeos é de quatro a sete vezes maior [15].

A variabilidade no número de cópias repetidas é devida ao mecanismo mutacional deslizamento da fita durante a replicação (*Strand-slippage replication*) [15, 20], o qual foi proposto por Kornberg et al. em 1964, diz que, durante a replicação do DNA ocorre um deslizamento entre a cadeia molde e a cadeia nova tal como mostra a Figura 2 [15]. Esse desalinhamento leva à inserção ou deleção de uma ou mais unidades de repetição no produto final da síntese. A consequência molecular imediata é uma variação no número total de repetições no *locus* afetado, originando os diferentes alelos detetáveis nas análises forenses.

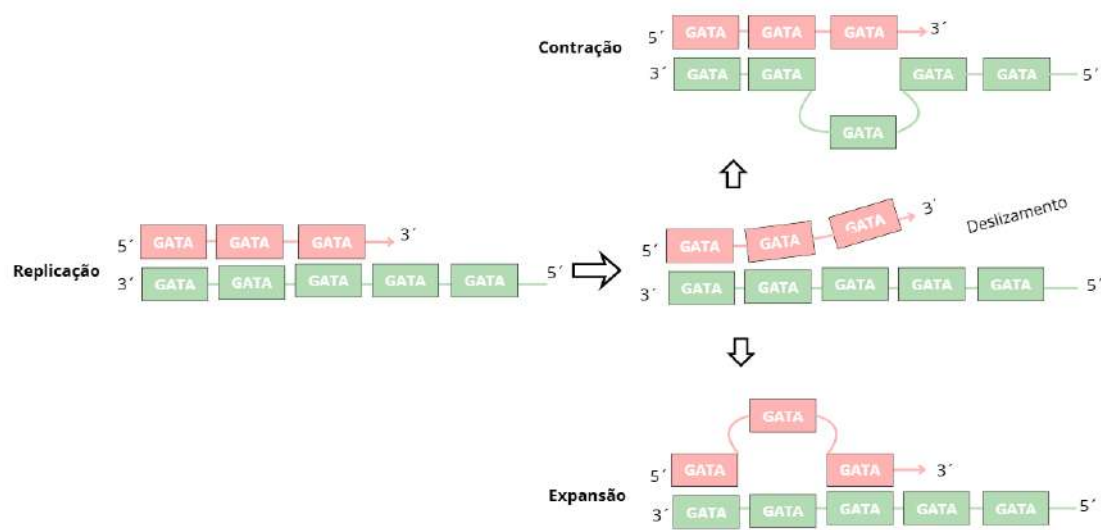


Figura 2. Ilustração esquemática de deslizamento da fita durante a replicação

2.3. Tipos de STRs: características e localização

STRs autossômicos

Os STRs autossômicos estão localizados nos cromossomos autossômicos, e cada indivíduo carrega dois alelos (um materno e um paterno) para cada *locus*, o que aumenta substancialmente a variabilidade a nível populacional e potencializa o poder discriminatório [25].

Apresenta uma elevada taxa de polimorfismo, o que significa que num determinado *locus*, indivíduos diferentes podem ter diferentes números de repetições, permitindo comparar diretamente perfis genéticos completos de indivíduos e calcular probabilidades combinadas de coincidência [17].

São amplamente aplicados em estudos forenses para identificar indivíduos, a partir de amostras biológicas, como cabelo ou sangue, em análises de paternidade e em investigação onde a comparação de perfis de DNA pode revelar relações biológicas. Para quantificar a eficácia dos STRs em testes de parentesco são utilizadas métricas como poder de exclusão (PE) e o índice de paternidade (IP) [20]. A aplicabilidade prática dos STRs autossômicos é favorecida pelo tamanho reduzido das regiões amplificadas por PCR, com elevada taxa de sucesso mesmo em DNA degradado ou fragmentado, com tamanho inferior a 400 pb [7].

A característica codominante destes marcadores assegura que ambos os alelos de um *locus* heterozigótico sejam detetados, garantindo clareza na distinção entre diferentes genótipos. Do ponto de vista molecular, a maior parte dos sistemas comerciais utilizados para análise forense de STRs autossómicos envolve motivos tetranucleótidos ou pentanucleótidos devido ao seu alto polimorfismo e variabilidade alélica [21, 26].

A estabilidade dos STRs autossómicos face a eventos mutacionais também influencia a sua utilização. As taxas médias de mutação dos STRs autossómicos situam-se entre 10^{-3} a 10^{-4} por geração, valores suficientemente elevados para manter diversidade significativa na população [20].

A combinação de múltiplos *loci* STR aumenta significativamente o poder discriminatório, e a escolha dos *loci* para um painel forense autossómico tem em consideração a frequência alélica esperada nas principais populações alvo. Para otimizar este aspeto, estudos populacionais prévios geram estimativas confiáveis usadas tanto nos cálculos individualizados quanto na composição das bases estatísticas internas dos laboratórios [27]. Consequentemente, torna-se necessário a criação constante de bases de dados populacionais amplas que registam as frequências alélicas, considerando a distribuição geográfica e a composição étnica relevante [28].

Os STRs autossómicos são analisados em sistema multiplex. Os *kits* comerciais incluem *primers* otimizados e fluorocromos distintos para permitir leitura simultânea por eletroforese capilar. Tal como referido anteriormente, a padronização da nomenclatura assegura comparabilidade internacional, classificando cada alelo pelo número exato (ou decimal em caso de repetições parciais) das unidades repetitivas detetadas [5]. Alguns *kits* presentes no mercado são: *kit PowerPlex System* (Promega, EUA), *kit GlobalFiler* (Thermo Fisher Scientific, Reino Unido), *kit Aggun expressmarker* (AGCU ScienTech Incorporation, China), *kit goldeneye* (Peoplespot Corporation, China), *kit AmpFISTR® Identifier™* (Thermo Fisher Scientific, EUA).

STRs do Cromossoma X (X-STRs)

Os X-STRs são marcadores genéticos localizados no cromossoma X, possuem um padrão de hereditariedade único, diferente dos autossómicos e dos Y-STRs, onde os indivíduos do sexo masculino herdam apenas uma cópia da mãe e os indivíduos do sexo feminino herdam as duas cópias, uma de cada progenitor [18, 29, 30]. Devido a esse padrão de hereditariedade, é bastante utilizado em análises forenses

especialmente em testes de parentesco mãe-filho, em meias-irmãs na ausência do material genético do progenitor e também avós e neta [18, 31].

A análise de X-STRs requer ajustes específicos devido ao potencial desequilíbrio de ligação significativo entre os *loci*, que resulta da ausência de recombinação em indivíduos do sexo masculino. Por se encontrarem fisicamente próximos, estes loci podem manter associações alélicas detetáveis [23].

A localização dos STRs no cromossoma X também tem implicações no contexto forense. Eles são, particularmente, valiosos quando combinados com dados autossômicos e Y-STRs para resolver casos complexos envolvendo múltiplos relacionamentos familiares simultâneos ou suspeitos relacionados parcialmente à vítima [8, 32].

Estudos populacionais para estes marcadores são imprescindíveis, uma vez que as frequências alélicas e haplotípicas variam entre diferentes regiões e podem afetar significativamente parâmetros forenses como o CPD e o CPE [26, 33].

Os painéis de X-STR atualmente descritos na literatura apresentam elevada diversidade genética (DG) e elevado conteúdo de informação de polimorfismo (PIC), com valores frequentemente comparáveis aos dos marcadores autossômicos mais informativos [34].

Tal como acontece nos STRs autossômicos, os *kits* comerciais disponíveis permitem multiplexar uma dúzia ou mais destes *loci* num único ensaio PCR. Essa configuração possibilita a leitura simultânea dos marcadores, aumentando eficiência na rotina forense [34, 35].

Aplicações práticas mostram que estes marcadores são úteis não só na identificação humana direta, mas também complementam investigações forenses sensíveis com viés sexual claro nos vínculos investigados. Por exemplo, em casos onde o material genético disponível provém, predominantemente, da vítima feminina e há necessidade de investigar o parentesco entre suspeitos do sexo masculino vinculados por via paterna indireta, os marcadores X-STR podem elucidar ambiguidades que os marcadores autossômicos, isoladamente, não conseguem resolver [35].

STRs do Cromossoma Y (Y-STRs)

Os Y-STRs estão localizadas no cromossoma Y e são transmitidos exclusivamente de pai para filho, refletindo sua característica haplóide. Isso implica ausência de

recombinação meiótica na maior parte da extensão do cromossoma, com exceção das regiões pseudoautossômicas [14, 36, 37]. Essa característica permite traçar a linhagem masculina direta, de um indivíduo, o que se revela particularmente valioso em contextos de investigação forense. Em casos que envolvem apenas DNA masculino ou vestígios compostos por grandes quantidades de DNA feminino misturado com contribuições masculinas mínimas. Como por exemplo em casos de agressão sexual, os *kits* comerciais validados de Y-STR demonstraram capacidade de detetar perfis masculinos mesmo quando o DNA proveniente do homem está presente em proporções ínfimas relativamente ao material feminino [38]. Essa ausência de separação meiótica implica que o poder discriminatório dos Y-STRs seja menor para distinguir indivíduos não aparentados dentro de populações pequenas ou isoladas com níveis elevados de endogamia.

Devido ao padrão de hereditariedade, formam haplótipos estáveis transmitidos intactos ao longo das gerações. Esta característica torna-os muito úteis em investigações genealógicas, análise de origem ancestral e identificação de misturas de DNA masculino/feminino e na genética populacional é utilizada para estudar padrões evolutivos e de fluxo genético devido à diversidade dos haplótipos [14].

