International Journal of
Molecular Sciences

MDPI

Review

Single Nucleotide Polymorphisms in Distant Kinship Inference and Forensic Genetic Genealogy

Denisse Stephania Becerra-Loaiza [1,2,3], Nayeli González-Ortiz [3,4], Yolanda Puga-Carrillo [3,5], Joel Alberto Aguilar-Velázquez [6], Itzae Adonai Gutiérrez-Hurtado [7,*] and José Alonso Aguilar-Velázquez [2,3,*]

1 Estancia Posdoctoral Avalada por la Secretaría de Ciencias, Humanidades, Tecnología e Innovación (SECIHTI), Guadalajara 44540, Jalisco, Mexico
2 Departamento de Morfología, Centro Universitario de Ciencias de la Salud, Universidad de Guadalajara, Guadalajara 44340, Jalisco, Mexico
3 Instituto de Genética Humana "Dr. Enrique Corona Rivera", Departamento de Biología Molecular y Genómica, Centro Universitario de Ciencias de la Salud, Universidad de Guadalajara, Guadalajara 44340, Jalisco, Mexico
4 Doctorado en Genética Humana, Departamento de Biología Molecular y Genómica, Centro Universitario de Ciencias de la Salud, Universidad de Guadalajara, Guadalajara 44340, Jalisco, Mexico
5 Maestría en Genética Forense e Identificación Humana, Departamento de Morfología, Centro Universitario de Ciencias de la Salud, Universidad de Guadalajara, Guadalajara 44340, Jalisco, Mexico
6 Departamento de Matemáticas, Universidad Autónoma Metropolitana Unidad Iztapalapa, Mexico City 09340, Mexico
7 Departamento de Biología Molecular y Genómica, Centro Universitario de Ciencias de la Salud, Universidad de Guadalajara, Guadalajara 44340, Jalisco, Mexico
* Correspondence: itzae.gutierrez@academicos.udg.mx (I.A.G.-H.); josealonso.aguilarvelazquez@academicos.udg.mx (J.A.A.-V.)

Abstract

Forensic genetics is moving from locus-based DNA profiling toward genome-wide inference enabled by high-density single-nucleotide polymorphism (SNP) data. While short tandem repeats remain central to routine human identification, SNP-based technologies and massively parallel sequencing have expanded the analysis of distant kinship through detection of identity-by-descent (IBD) segments and shared autosomal DNA. This narrative review synthesizes the biological basis of SNP-based distant kinship inference, the statistical and computational frameworks used to model genomic relatedness, and the operational transition from relatedness detection to forensic genetic genealogy (FGG). It distinguishes genetic genealogy database matching from formal forensic kinship testing, targeted SNP panels, SNP capture, low-coverage sequencing, Bayesian and machine-learning approaches, and independent forensic confirmation. Applications in criminal investigations, unidentified human remains, historical identifications, and broader relationship-inference contexts are discussed. The review also examines limitations related to recombination, stochastic inheritance, marker density, genotype quality, degraded or mixed forensic samples, population structure, endogamy, database composition, and genealogical record availability. Ethical and regulatory issues involving consent, privacy, database governance, law-enforcement access, data retention, and non-consenting relatives are considered. Overall, SNP-based forensic genomics can generate powerful investigative leads, but its outputs must be interpreted within method-specific analytical and evidentiary boundaries.

Keywords: single-nucleotide polymorphism; forensic genetic genealogy; investigative genetic genealogy; distant kinship; identity by descent; genetic genealogy database matching; massively parallel sequencing

Check for updates

Academic Editor: Richard Jäger

Received: 10 July 2026
Revised: 6 August 2026
Accepted: 13 August 2026
Published: 18 August 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Since the early 1990s, forensic genetics has relied heavily on DNA polymorphisms—commonly referred to as molecular markers—for individual identification, kinship testing, and the resolution of criminal and humanitarian cases [1,2]. Among these markers, short tandem repeats (STRs) became the cornerstone of human identification (HID) and gold standard in forensic genetics [3,4]. Their widespread forensic adoption is largely attributable to their broad genomic distribution [5,6], predominance in non-coding regions [6], and high levels of polymorphism, with heterozygosity often exceeding 70%, thereby providing strong discriminatory power for individual identification [5–7].

Despite their success, STR markers present several intrinsic limitations. Polymerase Chain Reaction (PCR)-based STR typing is susceptible to stochastic artifacts, including stutter peaks, allelic imbalance, and allele drop-out, particularly in low-template or degraded DNA samples [4]. In addition, the relatively large amplicon sizes required for conventional STR assays can compromise analytical performance when DNA is highly fragmented. These limitations have motivated the development and forensic implementation of alternative genetic marker systems [4,8].

Among these alternatives, Single-Nucleotide Polymorphisms (SNPs) have emerged as a powerful complementary—and, in some contexts, alternative—class of markers. SNPs consist of single-base substitutions in the DNA sequence [4,8] and represent the most abundant form of human genetic variation, occurring approximately once every 1000 base pairs (bp) throughout the genome [9,10]. They are distributed across coding and non-coding regions of the autosomal, sex-chromosomal, and mitochondrial genomes (Figure 1a) [9,10].

In forensic genetics, SNPs are commonly classified according to their principal forensic utility in individual identification, lineage analysis, biogeographic-ancestry inference, and prediction of externally visible characteristics (Figure 1b) [10–13]. Identity-informative SNPs (iiSNPs) are primarily used for individual identification and differentiation and may be located on autosomes or sex chromosomes [10]. Lineage-informative SNPs (liSNPs), located in mitochondrial DNA (mtDNA) and the Y chromosome, are useful for evolutionary analyses and kinship investigations involving distant relationships, including missing-person identification across multiple generations [10,11]. Ancestry-informative SNPs (aiSNPs) enable inference of biogeographic ancestry [10,12]. Phenotype-informative SNPs (piSNPs), typically located in coding or regulatory regions, are used to predict externally visible characteristics such as eye, hair, and skin pigmentation, as well as selected morphological traits [3,10,13]. In addition, pharmacogenetic SNPs have emerged within molecular forensic pathology and are increasingly applied in post-mortem "molecular autopsies" to investigate genetically mediated causes of death [10].

Several characteristics make SNPs particularly attractive for forensic applications [10]. First, their short amplicon requirements (typically 50–150 bp) facilitate successful amplification of highly degraded DNA fragments [8]. Second, SNPs exhibit substantially lower mutation rates than STRs (~10^{-8} vs. ~10^{-3} per generation) [1,14], increasing marker stability in kinship analyses across multiple generations [9,15]. Third, SNP genotyping is highly compatible with massively parallel sequencing (MPS) technologies, enabling the simultaneous analysis of hundreds to millions of loci within a single assay [8,16,17]. Although individual SNPs are typically biallelic and therefore less informative per locus than STRs, their high genomic density enables the interrogation of large marker panels, substantially increasing cumulative statistical power [8,18]. This technological progression from targeted SNP typing to high-density genome-wide genotyping and sequencing-based forensic profiling is summarized in Figure 1c [19–24].

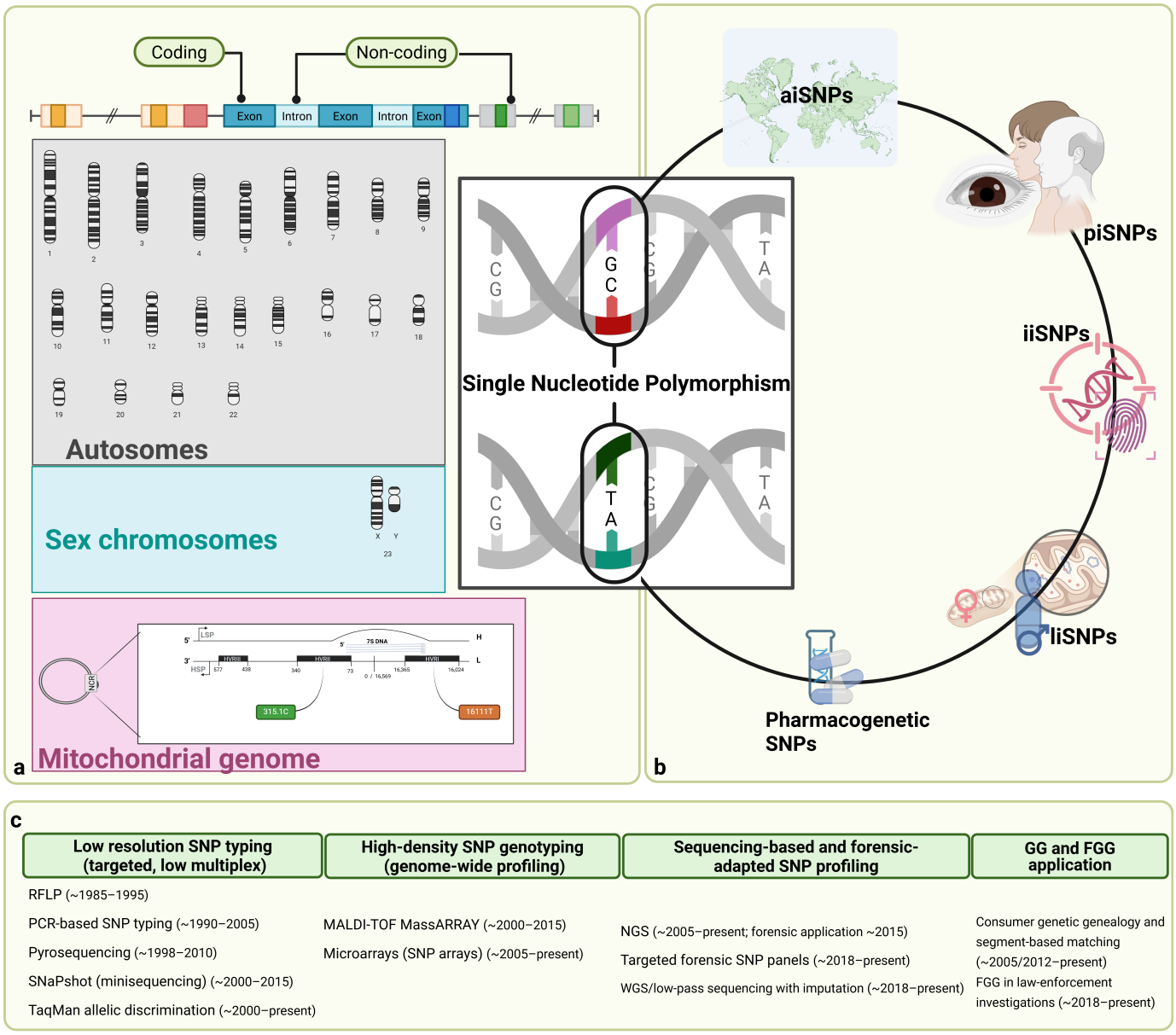


Figure 1. Overview of the genomic distribution, forensic classification, and technological evolution of SNPs. (**a**) SNPs occur throughout coding and non-coding regions of autosomal, sex-chromosomal, and mitochondrial DNA, supporting their use as genome-wide forensic markers. (**b**) According to forensic utility, SNPs can be classified as identity-informative (iiSNPs), ancestry-informative (aiSNPs), phenotype-informative (piSNPs), lineage-informative (liSNPs), and pharmacogenetic SNPs. (**c**) SNP-genotyping technologies have evolved from targeted assays to high-density genome-wide and sequencing-based approaches, enabling identity-by-descent segment detection, genetic genealogy database matching, and forensic genetic genealogy applications for distant kinship inference [19–24]. Created in BioRender. Becerra, D. (2026) https://BioRender.com/f7zbpn1.

