Arch Med Sadowej Kryminol 2022; vol. 72 (4): 211-222 DOI: 10.4467/16891716AMSIK.22.023.17623

archiwum medycyny sądowej i kryminologii

Praca przeglądowa
Review paper

Urszula Rogalla-Ładniak

The overview of forensic genetic genealogy

Sądowa genealogia genetyczna – przegląd zagadnień

Department of Forensic Medicine, Collegium Medicum in Bydgoszcz, Nicolaus Copernicus University in Toruń

Abstract

Forensic genetic genealogy (FGG) benefits largely from popularity of genealogical research within (mostly) American society and the advent of new sequencing techniques that allow typing of challenging forensic samples. It is considered a true breakthrough for both active and especially cold cases where all other resources and methods have failed during investigation. Despite media coverage generally highlighting its powers, the method itself is considered very laborious and the investigation may easily got suspended at every stage due to many factors including no hits in the database or breaks in traceable lineages within the family tree. This review summarizes the scope of FGG use, mentions most concerns and misconceptions associated with the technique and points to the plausible solutions already suggested. It also brings together current guidelines and regulations intended to be followed by law enforcement authorities wishing to utilize genetic genealogy research.

Key words: forensic genetics, investigative genetic genealogy, familial searching

Streszczenie

Sądowa genealogia genetyczna korzysta w głównej mierze z popularności badań genealogicznych (zwłaszcza) wśród Amerykanów oraz z pojawienia się nowych technik sekwencjonowania umożliwiających analizę trudnych próbek pochodzących z miejsca zdarzenia. Metoda uważana jest za prawdziwy przełom zarówno dla bieżących śledztw, jak i tych, w których wszystkie inne zasoby i techniki zawiodły podczas toczącego się śledztwa. Pomimo, że doniesienia medialne w sposób istotny podkreślają jej wyjątkową wartość, sądowa genealogia genetyczna jest bardzo pracochłonna a dochodzenie może zostać zawieszone na każdym etapie ze względu na wiele czynników w tym brak trafień w dostępnych bazach danych lub niekompletność drzew genealogicznych. Niniejszy przegląd podsumowuje zakres zastosowań sądowej genealogii genetycznej, wymienia wiele związanych z nią wątpliwości i nieporozumień a także wskazuje dostępne rozwiązania dla poruszanych problemów. Zawiera również aktualne wytyczne i przepisy, których powinny przestrzegać organy ścigania chcące wykorzystywać sądową genealogię genetyczną.

Słowa kluczowe: genetyka sądowa, sądowa genetyka genealogiczna, przeszukiwanie rodzinne

Nadesłano / Submitted: 2023-02-06
Zaakceptowano / Accepted: 2023-03-29

Introduction

Forensic Genetic Genealogy (FGG) also referred to as Investigative Genetic Genealogy (IGG) is a relatively new branch of forensics that brings together traditional genealogical investigations, latest developments in DNA analysis, public DNA profiles repositories and a magnitude of statistical considerations to achieve simple goal – a DNA match. Traditional genetic identification relies upon a very well-known principle of DNA sharing between related individuals that may be easily estimated on the basis of known relationships. As long as the kinship is relatively close, e.g. parents with offspring or siblings, usually use of a typical set of 15-21 STRs is enough for relationship inference, although in some cases additional STRs need to be included. More distant relationships require extension of DNA markers scope to be analyzed or inclusion of profiles of additional relatives, preferably as immediate as possible. The more distant the relatives are, the more dense the markers ought to be, therefore a switch from STRs to SNPs (or at least a combination of these) needs to be made. What makes SNPs better markers for distant relationship inference is their much lower (compared to STRs) mutation rate, high multiplexing capability (as a consequence of short amplicon sizes) and their huge number spread across the entire genome. This raises some concerns within forensic geneticists community as the markers used for identification should not be linked to any visual or more generally phenotypic or bearing any medical information traits and these criteria are definitely not met if a huge number of SNPs is to be analyzed (1). One needs to bear in mind, however, that this reservation does not apply to forensic DNA phenotyping (FDP) that may be basically inferred from SNPs in coding regions. Although most of the European countries have no explicit regulations on the FDP use, there are several exceptions: Germany, Slovakia, the Netherlands (2) and Switzerland among which only in the latter inference of "personal characteristics" on the basis of DNA testing is still forbidden (3).

However, most European countries have no regulations.

In the case of forensic genetic genealogy promising leads may only be generated if thousands of single nucleotide markers are typed – in fact this indirect kind of matches search is based on over 600 thousand SNPs data (4) typically stored for each sample by the high-resolution direct-to-consumer genetic databases (DTC). Nevertheless this huge number of SNPs was traditionally sufficient only for reliable searches including up to first cousins, who share on average 12,5% of their DNA. If pairwise comparisons included some more distant relatives like second cousins, one needed to bear in mind it was only on average 3,13% of DNA shared between them. So the more distant the relatives are, the less DNA is shared and the probability of a direct genetic match inevitably drops to zero at some certain point. But if these SNPs are analyzed as linked variation where whole sections of chromosomes are shared among relatives, then these carry much more information on genetic relativeness than simple STR typing and adding external information gathered with the use of genetic genealogy techniques may vitally aid the investigation by building complex family trees and providing matches with quite distant relatives. The true problem with FGG arises however when a potential link leads to nowhere and the scope of analyzed marker becomes irrelevant – it is quite obvious that by no means are all genealogical relationships also genetic ones (5). Other potential problems with reliable genealogy reconstruction include endogamy, misattributed parentage, adoption or gamete donation.