Com base nas taxas de mutação os Y-STR são classificados como Y-STRs de mutação rápida (RM) e Y-STRs de mutação lenta (SM). Sendo que nos RM Y-STR a taxa de mutação encontra-se na ordem de 10^{-2} por geração, fazendo com que, sejam mais indicados para identificação forense de linhagem paterna ou outra identificação complexa de parentesco, estes marcadores introduzem variação suficiente para diferenciar inclusive irmãos ou primos paternos próximos. Os SM Y-STRs são marcadores mais estáveis, com uma taxa de mutação de aproximadamente 10^{-3} , preservando melhor a assinatura genética ao longo das gerações e tomando-se, mais adequados para estudos genealógicos [13].

Os parâmetros forenses associados aos Y-STRs incluem a diversidade haplotípica e a capacidade de discriminação (DC). Estudos populacionais, como os realizados na população Lisu na China, identificaram centenas de haplótipos distintos a partir destes *loci*, com valores de diversidade haplotípica superiores a 0,99 em determinadas populações. Esses resultados evidenciam elevada variabilidade genética, mesmo considerando a transmissão uniparental do cromossoma Y [39].

Os *kits* amplamente usados combinam conjuntos “core” herdados do sistema *Yfiler* (17 *loci* tradicionais), loci altamente discriminantes adicionais e RM Y-STRs para maximizar eficiência tanto em contexto internacional quanto regional [13].

2.4. Técnicas de detecção e análise

Embora sejam utilizadas muitas técnicas de detecção, a combinação do PCR *multiplex* com detecção por eletroforese capilar é a espinha dorsal da genética forense e populacional devido à sua eficiência, sensibilidade, robustez e padronização, permitindo a genotipagem simultânea de dezenas de marcadores STR de forma rápida e confiável [7].

PCR multiplex

A reação em cadeia da polimerase (PCR) é a técnica primordial para a detecção e análise dos STRs, permitindo a cópia seletiva de regiões genômicas específicas a partir de quantidades mínimas de DNA. O processo envolve três fases cíclicas principais [8]:

- Desnaturação da dupla hélice
- Ligação dos *primers* às sequências flanqueadoras
- Extensão catalisada pela DNA polimerase

O desenvolvimento do PCR *multiplex* permitiu com que múltiplos *loci* STRs fossem amplificados, simultaneamente, numa mesma reação, com elevada eficiência, permitindo que mesmo pequenas quantidades de amostra sejam suficientes para gerar informações genéticas em menor tempo, simplificando o processamento laboratorial [40]. Para isso, é necessário um desenho criterioso dos *primers* para evitar formação de dímeros e sobreposição de fragmentos no mesmo espectro de detecção, além da combinação estratégica de marcadores com tamanhos e fluorocromos distintos para otimizar a resolução eletroforética [20, 40].

Eletroforese capilar

A eletroforese capilar consolidou-se como a principal técnica para análise de STRs em genética forense, substituindo métodos anteriores como a eletroforese em gel de

poliacrilamida. O princípio fundamental baseia-se na migração diferencial de fragmentos de DNA carregados negativamente através de um capilar com polímero sob a ação de um campo elétrico. O tamanho do fragmento é inversamente proporcional à sua mobilidade: moléculas menores migram mais rapidamente que moléculas maiores, permitindo inferir o número de repetições no *locus* STR a partir da distância migrada [5].

No contexto forense, essa separação é associada à detecção por fluorescência induzida a laser. Cada fragmento recebe um fluoróforo específico durante a PCR *multiplex*, e sistemas multicanais, permitem leitura simultânea em diferentes comprimentos de onda, facilitando a interpretação de múltiplos *loci* amplificados numa única corrida [20].

A precisão da técnica decorre não apenas da física da migração eletroforética, mas também da química do polímero e do controlo experimental. A matriz capilar, geralmente composta por um polímero líquido substituível, cria um ambiente viscoso que discrimina diferenças inferiores a um par de base, o que é essencial para distinguir alelos que diferem por apenas uma repetição [5]. A estabilidade térmica, garantida por sistemas de controle ativo de temperatura, reduz variações de mobilidade entre corridas.

O processo envolve três etapas principais:

- Injeção da amostra – eletrocinética (sensível à força iónica) ou hidrodinâmica (mais uniforme, exige calibração).
- Detecção – lasers excitam os fluoróforos e detetores captam a fluorescência de cada canal.
- Processamento – softwares como *GeneMapper ID-X* geram eletroferogramas, calibrados com marcadores internos.

Laboratórios forenses adotam controlos internos rigorosos, incluindo negativos e positivos calibrados, com protocolos corretivos em caso de divergências [20]. Apesar da elevada resolução, rapidez e compatibilidade com *multiplex* complexos, a técnica tem custos elevados e exige manutenção frequente. Alternativas existem, mas nenhuma iguala a sensibilidade e padronização internacional da eletroforese capilar [5, 20, 41].

2.5. Kits comerciais

Atualmente, a amplificação dos STRs é realizada utilizando diversos *kits* comerciais, de acordo com o tipo de marcador e o objetivo da análise. Por exemplo, o *kit PowerPlex*

Y23 é amplamente utilizado para Y-STRs [42, 43], o *Investigator® Argus X-12* para X-STRs [11], e o *AmpFISTR Identifier Plus* ou o *PowerPlex 21/23* para STRs autossómicos [44, 45, 46].

Investigator Argus X-12

O kit *Investigator Argus X-12*, desenvolvido pela QIAGEN, é um dos kits mais utilizados para análise de X-STRs e inclui 12 marcadores X-STR (DXS10074, DXS10079, DXS10101, DXS10103, DXS10134, DXS10135, DXS10146, DXS10148, DXS7132, DXS7423, DXS8378, HPRTB) e o *locus* da amelogenina. [31], [47].

De acordo com a localização no cromossoma X, os 12 marcadores X-STR estão associados em quatro grupo de ligação [47]:

- LG1 (Xp22): Composto pelos marcadores DXS8378, DXS10135 e DXS10148.
- LG2 (Xq11): Abrange os marcadores DXS7132, DXS10074 e DXS10079.
- LG3 (Xq26): Inclui os marcadores DXS10101, DXS10103 e HPRTB.
- LG4 (Xq28): Contém os marcadores DXS7423, DXS10134 e DXS10146

Possui marcadores altamente polimórficos que podem ser usados para identificação individual, testes de paternidade e análise de estrutura populacional. A combinação destes marcadores oferece uma ferramenta poderosa para estudos forenses e genéticos, especialmente em casos onde a informação dos marcadores autossómicos ou do cromossomo Y pode ser insuficiente [18, 48].

Kit PowerPlex Y23

O kit *PowerPlex Y23*, desenvolvido pela Promega Corporation, permite co-amplificar 23 *loci* de Y-STR em uma única reação *multiplex*. Este kit inclui 17 *loci* Y-STR típicos do kit Y Filer, além de quatro Y-STRs altamente discriminantes e dois Y-STRs com elevadas taxas de mutação. A combinação destes marcadores permite uma análise mais abrangente e discriminativa das linhagens paternas, sendo amplamente utilizado em estudos populacionais, forenses e de história evolutiva [49].

Os *loci* incluídos no kit *PowerPlex Y23* são: DYS533, DYS438, DYS437, DYS570, DYS635, DYS390, DYS439, DYS392, DYS643, DYS576, DYS389I, DYS448, DYS389II, DYS19, DYS391, DYS481, DYS549, DYS393, DYS458, DYS385a/b, DYS456 e GATA-H4. Esses marcadores foram selecionados pela elevada variabilidade genética e

capacidade de discriminação entre indivíduos e populações, tornando o *kit* mais eficaz para análises de paternidade, estudos populacionais e investigações forenses.

2.6. Banco de dados STRs

As bases de dados de STRs são cruciais tanto na genética forense quanto na populacional. No contexto forense, eles organizam perfis genéticos individuais como se cada amostra contasse sua própria história, permitindo calcular parâmetros estatísticos, como PD e CPD e garantindo que as evidências não se percam no mar de informações [20]. Ferramentas como STRBase, CODIS e YHRD permitem que um perfil de DNA possa ser rastreado, comparado e analisado, até mesmo em casos complexos de misturas genéticas [27, 50].

Estas bases de dados permitem estudar a variabilidade genética, estrutura de populações, diferenciação populacional e trajetórias migratórias. Os dados de marcadores autossômicos, X e Y funcionam como mapas secretos, mostrando afinidades e distâncias entre diferentes grupos [20]. Contudo, sem representatividade e atualização contínua, estas bases podem se tornar espelhos distorcidos da realidade, comprometendo análises e inferências. Por isso, a manutenção e expansão destes repositórios não só fortalecem a genética forense, como também contribuem para o conhecimento da diversidade genética humana [51].

3. Genética populacional

A genética populacional é o ramo da biologia que mostra como as forças evolutivas como mutação, seleção, deriva e migração moldam a distribuição dos alelos nas populações ao longo do tempo, fornecendo informações valiosas sobre a diversidade genética, a variabilidade e a estrutura de populações [52].

3.1. Frequência alélica

Estimar as frequências alélicas compõe uma etapa fundamental em análises genéticas populacionais com relevância forense direta, visto que esses valores sustentam cálculos de probabilidades de correspondência, razões de verossimilhança e estatísticas

relacionadas. As frequências alélicas representam a proporção relativa de cada alelo observada num determinado *locus* dentro de uma população [53].

Esse conceito é essencial para estimar a distribuição genotípica esperada segundo as premissas do equilíbrio de Hardy-Weinberg (HWE), que estabelece que, numa população panmítica sem influências de seleção, migração ou mutação, as frequências alélicas permanecem constantes ao longo das gerações e as frequências genotípicas podem ser determinadas diretamente a partir delas [5, 52].

3.2. Princípio de Hardy-Weinberg

Em 1908, o matemático Godfrey H. Hardy e o médico alemão Wilhelm Weinberg formularam, independentemente, um princípio que se tornou a base de toda a genética populacional. O Equilíbrio de Hardy-Weinberg (HWE) estabelece que, na ausência de fatores evolutivos, as frequências alélicas e genotípicas numa população mantêm-se constantes ao longo das gerações. Matematicamente é ilustrado pela equação $p^2 + 2pq + q^2 = 1$, onde p e q representam a frequência dos alelos [52, 54].

