Using genetic genealogy databases in missing persons cases and to develop suspect leads in violent crimes

Debbie Kennett
Research Department of Genetics, Evolution and Environment, University College London, Gower Street, London WC1E 6BT, United Kingdom.

E-mail: debbie.kennett@ucl.ac.uk.

Highlights

- Genetic genealogy databases produce investigative leads in cold and active cases.
- Many people in the databases have not given informed consent for law enforcement use.
- There is an unknown number of minors in the databases.
- You are affected if you have a relative in the databases even if you have not tested.
- International interdisciplinary best practice guidelines are needed.

Abstract

In the last year direct-to-consumer (DTC) genetic genealogy databases have been used to identify suspects and missing persons in over fifty cold cases, many of which have been unsolved for decades. Genealogists worked on these cases in collaboration with law enforcement agencies. Raw DNA data files were uploaded to the genealogy websites GEDmatch and FamilyTreeDNA, and identification was made by tracing the family trees of relatives who were predicted to be close genetic matches in the database. Such searches have far-reaching consequences because they affect not just those who have consented to upload their DNA results to these databases but also all of their relatives, regardless of whether or not they have taken a DNA test. This article provides an overview of the methods used, the potential privacy and security issues, and the wider implications for society. There is an urgent need for forensic scientists, bioethicists, law enforcement agencies, genetic genealogists and other interested parties to work together to produce international guidelines and policies to ensure that the techniques are used responsibly and effectively.

Keywords

cold cases; DTC genetic testing; ethics; genetic genealogy, genetic privacy; security

Introduction

In April 2018 the news broke that two long-standing cold cases in the US had potentially been solved by the use of GEDmatch, a genetic genealogy database well known to genealogists, but which had previously attracted little attention from other researchers.

Buckskin Girl, named after the distinctive jacket she was wearing, was a murder victim and all previous attempts to identify her had failed over the last 37 years. The case was taken on by the DNA Doe Project, a new not-for-profit organisation which is using DNA to identify missing persons. They uploaded a DNA data file to GEDmatch and, after receiving a first cousin once removed match in the database, a potential identification was made within a matter of hours.[1] The DNA Doe Project has since gone on to make identifications in nine other cold cases.[2]

The Golden State Killer was a serial killer and rapist who went on a ten-year crime spree in California between 1976 and 1986 but had evaded detection. Genetic genealogist Barbara Rae-Venter, working with the FBI and law enforcement officers from Contra Costa County District Attorney's office, was able to identify a suspect after a DNA data file was uploaded to GEDmatch, though in this case the matches were more distant and it took several thousand hours of genealogical detective work.[3–6] Rae-Venter had previously worked with law enforcement agencies and used genetic genealogy techniques to identify an abducted girl, Lisa Jensen, and serial killer, Terry Peder Rasmussen, who was responsible for the Allenstown murders in New Hampshire.[7,8] The use of genetic genealogy in these cases received little coverage at the time and it was the Golden State Killer case which brought the technique to public attention. Rae-Venter has since worked on multiple additional cases, both perpetrator and unidentified victim, including some actives cases. She has also helped to train law enforcement agents.[9–11]

In May 2018 Parabon NanoLabs announced the launch of a forensic genetic genealogy service.[12] By the end of January 2019 they had identified suspects or victims in 25 cold cases.[13] The methodology is described in a paper by Greytak et al.[14]

In January 2019 the DTC genetic testing company FamilyTreeDNA (FTDNA) announced that it was testing samples for the FBI and allowing them to upload profiles to its database.[15] Bode Technology announced the launch of a forensic genealogy service in February 2019.[16]

A survey of 1,587 US residents over the age of 18 found that the majority of respondents supported the police use of genealogy databases to identify perpetrators of violent crimes, perpetrators of crimes against children, and missing persons. The majority of respondents were not in favour of such usage to identify perpetrators of non-violent crimes. [17] Since then, genetic genealogy databases have been used to identify the mothers of two abandoned babies and for some people this is a step too far.[18,19]

The use of consumer DNA databases by law enforcement has sparked a privacy debate, and there have been calls for regulation and oversight. [20] [21] [22] [23] [24]. In order to have an informed debate on the subject and to shape public policy internationally it is important to have an understanding of the methodology involved and to appreciate the implications. This article attempts to lay the foundations for further discussion.

What is genetic genealogy?

Genetic genealogy is the term used to describe the combination of genealogical research with DNA records to form conclusions about relationships.[25] There is a long history of interdisciplinary research on genetic genealogy based on Y-chromosomal DNA (Y-DNA) data, including in forensic research.[26,27] Mitochondrial DNA (mtDNA) also has a long tradition in forensic science and has been used in combination with genealogical techniques in a number of identification studies, most notably that of Richard III.[28] The first DTC Y-DNA and mtDNA tests became available in the year 2000.[29] Y-DNA testing has inspired a "genetic genealogy revolution" with many amateur family historians running their own surname studies and contributing to the scientific discovery process.[30]

The first autosomal SNP DNA tests for genealogy purposes became available in 2009 with the launch of 23andMe's Relative Finder feature.[31] The methodology behind the feature

was published in 2012.[32] An autosomal DNA test provides the user with a list of DNA matches and a prediction of the possible relationship or range of relationships based on the amount of DNA shared. Relationships can be reliably predicted within relatively narrow ranges up to about the second cousin level but become more difficult to predict for more distant relationships because of the increasing spread of values around the mid-point estimate for ever more distant relationships.

Autosomal DNA relative-matching tests are now also offered by AncestryDNA, FTDNA (the Family Finder test) and MyHeritageDNA. Living DNA, a UK-based company is currently beta testing a relative-matching feature known as Genetic Networks.[33] The methodology used by these companies has not been subjected to peer review. AncestryDNA has published a white paper explaining the matching process.[34] FTDNA has provided some technical information in its Learning Center.[35] MyHeritage has described its methods in a blog post.[36] Each company uses different methods for IBD (identical by descent) detection and sets different match thresholds. As a result, the relationship predictions can vary from company to company.

As a check on the relationship predictions provided by the companies, genealogists led by Blaine Bettinger have set up a collaborative citizen science project known as the Shared cM (centimorgan) Project.[37] By August 2017 the project had collected empirical data on over 25,000 known genealogical relationships.[38] Problems of data input errors and misattributed relationships are offset by the large number of submissions.

The power of genetic genealogy lies in the comparison process and the ability to search for genetic matches in a database.[39] As the cost of testing has come down and the DTC databases have grown in size, DNA testing has become an increasingly useful tool for genealogy. By February 2019 it was estimated that more than 26 million people had taken a DTC genetic test.[40] By 2021 there are likely to be over 100 million people in the DTC databases.[41]

The growth of the DTC DNA databases has been accompanied by a corresponding increase in the availability of online genealogical records. Large quantities of records from around the world have been digitised and indexed and made freely available online on the FamilySearch website. Subscription websites such as Ancestry.com, Findmypast, Geneanet and MyHeritage provide access to censuses, birth, marriage and death records, electoral registers, newspaper articles and a variety of other historical records from many different countries. The FamilySearch Wiki (https://www.familysearch.org/wiki) provides a useful guide to the availability of records around the world. There are many websites which allow users to upload family trees, and trees are often uploaded to the DNA websites as well. In addition, Facebook and other social media sites can be used, along with people search websites (eg, BeenVerified, Intelius, 192.com) and directories, to find information about living people. Research which once used to take months or years to do by visiting the repositories in person can now be done at home on a computer in a matter of hours.

The growth of the genetic genealogy databases, combined with the greater availability of genealogy records has been a boon for genealogists, and has also been instrumental in helping people with unknown parentage, such as adoptees, donor-conceived individuals and foundlings, to identify their biological parents.[42–44] Success stories, mostly in the US, have attracted a lot of media attention. The methods used in unknown parentage searches are equally applicable to identify criminals and missing persons and it was therefore inevitable

that law enforcement agencies would seek to access the genetic genealogy databases if no matches could be found in their own databases.

