



Forensic DNA traces and genealogy

Use of investigative genetic genealogy in criminal investigations

Authors

Siri Aili Fagerholm (NFC)

Andreas Tillmar (RMV)

Jan Staaf (region Öst)

Mikael Rying (Noa)

Ricky Ansell (NFC)

Publisher

Swedish National Forensic Centre – NFC

SE-581 94 Linköping

Phone: +46 10 562 80 20

E-mail: registrator.nfc@polisen.se

www.nfc.polisen.se

Polismyndigheten, November 2021

Dnr A594.805/2021

Summary

In this report a new and most powerful tool for Swedish criminal investigations and forensic science is presented. With the support of legal inquiries a pilot case study has been successfully completed and with the use of the tool *Investigative Genetic Genealogy* a serious crime has been solved.

Two criminal cases were planned to be included in the pilot case study but as one of them was solved with familial searches only one case was later used, the unsolved double murder at Åsgatan, Linköping, from October 2004. In 2019 the investigation of the murders had become the second largest ever in Sweden. In the pilot, DNA traces were analysed and SNP datasets produced that were searched in genetic genealogy databases with resulting list of matches. Matches were then evaluated by investigative genetic genealogy. In June 2020, more than 15 years after the crimes were committed; one person was arrested for the murders. Upon comparative routine STR-profiling the person was confirmed to match the crime scene samples. The pilot case study was the product of a successful cooperation between different parts of the Swedish Police Authority, including the Legal Affairs Department, region Öst and the National Forensic Centre, and expertise from the National Board of Forensic Medicine, an external laboratory and a contracted genealogist.

Herein, an overview of investigative genetic genealogy is presented. In addition, the case study is described in some detail as well as experiences gained and conclusions drawn. The Swedish Police Authority aim to implement and use this novel and valuable tool in future criminal investigations and the ongoing work is described, including an overview of legal settings and establishment of national guidelines.

Contents

- SUMMARY I**
- CONTENTS..... III**
- GLOSSARY OF TERMS AND ABBREVIATIONS V**

- 1 INTRODUCTION 1**
 - 1.1 Background 1
 - 1.2 Aim 2

- 2 OVERVIEW OF INVESTIGATIVE GENETIC GENEALOGY 2**
 - 2.1 Genetic genealogy databases 4

- 3 PILOT CASE STUDY 7**
 - 3.1 Initial process description 7
 - 3.2 Legal inquiries 7
 - 3.3 Aim of the pilot case study 11
 - 3.4 Criminal cases used in the pilot study..... 11
 - 3.5 Data processing and searches in genealogy databases 12
 - 3.6 Genetic genealogy in the pilot case..... 16
 - 3.7 Reflections from the pilot case study 18
 - 3.8 Summary of the pilot case study 22

- 4 IMPLEMENTATION OF THE IGG METHOD 24**
 - 4.1 National guidelines 25
 - 4.2 National inventory of criminal cases 25
 - 4.3 Legal considerations 25

- 5 FUTURE PERSPECTIVES..... 28**

- 6 CONCLUDING REMARKS 29**

- 7 ACKNOWLEDGEMENTS 30**

- 8 REFERENCES 31**

Glossary of terms and abbreviations

Meanings and explanations of terms and abbreviations used in this report.

DTC	Abbreviation for <i>Direct-to-consumer</i> , used for companies offering home delivered DNA tests including FTDNA, AncestryDNA, MyHeritage.
Familial search	Searches of a STR DNA profile in the authority's national DNA database. The search results in a list of candidates that either could have a parent-child relation or a sibling relation with the donor of the crime scene trace.
FTDNA	Abbreviation for the commercial genetic testing company FamilyTree DNA.
Genetic genealogy	The term includes searches, comparisons and other measures taken in DNA-based genealogy databases in combination with traditional genealogy, aiming at clarifying close and/or distant family relationships.
GEDmatch	Online database that offers upload of genetic data from a range of DTC companies and user tools for genetic genealogy purposes.
Investigative genetic genealogy	The use of genetic genealogy in criminal investigations, abbreviation IGG. Also named <i>forensic genetic genealogy</i> .
LE	Law Enforcement
mtDNA	Mitochondrial DNA
NFC	Swedish National Forensic Centre
RMV	National Board of Forensic Medicine, Sweden
SNP	Abbreviation for <i>Single Nucleotide Polymorphism</i> . A SNP is a variation at a specific position in the genome where at least 1% of the population display the nucleotide-variation (polymorphism).
SPA	Swedish Police Authority
STR	Abbreviation for <i>Short Tandem Repeat</i> . The number of repetitions of a specific sequence in the genome. At present, NFC analyses 15 STR markers and one additional marker for gender determination to render a STR-based DNA profile that can be used in everyday work at NFC.
WGS	Whole Genome Sequencing
Y-DNA	Y-chromosomal DNA

1 Introduction

1.1 Background

Genealogy based on DNA data or *genetic genealogy*, comprises the use of DNA testing and analysis in combination with traditional genealogy work. DNA tests offered by direct-to-consumer (DTC) companies are primarily used by people who want to know more about their genetic origin or perform genealogy work from DNA data as a complement to traditional genealogy work.

Commercially available genetic genealogy databases and DTC companies have offered their services to the public during some years now. Two of the largest DTC companies, 23andMe and FamilyTreeDNA (FTDNA) started to offer cheaper tests of autosomal SNPs (Single Nucleotide Polymorphisms) to the wider public in 2009 and 2010 respectively. They were followed in 2012 by AncestryDNA and MyHeritage in 2016 [1]. A number of years would pass though, before such databases were first used by law enforcement (LE) to obtain investigative leads in both cold criminal cases or in the purpose of identifying deceased [2,3].

In spring 2018, "the Golden State Killer", a Californian serial killer and rapist in the 1970-1980's, was reported by US media to finally have been arrested. The perpetrator was identified with the use of comparative DNA searches in a commercially available genealogy database (GEDmatch) and extensive genealogy work¹.

Some weeks before, the DNA Doe Project made public that genetic genealogy had been used to identify a 21-year-old woman, the so called "Buckskin Girl", that was found dead 1981 in Ohio, US². The DNA Doe Project is a non-profit organization that ever since has used genealogy databases for the identification of human remains.

Following these successfully solved cases genetic genealogy has been used by LE, primarily in the US, to identify human remains or suspects of a crime. More than 200 cold cases have been solved, where the majority so far concerns identification cases [4].

A number of scientific papers on the use of investigative genetic genealogy have been published since, that covers methodological issues as well as ethical and judicial aspects [e.g. 3-8]. In LE and the forensic society the terms *investigative genetic genealogy* or *forensic genetic genealogy* have been used interchangeably, even if the first term has been used most frequently and with the acronym IGG.

In September 2019 the American Department of Justice published an interim policy on the use of genetic genealogy for forensic purposes [9,10]. The policy covers when and how the method should be used, its limitations, case criteria (unsolved violent crime with a candidate forensic sample from a putative perpetrator or unidentified human remains of a suspected murder victim), investigative caution and data handling. In February 2020 the Scientific Working Group on DNA Analysis Methods³ (SWGDM) published recommendations on the use of IGG [11]. In the recommendations the method

¹ The scientific journal Science ranked the use of forensic genealogy top ten out of the most important scientific breakthroughs of the year 2018, <https://vis.sciencemag.org/breakthrough2018/finalists/#forensic-genealogy>.

² <https://dnadoeproject.org/case/buckskin-girl/>

³ SWGDAM (Scientific Working Group on DNA Analysis Methods) is a scientific working group with members from Federal, State and Local forensic laboratories and FBI. SWGDAM address issues that are of focus for the DNA community such as quality assurance standards, forensic biology methods, training and research. The organization also provides recommendations and conduct research to develop and/or validate forensic biology methods. Source: swgdam.org/about-us

is described, including differences between forensic familial searches and IGG, as well as a number of considerations for LE⁴.

1.2 Aim

The aim of this report is threefold:

- to give an overview of investigative genetic genealogy including available databases
- to give an overview of the work performed in a pilot case study in which the IGG method was used as well as to summarize reflections and considerations
- to present ongoing work for the implementation of the IGG method within the Swedish Police.

2 Overview of investigative genetic genealogy

In the following section a general workflow of genetic genealogy as used in the IGG method (primarily using autosomal DNA) is described.

At the Swedish National Forensic Centre (NFC) autosomal STR (Short Tandem Repeat) genotyping is performed routinely on crime scene samples. In STR genotyping repetitions of DNA at 15 different sites in the genome are analysed (plus a gender marker), ending up with a number series referred to as a DNA profile. The DNA profile can then successfully be used for comparison against known individuals and in familial searches to associate first degree relatives (although not always informative enough). However, when it comes to more distant relationships far more information is needed, leading to the use of SNP datasets in genetic genealogy work. These datasets can for example be produced by using microarrays for hundreds of thousands of SNPs or using Whole Genome Sequencing (WGS) from which raw data can be extracted and processed to a usable SNP dataset format.

The basis of genetic genealogy is the knowledge of how DNA is inherited through generations [8]. Autosomal DNA is inherited from both parents, and a child will inherit half of its DNA from its mother and father respectively, see Table 1. Because of a mechanism called recombination, an exchange of DNA between chromosome pairs or different regions on the same chromosome can occur, and DNA will not be inherited intact from a parent to a child. Recombination events will result in that siblings will inherit slightly different settings of DNA. The same parts, or segments, of DNA will thus not be inherited by all relatives to a common ancestor. In general, a child will inherit approximately 25% of its four grandparents DNA respectively, and approximately 12.5% of the great-grandparents DNA. The amount of DNA shared will decrease rapidly with increased distant relationships. For third cousins shared DNA will only constitute 0.781% of the total amount. These numbers are averages and the actual numbers vary in each case. The amount of shared DNA is measured in centimorgan (cM)⁵.

This “dilution” of shared DNA between generations means that extensive information on SNP variants is needed to be able to identify kinships. The genetic family tree that can be identified will be smaller than the true family tree since after some generations some relatives will not carry any DNA from their distant ancestors. In that case genealogy using other sources of information can potentially fill in the gaps.

⁴ Recommendations and considerations that can be found in both these reports are very much in line with the considerations and proposed method at the time already stated in the pilot legal inquiries.

⁵ 1cM corresponds to a recombination frequency of 1% for a segment. Single nucleotides that are positioned close in a segment has a greater probability to be inherited together compared to nucleotides that are far apart on a chromosome. With time, over generations, recombination events will break apart increasingly longer segments of shared DNA on a chromosome.

Table 1. Overview of different relationships, amount of shared DNA and kinship.

Relatives*	Amount of shared DNA	Shared cM (average)**	Kinship**
Parent/child	1/2	3400	
Sibling	1/2 (50%)	2613	
Half sibling, grandparents, aunt/uncle, niece	1/4 (25%)	1759	
1 st cousin, great-grandparents	1/8 (12.5%)	866	1c
2 nd cousin	1/32 (3.125%)	229	2c
3 rd cousin	1/128 (0.781%)	73	3c
4 th cousin	1/512 (0.195%)	35	4c
5 th cousin	1/2048 (<0.05%)	25	5c

* Some categories represent several individuals in a family tree, examples on relationships are given.

** Adapted from <https://thegeneticgenealogist.com/2020/03/27/version-4-0-march-2020-update-to-the-shared-cm-project/>.

The IGG method is schematically described in Figure 1. When searches are made in databases accepting LE, matching segments of SNPs are compared [1]. In the list of matches potential relatives will be presented as well as their estimated kinship together with their longest shared segment and the total number of cM shared. The kinship is presented as e.g. “1c”, first cousin (which could correspond to both a cousin and a great-grandfather). User contact details are also presented though these can be difficult for the genealogist to use if only abbreviations for the person’s name or pseudonyms are used.

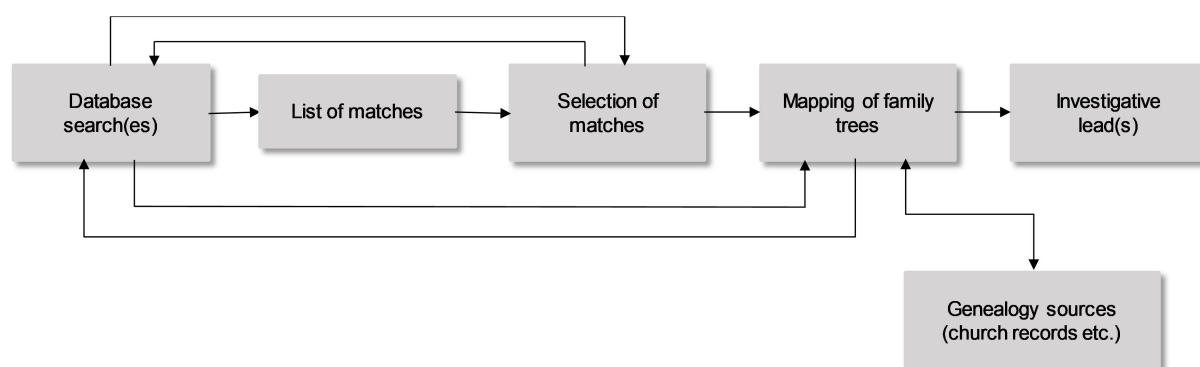


Figure 1. Overview of the IGG process. At first, following data upload and acceptance, database searches are being made and match lists generated. Matches are evaluated by the genealogist and some of the top matches are selected for further evaluation. Using these matches and genealogy sources the genealogist starts to map family trees backward in time. When common ancestral couples are identified family trees can be mapped forward in time to present day. Taken into account expected age, gender and geographic belonging of the DNA crime scene sample donor a person of interest can be identified, and the process has thus generated an investigative lead.

