Forensic Genetics and Its Applications: A Survey

Abstract

Forensic genetics represents a pivotal intersection of genetic analysis and legal application, crucial for individual identification and familial relationship establishment in legal contexts. This survey paper provides a comprehensive exploration of forensic genetics, emphasizing its historical development, methodologies, and applications. Key techniques such as short tandem repeats (STRs) and single nucleotide polymorphisms (SNPs) are discussed, highlighting their roles in DNA profiling and kinship analysis. The survey also examines the integration of genetic genealogy within law enforcement, underscoring its effectiveness in resolving cold cases and tracing ancestry. The ethical considerations, particularly regarding privacy and data security, are critically analyzed, advocating for robust regulatory frameworks to safeguard genetic data. Innovative advancements in DNA profiling, including the use of massively parallel sequencing (MPS) and computational tools, are explored, showcasing their potential to enhance forensic analyses. The paper concludes by reflecting on the transformative impact of forensic genetics on modern legal systems and the need for continued innovation and ethical oversight to ensure its responsible application. Overall, this survey highlights the significant contributions of forensic genetics to legal investigations and its potential for future advancements.

1 Introduction

1.1 Concept of Forensic Genetics

Forensic genetics applies genetic analysis within legal contexts, playing a critical role in identifying individuals and establishing biological relationships [1]. This field encompasses the analysis of genetic material to address legal issues, significantly contributing to paternity cases and familial relationship identification [2]. Central to forensic investigations, it focuses on DNA analysis, essential for examining biological evidence such as mixed DNA stains to identify contributors, thereby enhancing forensic accuracy [3, 4]. Through these methodologies, forensic genetics serves as a cornerstone of justice, providing reliable identification methods for victims and suspects alike.

1.2 Scope and Objectives

This survey offers a comprehensive exploration of forensic genetics, focusing on its historical development, methodologies, and diverse applications. It aims to analyze DNA profiling techniques, including short tandem repeats (STRs), single nucleotide polymorphisms (SNPs), insertion deletions, and sequencing methods [5]. A key objective is to elucidate the forensic applications of Y-chromosome DNA, particularly in cases where standard autosomal DNA profiling is inadequate, alongside reviewing advancements in X-chromosomal markers in population and forensic genetics over the past two decades. The integration of direct-to-consumer genetic genealogy databases within law enforcement is also examined, highlighting their effectiveness in solving cold cases and offering insights into heritability and relatedness via genome-wide SNP markers. Furthermore, the survey addresses forensic serology, incorporating RNA, proteomic, and epigenetic markers, and discusses the implications of genetic data use in forensic and medical genetics. It includes methods for victim
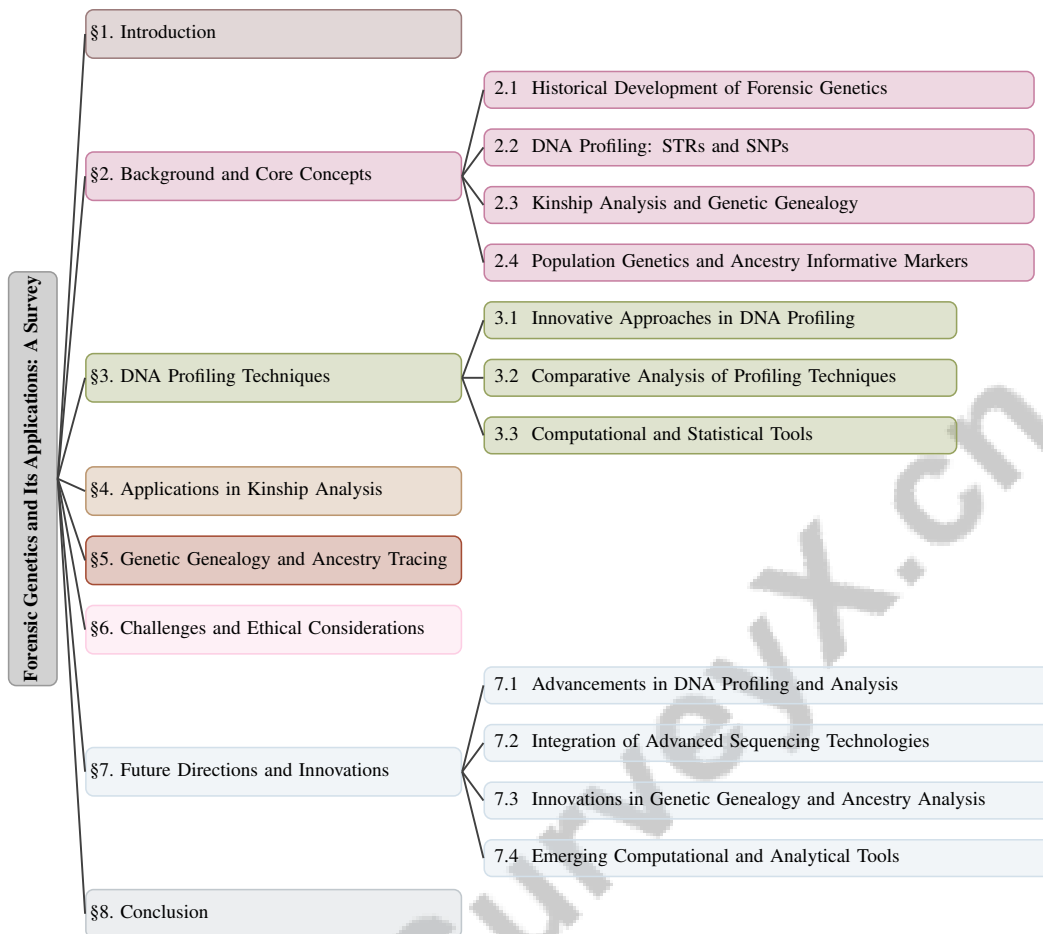
Figure 1: chapter structure

identification in both human and natural catastrophes, utilizing genetic identification techniques and kinship analysis [6]. Additionally, the review explores DNA transfer, persistence, prevalence, and recovery (DNA-TPPR), focusing on factors influencing DNA transfer and detection likelihood post-contact [7]. The primary objectives are to provide an overview of forensic genetics, evaluate methodologies and applications, and discuss ethical considerations [1].

1.3 Significance in Legal Contexts

Forensic genetics is crucial in legal proceedings due to its ability to provide reliable, scientifically robust evidence that can significantly influence the outcomes of criminal investigations and court cases [1]. Its application is especially vital in sexual assault and paternity cases, where it assists in distinguishing male DNA from mixed samples for precise identification [8]. The unique inheritance characteristics of X-chromosomal markers enhance its utility in kinship testing, offering solutions in complex familial scenarios [9].

The field's importance is underscored by its capacity to accurately infer genetic relationships and identities, even when foundational genetic assumptions may not hold [3]. This capability is essential for establishing identities at crime scenes and identifying victims in disasters, where traditional methods may be insufficient. The integration of forensic genetics into legal contexts aids in resolving cases involving suspects and unknown perpetrators while playing a critical role in humanitarian efforts, particularly when human remains are severely degraded.

Moreover, the incorporation of genetic genealogy databases into forensic investigations has expanded law enforcement's capabilities in identifying suspects and missing persons. While these databases present new opportunities for solving cases, they also raise ethical concerns regarding informed

consent and privacy, necessitating a careful balance between investigative benefits and individual rights. Understanding variables affecting DNA transfer is essential for forensic investigations, as it influences evidence interpretation and the likelihood of DNA detection post-contact. With ongoing advancements in forensic genetics, there is a pressing need for evaluations to incorporate activity level propositions in legal contexts, emphasizing the necessity for improved clarity and reliability in evidence interpretation during legal proceedings. Enhanced methodologies, such as probabilistic genotyping and Bayesian networks, are vital for forensic biologists to effectively assess and communicate their findings, ensuring the judiciary can make informed inferences regarding DNA evidence activity levels. Continued research into DNA transfer and persistence is crucial for developing accurate probability estimates and ensuring forensic practitioners are adequately trained to navigate these complexities [10, 11, 7, 12].

1.4 Structure of the Survey

This survey is meticulously organized to provide a comprehensive exploration of forensic genetics and its multifaceted applications within legal contexts. The paper begins with an **Introduction**, establishing the foundation by defining forensic genetics, outlining the scope and objectives, and emphasizing its significance in legal proceedings.

Following the introduction, the survey delves into the **Background and Core Concepts**, offering an overview of the historical development of forensic genetics and elucidating key concepts such as DNA profiling, kinship analysis, and genetic genealogy, which collectively provide the necessary foundation for understanding the methodologies employed in forensic investigations.