Some studies have shown a general utility of dense marker typing for distant relatives search reaching 80% correct inferences for sixth and seventh-degree cousins (6) but it is very dependent on the population and allele frequencies and informativeness of specific loci. As a matter of fact all inferences beyond a third or fourth-degree cousins become heavily complicated so further analyses are rarely performed. It has been shown, however, that an appealing proportion of third cousins (sharing great-great-grandparent) reaching 90% can be correctly matched (7) using identity-by-descent presumptions based on dense SNP data (from HGDP-CEPH – Human Genome Diversity Panel – Centre d'Étude du Polymorphisme Humain and 23andMe.com service in this case). There is a crowd-funded venture "The Shared cM Project" (8) by Blaine Bettinger, that summarizes empirical data from over

60 000 relationships on length of DNA segments shared between relatives expressed in centimorgans (cMs). It highlights an overlap between the amount of DNA shared between relatives of various degrees thus explaining why very precise inference is not likely to be possible ever.

The concept of database screening is not that new – for over 20 years forensic specialists have been using so called familial searching. It arose from the necessity of extending the scope of search from direct matching to including some close relatives who share a significant proportion of their DNA with a donor of the questioned sample. Its principle is different though, as what is analyzed is a standard autosomal STR profile or Y-STR profile and a rank of likelihood ratios is prepared that summarizes relationships between a questioned sample and national database's (usually curated by the police) hits (9). Its use is limited to close relatives including parents, children and siblings but false positives are also sometimes encountered – especially if the investigated number of STR is small and no ancestry inference on the query profile was performed (10).

Although the Interpol has been carrying out direct identification using DNA database since 2004, only recently it has launched an I-Familia global database collecting profiles from missing persons and allowing for identification using DNA kinship matching on the basis of international allele frequencies coupled with efficient LR algorithms performing millions of calculations even for partial profiles (11). As a matter of fact it's role is to fill the gap being a consequence of the Prüm Decisions from 2008 (on exchanging information by the law enforcement authorities across the EU) that were specifically lacking regulations on automatic exchange of kinship data.

Databases

Among resources being of a great use for FGG there are both publicly-available open-data personal genomics databases as well as so called direct-to-consumer genetic genealogy services. First products aiming at genetic genealogy inference were offered fifteen years ago by the 23andMe company. Since then they performed over 12 million tests and were overtaken only by the AncestryDNA that stores information from over 21 million users in their database. The numbers are not surprising taking into account that searching for relatives and genealogical research are said to be second most popular hobby within the US (12) as stated on Centers for Disease Control and Prevention website. One need to note, however, that these records are not directly available for legal authorities as these companies do not allow usage for FGG purposes without a warrant.

An up-to-date comparison chart of specific databases content, scope, number of records, offered services etc. is available from International Society of Genetic Genealogy Wiki curated by Dr. Tim Jenzen (13). Among the biggest players being 23andMe, AncestryDNA, MyHeritage and FamilyTreeDNA (FTDNA) only the latter allows for searching its content by legal authorities and provides information on users. Its cooperation with legal authorities is not limited to the US only and in case of other countries access in granted on case-by-case basis upon filling basic criteria. Database content corresponding to about 1.2M users is the smallest, though. Noteworthy, the parent company of FTDNA (Gene By Gene) established its own forensic laboratory performing genetic testing.

A huge difference in genetic genealogy was brought by the GEDmatch – a service that gathers information extracted from virtually all biggest players on the market (users of all of these may download raw data files). It is worth noting that the company was taken over by Verogen (forensic genomics company) in 2019. A customer may upload his own data in a specific file format GEDCOM prepared by external company like 23andMe and others and choose from a variety of tools offered by the GEDmatch to analyze ancestry admixture, search for relatives or build his family tree on the basis of other public profiles stored within the database. Upon registration, the user needs to choose privacy options and if he is after the "public opt-in" option (which is the default option – highly recommended), his data is not only to be processed for personal genetic genealogy research but also by law enforcement agencies solving violent crimes. One need to note, however, that all European users are automatically opted out in reference with the EU General Data Protection Regulation. Investigations associated with identifying unidentified human remains may freely use

matching options encompassing all GEDmatch user data. There is a specific tool – GEDmatchPRO designed for the police and forensic specialists that aids investigative comparisons. It allows for uploading suspect's profile generated with the ForenSeq Kintelligence Kit (14) (Verogen) or any other data source accepted by GEDmatch and processing it for FGG purposes using solutions provided by the service.

However, all DTC genetic genealogical services must respond and comply with court orders, subpoenas or search warrants for data as required by applicable law. Interestingly, 23andMe publishes transparency reports stating how many times legal authorities requested access to the data and if they were given any. As for January 13th 2023 there were 11 requests (all from the US authorities) and all were rejected by the company (15). Among all DTC services AncestryDNA has been the only that defined criteria and the process of IBD search and made it public (16).

It has been shown, that using "genealogical triangulation" (intersections), for over 50% of targets, their DNA can be identified using genetic genealogy databases record even when the database contains as little as 1% of the population (17). Recommendation by SWGDAM (Scientific Working Group on DNA Analysis Methods) has been published (18), stating that IGG technique shall only be used if all searches using conventional typing based on STRs fail. Forensic genetic genealogy is to be utilized for the purposes of investigations associated with violent crimes and identification of unidentified human remains – after fulfilling several country-specific criteria. Despite high utility of FGG, its usage faces certain doze of criticism raised by different groups of practitioners, specialists and members of public. If an individual makes his DNA available for searching within the database, he indirectly makes decision for his relatives, who share part of his genome and so they can get involved in an investigation even if they have never undergone any DNA testing (19). Moreover, data stored by DTC genetic genealogy databases are not a subject to a quality control widely spread within the forensic setting, so the question arises what is the confidence of the results obtained for FGG. However, the microarray method itself has finally been validated using the Scientific Working Group on DNA Analysis Methods (SWGDAM) Validation Guidelines for DNA Analysis Methods and Federal Bureau of Investigation Quality Assurance Standards (20)

Another doubt concerns bias introduced by the ethnic content of genetic genealogy databases. Most of the users are of Western European origin and others are significantly underrepresented, so the chance of success is high only under specific circumstances.