Beyond their role in individual identification, genome-wide SNP data enable the detection of chromosomal segments shared identically by descent (IBD), providing a robust framework for kinship inference across multiple generations [25,26]. These developments have gained relevance with the emergence of forensic genetic genealogy (FGG), in which SNP profiles derived from unknown forensic samples are compared against large genetic genealogy databases to identify potential relatives and generate investigative leads [23,27,28]. Given the rapid methodological, computational, and ethical evolution of this field, a comprehensive synthesis of SNP-based approaches for distant kinship inference

is warranted [23,28]. This narrative review was based on peer-reviewed literature and selected authoritative institutional or policy sources addressing SNP-based distant kinship inference, IBD detection, genetic genealogy database matching, FGG, targeted forensic SNP panels, low-coverage sequencing, and forensic kinship likelihood-ratio frameworks. Priority was given to peer-reviewed studies, validation papers, case reports, and policy or regulatory documents directly relevant to forensic implementation. Accordingly, this review examines SNP-based distant kinship inference as a broader forensic-genomic framework and distinguishes among three related but non-equivalent analytical contexts: general statistical and computational methods for modeling genomic relatedness, segment-based matching approaches used in genetic genealogy databases, and forensic SNP-based kinship methods applied outside genetic genealogy. Unlike previous reviews focused primarily on FGG practice, case profiles, ethical implications, or SNP technologies, the added value of this review lies in integrating the biological basis of IBD sharing, SNP marker-density requirements, statistical and computational relationship-inference frameworks, and the operational transition to FGG workflows, while clarifying the conceptual boundaries among genetic genealogy database matching, formal forensic kinship testing, and other SNP-based relationship inference approaches. Within this framework, FGG is discussed as a major operational application in which genome-wide SNP matches generate investigative leads requiring genealogical reconstruction and independent forensic confirmation.

2. Terminology and Scope

Terminology in this field remains heterogeneous. The terms forensic genetic genealogy (FGG), investigative genetic genealogy (IGG), and forensic investigative genetic genealogy (FIGG) are commonly applied to highly overlapping practices that integrate genome-wide SNP analysis, genetic genealogy database matching, genealogical research, and investigative follow-up. In this review, FGG is adopted as the primary term, whereas IGG and FIGG are recognized as related designations used across different scientific, legal, and professional contexts. These terms should therefore be understood primarily as heterogeneous terminology rather than strictly distinct technical methodologies.

The main conceptual distinction is not between FGG, IGG, and FIGG themselves, but between SNP-based genetic genealogy lead generation, conventional STR-based familial searching, formal forensic kinship testing, and independent forensic confirmation. Traditional familial searching generally relies on STR profiles and is optimized for the detection of close relatives, whereas FGG is based on genome-wide SNP profiles, shared-segment detection, genetic genealogy database matching, and extended pedigree reconstruction [28–32]. Formal forensic kinship testing, in contrast, evaluates explicit relationship hypotheses using reference samples and statistical interpretation. Maintaining these boundaries avoids treating overlapping terminology as separate technical categories and helps distinguish investigative lead generation from evidentiary identification.

3. Genetic Basis of Distant Kinship Inference

Genome-wide SNP interrogation has transformed kinship inference by enabling the detection of chromosomal segments shared IBD. These segments are inherited from common ancestors and progressively fragmented through recombination across successive meioses. Consequently, the amount, number, and length of shared IBD segments provide a genomic framework for estimating biological relatedness across multiple degrees of separation, extending analytical capabilities beyond conventional STR-based kinship testing [1,14,25,26,28,31,33].

This segment-based framework differs fundamentally from traditional STR-based comparisons, which rely on a limited number of largely independent loci. In contrast,

dense SNP datasets reconstruct patterns of shared genomic segments across chromosomes while incorporating genetic linkage, recombination, marker density, and genotype quality [25,26,33–35]. Within this context, the distinction between identity-by-state (IBS) and IBD is critical: IBS refers to observed allele sharing, whereas IBD implies inheritance from a recent common ancestor and therefore requires interpretation within an appropriate population and genealogical framework [25,26,34].

The expected proportion of autosomal DNA shared between relatives decreases with each meiosis separating them from a common ancestor (Figure 2). Close relatives generally share larger genomic proportions in longer chromosomal segments, whereas distant relatives share progressively fewer and shorter IBD tracts [25,26,33,36]. However, realized genomic sharing may deviate substantially from theoretical expectations because recombination, chromosomal segregation, crossover interference, and sex-specific recombination patterns introduce variability in both the total amount of shared DNA and the distribution of IBD segment lengths [36,37]. Consequently, adjacent relationship categories may exhibit overlapping distributions of shared DNA, particularly as genealogical distance increases [33,38,39].

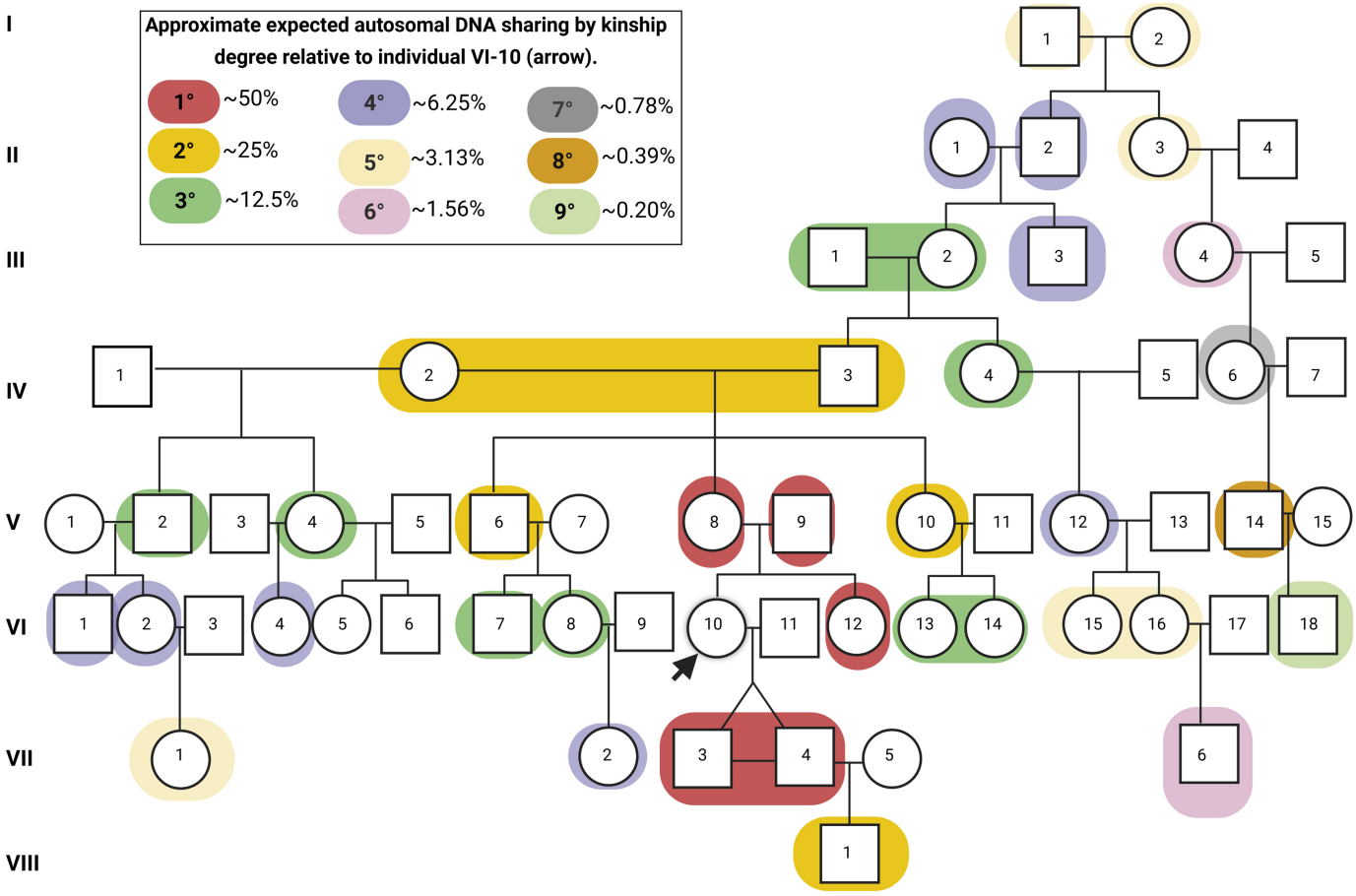


Figure 2. Expected autosomal DNA sharing across first- to ninth-degree relationships relative to the proband VI-10 (black arrow). The schematic illustrates the progressive decline in shared autosomal DNA and the fragmentation of inherited IBD segments with increasing genealogical distance. Roman numerals indicate generations. Individuals VII-3 and VII-4 are monozygotic twins; the horizontal line connecting their twin branches denotes monozygosity and does not represent a mating relationship. Created in BioRender. Becerra, D. (2026) https://BioRender.com/1olkp0u.

In genetic genealogy, shared autosomal DNA is commonly summarized in centiMorgans (cM), a unit of genetic distance reflecting recombination probability. Approximately

1 cM corresponds to a 1% probability of recombination between two loci during a single meiosis. Total shared cM, together with the number of shared segments and maximum segment length, provides an empirical indicator of relationship proximity. However, these metrics remain probabilistic rather than definitive, and shared-cM ranges may overlap substantially across relationship categories [25,26,33,38–42].

As genetic distance increases, a genealogical–genetic tradeoff becomes apparent: the number of potential genealogical relatives expands with each generation, whereas the probability of detecting informative autosomal IBD sharing progressively decreases [25,26,38]. Some distant relatives may retain detectable IBD segments, whereas others may share little or no detectable autosomal DNA depending on recombination history, marker density, genotyping quality, and analytical sensitivity [25,26,37,38,43,44]. This stochasticity represents a fundamental limitation of distant kinship inference and explains why genetic genealogy matches should be interpreted as probabilistic indicators of relatedness rather than direct evidence of a specific relationship category [28,29,31,38].

At a population-genomic scale, ancestry-specific IBD sharing and ancestry-tract-length distributions have also been used to infer the source and timing of historical admixture. Ioannidis et al. [45] applied these approaches to support pre-European Native American gene flow into eastern Polynesia. Although this differs from forensic pedigree analysis in inferential scale and purpose, it illustrates how recombination progressively shortens inherited tracts and thereby preserves temporal information about shared ancestry.

Marker density and data quality are therefore critical determinants of performance. Segment-based IBD detection for genetic genealogy database matching generally requires dense genome-wide SNP profiles comprising hundreds of thousands of markers to provide sufficient sensitivity for distant relationships (Figure 1c) [23,28,31,38,42,44,46]. Lower-density forensic SNP panels may support targeted kinship testing within defined analytical frameworks, particularly for closer or moderately distant relationships, but they should not be considered equivalent to high-density genetic genealogy database searches [24,38,43,47,48]. This distinction is especially relevant in forensic contexts, where evidentiary samples may be degraded, low-template, inhibited, contaminated, mixed, or incomplete, thereby reducing SNP recovery, increasing genotyping error, fragmenting true IBD segments, and decreasing the sensitivity of segment-based matching [24,28,38,44].

4. SNP-Based Approaches for Distant Kinship Inference

The biological principles underlying identity-by-descent (IBD) sharing described in the previous section provide the foundation for several analytical strategies used to infer distant genetic relationships. However, these approaches differ substantially in their objectives, statistical frameworks, marker-density requirements, and forensic applications. In forensic genetic genealogy (FGG), the initial analytical step typically consists of segment-based matching against genetic genealogy databases to identify potential biological relatives. In contrast, formal forensic kinship testing evaluates predefined relationship hypotheses using likelihood-based, pedigree-based, genotype-likelihood, or segment-based statistical models. Additional SNP-based methodologies—including targeted capture panels, forensic SNP panels, low-coverage sequencing strategies, Bayesian inference, and machine-learning approaches—also contribute to relationship inference but should not be regarded as interchangeable with routine genetic genealogy database matching. Distinguishing among these analytical frameworks is essential because they differ in their assumptions, statistical interpretation, forensic validation, and operational objectives [25,26,28,29,31,33,34,38].