The third section, **DNA Profiling Techniques**, focuses on methodologies used in DNA profiling, particularly short tandem repeats (STRs) and single nucleotide polymorphisms (SNPs), discussing their application in forensic contexts. It also explores innovative approaches, comparative analyses of different profiling techniques, and the computational tools utilized for data interpretation.

In **Applications in Kinship Analysis**, the paper examines the use of forensic genetics in determining familial relationships, emphasizing methodologies, case studies, and challenges faced in this domain. This section highlights the practical applications of forensic genetics in paternity testing and familial relationship determination.

The survey transitions to **Genetic Genealogy and Ancestry Tracing**, investigating the role of genetic genealogy in law enforcement, the integration with direct-to-consumer genetic databases, and the implications for ancestry tracing. This section addresses both technological advancements and ethical considerations involved.

Challenges and Ethical Considerations are discussed in the subsequent section, identifying issues such as privacy concerns, data security, and potential misuse, underscoring the need for regulatory frameworks and guidelines governing the application of forensic genetics.

The penultimate section, **Future Directions and Innovations**, explores emerging technologies and methodologies in forensic genetics, highlighting advancements in DNA profiling and the integration of advanced sequencing technologies, as well as innovations in genetic genealogy and the development of new computational tools.

The summarizes the key discussions presented throughout the paper, emphasizing the critical role of forensic genetics in the legal system. It reflects on the implications of advanced forensic techniques, such as probabilistic genotyping and activity level propositions, which enhance the interpretation of DNA evidence in court. Additionally, it considers the potential for future advancements in forensic genetics, highlighting the need for ongoing research, training, and the establishment of robust guidelines to improve the reliability and transparency of evidence evaluation in legal contexts [10, 7, 11, 13, 14]. This structured approach ensures a thorough examination of forensic genetics, providing insights into its current applications and future possibilities.The following sections are organized as shown in Figure 1.

2 Background and Core Concepts

2.1 Historical Development of Forensic Genetics

Forensic genetics has evolved significantly since the introduction of DNA profiling in the 1980s, which revolutionized forensic science by establishing DNA analysis as a critical component of legal investigations. This advancement enhanced the accuracy and reliability of genetic evidence in courtrooms [15, 16]. Over the years, DNA profiling techniques have become more sophisticated, expanding forensic applications and strengthening the credibility of genetic evidence [17]. The focus on single nucleotide polymorphisms (SNPs) has enriched forensic analyses by enabling precise trait associations, though challenges remain in extracting SNP-trait relationships from biomedical literature [18]. Y-chromosome analysis has been pivotal in tracing paternal lineages and solving complex cases where traditional methods are inadequate [8]. The integration of molecular genetics into various domains, such as livestock breeding and disease management, underscores the versatility of genetic technologies beyond human forensics [19]. These developments illustrate the dynamic progression of forensic genetics and its foundational role in modern legal systems.

2.2 DNA Profiling: STRs and SNPs

DNA profiling is central to forensic genetics, employing short tandem repeat (STR) and single nucleotide polymorphism (SNP) analyses to generate unique genetic profiles crucial for individual identification. STR analysis involves polymerase chain reaction (PCR) amplification and capillary electrophoresis (CE) to detect variations in repeat units, making it effective for human identification and resolving familial relationships due to STR loci's high polymorphism [1]. However, it struggles with degraded samples with low DNA content [20]. Advanced methodologies like the ForenSeq Kintelligence kit address these challenges by enabling high-density SNP profiling from low-quality samples, enhancing forensic robustness [21]. SNP analysis complements STR profiling by offering insights into genetic diversity and trait prediction, though SNP chips may underperform with rare variants, risking clinical misdiagnoses [22]. Y-STR haplotyping is vital for tracing paternal lineages and resolving mixed DNA samples where autosomal STR profiling falls short [8]. Innovations such as SNP-SNP microhaplotype detection and SNP-STR marker integration in multiplex PCR assays improve minor DNA component detection in mixtures [23, 4]. Statistical analysis of Y chromosome STR loci mixtures is challenging due to locus linkage, necessitating probabilistic models and Bayesian networks for effective management of complex mixtures [3, 24]. Evaluating highly informative SNPs, as demonstrated in the Persian population, enhances forensic tests and paternity analysis, underscoring DNA profiling's critical role in forensic science [25].

2.3 Kinship Analysis and Genetic Genealogy

Kinship analysis and genetic genealogy are crucial in forensic genetics, offering advanced methods for establishing familial relationships and tracing ancestry, particularly in catastrophic events. These techniques utilize SNP genotyping and genome-wide relatedness methods to accurately identify unknown remains and resolve cases with degraded samples. Forensic genetic genealogy enhances the connection of unidentified victims with their families, even when traditional methods fail [26, 6]. Genetic markers, including autosomal, Y-chromosomal, and X-chromosomal markers, provide unique insights into genetic relationships. X-chromosomal markers are particularly useful in complex kinship scenarios, such as mother-son or father-daughter relationships, due to their distinct inheritance patterns [6, 27, 9, 28, 2]. Evaluating genome-wide relatedness methods and identity by descent (IBD) segment approaches is critical, especially given challenges like missing data and genotyping errors in forensic datasets [26]. Accurate detection of Mendelian inconsistencies in SNP data is essential to prevent erroneous genotypes, ensuring kinship analysis reliability [29]. Kinship analysis is vital for victim identification, especially when family member samples are limited [6]. Integrating X-STR loci analysis provides additional genetic diversity information, enhancing kinship inference accuracy in diverse populations [27]. Genetic genealogy assists law enforcement in resolving cold cases by linking profiles to genealogical databases. The inference of relationships from DNA mixtures underscores the necessity of robust statistical and computational tools in forensic genetics [30]. Recent advancements in kinship analysis and genetic genealogy methodologies have significantly improved forensic genetics' capacity to deliver reliable insights into familial relationships and ancestry. These developments are vital in legal and investigative contexts, enhancing victim identification accuracy

4

and cold case resolution, thereby underscoring genetics' crucial role in modern forensic investigations [26, 6].

2.4 Population Genetics and Ancestry Informative Markers

Population genetics is integral to forensic applications and ancestry tracing, offering insights into genetic diversity and evolutionary history. SNP analysis is central to understanding population structure and genetic stratification, critical for accurate forensic investigations and ancestry analysis [31]. SNP genotyping technologies, such as the SBE/SBH architecture, enable high-throughput analysis of specific SNPs, identifying population-specific genetic markers and enhancing forensic resolution [32]. Ancestry informative markers (AIMs) differentiate populations with distinct ancestral backgrounds, aiding forensic cases involving unknown individuals and providing biogeographical context to genetic evidence [5]. Advanced sequencing technologies, including massively parallel sequencing, refine complex DNA sample analysis, improving ancestry inference accuracy [20]. Addressing challenges associated with low template DNA (LTDNA) samples, which often contain complex mixtures, is crucial. Advanced statistical models, such as Bayesian networks and the multivariate Dirichlet-multinomial distribution, enhance mixture analysis, enabling precise likelihood ratio calculations that account for genotyping errors and sub-population effects [33, 34, 30]. Computational tools, such as suffix arrays, facilitate efficient genomic data storage and retrieval, expediting SNP database analysis and enhancing forensic scalability [35]. Benchmarking toolkits standardize genotyping technology evaluations, ensuring forensic genetic analyses' reliability [36]. Incorporating population genetics into forensic science enhances evidence interpretation by emphasizing genetic diversity and stratification roles, employing advanced probabilistic models and Bayesian networks. These methodologies enable forensic biologists to assess DNA evidence more accurately, aiding the judiciary in making informed decisions based on robust scientific findings [10, 37]. As forensic genetics evolves, developing sophisticated tools and methodologies will further refine ancestry inference and population-based analyses, enhancing the capability to resolve complex cases and trace human ancestry.

3 DNA Profiling Techniques

Category	Feature	Method
Innovative Approaches in DNA Profiling	Sensitivity and Amplification Techniques	TX19[2], SNP-SNP[23]
	Computational and Data Management	CSASD[35]
Comparative Analysis of Profiling Techniques	Evolutionary Relationships	MYC[38]
Computational and Statistical Tools	Genetic Variation Analysis	ILSDS[24]

Table 1: This table provides a comprehensive overview of the key methodologies employed in recent advancements in DNA profiling. It categorizes these methodologies into innovative approaches, comparative analysis of profiling techniques, and computational and statistical tools, highlighting specific features and methods utilized within each category. The table serves as a succinct summary of the diverse strategies enhancing the precision and applicability of forensic genetic analyses.

