THE ROLE OF BIOINFORMATICS IN FORENSIC SCIENCE

[1]MUNEERA FAHAD AL-HULAIBAH, [2]BDOR HIRAB AL-MUTAIRI, [3]AMAL ABDULLAH AL-WADAH

Dept. of Computer Sciences, College of Computer Engineering and Sciences Prince Sattam Bin Abdul Aziz University Al Kharj, 16273, Saudi Arabia
E-mail: [1]446540128@psau.edu.sa, [2]446540191@psau.edu.sa, [3]446540057@psau.edu.sa

Abstract - Bioinformatics has become an essential component of modern forensic science, enabling the efficient analysis of large-scale biological data—particularly DNA—in criminal investigations. This paper examines the contemporary applications of bioinformatics in forensic science, with a focus on DNA profiling, genetic marker analysis, and mitochondrial DNA analysis. It highlights how advanced computational tools and technologies, such as massively parallel sequencing (MPS), the Basic Local Alignment Search Tool (BLAST), and genomic databases, enhance the accuracy, speed, and reliability of forensic analyses.

In addition, the study discusses real-world applications, including high-profile criminal cases and disaster victim identification, illustrating the practical impact of bioinformatics in forensic contexts. Key advantages, such as improved data processing, reduced human error, and faster case resolution, are examined alongside challenges related to data complexity, ethical considerations, and high computational demands.

Finally, the paper explores future trends in the field, particularly the integration of artificial intelligence and machine learning into forensic analysis, which is expected to revolutionize forensic science further. Overall, bioinformatics plays a critical role in advancing forensic investigations by making them more accurate, efficient, and reliable.

Keywords - Bioinformatics, Forensic Science, DNA Analysis, DNA Profiling, Genetic Markers, Next Generation Sequencing (NGS).

I. INTRODUCTION

Forensic science involves the use of biology, chemistry, and technology in criminal investigation and legal matters. Among these, DNA profiling plays an important role due to the uniqueness of genetic profile for every person. The field of bioinformatics has proved to be indispensable due to the shortcomings posed by the conventional approaches in terms of speed, errors, and ethical issues [1, 2].



Fig. 1. Forensic DNA Analysis in Criminal Investigation.

II. SECTOR OVERVIEW: FORENSIC MEDICINE

A. Current Challenges in Forensic Medicine

Although forensic medicine is effective, it faces several challenges. Traditional methods of DNA analysis are often slow and require a lot of manual effort, which may delay the results of the investigation. The significant increase in the volume of genetic data also poses challenges related to storage, processing, and analysis. In addition, manual operations may lead to human errors that affect the accuracy of the results. Ethical issues related to the privacy of genetic data also arise, especially when using DNA databases to identify individuals. All these challenges confirm the need for advanced technical solutions to improve the efficiency and accuracy of forensic analysis [2][5].

B. Importance of Bioinformatics in Forensic Medicine

Bioinformatics is a key component in the development of forensic medicine, providing advanced computer tools to analyze big genetic data quickly and efficiently. This field contributes to accelerating the process of DNA sequence analysis, comparing samples, and detecting genotypes with high accuracy. It also reduces reliance on manual analysis, which contributes to reducing human errors and increasing the reliability of results. In addition, bioinformatics allows the linking and analysis of large DNA databases, which helps in quickly identifying suspects or victims. Therefore, this field has become an indispensable part of modern criminal investigations [3][9].

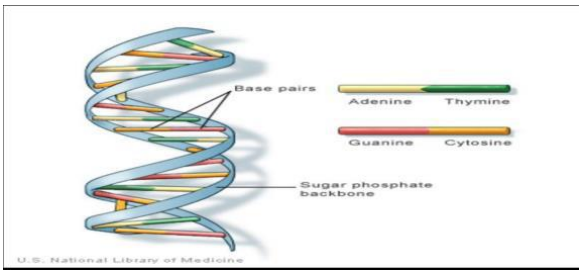
III. FUNDAMENTALS OF BIOINFORMATICS

A. Bioinformatics

Bioinformatics is defined as a scientific field that combines biology, computer science, and statistics with the aim of analyzing and interpreting biological data, especially genetic data such as DNA. This field relies on the use of advanced algorithms and software databases to process large amounts of genetic information, making it an essential tool in many scientific and medical applications, including forensic medicine [3].

