# BIOINFORMATICS AND FORENSICS GENOMICS: EVIDENTIARY STANDARDS, ADMISSIBILITY, AND PRIVACY IMPLICATIONS IN CRIMINAL JUSTICE

Dr. Madhuri Sharma[1], Dr. Sonam Ratnu[2], Dr. Aishwarya Singh[3], Dr. Kusum Saini[4], Hemangini Shekhawat[5], Chanchal Meena[6]

[1] Assistant professor, Biyani Law College, Jaipur
[1] Assistant professor, Biyani Law College, Jaipur
[1] Assistant professor, Law College Dehradun, Uttaranchal University, Dehradun
[1] Assistant professor, Biyani Law College, Jaipur
[1] Assistant professor, Biyani Group of Colleges, Jaipur
[1] Assistant professor, Biyani Law College, Jaipur

## ABSTRACT

Tools of bioinformatics have transformed criminal investigations by harnessing genomic sequencing, typing analysis, and investigative genetic genealogy (IGG). The power of this modern-day technology allows accurate identification from extremely small biological samples, reopens old cases, and exonerates wrongfully convicted individuals. The application of this powerful science into our criminal justice systems poses unique concerns over evidence of admissibility, quality, and privacy.

The study explores bioinformatics pipelines for forensic DNA profiling and includes data analysis, variant calls, and probabilistic genotyping. Admissibility into legal jurisdictions such as the Daubert standard that federal courts use (testability, peer review, error rate, general acceptance), and Frye for general acceptance in many states is discussed. While the more traditional DNA profiling of STRs remains widely accepted, recent DNA technologies including IGG and whole genome sequencing come with more stringent acceptance issues due to a lack of validation and risk of error. Privacy is a significant issue as genomic information can reveal personal details and family members.

This brings up potential issues regarding searching for relatives of defendants, the role of consent in consumer-based databases such as GEDmatch, and the long-term storage of data within CODIS. Inadequate protections against data misuse and breach may facilitate unwanted surveillance or discrimination under GDPR and U.S. Constitutional privacy standards.

The paper proposes interdisciplinary solutions: standardized bioinformatics validation protocols, judicial training on genomic complexities, tiered data access policies, and ethical frameworks balancing investigative needs with individual rights. As forensic genomics evolves with AI integration, proactive legal and technical governance is essential to maintain public trust and evidentiary integrity in criminal justice.

**Keywords: Forensic Genomics, Bioinformatics, DNA Evidence Admissibility, Privacy Implications, Investigative Genetic Genealogy**

## INTRODUCTION

Bioinformatics has an increasingly important function in the context of forensic genomics through various methodologies and interpretation. For example, with modern high-throughput technology, the interpretation of genetic data is becoming more challenging due to the sheer amount of raw genetic information available, and hence the development of machine learning-based methods for the interpretation is now emerging. The analysis of biological information is becoming essential for numerous research fields, among others, for the context of legal procedures and criminal proceedings in the area of forensic genomics.

Forensic genomics is the utilization of genomic science to study an organism's entire genome (set of DNA) within a legal and forensic context, thus making use of all the DNA that is left at the scene, in form of blood, hair, saliva, skin cells, semen or others.

Forensic genomics is facilitated and largely driven by bio-informatics which comprises elements from biology, statistics and computer sciences and the bio-informatic analysis tools and algorithms make it possible to work and interpret massive volumes of raw sequence data that can be acquired from modern sequencers.

To grasp this further, take this illustration: at the scene of a crime, investigators collect a few cells-a minuscule piece of biological evidence. Lab techs isolate DNA, read it with new high-throughput sequencing (NGS), producing millions of data bits that are otherwise uninformative. Software is then used to stitch these sequence fragments together, locate individual variations such as short tandem repeats or SNPs, search these against huge data banks, and produce the statistical probability of a match. These techniques may also analyze mixtures of DNA (more than one person's DNA), for example using probabilistic genotyping, to assign DNA profiles to individual contributors. This synergy of genomics and forensics has changed crime solving. It has solved cold cases by enabling the use of investigative genetic genealogy (IGG) and crime scene DNA is compared with large consumer genealogy databases to search for the identity of potential culprits, and in the exoneration of falsely accused persons, by disproving matches between defendant's and criminal-evidence DNA. Nevertheless, these exciting tools come with major responsibilities that have to be settled by the courts with regards to the legal standards for evidence (admissibility of results) and society, about the highly private and confidential nature of one's DNA data.