Os pressupostos inerentes ao HWE são os seguintes [23, 31, 55]:

- O acasalamento ocorre aleatoriamente (panmixia)
- População infinitamente grande, elimina os efeitos da deriva genética aleatória
- Ausência de mutação, não surgindo novos alelos
- Ausência de migração (fluxo génico), ou seja, não há entrada ou saída de indivíduos da população
- Ausência de seleção natural, isso é, todos os genes têm igual probabilidade de serem passados à descendência

O teste do desvio do HWE é feito usando o teste de Qui-quadrado, cuja fórmula é $\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$, ou pelo teste exato podendo utilizar ferramentas como Arlequin v^{3.5} ou o STRAF [56].

Desvio de HWE, feito com recurso, por exemplo, ao teste do Qui-quadrado, pode indicar erros de genotipagem, subestrutura populacional ou ocorrência de seleção natural e também é usado para calcular a variabilidade. Subestrutura populacional (o chamado princípio de Wahlund) leva frequentemente a excesso de homozigóticos e défice relativo

de heterozigóticos quando subgrupos com estruturas genéticas distintas são tratados como uma única população [22].

3.3. Diversidade e estrutura populacional

A diversidade genética humana resulta de processos históricos e demográficos que moldam a distribuição e a frequência dos alelos, sendo a deriva, a migração e o isolamento geográficos fatores-chave na configuração atual [52].

O termo “estrutura populacional” traduz precisamente as correlações entre frequências alélicas, decorrentes da organização não aleatória dos indivíduos no espaço físico ou social [57]. Medidas como F_{ST} (Índice de Diferenciação Populacional) quantificam quanto da variância total pode ser atribuída à diferenciação entre subpopulações; valores próximos de zero indicam elevada homogeneidade genética, enquanto valores altos refletem segregação marcada [58].

De forma complementar, o coeficiente de endogamia total (F_{ST} - Mede a redução da heterozigosidade na população total) resume a redução da heterozigosidade total, enquanto o coeficiente de endogamia Intra populacional (F_{IS} - Mede o desvio da heterozigosidade dentro de subpopulações) foca no desvio dentro de subpopulações específicas. A relação entre estes coeficientes é expressa matematicamente como: $1-F_{IT}=(1-F_{IS})(1-F_{ST})$ ou, $F_{IT} = 1 - (H_I / H_T)$, onde H_I representa a heterozigosidade média observada nos indivíduos e H_T a heterozigosidade esperada calculada considerando toda a população como panmítica única.

Essa equação destaca que a endogamia local e a diferenciação populacional não são independentes, mas sim forças que atuam em conjunto modulando a diversidade observada [33].

Atualmente, a estimativa de F_{ST} , F_{IS} e F_{IT} baseia-se em métodos estatísticos derivados de heterozigosidade esperada e observada, ajustados por amostras multilocus. Ferramentas como AMOVA (*Analysis of Molecular Variance*) permitem decompor a variância genética em níveis hierárquicos (indivíduos, subpopulações, população total), obtendo valores consistentes com diferentes modelos mutacionais [57].

No contexto forense, valores empíricos de F_{ST} são aplicados como fatores de correção probabilística, ajustando a estimativa de raridade de perfis em populações estruturadas. Assim, o uso adequado destes coeficientes evita superestimações e fortalece a robustez estatística em análises periciais e populacionais [17, 20, 57].

4. Genética forense

4.1. Evolução metodológica e integração de marcadores

A evolução da genética forense reflete uma confluência de avanços científicos, mudanças legais e necessidades crescentes do sistema de justiça criminal [2]. Desde a introdução inicial do uso do DNA como ferramenta de identificação humana na década de 1980, houve um progresso notável tanto nas metodologias laboratoriais quanto nos modelos estatísticos empregados para interpretação dos resultados [59, 60].

As primeiras aplicações ocorreram com técnicas baseadas em RFLP (*Restriction Fragment Length Polymorphism*), que exigiam quantidades consideráveis de DNA não degradado e o processamento era demorado [2, 60]. Foi precisamente esta limitação que impulsionou o desenvolvimento do PCR. Essa técnica, permitiu a amplificação de regiões curtas do genoma e abriu caminho para a aplicação usual dos STRs [8].

A consolidação dos STRs coincidiu com a criação das primeiras bases de dados genéticos criminais, que reuniam perfis obtidos a partir de amostras conhecidas e evidencias recolhidas em cenas de crime. Nos Estados Unidos, a década de 1990 presenciou o surgimento do CODIS, cujo crescimento foi impulsionado pela legislação que expandiu gradualmente as categorias de crimes para os quais era obrigatória a recolha de amostra de DNA. Em 1999, nos EUA, havia apenas um Estado que exigia a recolha, mas esse número aumentou para 32 em 2014, acompanhado por investimentos federais substanciais que possibilitaram processar cerca de 1,5 milhões de amostras anuais [34].

A incorporação dos STRs autossômicos nos perfis forenses constitui um avanço significativo, revolucionando não apenas as investigações criminais, mas também outras áreas, como a identificação de vítimas em desastres e a resolução de casos civis [34, 8].

Atualmente, os STRs autossômico constituem a espinha dorsal da identificação individual. A elevada heterozigotia e ampla distribuição nas populações garantem fiabilidade, sustentam a sua aplicação no contexto de genética forense. Mesmo assim, os STRs autossômicos apresentam limitações em cenários específicos. Em amostras com misturas onde há predominância de material feminino face à fração masculina, a identificação pode ser desafiadora. Situações de vínculos biológicos específicos como

irmãos com o mesmo pai ou reconstrução genealógica de linhas maternais, também podem carecer de resolução adequada apenas com dados biparentais [8].

É justamente para preencher essas lacunas que entram os marcadores sexuais. Os Y-STRs permitem rastrear linhagens patrilineares mesmo diante de grande quantidade de DNA feminino, enquanto os X-STRs oferecem uma perspectiva complementar em análises de vínculos familiares [26].

Do ponto de vista interpretativo, a análise qualitativa baseada em limiares fixos deu lugar à razão de verossimilhança (LR), que quantifica a probabilidade relativa dos dados sob hipóteses alternativas e reduz a arbitrariedade de limites rígidos [32]. Em paralelo, protocolos rigorosos contra contaminação e transferência secundária foram instituídos após evidências de que tarefas laboratoriais rotineiras podem introduzir DNA indesejado [8].

A evolução tecnológica trouxe consigo debates éticos. Estudos relacionados com procura familiar ampliam a capacidade de investigação, mas levantam algumas preocupações [34]. Da mesma forma, predições fenotípicas permanecem confiáveis apenas para poucas características, como cor dos olhos ou cabelo, não sendo adequadas para traços complexos, como altura, índice de massa corporal [20, 61].

Em síntese, a trajetória da genética forense desde primeiros perfis RFLP até os painéis de *multiplex* atuais, revela um equilíbrio contínuo entre ganhos técnicos e desafios legais, acompanhado pela expansão das redes de dados nacionais e internacionais que potencializam sua aplicação investigativa.

A genética forense tem como objetivo gerar e interpretar perfis de DNA para identificação humana ou elucidação de relações biológicas em investigações criminais e civis, traduzindo vestígios biológicos em informações comparáveis a suspeitos, vítimas, familiares ou bases de dados, mediante protocolos padronizados que previnem erros e contaminação [8]. Após recolha e transporte adequados, realiza-se identificação presumtiva e confirmatória do vestígio [59], seguida da extração e quantificação do DNA por qPCR, que avalia quantidade e qualidade, orientando o uso de miniSTRs ou SNPs [8].

A amplificação *multiplex* por PCR com STRs autossômicos, Y-STRs ou X-STRs, conforme o caso gera fragmentos separados por eletroforese capilar, interpretados por *softwares* que auxiliam na análise de perfis simples, mistos ou degradados [26, 32, 34].

Os genótipos são avaliados estatisticamente pelo *Likelihood Ratio* (LR), que compara hipóteses concorrentes considerando frequências alélicas populacionais [33]. Em perfis complexos, oriundos de múltiplos dadores, de amostras degradadas ou com baixo conteúdo de DNA, os valores de *Combined Probability of Inclusion* (CPI) tornam-se limitados, pois ignoram variações na altura dos picos e contribuições minoritárias. Modelos probabilísticos contínuos superam essas limitações ao integrar intensidades relativas, proporções de DNA, artefatos como *stutter*, taxas de *drop-out* (falha na detecção de um alelo que está presente) /*drop-in* (inclusão de um alelo estranho na amostra) e a presença ou ausência do perfil da vítima, permitindo LR mais informativos e robustos [32, 62]. A repetição experimental e o uso de *softwares* validados aumentam a estabilidade metodológica, assegurando interpretação probabilística equilibrada entre prática laboratorial e inferência estatística, essencial na análise de misturas heterogêneas [32, 34, 62].

5. Objetivos

A genética populacional e a genética forense têm beneficiado significativamente dos avanços na análise de marcadores moleculares, especialmente os STRs. Este trabalho propõe como objetivo central analisar a literatura existente no sentido de identificar e analisar marcadores STRs relevantes, com foco na sua aplicabilidade em diferentes contextos populacionais e forenses.