From the list of matches the genealogist will look for clusters of related matches and map the relationships in the family trees back in time to find common ancestors. When common ancestors have been found the genealogist build the trees until present day, which can be a tedious and sometimes unfruitful work. If different clusters are found the genealogist can try and find intersections (triangulation) of the clusters and continue to build the family tree until common candidate(s) can be identified. Several hundred to thousands of people may need to be considered in the process. Considering that in general an individual has 1.000 fourth cousins and 5.000 fifth cousins a genealogist will prefer to focus on third degree relatives or closer - if found - to minimize the workload [12]. The time needed can vary from a few hours to hundreds of hours. With the use of so called “target testing”, where possible closer relatives (e.g. individuals from the specific region of interest) are asked to take a test and aid in the investigation, family trees can be confirmed and substantially narrow the searches and efforts needed [1].

Finally, the genealogy research can produce candidates, so called *investigative leads*, to the investigators. Candidates that can be inquired by the police, swabbed for DNA and the DNA profile (STR markers) compared to the crime scene sample confirming a match or mismatch⁶. If several candidates are given, further investigative information such as the gender, age and likely geographic belonging/location of interest can narrow down the candidates relevant for DNA testing.

Except for autosomal DNA data the genetic genealogist can use Y-chromosomal DNA (Y-DNA) and/or mitochondrial DNA (mtDNA) analysis data. The first commercial tests on Y-DNA and mtDNA were launched already in 2000 [7]. Due to their paternal and maternal inheritance respectively, this type of data may imply distant relationships, facilitate construction of family trees or indicate a geographic region.

2.1 Genetic genealogy databases

A number of genealogy databases can be used by a private customer, though for LE only a few are available. In addition to the publically available personal genomics database GEDmatch and the DTC genetic testing company FTDNA, a smaller database named DNASolves accepts LE cases. DNASolves is owned by the private forensic laboratory Othram, US, and is focused on providing IGG services to LE in human identification investigations. Their services were not used in the pilot and will not be described further in the following. Other DTC companies such as AncestryDNA, 23AndMe and MyHeritage do not allow upload of data and comparative searches from LE⁷ [13,14].

GEDmatch

GEDmatch is an online database offering genealogy services to the public and was at the time for the pilot case study based in Lake Woth, Florida.

In spring 2018 “*the Golden State Killer*” was arrested due to IGG work using GEDmatch, which drew a lot of attention to the company. Also at that time, GDPR regulations became effective in Europe. GEDmatch then re-formulated the user terms, making them stricter. User terms were changed once more in May 2019 following strong and extensive criticism from influencers and other public interest groups within the genetic genealogy community [15]. This was the result of the use of the database by the company Parabon NanoLabs that had been allowed to use it for a case that was neither a murder nor a sexual assault (the only crime categories covered by the user terms) but a violent assault on an elderly women in Utah, US. IGG resulted in the arrest of a 17 year old boy that was later sentenced for the assault. GEDmatch’s exception sparked a vivid debate resulting in an adjustment of the user terms,

⁶ If the biological father of a person is not the one officially named it may not be possible to find the sought after person.

⁷ For example in MyHeritage’s privacy policy it its stated that “MyHeritage prohibits law enforcement use of its DNA Services” and “We will not provide information to law enforcement unless required by a valid court order or subpoena for genetic information”. <https://www.myheritage.se/FP/Company/popup-privacy-policy.php?dispLang=EN>, 2021-06-23.

including the allowed searchable crime categories. Ever since, all users must agree to LE matching, a so called “*opt-in*” or choose “*opt-out*”. Also, existing users from the EU were at the time automatically assigned to “*opt-out*” and had to actively change their status to “*opt-in*”. Notable is that the default setting as of today is “*opt-in*”. However, in a day, these changes dramatically made the searchable numbers shrink from around 1.3 million users to considerably less. Numbers of “*opt-in*” users has increased steadily since and is now reaching approximately 325.000 users [16]. For LE the usability of the database was of course very much reduced, while from a data impact assessment view the integrity of the individual user was significantly improved.

GEDmatch was acquired by the company Verogen in December 2019. Verogen, a San Diego-based company founded by Illumina, sells products and services to forensic laboratories. In connection to the acquisition, GEDmatch once more updated its user terms, as specified in “*Terms of Service and Privacy Policy*”⁸. The policy covers data protection terms and different types of DNA (whether it’s for example personal DNA, DNA from a person known to be deceased, DNA obtained and authorized from LE) that can be uploaded. During the pilot study, the service was free of charge, though nowadays a small fee is required for LE use of the service.

Recently the web interface “GEDmatch Pro” has been launched for LE. According to present user terms searches can be made in the purpose of identifying human remains and perpetrators in violent crimes such as murder, manslaughter, rape, robbery and violent assault cases. Data uploaded by LE will not be visible as matches for GEDmatch users.

In July 2020 GEDmatch experienced a security breach resulting in all data from every account (including LE accounts) under 3 hours being available for search and comparison by all users [16,17]. At this point the account used in the pilot case (to be described) had already been removed. Other possible security risks have been pinpointed before regarding GEDmatch [18,19] as well as for FTDNA, Myheritage and LivingDNA [19]. The security breach at GEDmatch was followed by a phishing attack on the genetic genealogy database MyHeritage, and it was announced that data could have been acquired during the previous attack on GEDmatch [20].

FTDNA

FamilyTree DNA, FTDNA, a division of Gene by Gene Ltd, is a commercial DTC based in Houston, Texas, US. The company offers autosomal DNA, Y-DNA and mtDNA tests and a number of genealogy tools at their website. At present, the company has around 1.5 million global users, with a heavy representation of users from the US and Europe.

According to the company’s user terms DNA profiles can be uploaded and searched by LE or a third-party representative in order to identify the remains of an unknown individual or to identify a perpetrator of a murder, sexual assault or child abduction.

Since March 2019 an application must be sent to the company from LE or a third-party representative. If the application is approved an account is set up by FTDNA. The account will be tracked by an “in-house”-identification system which will control correct searching and use of uploaded data. Data uploaded by LE will not be visible as matches for any other FTDNA users.

In addition, since March 2019 new users must actively approve (“*opt-in*”) that their DNA data can be compared towards LE uploaded profiles. American users that created their account before March 2019 were given the status “*opt-in*” but could actively choose to “*opt-out*”. As with GEDmatch, European users that created their account before March 2019 were given the status “*opt-out*” due to GDPR regulations but could actively choose to “*opt-in*”. According to the company approximately 5% of the users have opted out [written communication with FTDNA, May 2021].

⁸ <https://www.gedmatch.com/tos.htm>

Also FTDNA aim to protect the individual user integrity with clear and transparent user terms. These are specified in “*Terms of Service*” (TOS), “*Privacy Statement*” and additionally for law enforcement in “*Law Enforcement Guidelines*”⁹.

Comparison of FTDNA and GEDmatch Pro

FTDNA and GEDmatch, with the newly implemented law enforcement interface GEDmatch Pro, have many similarities but also some differences, which are highlighted in Table 2.

Table 2. Comparison of GEDmatch Pro and FTDNA.

	GEDmatch Pro	FTDNA
Size of database and number of users searchable for LE	In total: 1.2 – 1.3 million users, approx. 1% Swedish. Searchable: 325 000 profiles.	In total: 1.5 million users, majority from North America. Searchable: a majority of North Americans and a minor part of Europeans. In total ~ 5% opt-out.
Access of LE datasets for other users	Datasets and/or kit number are not shown on other users list of matches.	Datasets are not shown on other users list of matches.
Account details	Account is created through the GEDmatch Pro’s web interface, including complying with a TOS agreement.	Initial mail correspondence with the company including admission of a formal application form. A TOS contract needs to be signed by LE, if approved, LE accounts are created by FTDNA.
Searchable DNA profiles	Searches of autosomal DNA and X-chromosomal DNA. Y-SNP profiles and mtDNA can be used in order to narrow down the search.	Searches of autosomal DNA and Y-SNP profiles. Y-STR profiles can be searched manually by the company. mtDNA data can only be searched if generated by the Gene by Gene lab.
Presentation of the comparative searches	List of matches* – depending on applied search filter.	List of matches* – depending on applied search filter.

* List of matches does not contain any DNA data but contain information on name/alias, e-mail addresses and an estimated relationship between the uploaded DNA and the “match”. Using other tools available in each database makes it possible to evaluate each potential match more closely.

⁹ Documents available at: <https://www.familytreedna.com/legal/terms-of-service>
<https://www.familytreedna.com/legal/privacy-statement>
<https://www.familytreedna.com/legal/law-enforcement-guide>

3 Pilot case study

3.1 Initial process description

After the arrest of “*the Golden State Killer*” in 2018, which was intensely reported by media and increased the attention on IGG world-wide, the question was raised whether this tool could be used also by Swedish law enforcement.

In the beginning of May 2018, NFC requested a legal inquiry at the Legal Affairs Department within the Swedish Police Authority (SPA), on the possibility to use DNA data in genetic genealogy databases to obtain investigative leads in criminal investigations.

The Legal Affairs Department together with NFC and the National Board of Forensic Medicine (RMV) proposed such a method. The police officer in charge of the preliminary investigation of the double murder in Linköping 2004 was involved in the work process. The approach was documented in a legal inquiry “*Dna-spår och släktforskning*” (eng. “*DNA traces and genealogy*”), signed in January 2019¹⁰. A summary of the different steps in the method was described in the inquiry (Figure 2). It was recommended to test the proposed method in a few criminal cases, in terms of a pilot case study, and at a later stage decide if the method could have a future and wider use within Swedish law enforcement.

During the inquiry process ethical issues were raised and discussed in-house. The method, its implications, and how to proceed, were also discussed at the Swedish Police Authority’s Ethical Boards meeting in December 2018.

IGG had actually been used previously in another Swedish criminal case, though the case concerned an identification of a man found murdered in 2003, the so called “*Ekebymannen*”. The man had not been identified despite large efforts by the police, including the work performed by a cold-case team. The team asked RMV in 2018 for help to extract DNA from the human remains and produce SNP datasets that could be used for searches in genetic genealogy databases¹¹. RMV produced the desired files using WGS followed by bioinformatics and the processing of data [3]. With searches in GEDmatch combined with genealogy the consulted genealogist could narrow in on a geographic area in the northwest parts of Croatia as possible origin for the deceased and his relatives. Since then, the cold case team has identified some possible candidates that are to be investigated further. Unfortunately the work has been delayed due to the covid-19-pandemic¹².

In February 2019 the Director of NFC decided that a pilot case study was to be initiated. It was decided to use the IGG method in a few criminal cases. The pilot performed in 2019-2020 came to be a cooperation project between NFC, the investigation team in the Linköping double murder (with, at the time, the officer in charge of the investigation as well as crime investigators, crime scene investigators and the senior public prosecutor from Linköping Local Public Prosecution Office), legal advisors from the Legal Affairs Department, and expertise from RMV. External experts were contracted for DNA analysis and genealogy.

3.2 Legal inquiries

The Legal Affairs Department of the SPA compiled a legal inquiry¹³ covering the judicial framework and a data protection impact assessment, which was during the course of the pilot study supplemented with two additional inquiries¹⁴. The most important findings of the legal inquiries are summarized in

¹⁰ Dnr A637.388/2018. The legal inquiry was supplemented at two times during the pilot case study.

¹¹ DNA based identification of deceased is performed by forensic genetics at RMV.

