

Comparison of MoM method and IBD segment method for kinship inference in investigative genetic genealogy

Ran Li^{1,2,3,†}, Yu Zang^{1,3,†}, Jingyi Yang^{1,3}, Nana Wang^{1,3}, Jiajun Liu^{1,3}, Enlin Wu^{1,3}, Riga Wu^{1,3}, Hongyu Sun^{1,3,*}

¹ Faculty of Forensic Medicine, Zhongshan School of Medicine, Sun Yat-sen University, Guangzhou, China

² School of Medicine, Jiaying University, Meizhou, China

³ Guangdong Province Translational Forensic Medicine Engineering Technology Research Center, Sun Yat-sen University, Guangzhou, China

*Corresponding author. E-mail: sunhy@mail.sysu.edu.cn

† These authors contributed equally to this work.

Abstract

Method of moments (MoM) and identity by descent (IBD) segment methods are two popular algorithms for kinship inference in investigative genetic genealogy (IGG). However, there is no consensus on how or when to use them, and different criteria of IBD lengths or kinship coefficients are applied to consider a true match. In this study, we compared the performance of KING (representing the MoM method) and IBIS (representing the IBD segment method) for kinship inference in homogeneous populations, admixed populations, and sparse SNP panels. Both simulated and real family data were used. In addition, an equivalent threshold-based method for kinship inference was also proposed, based on either the kinship coefficients or the lengths of IBD segments. Results showed that the overall accuracies were 64.97% for homogeneous population and 54.71% for admixed population with KING while they were 72.92% and 70.88%, respectively, with IBIS. IBIS showed low recall and precision rates when the SNP number was below 164k. In contrast, KING performed still robustly with SNP number as low as 10k. Similar results were obtained with real family data. In conclusion, the two methods perform differently and different methods should be used for different scenarios. If the DNA is of high quality or the samples are from admixed populations, the IBD segment method is recommended. If the samples are of low quantity and/or quality, MoM is more appropriate. For more complex scenarios, both methods can be tried. More importantly, there is an urgent need to develop new algorithms to address related issues.

Keywords: forensic sciences; method of moments (MoM); identity by descent (IBD) segment method; kinship inference; investigative genetic genealogy (IGG)

Introduction

Investigative genetic genealogy (IGG), also known as forensic genetic genealogy (FGG) and long-range familial search, has attracted widespread attention in recent years [1–3]. In this technique, kinship inference is performed based on genomic data from hundreds of thousands of autosomal single nucleotide polymorphisms (SNPs) typically generated with SNP microarray technology and whole genome sequencing (WGS). Since the successful application in the Golden State killer case in the USA [4], IGG has assisted in hundreds of cases, mainly involving serial/recidivist offenders and sexual violence [5, 6]. While IGG has been successful in resolving many cold cases, it can also be applied to more recent unresolved cases, especially when conventional familial searching yields no hits. Familial searching is typically performed using about 20 common short tandem repeat (STR) loci and focuses primarily on 1st degree relatives (i.e., parent–child and full siblings) [3, 7]. Due to the low power of common forensic STRs, familial searching cannot be extended to more distant relatives, e.g., first cousins, second cousins, third cousins and beyond [2, 3].

Different from familial searching, IGG has a wider scope and can search for relationships ranging from 1st to 7th degrees. As estimated by Erlich et al. [1], with a database size of ~3 million U.S. individuals of European descent (2% of

the adults of this population), more than 99% of the people would have at least a single third-cousin (7th degree relatives) match and more than 65% are expected to have at least one second-cousin (5th degree relatives) match. Therefore, there is a potentially higher probability of a match with IGG. In addition, IGG can also be combined with Y-STR database searching, with which a 14-year-old case was solved in China [8]. In this case, the 41 Y-STR profile of the suspect was first searched against the national Y-STR database and 11 male lineages were found to match the profile perfectly. However, as there were many different relationships within each male lineage (ranging from father–son to distant cousins), investigators were not able to position the suspect precisely. IGG was therefore performed to search for potential distant relatives and investigators found that one individual shared three identity by descent (IBD) segments (66 cM in total) with the suspect, thus indicating an approximate 7th degree relationship. Finally, extended sampling was conducted and the suspect was identified based on autosomal STR genotyping. Given this, IGG has advantages and can help to narrow the scope of Y-STR database searching, thus improving the efficiency of forensic investigation.

Many methods have been applied in IGG, and they can be divided into three categories: likelihood-based approach [9, 10], method of moments (MoM) [11–13], and IBD

Received: December 28, 2023. Accepted: April 8, 2025

©The Author(s) 2026. Published by the Academy of Forensic Science.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

segment method [14, 15]. The likelihood approach is well known to forensic scientists, but has been less used in IGG due to its high computational burden. The MoM method estimates unknown parameters by equating theoretical moments under the assumed statistical model to empirical moments calculated from the observed data [13]. Several parameters have been used, including the kinship coefficient (θ) [16], Cotterman's coefficient, and the coefficient of relatedness [12]. Among the commonly used software are PLINK [17] and KING [11]. The IBD segment method identifies long stretches of DNA segments that are inherited from the common ancestor(s), i.e., IBD segments, for a pair of individuals. IBD segments can be further separated into IBD1 and IBD2 segments. An IBD1 segment is a haploid match between any pair of individuals where only one pair of haplotypes is involved, whereas an IBD2 segment is a diploid match where both haplotypes of a pair of individuals match [18]. A region with no sharing is termed IBD0. Many methods have been developed for the detection of IBD segments, such as GERMLINE [14], TRUFFLE [19] and IBIS [15].