Due to a huge number of markers analyzed, genetic genealogy typically requires certain amount of good quality DNA making it a little cumbersome in the case of crime scene samples. DTC services usually perform SNP typing using high-density microarrays being typically Illumina customized GSA (Global Screening Assays) chips (or Affymetrix Axiom Array) containing primers for over 600K autosomal SNPs typing (together with some sex chromosome and mitochondrial DNA SNPs). According to the manufacturer, in case of the Infinium Global Screening Array Kit, 200ng DNA is required as an input. Analysis of this kind gives no satisfactory results if the quality and/or quantity of DNA is poor, although successful kinship classification from as little as 250pg input DNA using microarrays has been announced (21).

For crime scene samples a different approach that utilizes whole genome sequencing (WGS) may be applied, although a typical WGS procedure requires as much as 1µg of input DNA. The procedure begins with initial whole-genome enrichment step coupled with low-coverage imputation. For 1x coverage data imputation became almost routine for many uses but forensic sample often yield far less coverage – for these a two-step pipeline, so far proven to be very useful for ancient samples, may be applied (22). It has been shown, that imputation from low-coverage whole-genome sequencing coupled with post-imputation filtering for forensic purposes considerably improves the accuracy of kinship inference (23). Nevertheless, whole genome sequencing raises privacy concerns in forensic setting and these should not be ignored.

Another option, especially useful for low quality samples, is the usage of newly developed ForenSeq Kintelligence Kit that targets 10 230 SNPs present also within the DTC genetic genealogy datasets and that requires only 1ng of input DNA. The study by

Peck et al (14) indicates it gives reliable results down to 0,05ng DNA. The SNP set has been carefully selected in order to pick only those that are 1) not known to be linked to any medical condition, 2) are covered by most commonly used microarrays and 3) most informative in terms of genetic relatedness inference. Moreover, owing to small amplicon size of less than 150bp it has already been proven ideal for degraded samples. However, a reduced set of SNPs comprising of approximately 10K polymorphisms will severely restrict relationship inference as it limits discovering matches beyond first cousin level (24,25) and linkage needs to be accounted for. Interpretation in this case is also heavily influenced by the statistical approach applied. It is worth noting, however, that as the ForenSeq Kintelligence Kit manufacturer (Qiagen) is the owner of the GEDMatch PRO, reagents' buyers are eligible for the free access to the database resources. The ForenSeq Kintelligence Kit may be used with MiSeq FGx machine only but the report obtained with Universal Analysis Software (UAS) Forenseq Kintelligence Analysis Module is fully compatible with GEDmatch PRO database format.

In any case quality check may indicate that the data is not good enough to be uploaded to GEDmatch or FamilyTreeDNA. The reason for that is not the DNA fragmentation or minute quantity but rather contamination originating from other person or bacterial source. These heavily hamper obtaining reliable and good quality profiles suitable for subsequent use by the matching algorithms. In general, analysis of mixtures from several contributors is a part of ordinary forensic work. Unfortunately, in case of forensic genetic genealogy not much is still known and reliable scientific evidence is still scarce. Currently, FGG searches require single source DNA, so a sample data needs initial preprocessing before an upload to any database. What is usually done is a deconvolution of a mixture by conditioning on known contributors, by simulation or application of a variety of statistical algorithms. As with other deconvolution applications, the profile generated is somewhat uncertain and the analyst needs to account for that. Recently published method (26) utilizes ForenSeq Kintelligence Kit and produces informative matches for two-contributor DNA mixtures for both major and minor contributors.

Statistics

Identical by descent approach relies on the fact that segments of DNA remain conserved over generations. Therefore one may assume, that the more distant the relatives, the shorter the common stretches of DNA as the growing number of recombination events introduce increasing number of breaks in those segments. The length, the number of shared DNA stretches expressed in cM and the count of shared alleles are the data to be used by various statistical approaches and software applied in FGG searches.

Plausible matches ale calculated using internal DTC databases algorithms for aligning chromosome segments what sometimes is considered a major limitation within the forensic setting (27). Drawing any conclusions requires understanding of the underlying statistics, yet most of the DTC genetic genealogy services use undisclosed proprietary algorithms that aid relatedness inference. Plenty of approaches were described that allow for relationship inference (28–35) and in the papers listed detailed explanations and summaries are given. In any case the goal of the analysis is to determine if a DNA segment shared by individuals is identical by descent (IBD) meaning they have a common ancestor. Two most common approaches include exploratory and likelihood ones – both giving similar results in case of distant relatives (36) with likelihood method falsely including slightly more unrelated individuals. Neither of the approaches provide exact degree of kinship but rather a range of plausible relationships that eventually need to be verified by classical genealogical research. As exploratory approach delivers no likelihood ratio, it cannot be directly presented to the court so its role is limited to investigative leads generation and further conventional STR typing needs to be performed. All dependencies and influence of additional factors including very subtle ones is very straightforwardly summarized in Kling et al (27).

Erturk et al (37) presented a noteworthy approach allowing for much faster processing of the FGG cases. The idea is based on modeling and analysis of the FGG as a stochastic dynamic process where tracking of typing the most recent common ancestor is vital and allows a tenfold reduction of time needed for filling FGG procedures.