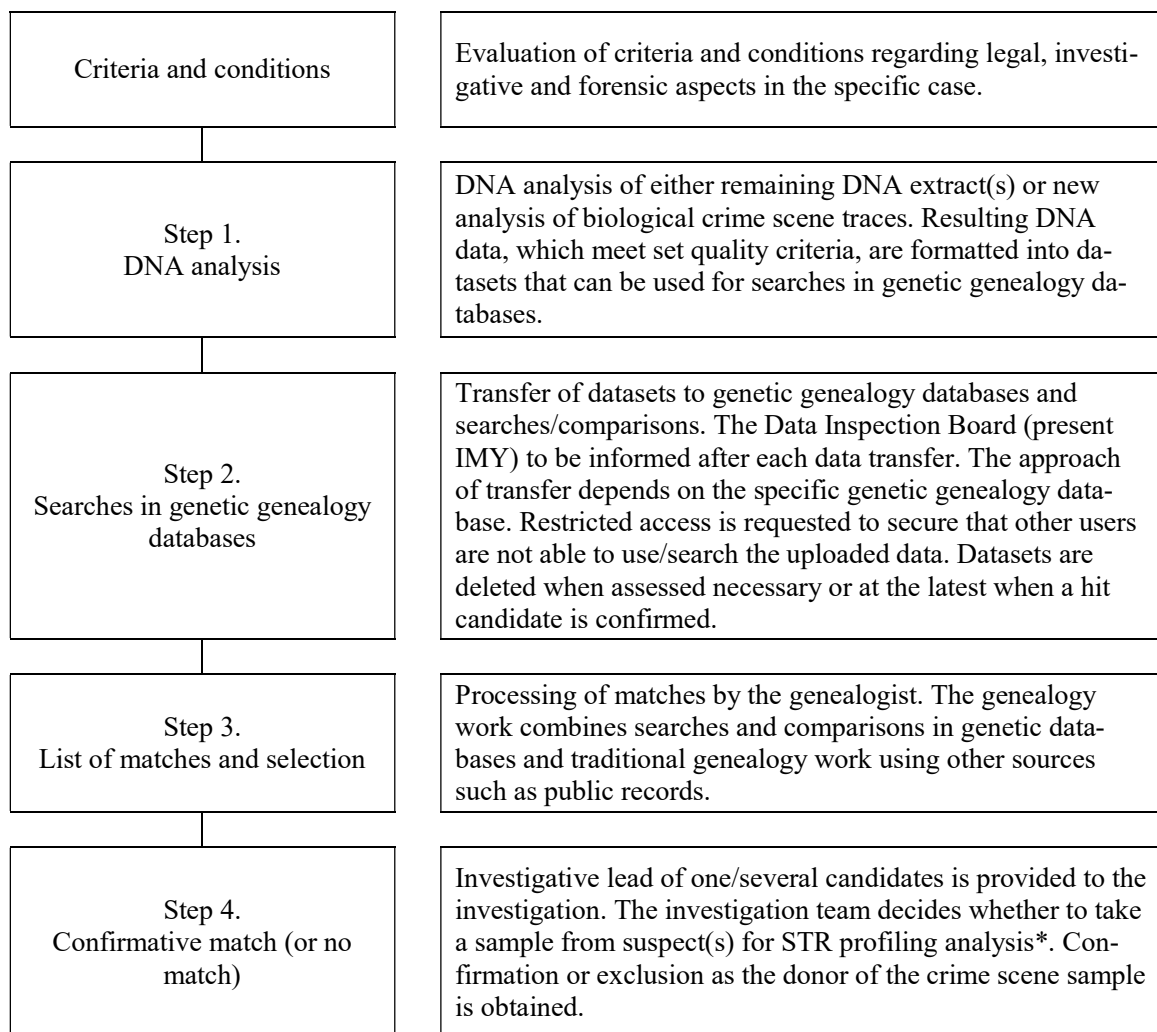
¹² Information provided by police superintendent Bo Lundqvist, Cold Cases, Region South, SPA.

¹³ Finalized in January 28th 2019.

¹⁴ Finalized in May 27th 2019 and Feb 5th 2020.

the following. In this context, it might be appropriate to point out that the Swedish data protection legislation in the area of law enforcement, including the *Criminal Data Act (2018:1177)*, implements Directive (EU) 2016/680 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data, and repealing Council Framework Decision 2008/977/JHA, hereinafter *Directive 2016/680*. However, Directive 2016/680 and national legislation are not identical.

In the first inquiry the four main steps of the method were described in some detail, as outlined in Figure 2.



* According to chapter 28, §12 a (“suspected on reasonable grounds”) or §12 b (other persons of interest in the case, in Swedish: “*annan*”) Code of Judicial Procedure.

Figure 2. Description of the workflow presented in the first legal inquiry with the different steps of the method. As it turned out step 2 and 3 in practice cannot be separated since numerous comparative searches need to be performed in the genetic genealogy databases with the candidates on the match lists.

Regarding step 1, DNA analysis: DNA analysis of a trace as well as the documentation of the analysis falls within the scope of *genetic data*. Genetic data refer to personal data that relate to the inherited or acquired genetic characteristics of a natural person and is a special category of personal data¹⁵, in Swedish law also considered as *sensitive personal data*¹⁶. The processing of special categories of personal data is allowed only when strictly necessary, and under further specific conditions¹⁷. Not every step concerning genetic data involves processing of personal data. Processing of existing documentation is not considered as processing of genetic data, solely as processing of personal data. According to Swedish law, there are legal grounds for NFC to process genetic data for forensic purposes while other units within the SPA do not have the same legal grounds. Thus, measures which concern processing of genetic data, needed to be handled by NFC¹⁸.

In addition, DNA analysis can be performed in-house at NFC but the laboratory also has legal grounds to contract other laboratories or other Swedish expert bodies for expertise assistance.

Regarding step 2, searches in genetic genealogy databases: the first legal inquiry included legal assessments concerning if this step included processing of genetic or biometric data. These were made with some difficulty, due to the legislation not being clear on this matter. It was deemed more likely to be a matter of processing of genetic data, and as a consequence, only NFC had legal grounds to perform step 2.

The legal inquiry also covered legal grounds for the transfer of personal data to *third countries* (i.e. a country outside the European Union), since genetic data was to be transferred to databases in the US. Conditions for transfer of personal data to third countries are regulated in Directive 2016/680 as well as in Swedish legislation. Under the circumstances present in this case, it was not a matter of using the so called Privacy Shield for the transfers. Instead, special conditions for a transfer to be lawful had to be met for each transfer, including that the transfer is deemed strictly necessary¹⁹. In addition, after each transfer of personal data to a third country the supervisory authority IMY (Swedish Authority for Privacy Protection) needed to be informed in writing²⁰.

Furthermore, the question of applicability of a restriction in the Criminal Data Act concerning searches among sensitive personal data was addressed, since this step includes searches among sensitive personal data in genetic genealogy databases. The restriction was found to be applicable only concerning personal data within the police, and thus not concerning personal data in commercial databases, accessible on the Internet, containing personal data being approved by individuals to be used for the purposes in this case. The Legal Affairs Department highlighted that the Swedish data protection legislation unfortunately is not clear on this matter.

In addition, a prior consultation of the supervisory authority²¹ IMY was considered. Though the method was new, it was deemed that since risks of integrity infringement were taken care of and as IMY was to be informed after every third country-transfer a consultation was not necessary at that stage of the process.

A judicial check-list was compiled by the legal advisors as a supplement to the legal inquiry and was written as a simple user-support that could be used during the course of the pilot study. The check-list was primarily used by NFC and to some extent by the investigation team.

¹⁵ Article 3 and 10 Directive 2016/680.

¹⁶ Chapter 2, §12 and §13 Criminal Data Act.

¹⁷ Article 10 Directive 2016/680 and chapter 2, §11 och §12 Criminal Data Act.

¹⁸ Legislation affecting the processing of genetic data at SPA can be found in, inter alia, chapter 2, 12§ Criminal Data Act and chapter 6, 4§ Police Criminal Data Act (2018:1693).

¹⁹ Chapter 8, §8 Criminal Data Act.

²⁰ Chapter 7, §5 Criminal Data Ordinance (2018:1202).

²¹ Article 28 Directive 2016/680 and Chapter 3, §7 second paragraph Criminal Data Act.

A number of ethical concerns do arise in connection to the use of commercial genetic genealogy databases. Thus, prior to the pilot case study the Police Authority's Ethical Board was consulted in December 2018. The discussion covered the methodology as specified in the first legal inquiry as well as possible infringement on privacy rights. The Board supported the continuation of the method development work with a so-called pilot case study.

In the first inquiry it was recommended by the legal advisors to await a more general use of the method within the authority and first evaluate if the method could be considered practically feasible and workable. It was suggested to continue with an evaluation of the proposed method after using a few so called pilot cases.

During the course of the pilot case study some further legal issues arose that had to be addressed. One of them was the updated user terms introduced by GEDmatch, and a legal assessment was made of the consequences. It was noted that the updated terms of service decreased the number of people involved to those actively granting a LE search. During the course of the pilot case study FTDNA also opened up for LE searches and a legal assessment was made on using this database. As suggested in the inquiry a comparative search would first be performed in one of the sanctioned databases and evaluated, and if the need persisted searches could be made also in the other database. It was not specified which database should be used first. However, it was noted that naturally the use of two databases, instead of one, would increase the number of affected individuals.

An assessment was also made of the consequences of the acquisition of GEDmatch by Verogen in December 2019. It was assessed that SNP datasets could be stored and handled by GEDmatch/Verogen. Also, an assessment was made on the fact that datasets needed to be uploaded and handled in the databases during a longer period of time than was anticipated beforehand.

Terms and conditions for suitable cases

A number of terms/criteria and conditions were set up as grounds for selection of suitable pilot cases. These covered legal as well as investigative and forensic perspectives. A majority were considered already in the first legal inquiry and aimed to minimize risks of compromised integrity, ensure proportionality and protection of personal data. Selection of criminal cases suitable for the pilot case study was based on:

- that it was an unsolved serious crime such as a murder or a severe sexual assault
- genetic genealogy searches had to be assessed to bring the investigation forward
- extensive investigative work had already been performed without success. Including DNA analyses and searches in the authority's own national DNA databases as well as international DNA searches including other European countries databases (through the Prüm Treaty), familial searches²² and mass DNA screening
- available and relevant crime scene samples/items and/or DNA extracts of good enough quality and quantity for renewed DNA analysis
- DNA analysis, subsequent bioinformatics and data processing performed by external laboratories needed to follow criteria of quality and secrecy specified by NFC

²² Familial searches have been used as a tool within the Swedish police since January 2019 (Government bill 2017/18:269 pp 172-173). The tool is used only in serious crime investigations. A search is made of a STR-based DNA profile in the authority's national DNA databases. A familial search results in hit candidates that can either be a parent, child or full brother or full sister to the person who left the trace at the crime scene. Hit candidates are presented in a list reported from NFC to the investigators in the specific case, the investigator's work will then show if any of the hit candidates indeed are true relatives or not. The DNA data used in familial searches is sparse compared to the information needed and used in genetic genealogy work.

- favourable circumstances for successful genealogy searches. For example, a witness or DNA-based assessment of the biogeographic origin of the crime scene stain donor, e.g. northern Europe, the British Isles or the US. Or in other words, from regions where many people use genetic genealogy databases and where extensive, historical population registers and records can be found (a prerequisite for the genetic genealogy work).

Regarding the choice of databases

A number of conditions were set up as grounds for selection of suitable databases:

- comparisons were to be made only in databases in which the users had granted LE searches and accepted their data being publicly used according to explicit terms and conditions
- database companies had to ensure that data uploaded was not to be used for any other purpose than the actual search/comparisons. Alternatively, that the SPA could limit the use so that uploaded data only could be used for the specific search
- uploaded data, matching results and other data had to be erasable after the searches had been completed.

3.3 Aim of the pilot case study

Specifically, the aim of the pilot case study was to evaluate and further develop the proposed method, as it was described in the legal inquires, by applying it on one or a few criminal cases. Furthermore, the pilot case study covered a number of key steps, including:

- select criminal cases suitable for the study
- to establish DNA datasets that could be used for searches in genetic genealogy databases
- to transfer DNA data and search in genetic genealogy databases available for law enforcement
- perform genealogy work based on information gathered from the database search (in combination with information from other sources)
- generate investigative leads and provide these to the crime investigation
- to illustrate different aspects concerning the handling of sensitive personal data including genetic information, the processing of data and security issues, contracted genealogy expertise etc.
- to elucidate the division of responsibilities for different actions
- to evaluate if the procedure is practical and workable

3.4 Criminal cases used in the pilot study

Initially two long-standing unsolved criminal cases were discussed. Both fulfilled the terms and conditions and could be selected as pilot cases to test the proposed IGG method. In both cases a child was involved.

The first case considered was an unsolved rape from 1995, the so called “Billdal-case”. The case was soon to reach its period of limitation. A girl had been brutally attacked and raped by an unknown perpetrator and ever since that day there was a high social interest for the case to be solved, but all police investigation efforts made had been in vain. Before the use of IGG a so called familial search was performed early 2019 leading to the arrest of a suspect whose DNA matched the crime scene traces [21]. Thus, the case was no longer relevant to use in the pilot study. In subsequent discussions regarding a replacement with another pilot case it was decided to continue with one case only, an unsolved double murder from 2004.