4.1. Segment-Based Matching in Genetic Genealogy Databases

Genetic genealogy (GG) databases identify potential biological relatives through segment-based comparison of genome-wide autosomal SNP profiles. Publicly described matching frameworks compare pairs of individuals across the autosomes to detect long chromosomal segments that are likely to be identical by descent (IBD). Detected matches are typically summarized using total shared centiMorgans (cM), the number of shared segments, the length of the longest shared segment, and an estimated relationship range [25,26,29,31,33,38,40–42,46]. Importantly, these outputs should not be interpreted as forensic likelihood ratios (LRs). Instead, they represent genetic genealogy match indicators that support genealogical triangulation, documentary research, and pedigree reconstruction [28–31,49].

This analytical strategy differs fundamentally from formal forensic kinship testing. Rather than evaluating a predefined relationship hypothesis, GG database searches compare an unknown SNP profile against large collections of voluntarily contributed genetic profiles to identify potential relatives. Relationship estimates are generally derived from empirical shared-segment metrics together with platform-specific algorithms, many of which remain proprietary or only partially described. Consequently, a genetic genealogy match constitutes an investigative lead rather than a forensic identification and must subsequently be evaluated through genealogical reconstruction, demographic and historical information, investigative follow-up, and independent forensic confirmation [28–31,49].

As discussed in Section 3, the sensitivity of this approach depends strongly on marker density and genotype quality. Accordingly, the output generated by GG databases should be interpreted as evidence of potential biological relatedness rather than proof of identity. Its principal forensic value lies in identifying relatives who may facilitate reconstruction of ancestral lineages and guide subsequent investigative and forensic analyses.

4.2. Identity-by-Descent Sharing and Pedigree Expectations

The analytical approaches described in the following sections are founded on the biological principles of Mendelian inheritance and identity-by-descent (IBD) sharing. As discussed in Section 3, genomic segments inherited from a common ancestor are progressively fragmented by meiotic recombination, resulting in shorter and fewer detectable IBD segments as genealogical distance increases. Consequently, distant kinship inference relies on probabilistic rather than deterministic patterns of genomic sharing [25,26,29,31,33,36,38].

A fundamental concept underlying pedigree-based relationship inference is the coefficient of relationship (r), which represents the expected proportion of the genome shared IBD between two individuals through a given genealogical pathway. If k denotes the number of meioses separating two individuals along a single transmission path, the contribution of that path is given by:

$$r = \sum (1/2)^k$$

where the summation accounts for all independent inheritance pathways connecting the two individuals within a pedigree [25,26].

This formulation reflects the principle that each meiosis transmits approximately one-half of the parental genome to the offspring. Consequently, first-degree relatives are expected to share approximately 50% of their autosomal genome despite differing pedigree structures. For example, parent–offspring pairs are connected through a single meiotic transmission, whereas full siblings share two independent parental inheritance pathways that together produce the same expected coefficient of relationship [25,26].

Although the coefficient of relationship provides the theoretical expectation of genomic sharing, the realized amount of shared DNA varies among individuals because recombination occurs stochastically. As a result, close relatives generally share multiple long IBD

segments, whereas distant relatives tend to share fewer and shorter segments distributed across the genome. Moreover, the distributions of total shared DNA, segment number, and segment length increasingly overlap with genealogical distance, limiting the precision of relationship classification based solely on empirical sharing metrics [25,29,31,33,38–42].

Modern computational methods therefore estimate relatedness by integrating observed genomic sharing with probabilistic models of inheritance rather than relying exclusively on empirical shared-DNA thresholds. In genome-wide datasets, IBD detection is commonly performed using hidden Markov models (HMMs), which jointly model recombination patterns, allele frequencies, marker density, and genotyping uncertainty to infer the most likely distribution of IBD segments across the genome [25,29,33,38,44]. These probabilistic approaches improve the detection of short shared segments but remain sensitive to population structure, endogamy, and genotype quality, particularly when evaluating distant relationships [29,37,38,46].

Accordingly, the coefficient of relationship should be regarded as the theoretical foundation of pedigree-based kinship inference, whereas the observed genomic sharing detected in modern SNP datasets represents a stochastic realization of that expectation. This distinction underlies the statistical frameworks discussed in the following sections and explains why relationship inference requires explicit probabilistic modeling rather than simple comparison of shared DNA values.

4.3. Likelihood-Based Frameworks for SNP-Based Kinship Inference

Formal forensic kinship analysis differs fundamentally from genetic genealogy database matching because it evaluates predefined relationship hypotheses rather than performing exploratory searches across large collections of genetic profiles. Consequently, evidentiary interpretation relies on statistical models that quantify the probability of the observed genetic data under competing pedigree hypotheses rather than on empirical measures of shared DNA alone [25,26,29,31,33,34,38].

The principal statistical framework used in forensic kinship analysis is the likelihood ratio (LR), which compares the probability of observing the genetic evidence under two competing hypotheses. For example, one hypothesis may represent a specified biological relationship, whereas the alternative assumes that the individuals are unrelated. The LR is defined as:

$$LR = \frac{P(G \mid H_1)}{P(G \mid H_2)}$$

where G denotes the observed multilocus genotype data.

For genome-wide SNP datasets, likelihood calculations are considerably more complex than those used for conventional STR panels because dense SNP markers cannot generally be considered independent. Instead, analytical models must account for allele frequencies, linkage disequilibrium (LD), recombination, and genotyping uncertainty when estimating the probability of the observed data [25,26,33,34,38].

Several complementary analytical strategies have therefore been developed. Some methods reduce marker dependence through marker selection or LD pruning when approximate locus independence is assumed. Others explicitly incorporate chromosomal linkage and recombination into pedigree-based or segment-based models that evaluate the probability of observing the detected IBD sharing under alternative relationship hypotheses [25,26,33,34,38]. These approaches are particularly important for genome-wide datasets because recombination patterns contain information that contributes directly to relationship inference.

Genome-wide IBD detection is commonly implemented using hidden Markov models (HMMs), which estimate the most likely sequence of IBD and non-IBD states along the genome while accounting for recombination, allele frequencies, marker density, and

genotyping uncertainty [25,29,33]. These probabilistic models substantially improve the detection of shared chromosomal segments, particularly shorter IBD tracts that are informative for more distant relationships, although their performance remains influenced by marker density, genotype quality, and population characteristics [33,38,44].

Bayesian inference represents a complementary probabilistic framework in which genotype data are combined with prior information regarding pedigree structure or other model parameters to estimate the posterior probability of competing relationship hypotheses. Hierarchical Bayesian models may additionally incorporate uncertainty in allele frequencies, recombination rates, mutation processes, and population structure, providing a flexible framework for modeling complex genealogical scenarios [25,29,31,38,50,51]. Although these approaches are computationally demanding and remain less common in routine forensic casework than likelihood-based methods, they have become increasingly relevant for uncertainty modeling and methodological research in genome-wide kinship inference [29,38,51].

Accurate modeling of genotyping uncertainty is particularly important in forensic applications because degraded, low-template, inhibited, contaminated, or incomplete samples may accumulate small genotyping errors across thousands of SNPs, potentially influencing relationship inference [24,28,38,44]. Consequently, robust likelihood-based analyses require appropriate consideration of marker dependence, population structure, genotype quality, and analytical uncertainty when evaluating genome-wide evidence [29,36,38,52–54].

Unlike formal forensic kinship testing, genetic genealogy database searches do not typically generate likelihood ratios for predefined pedigree hypotheses. Instead, they identify potential relatives through shared-segment matching, after which formal statistical evaluation may be performed if suitable reference samples become available. Accordingly, likelihood-based analysis should be viewed as a complementary forensic framework that supports evidentiary evaluation following the generation of investigative leads through genetic genealogy [28,29,31,54].

4.4. Forensic SNP-Based Kinship Methods Beyond Genetic Genealogy Databases

Not all forensic SNP-based kinship inference relies on genetic genealogy (GG) databases. Several analytical approaches have been developed for forensic scenarios involving predefined reference samples, historical remains, missing-person investigations, low-coverage sequencing data, or direct candidate comparisons rather than open-ended database searches. Although these methodologies also exploit genome-wide SNP information, they operate within controlled forensic comparison frameworks and should therefore be distinguished from the investigative workflow of forensic genetic genealogy (FGG) [24,36,38,47,48,52,53].

One important application is targeted SNP capture combined with massively parallel sequencing (MPS), which enables the recovery of informative nuclear SNPs from highly degraded skeletal or historical remains. Gorden et al. [55], for example, demonstrated that targeted SNP capture can support extended kinship analysis even when DNA quality is insufficient for conventional genome-wide microarray genotyping. Such approaches have expanded the applicability of SNP-based kinship inference to forensic contexts in which highly fragmented DNA would otherwise preclude genome-wide analysis.

Similarly, low-coverage sequencing strategies have broadened the range of forensic samples suitable for relationship inference. Nguyen et al. [56] developed IBDGem, a computational framework designed to detect IBD sharing from low-coverage shotgun sequencing data, thereby facilitating kinship estimation when complete, high-quality genotype profiles cannot be obtained [56].

Targeted forensic SNP panels represent another important analytical strategy. Systems such as the ForenSeq™ Kintelligence assay, together with related computational approaches, provide optimized marker sets for forensic kinship analysis in defined casework scenarios, particularly for close and moderately distant relationships [24,43,47,48]. Although these

panels require substantially fewer markers than consumer genome-wide SNP arrays, they have been specifically designed to maximize kinship information from forensic samples while maintaining compatibility with degraded DNA [24,38,43,47,48]. Nevertheless, their marker density and analytical objectives differ from those of high-density genetic genealogy database searches, and they should therefore be regarded as complementary forensic tools rather than interchangeable substitutes for GG database matching.

Collectively, these approaches illustrate that SNP-based distant kinship inference extends well beyond forensic genetic genealogy. In addition to database-assisted investigations, genome-wide SNP analyses increasingly support historical identifications, missing-person investigations, targeted kinship testing, and other forensic applications in which relationship hypotheses are evaluated within controlled analytical frameworks rather than through exploratory genealogy database searches.

4.5. Emerging Computational Approaches

Recent advances in artificial intelligence have introduced machine-learning approaches as complementary tools for SNP-based kinship inference. Rather than explicitly modeling inheritance probabilities, these methods learn patterns of genomic similarity from training datasets and use multiple genomic features—including shared DNA, IBD segment characteristics, allele-sharing statistics, and population-genetic information—to predict relationship categories [57,58].

Algorithms such as random forests, gradient-boosting methods, and neural networks have shown promising performance for distinguishing distant relationships by capturing complex interactions among genomic variables that may not be fully represented by conventional statistical models [57,58]. These approaches may therefore improve relationship classification, particularly when genomic sharing falls within overlapping ranges characteristic of more distant relatives.

Despite these advances, machine-learning methods remain complementary to established forensic statistical frameworks. Unlike likelihood ratios, which provide formally interpretable measures of evidentiary strength, many machine-learning models function primarily as predictive classifiers whose outputs may be difficult to translate into legally interpretable evidence. Consequently, their current value lies in supporting relationship prioritization and methodological development rather than replacing formal forensic hypothesis testing or genetic genealogy database matching [59].

4.6. Analytical Limitations and Sources of Uncertainty

Despite the considerable advances enabled by genome-wide SNP analysis, distant kinship inference remains subject to several biological, statistical, and technical limitations that influence the interpretation of relatedness estimates. These limitations affect both formal forensic kinship testing and genetic genealogy database matching, although their impact differs according to the analytical framework employed [28,29,38,39,44].

One of the principal sources of uncertainty arises from the stochastic nature of meiotic recombination. As genealogical distance increases, the realized proportion of shared DNA varies substantially among relatives belonging to the same pedigree class, resulting in broad and overlapping distributions of shared genomic segments. Consequently, increasingly distant relationships become progressively more difficult to distinguish using shared DNA metrics alone [25,26,29,33,36–39].

Population history further complicates relationship inference. Population structure, founder effects, endogamy, and pedigree collapse may increase background levels of genomic sharing or create multiple inheritance pathways between individuals, producing greater IBD sharing than predicted by simplified pedigree models. In formal forensic analyses, these factors can be partially addressed through the use of population-specific allele

frequencies and appropriate statistical modeling. In contrast, genetic genealogy database matching relies primarily on genealogical interpretation to recognize such population-specific patterns and avoid overestimating biological relatedness [29,37,38,46,60].

