

RESSALVA

Atendendo solicitação do(a) autor(a), o texto completo deste documento será disponibilizado somente a partir de 19/02/2028.

UNIVERSIDADE ESTADUAL PAULISTA

“Júlio de Mesquita Filho”

INSTITUTO DE BIOCÊNCIAS DE BOTUCATU

Programa de Pós-Graduação em Ciências Biológicas (Genética)

**Painel de marcadores InDel autossômicos para inferência de
ancestralidade humana em genética forense**

João Victor Hamamura Tolo

Botucatu - SP

2026

UNIVERSIDADE ESTADUAL PAULISTA

“Júlio de Mesquita Filho”

INSTITUTO DE BIOCÊNCIAS DE BOTUCATU

Programa de Pós-Graduação em Ciências Biológicas (Genética)

Painel de marcadores InDel autossômicos para inferência de ancestralidade humana em genética forense

João Victor Hamamura Toloi

Orientador: Prof. Dr. Erick da Cruz Castelli

Dissertação apresentada ao Instituto de Biociências de Botucatu, Universidade Estadual Paulista “Júlio de Mesquita Filho”, Câmpus de Botucatu, para obtenção do título de Mestre em Ciências Biológicas (Genética).

Botucatu - SP

2026

T653p

Toloi, João Victor Hamamura

Painel de marcadores InDel autossômicos para inferência de ancestralidade humana em genética forense / João Victor Hamamura

Toloi. -- , 2026

77 p. : il., tabs.

Dissertação (mestrado) - Universidade Estadual Paulista (UNESP), Instituto de Biociências, Botucatu,

Orientador: Erick da Cruz Castelli

1. Ancestralidade. 2. InDels. 3. Genética forense. 4. Bioinformática. 5. Micro-haplótipos. I. Título.

ATA DA DEFESA PÚBLICA DA DISSERTAÇÃO DE MESTRADO DE JOÃO VICTOR HAMAMURA TOLOI, DISCENTE DO PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS BIOLÓGICAS (GENÉTICA), DO INSTITUTO DE BIOCIÊNCIAS - CÂMPUS DE BOTUCATU.

Aos 19 de fevereiro de 2026, às 8h30min, por meio de Videoconferência, realizou-se a defesa de DISSERTAÇÃO DE MESTRADO de JOÃO VICTOR HAMAMURA TOLOI, intitulada "Painel de marcadores InDel autossômicos para inferência de ancestralidade humana em genética forense". A Comissão Examinadora foi constituída pelos seguintes membros: Prof. Dr. ERICK DA CRUZ CASTELLI (Orientador(a) - Participação Virtual) do(a) Departamento de Patologia / UNESP / Câmpus de Botucatu - FMB, Prof. Dr. CELSO TEIXEIRA MENDES JÚNIOR (Participação Virtual) do(a) Departamento de Química / FFCL/Ribeirão Preto - Universidade de São Paulo, Profa. Dra. SILVIENE FABIANA DE OLIVEIRA (Participação Virtual) do(a) Departamento de Genética e Morfologia / Universidade de Brasília, Após a exposição pelo mestrando e arguição pelos membros da Comissão Examinadora que participaram do ato, de forma presencial e/ou virtual, o discente recebeu o conceito final: aprovado . Nada mais havendo, foi lavrada a presente ata, que após lida e aprovada, foi assinada pelo(a) Presidente(a) da Comissão Examinadora.

Documento assinado digitalmente
 **ERICK DA CRUZ CASTELLI**
Data: 08/05/2026 10:43:09 -0300
Verifique em <https://validar.itb.gov.br>

Prof. Dr. ERICK DA CRUZ CASTELLI

Dedicatória

Dedico esta dissertação à minha família, fonte inesgotável de apoio, amor e compreensão ao longo desta jornada. A vocês, que sempre estiveram ao meu lado, meu profundo agradecimento por cada incentivo e sacrifício que tornaram possível este momento. Ao meu orientador, pela orientação sábia e pela paciência incansável, meu sincero reconhecimento. A todos os que de alguma forma contribuíram para este trabalho, meu eterno agradecimento. Este trabalho é dedicado a vocês, com gratidão e carinho.

Agradecimentos

✿ À minha família, pelo apoio incondicional, paciência e amor ao longo dos anos.

✿ Ao meu orientador Erick, pela orientação perspicaz e dedicação incansável

✿ Aos membros da banca examinadora, pelo tempo dedicado à avaliação deste estudo e pelos valiosos comentários fornecidos.

✿ Aos colegas de curso, pelo compartilhamento de ideias, discussões estimulantes e apoio mútuo ao longo desta jornada acadêmica.

✿ Aos amigos que estiveram ao meu lado, oferecendo suporte emocional e encorajamento nos momentos desafiadores.

✿ A Unesp e seus funcionários, pela infraestrutura e recursos que foram essenciais para a realização desta pesquisa.

✿ A todas as pessoas que participaram das entrevistas e colaboraram com informações importantes para este estudo.