B. DNA and Sequence Analysis

DNA is the basic genetic material in all living organisms, and it carries information that distinguishes one individual from another. DNA consists of a sequence of small units known as nucleotides, which are represented by four basic bases: adenine (A), thymine (T), cytosine (C) and guanine (G). The arrangement of these bases forms what is known as genetic sequencing, which is a unique biological fingerprint for each human being. In bioinformatics, these sequences are analyzed and compared between different samples with the aim of identifying or knowing biological relationships. In forensic medicine, this process is used to match DNA found



relies heavily on DNA analysis to determine the identity of individuals and solve crimes. However, traditional methods face challenges related to time, analysis accuracy, and data volume. This is where bioinformatics comes in, providing advanced technical solutions to analyze genetic data quickly and accurately, making it an essential element in the development of modern criminal investigations.

IV. CURRENT BIOINFORMATICS APPLICATIONS IN FORENSIC SCIENCE

Bioinformatics plays a crucial role in modern forensic science, as it facilitates the rapid and precise analysis of biological evidence, particularly DNA, for purposes of individual identification and criminal investigation. It relies on advanced computational methodologies and specialized databases to efficiently manage and interpret large-scale and complex genetic datasets. Through the interdisciplinary integration of computer science, biology, and statistics, bioinformatics substantially strengthens forensic capabilities, especially in the accurate examination of DNA obtained from crime scenes. [10].

A. Real-World Applications (Use cases) in Forensic Science

a) DNA Profiling

DNA profiling is one of the most critical applications of bioinformatics in forensic science, employing bioinformatics tools to analyze biological samples such as blood, hair, and saliva for the purpose of individual identification. The application of advanced technologies, including Massively Parallel Sequencing (MPS), has significantly improved the speed, accuracy, and reliability of genetic analyses, thereby enhancing the effectiveness of forensic investigations [10].

A real-world example is the Golden State Killer case, in which DNA evidence combined with genetic genealogy databases was utilized to identify a suspect more than 40 years after the crimes were committed [11].

b) Genetic Markers (STRs and SNPs)

Genetic markers are distinct DNA sequences that are employed for the identification of individuals and species, constituting a crucial element in forensic science and criminal investigations. Bioinformatics tools facilitate the rapid and precise analysis of these markers, enabling forensic specialists to generate accurate and reliable genetic profiles. Among the most widely used markers are Short Tandem Repeats (STRs), which exhibit high levels of polymorphism and are extensively applied in identity testing, and Single Nucleotide Polymorphisms (SNPs), which provide detailed and informative insights into genetic variation [12].

A real-world example: STR analysis is used in the CODIS database (FBI), where DNA profiles from crime scenes are compared with millions of stored records to identify suspects.

c) Mitochondrial DNA (mtDNA) Analysis

Mitochondrial DNA (mtDNA) analysis represents a powerful and effective tool in forensic investigations, particularly in cases where nuclear DNA is unavailable or severely degraded. Unlike nuclear DNA, mtDNA is maternally inherited and exists in multiple copies per cell, increasing the likelihood of successful recovery from limited or compromised biological samples.

Consequently, mtDNA analysis is especially advantageous in challenging forensic contexts, including the examination of aged or skeletal remains, hair shafts, mass disaster victim identification, and historical or cold case investigations. Recent advances in bioinformatics have further enhanced the accuracy, efficiency, and interpretative power of mtDNA analysis, reinforcing its reliability as a source of forensic evidence [13].

Real-world example: mtDNA analysis was used in identifying victims of the September 11 attacks, especially when DNA samples were severely damaged. [14].

B. Organizations and Institutions

Bioinformatics applications in forensic science are widely utilized by various organizations involved in criminal investigation and scientific research:

- Federal Bureau of Investigation (FBI): Uses CODIS database for DNA profiling and matching.
- Forensic laboratories: Analyze biological samples using bioinformatics tools
- Research institutions and universities: Develop and improve advanced DNA analysis techniques.

C. Technologies, Tools, and Datasets

Various technologies and tools are used in forensic bioinformatics to ensure accurate and efficient DNA analysis.

a) Technologies:

- Massively Parallel Sequencing (MPS)

An advanced sequencing technology enabling rapid and high-throughput DNA analysis, even from small or degraded samples.

- Polymerase Chain Reaction (PCR)

A key technique used to amplify small amounts of DNA for accurate forensic analysis and downstream genetic examination.

- Variant Calling

The process of comparing sequenced DNA with a reference genome to detect genetic variants essential for forensic identification and relationship analysis.