Analytical performance is also strongly influenced by DNA quality and data completeness. Degraded or low-template forensic samples, low sequencing coverage, genotyping errors, and phasing inaccuracies may reduce the detection of true IBD segments or alter their estimated boundaries, thereby decreasing sensitivity for distant relationship inference [24,28,38,39,44,55,56,61,62]. Likewise, harmonization among different SNP platforms, genotype-calling strategies, and bioinformatic pipelines remains an important consideration when integrating data generated using different technologies [28,29,38,39,44].

Finally, genome-wide analyses inherently involve the accumulation of uncertainty across thousands or hundreds of thousands of markers. Small systematic biases associated with allele frequency estimation, recombination maps, genotype calling, imputation, or marker selection may collectively influence likelihood ratios, posterior probabilities, or relationship classification. Accordingly, relatedness estimates should always be interpreted within the assumptions, validation limits, and intended application of the analytical framework from which they were generated [28,29,38,39,44].

Recognizing these limitations is essential for the appropriate interpretation of SNP-based kinship inference. Whether applied to formal forensic comparisons or forensic genetic genealogy, genomic relatedness estimates should be regarded as probabilistic evidence that must be evaluated together with genealogical, investigative, and independent forensic information. This integrated perspective provides the foundation for the operational FGG workflow described in the following section.

5. From Distant Kinship Inference to Forensic Genetic Genealogy

Transitioning from SNP-based distant kinship inference to FGG requires translating genetic relatedness signals into an operational investigative workflow integrating genome-wide SNP profiling, genetic genealogy database searches, genealogical reconstruction, and independent forensic confirmation (Figure 3).

At the laboratory level, FGG generally begins with the same evidentiary material analyzed in conventional forensic DNA casework. Biological evidence is recovered, DNA is extracted and quantified, and STR profiling is typically performed first to search national or regional forensic databases such as the Combined DNA Index System (CODIS), or equivalent systems where applicable [28,54,63,64]. If no direct STR match or conventional investigative lead is obtained, genome-wide SNP profiling may subsequently be pursued using high-density SNP microarrays, whole-genome sequencing, or validated forensic SNP workflows, depending on sample quality, case requirements, and database compatibility [17,18,23,24,28,31,38,44]. The selected profiling workflow must satisfy the profile-completeness, marker-density, and database-compatibility requirements summarized in Section 3 and Table 1.

Once uploaded or compared within an authorized genetic genealogy environment, the SNP profile is evaluated against database participants who have consented—according to the platform's applicable policies—to forensic or law-enforcement matching. The resulting match list does not provide a final identification; instead, it generates a ranked set of potential biological relatives whose shared segments may indicate common ancestry [27–31,49]. The strength and utility of these matches depend on the amount of shared DNA, number and length of shared segments, presence of endogamy or background relatedness, representativeness of the database, and availability of documentary genealogical records [28,29,38,49].

Once candidate genetic matches are identified, the investigation transitions from molecular genetics to genealogical reconstruction. This stage integrates archival research, public and vital records, obituaries, census documents, demographic contextualization, historical

documentation, and pedigree expansion to identify shared ancestors and reconstruct extended family networks [28–30,49]. Key strategies include familial triangulation and cascade searching, in which multiple genetic matches are evaluated jointly to identify overlapping lineages, common ancestral couples, and descendancy pathways potentially connecting the unknown forensic profile to a candidate individual or family branch [28–31,49].

Within this framework, genealogical reconstruction functions primarily as a mechanism for hypothesis generation. Candidate identities are progressively refined using demographic and investigative filters, including age, sex, geographic location, migration history, temporal plausibility, family structure, and case-specific investigative information [28–30,49]. When necessary, targeted testing of additional relatives may further refine pedigree reconstruction or evaluate competing hypotheses, although such testing should be distinguished from the initial GG database match [28,29,49].

Final identification in FGG workflows requires independent forensic confirmation, commonly through STR profiling, direct comparison with a reference sample, or another validated forensic identification procedure [28,29,44]. Accordingly, the operational boundary between genetic genealogy lead generation and forensic confirmation should remain explicit in both casework practice and evidentiary interpretation. Because FGG involves voluntarily contributed genealogical data and may affect biological relatives who did not directly participate in testing, implementation also requires privacy safeguards, informed consent, database governance, law-enforcement access policies, and jurisdiction-specific oversight frameworks [29,30,49,54,63,65–72].

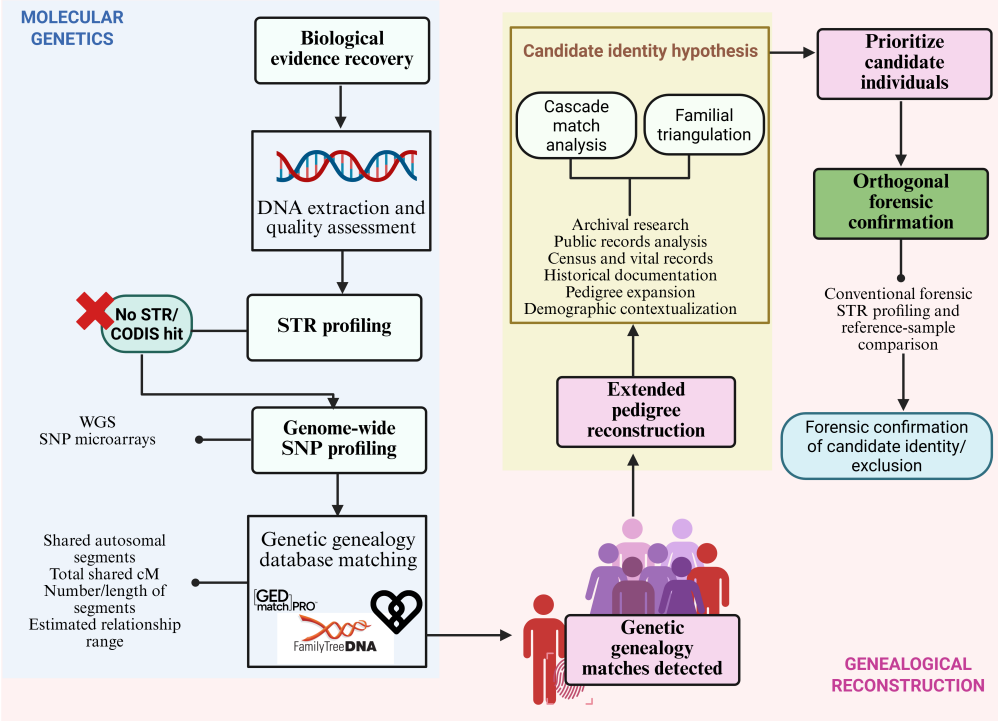


Figure 3. Operational workflow of FGG. The workflow begins with biological evidence recovery, DNA extraction, quality assessment, and conventional STR profiling. When no STR/CODIS hit or conventional investigative lead is obtained, genome-wide SNP profiling—using whole-genome sequencing [WGS], SNP microarrays, or validated forensic SNP workflow—may be used for authorized GG database matching. Genetic matches support genealogical reconstruction and candidate identity hypotheses, but final identification or exclusion requires independent forensic confirmation, typically through STR profiling and reference-sample comparison [27–31,44,46,49]. Created in BioRender. Becerra, D. (2026) https://BioRender.com/h8bchs5.

Table 1. Representative publicly reported FGG milestones in criminal investigations and unidentified human remains identification. Cases were included when publicly available sources reported the use of FGG involving genome-wide SNP-based genetic genealogy database searching for suspect identification or UHR resolution. This table is representative rather than exhaustive; for a full list of publicly reported cases, see Dowdeswell [32].

A. Criminal Investigations Resolved or Advanced Using FGG					
Suspect/Case	**Country**	**Incident Year**	**Resolution Year**	**Lead-Generating Entity, If Publicly Reported**	**Significance**
Joseph James De Angelo/Golden State Killer	USA	1974–1986	2018	Barbara Rae-Venter/law-enforcement collaboration	First widely recognized law-enforcement use of SNP-based FGG [23,30,31,73].
William Earl Talbott II/Cook–Van Cuylenborg murders	USA	1987	2018	Parabon NanoLabs/genetic genealogy collaboration	Early demonstration of FGG scalability in a cold-case homicide [30,32].
Raymond Charles Rowe/Christy Mirack murder	USA	1992	2018	Parabon NanoLabs/law enforcement	FGG lead followed by independent forensic confirmation [30,32].
Roy Charles Waller/NorCal Rapist	USA	1991–2006	2018	Law enforcement/genetic genealogy collaboration	Serial sexual-assault investigation advanced through FGG [32,74].
Brian Leigh Dripps/Angie Dodge murder	USA	1996	2019	Parabon NanoLabs/Idaho Falls Police Department	First reported FGG-enabled exoneration [30,32].
Daniel Nyqvist/Linköping double murder	Sweden	2004	2020	Swedish police	First confirmed European FGG case [75].

Table 1. *Cont.*

B. Identification of Unidentified Human Remains Using FGG					
Case Name	**Country**	**Found Year**	**Identification Year**	**Lead-Generating Entity, If Publicly Reported**	**Significance**
Bear Brook/Allenstown Four	USA	1985/2000	2017–2019	DNA Doe Project/Parabon involvement reported in public sources	Early complex UHR identification integrating genealogy and database matching [30,32,74].
Boy in the Box/Joseph Augustus Zarelli	USA	1957	2022	Not consistently specified in scientific literature	Long-unidentified child resolved through FGG [74,76].
Lady of the Dunes/Ruth Marie Terry	USA	1974	2022	Not consistently specified in scientific literature	Longstanding homicide victim identified through SNP-based genealogy [74,76].
Sandy Point Man/Christopher Luke Moore	Australia	2017	2023; case report published 2024	Victorian Institute of Forensic Medicine-led multidisciplinary collaboration; FGG work led by Runa Daniel according to the case report	Multidisciplinary Australian UHR identification integrating FGG and historical evidence [77].
Nogales John Doe/Donald Hadland Jr.	USA	2002	2023	Not consistently specified in scientific literature	Long-term UHR case resolved through FGG [74,76].
Rhinelander John Doe/Norman Grasser	USA	1980	2023	Not consistently specified in scientific literature	Missing-person identification through genealogical reconstruction [74,76].

FGG: forensic genetic genealogy; SNP: single nucleotide polymorphism; UHR: unidentified human remains.

6. Operational Applications of Forensic Genetic Genealogy

In 2018, the identification of the Golden State Killer suspect marked the first widely recognized law-enforcement application of FGG based on genome-wide SNP data and genetic genealogy database searching [23,30,31,73]. This milestone demonstrated that genetic genealogy matches, when integrated with documentary research and extended pedigree reconstruction, could generate investigative leads in cases where conventional STR database searches had failed to produce direct matches [28,30,31,49]. Since then, FGG has expanded across criminal investigations and unidentified human remains (UHR) casework and has consolidated as an emerging interdisciplinary forensic field, although its operational success remains strongly influenced by database accessibility, sample quality, SNP density, legal frameworks, and the availability of genealogical records [28,30–32,76,78].

Importantly, FGG should be distinguished from earlier STR-based familial searching approaches. Pre-FGG kinship-based strategies, including familial searches in forensic STR databases, demonstrated the investigative utility of indirect genetic matching but did not employ the genome-wide SNP-based genetic genealogy workflow that characterizes contemporary FGG [30,54,63,71]. Accordingly, pre-2018 STR familial-searching cases are discussed here as methodological precursors rather than operational FGG investigations and are therefore excluded from the revised FGG case table (Figure 4, Table 1). Table 1 lists representative milestone FGG cases; for a full list of publicly reported cases, see Dowdeswell [32].