✿ A CAPES e Fapesp, pelo suporte financeiro que viabilizou a realização deste projeto.

✿ A todos aqueles que, de alguma forma, contribuíram para o sucesso deste trabalho, meu sincero agradecimento.

Lista de figuras

Figura 1. Valores de F_{ST} entre os grupos AFR, AMR, EAS, EUR, NAT e SAS. Representação dos valores de F_{ST} como sendo a intensidade da coloração vermelha. Cada heatmap apresenta dois marcadores espelhados, com o objetivo de otimizar o espaço da figura. Os nomes dos marcadores estão indicados nos retângulos azuis.

53

Figura 2. Gráfico de barras representando as proporções de ancestralidade inferidas para cada indivíduo com base no output do *Structure* do painel 83 SNPs AIMs.

59

Figura 3. Gráfico de barras representando as proporções de ancestralidade inferidas para cada indivíduo com base no output do *Structure* do painel GeMBio.

59

Figura 4. Gráfico de barras representando as proporções de ancestralidade inferidas para cada indivíduo com base no output do *Structure* do painel GeMBio + 16 AIMs.

60

Figura 5. Análise de componentes principais baseada nos 83 marcadores SNPs AIMs presentes nos Painéis SNPforID 34-plex, *Eurasiaplex* e PIMA, considerando populações europeias (laranja), leste-asiáticas (verde-musgo), sul-asiáticas (verde-esmeralda), africanas (azul), e nativo-americanas (roxo) presentes nos bancos de dados 1KGEN e HGDP.

61

Figura 6. Análise de componentes principais baseada nos 83 marcadores SNPs AIMs presentes nos Painéis SNPforID 34-plex, *Eurasiaplex* e PIMA, considerando populações europeias (laranja), leste-asiáticas (verde-musgo), sul-asiáticas (verde-esmeralda), africanas (azul), e nativo-americanas (roxo) presentes nos bancos de dados 1KGEN e HGDP.

62

Figura 7. Análise de componentes principais baseada nos 28 marcadores de identificação humana, considerando populações europeias (laranja), leste-asiáticas (verde-musgo), sul-asiáticas (verde-esmeralda), africanas (azul), e nativo-americanas (roxo) presentes nos bancos de dados 1KGEN e HGDP.

62

Figura 8. Análise de componentes principais baseada nos 28 marcadores de identificação humana, considerando populações europeias (laranja), leste-asiáticas (verde-musgo), sul-asiáticas (verde-esmeralda), africanas (azul), e nativo-americanas (roxo) presentes nos bancos de dados 1KGEN e HGDP.

63

Figura 9. Análise de componentes principais baseada nos 28 marcadores de identificação humana em conjunto com os 16 AIMs selecionados, considerando populações europeias (laranja), leste-asiáticas (verde-musgo), sul-asiáticas (verde-esmeralda), africanas (azul), e nativo-americanas (roxo) presentes nos bancos de dados 1KGEN e HGDP.

63

Figura 10. Análise de componentes principais baseada nos 28 marcadores de identificação humana em conjunto com os 16 AIMs selecionados, considerando populações europeias (laranja), leste-asiáticas (verde-musgo), sul-asiáticas (verde-esmeralda), africanas (azul), e nativo-americanas (roxo) presentes nos bancos de dados 1KGEN e HGDP.

64

Figura 11. Distribuição dos 44 marcadores no painel conjunto (GeMBio + 16 AIMs). Cada faixa representa o tamanho dos alelos de cada marcador. Os tamanhos são representados pelo eixo x e a cor do fluoróforo pelo eixo y.

65

Lista de tabelas

Tabela 1. Distribuição das populações analisadas, bancos de dados de origem e número de indivíduos por grupo continental.	43
Tabela 2. Frequências alélicas dos 16 AIMs nos grupos populacionais continentais AFR, AMR, EAS, EUR, NAT, SAS.	46
Tabela 3. Frequências alélicas dos 28 marcadores de identificação humana nos grupos populacionais continentais AFR, AMR, EAS, EUR, NAT, SAS.	48
Tabela 4. Valores máximos do grau de correlação (r^2) observado entre pares de marcadores localizados no mesmo cromossomo.	52
Tabela 5. Parâmetros estatísticos de F_{ST} , I_n , A_e , PM , PD e PE dos 16 AIMs selecionados considerando as populações africanas, europeias, leste-asiáticas, nativo-americanas e sul-asiáticas presentes nos bancos de dados 1KGEN e HGDP.	53
Tabela 6. Parâmetros estatísticos de F_{ST} , I_n , A_e , PM , PD e PE dos 28 marcadores de identificação humana considerando as populações africanas, europeias, leste-asiáticas, nativo-americanas e sul-asiáticas presentes nos bancos de dados 1KGEN e HGDP.	54
Tabela 7. Comparação entre os parâmetros de F_{ST} , I_n , A_e , PMC e PEC entre diferentes conjuntos de marcadores.	56
Tabela 8. Proporções médias de ancestralidade estimadas pelo Structure a cada população com base nos marcadores do painel de referência 83 SNPs AIMs (SNPforID 34-plex, Eurasiaplex e PIMA) ($K = 5$).	59
Tabela 9. Proporções médias de ancestralidade estimadas pelo Structure a partir do painel de 28 micro-haplótipos de identificação humana ($K = 5$).	59

Tabela 10. Proporções médias de ancestralidade estimadas pelo Structure com o painel combinado de 28 micro-haplótipos de identificação humana e os 16 AIMS selecionados ($K = 5$).