This paper explores these intersections in detail: the technical foundations of bioinformatics in forensics, the legal standards for admitting such evidence in court, and the privacy risks involved when genetic data is used by law enforcement. By making these concepts accessible, it aims to bridge the gap between scientific innovation and legal practice.

## BIOINFORMATICS IN FORENSIC GENOMICS

A forensic workflow consists of sample acquisition, DNA extraction, library preparation, sequencing, and bioinformatic analysis (alignment, variation of calling, and statistical interpretation). STRmix and TrueAllele use probabilistic interpretation of complex DNA mixtures. IGG matches remote familial relatives to forensic profiles using open-source public genealogy datasets.

Bioinformatics is the computation that supports the analysis of molecular information in genomics and related biological disciplines. The main function of bioinformatics is to use computational software, algorithmic approaches, and statistical methods to analyze, manipulate, and manage very large volumes of biological information. In forensics, the data produced from lab tests are converted to high-quality evidentiary profiles that will hold up in the courtroom.

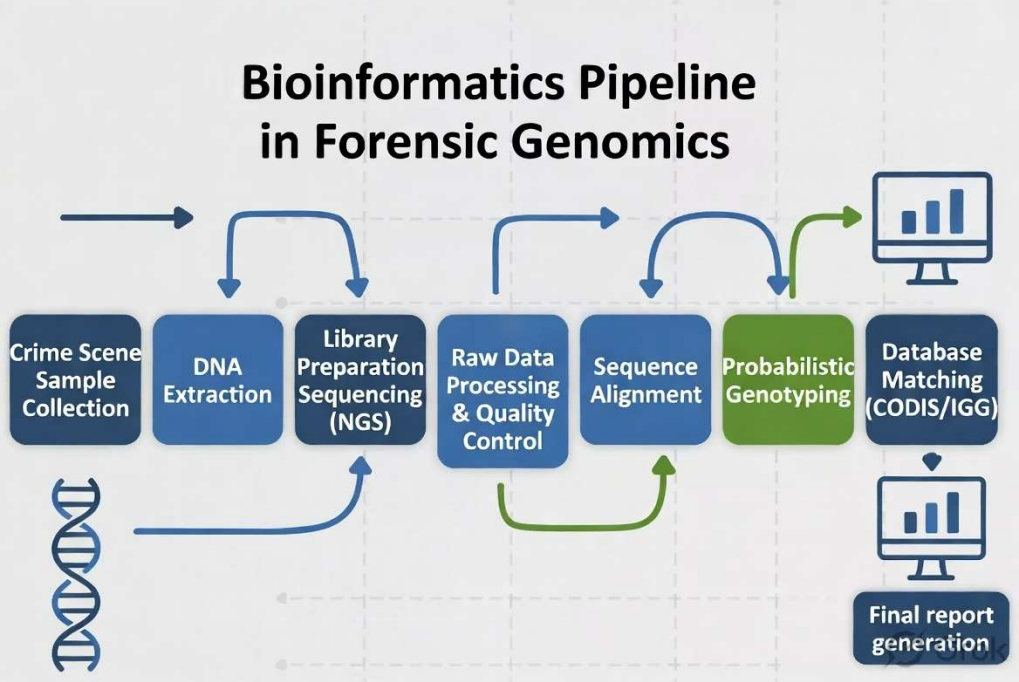
### Core Concepts and Workflow

The starting point for forensic DNA analysis comes in the form of trace biological material obtained from the scene of the crime. Given the minuscule amounts and often damaged condition of samples, high-throughput sequencing methods, including Next-Generation Sequencing (NGS), are used to obtain millions of very short DNA sequences called reads. This deluge of short sequences is noisy and unorganized, requiring bioinformatic analysis to generate useful genetic profiles.

A typical forensic bioinformatics workflow includes the following key stages:

1. **Pre-processing and Quality Control**: Raw sequencing data undergoes rigorous quality assessment using tools such as FastQC. Low-quality bases, adapter sequences, and contaminants are trimmed with programs like Trimmomatic or Cutadapt. This step ensures data integrity and minimizes downstream errors.
2. **Sequence Alignment and Mapping**: Cleaned reads are aligned to a standardized human reference genome (e.g., GRCh38) using aligners such as Burrows-Wheeler Aligner (BWA) or Bowtie2. Alignment identifies the genomic location of each read, forming the foundation for variant detection.
3. **Variant Calling**: Bioinformatics software identifies genetic variations, primarily Short Tandem Repeats (STRs) and Single Nucleotide Polymorphisms (SNPs). Forensic-specific tools have been optimized for the challenges of degraded or low-input DNA. STRs, consisting of repetitive DNA motifs, remain at the gold standard for individual identification due to their high discriminatory power.
4. **Probabilistic Genotyping and Mixture Interpretation**: Modern crime scene samples frequently contain DNA mixtures from multiple contributors. Traditional binary interpretation methods are inadequate for complex mixtures. Advanced bioinformatics platforms, such as STRmix (used by the FBI and many laboratories) and TrueAllele, apply Bayesian statistical models and Markov Chain Monte Carlo (MCMC) algorithms to compute likelihood ratios. These tools assess the probability of a suspect's contribution to the evidence while accounting for allele drop-out, stutter artifacts, and contributor numbers.

5. **Database Searching and Advanced Applications**: Processed profiles are queried against national databases like the FBI's Combined DNA Index System (CODIS). In cold case investigations, Investigative Genetic Genealogy (IGG) integrates forensic SNP profiles with commercial genealogy databases (e.g., GEDmatch or FamilyTreeDNA). Bioinformatics algorithms perform kinship inference and family tree reconstruction, significantly expanding the reach of forensic identification.




**EVIDENTIARY STANDARDS AND ADMISSIBILITY**

Daubert Trilogy in U.S Courts The main criteria the courts look into under the Daubert Trilogy are as follows; 1. Testability (if the theory or technique can test) 2. Peer Review (if the theory or technique has been published in the peer reviewed scientific journal) 3. Error Rate (if the rate of error in applying the theory or technique can be established) 4. Standards (if standards have been developed for applying the theory or technique) 5. Acceptance (if the theory or technique has generally accepted within the relevant scientific community) 6. General Acceptance, Testability and peer reviewed are not mandatory criteria. 7. Traditional STR methodologies fulfil all these readily.

It may be the cases when new methodologies and applications are developed such as; application to low template DNA and/or to analyze new biological markers, such as epigenetic markers in criminal case and it needs more validation to prevent it from any attack from the courts. The application of bioinformatics-driven forensic genomics within the context of criminal procedures and criminal proceedings would depends on whether the presented scientific evidence measures up to these legal standards required to admissibility and reliability in court proceedings. Courts act as gatekeepers for scientific evidence to ensure that only scientifically reliable information are presented to judges and juries, and the admissibility criteria to determine if the evidence will be allowed in the court may include two main legal standards that address novel scientific evidence in general, particularly concerning to testimony related to DNA.

**Major Legal Frameworks**

**1. The Frye Standard ("General Acceptance" Test)**

Originating from *Frye v. United States* (1923), this standard requires that a scientific technique or theory must be "sufficiently established to have gained general acceptance in the particular field in which it belongs." Under Frye, courts defer largely to the relevant scientific community (e.g., forensic geneticists and bioinformaticians) rather than conducting an independent evaluation. Several states, including California and New York, still primarily follow this test for novel evidence.

**Application to Forensic Genomics**: Traditional Short Tandem Repeat (STR) profiling has long satisfied the Frye test due to decades of widespread use, peer-reviewed validation, and laboratory standardization. However, newer bioinformatics methods such as complex mixture interpretation algorithms or Investigative Genetic Genealogy (IGG) may face challenges if they lack broad consensus within the scientific community.

### 2. The Daubert Standard (Reliability Test)

Established by the U.S. Supreme Court in *Daubert v. Merrell Dow Pharmaceuticals* (1993) and refined in subsequent cases (*Joiner* and *Kumho Tire*), this standard is used in federal courts and many states. It requires judges to act as active gatekeepers by evaluating the reliability of the methodology based on several non-exclusive factors:

- Whether the theory or technique can be (and has been) tested (falsifiability);
- Whether it has undergone peer review and publication;
- The known or potential rate of error;
- The existence and maintenance of standards controlling the technique's operation;
- Whether the method is generally accepted in the relevant scientific community.

**Application to Forensic Genomics**: Core STR profiling and standard CODIS database matching are routinely admitted under Daubert. However, advanced bioinformatics tools—such as probabilistic genotyping software for low-template or mixed DNA samples—require detailed validation studies, error rate disclosures, and peer-reviewed publications. Courts may hold *Daubert* hearings (pre-trial) where experts debate the reliability of specific bioinformatics pipelines.