Os objetivos específicos do trabalho são:

- Analisar a variabilidade genética de diferentes populações por meio de marcadores STR
- Examinar a variação genética entre distintos grupos populacionais
- Estudar a aplicabilidade dos diferentes tipos de marcadores STR em diferentes populações

6. Metodologia

A pesquisa e seleção dos artigos foi realizada usando exclusivamente o base de dados PubMed. Para o presente estudo a pesquisa dos artigos foi feita utilizando a combinação de palavras-chaves com operadores booleanos e ternos da pesquisa. Para a primeira

triagem a combinação foi feita usando palavras-chaves: “STR”, “*population genetic*” e “*human*” com o operador booleano AND e o termo de pesquisa usado foi *Title/Abstract*, com “*publication date: 10 years*” e “*free full text*” como critérios de inclusão. No total foram encontrados 88 artigos. Após a leitura dos títulos e resumos foram selecionados artigos relevantes para o estudo e foram excluídos artigos que não têm foco em marcadores genéticos, artigos de revisão, estudos em animais, plantas e outras espécies, e artigos sem acesso ao texto integral. Após esta filtragem, ficaram 41 artigos no total para análise (tabela em anexo).

Da análise dos 41 artigos, verificamos se cada um dos estudos se enquadrava no âmbito deste projeto, nomeadamente em relação a compreender estudo de genética populacional, identificação pessoal, teste de paternidade, estrutura populacional e parâmetros forenses.



Figura 3. Esquema ilustrativo da metodologia utilizada neste trabalho

7. Caracterização dos estudos incluídos

A análise das populações investigadas nos 41 artigos revelou forte concentração de estudos em populações asiáticas (30 artigos, 73%), sendo a população chinesa responsável por 49% destes estudos. A população do Médio Oriente aparece em seis artigos (15%), e a população europeia está representada por sete trabalhos (17%). A população africana encontra-se representada em apenas dois estudos e a população da Oceânia foi estudada em apenas um artigo. As populações americanas estão representadas num único estudo, mas também estão referidas de forma pontual e em contextos comparativos. Essa distribuição demonstra uma forte concentração regional, indicando lacunas de investigações em populações africanos e latino-americanos, que permanecem sub-representados em estudos de validação forense.

7.1. STRs autossómicos

Os artigos que utilizaram STRs autossómicos confirmam seu papel central na identificação individual e em testes de paternidade, além de contribuírem para estudos de genética populacional. *Kits* como o SureID PanGlobal (*HEALTH Gene Technology Co*, China), SureID 27comp e o AGCU Expressmarker 22 foram aplicados em populações da Europa (Inglaterra, Irlanda, Romênia) e da Ásia (diferentes regiões da China). Esses estudos destacaram parâmetros forenses como poder de discriminação, probabilidade de exclusão (PE) e equilíbrio de Hardy-Weinberg, confirmando a elevada eficiência destes painéis e a importância da validação local das frequências alélicas para assegurar a robustez probatória em cada população analisada.

7.2. X-STRs

Os X-STRs mostraram relevância especial em investigações de parentesco envolvendo vínculos maternos e em cenários de complexidade genealógica. Populações como a da Eslováquia e da China foram analisadas com *kits* como o Investigator Argus X-12 e o Goldeneye 17X, evidenciando seu valor tanto em aplicações forenses familiares quanto em caracterizações populacionais. Esses marcadores ampliam a resolução em casos onde os marcadores autossómicos não são suficientes.

7.3. Y-STRs

Os Y-STRs tiveram papel destacado em estudos de estrutura populacional masculina, fluxo génico e genealogia patrilinear, além de aplicações forenses em crimes sexuais, nos quais o DNA masculino está em baixa proporção. Populações como a de Qatar, China (Lisu, Dong, Uygur, Han) e Cazaquistão foram investigadas com *kits* como o

PowerPlex® Y23 e o Goldeneye Y Plus, que demonstraram elevado poder de resolução. No Médio Oriente, casos como o da população do Qatar e Iraque revelaram a utilidade do Y-STR na compreensão de padrões migratórios e na diferenciação intra-regional.

Tabela 2. Caracterização dos estudos e a sua aplicação

Doi	Nº e tipos STR	Kit utilizado no estudo	Números de alelos	Aplicação
https://doi.org/10.1080/03014460.2023.2265814	44 STR	<i>SureID® PanGlobal Human DNA Identification Kit and SureID® 27comp Human DNA Identification Kit</i>	—	Teste de Paternidade Identificação Pessoal Genética populacional
https://doi.org/10.1080/03014460.2023.2264769	12 X-STR	<i>Investigator Argus X-12 STR kit</i>	178	Análises forenses Teste de Paternidade
https://pubmed.ncbi.nlm.nih.gov/37780380/	44 Y STR	<i>Goldeneye DNA Identification System Y Plus kit</i>	393	Análises forenses
https://doi.org/10.1080/03014460.2023.2213478	16 X STR	<i>Goldeneye 17X kit</i>	149	Análises forenses Genética populacional
https://doi.org/10.1080/03014460.2023.2213478	23 Y-STR	<i>PowerPlex Y23 kit</i>		Análises forenses Estrutura populacional Genética populacional

https://doi.org/10.1080/03014460.2023.2224972	23 Y-STR	<i>PowerPlex® Y23 PCR amplification kit</i>	189	Análises forenses Estrutura populacional
https://pubmed.ncbi.nlm.nih.gov/37510323/	21 A-STR	<i>AGCU Expressmarker 22 kit</i>	333	Teste de Paternidade Identificação Pessoal

Limitações metodológicas

É fundamental reconhecer que os valores usados no presente estudo provêm de estatísticas e parâmetros já reportados em artigos científicos, e não de dados brutos individuais. Uma limitação adicional reside no facto de nem todos os artigos consultados reportarem a totalidade dos parâmetros genéticos essenciais, sendo frequente a omissão de indicadores cruciais como HD ou DC em estudos focados em populações isoladas ou específicas.

8. Marcadores X-STR

Tal como já foi referido anteriormente, os X-STRs têm sido amplamente utilizados em estudos forenses e populacional devido à elevada variabilidade e capacidade de fornecer informações sobre relações de parentesco complexas. Por apresentarem um padrão de hereditariedade único, onde o indivíduo do sexo masculino tem apenas uma cópia herdada da mãe e o indivíduo do sexo feminino tem duas cópias, torna-se uma ferramenta útil e mais preciso para análise de paternidade envolvendo filhas, estabelecer relacionamentos mãe-filho e relacionamentos avó-neta.

Nas secções seguintes analisaremos seis estudos sobre os X-STRs de diferentes populações, que utilizam o mesmo *Kit Investigator Argus X-12*, com o objetivo de fazer uma análise comparativa entre eles, extraindo conclusões sobre a diversidade e estrutura populacional.

8.1. Populações e Amostragem

Para uma análise mais detalhada dos marcadores X-STR, as populações estudadas encontram-se representadas na Figura 4. Na Europa foram analisadas as populações da Eslováquia, Noruega e Suíça [11, 18, 63]. Na Ásia, no Paquistão foram estudadas as populações Punjabi e Kashmiri das regiões Punjabi e Azad Jammu e Caxemira [47]. Na China, populações Mongol e Han Chineses da Mongólia Interior e do leste da China [31]. E por fim, oito grupos étnicos Eritreus (Afar-Cushitic, Bilen-Cushitic, Hedareb-Cushitic, Kunama Nilo-Saharan, Nara Nilo-Saharan, Saho-Cushitic, Tigre-Semitic e Tigrinya-Semitic), que pertencem a Eritreia, África [64].



Figura 4. Ilustração geográfica das populações estudadas na análise dos marcadores X-STRs

As amostras coletadas variam dependendo da população e do estudo de referência. O número de indivíduos analisados também apresentou grande variação, de 125 a 1.198 participantes, permitindo uma cobertura representativa para análise genética comparativa (tabela 3).

Tabela 3. Amostras das populações estudadas; NR: não relacionados; AN: anónimos

População	Números de indivíduos	Material recolhido	Ref.
Punjabi e Caxemira	125 doadores voluntários	Sangue	[47]
Eritreia	255 indivíduos NR	DNA	[64]
Mongol e Han Oriental	541 indivíduos NR	Sangue	[31]
Eslovaca	378 indivíduos AN	Esfregaço bucal	[18]
Suíços	1198 indivíduos	DNA	[63]
Noruegueses	631 homens NR	Esfregaço bucal	[11]

8.2. Metodologias usadas pelos artigos

Os X-STR foram amplificados com o *kit Investigator Argus X-12* (Qiagen) e genotipados por eletroforese capilar em sequenciadores ABI, utilizando o *software* GeneMapper para designação de alelos (tabela 4).

Tabela 4. Metodologia utilizada nos estudos analisados

População	Amplificação	Genotipagem	Ref.
Punjabi e Caxemira	<i>Kit Investigator Argus X-12</i> (Qiagen), <i>GeneAmp PCR System 9700</i>	Eletroforese capilar ABI 3500; genotipagem por <i>GeneMapper</i> v4.0, seguindo protocolo ISFG	[47]
Eritreia	<i>Kit Investigator Argus X-12</i> (Qiagen)	ABI PRISM 310 <i>Genetic Analyzer</i> ; alelos designados no <i>GeneMapper</i> ID v3.2	[64]
Mongol e Han Oriental	<i>Kit Investigator Argus X-12</i> (Qiagen) <i>GeneAmp PCR System 9700</i>	Eletroforese capilar ABI PRISM 3130XL; designação de alelos <i>Genemapper</i> ID v3.2.1	[31]
Eslováquia	<i>Kit Investigator Argus X-12</i> (Qiagen)	Analizador 3500 (<i>ThermoFisher</i>); genotipagem <i>GeneMapper</i> ID-X v1.4	[18]
Suíça		Eletroforese capilar 3500xl; interpretação <i>Genemapper</i> ® ID-X v1.4	[63]

Noruega	<i>kit Investigator Argus X-12 QS (Qiagen)</i>	Eletroforese capilar em 3500xl (<i>Thermo Scientific</i>). Genótipos analisados no <i>GeneMapper ID-X</i> 1.4, RFU ≥ 100	[11]
---------	--	---	------

8.3. Análise estatística

8.3.1. Equilíbrio de Hardy Weinberg e desequilíbrio de ligação

Equilíbrio de Hardy-Weinberg (HWE) numa análise comparativa destes estudos mostraram resultados distintos. Na população Punjabi e Caxemira, nove e sete *loci*, respetivamente, mostraram desvios de HWE, no entanto, após a correção de Benferroni apenas HPRTB e DXS10148 na população Caxemira, e DXS10079, DXS10103 e DXS10134 na população Punjabi revelaram desvio ao HWE. Já na população Mongol e Han oriental, com exceção do *locus* DXS10134, não foi observado nenhum desvio estatisticamente significativo do HWE. Na população norueguesa, eslovaca e Suíça não foram observados desvios estatisticamente significativos relativos ao HWE.