- Quality Assessment

Involves filtering and trimming low-quality sequence data to ensure the accuracy, reliability, and legal credibility of forensic DNA analysis [10].

b) Tools:

- Genomics Viewer:

Used to visualize DNA sequences and identify genetic markers or anomalies quickly.

- Geneious:

An integrated platform for sequence analysis, alignment, and primer design in one place.

- NextGENe:

Software for analyzing next-generation sequencing data with a user-friendly interface [10].

c) Datasets:

- Forensic DNA Databases:

These databases are maintained by law enforcement agencies and contain DNA profiles obtained from crime scenes and convicted offenders. They are used to compare and match DNA samples for suspect identification.

A well-known example is the Combined DNA Index System (CODIS), which is widely used to link suspects to criminal cases.

- ExAC Database:

The Exome Aggregation Consortium (ExAC) database provides large-scale data on human genetic variation. Although it is not specifically designed for forensic use, it supports the analysis of rare genetic profiles and helps in population-based genetic interpretation, which can be useful in complex forensic investigations [10].

V. BENEFITS AND OPPORTUNITIES

There are many benefits of using bioinformatics in forensic science, including:

- Dealing with damaged samples, such as victims of disasters and mass accidents using mtDNA.
- Decreasing mistakes made by people, as automated analysis reduces human errors.
- Speed of sample analysis, as bioinformatics enables faster processing of large DNA datasets compared to traditional methods.
- Solving cold cases, such as the Golden State Killer case that was solved more than 40 years later.
- Higher accuracy in identification using techniques such as mtDNA, STRs, and SNPs [1],[15].

VI. CHALLENGES AND LIMITATIONS

The most important potential challenges: First, big data requires very powerful processing and storage capabilities. In addition, MPS and NGS technologies are relatively expensive. The use of DNA databases sparks ethical and legal issues, making privacy issues a real challenge. Furthermore, some new evidence needs strong proof in court, which requires that the results be legally accepted. Finally, analysis requires dual expertise of biology and computer science, which requires specialized experts.[2],[5],[16].

VII. FUTURE TRENDS

The future of forensic science is expected to be significantly influenced by the integration of artificial intelligence (AI) and advanced computational technologies.

The use of AI can help detect complex patterns, predict analysis results, reduce investigation time, and improve sample matching. In addition, neural networks can be used to analyze genetic sequences through deep learning. AI can also support predictive forensics by identifying characteristics such as geographical origin, eye color, and hair color.

Moreover, forensic laboratories may become fully automated, allowing processes to operate with minimal human intervention. Finally, the integration of global DNA databases will help speed up the identification of criminals [18],[17],[19].

VIII. PRACTICAL IMPLEMENTATION OF FORENSIC DNA MATCHING

In this implementation, a simulated forensic DNA matching scenario was conducted using real DNA sequences retrieved from the NCBI database. One sequence was selected as the crime scene sample,

while three additional sequences were considered as potential suspects. The analysis was carried out in two main stages. First, BLAST analysis was performed to validate the DNA sequence and confirm its similarity to known human mitochondrial DNA references, yielding approximately 98% similarity and verifying the biological origin of the sample. Second, pairwise sequence alignment was conducted using the EMBOSS Needle tool to directly compare the crime scene sample with each suspect sequence, allowing precise calculation of sequence similarity percentages.

a) Use these sequences

- Crime Scene Sample

TTCTTTCATGGGGAAGCAGATTTGGGTACCAC
CCAAGTATTGACTCACCCATCAACAACCGCT
ATGTATT
TCGTACATTACTGCCAGCCACCATGAATATTG
TACAGTACCATAAATACTTGACCACCTGTAGT
ACATAA
AAACCCAATCCACATCAAAACCCTCCCCCCA
TGCTTACAAGCAAGTACAGCAATCAACCTTC
AACTGTCA
CAATCAACCGCAACTCCAAAGCCACCCCTCA
CCCACTAGGATACCAACAAACCTACCCACCC
TTAACAGT
ACATAGCACATAAAGCCATTTACCGTACATA
GCACATTACAGTCAAATCCCTTCTCGTCCCCA
TGGATGA
CCCCCCTCAGATAGGATAGCATTGCGAGACG
CTGGAGCCGGAGCACCCTATGTCGCAGTATC
TGTCTTTGATTCCTGCCCCATC
CCATTATTTATCGCACCTACGTTCAATATTAC
AGGCGAACATACCTATTAAAGTGTGTTAATT
AATTAAT
GCTTGTAGGACATAATAATAACAATTAAATG
TCTGCACAGCCACTTTCCACACAGACATCATA
ACAAAAA
ATTTCCACCAAACCCCCCCCTCCCCCCGCTTC
TGGCCACAGCACTTAAACACATCTCTGCCAA
ACCCCAA
AAACAAAGAACCCTAACACCAGCCTAACCAG
ATTTCAAATTTTAT