Practical use

Forensic genetic genealogy becomes more and more popular among law enforcement agencies in resolving cold cases and bringing back identity to unidentified human remains and has already proven its high success rate. For example, Oregon State Medical Examiner's Office together with Parabon NanoLabs managed to identify 30 out of 43 unidentified human remains using high-resolution SNP microarrays or whole genome sequencing and uploading the results to GEDmatch only (38).

It is estimated over four hundred cases were resolved only within the US with the aid of FGG (and the number is growing each week). Interestingly, the method has also been used for proving innocence of already convicted felons (39,40). So far, as the systematic review by Dowdeswell (39) revealed, FGG has been mostly used for solving crimes involving serial and sexual violence against females. Most of the suspects were of European descent and this should not be surprising as most databases contain mainly data from individuals of Western European origin.

First and by far most prominent use of forensic genetic genealogy is associated with the search of "The Golden State Killer" – a man responsible for several murders and numerous rapes in California, taking place between mid-seventies and eighties (41). Due to the severity of accusations significant resources and efforts were used resulting in pointing over 8000 suspects. In the course of investigation there was a 100% Y-chromosome match found within the Ysearch.org database (permanently closed now). Even though the company (Gene by Gene – owner of the Ysearch.org) was ordered by the court to provide its customers details, the investigation surprisingly stalled as the autosomal STR typing resulted in exclusion. A true breakthrough occurred when the investigators decided to reanalyze one of the crime scene samples (two swabs from a sexual assault case) in order to obtain a dense SNP profile of the perpetrator and uploaded it to the GEDmatch database. As a matter of fact, this first case violated terms of GEDmatch database use as each user had to declare that the DNA record he uploads refers to his own genome (or he "is a legal guardian of the DNA donor or otherwise authorized"). The matter was largely ignored as the perpetrator was finally found. Forensic genetic genealogy search allowed for identification of around 20 individuals being related with the Golden State Killer as third or fourth cousins but a second utilized database (MyHeritage) allowed for generation a much closer hit. The procedure was followed by the intense genealogical research that ended up with identification of a likely suspect being James Joseph DeAngelo, who after STR typing of a tissue sample from his garbage and a swab from his car doorhandle was finally charged and is currently serving 26 life sentences.

Despite the ethical issues raised upon arresting DeAngelo, a significant number of cold cases were re-opened and investigated using FGGS methods. The other outcome of this particular case was a wave of revisions of the privacy policy among direct-to-consumer genetic databases companies and a rapid introduction of restrictions for their use by legal authorities (40).

In 2020 investigation that used extensive FGG searches ended up with a man being sentenced for 897 years in prison for committing a series of mostly sexual assaults that took place between 1991 and 2006. The police managed to link six of the cases to the same perpetrator on the basis of routine DNA testing in 2006, but no hits from the state's criminal offender databases were generated and despite its weight, the case went cold. After DeAngelo conviction investigators decided upon attempt to use IGG in the NorCal rapist case and it turned to be a perfect move. As the perpetrator left plenty of DNA, a full SNP profile has been generated with ease and an IGG specialist managed to build a genealogical tree from the database hits that allowed tracking Roy Charles Waller (42).

One of the oldest cases solved by the use of forensic genetic genealogy refers to the "Boy in the Box". The four-year old boy's body was found in the cardboard box in 1957 and the autopsy has revealed he was malnourished and had suffered physical abuse prior to death. His identity remained unknown although a lot of effort has been made to solve the mystery for years. He was exhumed twice and in 2019 DNA analysis was performed but the DNA quantity and quality was poor. Nevertheless in December 2022 law enforcement agency with the

help from Identifinders International (forensic genetic genealogical company) 65 years after the body was found, was able to identify the victim's name being Joseph Augustus Zarelli. Still, the investigation on the homicide and the boy's death circumstances remains open (43).

It is worth noting, that the founder of the IdentiFinders, Dr. Colleen M. Fitzpatrick has also established the DNA Doe Project – an American nonprofit organization that aids identification of human unidentified remains using forensic genealogy. As for the time writing, they successfully identified 78 persons, including five hits from the beginning of the year (January 2023).

IGG worfklow

Although used for different types of cases, IGG workflow remains quite uniform with only minute modifications depending on the sample type and depth of genealogical insight. Nevertheless the process itself may be considered difficult with some of the problems being unique to FGG only and requires both specific professional skills and lots of determination. Forensic Genetic Genealogy pipeline begins with obtaining good quality SNP profile from forensic sample, both collected from the crime scene as well as from unidentified human remains. As discussed above, various approaches dependent on quality and quantity of DNA might be applied in order to achieve satisfactory final result – a subject's profile. Once this technically demanding part is complete, the profile is uploaded to a specific publicly available personal genomics database (one or more), that allows for the data being used by legal authorities for investigative purposes. Currently, only FamilyTreeDNA, DNASolves and GEDmatch/ GedmatchPRO meet these criteria. If FGG search results in discovering a genetic association between the forensic sample and a database record, DTC service provides the user with a list of associations containing usernames, expected relationship and the amount of DNA expressed in centimorgans (cM) shared by indicated samples/ user data. This step is often performed by an experienced IGG practitioner who can rank the discovered alleged associations from most to least likely on the basis of the amount of DNA shared. Relationships between the subject and the matches are established on the basis of known inheritance patterns and their statistical representation. Genealogical research encompasses usage of publicly accessible data including birth and marriage records, obituaries, census, digitized newspapers, social media and community forums to establish likely family tree with emphasis on finding most recent common ancestor (as well as the descendants later on). The ascendancy search often results in pointing to more than one person, thus several lineages need to be subsequently checked and revised. The process of finding descendants typically ends once a currently living generation is reached to narrow the number of potential leads (44). At this point the results are given to the legal authority that needs to undertake classical investigation – if a link to specific individual is confirmed, a reference sample needs to be legally collected and analyzed with standard short tandem repeat set to assure high confidence of a genetic match.