Recent advancements in DNA profiling have significantly bolstered the accuracy and reliability of forensic genetic analyses. These developments address the complexities inherent in forensic investigations, offering more precise tools for genetic analysis. As illustrated in Figure 2, the hierarchical structure of advancements in DNA profiling techniques can be categorized into three main areas: innovative approaches, comparative analysis, and computational tools. Table 4 provides a comprehensive summary of the methodologies underpinning these advancements, categorizing them into the same three areas, which are instrumental in shaping modern forensic genetics. Key innovations highlighted in the figure include sequencing technologies and analytical methods, while the comparative analysis contrasts traditional and modern techniques, emphasizing SNP and haplotype analysis. Furthermore, the computational tools are divided into frameworks and statistical models, underscoring their critical role in enhancing forensic genetics. The subsequent subsections explore these innovative approaches and comparative analyses in DNA profiling, thereby underscoring key advancements shaping the future of forensic genetics.
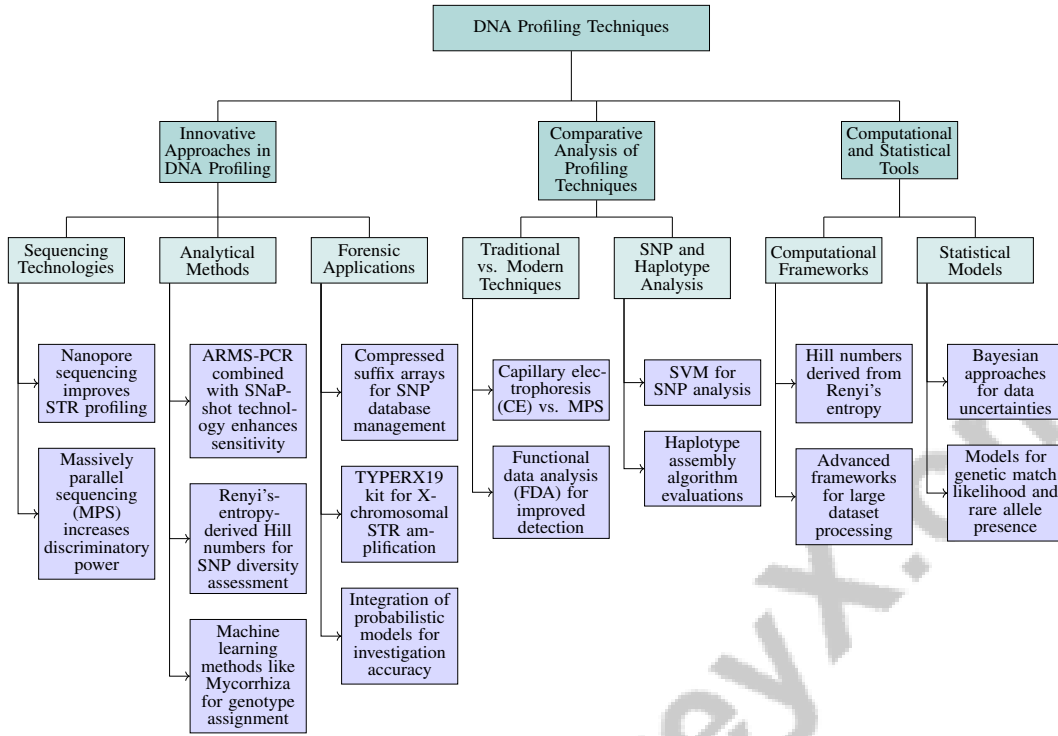
Figure 2: This figure illustrates the hierarchical structure of advancements in DNA profiling techniques, categorized into innovative approaches, comparative analysis, and computational tools. Key innovations include sequencing technologies and analytical methods, while comparative analysis contrasts traditional and modern techniques, emphasizing SNP and haplotype analysis. Computational tools are divided into frameworks and statistical models, highlighting their role in enhancing forensic genetics.

Method Name	Technological Innovations	Computational Methods	Forensic Applications
SNP-SNP[23]	Massively Parallel Sequencing	Advanced Algorithms	Forensic Scenarios
ILSDS[24]	-	Hill Numbers	-
MYC[38]	-	Machine Learning Method	Wildlife Forensics
CSASD[35]	Nanopore Sequencing	Advanced Algorithms	Improved Genotype Assignment
TX19[2]	Typerx19 Kit	-	Kinship Case Resolution

Table 2: Overview of innovative DNA profiling methods highlighting technological advancements, computational techniques, and forensic applications. The table delineates various methods, detailing their specific innovations and applications in forensic genetics.

3.1 Innovative Approaches in DNA Profiling

Innovative DNA profiling techniques are enhancing forensic genetic analysis precision and applicability. Nanopore sequencing has improved STR profiling, particularly for loci with homopolymers and complex structures, overcoming traditional sequencing limitations [39]. Massively parallel sequencing (MPS) increases discriminatory power, aiding human identification and mixture deconvolution [21, 20]. The ARMS-PCR combined with SNaPshot technology enhances sensitivity and specificity in detecting minor DNA components in degraded samples [23]. Advanced computational methods, such as Renyi's-entropy-derived Hill numbers, offer new metrics for assessing SNP diversity and similarity, crucial for precise individual identification [24]. Machine learning methods, like the Mycorrhiza approach, use phylogenetic networks for improved genotype assignment [38]. Compressed suffix arrays efficiently manage SNP databases, enhancing forensic scalability [35]. The TYPERX19 kit advances X-chromosomal STR amplification, aiding complex kinship case resolution [2].

Table 2 provides a comprehensive overview of the innovative methods in DNA profiling, showcasing the integration of technological innovations, computational methods, and their specific forensic applications. Figure 3 illustrates the key innovative approaches in DNA profiling, categorizing them

into sequencing technologies, computational methods, and forensic applications. Each category highlights specific advancements and their contributions to forensic genetics. These innovations expand forensic genetics' capabilities, ensuring genetic evidence remains central to modern forensic practice. The integration of probabilistic models and activity level propositions further improves forensic investigation accuracy and reliability, supporting informed legal decisions [37, 10, 11].
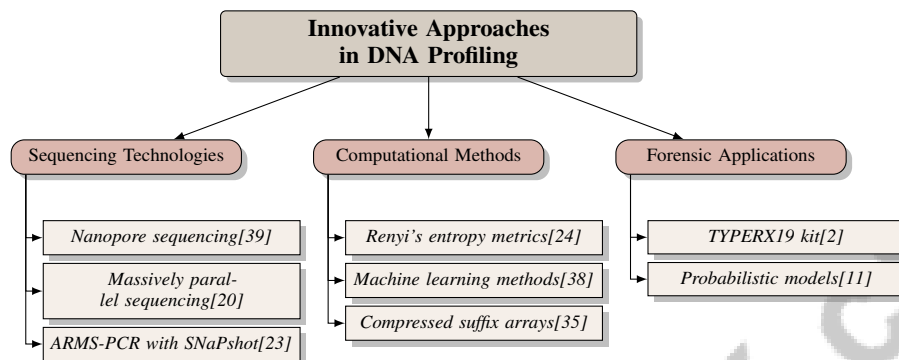

Figure 3: This figure illustrates the key innovative approaches in DNA profiling, categorizing them into sequencing technologies, computational methods, and forensic applications. Each category highlights specific advancements and their contributions to forensic genetics.

3.2 Comparative Analysis of Profiling Techniques

Benchmark	Size	Domain	Task Format	Metric
STR-NP[39]	726,328	Forensic Genetics	Str Genotyping	True Allele Alignment, Genotyping Accuracy
PENACC:E[26]	264	Forensic Genetics	Kinship Analysis	Classification Accuracy, RMSE
Y-STR-HG[40]	20,403	Forensic Genetics	Haplogroup Prediction	Accuracy, Error Rate
STR-SinoTibetan[41]	803	Forensic Genetics	Population Affinity Analysis	Discrimination Power, Hardy-Weinberg Equilibrium
MEC[42]	1,000,000	Genomics	Haplotype Assembly	MEC, Accuracy
SMASH[36]	5,000,000	Variant Calling	Variant Detection	Recall, Precision
SNP-Benchmark[22]	49,908	Genetics	Genotyping	Sensitivity, Specificity
MTB[1]	100,000	Text Classification	Multi-task Evaluation	Accuracy, F1-score

Table 3: This table presents a comprehensive overview of various benchmarks used in genomic and forensic genetics research. It details the size, domain, task format, and evaluation metrics for each benchmark, facilitating a comparison of their applicability and effectiveness in different contexts. The information is crucial for selecting appropriate benchmarks for specific DNA profiling and genetic analysis tasks.