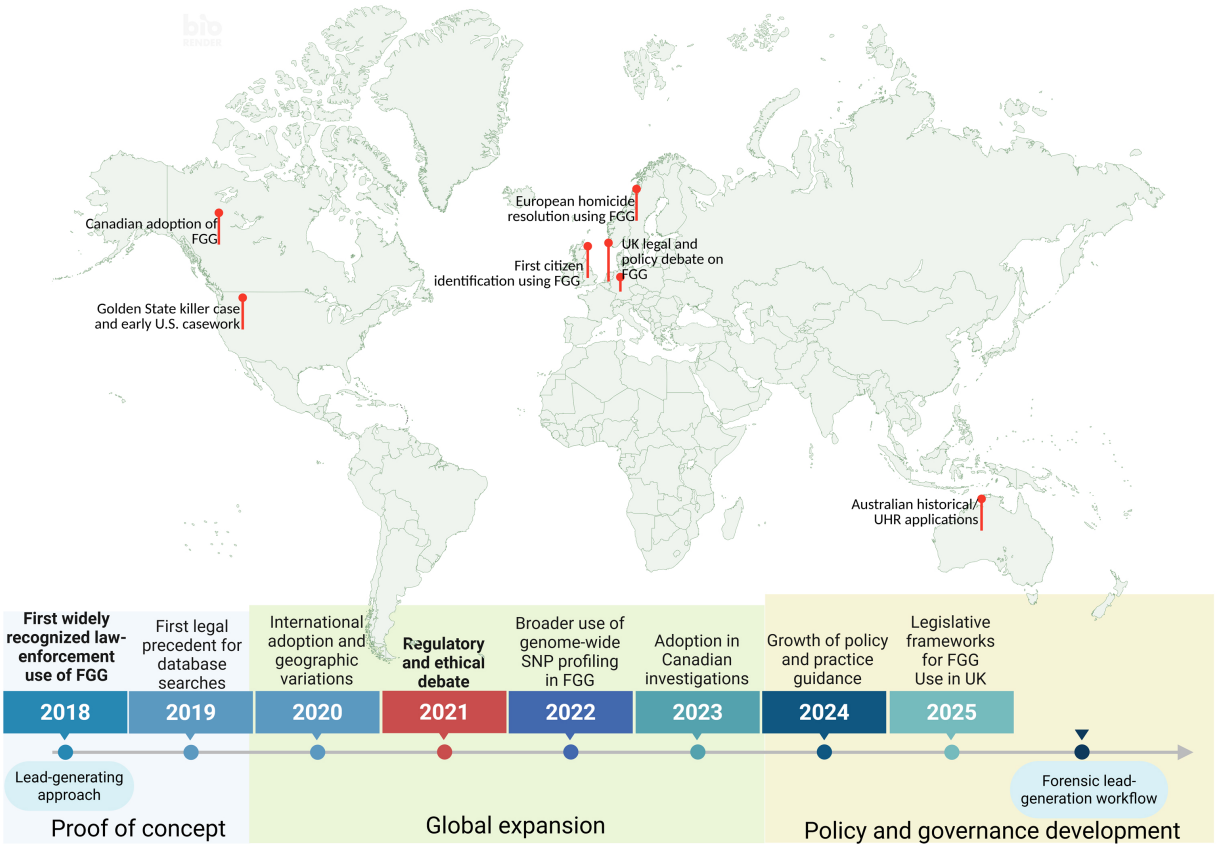


Figure 4. Selected milestones in the forensic implementation of genetic genealogy. The figure summarizes representative developments discussed in this review, including the 2018 Golden State Killer investigation, early legal and regulatory developments, international adoption of FGG, and the expansion of SNP-based workflows for criminal investigations and unidentified human remains casework. The timeline is representative rather than exhaustive [23,24,30–32,61,62,68–70,73,75,76,79–81]. Created in BioRender. Becerra, D. (2026) https://BioRender.com/b3gkggy.

6.1. Criminal Investigations

FGG has been used to generate investigative leads in both cold and active criminal investigations, particularly homicides and sexual assaults in which biological evidence is attributable to a single unknown contributor and conventional STR database searches have failed to identify a suspect [28,31,32,82,83]. In these investigations, a genome-wide SNP profile generated from crime-scene DNA is searched against genetic genealogy databases that permit forensic or law-enforcement use. The resulting genetic matches are interpreted as potential biological relatives, and genealogical reconstruction is subsequently used to identify common ancestors, reconstruct descent lines, and prioritize candidate individuals [28,29,31,49].

The Golden State Killer investigation served as the operational proof of concept for this approach, but subsequent investigations demonstrated that FGG could be applied beyond a single landmark case, including serial sexual assault investigations, cold-case homicides, and cross-jurisdictional investigations [30–32,73]. FGG may also contribute to the exclusion or correction of investigative assumptions when reconstructed pedigrees fail to support a candidate individual or when independent forensic confirmation excludes a proposed suspect [29,32]. For example, in the Angie Dodge homicide case, FGG was used to help identify crime-scene DNA that did not match the individual who had previously been convicted of the crime, resulting in the first FGG-based exoneration [30,32]. In criminal investigations, FGG functions as a lead-generation and hypothesis-development framework; final attribution requires independent forensic confirmation, typically through STR profiling or direct comparison with a reference sample [28,29,44].

6.2. Identification of Unidentified Human Remains

FGG has also emerged as an important tool in unidentified human remains (UHR) investigations when conventional missing-person database searches are negative. A major advantage of FGG in UHR casework is its ability to generate investigative leads even in the absence of direct family reference samples [61,62,76]. In these contexts, SNP profiles generated from skeletal or otherwise compromised remains may be searched against genetic genealogy databases to identify distant relatives, reconstruct family lineages, and generate identity hypotheses [61,62,76]. Greytak et al. [76] reviewed the laboratory, bioinformatic, and genealogical methodologies used in FGG for human remains identification and highlighted its growing role in UHR investigations.

Implementation in UHR investigations nevertheless remains highly dependent on DNA preservation and SNP profile completeness. Microarray-based FGG studies emphasize that sufficiently complete SNP profiles are necessary for reliable database searching, whereas compromised DNA may reduce SNP recovery, increase missing data rates, and decrease kinship-classification performance [61,62]. More recent forensic workflows using the ForenSeq™ Kintelligence system, targeted SNP panels, and capture-enrichment strategies further illustrate ongoing efforts to adapt SNP-based approaches to degraded or limited forensic material [24,62,79–81].

The Sandy Point case in Australia provides a useful example of why technically relevant and well-documented UHR investigations should be included in representative FGG milestone summaries. In this case, the absence of contextual information initially prevented the development of a viable identification hypothesis, and the eventual identification required integration of forensic anthropology, odontology, molecular biology, radiocarbon dating, historical investigation, and genealogical reconstruction [77]. The investigation ultimately resulted in the identification of Christopher Luke Moore and illustrates the multidisciplinary nature of UHR identification using FGG [77].

As in criminal investigations, genetic genealogy in UHR casework provides an investigative identity hypothesis that must subsequently be evaluated using historical records, missing-person information, biological profile consistency, DNA confirmation, and other independent lines of evidence [29,44,76]. This distinction remains central to forensic FGG practice: the database match guides the investigation, whereas final identification requires multidisciplinary evaluation and independent confirmation.

6.3. Broader Applications Related to SNP-Based Distant Kinship Inference

Although genetic genealogy originally emerged in non-forensic contexts such as ancestry exploration and unknown-parentage searches, FGG specifically refers to the forensic adaptation of these methodologies for law-enforcement and human-identification investigations [30,78]. Consequently, civil and administrative applications of genetic genealogy should be distinguished from operational FGG. Nevertheless, these adjacent applications remain relevant to the broader scope of this review because they illustrate how genome-wide SNP data, shared-segment analysis, and distant kinship inference can support biological relationship reconstruction when close relatives are unavailable.

Within this broader context, SNP-based distant kinship inference may contribute to unknown-parentage investigations, adoption-related searches, complex missing-person cases, historical identifications, and selected civil or administrative kinship questions. In contrast, inheritance disputes, citizenship claims, and military repatriation programs are more accurately characterized as kinship-comparison or human-identification contexts rather than FGG unless they specifically involve genetic genealogy database matching, genealogical reconstruction, and independent forensic confirmation [2,28,30,84]. This distinction is important because FGG primarily functions as an investigative lead-generation workflow, whereas formal kinship testing generally evaluates predefined relationship hypotheses using reference samples and statistical interpretation.

The analytical power of dense genome-wide SNP datasets enables increasingly distant biological relationships to be evaluated through identification of shared autosomal segments, although performance depends strongly on marker density, genotype quality, database composition, population structure, and the analytical framework employed (Figure 2, Table 2) [25,26,28,31,38,44]. Segment-based genetic genealogy searches generally require dense genome-wide datasets comprising hundreds of thousands of SNPs, whereas lower-density targeted forensic SNP panels may support defined kinship comparisons involving close or moderately distant relationships but should not be considered analytically equivalent to high-density genetic genealogy database matching [24,28,31,38,43,44,47,48]. These analytical distinctions are also relevant to immigration and refugee family-reunification proceedings, in which DNA-based relationship testing may be used to evaluate close biological relationships when documentary evidence is insufficient [85,86]. Although genome-wide SNP approaches could potentially support relationship assessment when only distant relatives are available, they should not be presented as an established routine replacement for validated STR-based testing. Any future or case-specific implementation should rely on validated decision thresholds, population-appropriate reference data, transparent reporting of uncertainty, and procedures that are understandable to applicants and legal representatives. Appropriate safeguards would also be required to minimize false inclusion or exclusion and to avoid conflating genetic kinship with legal, cultural, or social definitions of family.

Complementary marker systems such as microhaplotypes and compound markers may further strengthen forensic kinship analysis by increasing allelic diversity, improving performance with degraded DNA, and supporting interpretation in genetically diverse populations [17,87–89]. In addition, SNP-mixture deconvolution workflows are being devel-

oped to address mixed forensic samples, where multiple contributors may otherwise obscure kinship signals [90]. These approaches are best understood as complementary forensic-genomic developments rather than standard genetic genealogy database-matching methodologies.

Accordingly, SNP-based distant kinship inference extends beyond operational FGG. Its applications should be differentiated according to analytical objective, marker density, data quality, and interpretive framework. Genetic genealogy database matching, formal kinship testing, targeted SNP panels, low-coverage sequencing, SNP capture, microhaplotypes, and mixture-deconvolution workflows can all contribute to forensic relationship analysis, but they should not be treated as interchangeable methods or as providing equivalent evidentiary outputs. Maintaining these distinctions helps preserve the broader relevance of SNP-based kinship inference without overstating the scope of FGG itself.

Table 2. Comparison of SNP-based approaches for distant kinship inference and forensic relationship analysis.

Approach	Typical Analytical Context	Approximate Relationship Range	Key Observations
Conventional STR kits	Routine forensic identification and close-kinship testing	Primarily close relationships	STRs remain highly informative for individual identification and close kinship but have limited resolution for distant relationship inference because of the relatively small number of loci and higher mutation rates compared with SNPs [1,14,15].
100–200 independent SNPs	Parentage, sibship, and pedigree reconstruction under controlled conditions	Parent-offspring and full siblings; selected close relationships	Low-density independent SNP sets can support close relationship inference or pedigree reconstruction but are not designed for GG database matching [34,47,91,92].
~4000–10,000 targeted forensic SNPs	Targeted forensic kinship analysis and sparse SNP database searches	Up to third- or fourth-degree relationships under favorable conditions	Targeted panels such as ForenSeq™ Kintelligence (Verogen, Inc., San Diego, CA, USA) can support kinship estimation in defined forensic contexts, but their performance is not equivalent to dense genome-wide GG database matching [24,43,47,48].
Dense SNP arrays/Whole-Genome Sequencing	Segment-based genetic genealogy database matching and distant kinship inference	Useful for more distant relationship inference; performance varies by SNP density, DNA quality, database composition, endogamy, and genealogical record availability	High-density genome-wide profiles enable shared-segment detection, total shared cM estimation, and relationship-range inference in GG database searches. These outputs generate investigative leads rather than stand-alone forensic conclusions or formal identity determinations [23,25,28,31,38,44,93].
IBD coefficient- and segment-based statistical approaches	Formal distant relationship inference, simulation studies, and comparative evaluation of relationship-degree resolution	Variable; depends on relationship degree, marker density, recombination, and analytical model	IBD-based statistical approaches help evaluate the limits of relationship-degree classification and demonstrate why adjacent distant relationship classes may overlap. These approaches support interpretation of SNP-based distant kinship inference but should be distinguished from routine GG database matching [25,26,38,94,95].
SNP capture/low-coverage sequencing	Compromised remains, historical samples, low-input or low-coverage forensic material	Variable; depends on recovered SNPs, coverage, references, and analytical method	SNP capture and low-coverage sequencing methods can support identity or relatedness inference when microarray-quality profiles are unavailable, but they are best described as forensic SNP-based kinship approaches outside routine GG database matching [55,56].
Microhaplotypes, compound markers, and SNP-mixture workflows	Complementary forensic-genomic analysis, degraded DNA, genetically diverse populations, and mixture interpretation	Context-dependent	These approaches may increase informativeness or preserve kinship signal in difficult samples, but they should not be presented as standard FGG database-matching methods [17,87–90].