Despite many attempts at these methods, there is no consensus on how and when to use them, and different researchers and direct-to-consumer (DTC) companies have used different criteria of IBD lengths or kinship coefficient to consider a true match [2, 11, 18]. In this study, we compared the performance of two methods, KING (representing the MoM method) and IBIS (representing the IBD segment method), on homogeneous population (population possesses genetic ancestry from one source group), admixed population (population possesses genetic ancestry from multiple source groups), and sparse SNP panels (panels with decreasing numbers of SNPs). Both simulated and real family data were used. We found that different methods should be used for different scenarios.

Materials and methods

Sample collection and DNA extraction

Seventy-one individuals from a multi-generation family from Han Chinese population were recruited and 5 mL of peripheral blood was collected from each member with informed consent. DNA was extracted from 200 μ L of the blood from each sample using the QIAamp DNA Blood Mini Kit (Qiagen, Hilden, Germany). DNA was quantified using the Qubit[®] dsDNA BR Assay Kit (Thermo Fisher Scientific, San Francisco, CA, USA) with a Qubit[®] 3.0 fluorometer (Thermo Fisher Scientific) following the manufacturer's instructions. This study protocol was approved by the Ethics Committee of Zhongshan School of Medicine, Sun Yat-sen University (Guangzhou, China) (Approval No. [2020]044).

SNP genotyping

SNPs were genotyped using the Infinium Asian Screening Array (ASA; Illumina Inc., San Diego, CA, USA). This panel was designed by Illumina Inc. mainly for East and Southeast Asian populations, and 659 184 SNPs could be detected simultaneously. VCFtools [20] was then employed to exclude SNPs that deviated from Hardy-Weinberg Equilibrium or had a low minor allele frequency (MAF). We also excluded non-biallelic and non-autosomal SNPs. Following these filtering steps, approximately 300 000 (~300k) SNPs were retained for subsequent analysis.

Simulation of family data

Previous studies have shown that relatedness estimation that ignores population structure in admixed samples can produce either a positive or negative bias [21, 22]. To explore the influence of population homogeneity/admixture on kinship inference, and to assess the robustness of the two methods, we compared their performance on two distinct populations: Han Chinese and Mexicans (Mexican Ancestry from Los Angeles, MXL). Given the reported low levels of genetic admixture and limited population substructure among Han Chinese [23, 24], we considered this a relatively homogeneous population. In contrast, the MXL is a well-known admixed population, as demonstrated in many studies [24–26]. To obtain family data from the two populations, we downloaded haplotype data for 208 Han Chinese individuals (CHB and CHS) and 64 MXL individuals from the 1000 Genome Project dataset (<http://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/>). The same filtering steps described earlier were applied, resulting in over five million SNPs. We then randomly selected one-eighth (~329k SNPs) for pedigree simulation. Pedigrees were finally simulated using Ped-sim [27] with a sex-average genetic map [28], generating various relationships ranging from parent-child and full siblings to 2nd through 7th degree relatives, as well as unrelated pairs. To mimic sparse SNP panels, we also randomly selected one half of the original simulated family genetic data and generated SNP panels of approximately 164k, 82k, 41k, 20k, and 10k.

IBD segment detection and kinship coefficient estimation

The output VCF files from Ped-sim were converted to PLINK binary format (BED) using PLINK. IBD segments were detected using IBIS with command line: `IBIS -ibd2 -threads 10 -bfile simulated_family -f out`. By default, segments shorter than 7 cM were excluded to reduce spurious signals. Kinship coefficients were estimated using KING with command line: `KING -b simulated_family.bed--kinship--cpus 10--prefix out`.

Kinship inference

In this study, each simulated sample had a unique family ID, and no prior pedigree information was provided for inference. Kinship coefficients and IBD segments were therefore estimated for all pairs in the simulated dataset. However, only sample pairs from the same family were selected, given the small sample size of the reference dataset. According to the criteria in Manichaikul et al.'s study [11], we used cutoffs lying halfway between the expected kinship coefficients for different degrees of genetic relationships. Although only 1st – 3rd degree relationships were studied in Manichaikul et al.'s study, we extended the boundary to 7th degree relationships.

A pair of individuals with θ less than $0.00276 \left(\frac{1}{2^{17/2}} \right)$ was classified as unrelated. In addition, θ can also be estimated using the formula [18]:

$$\theta = \frac{L_{\text{IBD1}}}{4L_0} + \frac{L_{\text{IBD2}}}{2L_0},$$

where L_{IBD1} and L_{IBD2} denote the total lengths of IBD1 and IBD2 segments, respectively. L_0 is the total genetic length of the whole genome and $L_0 = 3\,346$ cM [28] was applied in this study. Since each person has two copies of the autosomal

chromosome, the total haplotype genetic length (L_{hapIBD}) between a pair of individuals can be calculated as $L_{hapIBD} = L_{IBD1} + 2 \times L_{IBD2}$. So, we have $\theta = \frac{L_{hapIBD}}{4L_0}$. Specifically, we may also have $\theta = \frac{L_{IBD1}}{4L_0}$ for all relationships except full siblings

because IBD2 is only observed in full siblings in non-inbred populations. Now, the cutoffs of IBD segment (t_{IBD}) to determine a relationship can be deduced by simply multiplying the kinship coefficient cutoffs (t_θ) by $4L_0$ (Table 1). This demonstrates that it is equivalent to use either kinship coefficients or the lengths of IBD segments for kinship inference.