Comparative analysis of DNA profiling techniques is crucial for evaluating their forensic effectiveness. Traditional capillary electrophoresis (CE) offers reliable results but limited genetic variation detection [5]. In contrast, MPS uncovers a broader spectrum of variations, essential for complex investigations [21]. Functional data analysis (FDA) improves detection power and association identification in genetic data [38]. Support vector machines (SVM) enhance SNP analysis, facilitating genetic variation classification [43, 37, 44]. In Y-chromosome analysis, models focus on minimizing haplogroup misclassification errors [39]. SNP identification methods, like DUSI, emphasize precision and recall [21]. Haplotype assembly algorithm evaluations highlight sequencing technology performance differences, crucial for degraded sample analysis [33, 10, 21, 26]. High-throughput SNP genotyping technologies assess decodable SNP pools and heuristic efficiency [38]. Table 3 provides a detailed overview of representative benchmarks used in the comparative analysis of DNA profiling techniques, highlighting their size, domain, task format, and relevant metrics. This comparative analysis guides forensic investigators in selecting methodologies tailored to specific case demands, ensuring reliable DNA evidence interpretation [7, 11]. Advancements in sequencing technologies and computational methods continue to enhance DNA profiling effectiveness and reliability.

As shown in Figure 4, DNA profiling techniques are pivotal in forensic science and genetic research for individual identification and understanding genetic relationships. The comparative analysis

(a) Overall accuracy by population across different missingness and error levels[26]
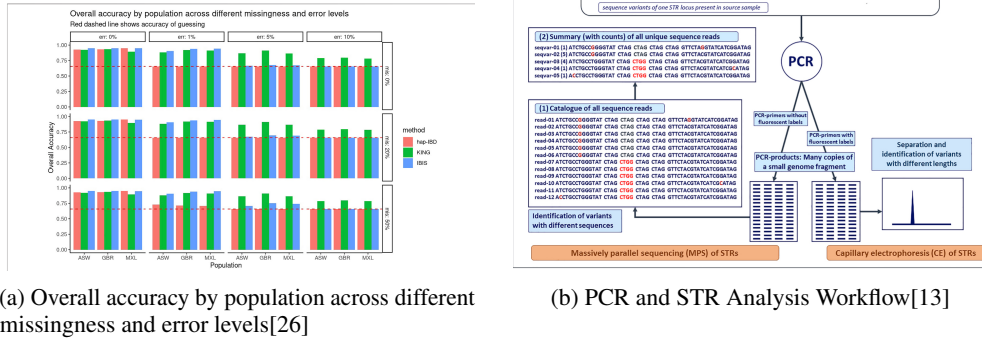

(b) PCR and STR Analysis Workflow[13]

Figure 4: Examples of Comparative Analysis of Profiling Techniques

highlights the overall accuracy of different methods across populations and the PCR and STR analysis workflow. These visual representations provide valuable insights into the effectiveness and procedural intricacies involved in DNA profiling [26, 13].

3.3 Computational and Statistical Tools

Sophisticated computational and statistical tools underpin DNA profiling data analysis in forensic genetics, managing genetic information complexity and volume. Hill numbers derived from Renyi's entropy define SNP diversity, enhancing individual genetic analysis precision [24]. Advanced computational frameworks process large forensic datasets efficiently, enabling meaningful conclusions from complex DNA profiles. Machine learning techniques, particularly SVM, enhance forensic genetic analyses by classifying individuals based on unique genetic markers, improving evidence interpretation accuracy [43, 3, 44, 37, 34]. Statistical models assess genetic match likelihood and rare allele presence, establishing a comprehensive framework for forensic investigation significance [10, 30, 3, 34, 37]. Bayesian approaches accommodate genetic data uncertainties, enhancing forensic conclusions' reliability. The integration of computational and statistical tools ensures DNA evidence is analyzed accurately and reliably, supporting robust legal context application [10, 37].

Feature	Innovative Approaches in DNA Profiling	Comparative Analysis of Profiling Techniques	Computational and Statistical Tools
Technological Innovation	Nanopore Sequencing	Massively Parallel Sequencing	Advanced Frameworks
Computational Method	Renyi's Entropy	Support Vector Machines	Bayesian Approaches
Forensic Application	Str Profiling	Snp Analysis	Genetic Match Likelihood

Table 4: This table presents a comparative overview of the key methodologies in DNA profiling, categorized into technological innovations, computational methods, and forensic applications. It highlights the distinct approaches and tools used in modern forensic genetics, emphasizing their roles in enhancing the accuracy and reliability of genetic analysis.

4 Applications in Kinship Analysis

4.1 Methodologies and Technologies in Kinship Analysis

Kinship analysis in forensic genetics relies on advanced methodologies and technologies to ascertain familial relationships and lineage through genetic markers. Key genetic markers, such as short tandem repeats (STRs) and single nucleotide polymorphisms (SNPs), are indispensable for DNA profiling, particularly in challenging samples [1]. These markers are foundational in establishing genetic relationships, a critical application in forensic genetics.

As illustrated in Figure 5, the hierarchical structure of methodologies and technologies utilized in kinship analysis is categorized into genetic markers, technologies, and computational methods. This categorization underscores how these components enhance the precision and robustness of forensic genetics.

Y-chromosome haplotyping is crucial for identifying male contributors in complex mixtures or degraded samples, leveraging the Y chromosome's unique inheritance to trace paternal lineages

8

when autosomal markers are inadequate [8]. Technologies like the TYPERX19 method enhance kinship analysis precision by improving genetic profiling through specialized kits for blood sample amplification [2]. This yields detailed genetic profiles essential for resolving intricate familial relationships.

SNP-SNP detection methodologies exhibit high sensitivity in identifying minor DNA components, proving effective in scenarios where traditional methods falter [23]. This bolsters forensic analyses' robustness by detecting low-abundance genetic material. Multiplex PCR assays targeting SNP-STR regions mitigate major DNA masking effects, enabling low-quantity genetic material detection [4], crucial in compromised sample quality contexts.

Phylogenetic methods, including phylogenetic split systems, train Random Forest classifiers to predict population origins, combining phylogenetics and machine learning for a robust framework in understanding genetic relationships [38]. Constrained steepest descent and linear fractional programming methodologies enhance kinship analysis sensitivity, ensuring accurate genetic data interpretation [3]. These computational techniques improve analytical precision in kinship studies.

Advancements in genetic research and computational techniques, including unified formulas for likelihood ratio calculations and improved data sharing practices, drive rapid progress in kinship analysis methodologies and technologies [28, 45, 6]. These developments enhance relationship identification accuracy in forensic contexts, facilitating complex genetic relationship analyses and improving kinship analysis efficiency.
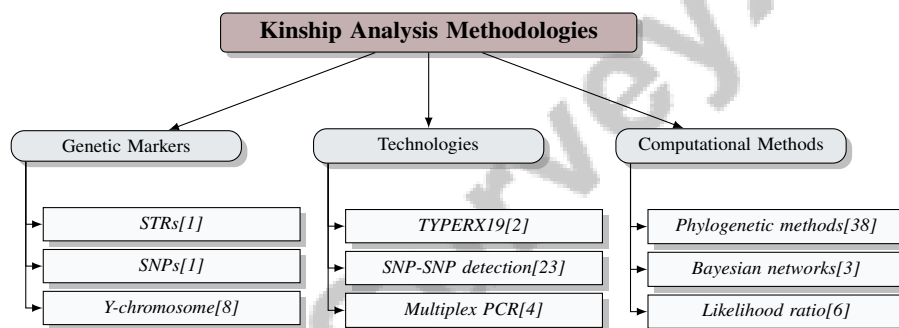
Figure 5: This figure illustrates the hierarchical structure of methodologies and technologies utilized in kinship analysis, categorizing them into genetic markers, technologies, and computational methods. These components are foundational in enhancing the precision and robustness of forensic genetics.

4.2 Case Studies and Practical Applications

Kinship analysis's practical applications in forensic genetics are demonstrated through case studies showcasing methodologies' effectiveness in real-world scenarios. One study analyzed 226 saliva samples for age prediction, highlighting the method's high reproducibility with minimal DNA amounts [46]. This underscores forensic genetics' utility in reliable age predictions, crucial for unidentified remains or age disputes.