SNP: single-nucleotide polymorphism; STR: short tandem repeat; FGG: forensic genetic genealogy; GG: genetic genealogy; IBD: identity-by-descent; cM: centiMorgan. Relationship ranges are approximate and should not be interpreted as deterministic limits. Performance depends on marker density, DNA quality, population context, database composition, endogamy, recombination, genotyping error, and the analytical framework used.

7. Limitations and Challenges

A central limitation in the forensic implementation of SNP-based distant kinship inference is the difference between high-quality reference genotypes and SNP profiles generated from forensic evidence. Genetic genealogy databases are populated primarily with high-quality consumer or reference profiles, whereas forensic samples may be degraded, low-template, inhibited, contaminated, mixed, or otherwise incomplete. These conditions can reduce SNP recovery, increase missing data and genotyping error, fragment true IBD segments, and decrease the sensitivity of segment-based matching. Suitability for GG database searching should therefore be evaluated based on profile completeness, genotype error, platform compatibility, and the marker-density considerations summarized in Section 3 and Table 2 [24,28,38,44,55,56,61,62].

Beyond laboratory limitations, fundamental biological processes also constrain the resolution of distant kinship inference. As genealogical distance increases, the expected proportion of shared genetic material declines rapidly across generations. Consequently, the probability of detecting informative IBD segments between distant relatives decreases, and the genetic signal of relatedness may become difficult to distinguish from background genomic similarity within the population [26,43,96–98]. A related challenge arises from the stochastic nature of meiotic recombination and inheritance. The amount of genomic sharing between relatives may deviate substantially from theoretical expectations, producing overlapping distributions of shared cM values and IBD segment counts across adjacent relationship categories. Such overlap can complicate discrimination among distant relatives and introduce uncertainty when relationship hypotheses involve individuals separated by multiple generations [29,39,43,94,95].

Population structure, endogamy, founder effects, and pedigree collapse represent additional interpretive challenges, although their relevance differs according to the analytical framework employed. In formal kinship testing, population structure may be addressed through allele-frequency databases, population-specific parameters, kinship coefficients, or structured population models. In contrast, GG database matching generally does not correct match lists using equivalent formal population-genetic approaches; instead, genealogical interpretation is required to recognize patterns suggestive of endogamy, background relatedness, or population-specific overmatching [29,37,38,46,60,98]. Accordingly, population context should be considered primarily as an interpretive factor influencing genealogical reconstruction rather than as a parameter automatically incorporated into GG database matching.

The composition and representativeness of genealogical databases further influence the effectiveness of FGG investigations. Consumer genetic genealogy databases consist largely of voluntarily contributed profiles and remain unevenly distributed across global populations. GG databases are enriched for US-based participants with European ancestry, and representation is markedly limited for other countries and ethnic backgrounds [23,30,49,99]. This demographic imbalance may reduce match probabilities and genealogical resolution for underrepresented populations, potentially creating disparities in investigative success among different population groups.

Even when informative genetic matches are identified, successful interpretation requires careful integration of genetic, genealogical, and investigative information. Genetic matches rarely provide direct identification; instead, they function as probabilistic leads that must be contextualized through genealogical reconstruction and independent investigative evidence [29,49]. Incomplete historical records, pedigree inaccuracies, adoption, misattributed parentage, non-paternity events, migration, name changes, and incomplete public documentation may complicate genealogical reconstruction and introduce uncertainty into identity-hypothesis generation [29,49].

Ethical and legal considerations add a further layer of complexity to FGG implementation. Because genetic data are inherently relational, analysis of one individual's DNA may indirectly reveal information about biological relatives who have not themselves consented to genetic testing [30,49,100]. This relational nature of genetic information raises important questions regarding privacy, informed consent, database governance, law-enforcement access, data retention, secondary data use, and the balance between investigative utility and individual or familial privacy [29,30,49,54,63,65–72,100]. Governance approaches differ substantially across jurisdictions and database platforms. In the United States, implementation has been strongly shaped by platform-specific consent models, law-enforcement access policies, and investigative guidelines. In the United Kingdom and the European Union, proportionality, privacy protection, data minimization, and oversight are more closely linked to human-rights and data-protection frameworks. In Australia and Canada, emerging casework experience and policy discussions have focused on missing-person and UHR investigations, database access, transparency, and public acceptability. Across these contexts, opt-in versus opt-out models, database platform policies, secondary use, data retention, cross-border access, oversight, and privacy implications for non-consenting relatives remain central governance issues [29,30,54,63,65–72].

Collectively, these limitations demonstrate that although SNP-based distant kinship inference has substantially expanded the capabilities of forensic genetics, its performance remains constrained by sample quality, marker density, recombination, stochastic inheritance, population structure, database composition, genealogical record availability, and governance frameworks. Maintaining clear distinctions among high-density GG database matching, lower-density forensic SNP panels, formal kinship testing, and complementary SNP-based approaches is essential to avoid overgeneralizing the performance or evidentiary significance of any single method.

8. Conclusions

Genome-wide SNP analysis has expanded forensic genetics from locus-based human identification toward distant kinship inference based on shared genomic segments. Dense SNP datasets can support IBD segment detection, shared-cM estimation, and relationship inference across degrees of relatedness that often exceed the practical resolution of conventional STR-based approaches. However, the value of SNP data depends strongly on marker density, genotype quality, population context, database composition, recombination, and the analytical framework used.

The central message of this review is that SNP-based distant kinship inference, genetic genealogy database matching, formal forensic kinship testing, and independent forensic confirmation are related but non-interchangeable processes. GG database matching primarily generates investigative leads through shared-segment metrics and genealogical reconstruction, whereas formal forensic kinship testing evaluates predefined hypotheses using reference samples and statistical interpretation. Targeted forensic SNP panels, SNP capture, low-coverage sequencing, Bayesian models, machine-learning classifiers, microhaplotypes, and mixture-deconvolution workflows may complement this broader field, but they differ in validation status, assumptions, outputs, and evidentiary meaning.

The effectiveness of SNP-based distant kinship inference remains influenced by recombination, stochastic inheritance, population structure, endogamy, database composition, marker density, genotype quality, and the availability of genealogical records. These challenges are amplified in forensic casework, where evidentiary samples may be degraded, low-template, mixed, or incomplete, while genealogy databases are populated primarily with high-quality consumer or reference genotypes. Uneven database representation across populations may further influence match probability and investigative success.

Future progress will require transparent validation, responsible use of genome-wide SNP data, improved methods for compromised forensic samples, and clearer standards for database matching, statistical interpretation, and independent confirmation. Maintaining conceptual boundaries among investigative leads, relationship hypotheses, and formal identification is essential for scientifically robust, ethically responsible, and legally defensible implementation of SNP-based forensic genomics.

Author Contributions: D.S.B.-L. and J.A.A.-V. (José Alonso Aguilar-Velázquez) conceived and designed the review. D.S.B.-L., N.G.-O., Y.P.-C. and J.A.A.-V. (Joel Alberto Aguilar-Velázquez) conducted the literature search, curated and integrated the scientific evidence, and prepared the original draft of the manuscript. D.S.B.-L. and J.A.A.-V. (José Alonso Aguilar-Velázquez) contributed to the conceptual development of the review, the interpretation of forensic and genealogical applications, and the critical revision of the manuscript. I.A.G.-H. and J.A.A.-V. (José Alonso Aguilar-Velázquez) supervised the study, contributed to the scientific direction and overall organization of the review, and critically revised the manuscript for important intellectual content. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Mexican Government's Secretaría de Ciencia, Humanidades, Tecnología e Innovación (SECIHTI), Mexico, under grant CF-2023-G-536, awarded to J.A.A.-V.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Acknowledgments: The authors acknowledge SECIHTI for the 2025 postdoctoral fellowship awarded to D.S.B.-L. (CVU 702574).

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

aiSNPs	Ancestry-informative SNPs
bp	Base pairs
cM	centiMorgans
CODIS	Combined DNA Index System
FGG	Forensic genetic genealogy
FIGG	Forensic investigative genetic genealogy
HID	Human identification
HMMs	Hidden Markov Models
IBD	Identical-by-Descent
IBS	Identical-by-State
IGG	Investigative Genetic Genealogy
iiSNPs	Identity-informative SNPs
liSNPs	Lineage-informative SNPs
LR	Likelihood Ratio
MPS	Massively Parallel Sequencing
mtDNA	mitochondrial DNA
PCR	Polymerase Chain Reaction
piSNPs	Phenotype-informative SNPs
SNP	Single Nucleotide Polymorphism
STR	Short Tandem Repeats
UHR	Unidentified Human Remains
WGS	Whole-Genome Sequencing