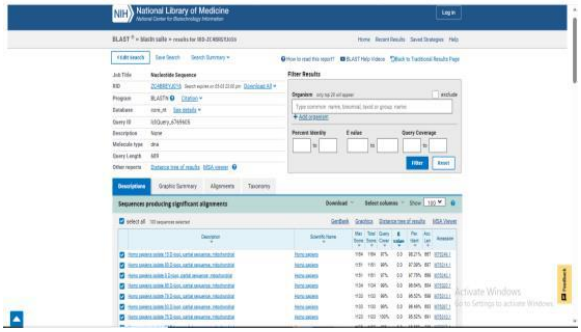


Fig. 4. BLAST analysis indicating ~98% identity with human mitochondrial DNA, confirming sample origin.

- Suspect 1

TTCTTTCATGGGGAAGCAGATTTGGGTACCAC
CCAAGTATTGACTCACCCATCAACAACCGCT
ATGTATT
TCGTACATTACTGCCAGCCACCATGAATATTG
TACAGTACCATAAATACTTGACCACCTATAGT
ACATAA
AAACCCAATCCACATCAAAACCCTCCCCCCA
TGCTTACAAGCAAGTACAGTAATCAACCTTC
AACTGTCA
CACATCAACCGCAACTCCAAAGCCACCCCTC
ACCCACTAGGATACCAACAAACCTACCCATC
CTTAACAG
TACATAGCACATAAAGCCATTTACCGTACAT
AGCACATTACAGTCAAATCCCTTCTCGTCCCC
ATGGATG
ACCCCCCTCAGATAGGTTGGTATTTTCGTCTG
GGGGGTGT
GCACGCGATAGCATTGCGAGACGCTGGAGCC
GGAGCACCCTATGTCGCAGTATCTGTCTTTGA
TTCCTGC
CCCATCCCATTATTTATCGCACCTACGTTCAA
TATTACAGGCGAACATACCTACTAAAGTGTG
TTAATTA
ATTAATGCTTGTAGGACATAATAATAACAAT
TAAATGTCTGCACAGCCACTTTCCACACAGA
CATCATAA
CAAAAAATTTCCACCAAACCCCCCCTCCCCC
CGCTTCTGGCCACAGCACTTAAACACATCTCT
GCCAAAC
CCCAAAAACAAAGAACCCTAACACCAGCCTA
ACC

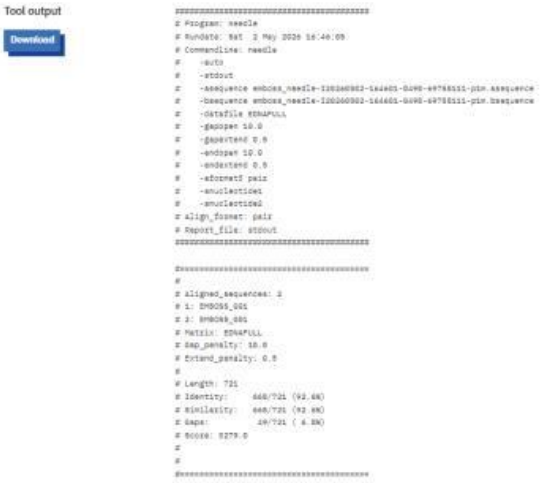


Fig. 5. Pairwise alignment showing 92.6% similarity with Suspect 1.

- Suspect 2

TTCTTTCATGGGGAAGCAGATTTGGGTACCAC
CCAAGTATTGACTCACCCATCAACAACCGCT
ATGTATT
TCGTACATTACTGCCAGCCACCATGAATATTG
TACAGTACCATAAATACTTGACCACCTGTAGT
ACATAA
AAACCCAATCCACATCAAAACCCTCCCCCCA
TGCTTACAAGCAAGTACAGCAATCAACCTTC
AACTGTCA
CACATTAACTGCAACTCCAAAGCCACCCCTC
ACCCACTAGGATACCAACAAACCTACCCATC
CTTAACAG