In another approach, the R package KINSIMU implemented a unified formula for large sample simulations in kinship analysis, outperforming traditional list-based methods in managing extensive genetic data [28]. KINSIMU exemplifies advancements in computational tools that enhance kinship analysis accuracy and scalability, providing robust solutions for complex familial relationship determinations.

These case studies illustrate kinship analysis's diverse applications in forensic contexts, from age estimation to resolving complex family relationships. Integrating advanced genetic markers, such as SNPs, and sophisticated computational methodologies, including probabilistic genotyping and Bayesian networks, significantly enhances forensic genetic analyses' accuracy and dependability. This evolution strengthens DNA findings' evidential weight in legal contexts, aiding forensic biologists in interpreting complex mixtures and low-quality samples, solidifying their role in criminal investigations and judicial proceedings [10, 30, 33, 11, 37].

4.3 Challenges in Kinship Analysis

Kinship analysis in forensic genetics faces challenges affecting result accuracy and reliability. A primary concern is dependency on DNA quality, as skewed mixtures complicate deconvolution [47]. Degraded samples or low DNA input can lead to inaccuracies, necessitating robust methodologies to address these limitations [48].

Genetic mutations, like uniparental disomy (UPD), pose challenges by potentially causing false paternity exclusions, complicating kinship data interpretation [49]. These mutations require careful analysis consideration to avoid erroneous conclusions and maintain result integrity.

While the ULR method is useful, it may not fully account for kinship analysis complexities, particularly in scenarios involving multiple individuals or linked genetic markers [28]. This highlights the need for sophisticated analytical approaches accommodating genetic relationships' intricacies.

Integrating new technologies into standard forensic practices remains a challenge, with comprehensive validation of emerging methodologies needed to ensure reliability in forensic applications [5]. Addressing these challenges through advanced techniques and rigorous validation processes is crucial for enhancing kinship analysis robustness in forensic genetics.

5 Genetic Genealogy and Ancestry Tracing

5.1 Role of Genetic Genealogy in Law Enforcement

Genetic genealogy has transformed law enforcement by facilitating the resolution of cold cases and suspect identification through genetic data analysis. This method uses direct-to-consumer (DTC) genetic databases, popular among genealogy enthusiasts, to link crime-scene DNA to potential suspects via relatives' genetic profiles [50]. This integration has significantly enhanced investigative capabilities, enabling the resolution of cases previously constrained by insufficient traditional forensic evidence.

A framework categorizing genetic genealogy's application in law enforcement highlights its investigative benefits and the ethical complexities it introduces [51]. Privacy concerns and informed consent are critical issues, as using genetic data from DTC databases may involve individuals who have not explicitly consented to their information being used in criminal investigations. This raises ethical questions about balancing public safety and individual privacy rights.

As illustrated in Figure 6, the hierarchical categorization of genetic genealogy's role in law enforcement emphasizes not only its applications but also the ethical concerns and technological advancements that accompany its use. High-profile cold cases have demonstrated genetic genealogy's successful application, supported by advancements in forensic genetics, including probabilistic models and activity level propositions [10, 11]. However, these successes underscore the need for clear guidelines and regulatory frameworks to govern genetic data use in forensic contexts. Such frameworks will ensure forensic biologists effectively communicate findings, aiding the judiciary in making informed decisions based on robust evidence assessments, particularly as analytical techniques become more sophisticated. Addressing these ethical considerations is essential as law enforcement increasingly adopts genetic genealogy, ensuring individual rights protection while maximizing genetic data's potential for public safety.

5.2 Integration with Direct-to-Consumer Genetic Databases

Integrating direct-to-consumer (DTC) genetic databases into forensic investigations presents innovative opportunities alongside significant challenges. Initially intended for personal ancestry exploration and health insights, these databases have become crucial resources for law enforcement in solving cold cases and identifying unknown suspects through genetic genealogy. The notable use of public genealogy databases in forensic contexts, exemplified by the Golden State Killer case, raises ethical and legal concerns regarding informed consent and individual privacy [50, 11, 12].

A primary challenge is the lack of informed consent from individuals whose genetic data is included in these databases, particularly minors and relatives who may not have agreed to law enforcement access [51]. This situation raises ethical questions about privacy and autonomy, as individuals may inadvertently expose their relatives to genetic scrutiny.
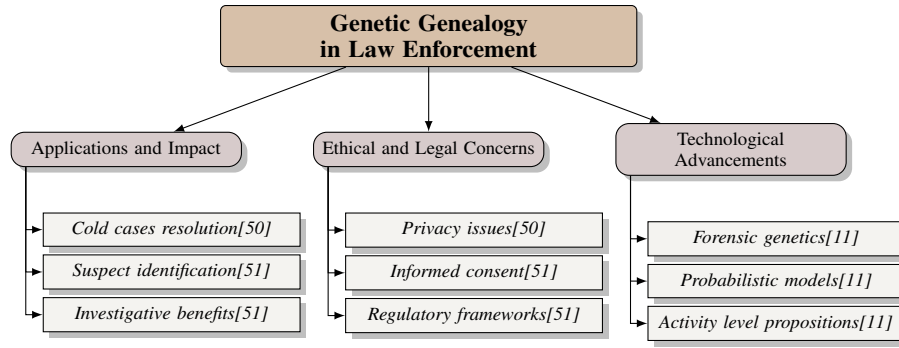
10

Figure 6: This figure illustrates the hierarchical categorization of genetic genealogy's role in law enforcement, highlighting applications, ethical concerns, and technological advancements.

Moreover, the increasing efficiency of methods used to extract genetic information from these databases poses risks, as adversaries with minimal technical expertise could access sensitive data [52]. Such vulnerabilities necessitate stringent data protection measures.

Genetic data in these databases is often represented as single nucleotide polymorphisms (SNPs), essential for algorithms identifying familial connections [53]. While these algorithms enhance forensic investigations, reliance on SNP data calls for careful consideration of genetic information's accuracy and reliability in legal contexts.

As DTC genetic databases' integration into forensic practices progresses, establishing comprehensive guidelines and regulatory frameworks becomes imperative. These frameworks should address complex ethical and legal implications concerning privacy, consent, and genetic data usage's societal impact, particularly given forensic and medical genetics convergence. Protecting individual rights while leveraging genetic data's investigative potential is crucial for balancing public safety with privacy concerns [51, 11, 12].

5.3 Implications for Ancestry Tracing

Genetic genealogy's application in ancestry tracing has significantly enhanced understanding of human history and lineage, providing insights into individual ancestry and population dynamics. Utilizing DTC genetic databases, individuals can trace lineage and uncover familial connections, exploring ancestral roots. These databases, rich in genetic information, allow users to identify relatives and construct detailed family trees based on shared genetic markers [51].

Genetic genealogy's integration into ancestry tracing has also facilitated biogeographical origins identification, offering insights into ethnic backgrounds and migration patterns. This capability is crucial for populations with complex historical admixtures, where traditional genealogical methods may be inadequate. Using SNPs and other genetic markers establishes a robust framework for distinguishing between ancestral lineages and inferring population structures [31].

Beyond individual exploration, aggregating genetic data from diverse populations enhances understanding of human evolution and migration, yielding insights into genetic diversity and history worldwide. Collective genetic knowledge, derived from advanced text mining methods like BioBERT-GRU and comprehensive genome-wide association studies (GWAS), improves understanding of human health, evolutionary adaptation, and complex traits' genetic foundations, particularly by elucidating SNPs' relationships with various phenotypic traits. High data sharing rates in ancient human DNA studies further emphasize transparency's importance in research, facilitating disease-associated SNPs discovery and enhancing genetic analyses' rigor [1, 31, 18, 45].

However, using genetic genealogy for ancestry tracing raises significant ethical concerns regarding privacy and data security. Sharing genetic information with DTC companies, combined with unauthorized access risk, necessitates comprehensive regulatory frameworks. With over 26 million individuals enrolled in DTC genetic testing services, allowing users to upload genetic data for genealogical purposes, vulnerability to adversarial attacks compromising sensitive information increases. For instance, adversaries can exploit genealogical databases like GEDmatch to reconstruct substantial

11

portions of individuals' genomes through strategic data uploads, potentially leading to unauthorized genetic relatives' identification and genetic data misuse. This evolving landscape underscores the urgent need for policies protecting individual privacy and maintaining public trust in genetic technologies, particularly in light of forensic genetics advances and medical and forensic applications intersection [52, 12]. Safeguarding individuals' genetic information while allowing genetic research's continued advancement is essential for maintaining public trust in genetic genealogy and ancestry tracing applications.