8.4. Parâmetros forenses

8.4.1. Conteúdo de Informação de Polimorfismo (PIC)

A Tabela 5 mostra os valores de PIC obtidos para 12 *loci* X-STR, para as populações da Mongólia e Han Oriental, Eslováquia, Punjabi e Caxemira. E também na população Eritreia os valores do PIC estão apresentados em grupos de ligação.

O *locus* DXS10135 foi o mais polimórfico para a população Caxemira, Mongol e Han Oriental e Eslovaca, com valores mais próximos da unidade. Na Eritreia o LG1 apresentou valores de PIC mais elevados, com o *locus* DXS10135 sendo mais polimórfico com um valor de PIC de 0,94, indicando uma maior variabilidade genética e na população norueguesa o LG1 mostrou maior diversidade genética. Em contrapartida, o *locus* DXS7423 apresentou menor informação forense para a maioria das populações analisadas, refletida por um valor de PIC de 0.4597 na população Mongol e Han Oriental, indicando poder discriminatório reduzido, o que reflete menor utilidade forense quando analisado isoladamente.

Tabela 5. Valores de PIC das populações analisadas com 12 X-STRs

Locus	Mongólia	Eslováquia	Punjabi	Caxemira	Eritreia
Han Oriental					
DXS8378	0,5558	0,6293	0,7289	0,7055	LG1 0,9936
DXS10135	0,9070	0,9274	0,9151	0,9267	LG2 0,9881
DXS10148	0,9043	0,8633	0,9377	0,9192	LG3 0,9901
DXS7132	0,6975	0,6892	0,7671	0,7267	LG4 0,9928
DXS10074	0,7444	0,8230	0,8505	0,8381	
DXS10079	0,8032	0,7954	0,8573	0,8418	
DXS10101	0,8889	0,8902	0,9179	0,9006	
DXS10103	0,7496	0,7100	0,7708	0,8165	
HPRTB	0,6594	0,7133	0,7984	0,7526	
DXS7423	0,4597	0,6699	0,8097	0,6938	
DXS10134	0,8323	0,8556	0,9240	0,8969	
DXS10146	0,8670	0,8797	0,8705	0,8460	

8.4.2. Poder de Discriminação (PD) e Poder Combinado de Discriminação (CPD)

Os painéis de 12 X-STRs mostraram um alto polimorfismo e elevado poder discriminatório nas populações abordadas (figura 5). Na população Mongol, Eslovaca, Punjabi e Caxemira, o poder combinado de discriminação para indivíduos do sexo masculino foi >0.9999999983, 0.999999983, 0.9999999984256, 0.9999999992276, respetivamente. O poder combinado de discriminação para indivíduos do sexo feminino é ligeiramente superior, refletindo o facto de possuírem duas cópias do cromossoma X.



Figura 5. Poder de discriminação dos 12X-STRs populações analisadas. A) Poder de discriminação dos 12X-STRs em amostras de mulheres das populações analisadas; B) Poder de discriminação de LG1-LG4 em mulheres da Eritreia e Noruega; C) Poder de discriminação dos 12X-STRs em homens das populações analisadas; D) Poder de discriminação de LG1-LG4 em homens da Eritreia e Noruega.

Os figuras de PDF e PDM demonstram que os *loci* X-STRs apresentam elevada capacidade discriminatória em todas as populações avaliadas, com valores próximos de 1,0. A análise por grupos de ligação (Eritreia e Noruega) reforça essa tendência, destacando o LG1 como o mais informativo. O locus DXS10135 mostrou-se o mais polimórfico, enquanto o DXS7423 apresentou os menores valores de PDM. De forma geral, observa-se que, apesar de pequenas variações entre populações e *loci*, o

conjunto dos 12 X-STRs garante alto poder de individualização, consolidando sua relevância em estudos forenses e de genética populacional.

8.4.3. Diversidade Haplotípicas (HD)

A diversidade haplotípica dos X-STRs foi elevada em todas as populações analisadas, confirmando a utilidade destes marcadores em estudos forenses e populacionais. As populações asiáticas, como Mongólia e Han Oriental, apresentaram os valores mais altos ($>0,99$), refletindo elevada variabilidade genética. Em contraste, a população da Caxemira exibiu os menores valores ($\sim 0,96$). As populações europeias (Eslováquia, Noruega e Suíça) mostraram valores homogêneos e elevados ($>0,98$), consistentes com uma estrutura genética estável. Já a Eritreia apresentou diversidade comparável à europeia, indicando a influência de fluxos génicos regionais.

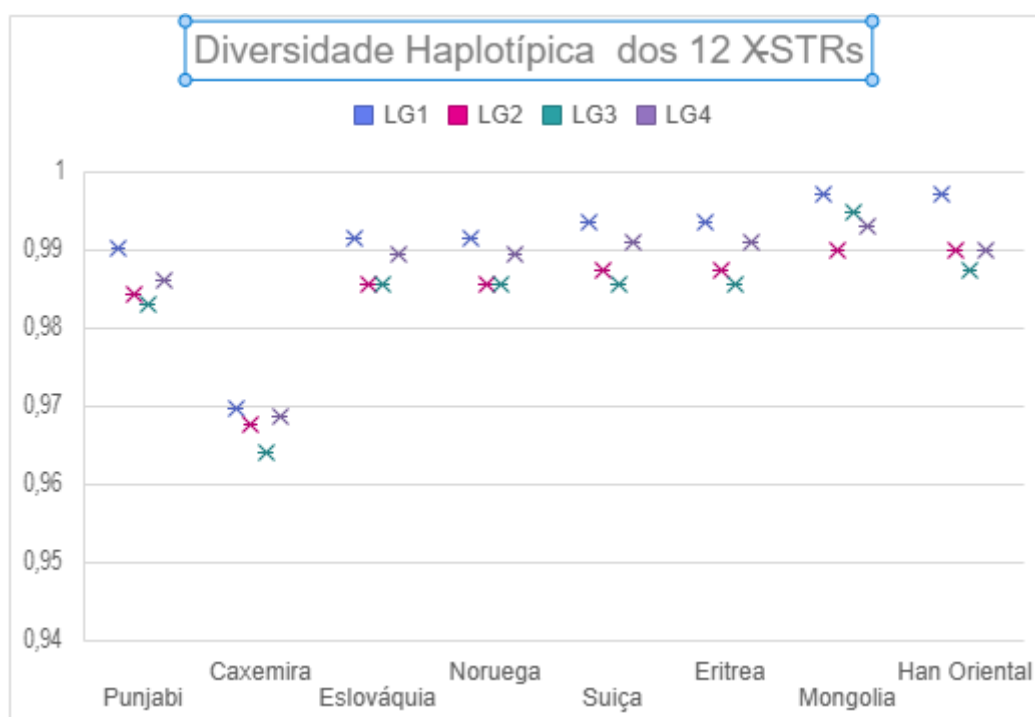


Figura 6. Diversidade haplotípica dos 12 X-STRs das populações analisadas

A distribuição da diversidade haplotípica (figura 6) evidencia de forma clara as diferenças entre as populações. Observa-se que o LG1 apresenta consistentemente os valores mais elevados de diversidade, confirmando ser o grupo de ligação mais polimórfico entre todos os analisados. Esses resultados reforçam a utilidade dos X-STRs como marcadores altamente informativos, capazes de refletir tanto a variabilidade genética quanto os processos históricos e demográficos que moldaram cada população.

8.4.4. Análise de Estrutura Populacional e diferenciação populacional

A estrutura genética mostra como padrões genéticos refletem a história e geografia das populações. Para a análise de estrutura genética utilizam a análise de componentes principais (PCA) e STRUCTURE.

Populações europeias, como Suíça e Eslováquia, mostraram elevada homogeneidade genética, sem sinais de estratificação relevante. Em contraste, populações do Sul da Ásia (Punjabi e Caxemira) apresentaram divergências mais marcantes. Já entre os grupos chineses (Mongol e Han), observou-se uma proximidade genética consistente, mas com diferenciações sutis. Esses resultados evidenciam que, embora algumas populações sejam geneticamente homogêneas, os X-STRs permitem detetar subestruturas relevantes em contextos mais complexos, reforçando sua aplicabilidade tanto em estudos forenses quanto em genética populacional.

8.5. Conclusão

Os resultados obtidos a partir da análise dos marcadores X-STRs confirmam a sua elevada utilidade em estudos de genética forense e populacional. A avaliação do poder de discriminação (PDM e PDF) evidenciou valores consistentemente altos nas diferentes populações analisadas, refletindo a capacidade destes *loci* em distinguir indivíduos não relacionados, tanto em contextos de identificação quanto em testes de parentesco.

A análise dos parâmetros forenses (PIC, PDM, PDF e CPD) mostrou que *loci* como DXS10135 e DXS10148 apresentam elevado polimorfismo, sendo os mais informativos para a maioria das populações.

Em conjunto, os resultados confirmam que os X-STRs constituem um painel altamente informativo, com aplicabilidade robusta tanto para identificação individual como para a compreensão da estrutura genética das populações. Para a prática forense, a elevada capacidade de exclusão e o poder discriminatório observados tornam esses marcadores ferramentas indispensáveis, devendo sua interpretação ser sempre contextualizada de acordo com as características específicas de cada população estudada.