To evaluate and compare the performance of the two tools for kinship inference, three parameters were used: recall rate, precision rate, and accuracy. Recall rate denotes the proportion of a known relationship that can be recalled with the cutoffs; precision rate indicates the proportion of a relationship that is correctly assigned; and accuracy is the proportion that the real and the predicted relationships are consistent. Supplementary Table S1 shows how the three parameters are calculated and readers can also refer to the website (<https://www.evidentlyai.com/classification-metrics/>)

Table 1. Cutoffs of kinship coefficient (θ) and total length of identity by descent (IBD) segments for kinship inference.

Degree of relatedness	Cutoffs	
	Kinship coefficient (t_θ)	Total length of IBD segments (t_{IBD})*
Monozygotic twins (identical individuals)	$\geq \frac{1}{2^{3/2}}$	$\geq 4\ 732$
1st	$\left[\frac{1}{2^{5/2}}, \frac{1}{2^{3/2}} \right)$	[2 366, 4 732)
2nd	$\left[\frac{1}{2^{7/2}}, \frac{1}{2^{5/2}} \right)$	[1 183, 2 366)
3rd	$\left[\frac{1}{2^{9/2}}, \frac{1}{2^{7/2}} \right)$	[591, 1 183)
4th	$\left[\frac{1}{2^{11/2}}, \frac{1}{2^{9/2}} \right)$	[296, 591)
5th	$\left[\frac{1}{2^{13/2}}, \frac{1}{2^{11/2}} \right)$	[148, 296)
6th	$\left[\frac{1}{2^{15/2}}, \frac{1}{2^{13/2}} \right)$	[74, 148)
7th	$\left[\frac{1}{2^{17/2}}, \frac{1}{2^{15/2}} \right)$	[37, 74)
Unrelated	$< \frac{1}{2^{17/2}}$	<37

* The total genetic length of the whole genome is $L_0 = 3\ 346$ cM.

multi-class-metrics) for more details. All data processing and statistical analysis were performed in R software v.3.6.1, and figures were generated using the ggplot2 package.

Results

Comparison of threshold and machine learning methods for kinship inference

In this study, 360 pairs were simulated for each relationship, including 1st – 7th degree relatives and unrelated pairs. In order to compare the performance of our threshold-based method (Table 1) and machine learning-based approaches for kinship inference, relationships were also predicted using a supervised k-Nearest Neighbour (k-NN) algorithm. Two-thirds of the simulated family data (based on the 208 Han Chinese individuals) were treated as the training set, and the remaining family data were treated as the testing set, with the number of neighbours (k) set to 5. The results are shown in Figure 1. Overall, our threshold-based method showed performance comparable to the k-NN method, with the main differences observed for distant relationships and unrelated pairs. When using IBD segments detected by IBIS, the threshold-based method had recall and precision rates similar to those of the k-NN method. The largest difference was observed for the 7th degree relationship. The recall rate for k-NN was 51.75%, about 1.68 times higher than that of the threshold-based method. The k-NN method also achieved higher precision rate than the threshold-based method for unrelated pairs (92.03% vs. 62.94%). When kinship coefficients estimated by KING were used for kinship inference, the two methods also showed similar performance. The threshold-based method had a higher recall rate for unrelated pairs (90.00% vs. 63.08%) but had a lower recall rate for the 7th degree relationship (16.67% vs. 28.95%). The overall accuracies were 64.97% (using kinship coefficients) and 72.92% (using IBD segments) for our threshold-based method, similar to those of the k-NN method (61.98% with KING and 76.28% with IBIS).

Performance on homogeneous and admixed populations

All samples were assigned a relationship based on the cutoffs in Table 1, and the results are shown in Figure 2. For the 1st degree relationships (parent-child and full siblings), recall rates were 100% with both KING and IBIS, but the precision rate was 99.45% with KING, although it remained 100% with IBIS. For the 2nd degree relationships, the two tools showed similar performance, with recall rates of 95.56% vs. 97.78%, and precision rates of 95.03% vs. 94.62%. However, for more distant relationships, both recall rates and precision rates decreased significantly, with IBIS outperforming KING. It appeared that the more distant the relationship, the greater the difference in recall rates and precision rates for the two methods. In particular, the recall rate was 100% with IBIS but only 90.00% with KING for unrelated individuals. This means 10% of unrelated pairs were incorrectly predicted to be related, mainly as 6th and 7th degree relationships. If one degree difference between the real and predicted relationship was permitted, both recall rates and precision rates increased significantly for both methods (Supplementary Figure S1A).