Y-STR databases

Y-chromosome STRs (short tandem repeats) can be used to find paternal relatives in a genealogy or forensic database. Y-STR haplotypes have the advantage that more distant relationships can be identified than those with autosomal genetic markers. However, the STR tests currently used both in forensics and genealogy contain some rapidly mutating Y-STRs and can reveal differences between closely related paternal relatives. The feasibility of using Y-STRs to predict surnames was investigated in a UK population by King et al in 2006. The approach was found to be most successful with rare surnames, which constitute around 42% of the UK population.[45] The link between the surname and the Y-chromosome was weak for more common surnames. A specific rare surname of interest is less likely to be represented in a DNA database so the approach would be most useful for intermediate frequency surnames.[30] Gymrek et al looked at the possibility of inferring surnames from the previously publicly available Y-STR databases, Ysearch (www.ysearch.org) and SMGF (www.smgf.org). These databases contained ~135,000 records representing ~39,000 unique surnames. They had a success rate of ~12% in recovering the surnames of US males of European ancestry, but in 5% of cases the wrong surname was returned and in 83% of cases no inference could be made.[46]. King and Jobling found that there are some common Y-STR haplotypes shared across many different surnames.[30] Other researchers have noted the finding of convergent haplotypes, belonging to different subhaplogroups (branches of the Y-chromosome tree).[47–49] These haplotypes would likely share a common ancestor over a thousand years ago prior to the adoption of surnames, thus confounding attempts to infer surnames. Genealogists using Y-STR testing in their surname projects generally use a 37-Y-STR test as a starting point, and can increase the number of markers up to 67, 111 or 700 STRs if need be.[50] In contrast, the Y-profiling kits for forensic use cover between 17 and 27 Y-STRs.

In view of the above statistics, it is not surprising that attempts to use Y-STR matches in genetic genealogy databases have not been very successful. Genetic genealogy Y-STR databases were first used by law enforcement in 2012 to investigate the 1991 murder of Sarah Yarborough.[51] A Y-STR match was said to indicate that the suspect's surname was Fuller. In an awkward coincidence there was a William Fuller who was a colleague of the victim's father. His daughter was the best friend of the victim. His name was mentioned in newspaper reports but he was the wrong age to be a suspect and had no family members who fitted the profile. The killer has never been found.[52] The Yarborough case sparked a debate in the genetic genealogy community about the reliability of the methods used and ethical concerns about the surreptitious way in which attempts were made to obtain genealogy information.[53]

In 2015 Michael Usry was targeted as a suspect in the 1996 murder of Angie Dodge in Idaho Falls, Idaho, after a partial Y-STR match (34/35 markers) was found between DNA from a semen sample taken from the crime scene and a profile uploaded by his father to the public SMGF database. Usry was subsequently eliminated from the enquiry after providing a DNA sample, but suffered the stress of waiting for nearly a month to receive the results before his name could be cleared.[54] The fallout from this case resulted in the closure of the SMGF database.[55] However, public Y-STR databases were successfully used in 2016 to identify the surname of a suspect in a 1991 cold case in Phoenix, Arizona.[56]

In the Golden State Killer case, investigators initially targeted a 73-year-old man in a nursing home in Oregon after identifying a 12-STR match in the public Ysearch database which included one rare marker in common with the suspect's DNA. FTDNA, the company who sponsored Ysearch, received a subpoena in March 2017 forcing them to reveal the identity of the customer and the payment method. The kit had been paid for by the man's daughter, but she was not informed of the investigators' decision to take a DNA sample from her father.[57]

Ysearch was shut down by FTDNA in May 2018 because of concerns about the European Union's General Data Protection Regulation.[58] AncestryDNA stopped offering Y-DNA tests in 2014 and shut down their Y-STR matching database.[59]. FTDNA hosts the only Y-STR matching database for genealogists and, as of April 2019, they have over 700,000 Y-STR records in their database.[60] FTDNA now allows law enforcement uploads and their database could potentially have much greater predictive power than the public databases previously used in forensic investigations.

If genealogical Y-STR databases are to be used by law enforcement agencies, it is important that the relevance of the match is not overstated and the uncertainty of the inference is quantified to avoid implicating private individuals unnecessarily in criminal investigations. There are well established statistical methods based on haplotype frequencies for estimating the weight of evidence of a Y-STR match which have been implemented in the forensic Y-chromosome haplotype reference database (YHRD).[61] Anderson and Balding have proposed a new technique to estimate the number of males with a Y-profile in a population based on population genetics methods. [62] These methods could potentially be adapted to interpret matches in genealogical Y-STR databases.

Forensic versus genetic genealogy autosomal DNA tests

Forensic DNA testing primarily uses autosomal STRs (short tandem repeats). The number of markers tested varies by jurisdiction, but current tests use between 16 and 23 autosomal STRs.[63,64] As of 30 September 2018, the UK National DNA Database had DNA samples from 5,436,235 individuals representing about 8% of the population.[65] In September 2018 the FBI's CODIS database had DNA profiles for 16,809,115 offenders and arrestees representing about 4% of the US population.[66] A familial search in one of these databases can most likely at best identify a potential sibling, parent, or child of the target.[67]

DTC genetic tests are done on DNA microarrays (chips) which sample several hundred thousand autosomal SNPs (single nucleotide polymorphisms) scattered across the genome. The companies have used different chips over the years, such as the Illumina OmniExpress and the Global Screening Array, with the number of SNPs varying from 550,000 up to about 900,000. The use of genome-wide SNP chips allows the companies to provide predictions of relationships for second, third and more distant cousins.[32,68]

The genetic genealogy market is dominated by four companies: AncestryDNA has now tested nearly 15 million people,[69] 23andMe has tested over ten million,[70] MyHeritage has tested 2.5 million people, and FTDNA is reported to have an ancestry database of 2 million people.[40] See Table 1 for a comparison of consumer and law enforcement databases.

Table 1. A comparison of consumer and law enforcement databases as of April 2019

Database	**Database size**	**Markers**	**Usage**
CODIS	16 million	STRs	Law enforcement

AncestryDNA	15 million	SNPs	Consumers
UK DNA Database	5.5 million	STRs	Law enforcement
23andMe	10 million	SNPs	Consumers
MyHeritage	2.5 million	SNPs	Consumers
FamilyTreeDNA*	2 million	SNPs	Consumers/law enforcement
GEDmatch	1 million	SNPs	Consumers/law enforcement

*FTDNA does not publish the size of their autosomal DNA database and this figure includes Y-DNA and mtDNA records, many of which are at low resolution.

In order for law enforcement agencies to use genealogy databases they have to get the forensic samples re-tested on a SNP chip. There are currently only a small number of companies that provide such a service. Parabon NanoLabs and Bode Technology offer microarray testing as part of their forensic genealogy services. Two other US companies, DNA Solutions and Gene By Gene, can do SNP microarray testing on forensic samples.[71] TothelettterDNA, a new Australian start-up company, is also offering a forensic genotyping and genetic genealogy service.[72] These methods currently have limited use in forensics and will be mostly restricted to high-level sources such as semen and saliva. Semen samples from sexual homicides are likely to be the best sources.[73]

The DNA Doe Project uses whole genome sequencing. The bioinformatics team then produces a SNP dataset of up 900,000 SNPs for upload to GEDmatch. [1,74] Proprietary methods are used to characterise the level of degradation and to assess the reliability of the matches.[75]

While the commercial autosomal DNA relative-matching tests have essentially been validated by usage by millions of genealogists, the methodologies have not been validated for forensic use. Forensic samples are likely to be degraded, producing a large a number of no calls, and it is not known what impact this will have on the relationship predictions. The proprietary techniques used by Parabon, the DNA Doe Project and the other companies are still experimental and have not been subjected to peer review, creating concerns about transparency and accountability.[67] [76] However, it is important to note that genetic genealogy is intended to be used to generate investigative leads and not for conviction.[13] Once the possible name of the suspect or missing person has been identified, standard forensic DNA testing is carried out to see if there is a match between the person of interest and the crime scene sample. It is also not known how the investigators were able to demonstrate that the DNA sample submitted for analysis was actually that of the person of interest, and what other evidence was available to support the identification. None of the cold cases where DTC databases were used have as yet come up for trial and it is possible that the methodology will come under scrutiny at that time.