U.S. Regulations

In the US all actions having association with forensic genetic genealogy and searching (FGGS) that are allowed to be legally undertaken are well covered in the interim policy document published by the United States Department of Justice (DJ) (45). The document is the first ever guideline that balances investigation's requirements and data privacy and safety. The guidance given emphasizes tight collaboration between agencies, prosecutors, investigators and laboratory personnel having first and foremost privacy interests in mind. It applies to both investigations carried out or funded by the US Department of Justice and investigations conducted by employees or contractors, based on leads obtained with FGGS and following genealogical research.

As a rule, criminal investigations utilize data collected by Combined DNA Index System (CODIS) – where DNA profiles obtained from known offenders (as well as missing persons, unidentified human remains and samples collected from crime scenes) comprising of 13-20 STRs are deposited. If a search of forensic profile against the database generates a match, a lead is generated. If not, investigators may use FGGS investigative technique, assuring constant data privacy at every step of investigation. Not in every case FGGS investigation may be utilized – currently it is reserved for the most violent crimes (sex-

ual assaults and homicides), attempts to commit violent crimes (with separate prosecutor's authorization if public or national safety is threatened) and identification of human remains (of a suspected homicide victim). As a matter of fact it opens possibility to use FGGS legally if any threat to public is suspected – and it is if a perpetrator walks free and may potentially commit another acts of crime. Before any FGGS procedure begins, both investigative agency and the prosecutor need to agree that all other options and reasonable scientific alternatives were checked and that the sample quality is sufficient prior to decision on FGGS implementation. The policy states, that police or any other agency has to get informed consent from FGG reference sample donors, unless "case-specific circumstances provide reasonable grounds to believe that this request would compromise the integrity of the investigation". Attempts of gathering any additional data on the basis of further genetic testing including medical conditions or genetic predispositions including psychological traits are forbidden. Biological material is usually outsourced to commercial laboratories as DJ does not handle high-resolution DNA typing so the procedure needs to be secured with additional contracts to assure privacy and security of all the information gathered and generated by these laboratories. Moreover, none of the information obtained using FGGS is uploaded to CODIS and cannot be stored within. Information provided by the DTC service on genetic associations has to be confirmed with traditional genealogical research and investigative research – during this step it is considered only as a lead. In any case it is not permitted to arrest any individual on the basis of FGGS testing only. Traditional STR typing absolutely needs to be performed for both the forensic sample and the suspect and uploaded for the comparison purposes to CODIS database. Classical investigative work needs to be done in order to confirm FGGS findings. All third-party reference samples (both extracts and associated data) have to be destroyed upon criminal prosecution as well in case when no criminal charges are filled.

In 2021 the state of Maryland passed the first law in the US (and in the world) that regulates the use of FGG. It encompassed seven important features: FGG may only be used after judicial authorization, judicial authorization is granted only if all other methods fail, only the databases getting affirmative consent and providing explicit notice to users about law enforcement potential use of the data may be utilized for FGGS, informed consent from all non-suspects need to be obtained for any DNA profiling, annual public report on FGG usage need to be prepared, FGG statue violations are to be penalized and laboratories performing typing used for FGG must be licensed.

It is worth noting that the same year a call for responsible genetic genealogy has been published in a Science editorial signed by a chief biometric scientist of FBI laboratory (46).

Law enforcement agencies in the US usually seek for support from companies specializing in genetic genealogy for forensic purposes. Two main players are Parabon NanoLabs and Bode Technology – both offering microarray SNP testing and WGS together with IGG specialist service and claiming high success rate reaching or exceeding 80%.

European Perspective

Not many countries utilize FGG approach and the factors responsible are limited access to appropriate DNA records and legal challenges (being mainly lack of any regulations).

The first National DNA Database was launched in United Kingdom in 1995, so one may expect this is the country that will remain a European leader in applying further identification advances. Meanwhile, the truth is somewhat different as legal regulations hold back the FGG use. The Biometrics and Forensics Ethics Group (UK) stated its opinion on feasibility of the use of such methods (47). The report points to several ethics and legal challenges that need to be addressed before any FGG investigation takes place. On the contrary to the US, where the police may actually collect any item that probably bears suspect's DNA as a reference sample, this kind of a proof collection would not necessarily be acceptable within the UK. Also a strength and quality of the national DNA database is highlighted, questioning advantages of FGG strategy based on US curated databases over "traditional" familial searching. In fact, the UK National DNA Database of England and Wales contains the profiles of about 6 million citizens corresponding to over 9% of the UK population

(48). The Group points to doubtful community support and virtual lack of regulations as well as ethical, legal and data safety issues that need to be solved. A general reception of the report is that it advises against FGG usage.

On the other hand, experimental use of FGG searches shows, that relatively large proportion of the UK citizens provided their DNA to the genealogical testing companies and the efficacy of the searches may be similar to these obtained in the US (49) as four out of ten volunteers were identified correctly even though one was not of European descent.