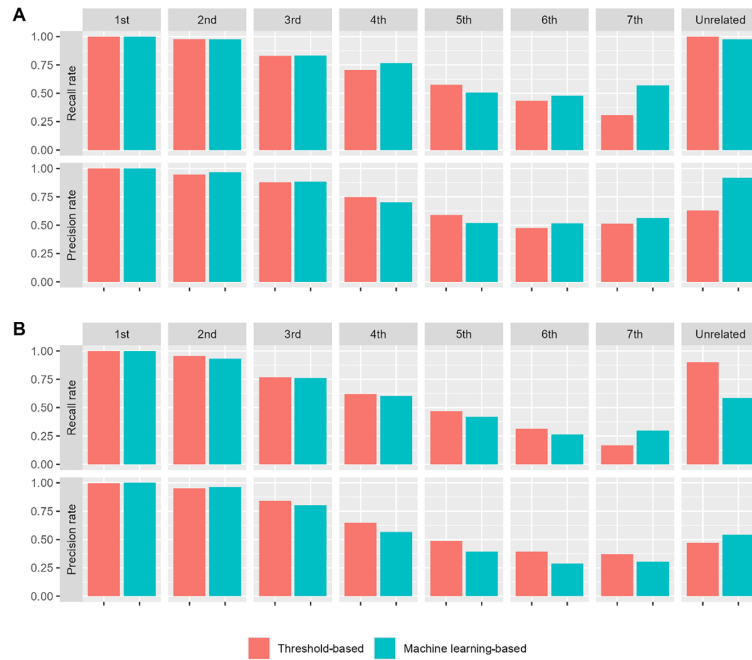


Figure 1 Comparison of the performance of threshold and machine learning methods for kinship inference. Recall rates and precision rates of kinship inference based on the identity by descent (IBD) segments detected by IBIS (A) and the kinship coefficients estimated by KING (B).

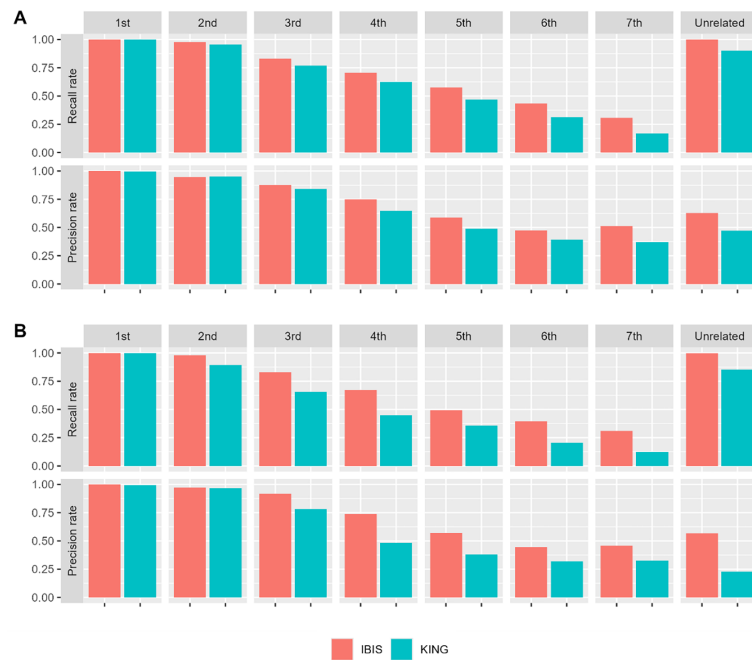


Figure 2 Performance of KING and IBIS for kinship inference on homogeneous and admixed population. Recall rates and precision rates of kinship inference on homogeneous population (A) and admixed population (B).

Both KING and IBIS had recall rates and precision rates over 95% for 1st – 4th degree relationships, although IBIS performed better than KING for 5th – 7th degree relationships. The overall accuracies increased from 64.97% to 90.94% with KING, and from 72.92% to 97.64% with IBIS.

For the admixed population (MXL), IBIS showed nearly identical performance, with an overall accuracy of 70.88%.

However, KING’s performance declined. Although the recall and precision rates were 100% and 99.34% when using KING, similar to those of the homogeneous population, both parameters decreased significantly for 2nd to 7th degree relationships. The greatest decrease was observed for unrelated individuals. The recall rate decreased from 90.00% to 85.33% and the precision rate decreased from 47.09% to 22.94%.

The overall accuracy decreased from 64.97% to 54.71%. A similar trend was also observed if one degree difference was permitted ([Supplementary Figure S1B](#)).

Performance on sparse SNP panels

Samples from the six simulated SNP panels (329k, 164k, 82k, 41k, 20k, and 10k) were assigned a relationship based on the cutoffs in [Table 1](#), and the results are shown in [Figure 3](#). With fewer SNPs, both recall rates and precision rates decreased for the two methods, but KING was far more robust than IBIS. With the 164k panel, the recall rate and precision rate were only slightly lower than with the original 329k panel for IBIS. However, as the number of SNPs decreased further, both rates decreased sharply. In particular, with only 20k SNPs, only about half of the 1st degree relationships were recalled, despite a high precision rate (100%). With only 10k SNPs, the recall rates and precision rates for 1st – 7th degree relationships all fell below 1%. Regardless of a high recall rate for unrelated individuals (100%), IBIS showed markedly reduced precision rates with these sparse panels. In contrast, KING performed quite robustly. Recall rates and precision rates remained >99% for 1st degree relationship and about 95% for 2nd degree relationship, even with only 10k SNPs. Although both rates decreased with fewer SNPs for more distant relationships, they were still much higher than those achieved by IBIS. Furthermore, we evaluated recall rates and precision rates when one degree difference was permitted. As expected, both rates increased significantly for both methods, and KING still outperformed IBIS ([Supplementary Figure S2](#)).