Double murder in Linköping, Sweden

It was in the morning of 19th October 2004 when an 8 year old boy was stabbed to death from behind on his way to school. A 56 year old woman who witnessed the attack was also stabbed and died from her injuries. The murder weapon, a butterfly knife, was left at the crime scene and subsequently seized. When examined at SKL (National Laboratory of Forensic Science, present day NFC) a DNA trace from three persons was found. The DNA was a mixture of DNA from the two victims and DNA from an unknown third person. The unknown person's DNA was found on more items that were seized, including a knitted cap.

The DNA profile from the unknown individual had been searched (and was continuously searched throughout the pilot case study) in national DNA databases and in international searches (e.g. in European countries implementing the Prüm Treaty). Early in the investigation the most likely geographic origin of the perpetrator was assessed to be north European and to have blond hair and blue eyes. These findings were a combination of genetically-based assessments made at the Erasmus MC, University Medical Center Rotterdam, Netherlands, as well as eye-witness reports and assessment of hair recovered from the knitted cap. A familial search was made in 2019 in the authority's own DNA databases without finding any true close matches. Some 6.500 people were heard and left a voluntary saliva sample to be compared to the DNA profile (STR markers) without a match.

The overall conclusion by the investigation team and NFC was that the DNA trace(s) and the rest of the case circumstances justified the use of this criminal case as a pilot case²³. Thus, an informed decision by the officer in charge of the preliminary investigation together with a forensic advisor at NFC to use the case in the pilot case study was taken in June 2019²⁴.

3.5 Data processing and searches in genealogy databases

Data protection impact assessments had been made in the legal inquiries of searches in GEDmatch and FTDNA, and it was considered possible to perform searches in both databases within the pilot study. DNASolves was not considered an option, partly due to its limited database size. Within the pilot study, SNP datasets were searched in both GEDmatch and FTDNA and upon every dataset transfer IMY was informed in writing by NFC.

To perform genetic genealogy SNP datasets are needed for the comparative searches. Different genealogy databases use slightly different, but overlapping, sets of SNPs. Data are handled in txt-files in which information on each SNP can be found (Figure 3). Usually 600-700k SNP positions are used for comparisons.

rsid	chromosome	position	genotype
rs3094315	1	742429	AA
rs12562034	1	758311	CT
rs3934834	1	995669	GC
rs9442372	1	1008567	AT
rs3737728	1	1011278	--
rs11260588	1	1011521	AC
rs6687776	1	1020428	AG
rs9651273	1	1021403	GG
rs4970405	1	1038818	--
rs12726255	1	1039813	GT

Figure 3. Example of a txt-file, each SNP identification number is stated (rsid) along with the chromosome, SNP position relative to the reference genome (position) and the specific genotype variant (nucleotides A, T, C and G). Usually datasets of several hundred thousands of SNPs are being used.

²³ This case had been discussed and used by the Legal Affairs Department and NFC as a "model case" and used in the process when the method was designed and described in the legal inquiry.

²⁴ The decision was documented 11th of June 2019 by the officer in charge of the preliminary investigation and NFC.

DNA analysis and bioinformatics

SNP datasets can be established using for example microarrays or through WGS followed by bioinformatics [1]. In the pilot case study, WGS was chosen for several reasons, including the fact that the crime scene samples available were of relatively low quantity, displayed some degree of degradation and presence of enzymatic inhibitors were expected [22].

Further, with WGS a complete characterization of the constituents (nucleotides) of an individual's complete DNA sequence (genome) at a specific time point is possible, which makes it possible to extract specific SNP data needed and produce different SNP datasets according to specific requirements. Since neither NFC nor RMV has the WGS technique at hand it was necessary to use an external supplier. WGS sequencing and some bioinformatics were performed in collaboration with the SNP&SEQ Technology Platform (SNP&SEQ) within the National Genomics Infrastructure (NGI) in Uppsala, Sweden. SNP&SEQ could offer an in part ISO/IEC 17025 accredited WGS procedure, while parts of the procedure that fell out of the scope of accreditation were performed according to validated best-practise procedures.

WGS data from SNP&SEQ were transferred to RMV, where the Forensic Genetics Department performed extensive bioinformatics data analysis including SNP analysis, SNP genotyping and bioinformatic processing²⁵. For some SNP datasets so called *imputation* was used, since imputation can expand datasets that lack information [22]. Final products were transferred to NFC before upload to the databases.

Data analysis and genealogy searches of autosomal SNP datasets

In the following section the different steps taken are described in some more detail, a flow chart of the work is presented in Figure 4.

In the beginning of the project an inventory of suitable DNA extracts was conducted by NFC. One DNA sample selected, *DNA extract 1*, had been recovered 2004 from blood on a knitted cap found close to the crime scene. One part of the sample was analysed at RMV, where a SNP panel of 129 SNPs (out of 131 targeted SNPs) was established²⁶ and used as a control in the following work. The other part of the DNA extract was analysed with WGS.

From the first WGS run, data were limited and reliable SNP datasets of less than 300 k SNPs could be established²⁷. Nevertheless, a match would be expected between the profile and near relatives (sibling, cousin or second cousin) if present in GEDmatch. When it comes to more distant relatives (third cousins and beyond) there was a risk for some of these to be missed out in the searches due to the lack of information. With this in mind, a first search in GEDmatch was made in August 2019. The genealogist and police investigators evaluated the list of matches and concluded that the search had led to a number of distant relatives only (>4c), and identified a cluster of potential though distant relatives from northern Germany, as described in more detail in section 3.6.

Different alternatives on how to proceed were discussed. With imputation experts at RMV tried to increase the number of SNPs in the dataset, and in parallel a new inventory of available DNA extracts was initiated at NFC as well as an investigation if a renewed WGS could be performed on existing material at SNP&SEQ.

During the autumn of 2019 SNP datasets were processed using imputation²⁸. The resulting imputed datasets contained many more SNPs, i.e. more informative data, and of higher quality (i.e. fewer SNPs

²⁵ RMV has legal grounds for the processing of personal data in Act 2020:421 Om Rättsmedicinalverkets behandling av personuppgifter (*eng.* "National Board of Forensic Medicine's process of personal data").

²⁶ Laboratory work performed at RMV was accredited according to ISO/IEC 17025.

²⁷ Corresponding to roughly a third to half of the markers normally used by DTC companies.

²⁸ When using imputation SNP genotypes are predicted with the use of a reference panel and expanded SNP datasets can be established, for additional information see e.g. [1] and [22].

that may be falsely assigned). Imputed datasets were thus searched in GEDmatch in October 2019. Upon evaluation of the list of matches, the genealogist once more found only distant connections to Germany (predicted relationships of >4c).

In October 2019 FTDNA changed to accept a setting of SNPs used by other DTC companies, meaning that the imputed dataset uploaded to GEDmatch could potentially be searched also in FTDNA. However, FTDNA had changed their policies and did not accept inquiries from outside the US. Discussions were held in the company's citizen panel whether or not to approve inquiries in murder and sexual assault cases from other countries than the US. At the time, only identification cases from Australia and Canada were accepted.

To expand the SNP datasets a new WGS run was performed (*DNA extract 1*), and followed by data processing, including data imputation. Resulting datasets were comprehensive and was uploaded to GEDmatch producing similar hit candidate lists as in previous attempts (>4c).

At the same time, the investigation team worked on the potential lead towards Germany, although that investigative lead proved hard to investigate, which is described in more detail in section 3.6.

As the produced SNP datasets contained far less information than aimed for a renewed inventory of stored crime scene items and traces were made in the hope of finding items that had not previously been analysed. Some items were identified and DNA analysis was performed only to find that the obtained DNA profiles (STR markers) did not correspond to the DNA profile of interest.

By the end of February, FTDNA informed NFC they had opened up for database searches in this specific case. Since searches in GEDmatch could not effectively lead the investigation forward a decision was taken also to make comparative searches in FTDNA. A TOS had previously been signed by NFC in preparation for the use of the service. The agreement contained restrictions on for example how search matches could be dealt with. Also the investigation was not to directly contact individuals from the match list unless it was absolutely necessary to bring the investigation forward. SNP datasets were transferred to FTDNA (from *DNA extract 1*) three times in March and April 2020. However, the datasets did not meet the company's quality requirements and no searches could be made.

Since remaining items from the crime scene did not give any usable DNA a renewed inventory of available DNA extracts at NFC was made. Two DNA samples were selected and pooled, purified and concentrated into one sample (*DNA extract 2*) which was subjected to WGS analysis and bioinformatics in spring 2020. Resulting SNP datasets were of good quality and coverage in sequence reads. One SNP dataset adapted for GEDmatch and FTDNA respectively was established and uploaded to the databases.

In GEDmatch comparative searches resulted in a similar list of matches as previously²⁹. No additional work was thereafter performed in GEDmatch.

In FTDNA the comparative searches, on autosomal DNA data (*DNA extract 2*), resulted in hit candidate lists with potentially closer relatives that became the basis for the genealogists work. In a "target testing" effort volunteers also sent saliva samples to FTDNA and one of them provided valuable information, as is more closely described in see section 3.6.

After the arrest of a suspect in June 2020 a formal request was sent from NFC to FTDNA and GEDmatch to remove accounts and data which was done and confirmed by the companies.

²⁹ This indicates that at the time there were no closer relatives registered as "opt-in"-users at GEDmatch.

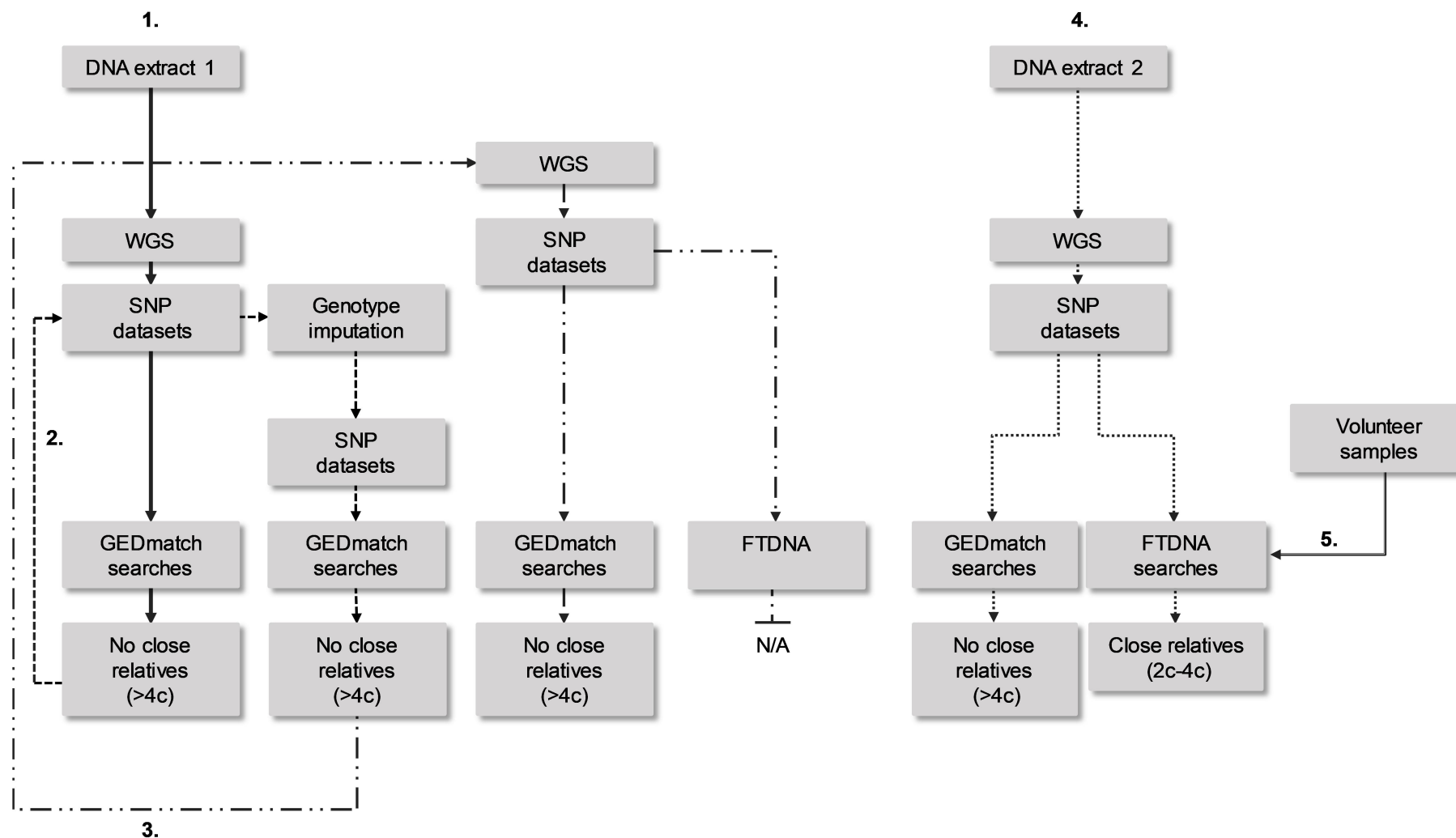


Figure 4. Flow chart of the different steps (timewise in order 1-5) in the pilot case study for searches made with the autosomal SNP datasets. Two different DNA extracts were used for WGS. WGS data was used to produce SNP datasets, and for some datasets genotype imputation was used for expansion. Datasets were searched in GEDmatch and/or FTDNA resulting in lists of matches, whereof a majority were not close relatives (fourth cousins and more distant, >4c). For the final searches in FTDNA, volunteer samples were used and these together with produced SNP datasets resulted in close relatives in the order of 2c-4c.