9. Marcadores Y-STR

Os marcadores Y-STR são particularmente úteis para estudos forenses e análise de genealogias paternas, pois permitem traçar linhagens masculinas e identificar relações de parentesco paterno e também identificar indivíduos do sexo masculinos.

A utilização de sistemas abrangentes, como o *kit PowerPlex Y23* tem sido fundamental para o enriquecimento das bases de dados de referência de haplótipos de cromossoma Y, como a YHRD, essenciais para a interpretação de perfis de DNA e para estudos genéticos em diversas populações.

Nas secções seguintes mostra-se a análise de sete estudos com populações diferentes a fim de comparar a diversidade genética, relações populacionais e outros parâmetros forenses.

9.1. População e amostragem

Os setes estudos abrangem distintas populações a nível mundial (figura 7): na Sibéria Oriental os Yakuts, no Sudeste da China, a população VA e o grupo étnicos Lisu, ambos da província de Yunnan, ainda no sudeste da China, grupos étnicos da província de Guizhou, Hans de Zunyi, Miaos de Qiandongnan e Bouyeis de Qiannan [10, 13, 39, 65]. Na Europa, em particular na Roménia, um grupo étnico Székelys. No Médio Oriente foram analisadas populações do Iraqui e Qatar [66, 67, 68].



População	Nº de indivíduos	Material recolhido	Ref.
Yakuts	130	Restos esqueléticos	[65]
Székely	92	Esfregaço bucal	[66]
Va (Yunnan)	368	Sangue	[69]
Lisu (Yunnan)	423	Sangue	[39]
Guizhou (Bouyei, Han e Miao)	308	Sangue	[4]
Qatar	379	Saliva	[68]
Iraque	254	Sangue	[67]

9.2. Metodologias usada nos artigos

As populações foram analisadas com amplificação pelo *kit Powerplex Y23 (Promega)* e genotipagem em sequenciadores automáticos (ABI 3130, 3130XL e 3500 *Genetic Analyser*), garantindo dados comparáveis entre os estudos (tabela 7).

Tabela 7. Metodologia usada para amplificação e genotipagem dos Y-STRs

População	Amplificação	Genotipagem	Ref.
Yakuts	<i>Powerplex Y23 System (Promega)</i>	3100 ou 3500 <i>genetic analyser (Life Technologies)</i>	[65]
Székelly	<i>Powerplex Y23 (Promega)</i>	ABI 3130 <i>Genetic Analyser</i>	[66]
Va (Yunnan)	<i>Kit Powerplex Y23 (Promega)</i>	ABI 3130XL <i>Genetic Analyser</i>	[69]
Lisu (Yunnan)	<i>Powerplex Y23 (Promega)</i>	ABI 3130XL <i>Genetic Analyser</i>	[39]
Guizhou (Bouyei, Han e Miao)	<i>Kit Powerplex Y23 (Promega)</i>	3500 <i>Genetic Analyser (Applied Biosystems)</i>	[4]
Qatar	<i>Kit Powerplex Y23 (Promega)</i>	ABI 3500 <i>Genetic Analyser (Thermo Fisher)</i>	[68]
Iraque	<i>Powerplex Y23 System (Promega)</i>	ABI 3500 <i>Genetic Analyser (Thermo Fisher)</i>	[67]

9.3. Análise estatística dos Y-STRs

9.3.1. Diversidade genética (GD)

A diversidade genética variou de forma significativa entre os *loci* analisados. O *locus* DYS385a/b foi consistentemente o mais informativo, atingindo valores de GD de 0,9520 nos Bouyei e 0,9468 nos Miao. Em contraste, *loci* como DYS391 e DYS392 apresentaram baixa variabilidade, refletindo menor poder discriminatório. Esses resultados encontram-se resumidos na Tabela 8 e na Figura 8, que apresentam os valores mínimos e máximos de GD por população.

Tabela 8. *Locus* mais e menos polimórfico das populações analisadas

População	<i>Locus</i> mais polimórfico (Maior GD)	<i>Locus</i> menos polimórfico (Menor GD)
Va (Yunnan)	DYS385a/b	DYS19
Guizhou – Bouyei	DYS385a/b	DYS391
Guizhou – Han	DYS385a/b	DYS438
Guizhou – Miao	DYS385a/b	DYS391
Lisu	DYS385a/b	DYS438
Iraque	DYS385	DYS392
Qatar	DYS458	DYS392

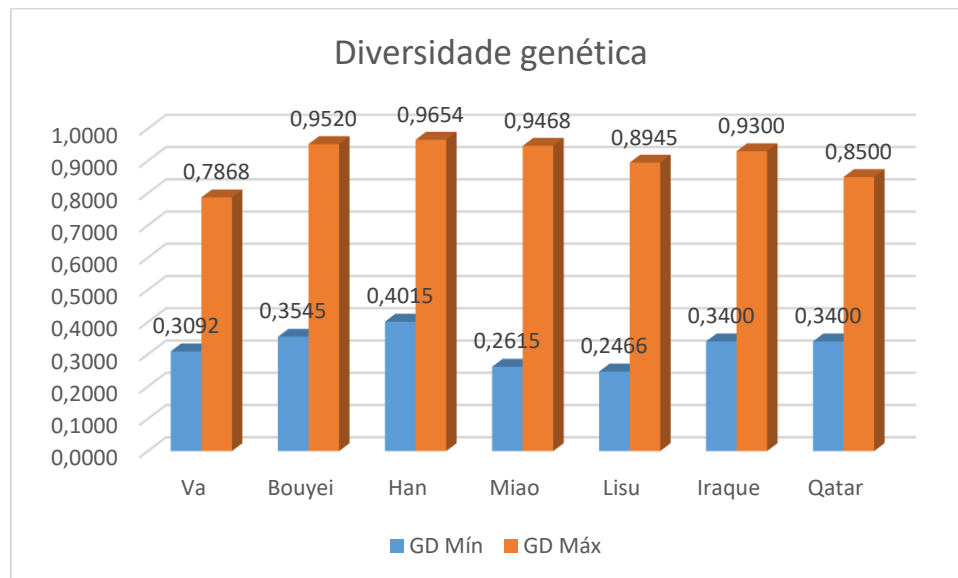


Figura 8. Diversidade genética das populações analisadas no estudo de marcadores do Y-STRs

9.3.2. Diversidade de Haplótipos (HD) e Capacidade Discriminatória (DC)

A diversidade haplotípica foi elevada em praticamente todas as populações (Tabela 9). Valores próximos de 1, como encontrados nos Han (0,9999), Miao (0,99983) e Bouyei (0,99979), confirmam a variabilidade intra-populacional. A única exceção relevante foi observada nos Yakuts, cujo HD mais baixo ($0,7726 \pm 0,0548$) sugere o impacto de gargalos demográficos e de deriva genética, de acordo com a história populacional da Sibéria. Essa diferença ressalta como fatores demográficos e históricos moldam diretamente a diversidade genética, influenciando a utilidade dos Y-STRs em diferentes contextos [65].

Tabela 9. Diversidade haplotípica (HD), capacidade de discriminação (DC) e número de haplótipos observados em populações analisadas por Y-STR

População	HD	DC	Haplótipos observados
Yakuts	$0,7726 \pm 0,0548$	-	19 haplótipos
Székelly	0,9995	0,978	72 haplótipos
Va (Yunnan)	0,9852	0,5543	204 haplótipos; 144 únicos
Catari	0,998924	86,20%	327 haplótipos únicos
Lisu	0,9977	0,7636	323 haplótipos; 271 únicos
Han	0,9999	0,9902	101 haplótipos; 100 únicos
Miao	0,99983	0,9908	108 haplótipos; 106 únicos
Bouyei	0,99979	0,97959	96 haplótipos diferentes
Iraque	-	96%	244 haplótipos distintos

A capacidade de discriminação acompanhou, em grande parte, os padrões de HD, confirmando a elevada eficácia de painéis Y-STR em diferenciar indivíduos na maioria das populações.

9.4. Conclusão

De forma geral, os resultados confirmam que os Y-STRs são ferramentas altamente informativas para estudos forenses e de genética populacional, especialmente em contextos de elevada diversidade haplotípica, como observado em várias populações asiáticas [38]. Esses padrões refletem a influência combinada de fatores históricos, geográficas e demográficas na configuração do perfil genético masculino, revelando tanto efeitos de deriva e isolamento como sinais de fluxos génicos regionais [26]. Para a aplicação forense, tais evidências reforçam a importância de considerar as especificidades de cada população, uma vez que a sobreposição de haplótipos em determinados grupos pode reduzir o poder de discriminação, ao passo que, em outros, a elevada variabilidade maximiza a aplicabilidade prática.

Apesar das vantagens, é importante destacar que, os Y-STRs descrevem apenas a linhagem paterna, restringindo interpretações quando não combinados a dados autossómicos.

Assim, a interpretação de dados baseados em Y-STRs deve ser sempre contextualizada e integrativa, incorporando informações autossómicas e demográficas. Essa abordagem aumenta a robustez analítica e garante maior confiabilidade nas aplicações forenses e populacionais.

10. Conclusões

A análise integrada dos marcadores genéticos, especialmente os STRs autossómicos, do cromossoma X e do cromossoma Y, revela a complexidade e a riqueza das ferramentas disponíveis para a genética forense e populacional. A diversidade estrutural e funcional dos marcadores STRs, somada à sua distribuição no genoma e a taxa mutacional, permite aplicações que vão da individualização de perfis à reconstrução de linhagens e à inferência de parentesco. Com bases de dados populacionais amplas e atualizadas, asseguram a confiabilidade estatística necessária para a interpretação forense e populacional.

A análise de várias populações permitiu ressaltar a importância dos STRs autossómicos para a distinção entre indivíduos, e a utilidade dos X-STRs e Y-STRs em situações mais complexas. As revisões bibliográficas dos estudos selecionados demonstraram as variações significativas na diversidade entre as populações, ativadas por fatores históricos, geográficas e demográficos.