Performance on real family data

From the 71 members of the studied family, 143, 207, 138, 53, 67, 101, 73 pairs of 1st to 7th degree relationships as well as 1 693 unrelated pairs were obtained. Although 10

pairs of 8th degree relationships were also yielded, they were not included in this study. After relationship assignment, the results were similar to those of the simulated family data. [Tables 2](#) and [3](#) show the details of the real and predicted relationships based on kinship coefficients and the total lengths of IBD segments. Recall rates and precision rates were 100% for the 1st degree relationships (parent–child and full siblings) when using KING and IBIS. Four out of 207 pairs of 2nd degree relationships were incorrectly predicted as 3rd degree relationships with KING, thus resulting in a recall rate of 98.07%. In contrast, only one pair was misclassified and a high recall rate (99.52%) was achieved with IBIS. Regardless, all the sample pairs predicted as 2nd degree relationships were correct for both methods, i.e., 100% precision rates. Of the 138 pairs of 3rd degree relationships, 29 pairs were incorrectly predicted as 4th degree relationships and 3 pairs were even incorrectly predicted as 5th degree relationships when KING was used. Thus, the recall rate was 76.81% with KING while it was 83.33% with IBIS. For 4th to 7th degree relationships, IBIS performed much better than KING, with recall rates about 2 to 5 times higher than those of KING, and precision rates about 1.7 to 3 times higher. Notably, among the 1 693 pairs of unrelated individuals, none were predicted as related with IBIS while 20 pairs were incorrectly predicted as 7th degree relationships and 8 pairs were even incorrectly predicted as 6th degree relationships when KING was used. The overall accuracies were 87.23% with KING and 91.84% with IBIS. If these unrelated pairs were excluded, the overall accuracies were 63.17% and 74.17% for KING and IBIS, respectively, very similar to those of the simulated family data. If one degree difference was permitted, both recall rates and precision rates increased significantly for the two methods. The overall accuracies increased to 93.09% with KING and to 98.30% with IBIS for all relationships.

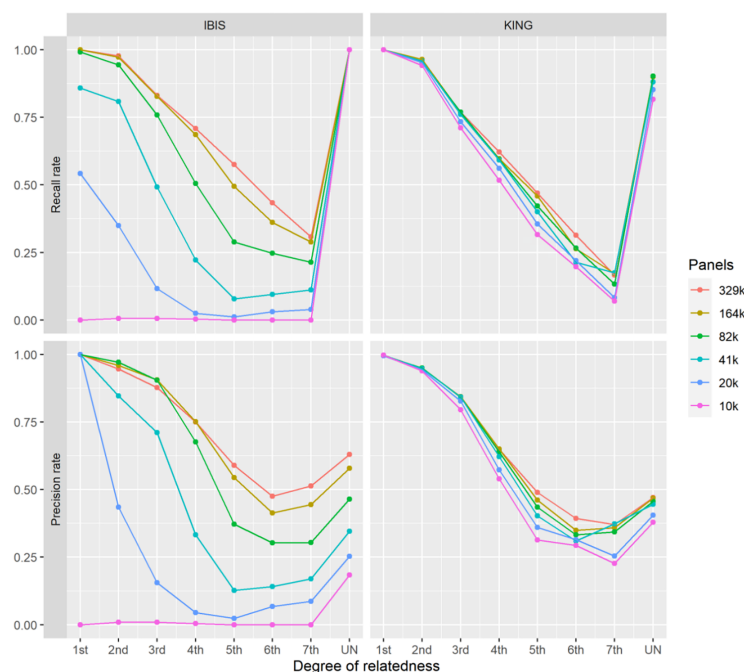


Figure 3 Performance of KING and IBIS for kinship inference on sparse SNP panels. UN: unrelated individuals.

Table 2. The real and predicted relationships based on kinship coefficients estimated by KING. Relationships were determined by using the cutoffs in Table 1.

Real	Predicted								Recall rate
	1st	2nd	3rd	4th	5th	6th	7th	Unrelated	
1st	143	0	0	0	0	0	0	0	1
2nd	0	203	4	0	0	0	0	0	0.980 7
3rd	0	0	106	29	3	0	0	0	0.768 1
4th	0	0	0	13	21	7	5	7	0.245 3
5th	0	0	0	0	13	12	5	37	0.194 0
6th	0	0	0	1	5	12	5	78	0.118 8
7th	0	0	0	0	0	0	4	69	0.054 8
Unrelated	0	0	0	0	0	8	20	1 665	0.983 5
Precision rate	1	1	0.963 6	0.302 3	0.309 5	0.307 7	0.102 6	0.897 1	Accuracy = 0.872 3

Table 3. The real and predicted relationships based on the total lengths of identity by descent (IBD) segments detected by IBIS. Relationships were determined by using the cutoffs in Table 1.