Genealogy searches of Y- and mtDNA datasets

NFC investigated the possibility to use Y-chromosomal and mtDNA data in FTDNA. Y-chromosomal DNA is paternally inherited while mtDNA is maternally inherited. This data could possibly aid the genealogist by making it easier to correctly place relatives in a family tree and also potentially give information on the geographic origin of the person of interest. In summary, this data could streamline the genealogy work and minimize time spent. FTDNA did not allow any direct upload of Y- or mtDNA-data and only offered manual comparisons performed by the company. A transfer of data was conducted in spring 2020 and the company initiated some data processing and manual comparisons. For the Y-SNP-profile a comparison was made, but a Y-STR-marker comparison was never performed as a suspect was found and arrested through the information brought forward by the genealogist. Regarding mtDNA data no comparison was made by the same reason. These results were not explored further and not communicated or used by the genealogist as the work was actively stopped upon the arrest of the suspect.

3.6 Genetic genealogy in the pilot case

In the beginning of the pilot case study, NFC and the police officer in charge of the Linköping double murder discussed if a genealogist could be contracted. The conclusion was that the genealogist already involved in the identification of “Ekebymannen” would be suitable. In fact, the genealogist himself later called the investigator before a formal contact was taken. The genetic genealogist was hired on hours. A non-disclosure agreement was signed. Primarily, the genealogy work was performed in the authority’s premises in Linköping, and in cooperation with two investigators. However, some of the genealogy work was performed at an external work place and in the home of the genealogist.

Analysis in GEDmatch

The first searches³⁰ in GEDmatch in the autumn of 2019 gave 3,000 matches (corresponding to a standard setting of no limit on the longest common segment for a match). The list of matches was analysed. The top match shared in total 86 cM and the longest common segment was in the order of 29 cM, this correspond to a very distant relative (>4c) or false relative. As all matches displayed a relationship that was distant, the genealogist requested new SNP datasets that could provide more information. While waiting for newer datasets the genealogist still analysed the list of matches and found potential clusters of relatives from the northern parts of Germany. The investigation team had information on a German student that had cut himself in the hand the day of the murder but was never swabbed for DNA. Also, another German man refusing to give his saliva test came in mind of the investigators. A list of a few male individuals from Germany staying in Linköping at the time of the murders was constructed and the genealogist task was to create their family trees. The task proved very hard as Germany practice strict confidentiality on relationships from the 20th century including birth records from within the last 110 years. Upon a renewed search in GEDmatch of expanded SNP datasets, and an evaluation of the list of matches, the genealogist once more found only distant relationships to Germany (predicted relationships of >4c).

Analysis in FTDNA

The genealogist had direct contact with an employee at FTDNA that was involved in the bioinformatics work regarding the specific samples. Searches were made in FTDNA in May/June 2020 and generated 890 matches. Matches were sorted according to total amount of shared DNA and the largest matching segment. The list of matches contained the stated name of each user, e-mail address and which part of the DNA that matched. Closest matches displayed around 60-100 cM which indicated a relationship in the range 3c-4c.

A selection of the most relevant matches was made based on total amount of matching DNA and the size of the largest segment. The genealogist selected 28 individuals whereof 18 Swedish residents and 10 US residents for additional analysis, processing and triangulation.

³⁰ Description of the genealogy work is based on oral/written communication with the genealogist and the novel Genombrottet [23].

Mapping of family trees

At first, individuals in the matching list were identified. Some, but not all, had entered information on their family relations (family trees and/or other information). Individuals without given information could be identified from their names or e-mail address. In two cases a formal request³¹ was sent to the e-mail address supplier. American residents were identified using public registers (such as whitepages.com, thatsthem.com and searchpeopleforfree.com), Facebook and also by public services at the DTC provider AncestryDNA. At Ancestry.com the genealogist used e.g. public family trees, Historical Residential Records and Birth, Marriage and Death Records. Also the Swedish service ArkivDigital (private online provider of Swedish Church Records and other Historical Records) and the Swedish National Archives digital research room were used.

The individuals' relations were researched back in time to seek connections between the family trees and to find common ancestors. In total around 600-700 relationships during the 20th century, 19th century and late 18th century were mapped. An intersection in the form of two sisters and their parents living in the late 18th century and early 19th century were found. Conclusions were drawn from the matching list that the person who left the crime scene sample came from the southern part of the county of Östergötland in Sweden. Since the common ancestors were distant it would have taken considerable time, probably several months, for the genealogist to track thousands of relatives building the family tree. It was decided by the police officer in charge to ask some former and current police staff members from the area if they could ease the process and voluntarily contribute with their reference samples to FTDNA, which some of them agreed to do. Thus, "target testing" was used to streamline and narrow the analysis.

In total, 15 persons with known origin in the county of Östergötland submitted their saliva sample to FTDNA for DNA analysis. Some volunteers did not work within the SPA but were known by the genealogist. All volunteers were informed that the investigation team used genetic genealogy in a pilot case study and that their DNA test possibly could help the investigation. They were also informed that their results would be visible for the police if they matched the potential perpetrator's DNA. Some of them accepted under the condition that their datasets would be deleted immediately after matching, which was also done.

FTDNA agreed upon receiving the samples and treating them with priority. One of these 15 individuals turned out to be a good match. The person was related with a total of 399 cM (largest segment was 96 cM), which when cleared from small random segments (under 5 cM which FTDNA includes in the total sum) resulted in 347 cM in common with the profile from the crime scene. This could imply a relationship closer than 3c (e.g. 3rd cousins)³². The match narrowed the research down from characterizing thousands of relations to analyse only hundreds.

For the common ancestors found in the 19th century the genealogist mapped around one hundred now living descendants. These descendants were mapped to see which of them fitted genealogically in the family tree and find present candidates. When possible candidates were found the genealogist researched back in time, confirming relationships in the constructed family trees according to shared DNA segments and genealogy records.

Finally, two brothers within the expected age range were found that matched the family tree, and thus were the only people that according to the genealogist could have left the DNA at the crime scene. The investigation was immediately informed and the brothers DNA profiles were compared to the crime scene sample, whereas one of them matched the crime scene sample.

³¹ A formal request was sent from the police authority.

³² Based on The Shared cM Project Version 4.0 (March 2020), from <https://thegeneticgenealogist.com/wp-content/uploads/2020/03/Shared-cM-Project-Version-4.pdf>.

3.7 Reflections from the pilot case study

Genetic genealogy databases accessible for law enforcement and IGG may constitute one of the most powerful tools that can be used in criminal investigations.

The pilot case study continued for almost two years, in between 2019-2020. A close collaboration between different parts of the SPA and external experts and suppliers was needed to fulfil the aim of the project and complete the different parts of the method.

In the legal inquiry covering the pilot case work only murder and sexual crime investigations were considered for IGG. A reflection in the aftermath of the pilot study was that the use of the method should also include identification of human remains as part of a criminal investigation. Thus, a legal revision on this topic was made after the end of the pilot (see section 4.3). Other legal issues remained as well and needed to be addressed as the legal basis for this tool needs to be rock-solid to motivate the use of this extremely powerful tool.

During the course of the pilot study several aggravating circumstances emerged. For example, GED-match changed database user terms which led to a drastic reduction in number of searchable user profiles from 1.3 million to considerably fewer. One database closed temporarily. Some potential security issues were pinpointed. These and other events can occur once more in the future, affecting the conditions for database searches and matching. In the following sections some of the circumstances and issues that have affected the work will be discussed, for an overview see Figure 5.

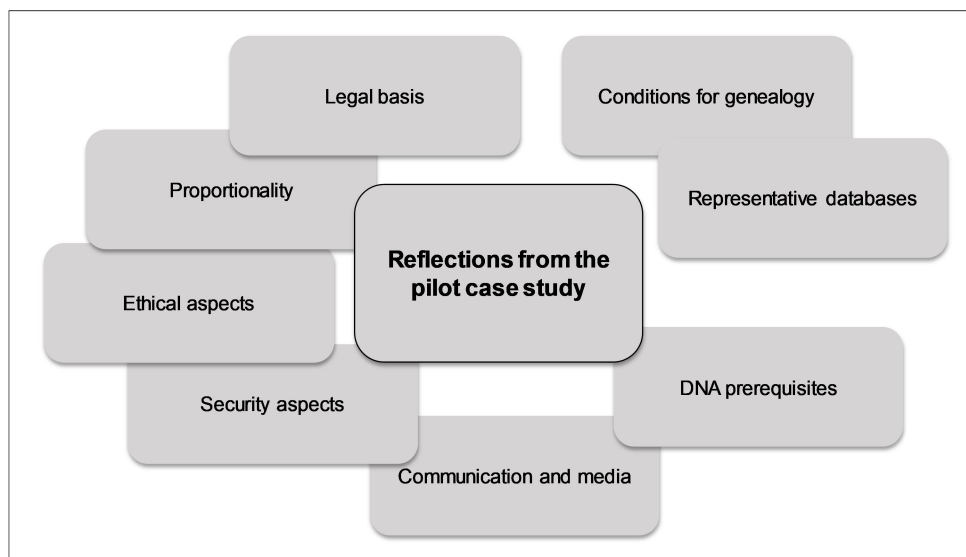


Figure 5. Reflections from the pilot case study concern the legal basis for the work, proportionality and ethics, security aspects, conditions for genealogy, the need of representative databases and DNA prerequisites as well as on communication and media.

Prerequisites on DNA samples

In the pilot study there were difficulties with producing high quality SNP data from available DNA extracts, as discussed in Tillmar et al. [22]. Partly, difficulties were due to the fact that the WGS analysis is adapted for good quality DNA of higher quantities. DNA from crime scene samples will in general be degraded to some extent and contain PCR inhibitory substances originating from the substrate. Many samples are also limited in DNA amount. At NFC current and past extraction protocols focus on gaining extracts suitable for routine STR analysis. As current commercial PCR-kits are highly tolerant to inhibitors the extraction protocols aim at minimizing the loss of DNA rather than mini-

mizing presence of inhibitors. Clearly, for future WGS analysis there is a need for high quality and quantity DNA to produce good settings for successful IGG.

Importance of a representative database

SNP datasets that were first searched in GEDmatch contained less information than aimed for and normally used in comparative searches, but was considered to be of good enough quality to motivate a search. Searches were thus performed resulting in distant potential relatives, which were explored by the genealogist. Unfortunately genealogy led the investigation team in the wrong direction. The investigation team started to follow a track to Germany, though this proved difficult to investigate. But if only comparative searches in GEDmatch would have been possible this direction would probably have been followed to this day.

Thus, there is a potential risk that an investigation may be led in a false direction by genealogy and much work made in vain. To counteract this risk SNP datasets used must be quality assured and the genealogists follow best practice.

In this context it is important to remember that a confirmation or exclusion through STR analysis is always needed as a final step to confirm a match (or no match) between the crime scene sample and the person(s) suggested by IGG. In the pilot case study, IGG suggested but could not separate between two brothers whereas STR analysis confirmed a match with one of them.

None of the searches made in GEDmatch resulted in any closer relatives. However the list of matches produced in FTDNA contained several potential closer relatives. The outcome of a search is dependent on user representatives in the databases. This can be considered obvious but still dictate the conditions for a successful search and has implications on the decision *to* or *not to* make searches in a specific database, or even if a specific case is even applicable for IGG at all.

Security aspects

In July 2020 GEDmatch was attacked through an existing user account³³ [17], which resulted in data from all accounts, including data uploaded by LE, being accessible for several hours allowing for searches (regardless of opt-in or opt-out settings). At this time point, data uploaded in the pilot case study had already been erased. Security risks at GEDmatch have previously been identified [18,19]. A so called "phishing-attack" on MyHeritage user accounts followed a couple of days after, where MyHeritage states that the information (e-mail addresses etc.) used, could have been found in the previous attack on GEDmatch³⁴. GEDmatch experienced new problems and was closed again for some time in late August 2020³⁵.

The storage of data in the databases poses an integrity risk for the person behind the crime scene sample and although security risks can be minimized they can probably never be totally removed.

Proportionality

That the method should be used only in serious felonies is considered reasonable with respect to proportionality. The treatment involves handling of sensitive personal data in considerable amounts and there is a risk for infringement on personal rights for the person behind the crime scene sample. In addition the IGG process involves handling of genetic data and other information from close and distant relatives to the same person but also from unrelated persons for example false matches that may appear on the list of matches.