### 3. Implications for Criminal Justice

While bioinformatics has strengthened the probative value of DNA evidence, admissibility disputes can significantly impact case outcomes. Defense attorneys increasingly challenge not just the laboratory results but the underlying computational methods. Judges must therefore possess sufficient understanding or rely on expert testimony to evaluate complex bioinformatics evidence.

This section highlights the critical need for clear validation of protocols and judicial education as forensic genomics continues to advance.

### 4. Privacy Implications in Criminal Justice

The use of bioinformatics in forensic genomics, while transformative for solving crimes, introduces profound privacy challenges. Unlike traditional fingerprints, genomic data contains highly sensitive personal and familial information that extends far beyond identification. This section examines the key privacy concerns arising from the intersection of bioinformatics, forensic genomics, and law enforcement practices.

Genomic data is inherently personal and immutable. A single DNA profile can reveal:

- **Identity and Familial Relationships**: Matches can identify not only the individual but also biological relatives (parents, siblings, children, and distant cousins).
- **Sensitive Attributes**: Through bioinformatics analysis, data may infer ancestry, physical traits (e.g., eye color, height), disease predispositions, or even behavioral tendencies.
- **Long-term Identifiability**: Unlike passwords or credit cards, DNA cannot be changed. Once compromised, privacy loss is permanent.

This creates a ripple effect: uploading a suspect profile to a database can inadvertently expose the genetic privacy of innocent relatives who never consented to testing.

#### Key Privacy Risks in Forensic Applications

1. **Familial Searching and Investigative Genetic Genealogy (IGG)**

Law enforcement increasingly uses consumer DNA databases (e.g., GEDmatch, FamilyTreeDNA) for IGG. A crime scene profile is compared against voluntary consumer uploads to find distant relatives. While effective for cold cases, this practice raises consent issues relatives of database users have no say in whether their genetic information is indirectly searched.

2. **CODIS and Government Databases**

The FBI's Combined DNA Index System (CODIS) stores millions of profiles from convicted offenders, arrestees, and crime scenes. Concerns include:

   a. Indefinite retention of profiles even after exoneration.
   b. Mission creep (expansion beyond violent crimes).
   c. Potential function creeps into non-criminal uses (e.g., immigration or research).

3. **Data Security and Breaches**

Bioinformatics pipelines involve large-scale data storage and transfer. Cyberattacks on forensic or consumer databases could expose sensitive genomic information on a massive scale.

4. **Surveillance and Discrimination Risks**

Widespread genomic surveillance could enable predictive policing or profiling based on genetic markers. There are also risks of genetic discrimination in employment, insurance, or social contexts, even though laws like GINA (Genetic Information Nondiscrimination Act) offer some protections in the U.S.

Within the US legal system, there are generally two opposing standards regarding the admissibility of bioinformatics forensic genomic evidence: the Frye general acceptance test and the Daubert reliability test. Established by the Supreme Court case Frye v. United States (1923), the Frye standard permits novel scientific evidence, provided the technique used to obtain evidence has gained "general acceptance" in the relevant scientific community. This standard is still used in some states in the context of DNA evidence.

In contrast, under the Daubert standard articulated by the US Supreme Court in Daubert v. Merrell Dow Pharmaceuticals (1993), trial judges act as "gatekeepers" in evaluating the reliability of a scientific methodology based on a flexible set of criteria: "whether the technique can be and has been tested; whether it has been subjected to peer review and publication; whether there is a known or potential rate of error; the existence and maintenance of standards controlling the technique's operation; and general acceptance in the relevant scientific community." These principles (expanded by Joiner and Kumho Tire) have since been adopted by federal courts and many states, providing a more robust and flexible framework to test advanced bioinformatics tools like probabilistic genotyping software and investigative genetic genealogy (IGG) techniques.

When applied in the forensic genomics arena, courts generally distinguish between established and emerging techniques. Longstanding Short Tandem Repeat (STR) profiling via capillary electrophoresis is widely accepted by both Frye and Daubert due to decades of rigorous validation, standardization via the CODIS loci and extensive peer-reviewed scientific publications. However, newer bioinformatics-derived evidence, including probabilistic genotyping of complex DNA mixtures using software (e.g., STRmix, TrueAllele), or the use of investigative genetic genealogy with consumer genealogy databases, may undergo higher judicial scrutiny.

Often, Daubert hearings will be held, allowing for an in-depth review of validation studies, potential error rates with low-level or mixed samples, algorithm transparency, and other factors that could introduce bias.