60

Tabela 11. Comparação das proporções de ancestralidade por indivíduo (média $\pm$ desvio padrão) estimadas pelo STRUCTURE entre os painéis de 83 SNPs AIMS, kit GeMBio e kit GeMBio + 16 AIMS nos grupos parentais AFR, EAS, EUR, NAT e SAS.

61

Tabela 12. Comparação das proporções de ancestralidade por indivíduo (média $\pm$ desvio padrão) estimadas pelo STRUCTURE entre os painéis de 83 SNPs AIMS, kit GeMBio e kit GeMBio + 16 AIMS nos grupos miscigenados ACB, ASW, CLM, MXL, PEL e PUR.

62

Lista de abreviaturas e siglas

Ae – Número Efetivo de Alelos

AIMs – *Ancestry Informative Markers*

AFR – *Africa*

AMR – *America*

BAM – *Binary Alignment Map*

BED – *Browser Extensible Data*

bp – *Base Pairs*

PMC – Probabilidade de *match* combinada

PEC – Poder de exclusão combinado

DIPs – *Deletion-Insertion Polymorphisms*

DNA – Ácido Desoxirribonucleico

EAS – *East Asia*

EUR – *Europe*

F_{ST} – *Fixation Index*

HW – *Hardy-Weinberg*

IGV – *Integrative Genomics Viewer*

InDels – *Insertions and Deletions*

K – Número de *Clusters* Genéticos

LD – *Linkage Disequilibrium*

NGS – *Next-Generation Sequencing*

PCA – *Principal Component Analysis*

PCR – *Polymerase Chain Reaction*

PD – *Power of Discrimination*

PE – *Probability of Exclusion*

PM – *Probability of Match*

RFLP – *Restriction Fragment Length Polymorphism*

SAS – *South Asia*

SNPs – *Single Nucleotide Polymorphisms*

STRs – *Short Tandem Repeats*

VNTR – *Variable Number Tandem Repeat*

VRFLP – *Variable Restriction Fragment Length Polymorphism*

VCF – *Variant Call Format*

Sumário

RESUMO	13
ABSTRACT	15
Capítulo I - Revisão da Literatura	16
RESUMO - Capítulo I	17
ABSTRACT - Chapter I	18
1. Introdução	19
1.1 Evolução Histórica da Genética Forense	19
1.2 Marcadores na Genética Forense para identificação humana	19
1.3 micro-haplótipos de InDels	20
1.4 AIMs em análises forenses de identificação humana	21
1.5 A População Brasileira e a Relevância do Estudo de Ancestralidade	22
1.6 Uso de Dados Genômicos	23
2. Justificativa	24
3. Objetivos	25
Referências	26
Capítulo II - Artigo	31
RESUMO - Capítulo II	32
ABSTRACT - Chapter II	33
1. Introdução	34
2. Materiais e Métodos	35
2.1 Amostras populacionais	35
2.2 Pré-processamento dos dados genômicos	35
2.3 Re-genotipagem e haplotipagem dos marcadores candidatos	36
2.4 Seleção inicial dos micro-haplótipos para análise de ancestralidade	37
2.5 Seleção final e montagem de primers	39
2.6 Parâmetros Forenses	40
3. Resultados	41
3.1 Amostra populacional	41
3.2 Regiões multi-InDels encontradas	43
3.3 Frequência alélica dos 16 AIMs	44
3.4 Desequilíbrio de ligação	49
3.5 Parâmetros forenses e populacionais dos marcadores selecionados	51
3.6 Análise de FST pareado entre grupos populacionais	54
3.7 Capacidade dos marcadores em identificar o perfil de ancestralidade genômica de cada amostra	56
3.8 Análise de componentes principais	63
3.9 Distribuição dos marcadores no painel combinado	66
3.10 Equilíbrio de Hardy-Weinberg	67
4. Discussão	68
5. Conclusão	72
6. Referências	73