At the same time, society's interest in solving the crime can be considered to outweigh potential risks for integrity infringement, which was discussed in the first legal inquiry and also in for example

³³ GEDmatch posted an update on its Facebook page July 20th 2020.

³⁴ MyHeritage blogg (2020) Security alert: malicious phishing attempt detected, possibly connected to GEDmatch breach.

³⁵ GEDmatch posted an update on their Facebook page August 26th 2020.

Wickenheiser et al. (2019) [6]. As mentioned before, according to EU and Swedish law the use of the method is conditional on the fact that sensitive personal data, such as genetic data, can only be processed if *strictly necessary*³⁶. Apart from integrity aspects the method is comparatively expensive and demanding. Altogether this only justifies the use of the tool in serious felonies when a majority of all other investigative measures and tools already have been used (without success).

Ethical aspects

Since ethical issues arise with the use of IGG discussions were held with the SPA's Ethical Board both in December 2018 and in September 2020 after the end of the pilot case study. The legal grounds for the method in Swedish law, the need of future regulation of the method in terms of which cases the method could be applied on and the upcoming prior consultation at IMY were discussed.

The use of genetic genealogy unarguably raises a number of ethical questions and has been discussed in the society and scientifically [e.g. 6,8,24-29]. Questions on personal integrity arise naturally. First, the integrity of the individual behind the crime scene sample must be considered, who is unidentified during the time of the investigation but for whom a large amount of genetic information is being processed without the person's knowledge.

Integrity of other users in the databases and their relatives can be discussed as well as already mentioned. An individual accepts searches by LE and with that acceptance comes an acceptance also for all biological relatives, close as well as distant. The acceptance may also be valid not only for past and living relatives but also for generations to come, introducing "transgenerational" issues [29]. Another opinion which has been put forth is that LE activity's does not meet the original purposes of the databases [24] and the commercial interests of the companies providing the databases for LE searches [29]. Also, risks when it comes to increased governmental power in totalitarian states have been a matter of debate³⁷.

IGG can result in the finding of family relations previously unknown and that current family relations are not correct biologically. In most cases, this information though will not be of interest to the investigation and will not be passed on to the individuals affected.

DNA data that are produced contain a large number of information that potentially could be linked to different medical predispositions as well as biogeographic ancestry and phenotype characteristics. Especially the first one is highly sensitive and could be debated from an ethical point of view. There is also a risk of wrongful investigative leads resulting in the arrest of a person innocent to the committed crime, which naturally will cause discomfort including unwanted media attention. This is the case already without the use of IGG and an important note is that the wrongfully identified person will rapidly be excluded when the STR profile is compared to the crime scene DNA profile.

There are also examples of individuals found guilty in court who later have been cleared of charges thanks to the use of IGG. In one case from 1985, a woman was stabbed to death. A man was later imprisoned though claiming his innocence. Recently, DNA recovered from under the fingernails and on the victim's nightgown could be found due to present days more advanced DNA analysis techniques. With IGG the unknown DNA could be linked to a man acquainted with the victim's daughter. The wrongfully convicted man was released after 15 years in prison [30].

In Sweden the magazine "Forskning & Framsteg" made a survey in 2019 on Swedes opinion on LE use of genealogy databases in the purpose of finding culprits [2]. In the survey around 90% of the Swedes were positive regarding the use of the method when it came to serious felonies as murder and

³⁶ Article 10 Directive 2016/680 and chapter 2, §11 och §12 Criminal Data Act.

³⁷ See e.g. a report from the Australian Strategic Policy Institute, ASPI, regarding the Chinese states forensic DNA database: https://s3-ap-southeast-2.amazonaws.com/ad-aspi/2020-06/Genomic%20surveillance_1.pdf?QhPFYrNVaSjvblmFT24HRXSuHyRfhpml.

rape. When it came to less serious crime categories, as car theft and possession of narcotics, 59% were still positive. Corresponding surveys have been conducted in the US showing similar results [24, 31].

Conditions for genealogy

In Sweden the conditions and settings for genealogy are very good as Sweden has comprehensive and easily accessible genealogy records covering several hundred years back in time [2, 23]. In addition, as many as 90,000 to 100,000 Swedes have taken a DNA test and uploaded their data to commercial genealogy services. Globally it is estimated that approximately 36 million people so far have taken a DNA test and used services at one of the four major DTC companies³⁸. The largest site is Ancestry.com with around 18 million posts in their database while FTDNA is considered to be the company with most Swedish users [32,33].

In an article by Erlich et al. (2018) [34] the power of IGG as a method is discussed and according to their estimations it is possible to find a third cousin or closer relative in 60% of the comparative searches made in a database with 1.28 million users. At the same time, a database that covers 2% of a population will result in a similar match rate (third cousin or closer) for almost all individuals in the population [34].

Comparably, GEDmatch contains data from approximately 10.000 Swedes, whereof some are searchable for LE [1]. According to Greytak et al. (2019) [8] the company Parabon NanoLabs has produced a match corresponding to a third cousin or closer for over 80% of their searched samples in more than 250 cases. Though, this automatically does not mean that family trees can be built since this depends on more factors (such as additional, not too distant, relatives and that obtained hits are possible to explore). When using Parabon's own scale on the >250 cases investigated (until 2019) around 80% of European crime scene samples and 60% of non-European samples had been possible to investigate using IGG³⁹.

In other words, the settings for genealogy can be less good if the person is from other parts of the world than Sweden, Europe or the US. Documents such as reliable public records, church records etc. need to be available, accessible and also thoroughly cover individuals back in time.

In the pilot case study, the lead pointing to Germany proved hard to investigate partly due to strict confidentiality of public records. Although, even if information had been made available to the genealogist, only relatively few Germans have performed DTC tests [23]. According to Kling et al. [1], only 1% of GEDmatch users are from Germany. This would have made family tree building using IGG troublesome for the genealogist. As it turned out the trail appeared to be a false lead, and this trail could have been pursued even longer by the investigation team if searches in FTDNA had not been the successful. Thus poorly represented databases should, depending on the circumstances, be avoided.

Communication and media

Ever since the news of the capture of the "Golden State Killer" and the use of IGG in a growing number of criminal investigations and identification of deceased, media attention and public interest for IGG has been extensive. Also, work performed in the pilot study as well as succeeding work attracted a lot of media attention [e.g. 2, 35]. When it comes to the criminal investigation of the double murder in Linköping the investigators have over the years had an intentional open approach towards media. A reporter at one of Sweden's largest daily papers, Dagens Nyheter, followed the work of the genetic

³⁸ User numbers (autosomal DNA-tests) provided by Ancestry, MyHeritage, 23andMe och FTDNA (<https://www.ancestry.com/corporate/about-ancestry/company-fact>, <https://mediacenter.23andme.com/company/about-us/>, https://isogg.org/wiki/Autosomal_DNA_testing_comparison_chart and <https://ggi2013.blogspot.com/2020/02/how-big-is-familyreedna-database.html>).

³⁹ Samples were classified as "European" if the person who left the crime scene sample was estimated to be of at least 80% European biogeographic origin.

genealogist [36,37] and later published the novel *Genombrottet*⁴⁰ [23] together with the genealogist. Throughout, a high level of transparency and sharing of details and findings from the pilot case study has been an actively chosen approach (see also ref. 22).

3.8 Summary of the pilot case study

In conclusion, the pilot case study was successfully conducted during 2019-2020, and a report covering the pilot case study (written in Swedish) was published in November 2020 [38]. DNA samples had been analysed to obtain usable DNA data, processed data had been uploaded and searched in genetic genealogy databases and using genetic genealogy methods hit candidates or investigation leads, had been presented to the investigation team. And then, in the beginning of June 2020, more than 15 years after the double murder in Linköping, a suspect could finally be arrested. After routine STR testing a DNA profile matching the DNA from the crime scene could be confirmed and the suspect was placed in custody. The suspect admitted committing the murders and was found guilty in court and sentenced to psychiatric care.

The work was done all in order to test the IGG method in real casework also aiming to solve the case. With some minor modifications the proposed method, as presented in the first legal inquiry, was found to be implementable and practically feasible. Thus, under the right conditions the IGG method can be an extremely powerful tool for Swedish police to solve previously unsolved cases. Recommendation from the group involved in the pilot case study was that the IGG method should be applied in Swedish serious crime investigations and in identification of human remains in investigations concerning serious crimes.

Based on the performed pilot study and the conducted legal inquiries a number of criteria and conditions were suggested to be fulfilled for a criminal case to be subject to IGG, as listed in Figure 6.

A number of additional considerations were addressed, including a proposition to establish national guidelines or equivalent documentation describing the method, means of regular overview and evaluation of the method. It was also recommended to establish routines for continuous follow up on security aspects and changes of user terms and conditions. Further, to coordinate media strategy in future cases as the work was foreseen to continue to attract media attention. It was suggested that the decision or advice to use the method in a specific case could be taken by a suitable body/subgroup within the SPA.

It was decided that concerned parties i.e. the National Operations Department (Noa), NFC, the Legal Affairs Department and representatives from investigative part of the SPA were to continue the work after the end of the pilot case study.

⁴⁰ Genombrottet: English "the Breakthrough", also "genom" may mean genome and "brottet" means crime.

Conditions and criteria

IGG should only be used in criminal investigations of very serious felonies as deadly violence and severe sexual assaults.

Previous investigative work has been extensive and included searches of STR profiles in national and international databases (according to the Prüm Treaty and via Interpol). Familial search has been conducted. Assessment of biogeographic origin has been made alternatively reliable information exists on an origin in e.g. Europe or North America.

Assessment together with a genealogist if there is a basis for successful genealogy. Based on expected biogeographic origin of the perpetrator, the population represented in searchable databases as well as accessibility of genealogy sources (public records etc.) from the region of interest.

Documentation of an informed decision by the officer in charge of the preliminary investigation. IGG should only be used in the specific case if deemed strictly necessary, proportional and can progress the investigative work.

DNA analysis

DNA extract from a relevant crime scene sample is available. Alternatively, crime scene samples or items can be examined and subjected to DNA analysis. The DNA extract must be from one person, and of high enough quantity and quality for a successful DNA analysis.

External suppliers used for DNA analysis and data processing fulfil demands set by NFC (including quality assurance and data protection demands).

DNA datasets need to be of high enough quality for comparative searches.

Comparisons in genealogy databases

Searches are only to be made in genealogy databases that allow law enforcement searches for the specific crime.

Searches are only to be made against users that have accepted law enforcement matching. For the user it must clearly be stated in the company's terms and conditions, or corresponding, that searches can be made by law enforcement.

It must be ensured that uploaded data cannot be used by other database users.

SPA is responsible for comparative searches made in genealogy databases. The database company is responsible for their data handling during the time when data is processed and stored.

Generation of investigative leads

The outcome of genetic genealogy measures is the production of investigative leads with one or more hit candidates.

Confirming match

A confirmative STR profiling of a suspect must always be made to confirm or exclude a hit candidate to the crime scene sample.

Processing and handling of data

Processing of data, in every step of the method, is made with respect to current laws and regulations as well as approved from a data security perspective. Upon the confirmation of a hit candidate all information and data used in genealogy databases and/or external suppliers should be erased.

SPA continuously oversees database terms and conditions as well as legislative development.

Figure 6. Criteria and conditions suggested to be fulfilled for the use of IGG within the SPA.

4 Implementation of the IGG method

Following the successful application of the IGG tool in the pilot case study it was decided within the SPA to initiate implementation of the method.

In September 2020, the pilot case study was once more discussed at a meeting with SPA's Ethical Board. The outcome of the pilot was presented to the Board and in the following discussion it was clearly emphasized that use of this extremely powerful method needs to be regulated in terms of for which criminal cases it can be applied on, possibly with the use of an oversight body/group. Also, the legal grounds for the application of the method were discussed, including the upcoming prior consultation at IMY.

In November 2020, a working group consisting of staff from NFC, the former police officer in charge of the criminal investigation team of the Linköping double murder, a forensic geneticist at RMV, and legal advisors from the Legal Affairs Department and Noa was established (Figure 7). The main task for the group was to suggest a process description of the method and organizational structure within the SPA for the procedure. In addition, national guidelines need to be established. Also, issues presented in the pilot case study report [38] were to be addressed as well as to perform a national inventory of criminal cases. Some legal considerations were addressed in a complementary fourth legal inquiry (described in the following).