The defense, in turn, often complains about proprietary "black box" software with limited code availability, hindering cross-examination and verification. Comparison to the international legal system highlights some contrasts, with the UK having strict data retention rules (Protection of Freedoms Act 2012) and the European Court of Human Rights stressing protective safeguards against the indefinite retention of DNA data. In essence, the law requires that bioinformatics evidence is not only scientifically sound but that its methodology is explainable, ensuring juries do not blindly rely on powerful statistical output such as likelihood ratios without sufficient understanding.

## SUGGESTION AND RECOMMENDATION

The convergence of bioinformatics and forensic genomics has undeniably strengthened the criminal justice system by providing highly accurate and objective evidence capable of solving previously intractable cases and exonerating the innocent. However, this technological progress has outpaced the development of appropriate legal safeguards, creating significant challenges in evidentiary reliability, judicial gatekeeping, and the protection of privacy rights. Courts continue to struggle with the evaluation of complex bioinformatics tools under both the *Frye* and *Daubert* standards, often lacking the specialized knowledge needed to assess algorithm transparency, validation studies, and potential sources of error. At the same time, the expansive reach of investigative genetic genealogy and large-scale DNA databases raises profound privacy concerns, as genomic information inherently reveals sensitive details not only about individuals but also about their biological relatives who never consented to testing or database inclusion.

A central tension lies in balancing the societal interest in effective crime-solving with the fundamental rights to privacy and due process. While traditional STR profiling enjoys broad acceptance, newer methods involving probabilistic genotyping and consumer database searches demand stricter scrutiny to prevent miscarriages of justice and unwarranted surveillance. The current fragmented legal landscape, characterized by inconsistent state laws and reliance on agency guidelines rather than comprehensive legislation, is inadequate for addressing the rapid evolution of technology. International comparisons reveal that jurisdictions such as the European Union and the United Kingdom have implemented more robust proportionality requirements and data retention limits, offering valuable models for reform. Without proactive intervention, risks such as algorithmic bias, function creep, and erosion of public trust in the justice system are likely to intensify.

To address these issues, several key recommendations emerge. First, federal legislation should be enacted to establish uniform national standards for the validation and use of bioinformatics tools in forensic contexts, including mandatory warrants for investigative genetic genealogy searches and clear rules for data retention and deletion. Second, judicial education programs must be developed to equip judges and attorneys with the necessary understanding of genomic science and computational methods, thereby improving the quality of *Daubert* hearings and evidentiary rulings. Third, forensic laboratories and software developers should prioritize transparency by adopting auditable algorithms and conducting independent validation studies, with full disclosure of limitations and error rates. Fourth, privacy protections should be strengthened through privacy-by-design principles in bioinformatics systems, tiered access controls for databases, and the creation of

independent oversight bodies that include civil liberties experts. Finally, sustained investment in interdisciplinary research and international cooperation is essential to develop bias-mitigating technologies and harmonized global standards for cross-border genomic data sharing.

In conclusion, while bioinformatics and forensic genomics hold tremendous potential to advance justice, realizing this potential requires deliberate legal and policy reforms. By implementing these recommendations, the criminal justice system can harness scientific innovation responsibly while safeguarding individual rights and maintaining public confidence.

### Recommendations

To address these challenges, the following recommendations are proposed:

1. **Legislative Reform**: Congress should enact comprehensive federal legislation governing the use of forensic genomics and bioinformatics. This law should establish uniform validation standards for bioinformatics pipelines, mandate warrants for IGG searches, and set clear rules for data retention and expungement.
2. **Judicial Education and Resources**: Courts should implement mandatory training programs for judges and legal practitioners on bioinformatics and forensic genomics. The development of bench guides and expert panels could assist judges during Daubert hearings.
3. **Transparency and Validation Standards**: Forensic laboratories and software developers must adopt open-source or auditable algorithms where possible. All bioinformatics tools used in criminal cases should undergo rigorous, independent validation according to NIST and SWGDAM guidelines, with full disclosure of error rates and limitations.
4. **Privacy Protections**: Adopt a tiered access model for genomic databases, requiring explicit consent for law enforcement use of consumer data. Implement privacy-by-design principles in bioinformatics tools and establish independent oversight boards with civil liberties experts.
5. **Research and Interdisciplinary Collaboration**: Fund research into bias mitigation in forensic algorithms and the development of privacy-preserving computational methods. Foster collaboration between forensic scientists, bioinformaticians, legal scholars, and ethicists.
6. **International Harmonization**: The United States should engage in international forums to develop global standards for cross-border genomic data sharing in criminal investigations.