What is GEDmatch?

GEDmatch is a privately owned website run as a hobby by two genealogists, Curtis Rogers and John Olsen, providing a range of tools for analysing and interpreting DNA matches and genealogy data. The site started out by providing a tool for comparing names in family trees to help with the interpretation of Y-DNA matches in surname projects, and the domain name was first registered in 2010.[77] GEDmatch grew by word of mouth and additional tools were added by user request to help with the analysis of autosomal DNA matches. The site now provides a variety of sophisticated tools (Figure 1). GEDmatch does not do any DNA testing of its own but accepts uploads of raw DNA data files from all the major testing companies, thus allowing users to make cross-platform comparisons between tests taken at different

companies. For example, if one relative has tested at AncestryDNA and another has tested at 23andMe, their results can be compared at GEDmatch. The data on GEDmatch therefore represents a publicly available subset of the data from each of the commercial companies.

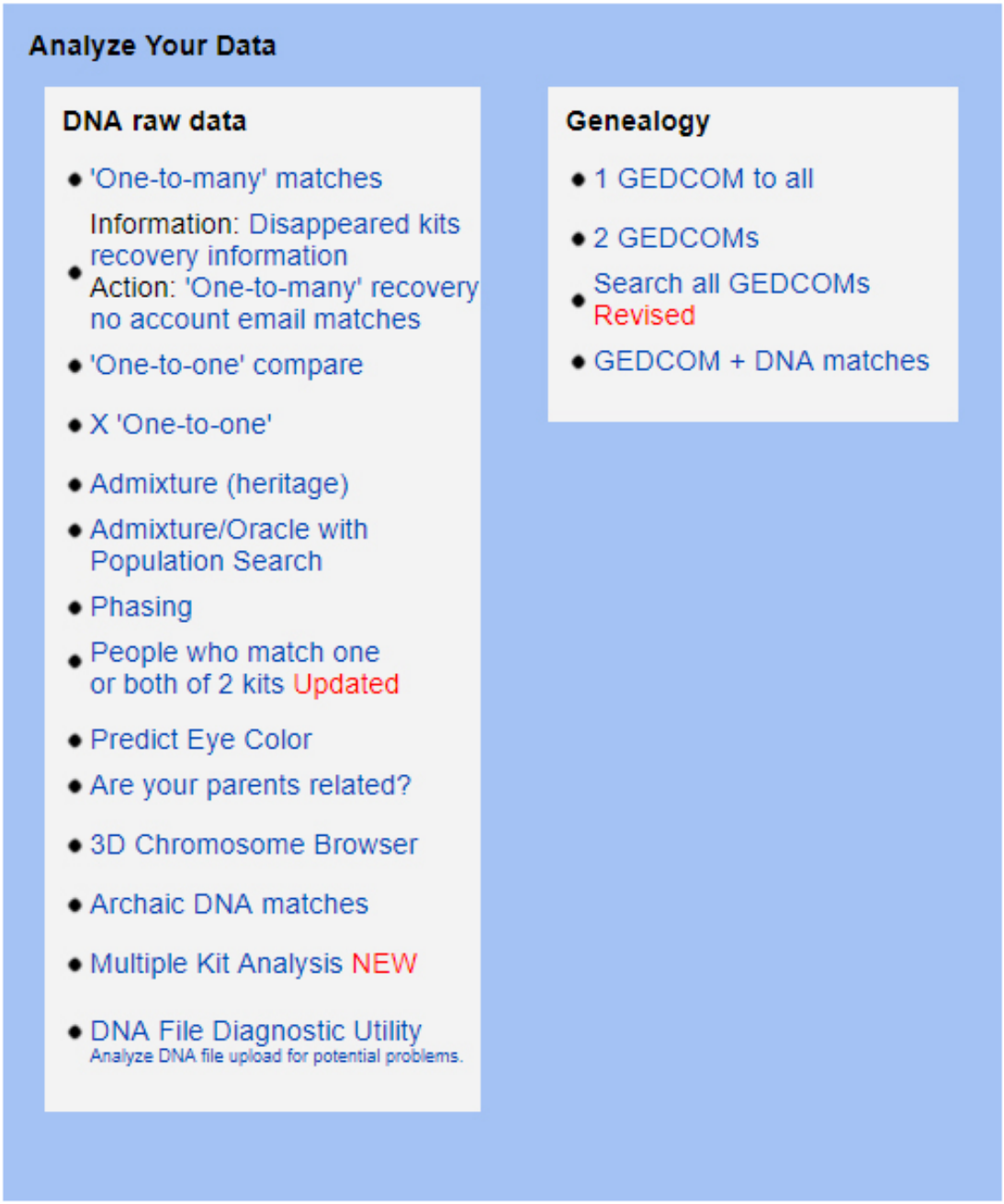


Figure 1. A screenshot showing the range of free tools available at GEDmatch. Additional tools are available with a Tier 1 subscription.

By November 2018 there were around one million people in the GEDmatch database with 1,800 users added every day.[77] The database is not publicly available. Users are required to set up a password-protected account and upload a raw DNA data file in order to view their matches. The site is free to use but there is a premium service costing $10 a month which provides access to additional tools for advanced users.[78] The site currently has over 7,000 subscribers and the subscription money is used to pay the hosting and server costs of $200,000 a year.[79]

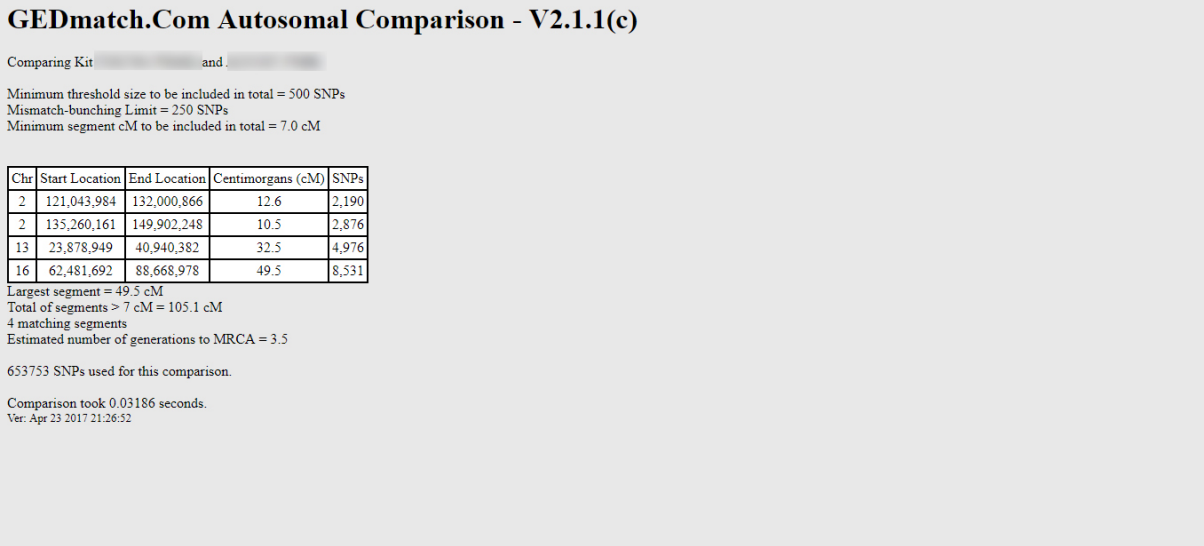
When uploading kits to GEDmatch there are three options: public, private and research mode.[80] The public option is the preferred mode for genealogists. It allows you to see your matches and allows your matches to see you. The kits uploaded to GEDmatch by law enforcement agencies use the research mode which means that the investigators can see the matches, but the kit will not be visible to other users.

The free one-to-many match tool provides the user with a list of the top matches in the database (Figure 2). In December 2018 GEDmatch launched a new Genesis database (https://genesis.gedmatch.com) and uploads to the old website were frozen.[81] GEDmatch Genesis allows comparisons between tests done on a wider range of chips, such as the Illumina Global Screening Array. With both the standard and Genesis databases, information is provided on the user name (either the real name or an alias) and e-mail address of the match and the amount of DNA shared.