References

1. Jobling, M.A.; Gill, P. Encoded evidence: DNA in forensic analysis. *Nat. Rev. Genet.* **2004**, *5*, 739–751. Correction in *Nat. Rev. Genet.* **2005**, *6*, 246. [CrossRef] [PubMed]
2. Gutiérrez-Hurtado, I.A.; García-Acéves, M.E.; Puga-Carrillo, Y.; Guardado-Estrada, M.; Becerra-Loaiza, D.S.; Carrillo-Rodríguez, V.D.; Plazola-Zamora, R.; Godínez-Rubí, J.M.; Rangel-Villalobos, H.; Aguilar-Velázquez, J.A. Past, present and future perspectives of forensic genetics. *Biomolecules* **2025**, *15*, 713. [CrossRef] [PubMed]
3. Kayser, M. Forensic DNA phenotyping: Predicting human appearance from crime scene material for investigative purposes. *Forensic Sci. Int. Genet.* **2015**, *18*, 33–48. [CrossRef] [PubMed]
4. Butler, J.M. *Advanced Topics in Forensic DNA Typing: Methodology*; Academic Press: San Diego, CA, USA, 2012. Available online: https://www.sciencedirect.com/book/monograph/9780123745132/advanced-topics-in-forensic-dna-typing-methodology (accessed on 17 March 2026).
5. Collins, J.R.; Stephens, R.M.; Gold, B.; Long, B.; Dean, M.; Burt, S.K. An exhaustive DNA micro-satellite map of the human genome using high performance computing. *Genomics* **2003**, *82*, 10–19. [CrossRef] [PubMed]
6. Ellegren, H. Microsatellites: Simple sequences with complex evolution. *Nat. Rev. Genet.* **2004**, *5*, 435–445. [CrossRef] [PubMed]
7. Panneerchelvam, S.; Norazmi, M.N. Forensic DNA profiling and database. *Malays. J. Med. Sci.* **2003**, *10*, 20–26. [PubMed]
8. Phillips, C. Application of autosomal SNPs and indels in forensic analysis. *Forensic Sci. Rev.* **2012**, *24*, 43–62. [PubMed]
9. Sobrino, B.; Brión, M.; Carracedo, A. SNPs in forensic genetics: A review on SNP typing methodologies. *Forensic Sci. Int.* **2005**, *154*, 181–194. [CrossRef] [PubMed]
10. Novroski, N.M.M.; Cihlar, J.C. Evolution of single-nucleotide polymorphism use in forensic genetics. *Wiley Interdiscip. Rev. Forensic Sci.* **2022**, *4*, e1459. [CrossRef]
11. Budowle, B.; van Daal, A. Forensically relevant SNP classes. *Biotechniques* **2008**, *44*, 603–610. [CrossRef] [PubMed]
12. Phillips, C. Forensic genetic analysis of bio-geographical ancestry. *Forensic Sci. Int. Genet.* **2015**, *18*, 49–65. [CrossRef] [PubMed]
13. Kayser, M.; Branicki, W.; Parson, W.; Phillips, C. Recent advances in forensic DNA phenotyping of appearance, ancestry and age. *Forensic Sci. Int. Genet.* **2023**, *65*, 102870. [CrossRef] [PubMed]
14. Kayser, M.; de Knijff, P. Improving human forensics through advances in genetics, genomics and molecular biology. *Nat. Rev. Genet.* **2011**, *12*, 179–192. Erratum in *Nat. Rev. Genet.* **2012**, *13*, 753. [CrossRef] [PubMed]
15. Butler, J.M.; Coble, M.D.; Vallone, P.M. STRs vs. SNPs: Thoughts on the future of forensic DNA testing. *Forensic Sci. Med. Pathol.* **2007**, *3*, 200–205. [CrossRef] [PubMed]
16. Seo, S.B.; King, J.L.; Warshauer, D.H.; Davis, C.P.; Ge, J.; Budowle, B. Single nucleotide polymorphism typing with massively parallel sequencing for human identification. *Int. J. Leg. Med.* **2013**, *127*, 1079–1086. [CrossRef] [PubMed]
17. de Knijff, P. From next generation sequencing to now generation sequencing in forensics. *Forensic Sci. Int. Genet.* **2019**, *38*, 175–180. [CrossRef] [PubMed]
18. Matute, S.P.; Iyavoo, S. Implementation of NGS and SNP microarrays in routine forensic practice: Opportunities and barriers. *BMC Genom.* **2025**, *26*, 541. [CrossRef] [PubMed]
19. Botstein, D.; White, R.L.; Skolnick, M.; Davis, R.W. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *Am. J. Hum. Genet.* **1980**, *32*, 314–331. [PubMed]
20. Syvänen, A.C. Accessing genetic variation: Genotyping single nucleotide polymorphisms. *Nat. Rev. Genet.* **2001**, *2*, 930–942. [CrossRef] [PubMed]
21. Shendure, J.; Ji, H. Next-generation DNA sequencing. *Nat. Biotechnol.* **2008**, *26*, 1135–1145. [CrossRef] [PubMed]
22. Metzker, M.L. Sequencing technologies—The next generation. *Nat. Rev. Genet.* **2010**, *11*, 31–46. [CrossRef] [PubMed]
23. Erlich, Y.; Shor, T.; Pe'er, I.; Carmi, S. Identity inference of genomic data using long-range familial searches. *Science* **2018**, *362*, 690–694. [CrossRef] [PubMed]
24. Antunes, J.; Walichiewicz, P.; Forouzmand, E.; Barta, R.; Didier, M.; Han, Y.; Perez, J.C.; Snedecor, J.; Zlatkov, C.; Padmabandu, G.; et al. Developmental validation of the ForenSeq® Kintelligence kit, MiSeq FGx® sequencing system and ForenSeq Universal Analysis Software. *Forensic Sci. Int. Genet.* **2024**, *71*, 103055. [CrossRef] [PubMed]
25. Browning, S.R.; Browning, B.L. Identity by descent between distant relatives: Detection and applications. *Annu. Rev. Genet.* **2012**, *46*, 617–633. [CrossRef] [PubMed]
26. Thompson, E.A. Identity by descent: Variation in meiosis, across genomes, and in populations. *Genetics* **2013**, *194*, 301–326. [CrossRef] [PubMed]
27. Ram, N.; Guerrini, C.J.; McGuire, A.L. Genealogy databases and the future of criminal investigation. *Science* **2018**, *360*, 1078–1079. [CrossRef] [PubMed]
28. Kling, D.; Phillips, C.; Kennett, D.; Tillmar, A. Investigative genetic genealogy: Current methods, knowledge and practice. *Forensic Sci. Int. Genet.* **2021**, *52*, 102474. [CrossRef] [PubMed]
29. Guerrini, C.J.; Wickenheiser, R.A.; Bettinger, B.; McGuire, A.L.; Fullerton, S.M. Four misconceptions about investigative genetic genealogy. *J. Law Biosci.* **2021**, *8*, lsab001. [CrossRef] [PubMed]

30. Katsanis, S.H. Pedigrees and perpetrators: Uses of DNA and genealogy in forensic investigations. *Annu. Rev. Genom. Hum. Genet.* **2020**, *21*, 535–564. [CrossRef] [PubMed]
31. Greytak, E.M.; Moore, C.; Armentrout, S.L. Genetic genealogy for cold case and active investigations. *Forensic Sci. Int.* **2019**, *299*, 103–113. [CrossRef] [PubMed]
32. Dowdeswell, T.L. Forensic genetic genealogy: A profile of cases solved. *Forensic Sci. Int. Genet.* **2022**, *58*, 102679. [CrossRef] [PubMed]
33. Kling, D.; Tillmar, A. Forensic genealogy-A comparison of methods to infer distant relationships based on dense SNP data. *Forensic Sci. Int. Genet.* **2019**, *42*, 113–124. [CrossRef] [PubMed]
34. Goudet, J.; Kay, T.; Weir, B.S. How to estimate kinship. *Mol. Ecol.* **2018**, *27*, 4121–4135. [CrossRef] [PubMed]
35. Ramstetter, M.D.; Dyer, T.D.; Lehman, D.M.; Curran, J.E.; Duggirala, R.; Blangero, J.; Mezey, J.G.; Williams, A.L. Benchmarking relatedness inference methods with genome-wide data from thousands of relatives. *Genetics* **2017**, *207*, 75–82. [CrossRef] [PubMed]
36. Qiao, Y.; Sannerud, J.G.; Basu-Roy, S.; Hayward, C.; Williams, A.L. Distinguishing pedigree relationships via multi-way identity by descent sharing and sex-specific genetic maps. *Am. J. Hum. Genet.* **2021**, *108*, 68–83. [CrossRef] [PubMed]
37. Caballero, M.; Seidman, D.N.; Qiao, Y.; Sannerud, J.; Dyer, T.D.; Lehman, D.M.; Curran, J.E.; Duggirala, R.; Blangero, J.; Carmi, S.; et al. Crossover interference and sex-specific genetic maps shape identical by descent sharing in close relatives. *PLoS Genet.* **2019**, *15*, e1007979. [CrossRef] [PubMed]
38. Ertürk, M.S.; Fitzpatrick, C.; Press, M.; Wein, L.M. Analysis of the genealogy process in forensic genetic genealogy. *J. Forensic Sci.* **2022**, *67*, 2218–2229. [CrossRef] [PubMed]
39. Tillmar, A.; Kling, D. Comparative study of statistical approaches and SNP panels to infer distant relationships in forensic genetics. *Genes* **2025**, *16*, 114. [CrossRef] [PubMed]
40. Bettinger, B.T. The Shared cM Project Version 4.0. The Genetic Genealogist. 2020. Available online: https://thegeneticgenealogist.com/wp-content/uploads/2020/03/Shared-cM-Project-Version-4.pdf (accessed on 5 May 2026).
41. Perl, J. DNA Painter Shared cM Project 4.0 Tool. DNA Painter. 2020. Available online: https://dnapainter.com/tools/sharedcmv4 (accessed on 5 May 2026).
42. AncestryDNA. AncestryDNA Matching White Paper. AncestryDNA. 2025. Available online: https://www.ancestrycdn.com/support/us/2025/01/2025matchingwhitepaper.pdf (accessed on 5 May 2026).
43. Daniel, R.; Raymond, J.; Sears, A.; Stock, A.; Scudder, N.; Padmabandu, G.; Kumar, S.; Snedecor, J.; Antunes, J.; Hartman, D. It's all relative: A multi-generational study using ForenSeq™ Kintelligence. *Forensic Sci. Int.* **2024**, *364*, 112208. [CrossRef] [PubMed]
44. Zang, Y.; Wu, E.; Li, T.; Liu, J.; Wu, R.; Li, R.; Sun, H. Evaluation of four forensic investigative genetic genealogy analysis approaches with decreased numbers of SNPs and increased genotyping errors. *Genes* **2024**, *15*, 1329. [CrossRef] [PubMed]
45. Ioannidis, A.G.; Blanco-Portillo, J.; Sandoval, K.; Hagelberg, E.; Miquel-Poblete, J.F.; Moreno-Mayar, J.V.; Rodríguez-Rodríguez, J.E.; Quinto-Cortés, C.D.; Auckland, K.; Parks, T.; et al. Native American gene flow into Polynesia predating Easter Island settlement. *Nature* **2020**, *583*, 572–577. [CrossRef] [PubMed]
46. Williams, C.M.; Scelza, B.A.; Slack, S.D.; Font-Porterias, N.; Al-Hindi, D.R.; A Mathias, R.; Watson, H.; Barnes, K.C.; Lange, E.; Johnson, R.K.; et al. A rapid accurate approach to inferring pedigrees in endogamous populations. *Genetics* **2025**, *230*, iyaf094. [CrossRef] [PubMed]
47. Huisman, J. Pedigree reconstruction from SNP data: Parentage assignment, sibship clustering and beyond. *Mol. Ecol. Resour.* **2017**, *17*, 1009–1024. [CrossRef] [PubMed]
48. Snedecor, J.; Fennell, T.; Stadick, S.; Homer, N.; Antunes, J.; Stephens, K.; Holt, C. Fast and accurate kinship estimation using sparse SNPs in relatively large database searches. *Forensic Sci. Int. Genet.* **2022**, *61*, 102769. [CrossRef] [PubMed]
49. Kennett, D. Using genetic genealogy databases in missing persons cases and to develop suspect leads in violent crimes. *Forensic Sci. Int.* **2019**, *301*, 107–117. [CrossRef] [PubMed]
50. van de Schoot, R.; Depaoli, S.; King, R.; Kramer, B.; Märtens, K.; Tadesse, M.G.; Vannucci, M.; Gelman, A.; Veen, D.; Willemsen, J.; et al. Bayesian statistics and modelling. *Nat. Rev. Methods Prim.* **2021**, *1*, 1. Correction in *Nat. Rev. Methods Primers* **2021**, *1*, 16. [CrossRef]
51. Nott, D.J.; Drovandi, C.; Frazier, D.T. Bayesian inference for misspecified generative models. *Annu. Rev. Stat. Appl.* **2024**, *11*, 179–202. [CrossRef]
52. Ge, J.; Budowle, B.; Cariaso, M.; Mittelman, K.; Mittelman, D. A likelihood ratio framework for inferring close kinship from dynamically selected SNPs. *Front. Genet.* **2025**, *16*, 1635734. [CrossRef] [PubMed]
53. Li, R.; Li, H.; Peng, D.; Hao, B.; Wang, Z.; Huang, E.; Wu, R.; Sun, H. Improved pairwise kinship analysis using massively parallel sequencing. *Forensic Sci. Int. Genet.* **2019**, *38*, 77–85. [CrossRef] [PubMed]
54. Ge, J.; Budowle, B. Forensic investigation approaches of searching relatives in DNA databases. *J. Forensic Sci.* **2021**, *66*, 430–443. [CrossRef] [PubMed]

