

Instituto de Biociências, Letras e
Ciências Exatas - IBILCE/UNESP
Departamento de Ciências da Computação e Estatística
Disciplina: Bioinformática - 1970SCC

Relatório do seminário: Bioinformática Forense

Grupo:

Heitor Xavier de Abreu

Lucas Shinji Yano

Vinícius Araújo Beltrami

Vitor Emanuel da Silva Rozeno

Professora: Vitória Zanon Gomes

Junho
2024

Instituto de Biociências, Letras e
Ciências Exatas - IBILCE/UNESP
Departamento de Ciências da Computação e Estatística
Disciplina: Bioinformática - 1970SCC

Relatório do seminário: Bioinformática Forense

Relatório do seminário sobre Bioinformática Forense,
apresentado no dia 07 de junho de 2024, parte da
disciplina de Bioinformática - 1970SCC.

Grupo:

Heitor Xavier de Abreu

Lucas Shinji Yano

Vinícius Araújo Beltrami

Vitor Emanuel da Silva Rozeno

Professora: Vitória Zanon Gomes

Junho
2024

Conteúdo

Resumo	1
1 Introdução	2
1.1 Ciência Forense	2
1.1.1 Grandes Contribuidores da Ciência Forense	2
1.1.2 Laboratórios de Crime	3
1.2 Funções de um Cientista Forense	3
1.3 Bioinformática Forense	3
2 Ferramentas e Softwares	5
2.1 CODIS	5
2.2 STRmix	5
2.3 Famílias	5
3 Trabalhos	7
3.1 A review of bioinformatic methods for forensic DNA analyses .	7
3.1.1 SNPs	7
3.1.2 STRs	8
3.1.3 mtDNA	8
3.1.4 Conclusão do artigo	9
3.2 Towards localizing head shots – Forensic sub-differentiation of anatomical brain regions by differential RNA expression	9
3.2.1 Materiais e métodos	9
3.2.2 Resultados	10
3.2.3 Conclusão	11
4 Mercado de Trabalho	12

Resumo

Este relatório apresenta um seminário sobre a correlação da área da bioinformática com a ciência forense. Exploramos alguns conceitos básicos e históricos da área forense e, posteriormente, mostramos o uso da bioinformática como ferramenta de apoio para analisar e interpretar dados biológicos, como DNA, em investigações criminais, o que possibilita formar perfis genéticos, árvores genealógicas e relações de biogeográficas entre indivíduos, além de explorar fenotípicos externos para diminuição do escopo de vítimas ou criminosos.

1 Introdução

1.1 Ciência Forense

Em sua definição ampla, a ciência forense é a aplicação da ciência às leis. De forma mais precisa, a ciência forense é a aplicação de ciência às leis criminais e civis que são aplicadas por agências policiais em um sistema de justiça criminal (Saferstein e Roy 2021). Qualquer tipo de ciência usado nos propósitos do direito é considerado como ciência forense. Esta é utilizada para resolver disputas civis, para aplicar as leis criminais e governamentais e para proteger a saúde pública (AAFS 2024). Algumas áreas que fazem parte da ciência forense são (Eckert 1997): patologia e biologia, toxicologia, criminalística, documentoscopia, odontologia forense, antropologia, psiquiatria, engenharias, geologia e microscopia.

Ainda, como definido anteriormente, quaisquer outras aplicações de ciência em termos jurídicos e penais são consideradas como aplicações de ciência forense.