						Haplogroup		Autosomal				X-DNA				
Kit Nbr	Type	List	Select	Sex	GED/WikiTree	Mt	Y	Details	Total cM	largest cM	Gen	Details	Total cM	largest cM	Name (* => alias)	Email
[redacted]	F2	L		F				A	106.7	49.5	3.5	X	6.2	6.2	[redacted]	[redacted]
[redacted]	F2	L		M				A	96	47.6	3.6	X	0	0	[redacted]	[redacted]
[redacted]	F2	L		M				A	52.3	25.0	4.1	X	0	0	[redacted]	[redacted]
[redacted]	F2	L		F				A	46.5	27.4	4.1	X	6.9	6.9	[redacted]	[redacted]
[redacted]	F2	L		M	GED			A	40.1	21.3	4.2	X	0	0	[redacted]	[redacted]
[redacted]	F2	L		F				A	31.5	21.2	4.4	X	0	0	[redacted]	[redacted]
[redacted]	F2	L		F				A	31.1	20.7	4.4	X	0	0	[redacted]	[redacted]
[redacted]	F2	L		F				A	30.9	25.7	4.4	X	0	0	[redacted]	[redacted]
[redacted]	F2	L		M	GED			A	29.7	29.7	4.5	X	0	0	[redacted]	[redacted]
[redacted]	F2	L		M				A	29.3	29.3	4.5	X	0	0	[redacted]	[redacted]
[redacted]	F2	L		M				A	28.7	15.8	4.5	X	0	0	[redacted]	[redacted]
[redacted]	V4	L		M		H3	I1	A	28.7	28.7	4.5	X	0	0	[redacted]	[redacted]
[redacted]	F2	L		U				A	28.5	22.6	4.5	X	7.3	7.3	[redacted]	[redacted]
[redacted]	F2	L		M	GED			A	28.1	17.7	4.5	X	0	0	[redacted]	[redacted]
[redacted]	F2	L		F				A	27.9	15.7	4.5	X	0	0	[redacted]	[redacted]
[redacted]	F2	L		M	GED		I-M253	A	27.6	27.6	4.5	X	0	0	[redacted]	[redacted]
[redacted]	F2	L		F				A	27	16.8	4.5	X	11.5	6.4	[redacted]	[redacted]
[redacted]	F2	L		F				A	26.9	14.2	4.5	X	0	0	[redacted]	[redacted]
[redacted]	F2	L		F				A	26.8	19.4	4.5	X	11.9	6.1	[redacted]	[redacted]
[redacted]	F2	L		M	GED			A	26.3	11.3	4.5	X	0	0	[redacted]	[redacted]
[redacted]	F2	L		F		K1a1b1f		A	29.3	9.9	4.6	X	0	0	[redacted]	[redacted]

Figure 2. The one-to-many matches tool at GEDmatch. "Type" refers to the type of kit uploaded to GEDmatch. F2 is an FTDNA, Ancestry or MyHeritage kit. V2, V3 and V4 are 23andMe kits. Clicking on "L" will provide a list of the one-to-many matches for that person. Clicking on A will give a one-to-one autosomal comparison showing the shared segments of DNA. Clicking on "X" will generate an X-chromosome comparison.

Some users provide a family tree in a GEDCOM file format or a link to a public family tree on Wikitree (https://www.wikitree.com), a free public website where genealogists collaborate to work on a single family tree. The one-to-one comparison tool at GEDmatch allows the user to see the chromosomal locations and sizes of the shared DNA segments (Figure 3). It is also possible to click through and see the match list of the people you match. By clicking on the matches in this way it is possible to access the match lists of large numbers of the people in the GEDmatch database who have opted in to sharing. In addition, many genealogists have shared their GEDmatch kit numbers in Facebook groups so once you have a GEDmatch ID it is possible to obtain a large amount of data about other users.



GEDmatch.Com Autosomal Comparison - V2.1.1(c)

Comparing Kit [redacted] and [redacted]

Minimum threshold size to be included in total = 500 SNPs
Mismatch-bunching Limit = 250 SNPs
Minimum segment cM to be included in total = 7.0 cM

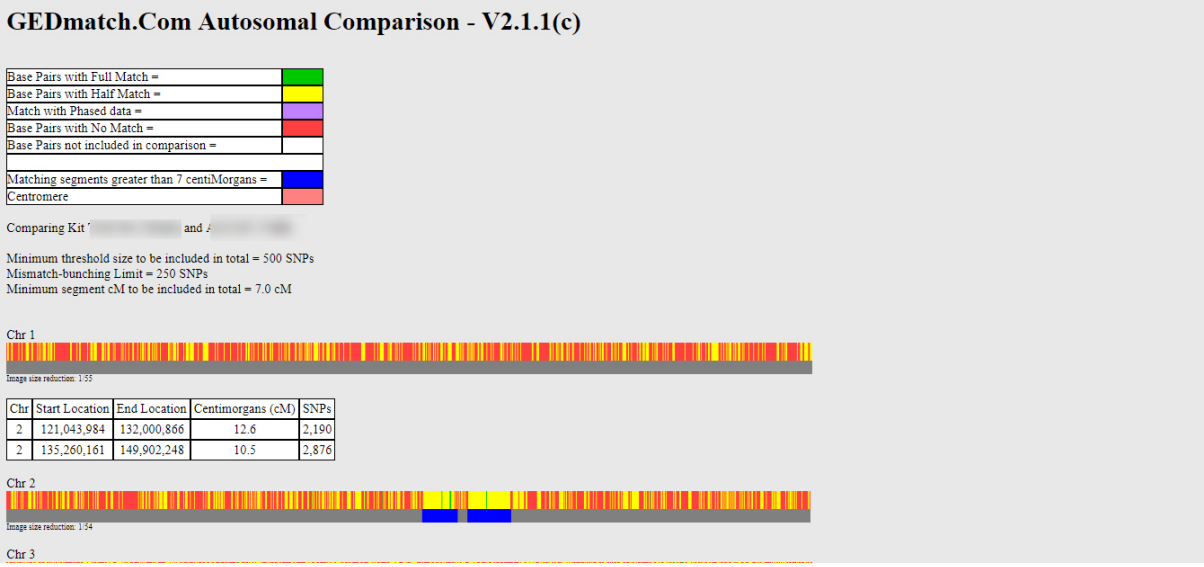
Chr	Start Location	End Location	Centimorgans (cM)	SNPs
2	121,043,984	132,000,866	12.6	2,190
2	135,260,161	149,902,248	10.5	2,876
13	23,878,949	40,940,382	32.5	4,976
16	62,481,692	88,668,978	49.5	8,531

Largest segment = 49.5 cM
Total of segments > 7 cM = 105.1 cM
4 matching segments
Estimated number of generations to MRCA = 3.5

653753 SNPs used for this comparison.

Comparison took 0.03186 seconds.
Ver: Apr 23 2017 21:26:52

Figure 3. The one-to-one comparison tool at GEDmatch allows the user to see how much DNA is shared between two people, the location of the shared segments on the individual chromosomes and the number of SNPs in each segment. The data can be viewed in a table (A) (above) or in a visual browser (B) (below) which provides detailed information about the sharing at the base-pair level.