Real	Predicted								Recall rate
	1st	2nd	3rd	4th	5th	6th	7th	Unrelated	
1st	143	0	0	0	0	0	0	0	1
2nd	0	206	1	0	0	0	0	0	0.995 2
3rd	0	0	115	23	0	0	0	0	0.833 3
4th	0	0	2	29	18	4	0	0	0.547 2
5th	0	0	0	2	31	22	7	5	0.462 7
6th	0	0	0	0	3	37	35	26	0.366 3
7th	0	0	0	0	0	5	19	49	0.260 3
Unrelated	0	0	0	0	0	0	0	1 693	1
Precision rate	1	1	0.974 6	0.537 0	0.596 2	0.544 1	0.311 5	0.954 9	Accuracy = 0.918 4

Discussion

IGG has attracted widespread attention in forensic genetics in the last 5 years [2, 3, 5, 29, 30]. MoM and IBD segment methods are two popular algorithms for kinship inference. In this study, we compared the performance of KING (representing the MoM method) and IBIS (representing the IBD segment method) and showed that different methods should be applied to different scenarios.

The main advantage of the MoM method is its high computational efficiency. It typically requires tens of seconds to a few minutes, whereas IBD segment methods often require several hours to days (including the time to phase genetic data for some methods) [22]. It has been widely used in many fields, such as genetic association studies, and population genetic analyses. Another advantage is its requirement for far fewer SNPs. As shown in this study, the recall rates and precision rates remained >99% for 1st degree relationship, and about 95% for 2nd degree relationship even with only 10k SNPs, whereas IBIS recalled only about 1% of relationships at this density. In addition, Turner et al. [31] showed that KING was more robust and outperformed IBD segment methods in the presence of genotyping error, especially at higher error rates (1%–5%). In this regard, the MoM method (e.g., KING) appears well-suited for IGG, because samples from the criminal scenes are often fragmented, of low quantity, and/or mixed with DNA from multiple individuals. It is error-prone when we generate genetic data from these challenging samples [32, 33]. However, this method has a high false positive rate, which is problematic for large database searches. As estimated in this study, about 10% of unrelated pairs were incorrectly predicted as related. This would yield nearly 100 000 false positive candidates when searching a database containing one million individuals, a number impractical

for subsequent extensive non-DNA investigative work. Furthermore, some MoM methods, e.g., KING, showed reduced performance in the presence of population structure, which might be unknown or poorly defined in forensic practice. Therefore, methods that account for population structure, such as PC-relate [21], are recommended.

In contrast, the IBD segment method has a high accuracy in kinship inference and an extremely low false positive rate for both homogeneous and admixed populations. Notably, among the 1 693 pairs of unrelated individuals from the family studied, none were predicted as related by IBIS, making it powerful for large database searches. It can also discriminate between relationships of the same degree based on different IBD signals [34, 35], and has the potential to ease the following non-DNA genealogy searching. However, IBD segment method generally requires a large number of SNPs to be analyzed and is vulnerable to genotyping error, which is common for crime scene samples [32, 33]. This means the IBD segment method may have a high false negative rate for genetic data with high genotyping error. Nevertheless, in 2022, Snedecor et al. [36] proposed an IBD-based method that was accurate out to 5th degree relatives using only 10 000 SNPs. The windowed kinship algorithm uses thresholds that are slightly lower than theoretical values, making it also relatively robust to genotyping error.

Generally, genotyping error is often unknown unless a reference sample is provided, but samples of high quality generally have low genotyping error [33]. Therefore, if a DNA sample meets the requirements of the platforms used, e.g., approximately 200 ng of genomic DNA and a good DNA Degradation Index (DI) score for SNP microarrays, high-quality SNP data can be obtained and IBD segment method is recommended, which is also widely used by many

DTC companies [2]. The IBD segment method is also preferred for samples from admixed populations. Conversely, for samples of low quantity and/or quality, the MoM method is more appropriate. In more complex scenarios involving low-quantity/quality samples from admixed populations, neither method may perform optimally. In such cases, we therefore recommend using both methods, given the high robustness of the MoM method and the high accuracy of the IBD segment method.

Furthermore, we showed that both methods had lower performance in admixed populations than in homogeneous populations, with a slight decrease in accuracy for IBIS (from 72.92% to 70.88%) and a significant decrease for KING (from 64.97% to 54.71%). As kinship coefficients or the total lengths of IBD segments were underestimated for the majority of relationships (Supplementary Figure S3), many related individuals were incorrectly classified as unrelated, which could explain the greatest decrease in precision rates for unrelated pairs. In addition, there were also more unrelated individuals that were assigned as related. Therefore, there is a higher risk of missing true matches and of falsely including unrelated individuals in admixed populations. Given this, genetic background should be taken into account and there is an urgent need to develop new algorithms to address these issues. Currently, IGG practitioners are dependent on the kinship algorithms chosen by database providers. However, as the need for IGG grows, there will be a large number of DNA profiles of low quality, and alternative methods or new algorithms may be implemented by the database owners. More importantly, forensic labs may have their own databases in the future (e.g., a small database of disaster victims). In such cases, kinship inference methods that are robust to challenging crime scene samples will be preferred and should be implemented.