11. Referências Bibliográficas

- [1] International Human Genome Sequencing Consortium, "Finishing the euchromatic sequence of the human genome," *Nature*, vol. 431, pp. 931–945, Oct. 2004, doi: 10.1038/nature03001.
- [2] A. Amorim, "Opening the DNA black box: Demythologizing forensic genetics," *New Genet Soc*, vol. 31, no. 3, pp. 259–270, Sep. 2012, doi: 10.1080/14636778.2012.687083.
- [3] H. Ellegren, "Microsatellites: Simple sequences with complex evolution," *Nature Reviews Genetics*, vol. 5, no. 6, pp. 435–445, Jun. 2004, doi: 10.1038/nrg1348.
- [4] Y. Zheng *et al.*, "Forensic efficiency evaluation of a novel multiplex panel of InDels and STRs in the Guizhou Han population and its phylogenetic relationships with other reference populations," *Ann Hum Biol*, vol. 50, no. 1, pp. 42–47, 2023, doi: 10.1080/03014460.2023.2168754.
- [5] I. A. Gutiérrez-Hurtado, C. I. Moreno-Betancur, J. C. González, and R. A. Ruiz-Garcia, "Past, present and future perspectives of forensic genetics," *Biomolecules*, vol. 15, no. 5, p. 713, May 2025, doi: 10.3390/biom15050713.
- [6] H. Ellegren, "Microsatellite mutations in the germline: implications for evolutionary inference," *Trends Genetics*, vol. 16, no. 12, pp. 551–558, Dec. 2000, doi: 10.1016/S0168-9525(00)02139-9.
- [7] P. Gill, L. Fereday, N. Morling, and P. M. Schneider, "The evolution of DNA databases - Recommendations for new European STR loci," *Forensic Sci Int*, vol. 156, no. 2–3, pp. 242–244, Jan. 2006, doi: 10.1016/j.forsciint.2005.05.036.
- [8] J. M. Butler, *Advanced Topics in Forensic DNA Typing: Methodology*, 1st ed. London, UK: Elsevier, 2012.
- [9] M. Oliveira and L. Azevedo, "Molecular Markers: An Overview of Data Published for Fungi over the Last Ten Years," Aug. 01, 2022, *MDPI*. doi: 10.3390/jof8080803.
- [10] S. Li *et al.*, "Genetic diversity of 23 STR loci of the Guizhou Tujia ethnic minority and the phylogenetic relationships with 22 other populations," *Ann Hum Biol*, vol. 50, no. 1, pp. 345–350, 2023, doi: 10.1080/03014460.2023.2224971.

- [11] E. F. Bergseth, A. Tillmar, P. J. T. Haddeland, and D. Kling, “Extended population genetic analysis of 12 X-STRs – Exemplified using a Norwegian population sample,” *Forensic Sci Int Genet*, vol. 60, Sep. 2022, doi: 10.1016/j.fsigen.2022.102745.
- [12] B. Sobrino, M. Brión, and A. Carracedo, “SNPs in forensic genetics: A review on SNP typing methodologies,” *Forensic Sci. Int.*, vol. 154, no. 2–3, pp. 181–194, Nov. 2005, doi: 10.1016/j.forsciint.2004.10.020.
- [13] P. Chen *et al.*, “Genetic diversities and phylogenetic analyses of three Chinese main ethnic groups in southwest China: A Y-Chromosomal STR study,” *Sci Rep*, vol. 8, no. 1, Dec. 2018, doi: 10.1038/s41598-018-33751-x.
- [14] M. R. Eskandarion *et al.*, “Haplotype diversity of 17 Y-STR in the Iranian population,” *BMC Genomics*, vol. 25, no. 1, Dec. 2024, doi: 10.1186/s12864-024-10217-1.
- [15] H. Fan and J. Y. Chu, “A brief review of short tandem repeat mutation,” *Genomics, Proteomics & Bioinformatics*, vol. 5, no. 1, pp. 7–14, Mar. 2007, doi: 10.1016/S1672-0229(07)60009-6.
- [16] A. Pilav, N. Pojskić, A. Ahatović, M. Džehverović, J. Čakar, and D. Marjanović, “Allele frequencies of 15 STR loci in Bosnian and Herzegovinian population,” *Croat Med J*, vol. 58, no. 3, pp. 250–256, 2017, doi: 10.3325/cmj.2017.58.250.
- [17] B. F. B. Algee-Hewitt, M. D. Edge, J. Kim, J. Z. Li, and N. A. Rosenberg, “Individual Identifiability Predicts Population Identifiability in Forensic Microsatellite Markers,” *Current Biology*, vol. 26, no. 7, pp. 935–942, Apr. 2016, doi: 10.1016/j.cub.2016.01.065.
- [18] Z. Červenák, F. Červenák, A. Choma, M. Baldovič, and S. Masnicová, “Population genetic analysis of 12 X-STR markers in Slovakia,” *Ann Hum Biol*, vol. 50, no. 1, pp. 431–435, 2023, doi: 10.1080/03014460.2023.2264769.
- [19] B. M. Albarzinji, F. M. Abdulkarim, S. A. Hussein, D. Rashid, and H. Lazim, “Population genetic study of 17 Y-STR Loci of the Sorani Kurds in the Province of Sulaymaniyah, Iraq,” *BMC Genomics*, vol. 23, no. 1, Dec. 2022, doi: 10.1186/s12864-022-09005-6.

- [20] J. M. Butler, "Genetics and genomics of core short tandem repeat loci used in human identity testing," *J. Forensic Sci.*, vol. 51, no. 2, pp. 253–265, Mar. 2006, doi: 10.1111/j.1556-4029.2006.00046.x.
- [21] N. S. Udogadi, M. K. Abdullahi, A. T. Bukola, O. P. Imose, and A. D. Esewi, "Forensic DNA profiling: Autosomal short tandem repeat as a prominent marker in crime investigation," *Malays. J. Med. Sci.*, vol. 27, no. 4, pp. 19–29, 2020, doi: 10.21315/mjms2020.27.4.3.
- [22] A. L. Edwards, H. A. Hammond, J. Li, C. T. Caskey, and R. Chakraborty, "Genetic variation at five trimeric and tetrameric tandem repeat *loci* in four human population groups," *Am. J. Hum. Genet.*, vol. 50, no. 4, pp. 862–869, Apr. 1992.
- [23] B. Song, J. Fu, J. Qian, L. Yang, J. Cheng, and J. Fu, "Genetic Polymorphism and Population Genetic Structure Analysis of 21 Autosomal STR Loci for a Han-Chinese Population from Luzhou of Southwest China," *Genes (Basel)*, vol. 14, no. 7, Jul. 2023, doi: 10.3390/genes14071419.
- [24] Q. Liu *et al.*, "Genetic polymorphism and forensic efficiency of 21 autosomal STR loci from Shandong Han population in Northern China," *Ann Hum Biol*, vol. 50, no. 1, pp. 341–344, 2023, doi: 10.1080/03014460.2023.2224057.
- [25] E. M. Shepard *et al.*, "Autosomal STR variation in five Austronesian populations," *Hum Biol*, vol. 77, no. 6, pp. 825–851, 2005, doi: 10.1353/hub.2006.0019.
- [26] M. A. Jobling and P. Gill, "Encoded evidence: DNA in forensic analysis," *Nat. Rev. Genet.*, vol. 5, no. 10, pp. 739–751, Oct. 2004, doi: 10.1038/nrg1455.
- [27] B. Budowle, T. R. Moretti, S. J. Niezgoda, and B. L. Brown, "CODIS and PCR-based short tandem repeat loci: Law enforcement tools," in *Second European Symposium on Human Identification*, Madison, WI, USA, Promega Corporation, 1998, pp. 73–88.
- [28] E. R. Mammadov, N. S. Mustafayev, A. C. Mammadov, A. B. Hasanov, and I. M. Huseynova, "Population-genetic and comparative interpopulation studies of Azerbaijan population based on the 15 autosomal STR markers," *Gene*, vol. 753, Aug. 2020, doi: 10.1016/j.gene.2020.144804.
- [29] C. Tomas, V. Pereira, and N. Morling, "Analysis of 12 X-STRs in greenlanders, danes and somalis using argus X-12," *Int J Legal Med*, vol. 126, no. 1, pp. 121–128, 2012, doi: 10.1007/s00414-011-0609-y.

- [30] M. G. García, L. Gusmão, C. I. Catanesi, G. A. Penacino, and N. Pinto, "Mutation rate of 12 X-STRs from investigator Argus X-12 *kit* in Argentine population," *Forensic Sci Int Genet Suppl Ser*, vol. 6, no. September, pp. e562–e564, 2017, doi: 10.1016/j.fsigss.2017.09.219.
- [31] R. Tao *et al.*, "Investigation of 12 X-STR loci in Mongolian and Eastern Han populations of China with comparison to other populations," *Sci Rep*, vol. 8, no. 1, Dec. 2018, doi: 10.1038/s41598-018-22665-3.
- [32] M. W. Perlin *et al.*, "Validating trueallele ® dna mixture interpretation," *J Forensic Sci*, vol. 56, no. 6, pp. 1430–1447, Nov. 2011, doi: 10.1111/j.1556-4029.2011.01859.x.
- [33] D. J. Balding and R. A. Nichols, "DNA profile match probability calculation: How to allow for population stratification, relatedness, database selection and single bands," *Forensic Sci. Int.*, vol. 64, no. 2–3, pp. 125–140, Aug. 1994, doi: 10.1016/0379-0738(94)90221-6.
- [34] J. M. Butler, "The future of forensic DNA analysis," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 370, no. 1674, 2015, doi: 10.1098/rstb.2014.0252.
- [35] T. M. Diegoli, "Forensic typing of short tandem repeat markers on the X and Y chromosomes," *Forensic Sci Int Genet*, vol. 18, pp. 140–151, 2015, doi: 10.1016/j.fsigen.2015.03.013.
- [36] R. Hodişan, D. C. Zaha, C. M. Jurca, S. Irina, and M. Bembea, "Genetic Diversity Based on the Analysis of 27 Y- Short Tandem Repetition (STR) Loci in Two Populations in the Apuseni Mountains, Romania," *Cureus*, Jun. 2024, doi: 10.7759/cureus.62505.
- [37] Y. Luo *et al.*, "Population genetic analysis of 36 Y-chromosomal STRs yields comprehensive insights into the forensic features and phylogenetic relationship of Chinese Tai-Kadai-speaking Bouyei," *PLoS One*, vol. 14, no. 11, Nov. 2019, doi: 10.1371/journal.pone.0224601.
- [38] M. Kayser, "Forensic use of Y-chromosome DNA: A general overview," *Hum. Genet.*, vol. 136, no. 5, pp. 621–635, May 2017, doi: 10.1007/s00439-017-1776-9.