1.1.1 Grandes Contribuidores da Ciência Forense

As contribuições para a ciência forense remontam a tempos antigos, refletindo uma longa história de desenvolvimento e aplicação de técnicas científicas na investigação de crimes. Alguns contribuidores podem ser destacados, considerando a importância de seus feitos. Mathieu Orfila (1787-1853) é considerado o pai da toxicologia forense, por conta de seus estudos sobre detecção de envenenamento e seus efeitos em animais. Alphonse Bertillon (1853-1914) concebeu o primeiro sistema de identificação pessoal, resultando na antropometria, em que são medidas várias partes do corpo, com o fim de distinguir indivíduos. Francis Galton (1822-1911) realizou um estudo sobre impressões digitais e desenvolveu uma metodologia de classificação destas. Leone Lattes (1887-1954) desenvolveu, em 1915, uma técnica para determinar o grupo sanguíneo de resíduos de sangue secos, o que foi amplamente aplicado em investigações criminais. Edmond Locard (1877-1966) introduziu o que ficou conhecido como princípio de troca de Locard. Esse princípio diz que sempre que dois objetos entram em contato, há uma troca de materiais entre eles. Assim, em um caso real, Locard foi capaz de determinar a presença de partículas de metal nas roupas dos suspeitos de um crime, o que foi determinante para a incriminação destes. Frances Glessner Lee (1878-1962) foi uma das pioneiras no desenvolvimento da ciência forense nos Estados Unidos. Seu trabalho "Estudos Resumidos de Mortes Não Explicadas" foi amplamente utilizado para estudos de casos de crimes reais e posteriormente,

para o ensino da ciência forense à novos profissionais.

1.1.2 Laboratórios de Crime

Com a incorporação de diversas técnicas, provenientes de diversos campos da ciência, a criação de laboratórios de crime se mostrou necessária, uma vez que para uma análise minuciosa e assertiva são necessários diversos equipamentos especializados, somados à uma equipe qualificada. O maior laboratório voltado à solução de problemas relacionados à ciência forense é o Laboratório Técnico do FBI, inaugurado em 1932, com foco inicial em áreas como sorologia, toxicologia, análise de escrita e metalografia (Fox).

As técnicas de impressão genética progrediram muito nos últimos tempos, de modo que, podem ser rastreados vestígios de sangue, manchas de sêmen, cabelo e saliva deixada em selos, copos e até marcas de mordida. Assim, muitos laboratórios são responsáveis por analisar estas amostras de DNA, e até armazená-las em bancos de dados, para fins de identificação de pessoas (Saferstein e Roy 2021).

1.2 Funções de um Cientista Forense

Majoritariamente, um cientista forense tem o trabalho de analisar evidências físicas em busca de vestígios. Esse trabalho de análise consiste em determinar quais testes ou examinações são apropriadas para o caso em mãos, de modo que, o conhecimento obtido neste processo possa ser aplicado na resolução de seu caso. Também, o julgamento de evidências científicas pode ser feito por este profissional, de modo que, uma evidência possa ser caracterizada como inconclusiva ou até falsa. Adicionalmente, poderá ser requisitado o testemunho do cientista forense, como uma testemunha técnica, sendo muito importante para a definição de sentenças (AAFS 2024, Saferstein e Roy 2021).

1.3 Bioinformática Forense

O papel da bioinformática no contexto da ciência forense é, inegavelmente, crucial. Diversas áreas têm a aplicação da bioinformática, com o intuito de facilitar investigações criminais e promover avanços no âmbito da justiça. A análise de DNA pode ser considerada como uma das principais funções da bioinformática forense, em que ferramentas e técnicas computacionais são aplicadas, a fim de realizar processamento e comparação de perfis genéticos. Assim, em identificação de indivíduos a partir de vestígios biológicos, como sangue, saliva, cabelo e células da pele, os processos advindos da bioinformática são indispensáveis.

Ademais, os bancos de dados biológicos no contexto de forense estão intrinsecamente relacionados com a bioinformática, visto o CODIS (*Combined DNA Index System*) nos Estados Unidos. Esses grandes bancos de dados possuem muitas informações sobre perfis genéticos de condenados, amostras de cenas de crime e também de indivíduos desaparecidos. O papel da bioinformática nesse contexto se dá pela rápida comparação entre sequência, proporcionada pelos avanços recentes no campo, gerando uma maior assertividade e facilidade na identificação de suspeitos e resolução de casos.