Another contribution of this study is our proposal of an equivalent threshold-based method for kinship inference using either kinship coefficients or the lengths of IBD segments. We also explored several machine learning approaches, finding k-NN to perform best among them. Nevertheless, our threshold-based method showed slightly superior performance to the k-NN approach. However, machine learning approaches (e.g., k-NN) have some limitations. They typically require large training datasets to achieve reliable performance, which may be a challenge for some laboratories. In addition, there may be inconsistent or even opposite conclusions with different models.

Although attractive, IGG also has some limitations. First, the technology is not compatible with existing platforms, and many forensic laboratories do not have the required infrastructure/instrumentation to perform SNP microarrays or WGS in-house [3, 30]. Furthermore, although SNPs have smaller amplicons than STRs and are expected to have better performance for degraded samples, it appears that these technologies do not outperform traditional forensic platforms for the analysis of degraded samples [32, 33]. Second, IGG is relatively expensive, not only in money but also in time. It has been estimated that about US\$3 000 to US\$10 000 per case is required for the IGG service, much higher than the cost of traditional familial searching and Y-STR database searching [3]. In addition, as many matches may be obtained through a database search and each match may have hundreds to thousands of distant relatives [9, 37], more efforts are required for the follow-up non-DNA investigations. The time taken for IGG

investigations to resolve cases is weeks to months whereas familial searching takes only days to weeks [3]. Third, there are also some ethical issues associated with this approach, which have been discussed widely [2, 37]. Nevertheless, Budowle et al. [38] argue that IGG is cost-effective at a system level, as it prevents police resources from being wasted on ineffective methods and helps prevent crimes and victimization.

There are some limitations of this study. First, we employed only one approach to represent the two types of kinship inference methods: KING for the MoM method and IBIS for the IBD segment method. However, it is worth noting that there are numerous other tools that utilize distinct principles, which may have different performance compared to the two methods studied. In particular, approaches used for the analysis of ancient DNA may be promising tools for challenging crime scene samples [39]. Second, the sample size of real family dataset was relatively small and, more importantly, the performance of related methods on real challenging samples was not explored, which will be studied in our future work.

Conclusion

The MoM method (e.g., KING) is fast and requires fewer SNPs. Nevertheless, it has reduced performance for distant kinship inference and in admixed populations. In contrast, the IBD segment method (e.g., IBIS) achieves high accuracy for both close and distant relationships, with an extremely low false positive rate across homogeneous and admixed populations. However, it is vulnerable to genotyping errors, and may not be suitable for challenging crime scene samples. Therefore, for high-quality samples, particularly those from admixed populations, the IBD segment method is recommended. In contrast, if the samples are of low quantity/quality, MoM is more appropriate. For more complex scenarios, both methods can be tried. More importantly, there is an urgent need to develop new algorithms to address these issues.

Acknowledgements

The authors would like to thank all the volunteers for their contributions.

Authors' contributions

Ran Li and Hongyu Sun were responsible for the study's conceptualization and methodology. Data collection was handled by Yu Zang. Data analysis was performed by Yu Zang, Jingyi Yang and Enlin Wu. Ran Li and Yu Zang drafted the manuscript. All authors participated in reviewing and editing the manuscript. Visualization was managed by Nana Wang and Jiajun Liu. Project administration was handled by Riga Wu and Hongyu Sun. Riga Wu provided project supervision. Ran Li and Hongyu Sun acquired funding. All authors reviewed and approved the final manuscript.

Compliance with ethical standard

Informed consent was obtained from all participants. This study was approved by the Ethics Committee of Zhongshan

School of Medicine, Sun Yat-sen University (Guangzhou, China) (Approval No. [2020]044).

Disclosure statement

The authors report there are no competing interests to declare.

Funding information

This research was funded by the National Natural Science Foundation of China [grant numbers 82302125 and 81971798] and the Natural Science Foundation of Guangdong Province [grant number 2019A1515011527].