TACATAGCACATAAAGCCATTTACCGTACAT
AGCACATTACAGTCAAATCCCTTCTCGTCCCC
ATCCNNNTATTTTCGTCTGGGGGGTGTGCACG
CGATAGCATTGCGAGACG
CTGGAGCCGGAGCACCCTATGTCGCAGTATC
TGTCTTTGATTCCTGCCCCATCCTATTATTTAT
CGCACC
TACGTTCAATATTACAGGCGAACATACCTACT
AAAGTGTGTTAATTAATTAATGCTTGTAGGAC
ATAATA
ATAACAATTAAATGTCTGCACAGCCACTTTCC
ACACAGACATCATAACAAAAAATTTCCACCA
AACCCCC
CCTCCCCCCCGCTTCTGGCCACAGCACTTAAAC
ACATCTCTGCCAAACCCCAAAAACAAAGAAC
CCTAACA
CCAGCCTAACCAGATTTCAAATTTTAT

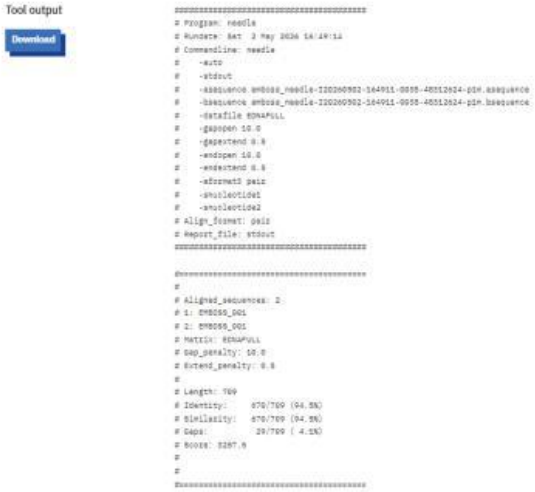


Fig. 6. Pairwise alignment showing 94.5% similarity with Suspect 2.

- Suspect 3

TTCTTTCATGGGGAAGCAGATTTGGGTACCAC
CCAAGTATTGACTCACCCATCAACAACCGCT
ATGTATC
TCGTACATTACTGCCAGCCACCATGAATATTG
TACAGTACCATAAATACTTGACCACCTGTAGT
ACATAA
AAACCCAATCCACATCAAAACCCTCCCCCCA
TGCTTACAAGCAAGTACAGTAATCAACCCTC
AACTGTCA
TACATCAACCGCAACTCCAAAGCCACCCCTC
AGCCACTAGGATACCAACAAACCTACCCACC
CTTAACAG
TACATAGCACATAAAGCCATTTACCGTACAT
AGCACATTACAGTCAAATCCCTTCTCGTTATT
TTCGTCTGGGGGGTGTGCACGCGATAGCATT
GCGAGACGCTGGAGCC
GGAGCACCCTATGTCGCAGTATCTGTCTTTGA
TTCCTGCCCCATCCCATTATTTATCGCACCTA
CGTTCA
ATATTACAGGCGAACATACCTACCAAAGTGT
GTTAATTAATTAATGCTTGTAGGACATAATAA
TAACAAT
TAAATGTCTGCACAGCCACTTTCCACACAGA
CATCATAACAAAAAATTTCCACCAAACCCCC
CCTCCCCC
CGCTTCTGGCCACAGCACTTAAACACATCTCT
GCCAAACCCCAAAAACAAAGAACCCTAACAC
CAGCCTA
ACCAGATTTCAAA

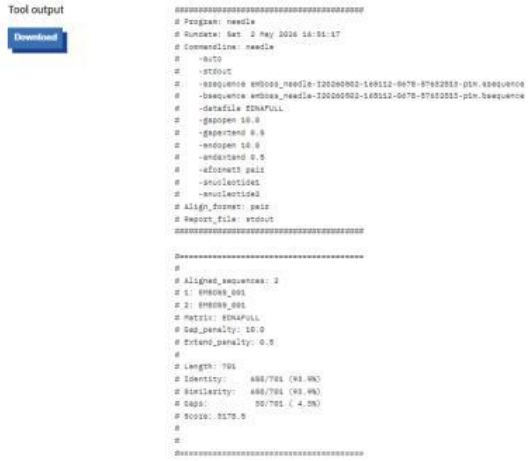


Fig. 7. Pairwise alignment showing 93.9% similarity with Suspect 3.

b) Results

Comparison	Similarity (%)
Crime vs Suspect 1	92.6%
Crime vs Suspect 2	94.5%
Crime vs Suspect 3	93.9%