The only European country so far that decided to apply FGG in a case pilot study was Sweden. Interestingly, Swedish profiles account for about 1% of GEDmatch records (27). The FGG search had been applied to one of the largest investigations ever held in Sweden (50). It related to a double murder of an eight-year-old boy and a 56-year-old woman who were found being stabbed to death in October 2004 in Linköping. During the course of investigation over 6000 men were screened using standard STR typing but no matches were found and no leads generated. Upon decision of forensic genetic genealogy use, whole genome sequencing was performed for crime scene samples. Further search of genetic genealogy databases resulted in two hits, one of which matched the profile found on the crime scene as confirmed with STR typing. A fruitful cooperation between various departments of Swedish Police Authority and the National Board of Forensic Medicine as well as a contracted laboratory and a FGG specialist resulted in resolving the homicide case after fifteen years of unsuccessful investigation. This unprecedented outcome resulted in a decision of the Swedish Police Authority on implementation and more common usage of FGG (51).

Misconceptions and Future Perspectives

There are several misconceptions that are sustained by the media and repeated by general community that need to be clarified as they seem confusing and lead to misunderstanding of basic concepts of the IFF searches (52). First of all the investigators (both the law enforcement agencies and the IGG professionals hired by them) have no special access to direct-to-consumer databases. Apart from the fact that only GEDmatch and FTDNA allow for its data to be used by legal authorities (after user's informed consent), investigators are given same access to data as regular users, so they do not obtain users' raw data and cannot directly download any further information beside the ranked links of the users and the subject (same as regular users do). Any participant may, in turn, use chromosome browser tool and at least partially infer DNA sequence of other users if a segment match between them was found. With a little bioinformatic knowledge individual SNP profile may be extracted, but this option is not reserved for law enforcement agencies and can be utilized by virtually any genetic genealogy enthusiast who uses DTC services.

Another misconception refers to the idea of collection and testing of more innocent people that it would normally be done in "classical investigation". In fact, when a DNA is found on the crime scene, it is inevitable that many suspects need to be excluded on the basis of genetic testing and so many people may undergo STR typing. What is more, FGG may dramatically narrow the number of suspects so in the end the number of individuals tested may be even smaller than it would be if no genetic genealogy investigation was applied.

As for urgent actions that need to be undertaken in the near future is the necessity of a definite addressing of several issues collectively referring to standardization and certification (53). There is an ongoing debate on how to resolve data privacy questions. As the IGG practitioner usually employs methodology similar to the one used in classical genealogy research, he also needs to dig into personal data including both general demographic data as well as information available from social media and community forums. These often include daily routines, photographs, relationships – and it would not raise major concerns if it was associated to suspects only. In the case of FGG however, all relatives undergo this one of a kind digital research and from this point of view their privacy may certainly be violated. Especially that these individuals have absolutely no clue they are being a part of an ongoing investigation (40,54). It should be paramount to FGG investigations to assure reasonable level of privacy to all innocent individuals whose data is relevant and included in the research.

Moreover, the power of FGG searches rely on the number of individuals that trust in the justice system and are willing to intentionally give access to their DNA profiles to the legal authorities and cooperating parties. Therefore IGG practitioners together with legal authorities need to build and sustain a public trust in both the work they do (to serve public safety) and the methods they use to achieve transparency and absolute flawlessness. Therefore no misuses and ignoring terms of use of the public databases may take place – even if these are extremely rare, they undermine justice system and makes it being perceived as operating outside of its own legal standards. The consequences are easy to be foreseen as the members of general public will try to protect their data by removal or hiding information from social media for instance thus severely hindering FGGS.

Third concern refers to skills and education needed to become an FGG specialist that are not currently systematized. There are courses and trainings that end with obtaining FGG professional certificate but these lack standardization. If a skilled genetic genealogist joins this track of education, there is a huge chance he will become a good FGG specialist if he manages to stick to general work ethics and data privacy regulations. Yet these courses are open to anyone and if no standards are directly applied to the education and certification itself, the quality of the FGG specialist's service may only be assessed by informal personal communications from other skilled professionals (one that solves numerous cases having many references), that practically have no impact at all. FGG specialist should be considered as a high stake professional, so to ensure his accountability some standards and certifications need to be introduced. Otherwise FGG specialists proficiency may be questioned as no sanctions for improper practices or falling below ethical or professional standards may be applied. What is more, given no certification is available, many proficient, talented and simply lesser known FGG specialists may have trouble being get to known by the agencies as these in turn have absolutely no means to objectively verify FGG specialists' skills.

Taken above into consideration, a creation on Board of Certification for Investigative Genetic Genealogy has been announced (55). The non-profit board consists of six individuals some of which are FGG professionals and aims at developing certification processes, exams and standards similar to those already established in 2019 by the Board of Certification of Genealogists for genealogical research.

Conclusion

As the forensic genetic genealogy is a new field, it suffers from lack of regulations, somewhat unclear rules and personal privacy data concerns. Even though, it should be noted that the FGG search is intended for generating investigative leads only, not conviction in any case. Nevertheless it has already proven its extreme usefulness in solving cold cases and bringing back identities to people whose fate has remained unknown for tens of years. In the right, skilled and certified hands of an FGG professional who works along ethical guidelines, it may become an irreplaceable forensic tool that stands in line with other groundbreaking techniques like elementary STR typing.