RESUMO

A genética forense evoluiu bastante ao longo do século XX, passando de marcadores de grupos sanguíneos a técnicas modernas de tipagem de DNA. Nesse percurso, os VNTRs deram lugar aos STRs, que se tornaram o padrão-ouro da identificação humana. Com o avanço das tecnologias de sequenciamento, SNPs e InDels passaram a ser explorados como complementos ou alternativas aos STRs. Os SNPs, embora informativos, são bialélicos e sua análise por eletroforese capilar (EC) é bastante complexa, método consolidado nos laboratórios forenses. Os InDels, por sua vez, combinam compatibilidade com a EC, baixa ocorrência de artefatos e bom desempenho em amostras degradadas. Quando múltiplos InDels próximos são analisados em conjunto, formando micro-haplótipos (multi-InDels), obtêm-se perfis multi-alélicos com maior poder de discriminação individual e maior resolução populacional, características especialmente relevantes para a inferência de ancestralidade genômica. A inferência de ancestralidade é uma ferramenta complementar nas investigações forenses, permitindo estimar a origem étnica de um indivíduo a partir de vestígios biológicos, o que pode orientar a busca por suspeitos quando não há perfis para comparação. Para isso, utilizam-se os chamados AIMs (*Ancestry Informative Markers*), marcadores selecionados com base em métricas como o F_{ST} , um índice que quantifica a diferenciação alélica entre populações. No contexto brasileiro, marcado por elevada miscigenação, com contribuições europeias, africanas, indígenas e asiáticas, a necessidade de painéis de AIMs adaptados a essa diversidade torna-se particularmente relevante. Partindo desse cenário, este trabalho teve como objetivo desenvolver um novo painel de marcadores InDels, únicos ou formando micro-haplótipos, voltado para a inferência de ancestralidade genômica. Em seguida, integrá-lo ao kit forense de identificação humana GeMBio, composto por 28 marcadores multi-InDels autossômicos atualmente em desenvolvimento. A estratégia metodológica baseou-se na análise computacional dos dados do Projeto 1000 Genomas, utilizando *scripts* personalizados para identificar regiões candidatas com padrões de frequência alélica altamente diferenciados entre populações africanas, europeias, americanas, leste-asiáticas e sul-asiáticas. Após etapas de pré-processamento, genotipagem e avaliação de diversos parâmetros, foram selecionados 16 AIMs. A avaliação de desempenho comparou três painéis: o painel de referência com 83 SNPs AIMs (SNPforID 34-Plex, Eurasiaplex e PIMA), o kit GeMBio com seus 28 marcadores de identificação humana e o painel combinado GeMBio + 16 AIMs selecionados. A inferência de ancestralidade foi realizada com o *software Structure*. Os resultados demonstraram que o kit GeMBio, embora eficaz na identificação humana, apresentou limitações na distinção entre indivíduos europeus e sul-asiáticos. A incorporação dos 16 novos AIMs aprimorou significativamente a precisão das estimativas, aproximando o desempenho do painel combinado ao do conjunto de 83 SNPs de referência. Os resultados reforçam que os InDels, isolados ou organizados em micro-haplótipos, possuem potencial real para a inferência de ancestralidade genômica em contextos forenses, oferecendo informações biogeográficas detalhadas compatíveis com os fluxos de trabalho laboratoriais já estabelecidos. A integração desses marcadores ao kit GeMBio representa um avanço concreto no desenvolvimento de ferramentas forenses adaptadas à complexidade genética da população brasileira.

Palavras-chave: ancestralidade, InDels, genética forense, bioinformática, micro-haplótipos

ABSTRACT

Forensic genetics has undergone substantial development over the last century, evolving from blood group markers to modern DNA typing technologies. During this process, VNTRs were progressively replaced by STRs, which became the standard markers for human identification. More recently, advances in sequencing technologies have expanded the use of SNPs and InDels as complementary or alternative forensic markers. Although SNPs are highly informative, their biallelic nature and the complexity of capillary electrophoresis-based analysis limit their routine forensic application. InDels, in contrast, are compatible with capillary electrophoresis workflows, generate fewer amplification artifacts, and perform well with degraded DNA samples. When closely linked InDels are analyzed jointly as microhaplotypes (multi-InDels), they produce multiallelic profiles with increased discriminatory power and improved population resolution, features that are particularly relevant for genomic ancestry inference. Ancestry inference represents an important auxiliary tool in forensic investigations, allowing the estimation of an individual's biogeographical origin from biological evidence when reference profiles are unavailable. This approach relies on ancestry informative markers (AIMs), selected according to population differentiation metrics such as F_{ST} . In highly admixed populations, such as the Brazilian population, which presents European, African, Native American, and Asian genetic contributions, the development of ancestry panels adapted to this complex demographic background is especially important. In this context, the present study aimed to develop a novel panel of InDel markers, either isolated or organized as microhaplotypes, for genomic ancestry inference and to integrate these markers into the GeMBio forensic human identification kit, currently composed of 28 autosomal multi-InDel markers. Candidate regions were identified through computational analyses of data from the 1000 Genomes Project using custom scripts designed to detect loci with highly differentiated allele frequency distributions among African, European, American, East Asian, and South Asian populations. After preprocessing, genotyping, and marker evaluation steps, 16 AIMs were selected. Performance analyses compared three marker sets: a reference panel containing 83 AIM-SNPs (SNPforID 34-Plex, Eurasiaplex, and PIMA), the GeMBio kit with its 28 identification markers, and the combined GeMBio + 16 AIMs panel. Ancestry inference analyses were performed using the STRUCTURE software. The results demonstrated that, although the GeMBio kit showed satisfactory performance for human identification, it presented limitations in distinguishing European and South Asian individuals. The incorporation of the 16 selected AIMs substantially improved ancestry estimation accuracy, resulting in performance comparable to that obtained with the 83-SNP reference panel. These findings demonstrate that InDels, either individually or organized as microhaplotypes, represent promising markers for forensic ancestry inference, providing detailed biogeographical information while remaining compatible with established forensic laboratory workflows. The integration of these markers into the GeMBio kit represents an important step toward the development of forensic tools adapted to the genetic complexity of the Brazilian population.