The results indicate that Suspect 2 has the highest similarity to the crime scene DNA.

c) Discussion

The findings demonstrate that DNA sequence comparison is an effective approach for identifying the closest genetic match in forensic investigations. Although none of the suspects exhibited a perfect match, Suspect 2 showed the highest similarity, indicating a greater likelihood of association with the crime scene sample. These results reflect natural genetic variation, where minor differences between DNA sequences are expected. This approach highlights the importance of bioinformatics tools in enabling accurate and efficient genetic data analysis. Furthermore, this implementation illustrates the practical application of bioinformatics in forensic science, where the integration of BLAST for sequence validation and pairwise alignment for direct comparison supports informed decision-making. Overall, this methodology reflects real-world forensic practices and underscores the critical role of computational tools in modern biological analysis.

This implementation demonstrates the educational application of bioinformatics tools in forensic DNA

analysis and is not intended to replace official forensic identification procedures.

IX. CONCLUSION

In conclusion, bioinformatics has profoundly transformed modern forensic science by significantly improving the accuracy, efficiency, and reliability of DNA-based analyses in criminal investigations. Techniques such as short tandem repeat (STR) profiling, single-nucleotide polymorphism (SNP) analysis, mitochondrial DNA testing, and artificial intelligence–driven tools have enabled forensic scientists to interpret complex biological data and resolve cases that were once considered unsolvable.

Although challenges remain—notably ethical considerations, data privacy issues, and the high computational demands of advanced analyses—ongoing technological advancements are expected to mitigate these limitations. The continued integration of artificial intelligence, big data analytics, and internationally connected DNA databases will further enhance the scope and effectiveness of forensic investigations.

REFERENCES

[1] J. M. Butler, Forensic DNA Typing: Biology, Technology, and Genetics of STR Markers. Academic Press, 2005.

[2] National Research Council, Strengthening Forensic Science in the United States: A Path Forward. National Academies Press, 2009.

[3] D. W. Mount, Bioinformatics: Sequence and Genome Analysis. Cold Spring Harbor Laboratory Press, 2004.

[4] A. Brazma et al., "Minimum information about a microarray experiment (MIAME)," Nature Genetics, 2001.

[5] E. R. Mardis, "Next-generation DNA sequencing methods," Annual Review of Genomics and Human Genetics, 2008.

[6] J. C. Venter et al., "The sequence of the human genome," Science, 2001.

[7] National Center for Biotechnology Information (NCBI), [Online]. Available: https://www.ncbi.nlm.nih.gov

[8] R. Durbin et al., Biological Sequence Analysis. Cambridge University Press, 1998.

[9] S. Karni et al., "Bioinformatics in forensic science," Forensic Science International, 2020.

[10] Riveraxe, "Application of Bioinformatics in Forensic Science," [Online]. Available: https://riveraxe.com/application-of-bioinformatics-in-forensic-science/

[11] A. Author et al., "Forensic DNA analysis and bioinformatics approaches," NCBI, 2020. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC7219171/

[12] M. A. Jobling, P. Gill, and C. Phillips, "Advances in forensic DNA analysis and bioinformatics," NCBI, 2022. [Online]. Available: https://pmc.ncbi.nlm.nih.gov/articles/PMC9529747/

[13] Microbe Notes, "Mitochondrial DNA Analysis," [Online]. Available: https://microbenotes.com/mitochondrial-dna-analysis/

[14] National Institute of Justice, "Identifying Remains: Lessons Learned from 9/11," [Online]. Available: https://nij.ojp.gov/topics/articles/identifying-remains-lessons-learned-911

[15] M. A. Jobling and P. Gill, "Encoded evidence: DNA in forensic analysis," Nature Reviews Genetics, vol. 5, no. 10, pp. 739–751, 2004.

[16] M. Kayser and P. de Knijff, "Improving human forensics through advances in genetics, genomics and molecular biology," Nature Reviews Genetics, vol. 12, no. 3, pp. 179–192, 2011.

[17] E. J. Topol, "High-performance medicine: the convergence of human and artificial intelligence," Nature Medicine, vol. 25, pp. 44–56, 2019.

[18] L. G. Biesecker and R. C. Green, "Diagnostic clinical genome and exome sequencing," New England Journal of Medicine, vol. 370, no. 25, pp. 2418–2425, 2014.

[19] A. Dhall et al., "Artificial intelligence in forensic science: A review," Wiley Interdisciplinary Reviews: Forensic Science, vol. 4, no. 2, 2022.

★★★