6   Challenges and Ethical Considerations

The intersection of scientific innovation and ethical responsibility in forensic genetics presents a complex landscape requiring comprehensive examination. The evolving field necessitates addressing challenges and ethical considerations related to privacy and data security, fundamental to ethical practices in this domain.

6.1   Privacy Concerns and Data Security

Privacy and data security are critical in forensic genetics due to the sensitive nature of genetic information and misuse risks. The use of genetic genealogy databases in forensic investigations amplifies concerns about unauthorized access and potential DNA profile exploitation [23]. Effective data protection strategies must address privacy challenges, such as data, query, and output privacy, which current methods often inadequately satisfy [54].

As illustrated in Figure 7, the critical privacy and data security concerns in forensic genetics can be categorized into three main areas: privacy challenges, security measures, and ethical considerations, each supported by relevant literature. Services like GEDmatch are vulnerable to API attacks, necessitating rigorous security evaluations to protect sensitive genetic data [53]. Ensuring data integrity requires stringent security measures. The reliability of forensic analyses depends on high-quality DNA inputs, correlating improved outcomes with meticulous sample preparation and evaluation of massively parallel sequencing (MPS) technologies [21]. However, challenges may arise in regions with varying allele frequencies, like East Asia [4].

Addressing privacy and security issues in forensic genetics demands a multifaceted approach, including advanced encryption, clear consent guidelines, and robust regulatory frameworks. Ethical considerations, such as informed consent, are crucial for maintaining public trust and balancing forensic genetics' investigative potential with personal privacy rights [23].
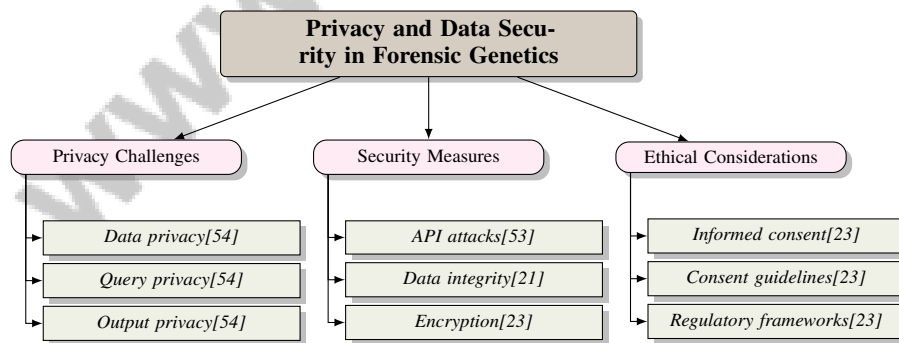

Figure 7: This figure illustrates the critical privacy and data security concerns in forensic genetics, categorizing them into privacy challenges, security measures, and ethical considerations, each supported by relevant literature.

6.2   Regulatory Frameworks and Guidelines

Regulatory frameworks and guidelines are essential for ensuring ethical compliance and safeguarding individual rights in genetic data usage. These frameworks balance forensic genetics' benefits in legal investigations with personal privacy protection and misuse prevention [15]. The convergence

of forensic and medical genetics highlights the need for protective measures to prevent genetic data exploitation beyond forensic applications [12].

Recent advancements include cryptographic frameworks enhancing genetic data security, such as systems encrypting DNA-based fingerprints and other biometric data, ensuring sensitive information disclosure only upon confirmed matches [55]. Direct-to-consumer (DTC) genetic services present unique privacy challenges, necessitating specific countermeasures against breaches [52].

Public discourse on genetic privacy is evolving, emphasizing informed discussions weighing genetic genealogy's benefits and risks [50]. Such dialogue is vital for developing comprehensive regulatory frameworks addressing genetic data use's ethical implications in forensic investigations.

Enforcing comprehensive regulatory frameworks is crucial for ensuring forensic genetics' integrity and reliability while safeguarding public trust amid emerging privacy, consent, and convergence challenges [11, 12]. These regulations support responsible genetic data use, advancing forensic science and protecting individual rights.

6.3 Potential for Misuse and Discriminatory Practices

Forensic genetics' application carries significant risks of genetic data misuse and discriminatory practices, given the sensitive nature of the information and its interpretative complexities. Forensic scientists must formulate unbiased propositions and accurately assess evidence within case contexts, especially with advancements like probabilistic genotyping and public genealogy databases, raising privacy and consent concerns [3, 10, 11, 12].

Genetic data misuse risks compromising privacy and leading to discrimination, particularly when models reliant on mutation rates and population dynamics assumptions are applied across diverse populations. Challenges in analyzing low-quality DNA samples, high genotyping error rates, and frequent locus drop-outs exacerbate misuse potential, leading to erroneous interpretations, especially when methodologies do not account for high error rates [20].

Statistical attacks on privacy-preserving methods due to inherent leakages underscore the need for rigorous security evaluations and robust privacy measures [53]. Economic costs, data misuse potential, and ethical dilemmas associated with data sharing complicate the forensic genetics landscape, raising conflicts of interest, especially concerning genetic discoveries' patenting.

Comprehensive ethical guidelines and robust regulatory frameworks are crucial to mitigate misuse and discrimination risks in forensic genetics, especially with emerging practices like genetic genealogy and trait prediction. Such frameworks are essential for safeguarding privacy and consent, ensuring forensic genetics' integration into the criminal justice system promotes justice and equity while maintaining public trust [11, 14, 12]. Addressing these concerns will enable valuable insights while safeguarding individual rights and promoting fairness across diverse populations.

7 Future Directions and Innovations

7.1 Advancements in DNA Profiling and Analysis

Recent advancements in DNA profiling have significantly enhanced the precision and efficiency of forensic analyses. Comprehensive Y-STR kits, incorporating rapidly mutating markers, improve resolution in tracing paternal lineages and complex familial cases [8]. These innovations address the demand for sensitive detection in degraded or low-quantity DNA scenarios, using primers with enhanced specificity and additional polymorphic markers [23]. X-chromosomal markers further enrich kinship analysis by providing insights into genetic relationships not easily resolved with autosomal markers [2].

Massively parallel sequencing (MPS) technologies are transforming DNA analysis, offering comprehensive genetic insights and enhancing forensic investigations [20]. The SMASH benchmarking toolkit, integrating synthetic, mouse, and human data, establishes robust metrics for sequencing accuracy and performance [36]. Future research should refine these benchmarks and explore diverse experimental conditions to strengthen MPS technologies in forensic applications [21].

Advanced computational methods, such as Mycorrhiza, demonstrate high accuracy in genotype assignment, surpassing traditional approaches, especially in high population differentiation contexts

[38]. Improvements in data processing, such as the use of compressed suffix arrays, enhance query execution time and manage large genomic databases efficiently [54, 35]. Analysis of informative SNPs, like rs1454361, underscores their utility in enhancing profiling accuracy [25]. Future research should expand SNP-STR marker sets and develop probabilistic evaluation approaches for imbalanced DNA mixtures [4].

As forensic genetics progresses, addressing security vulnerabilities in genetic genealogy services, such as the GEDmatch API, is crucial to safeguard against unauthorized access and potential misuse of genetic data [53]. Addressing these challenges will enable forensic genetics to provide valuable insights while maintaining genetic information integrity in legal contexts.

7.2 Integration of Advanced Sequencing Technologies

The integration of advanced sequencing technologies has revolutionized forensic genetics, enhancing resolution and precision in DNA analysis. Massively parallel sequencing (MPS), or next-generation sequencing (NGS), enables simultaneous sequencing of millions of DNA fragments, facilitating comprehensive analyses previously unattainable with traditional methods [20]. This technology expands forensic investigations by analyzing complex DNA mixtures and degraded samples.

MPS increases the discriminatory power of DNA profiling through SNPs and STRs analysis, with greater sensitivity and specificity for low-abundance allele detection in mixed DNA samples [21]. Targeted sequencing panels for forensic applications capture specific regions like ancestry informative markers, refining MPS utility while reducing costs [31].

Integrating MPS with advanced bioinformatics tools improves complex genetic data interpretation, allowing forensic scientists to draw accurate conclusions from DNA evidence. Machine learning and statistical models enhance the identification of genetic patterns, contributing to robust forensic analyses [38]. Implementing MPS requires careful validation and standardization to ensure reliability and reproducibility [5].

7.3 Innovations in Genetic Genealogy and Ancestry Analysis

Innovations in genetic genealogy and ancestry analysis are reshaping human lineage understanding, providing new methodologies for tracing ancestry and resolving genealogical queries. Advances focus on enhancing analysis accuracy through sophisticated models and computational tools. Bayesian models for genetic association analysis improve ancestry inference precision and genetic relationship identification across populations [56]. Robust models for complex DNA mixture analysis enhance genetic evidence reliability in forensic contexts [11].