Keywords: ancestry inference, InDels, forensic genetics, bioinformatics, microhaplotypes.

Capítulo I - Revisão da Literatura

RESUMO - Capítulo I

A genética forense evoluiu significativamente desde o início do século XX, com marcos como a descoberta dos grupos sanguíneos ABO e, posteriormente, a estrutura do DNA, que pavimentaram o caminho para as modernas técnicas de identificação genética. Inicialmente, as análises forenses baseavam-se em VNTRs (*Variable Number Tandem Repeats*), que, embora informativos, exigiam grandes quantidades de DNA e apresentavam limitações na análise de regiões degradadas. A migração para STRs (*Short Tandem Repeats*) trouxe maior eficiência, pois STRs são mais curtos, amplificam-se mais facilmente em DNA degradado e geram perfis altamente discriminatórios com menor quantidade de amostra. Atualmente, além dos STRs, marcadores como SNPs e InDels são utilizados conforme a necessidade da análise. Os SNPs, embora informativos e compatíveis com DNA degradado, têm baixo poder discriminatório por serem bialélicos. Os InDels combinam as vantagens dos SNPs com a detecção via eletroforese capilar típica dos STRs. Nesse contexto, os micro-haplótipos de InDels surgem como uma alternativa relevante. Essas regiões curtas do DNA, compostas por múltiplos InDels próximos, oferecem perfis multi-alélicos, compatibilidade com amostras degradadas e menor taxa de erro, refletindo uma menor ocorrência de artefatos ou inconsistências na genotipagem, além de menor taxa de mutação em comparação aos STRs. Por apresentarem múltiplas combinações haplotípicas, os micro-haplótipos multialélicos tendem a oferecer maior poder de discriminação individual, por outro lado, marcadores bialélicos apresentaram melhor desempenho em análises de ancestralidade, favorecendo a diferenciação entre populações com base em seus perfis genéticos. Sua capacidade de diferenciar populações os torna especialmente úteis como AIMs (*Ancestry Informative Markers*). A inferência de ancestralidade é relevante na genética forense, sobretudo quando não há correspondência entre os perfis obtidos a partir de vestígios e os perfis de suspeitos ou de indivíduos presentes em bancos de dados. Os AIMs podem ser escolhidos com base em métricas como o F_{ST} , que quantifica a diferença alélica entre populações. No caso específico do Brasil, a elevada miscigenação, envolvendo contribuições europeias, africanas, indígenas e, mais recentemente, asiáticas, torna essencial o uso de marcadores adaptados à diversidade local. A complexidade genômica da população brasileira evidencia a necessidade de desenvolver e validar painéis de AIMs Multi-InDels representativos dessa realidade, garantindo precisão nas análises forenses e ampliando as possibilidades de investigação em um país de alta heterogeneidade genética.

Palavras-chave: ancestralidade, InDels, genética forense, micro-haplótipos

ABSTRACT - Chapter I

Forensic genetics has evolved significantly since the early twentieth century, with milestones such as the discovery of the ABO blood groups and, later, the elucidation of DNA structure, which paved the way for modern genetic identification techniques. Initially, forensic analyses were based on VNTRs (Variable Number Tandem Repeats), which, although informative, required large amounts of DNA and showed limitations in the amplification of degraded regions. The transition to STRs (Short Tandem Repeats) brought greater efficiency, as STRs are shorter, amplify more readily from degraded DNA, and generate highly discriminatory profiles using smaller amounts of sample. Currently, in addition to STRs, markers such as SNPs and InDels are used according to analytical needs. SNPs, although informative and compatible with degraded DNA, have low discriminatory power because they are biallelic. InDels combine the advantages of SNPs with detection by capillary electrophoresis, as is typical for STRs. In this context, InDel microhaplotypes have emerged as a relevant alternative. These short DNA regions, composed of multiple closely spaced InDels, provide multiallelic profiles, compatibility with degraded samples, and lower error and mutation rates compared with STRs. By presenting multiple haplotypic combinations, microhaplotypes increase individual discriminatory power and population resolution, enhancing the ability to distinguish both individuals and populations based on their genetic profiles. Their capacity to differentiate populations makes them particularly useful as ancestry informative markers (AIMs). Ancestry inference is highly relevant in forensic genetics, especially when there is no match between profiles obtained from biological evidence and those of suspects or individuals in databases. AIMs can be selected based on metrics such as F_{ST} , which quantifies allelic differentiation among populations. In the specific case of Brazil, the high level of admixture between European, African, Indigenous, and more recent Asian contributions, makes it essential to use markers adapted to local diversity. The genomic complexity of the Brazilian population highlights the need to develop and validate panels of multi-InDel AIMs that are representative of this reality, ensuring accuracy in forensic analyses and expanding investigative possibilities in a country with high genetic heterogeneity.