GEDmatch.Com Autosomal Comparison - V2.1.1(c)

Base Pairs with Full Match =	
Base Pairs with Half Match =	
Match with Phased data =	
Base Pairs with No Match =	
Base Pairs not included in comparison =	
Matching segments greater than 7 centiMorgans =	
Centromere	

Comparing Kit [redacted] and [redacted]

Minimum threshold size to be included in total = 500 SNPs
Mismatch-bunching Limit = 250 SNPs
Minimum segment cM to be included in total = 7.0 cM

Chr 1

Image size reduction: 1:55

Chr	Start Location	End Location	Centimorgans (cM)	SNPs
2	121,043,984	132,000,866	12.6	2,190
2	135,260,161	149,902,248	10.5	2,876

Chr 2

Image size reduction: 1:54

Chr 3

Figure 3B

When news first broke that GEDmatch had been used by law enforcement agencies in the Golden State Killer case, many people expressed concern that the database was being used without the informed consent of the participants.[82–85] The owners themselves had not been notified in advance of this usage. However, the GEDmatch site policy, which was introduced on 18 August 2017, was very broad and, although it did not specifically permit access by law enforcement agencies, it did anticipate unexpected uses:

> "While the results presented on this site are intended solely for genealogical research, we are unable to guarantee that users will not find other uses. If you find the possibility unacceptable, please remove your data from this site."[86]

On 28 April 2018 GEDmatch posted a notice on the website to alert users to the use of the database by law enforcement agencies.[87]

GEDmatch updated the Terms of Service and Privacy Policy on 20 May 2018 to specifically allow law enforcement usage but only to "(1) identify a perpetrator of a violent crime against another individual; or (2) identify remains of a deceased individual". [88] They also took the additional step of e-mailing all users who had not been on the site for a while to alert them to the new ways in which the website was being used.[89] The decision to allow law enforcement use of the website was in part a recognition that GEDmatch "doesn't have the legal resources to prevent the site being used this way". Curtis Rogers, the co-founder of GEDmatch, noted "We could have published a statement that says we require a court order for police to use our site, but it's only words… They could go ahead and do it, and they probably would."[90] Although Rogers was initially "outraged" at the use of his website by law enforcement agencies he now feels "proud".[77]

Although the majority of GEDmatch users now appear to be aware that law enforcement agencies can use the database in specified scenarios, it is likely that many users do not fully understand the potential privacy issues. In addition, many genealogists upload raw data files on behalf of other family members, and it is possible that in some cases fully informed consent will not have been obtained. GEDmatch's site policy prohibits the use of the database by children under 13, but has no restrictions on minors aged between 13 and 18. Moray et al showed that many genetic ancestry companies do not properly address the issue of the testing of minors and that fathers can potentially use DTC databases for secret paternity testing.[91] There is therefore a risk that minors will be included in the GEDmatch database and included in law enforcement matching without their consent.

The vast majority of GEDmatch's users are in the US. This is mainly because of the way that the genetic ancestry testing market has developed. AncestryDNA, the market leader, launched their test in the US in 2012, three years before making it available in a limited number of other countries. It is now sold in 36 countries. The 23andMe test is sold in just over 50 countries. The other companies ship to most countries in the world. According to GEDmatch the other countries apart from the US where their service is most popular are Canada, Australia and the UK.[92] It is therefore possible that law enforcement agencies in these countries and elsewhere will consider using GEDmatch, though it is likely to be a few more years before the database has enough participants outside the US to make such searches worthwhile. The police in Sweden have already been given the go ahead to test similar methods.[93]

FamilyTreeDNA and the FBI

In January 2019 FTDNA announced that they were collaborating with the FBI and allowing them to upload DNA profiles and create accounts with the same level of access as ordinary users. Existing customers could choose to opt out of matching but this would mean that they would not benefit from the services they had paid for.[15] It was later revealed that the FBI had already been accessing the FTDNA database for an undetermined time without the company's knowledge.[94] Following a public backlash, FTDNA introduced an opt out from law enforcement matching in March 2019.[95] They then went on to launch a TV advertising campaign in the US to encourage people to upload their DNA profiles to the database to catch criminals.[96] However, concerns still remain about the lack of informed consent for participation in law enforcement matching.[97] FTDNA has an international database and it has been suggested that non-US customers should be required to opt in rather than opt

out.[98] As with GEDmatch, there are also concerns about the participation of minors in law enforcement matching. FTDNA's terms of service state that participants must be 13 years of age to participate in the database. Minors between the ages of 13 and 18 can only be tested with the permission of a parent or guardian.[99] It is not known how or if these terms are enforced.

FTDNA have published law enforcement guidelines.[100] They are currently only accepting uploads from agencies in the US.[101] FTDNA previously published a transparency report.[102] An updated report has been promised.

Other consumer DNA databases

The other genetic genealogy testing companies have all firmly resisted law enforcement access.[103]

My Heritage accepts uploads from other testing companies but does not allow law enforcement agencies to use its database without a court order or valid legal documentation.[104] However, it is quite possible that the raw data files could be manipulated for upload and the company would not realise that they were processing files from non-standard sources. Erlich et al have proposed that the testing companies should use a cryptographic signature so that GEDmatch and other third-party services could authenticate that the file was created by a legitimate provider.[46]

AncestryDNA and 23andMe do not accept transfers, and do not permit law enforcement agencies to access their databases unless required by valid legal process.[105,106] Both companies use saliva kits which would make it harder for illicit uses of the databases because of the quantity of saliva required. However, methods have been developed to produce artificial saliva in order to submit a DNA sample for those who have problems producing enough spit.[107] 23andMe and AncestryDNA publish guides for law enforcement agencies and transparency reports revealing details of requests for data.[108–111]

It is not clear how the companies are able to monitor or enforce their terms and conditions, either in their own jurisdictions or in other countries, and whether or not there are penalties for breaching them. It is quite possible that the policies could be breached without their knowledge, especially given that the FBI previously uploaded profiles to both GEDmatch and FTDNA prior to the changes in the terms and conditions which explicitly allowed such uploads.

Privacy and security implications

The autosomal STR markers used for forensic testing are chosen to provide enough information to identify a suspect but without revealing any personal information such as predisposition to disease.[67] In contrast, genome-wide microarray data are vastly more informative and have the potential to reveal sensitive information about a suspect or a victim and his or her family members. However, the raw genetic data is not disclosed to law enforcement agencies, and it is only the amount of DNA shared and the length of the shared segments that are used to infer relationships.[13] Nevertheless, the genotype data can sometimes reveal sensitive information, and in some cases the raw data has been made available to the genealogists working on the analysis.

In the Golden State Killer case, the suspect's genotype data was uploaded to Promethease, a literature retrieval service, one of a number of third-party services which can be used to

obtain health risk reports.[112] Promethease only appears to have been used to obtain trait reports about eye colour and propensity to baldness.[9,71] In theory such sites could also be used to look at a suspect's propensity for various diseases. However, the predictive power of genetic testing is very weak,[113] and such reports are unlikely to have any practical application in a criminal investigation. If an investigator is able to identify through Promethease that the suspect is a carrier for a disease such as cystic fibrosis then it would also be possible to make inferences about the people who match on the same segment at GEDmatch,[114], though again such a finding is unlikely to have any practical utility in an investigation.

GEDmatch has an Are Your Parents Related tool which is often used at the start of an unknown parentage search because consanguineous relationships and endogamy can affect the interpretation of the matches. As a result, this tool could reveal a history of incest in the family of a suspect or victim.[115]

The possibility of privacy breaches at GEDmatch was predicted in 2014 by Erlich and Narayanan.[116] Ney et al have identified a number of security risks of relative matching in DTC DNA databases including denial of service attacks and the creation of synthetic relatives to corrupt the database and thwart investigations.[117] DTC databases might also be the target of hacking attacks. As GEDmatch now enjoys a high profile as a result of the media publicity, it is potentially a target. As a small private company with limited resources, it is less likely to be equipped to prevent such attacks or to mitigate against the consequences. These possibilities create major governance issues. A data breach could potentially compromise an investigation, but would also be a privacy invasion for the users of the databases.

Genealogy is largely conducted as a hobby but some hobbyists turn professional and charge for their services, and there are now some university courses which offer genealogical training.[118] There has been a long-standing debate within the genealogical community about whether or not professional genealogists should be accredited, and there is no consensus. Although there are some professional organisations, they have limited powers to enforce standards, and there are many professional practitioners who do not belong to any organisation. There are many amateur family historians who do excellent work, but there are also genealogists who charge for their services who produce poor-quality work. Few professional genealogists are able to make a living solely from genealogy work unless they work for a major company.[119] A survey of family historians in the US who had paid for professional services found that only around a third of respondents were satisfied with the services they had received, including services from genealogists who were accredited with one of the two credentialing agencies in the US.[120] The application of DNA testing to genealogical research is a relatively new discipline, and there is no formal training or accreditation for genetic genealogists. There are currently very few genetic genealogists with the necessary skills to perform this type of work for law enforcement agencies. If there is public pressure for the police to incorporate genetic genealogy into their routine forensic work, there are likely to be many genealogists coming forward to offer their services, but the agencies will not have any reliable way of judging their skills and expertise. This lack of professionalisation and accountability increases the risk of ethical and privacy breaches.