De maneira auxiliar, a técnica Reação em Cadeira da Polimerase (PCR, do inglês *Polymerase Chain Reaction*) é amplamente utilizada com o objetivo de amplificar pequenas quantidades de DNA, muitas vezes uma realidade em cenas de crime. Este processo consiste em três etapas:

1. **Desnaturação:** É realizado o aquecimento do DNA a uma temperatura entre 94°C e 98°C, fazendo com que as fitas complementares do DNA se separem;
2. **Anelamento:** Resfriamento da mistura a uma temperatura entre 50°C e 65°C, para permitir a ligação entre os primers e as sequências complementares nas fitas simples de DNA;
3. **Elongation:** Ajuste da temperatura para 72°C, temperatura ideal para a enzima TAQ Polimerase, que irá adicionar nucleotídeos aos primers, sintetizando novas fitas de DNA.

Ainda, o surgimento do sequenciamento de nova geração (NGS, do inglês *Next-Generation Sequencing*), também conhecido como **Sequenciamento Massivo Paralelo** (MPS, do inglês *Massive Parallel Sequencing*) foi extremamente importante para a ciência forense, visto que a implementação do sequenciamento paralelo permitiu um aumento estrondoso na velocidade desses processos, possibilitando um contexto de *Big Data*, exemplificado pelo grandes bancos de dados moleculares, em especial o CODIS. Esse avanço tecnológico permite que alguns sequenciadores possam ser considerados como móveis, visto sua facilidade de transporte, assim, dispositivos compactos são extremamente úteis em contextos que necessitam de agilidade, como a resolução de crimes ou julgamentos.

Por fim, muitos softwares da bioinformática podem ser utilizados no contexto da forense, como o *GeneMapper* usado principalmente no contexto da genotipagem, o MEGA (do inglês *Molecular Evolutionary Genetics Analysis*) utilizado para análise de sequências, e o BLAST também utilizado no contexto de comparação de sequências biológicas.

2 Ferramentas e Softwares

Nesta seção será abordado algumas ferramentas e software comumente utilizadas na área de bioinformática forense de modo puramente informativo.

2.1 CODIS

CODIS é o acrônimo para Combined DNA Index System, uma base de dados de perfis genéticos amplamente utilizado para comparação entre amostras de DNA. Em 2012, o Brasil aprovou a Lei nº 12.654, que regula o uso de bancos de perfis genéticos para investigações criminais, marcando um avanço nas discussões sobre o uso de DNA na justiça. A lei coincide com a adoção do Combined DNA Index System (Codis) em 2010, parte de uma tendência global de expansão de bancos genéticos. Laboratórios forenses devem ser certificados para incluir dados no Codis, e o Núcleo de Perícia em DNA Forense da Pefoce é certificado pelo controle de qualidade do Grupo Iberoamericano de Trabalho em Análise de DNA (Gitad) desde 2007, garantindo alta qualidade na coleta e análise de dados genéticos para uso em investigações criminais no país.

2.2 STRmix

O STRmix é um poderoso software forense utilizado para interpretar perfis de DNA, especialmente em casos em que múltiplos indivíduos contribuíram para a amostra. O software foi amplamente adotado para interpretação de trabalhos rotineiros e é considerado o padrão na Australásia, subregião da Oceania, desde 2012. Ele foi usado com sucesso em casos envolvendo perfis de DNA mistos, auxiliando na identificação de contribuintes mesmo quando os métodos tradicionais falham em fazê-lo. O programa possui uma gama de recursos como Resolução de Perfis de DNA Mistos, Deconvolução com Faixa de Número de Contribuintes, Atribuição de Razões de Verossimilhança, etc.

2.3 Familias

O Familias é um software livre para cálculos de probabilidade ao inferir paternidade e identificação com base em dados de DNA. Ele pode lidar com casos complexos onde mutações potenciais são consideradas, além de ser capaz de lidar com múltiplas linhagens familiares simultaneamente. Desenvolvido inicialmente pelo Norwegian Computing Center em cooperação com o Instituto de Medicina Forense em Oslo, o programa tem sido atualizado ao

longo dos anos. Sua funcionalidade básica inclui determinar a paternidade em casos simples, como quando se deseja saber se um homem é o pai de uma criança, usando dados de DNA da criança, da mãe e do suposto pai, juntamente com frequências alélicas na população relevante. O software também pode calcular a probabilidade de outras relações familiares, como tios, irmãs, entre outros, considerando diferentes hipóteses. Além disso, ele leva em conta a possibilidade de mutações e oferece recursos como ajuste para parentesco.