Keywords: ancestry, InDels, forensic genetics, microhaplotypes

1. Introdução

1.1 Evolução Histórica da Genética Forense

A genética forense é um ramo da genética que aplica técnicas moleculares à resolução de questões legais, especialmente na identificação de indivíduos e na investigação criminal. Essa área surgiu a partir de marcos históricos importantes que, ao longo do século XX, estabeleceram os fundamentos científicos necessários para a análise forense do DNA⁽¹⁾.

O primeiro grande avanço ocorreu no início do século XX, com a descoberta dos grupos sanguíneos ABO por Karl Landsteiner em 1900, que possibilitou a individualização biológica com base em características genéticas⁽²⁾. Décadas mais tarde, a estrutura do DNA foi apresentada por James Watson e Francis Crick em 1953, um marco fundamental que permitiu a compreensão da herança molecular e pavimentou o caminho para a biologia molecular moderna⁽³⁾. A partir da década de 1980, a genética forense foi revolucionada pela introdução de métodos de análise direta do DNA, como o polimorfismo no comprimento de fragmentos de restrição (RFLP) e a reação em cadeia da polimerase (PCR)^(4,5).

Alec Jeffreys foi responsável por aplicar, pela primeira vez, técnicas de *DNA fingerprinting* baseadas em regiões altamente polimórficas, conhecidas como VNTRs (*Variable Number of Tandem Repeats*). Esses *loci* consistem em repetições de unidades de aproximadamente 10 a 100 pares de bases, cuja variação no número de repetições entre indivíduos gera perfis genéticos altamente específicos. Apesar de seu alto poder discriminatório, os VNTRs exigem grandes quantidades de DNA de boa qualidade (muitas vezes indisponível em casos forenses), são difíceis de amplificar por PCR devido ao tamanho das repetições e geram análises estatísticas mais complexas, uma vez que o número exato de unidades repetitivas nem sempre é claramente definido^(4,6).

A limitação dos VNTRs motivou a transição para os STRs (*Short Tandem Repeats*), marcadores que consistem em repetições curtas de 2 a 6 pares de bases, mais adequados para amplificação por PCR devido ao tamanho reduzido dos amplicons, o que aumenta o rendimento da PCR em amostras degradadas ou com baixa quantidade de DNA. Embora os STRs em geral sejam mais propensos a artefatos de *stutter*, os *loci* escolhidos para aplicações forenses são predominantemente tetra nucleotídeos, apresentando menor incidência de *stutter* em comparação com dinucleotídeos ou trinucleotídeos. Todas essas características contribuíram para que os STRs substituíssem os VNTRs como padrão-ouro na genética forense, consolidando sua adoção global em bancos de dados e em protocolos de identificação. (4,6,7).

Referências

1. BUTLER, John M. *Forensic DNA typing: biology, technology, and genetics of STR markers*. 2. ed. San Diego: Elsevier Academic Press, 2005.
2. LANDSTEINER, Karl. Zur Kenntnis der antifermentativen, lytischen und agglutinierenden Wirkungen des Blutserums und der Lymphe. *Zentralblatt für Bakteriologie*, v. 27, p. 357–362, 1900.
3. WATSON, James D.; CRICK, Francis H. C. A structure for deoxyribose nucleic acid. *Nature*, Londres, v. 171, n. 4356, p. 737–738, 1953.
4. JEFFREYS, Alec J.; WILSON, Victoria; THEIN, S. L. Hypervariable 'minisatellite' regions in human DNA. *Nature*, v. 314, n. 6006, p. 67–73, 1985.
5. MULLIS, Kary; FALOONA, Frederick. Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. *Methods in Enzymology*, v. 155, p. 335–350, 1987.
6. GILL, Peter; JEFFREYS, Alec J.; WERRETT, David J. Forensic application of DNA 'fingerprints'. *Nature*, v. 318, n. 6046, p. 577–579, 1985.