Furthermore, much of the work in cold case investigations is being done by volunteers. For example, the DNA Doe Project relies on a team of volunteers to work on the cases. The use of volunteers is inherently risky because of the lack of accountability and the potential for

leaks, such as the disclosure of the victim's GEDmatch ID or information on whether an individual has a match with the missing person. Many missing persons cases involve victims of crime. There is a possibility that the victim or their family might be known to one of the volunteers, which could potentially compromise an investigation, especially if a family member is a suspect.

There is also the risk of misidentification. In adoption searches volunteers have sometimes identified the wrong birth parent.[9] Incorrect family trees could cause confusion and confound investigations. Broken lineages in genealogies due to hidden adoptions or misattributed parentage could potentially lead to incorrect conclusions, putting innocent people under suspicion. Although the historical rate of misattributed paternity is low at around 1% per generation, it is likely to be higher in more recent times, especially in cosmopolitan populations.[121]

The investigative leads generated by genetic genealogy are always followed up by CODIS testing, but a misidentification could potentially breach the privacy of innocent people by putting them under suspicion or surveillance and subjecting them to police intrusion into their private lives. Misidentification of a murder or suicide victim could result in stress and anguish for the family members.

Privacy breaches could also occur as a result of the inadvertent disclosure of information by the police or the media. In one recent case, the name of an unwitting "genetic informant" was revealed in a search warrant and published in the media.[122]

However, it is the combination of genetic data with online genealogy records and social media which can be the most intrusive aspect of an investigation. In a novel experiment, the journalist Peter Aldhous tried to identify ten colleagues at Buzzfeed news from profiles uploaded to GEDmatch under a pseudonym. He successfully identified six of his ten colleagues but felt unease when trawling through intimate Facebook profiles to make the identifications.[57]

Surreptitious collection of DNA

In most of the cold cases where genetic genealogy has been used, law enforcement officers obtained discarded DNA from the suspects, which generally required putting them under surveillance for a few days. Surreptitious collection of DNA by law enforcement agencies is permissible in most US states without a warrant as long as there is probable cause. The rationale is that the DNA has been voluntarily discarded and there is therefore "no reasonable expectation of privacy".[21,76] In the cases reported to date it is unclear how many innocent individuals had their DNA sampled surreptitiously. The legality of surreptitious DNA testing in the US is unsettled and it has been suggested that the adoption of a DNA theft offence would "help clarify the appropriate Fourth Amendment characterization of genetic information that everyone sheds involuntarily".[123] Murphy has also suggested that there should be clearer rules about surreptitious sampling.[67] DNA theft is already a criminal offence in the UK as a result of the Human Tissue Act, but this legislation does not apply to law enforcement.[124] It is unclear whether a genealogical link would be sufficient justification in the UK for taking a DNA sample without consent.

Paradoxically, although it is permissible to take surreptitious DNA samples in most US states, there are many states where the police do not have an automatic right to take DNA from arrestees [125] or from convicted offenders.[125,126] This anomaly created a curious

situation in the case of Michael Henslick, who was identified through genetic genealogy as being a suspect in the 2009 murder of 22-year-old Holly Cassano of Mahomet, Illinois. The identification was confirmed after the police tested his DNA without his consent from a discarded cigarette butt.[127] However, Henslick had a string of convictions dating back to 2006 and had even spent 30 days in the county jail but the police had been unable to get a DNA sample from him.[128]

Familial searches

Familial searching of the UK National DNA Database was introduced in 2002. Familial searches are also used in a limited number of international jurisdictions, though some researchers have raised concerns about the ethics and governance of this technology. By 2014 familial searches had been used in around 210 cases in the UK and led to the identification of 41 perpetrators or suspects.[129] Some US states have developed policies for familial searches of the CODIS database, and other states are still investigating policy options, but the practice is banned in Maryland.[130]

It is interesting to note that in some cases where genetic genealogy techniques have been deployed, the suspects could potentially have been identified earlier through familial searches. Luke Fleming of St Petersburg, Florida, was arrested in September 2018 for the murder of 47-year-old Deborah Dalzell after being identified as a suspect through genealogy searches. A DNA sample was collected from Fleming using “investigative means”. However, Fleming’s brother was already in the CODIS database.[131] Florida does not have an official policy on the use of familial searching but these methods would have identified Fleming as a suspect many years earlier.

John DeAngelo, the brother of the Golden State Killer suspect Joseph DeAngelo, already had a criminal record.[132] His offences occurred prior to the implementation of Proposition 69 in California which requires the collection of DNA from all felons.[133] California implemented familial searching in 2008. If a policy had been in place earlier to collect DNA upon conviction and to allow familial searching, it might have been possible to identify Joseph DeAngelo as a suspect a decade ago.

It should also be noted that in the US there is a backlog of untested evidence from hundreds of thousands of sexual assault kits. DeLisi (2018) has argued that these kits should be tested as matter of urgency.[134] Speaker (2019) has shown that testing the backlog would provide a major return on investment.[135]

Societal implications

While the individuals who upload their data to GEDmatch are now doing so in the knowledge that their DNA could be used to assist with criminal investigations, there are broader implications for society. Even if you have not taken a DNA test yourself, or if you have tested but have decided not to share your results on GEDmatch, you could still be mixed up in an investigation because your sibling or cousin has tested. As Fullerton and Rohlfs comment “the decisions of individuals to contribute their own genetic information inadvertently exposes many others across their family tree who may not be aware of or interested in their genetic relationships going public.”[136] In some cases the investigators have approached relatives of the suspect for target testing. This usually happens when there are no matches at the second or third cousin level or closer, but the triangulation of the family trees of the more distant matches has generated a potential lead. Further testing of a closer relative is necessary to confirm or refute the hypothesis. This happened in the Golden State

Killer case where a relative of the then suspect was tested. Although the individual did not match as expected, he did turn out to be a second cousin of Joseph DeAngelo, the suspect who has now been arrested.[71] Family members were also voluntarily tested for elimination purposes in the 2001 murder of Christine Franke in Orlando, Florida. Three relatives were identified as matches on GEDmatch, but it was necessary to narrow down the search by voluntarily testing other family members for elimination purposes.[137]

The genetic genealogy databases are global and the effects are therefore far-reaching. With extended family members living around the world, the decision of an individual in one country to take a DNA test could mean that a relative in another country becomes involved in an investigation. In a recent case in Canada, immigration authorities tested the DNA of a refugee by the name of Frank Goodwin at FTDNA. Two of his close relatives in the UK were in the company's database and they were contacted by the authorities in an attempt to determine Goodwin's nationality.[138] In an investigation into the cold case of Annie Doe, a young girl whose skeletal remains were found in 1971 on the Oregon/California border, her matches at GEDmatch indicated that she probably had relatives in New Zealand and the UK. An appeal was launched through the media in New Zealand to encourage people to upload their DNA results to GEDmatch.[139]

But how effective are genetic genealogy searches likely to be? In order to answer this question, Edge and Coop used a simple model to determine how often a close match would be found in a genealogy database. They calculated that in a database of one million people everyone in the database would be likely to have 20 or more matches at the third and fourth cousin level and there was a 25% chance of matching with a second cousin.[140] Erlich et al used empirical data from 1.28 million people in the MyHeritage database, who were primarily Americans of European origin, to calculate the probability of finding a close match. They found that "over 60% of searches for individuals of European-descent would result in a match with a third cousin or closer, and 15% of matches corresponded to a relative who was a second cousin or closer". They also predicted that a genetic database would need to cover just 2% of a population to produce a third cousin match for nearly everyone in the database and that over 40% would be expected to have at least a second cousin match.[46] As the journalist Sarah Zhang has commented "Soon, it won't be hard to imagine a world where everyone can be found for whatever reason through a relative's DNA."[141]