- [39] X. Zhang, H. Zheng, and C. Liu, "Genetic diversity of 23 Y-STR *loci* of the Lisu ethnic minority residing in Chuxiong Yi Autonomous Prefecture, Yunnan province, Southwest China," *Ann Hum Biol*, vol. 50, no. 1, pp. 356–359, 2023, doi: 10.1080/03014460.2023.2224972.
- [40] J. M. Butler, "Constructing STR multiplex assays," in *Forensic DNA Typing Protocols*, W. Goodwin, Ed. Totowa, NJ, USA: Humana Press, 2007, pp. 53–66, doi: 10.1007/978-1-59745-278-9_4.
- [41] Y. Ashirbekov, A. Nogay, A. Abaildayev, A. Zhunussova, Z. Sabitov, and M. Zhabagin, "Genetic polymorphism of 27 Y-STR loci in Kazakh populations from Eastern Kazakhstan," *Ann Hum Biol*, vol. 50, no. 1, pp. 48–51, 2023, doi: 10.1080/03014460.2023.2170465.
- [42] N. Pinto *et al.*, "Paternal and maternal mutations in X-STRs: A GHEP-ISFG collaborative study," *Forensic Sci Int Genet*, vol. 46, no. October 2019, p. 102258, 2020, doi: 10.1016/j.fsigen.2020.102258.
- [43] J. Fu *et al.*, "Genetic Polymorphism Analysis of 24 Y-STRs in a Han Chinese Population in Luzhou, Southwest China," *Genes (Basel)*, vol. 14, no. 10, Oct. 2023, doi: 10.3390/genes14101904.
- [44] S. Füredi, J. Woller, Z. Pádár, and M. Angyal, "Y-STR haplotyping in two Hungarian populations," *Int. J. Legal Med.*, vol. 113, no. 1, pp. 38–42, 1999, doi: 10.1007/s004140050276.
- [45] K. Slooten and F. Ricciardi, "Estimation of mutation probabilities for autosomal STR markers," *Forensic Sci Int Genet*, vol. 7, no. 3, pp. 337–344, 2013, doi: 10.1016/j.fsigen.2013.01.006.
- [46] N. Wyner, M. Barash, and D. McNevin, "Forensic Autosomal Short Tandem Repeats and Their Potential Association With Phenotype," Aug. 06, 2020, *Frontiers Media S.A.* doi: 10.3389/fgene.2020.00884.
- [47] M. F. Khan *et al.*, "Genetic Diversity and Forensic Utility of X-STR Loci in Punjabi and Kashmiri Populations: Insights into Population Structure and Ancestry," *Genes (Basel)*, vol. 15, no. 11, Nov. 2024, doi: 10.3390/genes15111384.
- [48] Y. xin Guo *et al.*, "Genetic polymorphism analyses of a novel panel of 19 X-STR loci in the Chinese Uygur ethnic minority," *J Zhejiang Univ Sci B*, vol. 17, no. 5, pp. 367–374, May 2016, doi: 10.1631/jzus.B1500228.

- [49] J. M. Thompson *et al.*, “Developmental validation of the PowerPlex® Y23 System: A single multiplex Y-STR analysis system for casework and database samples,” *Forensic Sci Int Genet*, vol. 7, no. 2, pp. 240–250, Feb. 2013, doi: 10.1016/j.fsigen.2012.10.013.
- [50] S. Willuweit and L. Roewer, “Y chromosome haplotype reference database (YHRD): Update,” *Forensic Sci Int Genet*, vol. 1, no. 2, pp. 83–87, Jun. 2007, doi: 10.1016/j.fsigen.2007.01.017.
- [51] Federal Bureau of Investigation, “CODIS and NDIS Fact Sheet.” [Online]. Available: <https://www.fbi.gov/how-we-can-help-you/dna-fingerprint-act-of-2005-expungement-policy/codis-and-ndis-fact-sheet>. [Accessed: Jun. 15, 2024]
- [52] A. Iglesias-Ramas, S. P. Lipani, and R. J. Allen, “Population genetics: An introduction for physicists,” *SciPost Phys. Lect. Notes*, vol. 89, 2024, doi: 10.21468/SciPostPhysLectNotes.89
- [53] S. Wright, “The genetical structure of populations,” *Ann. Eugen.*, vol. 15, no. 4, pp. 323–354, Dec. 1951, doi: 10.1111/j.1469-1809.1949.tb02451.x.
- [54] J. L. Watson, K. Cho, K. Grisedale, J. Ward, and D. McNevin, “Characterisation of identity-informative genetic markers in the Australian population with European ancestry,” *Forensic Sci Int Genet*, vol. 74, Jan. 2025, doi: 10.1016/j.fsigen.2024.103169.
- [55] S. Iyavoo *et al.*, “Unveiling genetic diversity and forensic utility of SureID® human DNA identification kits: a comprehensive analysis of 44 autosomal STR loci in English and Irish populations,” *Ann Hum Biol*, vol. 50, no. 1, pp. 399–406, 2023, doi: 10.1080/03014460.2023.2265814.
- [56] S. Chai *et al.*, “Forensic efficiency and population genetic construction of Guizhou Gelao minority from Southwest China revealed by a panel of 23 autosomal STR loci,” *Forensic Sci Res*, vol. 9, no. 2, Jun. 2024, doi: 10.1093/fsr/owad058.
- [57] B. S. Weir and C. C. Cockerham, “Estimating F-statistics for the analysis of population structure,” *Evolution (N Y)*, vol. 38, no. 6, pp. 1358–1370, 1984, doi: 10.1111/j.1558-5646.1984.tb05657.x.
- [58] A. R. Templeton, “Biological races in humans,” *Studies in History and Philosophy of Science Part C :Studies in History and Philosophy of Biological and Biomedical*

Sciences, vol. 44, no. 3, pp. 262–271, Sep. 2013, doi: 10.1016/j.shpsc.2013.04.010.

- [59] P. Gill, “Misleading DNA Evidence: Reasons for Miscarriages of Justice,” *International Commentary on Evidence*, vol. 10, no. 1, Jun. 2017, doi: 10.1515/ice-2014-0010.
- [60] A.J. Jeffreys, V. Wilson, S.L. Thein. “Individual-specific ‘fingerprints’ of human DNA.” *Nature*, vol. 316, pp. 76–79, 1985. doi: 10.1038/316076a0.
- [61] M. Kayser and P. M. Schneider, “DNA-based prediction of human externally visible characteristics in forensics: Motivations, scientific challenges, and ethical considerations,” *Forensic Sci Int Genet*, vol. 3, no. 3, pp. 154–161, Jun. 2009, doi: 10.1016/j.fsigen.2009.01.012.
- [62] P. Gill et al., “DNA commission of the International Society of Forensic Genetics: Recommendations on the evaluation of STR typing results that may include drop-out and/or drop-in using probabilistic methods,” *Forensic Sci. Int. Genet.*, vol. 6, no. 6, pp. 679–688, Nov. 2012, doi: 10.1016/j.fsigen.2012.06.002.
- [63] M. Bottinelli, A. Gouy, S. Utz, and M. Zieger, “Population genetic analysis of 12 X-chromosomal STRs in a Swiss sample,” *Int J Legal Med*, vol. 136, no. 2, pp. 561–563, Mar. 2022, doi: 10.1007/s00414-021-02684-y.
- [64] C. Bini *et al.*, “Haplotype data and forensic evaluation of 23 Y-STR and 12 X-STR loci in eight ethnic groups from Eritrea,” *Int J Legal Med*, vol. 135, no. 2, pp. 449–453, Mar. 2021, doi: 10.1007/s00414-020-02446-2.
- [65] C. Keyser *et al.*, “The ancient Yakuts: A population genetic enigma,” *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 370, no. 1660, Jan. 2015, doi: 10.1098/rstb.2013.0385.
- [66] N. Borbély *et al.*, “High Coverage Mitogenomes and Y-Chromosomal Typing Reveal Ancient Lineages in the Modern-Day Székely Population in Romania,” *Genes (Basel)*, vol. 14, no. 1, Jan. 2023, doi: 10.3390/genes14010133.
- [67] H. Lazim, E. K. Almohammed, S. Hadi, and J. Smith, “Population genetic diversity in an Iraqi population and gene flow across the Arabian Peninsula,” *Sci Rep*, vol. 10, no. 1, Dec. 2020, doi: 10.1038/s41598-020-72283-1.
- [68] E. K. Almohammed, A. Hadi, M. Al-Asmakh, and H. Lazim, “The Qatari population’s genetic structure and gene flow as revealed by the Y chromosome,”

PLoS One, vol. 18, no. 9 September, Sep. 2023, doi: 10.1371/journal.pone.0290844.

- [69] J. Yuan, L. Huang, Y. Yin, and X. Zhang, "Genetic analysis of 23 Y-STR loci in the Va population from Yunnan Province, Southwest China," *Ann Hum Biol*, vol. 50, no. 1, pp. 206–210, 2023, doi: 10.1080/03014460.2023.2197654.

12. Anexos



Caracterização dos
41 artigos .pdf