References

1. Williams R, Wienroth M. Social and ethical aspects of forensic genetics: A critical review. Forensic Sci Rev. 2017 Jul;29(2):145–69.
2. Zieger M. Forensic DNA phenotyping in Europe: How far may it go? J Law Biosci. 2022 Jul 1;9(2):lsac024.
3. SR 363 – Federal Act of 20 June 2003 on the Use of DNA Profiles in Criminal Proceedings and for Identifying Unidentified or Missing Persons (DNA Profiles Act) [Internet]. [cited 2023 Mar 29]. Available from: https://www.fedlex.admin.ch/eli/cc/2004/811/en
4. Glynn CL. Bridging Disciplines to Form a New One: The Emergence of Forensic Genetic Genealogy. Genes. 2022 Aug 1;13(8):1381.
5. Edge MD, Coop G. Donnelly (1983) and the limits of genetic genealogy. Theor Popul Biol. 2020 Jun;133:23–4.
6. Huff CD, Witherspoon DJ, Simonson TS, Xing J, Watkins WS, Zhang Y, et al. Maximum-likelihood estimation of recent shared ancestry (ERSA). Genome Res. 2011 May;21(5):768–74.
7. Henn BM, Hon L, Macpherson JM, Eriksson N, Saxonov S, Pe'er I, et al. Cryptic Distant Relatives Are Common in Both Isolated and Cosmopolitan Genetic Samples. PLOS ONE. 2012 Apr 3;7(4):e34267.

8. Bettinger BT. The Shared cM Project Version 4.0 (March 2020). 2020;
9. Mateen RM, Sabar MF, Hussain S, Parveen R, Hussain M. Familial DNA analysis and criminal investigation: Usage, downsides and privacy concerns. Forensic Sci Int. 2021 Jan;318:110576.
10. Fortier AL, Kim J, Rosenberg NA. Human-Genetic Ancestry Inference and False Positives in Forensic Familial Searching. G3 GenesGenomesGenetics. 2020 Jun 25;10(8):2893–902.
11. Laurent FX, Fischer A, Oldt RF, Kanthaswamy S, Buckleton JS, Hitchin S. Streamlining the decision-making process for international DNA kinship matching using Worldwide allele frequencies and tailored cutoff log10LR thresholds. Forensic Sci Int Genet. 2022 Mar 1;57:102634.
12. Bowen S, Khoury M. Consumer genetic testing is booming: but what are the benefits and harms to individuals and populations? CDC, Cent DisCont rol: Genom Precis Health [Internet]. 2018 [cited 2023 Jan 31]; Available from: https://blogs.cdc.gov/genomics/2018/06/12/consumer-genetic-testing/
13. Autosomal DNA testing comparison chart [Internet]. Available from: https://isogg.org/wiki/Autosomal_DNA_testing_comparison_chart
14. Peck MA, Koeppel AF, Gorden EM, Bouchet JL, Heaton MC, Russell DA, et al. Internal Validation of the ForenSeq Kintelligence Kit for Application to Forensic Genetic Genealogy. Forensic Genomics. 2022 Dec;2(4):103–14.
15. Transparency report [Internet]. 2023 [cited 2023 Feb 3]. Available from: https://www.23andme.com/en-int/transparency-report/
16. Ball CA, Barber MJ, Byrnes J, Carbonetto P, Curtis RE, Granka JM, et al. Discovering genetic matches across a massive, expanding genetic database. 2020 [cited 2023 Feb 3]; Available from: https://www.ancestrycdn.com/support/us/2020/08/matching-whitepaper.pdf
17. Ellenbogen P, Narayanan A. Identification of Anonymous DNA Using Genealogical Triangulation [Internet]. bioRxiv; 2019 [cited 2023 Feb 3]. p. 531269. Available from: https://www.biorxiv.org/content/10.1101/531269v1
18. SWGDAM, Overview of investigative genetic genealogy [Internet]. 2020 [cited 2023 Jan 23]. Available from: https://www.swgdam.org/publications
19. Kennett D. Using genetic genealogy databases in missing persons cases and to develop suspect leads in violent crimes. Forensic Sci Int. 2019 Aug 1;301:107–17.
20. Russell DA, Gorden EM, Peck MA, Neal CM, Heaton MC, Bouchet JL, et al. Developmental Validation of the Illumina Infinium Assay Using the Global Screening Array on the iScan System for Use in Forensic Laboratories. Forensic Genomics. 2023 Mar;3(1):15–24.
21. de Vries JH, Kling D, Vidaki A, Arp P, Kalamara V, Verbiest MMPJ, 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 Jan 1;56:102625.
22. Hui R, D'Atanasio E, Cassidy LM, Scheib CL, Kivisild T. Evaluating genotype imputation pipeline for ultra-low coverage ancient genomes. Sci Rep. 2020 Oct 29;10(1):18542.
23. Nagraj V p., Scholz M, Jessa S, Ge J, Huang M, Woerner AE, et al. Relationship Inference with Low-Coverage Whole Genome Sequencing on Forensic Samples. Forensic Genomics. 2022 Sep;2(3):81–91.
24. Kling D. On the use of dense sets of SNP markers and their potential in relationship inference. Forensic Sci Int Genet. 2019 Mar 1;39:19–31.
25. Skare Ø, Sheehan N, Egeland T. Identification of distant family relationships. Bioinforma Oxf Engl. 2009 Sep 15;25(18):2376–82.
26. Mitchell R, Enke S, Eskey K, Ferguson T, Just R. A method to enable forensic genetic genealogy investigations from DNA mixtures. Forensic Sci Int Genet Suppl Ser. 2022 Dec;8:159–61.
27. Kling D, Phillips C, Kennett D, Tillmar A. Investigative genetic genealogy: Current methods, knowledge and practice. Forensic Sci Int Genet. 2021 May 1;52:102474.
28. Browning SR, Browning BL. High-resolution detection of identity by descent in unrelated individuals. Am J Hum Genet. 2010 Apr 9;86(4):526–39.
29. Kong A, Masson G, Frigge ML, Gylfason A, Zusmanovich P, Thorleifsson G, et al. Detection of sharing by descent, long-range phasing and haplotype imputation. Nat Genet. 2008 Sep;40(9):1068–75.
30. Li H, Glusman G, Huff C, Caballero J, Roach JC. Accurate and Robust Prediction of Genetic Relationship from Whole-Genome Sequences. PLoS ONE [Internet]. 2014 [cited 2023 Feb 1];9(2). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3938395/
31. Manichaikul A, Mychaleckyj JC, Rich SS, Daly K, Sale M, Chen WM. Robust relationship inference in genome-wide association studies. Bioinforma Oxf Engl. 2010 Nov 15;26(22):2867–73.
32. Weir BS, Anderson AD, Hepler AB. Genetic relatedness analysis: modern data and new challenges. Nat Rev Genet. 2006 Oct;7(10):771–80.
33. McPeek MS, Sun L. Statistical tests for detection of misspecified relationships by use of genome-screen data. Am J Hum Genet. 2000 Mar;66(3):1076–94.
34. Thompson EA. Statistical Inference from Genetic Data on Pedigrees. NSF-CBMS Reg Conf Ser Probab Stat. 2000;6:i–169.
35. Thompson EA. Identity by descent: variation in meiosis, across genomes, and in populations. Genetics. 2013 Jun;194(2):301–26.