## CONCLUSION

The integration of bioinformatics and forensic genomics into the criminal justice system represents one of the most significant technological advancements in law enforcement in recent decades. By enabling precise identification from trace evidence, solving long-standing cold cases, and helping to exonerate the wrongfully convicted, these tools have the potential to enhance the accuracy and fairness of judicial outcomes. However, as this paper has demonstrated, the benefits are accompanied by substantial legal and ethical challenges. Evidentiary standards, particularly under the *Daubert* and *Frye* frameworks, must evolve to adequately assess the reliability of complex bioinformatics methods such as probabilistic genotyping and investigative genetic genealogy. At the same time, the profound privacy implications of genomic data which extend beyond the individual to biological relatives and future generations demand robust legal protections that current frameworks often fail to provide.

Ultimately, the responsible use of bioinformatics in forensic genomics requires a balanced approach that prioritizes both public safety and individual rights. Without deliberate action, the rapid advancement of technology risks outpacing legal safeguards, potentially leading to unreliable evidence in courtrooms, unwarranted surveillance, and erosion of public trust in the justice system. To prevent these outcomes, policymakers, courts, forensic scientists, and bioinformaticians must collaborate to develop clear validation standards, enhance judicial literacy, strengthen privacy protections, and promote transparency in algorithmic tools.

In conclusion, forensic genomics powered by bioinformatics offers a powerful opportunity to strengthen justice, but only if accompanied by thoughtful legal and ethical governance. The coming years will be critical in determining whether these technologies serve as instruments of fairness and truth-seeking or become sources of new injustices. Proactive reform today will ensure that scientific progress aligns with constitutional values and societal expectations for many years to come.

## REFERENCES

1. Budowle, B., Schmedes, S. E., & Wendt, F. R. (2023). Revisiting informed consent in forensic genomics considering current practices and emerging technologies. *Forensic Science International: Genetics*, 62, 102789. https://doi.org/10.1016/j.fsigen.2022.102789

2. Butler, J. M. (2015). The future of forensic DNA analysis. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370(1674), 20140252. https://doi.org/10.1098/rstb.2014.0252
3. Cole, S. A. (2018). Forensic science and the law: The case of probabilistic genotyping. *Annual Review of Criminology*, 1, 141–160. https://doi.org/10.1146/annurev-criminol-032317-092220
4. FBI. (2025). *CODIS and NDIS Statistics*. Federal Bureau of Investigation. https://www.fbi.gov/services/laboratory/biometric-analysis/codis/ndis-statistics
5. Gill, P., Gusmão, L., Parson, W., & Phillips, C. (2022). The future of forensic DNA analysis: Challenges and opportunities. *Forensic Science International: Genetics*, 58, 102678. https://doi.org/10.1016/j.fsigen.2022.102678
6. Kaye, D. H. (2010). The admissibility of DNA evidence in court. *Journal of Law and Policy*, 18(1), 1–45.
7. Murphy, E. (2018). *Inside the cell: The dark side of forensic DNA*. Nation Books.
8. National Research Council. (2009). *Strengthening forensic science in the United States: A path forward*. The National Academies Press. https://doi.org/10.17226/12589
9. Phillips, C., et al. (2021). The role of SNPs in forensic genetics: Current applications and future directions. *Forensic Science International: Genetics*, 51, 102456. https://doi.org/10.1016/j.fsigen.2020.102456
10. Ram, N. (2021). The genealogy of genetic privacy. *Harvard Law Review*, 134(6), 1980–2048.
11. Roth, A. (2018). The admissibility of DNA evidence in court. *Berkeley Journal of Criminal Law*, 23(1), 1–45.
12. Saks, M. J., & Koehler, J. J. (2005). The coming paradigm shift in forensic identification science. *Science*, 309(5736), 892–895. https://doi.org/10.1126/science.1111565
13. SWGDAM. (2020). *Validation guidelines for DNA analysis methods*. Scientific Working Group on DNA Analysis Methods.
14. United States Department of Justice. (2022). *Interim policy on investigative genetic genealogy*. https://www.justice.gov/olp/page/file/1500001/download
15. van der Beek, K., et al. (2023). Ethical and legal challenges of forensic genetic genealogy. *Forensic Science International: Genetics*, 65, 102876. https://doi.org/10.1016/j.fsigen.2023.102876
16. Williams, R., & Wienroth, M. (2021). *Forensic genetics and the law: New technologies, old problems*. Routledge.