7. BUDOWLE, Bruce et al. Population data on the thirteen CODIS core short tandem repeat loci in African Americans, U.S. Caucasians, Hispanics, Bahamians, Jamaicans, and Trinidadians. *Journal of Forensic Sciences*, v. 44, n. 6, p. 1277–1286, 1999.
8. GOODWIN, William; LINACRE, Adrian; HADI, Sibte. *An introduction to forensic genetics*. 2. ed. Oxford: Wiley-Blackwell, 2007.
9. ALONSO, Antonio et al. Real-time PCR designs to estimate nuclear and mitochondrial DNA copy number in forensic and ancient DNA studies. *Forensic Science International*, v. 139, n. 2–3, p. 141–149, 2004.
10. THE 1000 GENOMES PROJECT CONSORTIUM. A global reference for human genetic variation. *Nature*, London, v. 526, n. 7571, p. 68–74, 2015. DOI: 10.1038/nature15393.
11. PEREIRA, R. et al. A new multiplex for human identification using insertion/deletion polymorphisms. *Electrophoresis*, v. 30, n. 21, p. 3682–3690, 2009.
12. KIDD, K. K. et al. Progress toward an efficient panel of SNPs for ancestry inference. *Forensic Science International: Genetics*, v. 10, p. 23–32, 2014.
13. KIDD, K. K. et al. Microhaplotype loci are a powerful new type of forensic marker. *Forensic Science International: Genetics Supplement Series*, v. 4, n. 1, p. e123–e124, 2013.
14. OLDONI, F.; KIDD, K. K.; PODINI, D. Microhaplotypes in forensic genetics. *Forensic Science International: Genetics*, v. 38, p. 54–69, 2019.
15. BØRSTING, C.; MORLING, N. Next generation sequencing and its applications in forensic genetics. *Forensic Science International: Genetics*, v. 18, p. 78–89, 2015.
16. RAMOS, L. C. et al. Assessing the performance of multi-InDel panels for human identification among admixed Brazilians. *Forensic Science International: Genetics*, [S.l.], v. 74, p. 103161, 2025. ISSN 1872-4973. <https://doi.org/10.1016/j.fsigen.2024.103161>.
17. LIU, J. et al. A new set of 20 Multi-InDel markers for forensic application. *Electrophoresis*, v. 43, n. 11, p. 1193–1202, 2022. DOI: <https://doi.org/10.1002/elps.202100361>.
18. LIU, J. et al. DIP-microhaplotypes: new markers for detection of unbalanced DNA mixtures. *International Journal of Legal Medicine*, v. 135, p. 13–21, 2021.

19. STAADIG, M.; TILLMAR, A.; NILSSON, M.; ALLEN, M. Evaluation of microhaplotypes for forensic analysis in a Swedish population. *International Journal of Legal Medicine*, v. 134, n. 6, p. 2103–2113, 2020.
20. PHILLIPS, C. Forensic genetic analysis of bio-geographical ancestry. *Forensic Science International: Genetics*, v. 18, p. 49–65, 2015.
21. WU, Riga; LI, Haixia; LI, Ran; PENG, Dan; WANG, Nana; SHEN, Xuefeng; SUN, Hongyu. Identification and sequencing of 59 highly polymorphic microhaplotypes for analysis of DNA mixtures. *International Journal of Legal Medicine*, [s. l.], v. 135, p. 1161-1173, 2021. DOI: <https://doi.org/10.1007/s00414-020-02485-0>
22. ROSENBERG, N. A. et al. Informativeness of genetic markers for inference of ancestry. *The American Journal of Human Genetics*, v. 73, n. 6, p. 1402–1422, 2003.
23. TOZZO, P. et al. External visible characteristics prediction through SNPs analysis in the forensic setting: a review. *Frontiers in Bioscience (Landmark Edition)*, 2021. DOI: 10.52586/4991.
24. KIDD, K. K.; SPEED, W. C. Criteria for selecting microhaplotypes: mixture detection and ancestry inference. *Forensic Science International: Genetics Supplement Series*, v. 5, p. e512–e514, 2015.
25. QU, Shengqiu et al. Multi-Indel: A microhaplotype marker can be typed using capillary electrophoresis platforms. *Frontiers in Genetics*, [S.l.], v. 11, p. 567082, 23 out. 2020. DOI: 10.3389/fgene.2020.567082. Disponível em: <https://doi.org/10.3389/fgene.2020.567082>.
26. XUE, Jiaming et al. MHBBase: A comprehensive database of short microhaplotypes for advancing forensic genetic analysis. *Forensic Science International: Genetics*, [S.l.], v. 71, p. 103062, jul. 2024. DOI: 10.1016/j.fsigen.2024.103062. Epub 17 maio 2024. Disponível em: <https://doi.org/10.1016/j.fsigen.2024.103062>.
27. WU, H. et al. Development and evaluation of an InDel-based microhaplotype panel for forensic ancestry inference. *Frontiers in Genetics*, v. 12, p. 673948, 2021.
28. BENNETT, L. et al. Microhaplotype markers: A new forensic DNA tool for ancestry inference and beyond. *Genes*, v. 11, n. 8, p. 887, 2020.