36. Kling D, Tillmar A. Forensic genealogy—A comparison of methods to infer distant relationships based on dense SNP data. Forensic Sci Int Genet. 2019 Sep 1;42:113–24.
37. Ertürk MS, Fitzpatrick C, Press M, Wein LM. Analysis of the genealogy process in forensic genetic genealogy. J Forensic Sci. 2022 Nov;67(6):2218–29.
38. Cady J, Greytak EM. Whole-genome sequencing of degraded DNA for investigative genetic genealogy. Forensic Sci Int Genet Suppl Ser. 2022 Dec;8:20–2.
39. Dowdeswell TL. Forensic genetic genealogy: A profile of cases solved. Forensic Sci Int Genet. 2022 May 1;58:102679.
40. Ram N, Roberts JL. Forensic genealogy and the power of defaults. Nat Biotechnol. 2019 Jul;37(7):707–8.
41. Wickenheiser RA. Forensic genealogy, bioethics and the Golden State Killer case. Forensic Sci Int Synergy. 2019;1:114–25.
42. Murray K. "NorCal" rapist Roy Waller sentenced to 897 years [Internet]. CNN. 2020 [cited 2023 Feb 3]. Available from: https://www.cnn.com/2020/12/18/us/norcal-rapist-roy-waller-sentenced-trnd/index.html
43. Boy in the box philadelphia homicide. [cited 2023 Jan 23]; Available from: https://www.nytimes.com/2022/12/08/us/boy-in-the-box-philadelphia-homicide.html
44. Greytak EM, Moore C, Armentrout SL. Genetic genealogy for cold case and active investigations. Forensic Sci Int. 2019 Jun 1;299:103–13.
45. United States Department of Justice Interim Policy: Forensic Genetic Genealogical DNA Analysis and Searching.
46. Callaghan TF. Responsible genetic genealogy. Science. 2019 Oct 11;366(6462):155.
47. Should we be making use of genetic genealogy to assist in solving crime? A report on the feasibility of such methods in the UK.
48. Maguire CN, McCallum LA, Storey C, Whitaker JP. Familial searching: A specialist forensic DNA profiling service utilising the National DNA Database® to identify unknown offenders via their relatives—The UK experience. Forensic Sci Int Genet. 2014 Jan 1;8(1):1–9.
49. Thomson J, Clayton T, Cleary J, Gleeson M, Kennett D, Leonard M, et al. The effectiveness of forensic genealogy techniques in The united kingdom – an experimental assessment. Forensic Sci Int Genet Suppl Ser. 2019 Dec 1;7(1):765–7.
50. Tillmar A, Fagerholm SA, 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 Jul 1;53:102525.
51. The Swedish Police Authority, National Forensic Centre. Forensic DNA Traces and Genealogy. Use of investigative genetic genealogy in criminal investigations [Internet]. 2021 [cited 2023 Feb 2]. Available from: https://polisen.se/SysSiteAssets/dokument/forensik/forensic-dna-traces-and-genealogy.pdf.
52. Guerrini CJ, Wickenheiser RA, Bettinger B, McGuire AL, Fullerton SM. Four misconceptions about investigative genetic genealogy. J Law Biosci. 2021 Apr 10;8(1):lsab001.
53. McEwen J, Pino N, Raphael A, Renna K, Boyer J, Brody LC. Investigative Genetic Genealogy: Ethical, Legal, and Social Issues and Directions for Future Research. Forensic Genomics. 2021 Sep;1(3):91–8.
54. Guerrini CJ, Robinson JO, Petersen D, McGuire AL. Should police have access to genetic genealogy databases? Capturing the Golden State Killer and other criminals using a controversial new forensic technique. PLoS Biol [Internet]. 2018 Oct [cited 2023 Feb 1];16(10). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6168121/
55. Gurney D, Press M, Moore C, Rolnick CI, Hochreiter A, Bossert BL. The need for standards and certification for investigative genetic genealogy, and a notice of action. Forensic Sci Int. 2022 Dec;341:111495.

ORCID

Urszula Rogalla-Ładniak – 0000-0002-2778-9844

AUTOR DO KORESPONDENCJI
CORRESPONDING AUTHOR

Urszula Rogalla-Ładniak
Department of Forensic Medicine, Collegium Medicum in Bydgoszcz, Nicolaus Copernicus University in Toruń
e-mail: urszula.rogalla@cm.umk.pl