Some researchers have suggested that the best way of protecting the privacy of innocent people and reducing the need for invasive investigative techniques is to implement a universal forensic DNA database.[142,143]

Conclusions

DTC genetic genealogy databases have proved to be a useful tool for generating suspect leads and for identifying crime victims and missing persons. There are outstanding issues relating to informed consent as not all of the people in the databases have actively opted in to participate in law enforcement matching. There are also concerns about the inclusion of minors in the searches. People who have never taken a DNA test could still be involved in an investigation because one of their cousins is in the database. The privacy rights of those who have not committed a crime need to be balanced against the victims' right to justice and the need to protect public safety.[144] The companies use proprietary methods for generating matches and the techniques have not been validated for forensic use. The technology has advanced faster than our ability to introduce safeguards. There is a need for oversight and regulation to ensure transparency and accountability. Forensic scientists, bioethicists, law

enforcement agencies, genetic genealogists and other interested parties should work together to produce international policies and best practice guidelines and to clarify the situations where it is appropriate to use this methodology. The ethical principles devised by the UK's Biometrics and Forensics Ethics Group provide a useful foundation for discussion.[145] These measures will ensure that the technology can be deployed responsibly and effectively for the benefit of society.

Declarations of interest

Declarations of interest: none.

Funding

None.

References

[1] S. Augenstein, "Buck Skin Girl" case break is success of new DNA Doe Project, Forensic Magazine. (2018). https://www.forensicmag.com/news/2018/04/buck-skin-girl-case-break-success-new-dna-doe-project (accessed November 5, 2018).

[2] Success stories, DNA Doe Project. (n.d.). http://dnadoeproject.org/project_category/success/ (accessed April 7, 2019).

[3] L. Wamsley, In hunt for Golden State Killer, investigators uploaded his DNA to genealogy site, NPR.Org. (2018). https://www.npr.org/sections/thetwo-way/2018/04/27/606624218/in-hunt-for-golden-state-killer-investigators-uploaded-his-dna-to-genealogy-site (accessed November 5, 2018).

[4] M. Gafni, Exclusive: The woman behind the scenes who helped capture the Golden State Killer, The Mercury News. (2018). https://www.mercurynews.com/2018/08/24/exclusive-the-woman-behind-the-scenes-who-helped-capture-the-golden-state-killer/ (accessed November 5, 2018).

[5] A. Lee, Critical genealogist in Golden State Killer case speaks out about her role, Fox KTVU. (2018). http://www.ktvu.com/news/critical-genealogist-in-golden-state-killer-case-speaks-out-about-her-role (accessed November 5, 2018).

[6] C. Phillips, The Golden State Killer investigation and the nascent field of forensic genealogy, Forensic Science International: Genetics. 36 (2018) 186–188. doi:10.1016/j.fsigen.2018.07.010.

[7] S. Augenstein, The tale of the abandoned girl's DNA that led to a notorious cold case, Forensic Magazine. (2017). https://www.forensicmag.com/article/2017/02/tale-abandoned-girls-dna-led-notorious-cold-case (accessed November 12, 2018).

[8] S. Augenstein, True ID of 'Chameleon' Killer revealed: Terry Peder Rasmussen, Forensic Magazine. (2017). https://www.forensicmag.com/news/2017/08/true-id-chameleon-killer-revealed-terry-peder-rasmussen (accessed November 12, 2018).

[9] H. Murphy, She helped crack the Golden State Killer case. Here's what she's going to do next, The New York Times. (2018). https://www.nytimes.com/2018/08/29/science/barbara-rae-venter-gsk.html (accessed November 5, 2018).

[10] M. Gafni, UC Berkeley employee arrested in notorious NorCal Rapist cases using genetic genealogy website, The Mercury News. (2018). https://www.mercurynews.com/2018/09/21/major-development-in-hunt-for-notorious-norcal-rapist/ (accessed November 12, 2018).

[11] M. Molteni, The Future of Crime-Fighting Is Family Tree Forensics, Wired. (2018). https://www.wired.com/story/the-future-of-crime-fighting-is-family-tree-forensics/ (accessed April 7, 2019).
[12] Parabon NanoLabs, Parabon® announces Snapshot® genetic genealogy service for law enforcement (press release), Cision PR Newswire. (2018). https://www.prnewswire.comparabon-announces-snapshot-genetic-genealogy-service-for-law-enforcement-300644394 (accessed November 6, 2018).
[13] E.M. Greytak, D.H. Kaye, B. Budowle, C. Moore, S.L. Armentrout, Privacy and genetic genealogy data, Science. 361 (2018) 857–857. doi:10.1126/science.aav0330.
[14] E.M. Greytak, C. Moore, S.L. Armentrout, Genetic genealogy for cold case and active investigations, Forensic Science International. (2019). doi:10.1016/j.forsciint.2019.03.039.
[15] S. Hernandez, One of the biggest at-home DNA testing companies is working with the FBI, BuzzFeed News. (2019). https://www.buzzfeednews.com/article/salvadorhernandez/family-tree-dna-fbi-investigative-genealogy-privacy (accessed April 7, 2019).
[16] S. Augenstein, Bode, Biggest Private Forensic DNA Lab, Launches Genealogy Service, Forensic Magazine. (2019). https://www.forensicmag.com/news/2019/02/bode-biggest-private-forensic-dna-lab-launches-genealogy-service (accessed April 7, 2019).
[17] C.J. Guerrini, J.O. Robinson, D. Petersen, A.L. McGuire, Should police have access to genetic genealogy databases? Capturing the Golden State Killer and other criminals using a controversial new forensic technique, PLOS Biology. 16 (2018) e2006906. doi:10.1371/journal.pbio.2006906.
[18] S. Zhang, An Abandoned Baby's DNA Condemns His Mother, The Atlantic. (2019). https://www.theatlantic.com/science/archive/2019/03/38-years-later-dna-leads-to-teenager-who-abandoned-her-baby-in-a-ditch/584683/ (accessed April 8, 2019).
[19] M. Molteni, DNA Crime-Solving Is Still New, Yet It May Have Gone Too Far, Wired. (2019). https://www.wired.com/story/dna-crime-solving-is-still-new-yet-it-may-have-gone-too-far/ (accessed April 8, 2019).
[20] D. McNevin, N. Scudder, Is your genome really your own? The public and forensic value of DNA, The Conversation. (2018). http://theconversation.com/is-your-genome-really-your-own-the-public-and-forensic-value-of-dna-95786 (accessed October 28, 2018).
[21] B.E. Berkman, W.K. Miller, C. Grady, Is it ethical to use genealogy data to solve crimes?, Annals of Internal Medicine. 169 (2018) 333. doi:10.7326/M18-1348.
[22] N. Ram, C.J. Guerrini, A.L. McGuire, Genealogy databases and the future of criminal investigation, Science. 360 (2018) 1078–1079. doi:10.1126/science.aau1083.
[23] T. May, Sociogenetic risks — Ancestry DNA testing, third-party identity, and protection of privacy, New England Journal of Medicine. 379 (2018) 410–412. doi:10.1056/NEJMp1805870.
[24] N. Scudder, J. Robertson, S.F. Kelty, S.J. Walsh, D. McNevin, Crowdsourced and crowdfunded: the future of forensic DNA?, Australian Journal of Forensic Sciences. 0 (2018) 1–7. doi:10.1080/00450618.2018.1486456.
[25] D. Kennett, Sense about genealogical DNA testing, Sense About Science. (2013). http://archive.senseaboutscience.org/blog.php/41/sense-about-genealogical-dna-testing.html (accessed November 6, 2018).
[26] F. Calafell, M.H.D. Larmuseau, The Y chromosome as the most popular marker in genetic genealogy benefits interdisciplinary research, Hum Genet. 136 (2017) 559–573. doi:10.1007/s00439-016-1740-0.