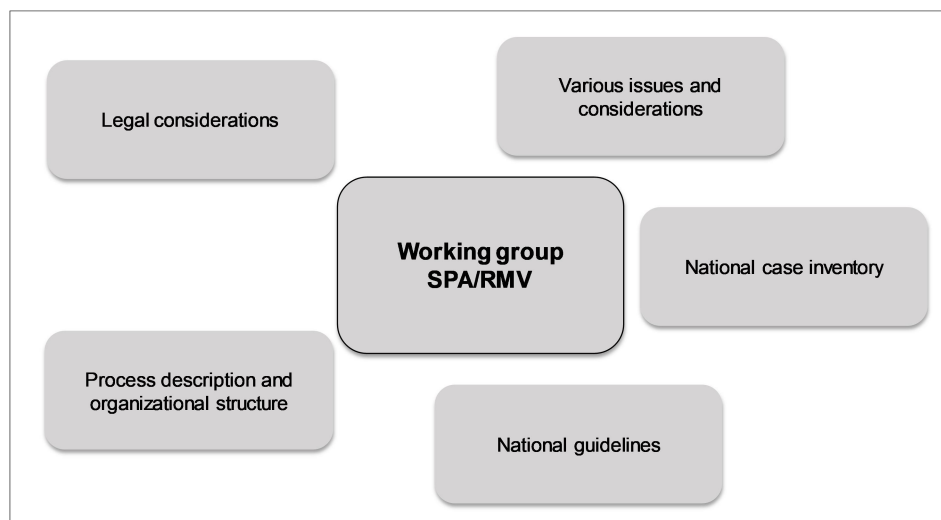


Figure 7. A working group with members from the SPA and RMV was established to work with the implementation of the IGG method. In the figure the different tasks of the group is schematically illustrated.

Regarding issues that were stated in the report [38] that needed to be addressed before implementation of the method, these comprise for example data protection issues as well as forensic and investigative issues. Some of these are:

- to establish requirements on quality and quantity of DNA extracts, as well as on the bioinformatics and data processing used in the SNP analysis (quality assurance)
- finalize information data protection routines including correct data processing and treatment, within the SPA as well as at external suppliers
- to assure quality and/or specify genealogy standards for best practice. This could include to systematize documentation including sources utilized and on what grounds (etc.)
- review and explore the potential benefits using Y-DNA and mtDNA data comparisons

- perform a legal inquiry of the potential use of public information in databases such as Ancestry, 23andMe and MyHeritage
- clarifying if the genealogist should be employed by the SPA, and if so at NFC or at another part of the organization, or can act as an external supplier.

4.1 National guidelines

A process for the development of national guidelines has been initiated. The objective is to provide a guide for the SPA, and especially for investigators within the SPA that are interested in using the method in a specific criminal case or in cases concerning the identification of deceased.

Guidelines will include a description of the IGG method under Swedish settings including suitable case criteria and conditions and references to relevant legal documentation. Further, the guide will contain both regulatory conditions as well as recommendations.

4.2 National inventory of criminal cases

It is presently unknown how many criminal cases in Sweden that fulfil proper criteria and conditions for the application of the IGG method. Since IGG is time consuming and a somewhat expensive tool and also raises ethical and integrity considerations, it is of value to estimate the number of cases that can be of interest for IGG.

Depending on the number of cases it would, following the implementation of IGG, be necessary to at least for a start prioritize between cases and direct the use of the authority's resources and efforts on cases that would benefit the most.

With this in mind the working group initiated a national survey in spring 2021 that was sent to all police regions. Investigators nationwide responded to the survey showing their interest for the method in a number of specific cases.

However, due to the outcome of the prior consultation at IMY, it was decided that the incoming data should not be summarized, in order to avoid creating a delicate list with no direct use. Depending on the timeline and outcome of the present situation a renewed inventory may again become relevant.

4.3 Legal considerations

In preparation for the implementation of the method and the prior consultation at IMY, a fourth legal inquiry was conducted. In the following sections, the fourth legal inquiry, the prior consultation by IMY and the present situation are summarized.

Supplementary legal inquiry on DNA traces and genealogy

The fourth supplementary legal inquiry was finalized by the Legal Affairs Department in February 2021. It covered specific data protection issues that arose during the course of and after the pilot study and were deemed necessary to address before a prior consultation at IMY.

One issue concerned the possibility to use IGG not only to find a perpetrator, but also to identify victims of crime. The Legal Affairs Department did not find any legal obstacle for the method to be used to identify human remains in the investigation of a serious crime, as long as the same terms and conditions apply as when the method is used to find a perpetrator of a serious crime. It can be noted that in Swedish law the Criminal Data Act applies to the processing of data concerning individuals that are *alive*⁴¹. Thus in identification cases the Criminal Data Act and Police Criminal Data Act (2018:1693) do not apply to the processing of data in the first step of the method. However, in the following steps

⁴¹ Although not expressly stated in Directive 2016/680, during the Swedish implementation process this was also found to apply to the Directive.

data concerning individuals who are alive might be processed (relatives to the deceased), which makes the legislation applicable to the processing of their personal data. Nor in this case any legal obstacles were found. In addition, the companies' terms of service specify the processing of personal data from human remains. At present, both FTDNA and GEDmatch allow the upload of data from the remains of a deceased with the purpose of finding a perpetrator. Of note from the new terms and conditions in GEDmatch, a user cannot "opt-out" from such a search while FTDNA still provides this opportunity to their users.

Second, the legal inquiry addressed the types of criminal offences which the method can be applied to. It is clear that IGG affects personal data from persons who may not have anything to do with the offence or investigation, is time demanding, should only be used when highly set criteria are fulfilled and that only a few pilot cases (one) have been executed in Sweden. All taken together, it is proposed that the method should only be used in specific serious types of crimes, but from a data protection point of view, there was no reason to point out which types of crimes that should come into question. In every case, a proportionality assessment needs to be performed and in that assessment the severity of the crime should be taken into consideration. However and noteworthy is that the method will be restricted to the crimes dictated by the database LE user terms.

A third issue concerned to what extent the conditions of transfer of data to a third country (in this case the US) are affected by the so called *Schrems II judgment* (C-311/18) by the Court of Justice of the European Union (CJEU). This relates to the step when data are transferred to a US based company. It is argued that the Schrems II judgment does not affect the previous legal inquiry assessments on the transfer of data to the US, since the assessments do not rely on transfers using the so called Privacy Shield, but on conditions of a transfer in the specific case. It should be noted that the conditions of a transfer need to be fulfilled in each specific case.

Prior consultation at IMY

IMY's role when it comes to prior consultations⁴² is to provide advice when a planned process would risk the processing of personal data being in conflict with law or other statutes⁴³. Their advice is not appealable.

A prior consultation at the IMY was requested in February 2021 by the SPA regarding the authority's planned implementation and altered processing of personal data at NFC, i.e. implementing IGG. Documents provided to IMY included all four legal inquiries made by legal advisors at the Legal Affairs Department as well as the pilot case study report [38].

IMY responded in May 2021 with a written statement which is summarized below⁴⁴. On several points the views and assessments made by IMY differ from the assessments made by the SPA. Accordingly, SPA and IMY make different interpretations of the data protection legislation.

On processing of genetic data

First, IMY deems that genetic data are being processed in the different method steps from DNA analysis of DNA traces to when a comparative search is made in genetic genealogy databases.

On legal basis for the processing of genetic data

In the legal inquiries SPA states that NFC can process genetic data on the basis of chapter 2, §12, Criminal Data Act and chapter 6, §4 Police Criminal Data Act. From these provisions IMY concludes that only NFC, as part of the SPA, has the right to process the intended genetic data, and that data need to be processed at NFC according to the purposes stated in chapter 6, §1 Police Criminal Data Act.

⁴² Article 28 Directive 2016/680 and Chapter 3, §7 second paragraph Criminal Data Act.

⁴³ Chapter 5, §4 Criminal Data Act and chapter 5, §3 Criminal Data Ordinance. See also article 28 and 46 Directive 2016/680.

⁴⁴ Report "Förhandssamråd enligt brottsdatalagen; användning av dna-baserade släktforskningsdatabaser", DI-2021-1521, 2021-05-03.

Further, IMY does not question SPA finding that the purposes stated in chapter 6, §1 Police Criminal Data Act cover the part of the method where DNA data are analysed and processed with the assistance of a Swedish expert laboratory and RMV (step one). However, for the next step in the method when DNA data are transferred to a privately owned foreign genealogy database, IMY finds that there are no valid purposes for the processing of data by NFC. It might be pointed out that SPA does not agree on this matter.

IMY also highlights the need of *strict necessity* for processing of genetic data to be allowed⁴⁵, although they do not go into detail of what would be needed to fulfil this requirement given their assessment of lack of legal grounds, see the following.

On prohibition of searches according to the Criminal Data Act

According to chapter 2, §14 Criminal Data Act, it is prohibited to conduct searches with the purpose of producing a selection of individuals based on sensitive personal data, such as genetic data. Both SPA and IMY state that searches made in genetic genealogy databases are conducted with that specific purpose. However, as mentioned above, SPA claims that the prohibition does not apply in this situation since it is deemed to apply only to searches *within* the police, and searches in the present case are made on data available on the Internet and only concerning users who themselves have granted searches. In turn, this makes the information clearly public and the police must be assumed to be able to process such data according to article 10.c Directive 2016/680. However, IMY disagrees on this point and finds that the prohibition is applicable in this case, and therefore that the planned processing is not in line with current national legislation. According to IMY a change of law is required for the police to be able to use genetic data as proposed in the method.

IMY also mentions that other provisions apply and must be fulfilled, such as provisions on transfer of personal data to third countries. Although IMY states that transfer of data may be performed under certain conditions, any transfer of data to a third country must be performed restrictively and in a small scale. In their opinion SPA aims to use the method in a larger extent and searches are to be performed repetitively which is not allowed using the specified legal basis, which only permits transfers in specific cases.

On the obligation to request a prior consultation

In addition to the above mentioned response IMY commented on some issues, one being that the SPA should have requested a prior consultation *before* the start of the pilot case study according to chapter 3, §7 in the Criminal Data Act, due to the fact that personal data was processed.

In summary, IMY assessed that the planned processing of data would result in the processing of personal data in contradiction with current legislation. Thus, IMY advised the SPA not to proceed with the planned method.

On the consequences of IMY's assessments

For the SPA, the consequences of IMY's consultation have been considerable and will delay the implementation and use of the method with several years. However, there is a strong need for this method to be available for the SPA when other measures in the course of a criminal investigation have proven not to be sufficient.

As mentioned above, SPA does not agree on the assessments made by IMY. One complicating factor is that the statement by IMY (prior consultation leading to advice) is not legally binding or appealable, despite its effect as a prohibition in practice. To reach an appealable decision, SPA would probably

⁴⁵ In chapter 2, §12 in the Criminal Data Act it is stated that biometric and genetic data can only be processed if specifically stated and strictly necessary for the purposes of the processing. In chapter 6, §4 in the Police Criminal Data Act it is stated that NFC may process biometric and genetic data for forensic purposes as stated in §1 if strictly necessary for the purposes of the processing.

need to proceed with the proposed method, which would provoke sanctions from IMY, such as a financial sanction. Such sanctions could then be appealed to court by SPA, and lead to a judicial review of the diverse assessments made by SPA and IMY. Though, such a process is expected to be prolonged and have an unknown outcome.

Further alternatives on how to proceed have been examined by SPA, such as initiating a formalized revision of legal statutory followed by necessary adjustments of statutory regulations, i.e. aim for the legislation to be changed, to adjust the method so that the obstacles addressed by IMY can be avoided or lastly, to decide on no further use of the method.

SPA finds a significant need to use the method in certain unsolved cases. Refraining from step two (the fundamental part of the method - searches in genealogy databases, which IMY found was not in line with national legislation) is not an option, thus adjusting the method was found not to be an appropriate way forward.

Instead, initiating a revision of legal statutory, and if necessary also adjustments of relevant legislation, was deemed to be the most constructive way forward. On 20 May 2021 the Swedish Government decided on Committee directives to review the prerequisites of the use of biometric tools by law enforcement⁴⁶. A special investigator will conduct the inquiry, which has the overall aim of more criminal suspects being identified through the use of fingerprints, DNA, facial images or similar information on individual characteristics. Current legislation and possibilities to collect, store and use such information will be analysed. SPA will be involved in the inquiry and will aim for necessary legal grounds to be clarified and specified for the future use of genetic genealogy. According to the Committee directives, the inquiry is to be finalized in February 2023. If needed, new legislation can follow thereafter, resulting in clear legal grounds for SPA to use IGG.

5 Future perspectives

In Sweden, future use of IGG in criminal cases will have to await findings of revision of legal statutory and, if deemed necessary, legislation adjustments. During this timeframe several other prerequisites for the use of the method may be changed. These include the availability of commercial databases for LE searches and for which type of criminal offences searches can be conducted. Also, the number of available database users may have changed, either to lower, no longer usable, numbers or expanded with even more users and perhaps also additional geographic regions represented.

It will also be of great interest to follow the work on IGG being performed in other European countries, including other police authorities' data protection impact assessments on legal grounds for transfer of data to third countries and on selection of individuals based on sensitive personal data.