3 Trabalhos

Nesta seção serão apresentados alguns trabalhos recentes na área de bioinformática forense, visando demonstrar o atual estado da arte e suas aplicações.

3.1 A review of bioinformatic methods for forensic DNA analyses

O artigo de Liu e Harbison 2018 aborda a crescente importância da bioinformática nas análises forense de DNA, especialmente com o advento das plataformas de Sequenciamento Massivo Paralelo (MPS, do inglês *Massive Parallel Sequencing*). A grande quantidade de dados gerados pelos MPS precisam ser processadas, analisadas e interpretadas para poder utilizar os marcadores genéticos forenses.

Os marcadores genéticos são sequências de nucleotídeos (curtas ou longas) com uma localização conhecida em um cromossomo que pode ser utilizado para identificar indivíduos ou espécies, sendo capaz de diferenciar os progenitores e ser reproduzido com precisão nos descendentes. Algumas características possuem uma alta polimorfia, o que permite identificar unicamente indivíduos e a estabilidade em diferentes ambientes.

No artigo são ressaltados três marcadores genéticos que são amplamente utilizados na análise de DNA forense com MPS:

- SNPs (*Single Nucleotide Polymorphisms*)
- SRTs (*Short Tandem Repeats*)
- mtDNA (*Mitochondrial Genome*)

3.1.1 SNPs

Esses marcadores são variações na sequência dos nucleotídeos que ocorrem quando um único nucleotídeo na sequência é diferente da norma em pelo menos 1% da população (medida necessária para inviabilizar a confusão com mutações pontuais). Eles podem ocorrer dentro ou fora dos genes (área responsável por um fenótipo), aparecendo um SNP a cada 100 a 300 bases nitrogenadas. Quando ocorre perto de um gene particular, o SNP age como um marcador genético molecular para o gene.

Devido ao grande volume de dados de SNPs gerados, é possível inferir informações dos SNPs relevantes para área forense. Assim, os SNPs são classificados em 4 tipos:

- SNPs informativos de identidade (IISNPs) para individualização;

- SNPs informativos de linhagem (LISNPs) para haplotipagem de parentesco
- SNPs informativos de ancestralidade (AISNPs) para inferência de ancestralidade biogeográfica
- SNPs informativos de fenótipos (PISNPs) para predição de fenótipos externos

O uso desse tipo de marcador cresce conforme é exequível o processamento de grandes quantidades de SNPs, visto que é necessário processar uma quantidade maior de loci comparado aos outros marcadores.

3.1.2 STRs

Também denominados Microssatélites, esses marcadores são unidades curtas de sequências repetitivas de nucleotídeos (2 a 6 pares de bases). Conforme o uso de MPS amplia-se, o método CE (do inglês, *Capillary Electrophoresis*) com STR tradicional é comparado, devido as novas vantagens emergentes, como a detecção de variações de sequência do STR e o maior número de marcadores de DNA que podem ser multiplexados juntos.

Um dos principais desafios da utilização de MPS é a busca de um formato que melhor combina a informação sequencial com as variações de sequência, enquanto se mantém a compatibilidade com base de dados do método CE.

3.1.3 mtDNA

O DNA mitocondrial pode ser utilizado para revelar a linhagem dos indivíduos e se utiliza em situações quando o DNA nuclear está indisponível, degradado - devido ao fato do DNA de estrutura circular que possui resistência a digestão enzimática - ou informações adicionais de linhagem são necessárias.

O processamento de mtDNA era realizado tradicionalmente pelo método de sequenciamento de Sanger, focado em regiões de controle e de HVS (Hypervariation). Entretanto, a análise nessas regiões ainda ocasionava vários erros, como *artificial recombination* causado pela mistura de segmentos das regiões hipervariáveis de diferentes indivíduos, dados de baixa qualidade que levavam erros de interpretação de sequência, etc. Mas a descoberta de polimorfismos fora dessas regiões e a queda nos custos de produção de informação tornou o MPS uma solução mais adequada que o método de Sanger.