29. Cheung E. Y. Y, et al. Performance of ancestry-informative SNP and microhaplotype markers. *Forensic Sci Int Genet.* 2019 Nov;43:102141. doi: 10.1016/j.fsigen.2019.102141. Epub 2019 Aug 8. PMID: 31442930.
30. ZHANG, S. et al. Genetic structure and forensic features of Chinese Han population using a panel of 45 microhaplotypes. *Forensic Science International: Genetics*, v. 48, p. 102361, 2020.
31. PAKSTIS, A. J. et al. Increasing the discrimination power of ancestry inference using microhaplotypes. *Genes*, v. 10, n. 3, p. 228, 2019.
32. PENA, S. D. J. et al. The genomic ancestry of individuals from different geographical regions of Brazil is more uniform than expected. *PLoS ONE*, v. 6, n. 2, e17063, 2011.
33. MANTA, F. S. N. et al. Revisiting the genetic ancestry of Brazilians using autosomal AIMS. *American Journal of Human Biology*, v. 25, n. 6, p. 915–917, 2013.
34. KEHDY, F. S. G. et al. Origin and dynamics of admixture in Brazilians and its effect on the pattern of deleterious mutations. *Proceedings of the National Academy of Sciences of the United States of America*, v. 112, n. 28, p. 8696–8701, 2015.
35. NASLAVSKY, M. S. et al. Whole-genome sequencing of 1,171 elderly admixed individuals from São Paulo, Brazil. *Nature Communications*, v. 13, n. 1, p. 100, 2022. DOI: 10.1038/s41467-022-28648-3.
36. RAMOS, Bruna Ribeiro de Andrade et al. Neither self-reported ethnicity nor declared family origin are reliable indicators of genomic ancestry. *Genetica, Dordrecht*, v. 144, n. 3, p. 259–265, 2016. DOI: 10.1007/s10709-016-9894-1.
37. BOND, J. W. Value of DNA evidence in detecting crime. *Journal of Forensic Sciences*, v. 52, n. 1, p. 128–136, 2007.
38. JOBLING, M. A.; GILL, P. Encoded evidence: DNA in forensic analysis. *Nature Reviews Genetics*, v. 5, n. 10, p. 739–751, 2004.
39. SUN, K.; YE, Y.; LUO, T.; HOU, Y. Multi-InDel Analysis for Ancestry Inference of Sub-Populations in China. *Scientific Reports*, v. 6, p. 39797, 2016.
40. HUANG, J.; LUO, H.; WEI, W.; HOU, Y. A novel method for the analysis of 20 multi-Indel polymorphisms and its forensic application. *Electrophoresis*, v. 35, n. 4, p. 487–493, 2014.

41. WEN, Y. et al. Forensic biogeographical ancestry inference: recent insights and current trends. *Genes & Genomics*, 2023. DOI: 10.1007/s13258-023-01387-5.
42. KAYSER, M. et al. Recent advances in Forensic DNA Phenotyping of appearance, ancestry and age. *Forensic Science International: Genetics*, 2023. DOI: 10.1016/j.fsigen.2023.102870.
43. RAMOS, Livia Carla et al. *Assessing the performance of multi-InDel panels for human identification among admixed Brazilians*. *Forensic Science International: Genetics*, Amsterdam, v. 74, p. 103161, 2025. DOI: 10.1016/j.fsigen.2024.103161.
44. Fairley, S. et al. The International Genome Sample Resource (IGSR) collection of open human genomic variation resources. 2020. *Nucleic Acids Research*. Volume 48, Issue D1, 08 January, Pages D941–D947, <https://doi.org/10.1093/nar/gkz836>
45. PRITCHARD, Jonathan K.; STEPHENS, Matthew; DONNELLY, Peter. Inference of population structure using multilocus genotype data. *Genetics*, [S.l.], v. 155, n. 2, p. 945–959, 2000. Disponível em: <https://web.stanford.edu/group/pritchardlab/structure.html>.
46. FONDEVILA, M. et al. Revision of the SNPforID 34-plex forensic ancestry test: Assay enhancements, standard reference sample genotypes and extended population studies. *Forensic Science International: Genetics*, [S.l.], v. 7, n. 1, p. 63–74, jan. 2013. DOI: 10.1016/j.fsigen.2012.06.007.
47. PEREIRA, R. et al. Development and validation of the SNaPshot® assay for the simultaneous typing of 44 ancestry-informative markers. *Electrophoresis*, v. 33, n. 21, p. 3681–3691, 2012. DOI: <https://doi.org/10.1002/elps.201200273>
48. PHILLIPS, C. et al. Eurasiaplex: a forensic SNP assay for differentiating European and South Asian ancestries. *Forensic Science International: Genetics*, v. 7, n. 3, p. 359–366, 2013. DOI: <https://doi.org/10.1016/j.fsigen.2013.02.010>
49. BERGSTRÖM, A. et al. Insights into human genetic variation and population history from 929 diverse genomes. *Science*, v. 367, n. 6484, eaay5012, 2020. DOI: 10.1126/science.aay5012.