Several European countries have considered the use of IGG, to our knowledge including the Netherlands and Great Britain. In the Netherlands, a pilot study is underway on the identification of human remains [29]. In Great Britain there where information in the press during spring 2020 which stated that the method would not be used in cold cases due to ethical issues [40], but in September 2020 a report was published by the British Biometrics and Forensics Ethics Group⁴⁷ (BFEG) with a slightly different angle. BFEG evaluated the use of IGG in criminal investigation and identification cases in Great Britain [41]. In the report it was concluded that the process is unregulated and that many unanswered questions remain. BFEG recommend the use of the method only in identification cases, and not in criminal investigations, at least initially to test the procedure in the UK settings. Further, the

⁴⁶ Biometri i brottsbekämpningen, Dir. 2021:34. <https://www.regeringen.se/rattsliga-dokument/kommittedirektiv/2021/05/dir.-202134/>.

⁴⁷ BFEG "Biometrics and Forensics Ethics Group is an advisory non-departmental public body, sponsored by the Home Office".

group recommends that the use of IGG in a specific case be authorized by the Forensic Information Database Strategy Board⁴⁸.

A framework for LE use of IGG in Australia, has been published [42] discussing the use of the method to solve cold cases or identify human remains. To our knowledge IGG has been used in a few cases in Australia. IGG has also been used in identification cases in Canada and in the Philippines to find the unknown fathers of a few children [29].

There are several companies offering to perform DNA extraction and produce searchable DNA data for genetic genealogy purposes, including Parabon NanoLabs, Bode Technology and Verogen [43-45]. At present, Parabon NanoLabs and Bode Technology also offer genetic genealogy services using commercially available genealogy databases [43,45]. This could be an alternative also for the SPA, however to use such services in the future must first be reviewed legally.

Verogen has developed their own laboratory kit⁴⁹ that potentially can be used on forensic samples to produce SNP datasets for genetic genealogy services. The kit contains a limited amount of SNPs (10k) which limit the tools that can be used for comparative searches and also, at present, to our understanding closes the door for searches in other databases than GEDmatch Pro.

Successful IGG depends on comprehensive and qualitative SNP datasets and currently searches are being made with DNA profiles from a single person. Thus, usually one uses crime scene samples indicating DNA from a single person. When it comes to crime scene samples that contain a mixture of DNA from several persons, the laboratory needs to extract the SNP profile of the person of interest from the mixture. In the literature some successful uses of mixed DNA samples have been described (where one DNA contributor is known and characterised) [1,8]. One will probably be able to use IGG on DNA mixtures more often in the future. In addition the overall technical progress may result in that DNA of less quantity and of lower quality can be used in the future as protocols for WGS or microarrays will be adapted for forensic samples. With method development projects in-house, the aim is to optimize DNA samples for SNP analysis. Also genotyping protocols can be adapted to the specific requirements of the forensic DNA traces.

6 Concluding remarks

In this report an overview of IGG has been presented as well as the performed pilot case study and measures taken for the implementation of the IGG method within the SPA.

The power of IGG has proven itself again and again. If necessary legal framework is in place the tool can be used to help solve serious cases as well as aid in the identification of human remains. Work performed within the pilot case study was not straightforward but successfully ended in the identification of a suspect, solving one of Sweden's most highlighted serious crimes. A number of valuable lessons were learnt during the course of the pilot case study and considerations for future use of the method are handled primarily by the established working group. A review of current legislation will hopefully be conducted so that the IGG tool can be used once again in Sweden.

⁴⁸ For more information see <https://www.gov.uk/government/groups/national-dna-database-strategy-board>.

⁴⁹ <https://verogen.com/products/forenseq-kintelligence-kit/>

7 Acknowledgements

We would like to thank people involved in the past and in present ongoing work. Many thanks to Helena Trolläng director of NFC, legal advisors Leena Mildemberger, Sara Markstedt, David Kummel and desk officer Lena Elfving Nixon (SPA, Legal Affairs Dept.), legal advisor Joakim Norenhag (SPA, Noa), police inspectors Anders Bergmark, Henry Jansén and Susanne Johansson (SPA, Region East), district prosecutor Britt-Louise Viklund (ÅM), genealogist Peter Sjölund/Släkt & DNA, superintendent Bo Lundqvist (SPA, Region South), SNP&SEQ Technology Platform in Uppsala, FTDNA and GEDmatch. Special thanks to legal advisor Leena Mildemberger (SPA, Legal Affairs Dept.) and senior advisor Christina Forsberg (SPA, NFC) for reviewing the manuscript.

8 References

- [1] Kling, D. et al (2021) Investigative genetic genealogy: current methods, knowledge and practice. *Forensic Sci. Int. Genet.* 52 (102474)
- [2] Snaprud, P. (2019) Mördare fångad av släktforskare. *Forskning & Framsteg* (nr 4/2019). <https://fof.se/tidning/2019/4/artikel/mordare-fangad-av-slaktforskare> [2021-09-12]
- [3] Tillmar A. et al. (2020) Whole-genome sequencing of human remains to enable genealogy DNA database searches - A case report. *Forensic Sci. Int. Genet.* 46 (102233)
- [4] Sederstrom, J. (2020) ‘We got the right ending’: Genetic genealogist who helped put Golden State Killer away reflects on his guilty plea. <https://www.oxygen.com/crime-news/barbara-rae-vente-reacts-to-golden-state-killer-joseph-deangelos-guilty-plea> [2021-09-12]
- [5] Thomson J. et al. (2020) An empirical investigation into the effectiveness of genetic genealogy to identify individuals in the UK. *Forensic Sci. Int. Genet.* 46 (102263)
- [6] Wickenheiser, R.A. (2019) Forensic genealogy, bioethics and the Golden State Killer case. *Forensic Sci. Int. Synergy* 1 (114-125)
- [7] Kennett, D. (2019) Using genetic genealogy databases in missing persons cases and to develop suspect leads in violent crimes. *Forensic Sci. Int.* 301 (107-117)
- [8] Greytak, E.M. (2019) Genetic genealogy for cold case and active investigations. *Forensic Sci. Int.* 299 (103-113)
- [9] United States Department of Justice (2019) Department of Justice Announces Interim Policy on Emerging Method to Generate Leads for Unsolved Violent Crimes. <https://www.justice.gov/opa/pr/departement-justice-announces-interim-policy-emerging-method-generate-leads-unsolved-violent> [2021-09-12]
- [10] United States Department of Justice (2019) Interim policy forensic genetic genealogical DNA analysis and searching. <https://www.justice.gov/olp/page/file/1204386/download> [2021-09-12]
- [11] SWGDAM (2020-02-18) Overview of Investigative Genetic Genealogy <https://www.swgdam.org/publications> SWGDAM Communications. [2021-09-12]
- [12] Ge et al. (2021) Forensic investigation approaches of searching relatives in DNA databases. *J. Forensic Sci.* 66 (430–443)
- [13] MyHeritage, <https://www.myheritage.se/privacy-policy> [2021-09-12]
- [14] Skeva, S. et al. (2019) Review of policies of companies and databases regarding access to customers’ genealogy data for law enforcement purposes. *Personalized Med.* 17 (2)
- [15] CNN (2019-06-09) How a Utah assault case upended the cutting-edge DNA website that caught the Golden State Killer <https://edition.cnn.com/2019/05/27/us/genetic-genealogy-gedmatch-privacy/index.html> [2021-09-12]
- [16] Wired (2020-12-16) Cops Are Getting a New Tool For Family-Tree Sleuthing <https://www.wired.com/story/cops-are-getting-a-new-tool-for-family-tree-sleuthing/> [2021-09-12]
- [17] Facebook (2020-07-20) <https://www.facebook.com/officialGEDmatch/posts/226701508802584> [2020-07-21]

- [18] Ney, P. et al. (2019) Genotype extraction and false relative attacks: security risk to third-party genetic genealogy services beyond identity inference. Network and Distributed Systems Security Symposium 2020
- [19] Edge, M.D. et al. (2020) Attacks on genetic privacy via uploads to genealogical data bases. eLife 9:e51810
- [20] Security alert: malicious phishing attempt detected, possibly connected to GEDmatch breach. MyHeritage Blog. <https://blog.myheritage.com/2020/07/security-alert-malicious-phishing-attempt-detected-possibly-connected-to-gedmatch-breach/> [2020-07-27]
- [21] Schander, A. (2019) I jakten på en våldtäktsman, Kriminalteknik 3 (14-19), <https://nfc.polisen.se/om-nfc/nyhetsarkiv/2020/februari/i-jakten-pa-en-valdtaktsman/> [2021-09-12]
- [22] Tillmar, A. et al. (2021) 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. 53 (102525)
- [23] Bodin, A. and Sjölund, P. (2021) Genombrottet. Stockholm: Norstedts
- [24] Samuel, G. & Kennett, D. (2020) The impact of investigative genetic genealogy: perception of UK professional and public stakeholders. Forensic Sci. Int. Genet. 48 (102366)
- [25] Court, D.S. (2018) Forensic genealogy: Some serious concerns. Forensic Sci. Int. Genet. 36 (203-204)
- [26] LA Times (2019-10-24) DNA genealogical databases are a gold mine for police, but with few rules and little transparency. <https://www.latimes.com/california/story/2019-11-24/law-enforcement-dna-crime-cases-privacy> [2021-09-12]
- [27] Nature (2020-09-09) The controversial company using DNA to sketch the faces of criminals. <https://www.nature.com/articles/d41586-020-02545-5> [2020-09-10]
- [28] The Biometrics and Forensics Ethics Group. Should we be making use of genetic genealogy to assist in solving crime? A report on the feasibility of such methods in the UK, September 2020. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/916364/BFEG_Genetic_Genealogy_Final.pdf [2021-09-12]
- [29] de Groot, N.F. et al. (2021) Commercial DNA tests and police investigations: a broad bioethical perspective J. Med. Ethics. 2021-107568
- [30] The Guardian (2020-02-14) Californian man freed after 15 years in prison thanks to genealogy website data. <https://www.theguardian.com/us-news/2020/feb/14/california-man-second-person-ever-exonerated-publicly-available-genetic-technology-dna> [2021-09-12]
- [31] Guerrini, C.J. et al. (2018) Should police have access to genetic genealogy databases? Capturing the Golden State Killer and other criminals using a controversial new forensic technique. PLOS Biology Oct 2
- [32] Aftonbladet (2020-06-17) <https://www.aftonbladet.se/kultur/a/qLjQAL/slakten-ar-varst> [2021-09-12]
- [33] Rötter.se (2019-02-12) Så mycket kostar det att släktforska. <https://www.rotter.se/senaste-nytt/2799-sa-mycket-kostar-dna-slaktforskning> [2021-09-12]
- [34] Erlich, Y. et al. (2018) Identity inference of genomic data using long-range familial searches. Science 362 (690-694)

- [35] Snaprud, P. (2019) Polisen ska använda släktforskarens dna. Forskning & Framsteg 2019-01-30. <https://fof.se/artikel/polisen-ska-anvanda-slaktforskarens-dna> [2021-09-12]
- [36] Bodin, A. (2020) Studio DN 12 juni: Ögonblicket då dubbelmordet i Linköping fick sin lösning. DN [2020-06-12]
- [37] Bodin, A. et al. (2020) I 16 år har Jan Staaf försökt lösa dubbelmordet - släktforskaren gav sista pusselbiten. DN [2020-06-09]
- [38] Fagerholm, S.A. et al. (2020). Pilot: Dna-spår och släktforskning, Användning av släktforskningsdatabaser i brottsutredande syfte. Report Swedish Police Authority, NFC, Biology Unit 2020
- [39] Hofreiter, M. et al. (2021) Progress in forensic bone DNA analysis: Lessons learned from ancient DNA. Forensic Sci. Int. Genet. 54 (102538)
- [40] The Times (2020-03-30) Police won't use genealogy sites for cold cases. <https://www.thetimes.co.uk/article/police-wont-use-genealogy-sites-for-cold-cases-vvk0rbhqq> [2021-09-12]
- [41] BFEG rapport (2020-09-09) Should we be making use of genetic genealogy to assist in solving crime? <https://www.gov.uk/government/publications/use-of-genetic-genealogy-techniques-to-assist-with-solving-crimes/should-we-be-making-use-of-genetic-genealogy-to-assist-in-solving-crime-a-report-on-the-feasibility-of-such-methods-in-the-uk-accessible-version> [2021-09-12]
- [42] Scudder, N. et al. (2020) Operationalising forensic genetic genealogy in an Australian context. Forensic Sci. Int. 316 (110543)
- [43] Parabon Nanolabs, <https://snapshot.parabon-nanolabs.com/genealogy> [2021-09-12]
- [44] Verogen, <https://verogen.com/applications/forensic-casework/> [2021-09-12]
- [45] Bode Technology <https://www.bodetech.com/pages/about-bode-fgs> [2021-09-12]

nfc.polisen.se

Nationellt forensiskt centrum – NFC
581 94 Linköping, Telefon 010-562 80 20

Swedish National Forensic Centre – NFC
SE-581 94 Linköping, Tel +46 10 562 80 20
e-mail registrator.nfc@polisen.se