3.1.4 Conclusão do artigo

A acessibilidade e desenvolvimento das plataformas e bases de dados MPS possibilita a aplicação do novo sequenciamento ao invés das técnicas tradicionais existentes para cada tipo de marcação, possibilitando a contribuição de novas informações valiosas provenientes das marcações genéticas para o campo forense.

3.2 Towards localizing head shots – Forensic sub-differentiation of anatomical brain regions by differential RNA expression

O artigo de Euteneuer, Moitinho-Silva e Courts 2022 explora a possibilidade de usar a análise de RNA para identificar regiões específicas do cérebro que foram atingidas em traumas severos, como aqueles causados por armas de fogo ou impactos contundentes.

A análise forense de RNA é uma técnica válida e rotineiramente utilizada para detectar tecidos de órgãos em amostras desafiadoras. Em casos de traumas cranianos severos, pode ser benéfico identificar qual área da cabeça foi atingida para fins de investigação criminal. O cérebro humano é uma estrutura altamente complexa composta por diferentes regiões anatômicas, cada uma com funções específicas e distintas. Essas regiões são o resultado de diferentes perfis de expressão gênica, que se traduzem em diversas populações de células com diferentes funções. O mRNA (RNA mensageiro) é um intermediário crucial entre o DNA (que contém a informação genética) e as proteínas (que desempenham funções celulares). A expressão diferencial de mRNAs em diferentes partes do cérebro reflete as necessidades funcionais específicas de cada região. Este estudo visa encontrar marcadores específicos de mRNA que possam diferenciar regiões anatômicas do cérebro.

Simultaneamente, é necessário uma análise completa do transcriptoma (Sagurthi, Setti e Pawar 2015) (conjunto completo de transcritos - RNAs mensageiros, RNAs ribossômicos, RNAs transportadores e os microRNAs - de um dado organismo, órgão, tecido ou linhagem celular. Portanto, ele é o reflexo direto da expressão dos genes) das quatro áreas do córtex cerebral de interesse forense e traumatológico (lobo frontal, temporal, parietal, occipital).

3.2.1 Materiais e métodos

O estudo utilizou 80 amostras coletadas de autópsias das 4 diferentes áreas do córtex cerebral, descartando amostras com traumas cranianos, sinais de putrefação ou doenças neurodegenerativas visíveis. A partir das amostras,

extraíram e sequenciaram o RNA para quantificar a expressão de milhares de genes simultaneamente, identificando quais mRNAs estão presentes em uma amostra de tecido.

A análise resultou em 1.030.286 RNAs. Para selecionar candidatos robustos de mRNA específicos para áreas cerebrais, critérios de seleção incluíram mudança de expressão significativa, valor $p > 0,01$, BaseMean > 50 e $\log_2\text{FoldChange} > 1$, resultando em apenas 29 RNAs.

3.2.2 Resultados

No estudo, após a identificação inicial dos candidatos a mRNAs específicos das diferentes regiões do cérebro, foi realizada uma análise mais detalhada utilizando bases de dados como BioGPS e ProteinAtlas para verificar os padrões de expressão desses mRNAs em todo o córtex cerebral. Essas bases de dados contêm informações extensivas sobre a expressão gênica em diversos tecidos e órgãos, incluindo o cérebro.

Os pesquisadores usaram essas ferramentas para confirmar se os candidatos a mRNAs identificados no estudo inicial tinham padrões de expressão que poderiam ser úteis para diferenciar entre as quatro regiões do cérebro (frontal, temporal, parietal e occipital). Os resultados das análises do BioGPS e ProteinAtlas revelaram dois cenários principais para os mRNAs candidatos:

1. Expressão Uniforme em Todo o Córtex Cerebral - alguns dos mRNAs candidatos mostraram uma expressão uniforme em todas as áreas do córtex cerebral. Isso significa que esses mRNAs não têm variação suficiente na sua expressão entre as diferentes regiões para serem úteis como marcadores específicos de área.
2. Regiões Diferentes de Upregulação - outros mRNAs mostraram um aumento de expressão (upregulação) em partes diferentes do cérebro, mas não de maneira específica o suficiente para distinguir claramente entre as quatro regiões estudadas. Essa variação não foi suficiente para permitir a sub-diferenciação precisa das áreas cerebrais em um contexto forense.