55. Gorden, E.M.; Greytak, E.M.; Sturk-Andreaggi, K.; Cady, J.; McMahon, T.P.; Armentrout, S.; Marshall, C. Extended kinship analysis of historical remains using SNP capture. *Forensic Sci. Int. Genet.* **2022**, *57*, 102636. [CrossRef] [PubMed]
56. Nguyen, R.; Kapp, J.D.; Sacco, S.; Myers, S.P.; Green, R.E. A computational approach for positive genetic identification and relatedness detection from low-coverage shotgun sequencing data. *J. Hered.* **2023**, *114*, 504–512. [CrossRef] [PubMed]
57. Wang, M.; Chen, H.; Luo, L.; Huang, Y.; Duan, S.; Yuan, H.; Tang, R.; Liu, C.; He, G. Forensic investigative genetic genealogy: Expanding pedigree tracing and genetic inquiry in the genomic era. *J. Genet. Genom.* **2025**, *52*, 460–472. [CrossRef] [PubMed]
58. Cosenza, G.; Castellino, L.; Morelli, S.; Alladio, E.; Pilli, E. Kinship analysis and machine learning algorithms in forensic contexts: A new NGS panel. *Expert Syst. Appl.* **2025**, *266*, 126161. [CrossRef]
59. Marsico, F.; Amigo, M. Ethical and security challenges in AI for forensic genetics: From bias to adversarial attacks. *Forensic Sci. Int. Genet.* **2025**, *76*, 103225. [CrossRef] [PubMed]
60. Li, R.; Zang, Y.; Liu, J.; Wu, E.; Wu, R.; Sun, H. Inferring the degree of relatedness and kinship types using an all-in-one marker set. *Genes* **2025**, *16*, 455. [CrossRef] [PubMed]
61. Davawala, A.; Stock, A.; Spiden, M.; Daniel, R.; McBain, J.; Hartman, D. Forensic genetic genealogy using microarrays for the identification of human remains: The need for good quality samples-A pilot study. *Forensic Sci. Int.* **2022**, *334*, 111242. [CrossRef] [PubMed]
62. de Vries, J.H.; Kling, D.; Vidaki, A.; Arp, P.; Kalamara, V.; Verbiest, M.M.P.J.; Piniewska-Róg, D.; Parsons, T.J.; Uitterlinden, A.G.; Kayser, M. Impact of SNP microarray analysis of compromised DNA on kinship classification success in the context of investigative genetic genealogy. *Forensic Sci. Int. Genet.* **2022**, *56*, 102625. [CrossRef] [PubMed]
63. Gershaw, C.J.; Schweighardt, A.J.; Rourke, L.C.; Wallace, M.M. Forensic utilization of familial searches in DNA databases. *Forensic Sci. Int. Genet.* **2011**, *5*, 16–20. [CrossRef] [PubMed]
64. Aguilar-Velázquez, J.A. Perspectiva general de la genética y la genómica forense en México. *Rev. Digit. Cienc. Forense* **2023**, *2*, 48–91. [CrossRef]
65. Budowle, B.; Baker, L.; Sajantila, A.; Mittelman, K.; Mittelman, D. Prioritizing privacy and presentation of supportable hypothesis testing in forensic genetic genealogy investigations. *Biotechniques* **2024**, *76*, 425–431. [CrossRef] [PubMed]
66. Claerhout, S.; Noppe, H.; Cohn, B.; Borry, P. Opt-in or out? Public perspectives on forensic DNA kinship investigations within the Dutch-speaking community. *Heliyon* **2024**, *10*, e30074. [CrossRef] [PubMed]
67. Dahlquist, J.; Robinson, J.O.; Daoud, A.; Bash-Brooks, W.; McGuire, A.L.; Guerrini, C.J.; Fullerton, S.M. Public perspectives on investigative genetic genealogy: Findings from a national focus group study. *AJOB Empir. Bioeth.* **2024**, *15*, 280–290. [CrossRef] [PubMed]
68. Ram, N.; Roberts, J.L. Forensic genealogy and the power of defaults. *Nat. Biotechnol.* **2019**, *37*, 707–708. [CrossRef] [PubMed]
69. Syndercombe Court, D. Forensic genealogy: Some serious concerns. *Forensic Sci. Int. Genet.* **2018**, *36*, 203–204. [CrossRef] [PubMed]
70. Wickenheiser, R.A. Forensic genealogy, bioethics and the Golden State Killer case. *Forensic Sci. Int. Synerg.* **2019**, *1*, 114–125. [CrossRef] [PubMed]
71. Mateen, R.M.; Sabar, M.F.; Hussain, S.; Parveen, R.; Hussain, M. Familial DNA analysis and criminal investigation: Usage, downsides and privacy concerns. *Forensic Sci. Int.* **2021**, *318*, 110576. [CrossRef] [PubMed]
72. Nogel, M. Some areas where digital forensics can support the addressing of legal challenges linked to forensic genetic genealogy. *Forensic Sci. Int. Digit. Investig.* **2024**, *49*, 301696. [CrossRef]
73. Phillips, C. The Golden State Killer investigation and the nascent field of forensic genealogy. *Forensic Sci. Int. Genet.* **2018**, *36*, 186–188. [CrossRef] [PubMed]
74. Law Enforcement Cases Solved Using Genetic Genealogy. International Society of Genetic Genealogy. Available online: https://isogg.org/wiki/Law_enforcement_cases_solved_using_genetic_genealogy (accessed on 18 February 2026).
75. Tillmar, A.; Fagerholm, S.A.; Staaf, J.; Sjölund, P.; Ansell, R. Getting the conclusive lead with investigative genetic genealogy—A successful case study of a 16 year old double murder in Sweden. *Forensic Sci. Int. Genet.* **2021**, *53*, 102525. [CrossRef] [PubMed]
76. Greytak, E.; Wyatt, S.; Cady, J.; Moore, C.; Armentrout, S. Investigative genetic genealogy for human remains identification. *J. Forensic Sci.* **2024**, *69*, 1531–1545. [CrossRef] [PubMed]
77. Blau, S.; Hartman, D.; Stock, A.; Leahy, F.; Leditschke, J.; Smythe, L.; Woodford, N.; Rowbotham, S. Moving from the unknown to the known: A multidisciplinary approach to the identification of skeletal remains from Sandy Point, Australia. *Forensic Sci. Res.* **2024**, *9*, owae032. [CrossRef] [PubMed]
78. Glynn, C.L. Bridging disciplines to form a new one: The emergence of forensic genetic genealogy. *Genes* **2022**, *13*, 1381. [CrossRef] [PubMed]
79. Watson, J.; McNevin, D.; Grisedale, K.; Spiden, M.; Seddon, S.; Ward, J. Operationalisation of the ForenSeq® Kintelligence Kit for Australian unidentified and missing persons casework. *Forensic Sci. Int. Genet.* **2024**, *68*, 102972. [CrossRef] [PubMed]
80. McNevin, D.; Watson, J.; Grisedale, K.; Dahal, A.; Goodwin, C.; Ward, J. Comparison of commercial targeted amplicon sequencing assays for human remains identification casework. *Int. J. Leg. Med.* **2025**, *139*, 49–60. [CrossRef] [PubMed]

81. Watson, J.L.; Grisedale, K.; McNevin, D.; Ward, J. Evaluation of the ForenSeq® Kintelligence Kit and the FORensic Capture Enrichment Panel for unidentified and missing persons casework. *Int. J. Leg. Med.* **2025**, *139*, 2047–2062. [CrossRef] [PubMed]
82. Rogalla-Ładniak, U. The overview of forensic genetic genealogy. *Arch. Med. Sadowej Kryminol.* **2022**, *72*, 211–222. [CrossRef] [PubMed]
83. Cholistani, M.S.; Raza, M.H.; Farhan, M.; Ashraf, B.; Rehman, H.U.; Arjamand, M.; Lugman, M.; Basit, A.; Shakoor, S.; Saleem, A.; et al. Forensic genealogy: A powerful investigation tool for resolving the cold cases. *J. Forensic Sci. Criminol.* **2024**, *12*, 301.
84. Wang, J.; Zhang, T.; Jia, X.; Zhang, Z.; He, J.; Ning, C.; Wang, J.; Yang, T.; Wang, G.; Pang, Y.; et al. Genomic profiling of a six-generation patrilineal family of the Ming-Qing dynasties in China. *iScience* **2025**, *28*, 112968. [CrossRef] [PubMed]
85. Ge, J.; Budowle, B. How many familial relationship testing results could be wrong? *PLoS Genet.* **2020**, *16*, e1008929. [CrossRef] [PubMed]
86. Farahany, N.; Chodavadia, S.; Katsanis, S.H. Ethical guidelines for DNA testing in migrant family reunification. *Am. J. Bioeth.* **2019**, *19*, 4–7. [CrossRef] [PubMed]
87. Kidd, K.K.; Pakstis, A.J. State of the art for microhaplotypes. *Genes* **2022**, *13*, 1322. [CrossRef] [PubMed]
88. Shabalala, S.; Ismail, N.; Ghai, M. Forensic applications of compound genetic markers: Trends and future directions. *Sci. Justice* **2025**, *65*, 101301. [CrossRef] [PubMed]
89. Carratto, T.M.T.; Moraes, V.M.S.; Recalde, T.S.F.; de Oliveira, M.L.G.; Teixeira Mendes-Junior, C. Applications of massively parallel sequencing in forensic genetics. *Genet. Mol. Biol.* **2022**, *45*, e20220077. [CrossRef] [PubMed]
90. Mitchell, R.; Peck, M.; Gorden, E.; Just, R. MixDeR: A SNP mixture deconvolution workflow for forensic genetic genealogy. *Forensic Sci. Int. Genet.* **2025**, *76*, 103224. [CrossRef] [PubMed]
91. Wang, J.; Santure, A.W. Parentage and sibship inference from multilocus genotype data under polygamy. *Genetics* **2009**, *181*, 1579–1594. [CrossRef] [PubMed]
92. Xu, H.; Liu, Y.; Lan, Q.; Yuan, X.; Lin, Y.; Wu, X.; Luo, L.; Liang, Q.; Zhu, B. Forensic application research of kinship identification disclosed by genetic polymorphisms of 150 loci in three panels. *Gene* **2025**, *965*, 149666. [CrossRef] [PubMed]
93. Mandape, S.N.; Budowle, B.; Mittelman, K.; Mittelman, D. Dense single nucleotide polymorphism testing revolutionizes scope and degree of certainty for source attribution in forensic investigations. *Croat. Med. J.* **2024**, *65*, 249–260. [CrossRef] [PubMed]
94. Wang, H.; Wei, Y.; Shan, T.; Yang, C.; Cai, Y.; Hu, Y.; Zhu, Q.; Zhang, J. Resolving kinship degrees with identity-by-descent coefficients: Limits and insights from a simulation study. *Forensic Sci. Int.* **2025**, *377*, 112647. [CrossRef] [PubMed]
95. Kruijver, M. An upper bound on the power of DNA to distinguish pedigree relationships. *Genes* **2025**, *16*, 492. [CrossRef] [PubMed]
96. Wei, Y.; Zhu, Q.; Wang, H.; Cao, Y.; Li, X.; Zhang, X.; Wang, Y.; Zhang, J. Pairwise kinship inference and pedigree reconstruction using 91 microhaplotypes. *Forensic Sci. Int. Genet.* **2024**, *72*, 103090. [CrossRef] [PubMed]
97. Tiret, M.; Hospital, F. Blocks of chromosomes identical by descent in a population: Models and predictions. *PLoS ONE* **2017**, *12*, e0187416. [CrossRef] [PubMed]
98. Nait Saada, J.; Kalantzis, G.; Shyr, D.; Cooper, F.; Robinson, M.; Gusev, A.; Palamara, P.F. Identity-by-descent detection across 487,409 British samples reveals fine scale population structure and ultra-rare variant associations. *Nat. Commun.* **2020**, *11*, 6130. [CrossRef] [PubMed]
99. He, G.; Wei, L.H.; Wang, M. Editorial: Forensic investigative genetic genealogy and fine-scale structure of human populations. *Front. Genet.* **2023**, *13*, 1067865. [CrossRef] [PubMed]
100. Edge, M.D.; Coop, G. Attacks on genetic privacy via uploads to genealogical databases. *eLife* **2020**, *9*, e51810. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.