Direct-to-consumer (DTC) genetic databases have transformed genetic genealogy, enabling extensive genetic information access and potential relative identification through shared markers. With over 26 million participants in services like 23andMe, these platforms facilitate familial connections but pose privacy challenges [52, 53]. Addressing these challenges is essential for responsible genetic data use.

Ongoing research should refine methodologies and explore new applications. Comprehensive genetic models, like the ForenSeq Kintelligence kit, combined with computational techniques like GemTools, will enhance ancestry analysis accuracy and reliability. This integration facilitates low-input and degraded sample analysis while providing insights into human history and lineage [48, 45, 57].

7.4 Emerging Computational and Analytical Tools

Emerging computational and analytical tools are revolutionizing forensic genetics by optimizing data analysis processes and enhancing forensic investigation effectiveness [15]. GenHap, a novel computational method for haplotype assembly, uses the Message Passing Interface (MPI) for parallel processing, improving reconstruction speed and accuracy [58].

Cryptographic techniques ensure genomic data confidentiality. Methods like PrivGenDB encode genomic data, enabling secure query processing [59]. Privacy-preserving algorithms, such as the selective hiding algorithm, offer near-optimal privacy-utility trade-offs [60]. Integrating these algorithms ensures sensitive information protection, even if servers are compromised [54].

14

The development of GemTools highlights potential for refining clustering algorithms and exploring applications in diverse datasets [57]. The FastID method demonstrates performance enhancements adaptability to various platforms [61].

The continuous evolution of computational tools underscores innovation's importance in enhancing genetic analyses' accuracy, efficiency, and security. Addressing genomic data management complexities and implementing strong privacy safeguards are crucial in advancing forensic science. They enhance genetic evidence reliability in legal proceedings while addressing privacy concerns raised by DNA databases, particularly in light of advances like genetic genealogy and trait prediction. Robust policy responses are necessary to protect privacy and maintain public trust in forensic and medical genetics [55, 53, 52, 12].

8   Conclusion

The survey of forensic genetics highlights its crucial role in contemporary legal systems, significantly improving the accuracy and reliability of legal investigations. Advanced DNA profiling techniques, such as short tandem repeats (STRs) and single nucleotide polymorphisms (SNPs), have markedly enhanced forensic analyses, facilitating the resolution of complex cases and more confident individual identifications. Beyond criminal investigations, forensic genetics is vital in disaster scenarios for victim identification, with ongoing technological advancements promising further improvements in accuracy and potential developments [6].

Additionally, the survey addresses the ethical challenges associated with genetic data use, particularly concerning privacy and data security. The necessity for a Genetic Data Protection Act is underscored to protect genetic information and maintain public trust in its application across forensic and medical contexts [12]. As forensic genetics evolves, establishing robust regulatory frameworks and guidelines will be essential for addressing these issues and ensuring the responsible use of genetic data.

Looking ahead, forensic genetics is set for further advancements, with new technologies and methodologies enhancing forensic investigations. The integration of advanced sequencing technologies and computational tools will continue to foster innovation, improving the efficiency and accuracy of genetic analyses. As these developments progress, forensic genetics will remain integral to modern legal systems, providing critical insights and supporting the pursuit of justice.

References

[1] Ruth M. Pfeiffer, Mitchell H. Gail, and David Pee. On combining data from genome-wide association studies to discover disease-associated snps, 2010.

[2] ZL Yu, Jing Sun, GS Ding, Xi Wang, Hong Liu, Xingchun Zhao, and Wanshui Li. The application of x-str: two case reports. *Journal of Pharmacy and Pharmacology*, 6(9):859–862, 2018.

[3] Peter J. Green and Julia Mortera. Sensitivity of inferences in forensic genetics to assumptions about founding genes, 2009.

[4] Tian Wei, Fei Liao, Yaowu Wang, Chao Pan, Chao Xiao, and Daixin Huang. A novel multiplex assay of snp-str markers for forensic purpose. *PLoS One*, 13(7):e0200700, 2018.

[5] Bruce R McCord, Quentin Gauthier, Sohee Cho, Meghan N Roig, Georgiana C Gibson-Daw, Brian Young, Fabiana Taglia, Sara C Zapico, Roberta Fogliatto Mariot, Steven B Lee, et al. Forensic dna analysis. *Analytical chemistry*, 91(1):673–688, 2018.

[6] Alejandra Real-Picado, Luis Díaz, and Cláudia Gomes. Relevance of genetic identification and kinship analysis in human and natural catastrophes—a review. *Genealogy*, 7(3):44, 2023.

[7] Roland AH Van Oorschot, Bianca Szkuta, Georgina E Meakin, Bas Kokshoorn, and Mariya Goray. Dna transfer in forensic science: a review. *Forensic Science International: Genetics*, 38:140–166, 2019.

[8] Manfred Kayser. Forensic use of y-chromosome dna: a general overview. *Human genetics*, 136(5):621–635, 2017.

[9] Iva Gomes, Nádia Pinto, Sofia Antão-Sousa, Verónica Gomes, Leonor Gusmão, and António Amorim. Twenty years later: a comprehensive review of the x chromosome use in forensic genetics. *Frontiers in Genetics*, 11:926, 2020.

[10] Duncan Taylor, Bas Kokshoorn, and Alex Biedermann. Evaluation of forensic genetics findings given activity level propositions: a review. *Forensic Science International: Genetics*, 36:34–49, 2018.

[11] Peter Gill, Tacha Hicks, John M Butler, Ed Connolly, Leonor Gusmão, Bas Kokshoorn, Niels Morling, Roland AH van Oorschot, Walther Parson, Mechthild Prinz, et al. Dna commission of the international society for forensic genetics: Assessing the value of forensic biological evidence-guidelines highlighting the importance of propositions: Part i: evaluation of dna profiling comparisons given (sub-) source propositions. *Forensic Science International: Genetics*, 36:189–202, 2018.

[12] Caitlin Curtis, James Hereward, Marie Mangelsdorf, Karen Hussey, and John Devereux. Protecting trust in medical genetics in the new era of forensics. *Genetics in Medicine*, 21(7):1483–1485, 2019.

[13] Peter De Knijff. From next generation sequencing to now generation sequencing in forensics. *Forensic Science International: Genetics*, 38:175–180, 2019.

[14] Helena Machado and Rafaela Granja. *Forensic genetics in the governance of crime*. Springer Nature, 2020.

[15] James Lindesay, Tshela E Mason, Luisel Ricks-Santi, William Hercules, Philip Kurian, and Georgia M Dunston. A new biophysical metric for interrogating the information content in human genome sequence variation: Proof of concept, 2011.

[16] Markus Scholz and Dirk Hasenclever. A canonical measure of allelic association, 2009.

[17] Jaya Lakshmi Bukyya, M L Avinash Tejasvi, Anulekha Avinash, HP Chanchala, Priyanka Talwade, Mohammed Malik Afroz, Archana Pokala, Praveen Kumar Neela, T K Shyamilee, and Vammi Srisha. Dna profiling in forensic science: a review. *Global medical genetics*, 8(04):135–143, 2021.

[18] Mohammad Dehghani, Behrouz Bokharaeian, and Zahra Yazdanparast. Biobert based snp-traits associations extraction from biomedical literature, 2023.

[19] Gregório Miguel Ferreira de Camargo. The role of molecular genetics in livestock production. *Animal Production Science*, 59(2):201–206, 2018.

[20] Elaine R Mardis. Dna sequencing technologies: 2006–2016. *Nature protocols*, 12(2):213–218, 2017.

[21] Vishakha Sharma, Hoi Yan Chow, Donald Siegel, and Elisa Wurmbach. Qualitative and quantitative assessment of illumina's forensic str and snp kits on miseq fgx™. *PLoS One*, 12(11):e0187932, 2017.

[22] MN Weedon, L Jackson, JW Harrison, KS Ruth, J Tyrrell, AT Hattersley, and CF Wright. Use of snp chips to detect rare pathogenic variants: retrospective, population based diagnostic evaluation. *bmj*, 372, 2021.

[23] Rangran Zhang, Yu Tan, Li Wang, Hui Jian, Jing Zhu, Yuanyuan Xiao, Mengyu Tan, Jiaming Xue, Fan Yang, and Weibo Liang. Set of 15 snp-snp markers for detection of unbalanced degraded dna mixtures and noninvasive prenatal paternity testing. *Frontiers in Genetics*, 12:800598, 2022.