Essa descoberta sublinha a complexidade da expressão gênica no cérebro e os desafios associados à tentativa de usar perfis de expressão de RNA para aplicações forenses. A uniformidade ou a upregulação não específica dos mRNAs nos candidatos analisados limita a capacidade de determinar a localização exata do trauma em diferentes regiões do cérebro apenas com base na expressão de mRNA.

3.2.3 Conclusão

A utilização de ferramentas como BioGPS e ProteinAtlas para confirmar os perfis de expressão gênica é um passo crítico na validação de potenciais biomarcadores. No entanto, no contexto deste estudo, esses recursos confirmaram que os mRNAs candidatos não eram adequados para a sub-diferenciação anatômica precisa das regiões cerebrais, destacando a necessidade de explorar outras abordagens ou identificar novos biomarcadores com maior especificidade regional para aplicações forenses.

4 Mercado de Trabalho

Os cientistas forenses trabalham em laboratórios e escritórios, podendo ter atividades em campo. Alguns dos lugares mais comuns que possuem cientistas forenses trabalhando são: laboratórios público e privados; cenas de crime; salas de aula; necrotérios; governos e organizações internacionais; hospitais; universidades; departamentos de polícia, entre outros. O salário médio de um cientista forense nos Estados Unidos é estimado entre 40 mil dólares e 100 mil dólares (AAFS 2024). No panorama brasileiro, uma grande parcela das vagas para especialista em ciência forense está concentrada em órgãos públicos, como a Polícia Científica e a Polícia Civil, necessitando da aprovação em concurso para admissão. A remuneração inicial média desses postos de trabalho é de R\$10.000,00. As provas para estes cargos são constituídas de etapas objetivas e discursivas, além de um teste de aptidão física (Souza).

Referências

- AAFS 2024 AAFS, A. A. of F. S. *What is Forensic Science?* 2024. Acessado em 23/05/2024. Disponível em: <https://www.aafs.org/careers-forensic-science/what-forensic-science>.
- Eckert 1997 ECKERT, W. G. *Introduction to forensic sciences, second edition*. [S.l.]: CRC Press, 1997.
- Euteneuer, Moitinho-Silva e Courts 2022 EUTENEUER, J.; MOITINHO-SILVA, L.; COURTS, C. Towards localizing head shots – forensic sub-differentiation of anatomical brain regions by differential rna expression. *Forensic Science International: Genetics Supplement Series*, v. 8, p. 178–180, 2022. ISSN 1875-1768. The 29th Congress of the International Society for Forensic Genetics. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1875176822000750>.
- Fox FOX, J. F. *The Birth of the FBI's Technical Laboratory - 1924 to 1935*. Acessado em 23/05/2024. Disponível em: <https://www.fbi.gov/history/history-publications-reports/the-birth-of-the-fbis-technical-laboratory1924-to-1935>.
- Liu e Harbison 2018 LIU, Y.-Y.; HARBISON, S. A review of bioinformatic methods for forensic dna analyses. *Forensic Science International: Genetics*, v. 33, p. 117–128, 2018. ISSN 1872-4973. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1872497317302818>.
- Saferstein e Roy 2021 SAFERSTEIN, R.; ROY, T. *Criminalistics An Introduction to Forensic Science*. [S.l.]: Pearson, 2021.
- Sagurthi, Setti e Pawar 2015 SAGURTHI, S.; SETTI, A.; PAWAR, S. Systems biology: A new frontier. *Plant Biology and Biotechnology: Volume II: Plant Genomics and Biotechnology*, p. 301–314, 01 2015.
- Souza SOUZA, D. *Concursos para Perito Criminal: 193 vagas autorizadas*. Acessado em 23/05/2024. Disponível em: <https://www.estrategiaconcursos.com.br/blog/concursos-para-perito-criminal-vagas-ja-autorizadas/>.