References

1. Erlich Y, Shor T, Pe'er I, et al. Identity inference of genomic data using long-range familial searches. *Science*. 2018;362:690–694.
2. Kling D, Phillips C, Kennett D, et al. Investigative genetic genealogy: current methods, knowledge and practice. *Forensic Sci Int Genet*. 2021;52:102474.
3. Ge J, Budowle B. Forensic investigation approaches of searching relatives in DNA databases. *J Forensic Sci*. 2021;66:430–443.
4. Kaiser J. We will find you: DNA search used to nab Golden State Killer can home in on about 60% of white Americans. *Science*. 2018. doi: 10.1126/science.aav7021.
5. Greytak EM, Moore CC, Armentrout SL. Genetic genealogy for cold case and active investigations. *Forensic Sci Int*. 2019;299:103–113.
6. Dowdeswell TL. Forensic genetic genealogy: a profile of cases solved. *Forensic Sci Int Genet*. 2022;58:102678.
7. Bieber FR, Brenner CH, Lazer D. Finding criminals through DNA of their relatives. *Science*. 2006;312:1315–1316.
8. Liu J, Zhang Y, Wang H, et al. Forensic SNP genealogical inference helping ferret out a 14-year long unsolved cold case. *Forensic Sci Technol*. 2021;46:652–656. Chinese.
9. 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.
10. Lander ES, Green P. Construction of multilocus genetic linkage maps in humans. *Proc Natl Acad Sci U S A*. 1987;84:2363–2367.
11. Manichaikul A, Mychaleckyj JC, Rich SS, et al. Robust relationship inference in genome-wide association studies. *Bioinformatics*. 2010;26:2867–2873.
12. Hauser SS, Galla SJ, Putnam AS, et al. Comparing genome-based estimates of relatedness for use in pedigree-based conservation management. *Mol Ecol Resour*. 2022;22:2546–2558.
13. Speed D, Balding DJ. Relatedness in the post-genomic era: is it still useful? *Nat Rev Genet*. 2015;16:33–44.
14. Gusev A, Lowe JK, Stoffel M, et al. Whole population, genome-wide mapping of hidden relatedness. *Genome Res*. 2009;19:318–326.
15. Seidman DN, Shenoy SA, Kim M, et al. Rapid, phase-free detection of long identity-by-descent segments enables effective relationship classification. *Am J Hum Genet*. 2020;106:453–466.
16. Thompson EA. Identity by descent: variation in meiosis, across genomes, and in populations. *Genetics*. 2013;194:301–326.
17. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analysis. *Am J Hum Genet*. 2007;81:559–575.
18. Naseri A, Shi J, Lin X, et al. RAFFI: accurate and fast familial relationship inference in large scale biobank studies using RaPID. *PLoS Genet*. 2021;17:e1009555.
19. Dimitromanolakis A, Paterson AD, Sun L. Fast and accurate shared segment detection and relatedness estimation in unphased denetic data *via* TRUFFLE. *Am J Hum Genet*. 2019;105:78–88.
20. Danecek P, Auton A, Abecasis G, et al. The variant call format and VCFtools. *Bioinformatics*. 2011;27:2156–2158.
21. Conomos MP, Reiner AP, Weir BS, et al. Model-free estimation of recent genetic relatedness. *Am J Hum Genet*. 2016;98:127–148.
22. Ramstetter MD, Shenoy SA, Dyer TD, et al. Benchmarking relatedness inference methods with genome-wide data from thousands of relatives. *Genetics*. 2017;207:75–82.
23. Wu W, Chen D, Fu Y, et al. Population structure of Han population in China revealed by 41 STR loci. *Ann Hum Biol*. 2020;47:65–69.
24. Bulbul O, Pakstis AJ, Soundararajan U, et al. Ancestry inference of 96 population samples using microhaplotypes. *Int J Legal Med*. 2018;132:703–711.
25. Thornton T, Tang H, Hoffmann TJ, et al. Estimating kinship in admixed populations. *Am J Hum Genet*. 2012;91:122–138.
26. McHugh C, Brown L, Thornton TA. Detecting heterogeneity in population structure across the genome in admixed populations. *Genetics*. 2016;204:43–56.
27. Caballero M, Scelza BA, Gusev A, et al. Crossover interference and sex-specific genetic maps shape identical by descent sharing in close relatives. *PLoS Genet*. 2019;15:e1007979.
28. Bherer C, Campbell CL, Auton A. Refined genetic maps reveal sexual dimorphism in human meiotic recombination at multiple scales. *Nat Commun*. 2017;8:14994.
29. Phillips C. The Golden State Killer investigation and the nascent field of forensic genealogy. *Forensic Sci Int Genet*. 2018;36:186–188.
30. Glynn CL. Bridging disciplines to form a new one: the emergence of forensic genetic genealogy. *Genes (Basel)*. 2022;13:544.
31. Turner SD, Nurkoltsev A, Sun Y, et al. Evaluating the impact of dropout and genotyping error on SNP-based kinship analysis with forensic samples. *Front Genet*. 2022;13:882268.
32. de Vries JH, Benschop CCG, van der Linden J, et al. 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.
33. Davawala A, Suresh Kumar G, Kher S, et al. Forensic genetic genealogy using microarrays for the identification of human remains: the need for good quality samples—a pilot study. *Forensic Sci Int*. 2022;334:111265.
34. Williams CM, Scelza BA, Slack SD, et al. A rapid accurate approach to inferring pedigrees in endogamous populations. *Genetics*. 2025;230:iyaf094.
35. Morimoto C, Manabe S, Fujimoto S, et al. Discrimination of relationships with the same degree of kinship using chromosomal sharing patterns estimated from high-density SNPs. *Forensic Sci Int Genet*. 2018;33:10–16.
36. Snedecor J, Gusev A, Buxbaum S, et al. Fast and accurate kinship estimation using sparse SNPs in relatively large database searches. *Forensic Sci Int Genet*. 2022;61:102770.
37. Syndercombe Court D. Forensic genealogy: some serious concerns. *Forensic Sci Int Genet*. 2018;36:203–204.
38. Budowle B, Arnette A, Sajantila A. A cost-benefit analysis for use of large SNP panels and high throughput typing for forensic investigative genetic genealogy. *Int J Legal Med*. 2023;137:1595–1614.
39. Childebayeva A, Zavala EI. Review: computational analysis of human skeletal remains in ancient DNA and forensic genetics. *iScience*. 2023;26:105805.