[24] Zhanshan, Ma, Lianwei Li, and Ya-Ping Zhang. Individual-level snp diversity and similarity profiles, 2019.

[25] Sajad Habibi, Amirhossein Ahmadi, Mehrdad Behmanesh, Ali Miri, and Mahmood Tavallaie. Evaluation of ten snp markers for human identification and paternity analysis in persian population. *Iranian Journal of Biotechnology*, 17(3):e2148, 2019.

[26] Stephen D Turner, VP Nagraj, Matthew Scholz, Shakeel Jessa, Carlos Acevedo, Jianye Ge, August E Woerner, and Bruce Budowle. Evaluating the impact of dropout and genotyping error on snp-based kinship analysis with forensic samples. *Frontiers in Genetics*, 13:882268, 2022.

[27] Xingyi Yang, Xiaofang Zhang, Junyong Zhu, Linli Chen, Changhui Liu, Xingling Feng, Ling Chen, Huijun Wang, and Chao Liu. Genetic analysis of 19 x chromosome str loci for forensic purposes in four chinese ethnic groups. *Scientific reports*, 7(1):42782, 2017.

[28] Guanju Ma, Qian Wang, Bin Cong, and Shujin Li. An approach to unified formulae for likelihood ratio calculation in pairwise kinship analysis. *Frontiers in Genetics*, 15:1226228, 2024.

[29] Yuri Tani Utsunomiya, Rodrigo Vitorio Alonso, Adriana Santana do Carmo, Francine Campagnari, José Antonio Vinsintin, and José Fernando Garcia. mendelfix: a perl script for checking mendelian errors in high density snp data of trio designs, 2013.

[30] Peter J. Green and Julia Mortera. Paternity testing and other inference about relationships from dna mixtures, 2017.

[31] Hanbin Lee. A genealogical interpretation of fixed and random effects models of complex traits, 2023.

[32] Ion I. Mandoiu and Claudia Prajescu. High-throughput snp genotyping by sbe/sbh, 2005.

[33] Mikkel Meyer Andersen, Marie-Louise Kampmann, Alberte Honoré Jepsen, Niels Morling, Poul Svante Eriksen, Claus Børsting, and Jeppe Dyrberg Andersen. Shotgun dna sequencing for human identification: Dynamic snp selection and likelihood ratio calculations accounting for errors, 2024.

[34] Torben Tvedebrink, Poul Svante Eriksen, and Niels Morling. The multivariate dirichlet-multinomial distribution and its application in forensic genetics to adjust for sub-population effects using the $\theta - correction$, 2014.

[35] Travis Gagie. Suffix arrays for spaced-snp databases, 2014.

17

[36] Ameet Talwalkar, Jesse Liptrap, Julie Newcomb, Christopher Hartl, Jonathan Terhorst, Kristal Curtis, Ma'ayan Bresler, Yun S. Song, Michael I. Jordan, and David Patterson. Smash: A benchmarking toolkit for human genome variant calling, 2014.

[37] Robert George Cowell. A unifying framework for the modelling and analysis of str dna samples arising in forensic casework, 2018.

[38] Jeremy Georges-Filteau, Richard C. Hamelin, and Mathieu Blanchette. Mycorrhiza: Genotype assignment usingphylogenetic networks, 2020.

[39] Article.

[40] Chuan-Chao Wang, Ling-Xiang Wang, Rukesh Shrestha, Shaoqing Wen, Manfei Zhang, Xinzhu Tong, Li Jin, and Hui Li. Convergence of y chromosome str haplotypes from different snp haplogroups compromises accuracy of haplogroup prediction, 2013.

[41] Hong-Bing Yao, Chuan-Chao Wang, Jiang Wang, Xiaolan Tao, Shao-Qing Wen, Qiajun Du, Qiongying Deng, Bingying Xu, Ying Huang, Hong-Dan Wang, Shujin Li, Bin Cong, Liying Ma, Li Jin, Johannes Krause, and Hui Li. Genetic structure of sino-tibetan populations revealed by forensic str loci, 2015.

[42] Sina Majidian, Mohammad Hossein Kahaei, and Dick de Ridder. Minimum error correction-based haplotype assembly: considerations for long read data, 2020.

[43] Seung chan Ahn, Gene Kim, and MyungHo Kim. A note on applications of support vector machine, 2001.

[44] Gene Kim and MyungHo Kim. Application of support vector machine to detect an association between a disease or trait and multiple snp variations, 2001.

[45] Paolo Anagnostou, Marco Capocasa, Nicola Milia, Emanuele Sanna, Daniela Luzi, and Giovanni Destro Bisol. When data sharing gets close to 100can teach the open science movement, 2014.

[46] Sae Rom Hong, Sang-Eun Jung, Eun Hee Lee, Kyoung-Jin Shin, Woo Ick Yang, and Hwan Young Lee. Dna methylation-based age prediction from saliva: High age predictability by combination of 7 cpg markers. *Forensic Science International: Genetics*, 29:118–125, 2017.

[47] Søren B. Vilsen, Torben Tvedebrink, and Poul Svante Eriksen. Dna mixture deconvolution using an evolutionary algorithm with multiple populations, hill-climbing, and guided mutation, 2020.

[48] Michelle A Peck, Alexander F Koeppel, Erin M Gorden, Jessica L Bouchet, Mary C Heaton, David A Russell, Carmen R Reedy, Christina M Neal, and Stephen D Turner. Internal validation of the forenseq kintelligence kit for application to forensic genetic genealogy. *Forensic Genomics*, 2(4):103–114, 2022.

[49] Andrzej Doniec, Wojciech Łuczak, Maria Wróbel, Miłosz Januła, Andrzej Ossowski, Paweł Grzmil, and Tomasz Kupiec. Confirmation of paternity despite three genetic incompatibilities at chromosome 2. *Genes*, 12(1):62, 2021.

[50] Teneille R Brown. Why we fear genetic informants: using genetic genealogy to catch serial killers. *The Columbia science and technology law review*, 21(1):114, 2019.

[51] Debbie Kennett. Using genetic genealogy databases in missing persons cases and to develop suspect leads in violent crimes. *Forensic science international*, 301:107–117, 2019.

[52] Michael D Edge and Graham Coop. Attacks on genetic privacy via uploads to genealogical databases. *Elife*, 9:e51810, 2020.

[53] Peter Ney, Luis Ceze, and Tadayoshi Kohno. Genotype extraction and false relative attacks: Security risks to third-party genetic genealogy services beyond identity inference. In *NDSS*, 2020.

[54] Mohammad Zahidul Hasan, Md Safiur Rahman Mahdi, and Noman Mohammed. Secure count query on encrypted genomic data, 2017.

[55] Taeho Jung, Ryan Karl, and Geoffrey H. Siwo. Genomic encryption of biometric information for privacy-preserving forensics, 2021.

[56] Xiaoquan Wen. Bayesian model comparison in genetic association analysis: Linear mixed modeling and snp set testing, 2015.

[57] Lambertus Klei, Brian P. Kent, Nadine Melhem, Bernie Devlin, and Kathryn Roeder. Gemtools: A fast and efficient approach to estimating genetic ancestry, 2011.

[58] Andrea Tangherloni, Simone Spolaor, Leonardo Rundo, Marco S. Nobile, Paolo Cazzaniga, Giancarlo Mauri, Pietro Liò, Ivan Merelli, and Daniela Besozzi. Genhap: A novel computational method based on genetic algorithms for haplotype assembly, 2018.

[59] Sara Jafarbeiki, Amin Sakzad, Shabnam Kasra Kermanshahi, Raj Gaire, Ron Steinfeld, Shangqi Lai, and Gad Abraham. Privgendb: Efficient and privacy-preserving query executions over encrypted snp-phenotype database, 2021.

[60] Nour Almadhoun Alserr, Gulce Kale, Onur Mutlu, Oznur Tastan, and Erman Ayday. Near-optimal privacy-utility tradeoff in genomic studies using selective snp hiding, 2021.

[61] Siddharth Samsi, Brian Helfer, Jeremy Kepner, Albert Reuther, and Darrell O. Ricke. A linear algebra approach to fast dna mixture analysis using gpus, 2017.

Disclaimer:

SurveyX is an AI-powered system designed to automate the generation of surveys. While it aims to produce high-quality, coherent, and comprehensive surveys with accurate citations, the final output is derived from the AI's synthesis of pre-processed materials, which may contain limitations or inaccuracies. As such, the generated content should not be used for academic publication or formal submissions and must be independently reviewed and verified. The developers of SurveyX do not assume responsibility for any errors or consequences arising from the use of the generated surveys.