[27] M. Kayser, Forensic use of Y-chromosome DNA: a general overview, Hum Genet. 136 (2017) 621–635. doi:10.1007/s00439-017-1776-9.
[28] T.E. King, G.G. Fortes, P. Balaresque, M.G. Thomas, D. Balding, P.M. Delser, R. Neumann, W. Parson, M. Knapp, S. Walsh, L. Tonasso, J. Holt, M. Kayser, J. Appleby, P. Forster, D. Ekserdjian, M. Hofreiter, K. Schürer, Identification of the remains of King Richard III, Nature Communications. 5 (2014) ncomms6631. doi:10.1038/ncomms6631.
[29] J.K. Wagner, J.D. Cooper, R. Sterling, C.D. Royal, Tilting at windmills no longer: a data-driven discussion of DTC DNA ancestry tests, Genetics in Medicine. 14 (2012) 586–593. doi:10.1038/gim.2011.77.
[30] T.E. King, M.A. Jobling, What's in a name? Y chromosomes, surnames and the genetic genealogy revolution, Trends Genet. 25 (2009) 351–360. doi:10.1016/j.tig.2009.06.003.
[31] 23andMe, Introducing Relative Finder: the newest feature from 23andMe, 23andMe Blog. (2009). https://blog.23andme.com/news/introducing-relative-finder-the-newest-feature-from-23andme/ (accessed November 6, 2018).
[32] B.M. Henn, L. Hon, J.M. Macpherson, N. Eriksson, S. Saxonov, I. Pe'er, J.L. Mountain, Cryptic distant relatives are common in both isolated and cosmopolitan genetic samples, PLoS One. 7 (2012). doi:10.1371/journal.pone.0034267.
[33] Living DNA, 2018 - Living DNA product updates, Living DNA. (2018). https://www.livingdna.com/blog/346-august-2018-living-dna-product-updates (accessed November 6, 2018).
[34] C.A. Ball, M.J. Barber, J.K. Byrnes, P. Carbonetto, R.E. Curtis, J.M. Granka, E. Han, A.R. Kermany, N.M. Myres, J. Qi, K. Rand, Y. Wang, L. Willmore, Ancestry DNA Matching White Paper, (2016). https://www.ancestry.com/dna/resource/whitePaper/AncestryDNA-Matching-White-Paper.
[35] FamilyTreeDNA Learning Center, Autosomal genealogy matching, (n.d.). https://www.familytreedna.com/learn/autosomal-genealogy-matching/ (accessed April 7, 2019).
[36] MyHeritage, Major updates and improvements to MyHeritage DNA matching, MyHeritage Blog. (2018). https://blog.myheritage.com/2018/01/major-updates-and-improvements-to-myheritage-dna-matching/ (accessed April 7, 2019).
[37] B. Bettinger, The Shared cM Project: a demonstration of the power of citizen science, Journal of Genetic Genealogy. 8 (2016) 38–42. http://jogg.info/pages/vol8/editorial/bettinger/bettinger-sharedcMProject.html (accessed November 11, 2018).
[38] B. Bettinger, August 2017 update to the Shared cM Project, The Genetic Genealogist. (2017). https://thegeneticgenealogist.com/2017/08/26/august-2017-update-to-the-shared-cm-project/ (accessed November 11, 2018).
[39] D. Kennett, Which DNA test should I take?, Who Do You Think You Are? Magazine. (2017) 17–23. http://www.whodoyouthinkyouaremagazine.com/blog/what-dna-testing-can-tell-you-about-your-family-history (accessed September 11, 2018).
[40] A. Regalado, More than 26 million people have taken an at-home ancestry test, MIT Technology Review. (2019). https://www.technologyreview.com/s/612880/more-than-26-million-people-have-taken-an-at-home-ancestry-test/ (accessed April 7, 2019).
[41] R. Khan, D. Mittelman, Consumer genomics will change your life, whether you get tested or not, Genome Biology. 19 (2018) 120. doi:10.1186/s13059-018-1506-1.
[42] C. Moore, The history of genetic genealogy and unknown parentage searches: an insider's view, Journal of Genetic Genealogy. 8 (2016) 35–37.

http://jogg.info/pages/vol8/editorial/moore/moore-history.html (accessed September 12, 2017).
[43] J. Petrone, As consumer genomics databases swell, more adoptees are finding their biological families, GenomeWeb. (2015). https://www.genomeweb.com/applied-markets/consumer-genomics-databases-swell-more-adoptees-are-finding-their-biological (accessed September 12, 2018).
[44] J.C. Harper, D. Kennett, D. Reisel, The end of donor anonymity: how genetic testing is likely to drive anonymous gamete donation out of business, Hum. Reprod. 31 (2016) 1135–1140. doi:10.1093/humrep/dew065.
[45] T.E. King, S.J. Ballereau, K.E. Schürer, M.A. Jobling, Genetic signatures of coancestry within surnames, Curr. Biol. 16 (2006) 384–388. doi:10.1016/j.cub.2005.12.048.
[46] M. Gymrek, A.L. McGuire, D. Golan, E. Halperin, Y. Erlich, Identifying personal genomes by surname inference, Science. 339 (2013) 321–324. doi:10.1126/science.1229566.
[47] M.H.D. Larmuseau, N. Vanderheyden, A.V. Geystelen, M. van Oven, P. de Knijff, R. Decorte, Recent radiation within Y-chromosomal haplogroup R-M269 resulted in high Y-STR haplotype resemblance, Annals of Human Genetics. 78 (2014) 92–103. doi:10.1111/ahg.12050.
[48] N. Solé-Morata, J. Bertranpetit, D. Comas, F. Calafell, Recent radiation of R-M269 and high Y-STR haplotype resemblance confirmed, Ann. Hum. Genet. 78 (2014) 253–254. doi:10.1111/ahg.12066.
[49] C.-C. Wang, L.-X. Wang, R. Shrestha, S. Wen, M. Zhang, X. Tong, L. Jin, H. Li, Convergence of Y chromosome STR haplotypes from different SNP haplogroups compromises accuracy of haplogroup prediction, Journal of Genetics and Genomics. 42 (2015) 403–407. doi:10.1016/j.jgg.2015.03.008.
[50] FamilyTreeDNA, Which Y-DNA STR markers do FamilyTreeDNA test?, FamilyTreeDNA Learning Center. (n.d.). https://www.familytreedna.com/learn/y-dna-testing/y-str/y-dna-str-markers-family-tree-dna-test/ (accessed November 10, 2018).
[51] CNN wire staff, DNA links 1991 killing to Colonial-era family, CNN. (2012). https://www.cnn.com/2012/01/09/justice/washington-cold-case/index.html (accessed November 6, 2018).
[52] G. Johnson, Cops hope colonial ties reheat cold case, The Spokesman-Review. (2012). http://www.spokesman.com/blogs/sirens/2012/jan/13/cops-hope-colonial-ties-reheat-cold-case/ (accessed November 6, 2018).
[53] C. Moore, Using public Y-DNA profiles to track down criminals: would you?, Your Genetic Genealogist. (2012). http://www.yourgeneticgenealogist.com/2012/01/using-public-y-dna-profiles-to-track.html (accessed November 6, 2018).
[54] J. Mustian, New Orleans filmmaker cleared in cold-case murder; false positive highlights limitations of familial DNA searching, The New Orleans Advocate. (2015). https://www.theadvocate.com/new_orleans/news/article_1b3a3f96-d574-59e0-9c6a-c3c7c0d2f166.html (accessed November 6, 2018).
[55] R. Estes, RIP Sorenson – a crushing loss, DNAeXplained. (2015). https://dna-explained.com/2015/05/15/rip-sorenson-a-crushing-loss/ (accessed November 6, 2018).
[56] M. Taylor, Forensic genealogist uncovers suspect's name in Phoenix cold case, Forensic Magazine. (2016). https://www.forensicmag.com/news/2016/12/forensic-genealogist-uncovers-suspects-name-phoenix-cold-case (accessed November 6, 2018).
[57] P. Aldhous, Cops forced a company to share a customer's identity for the Golden State Killer investigation, BuzzFeed News. (2018).

https://www.buzzfeednews.com/article/peteraldhous/family-tree-dna-subpoena-golden-state-killer (accessed November 6, 2018).
[58] D. Kennett, Genealogy and DNA casualties of GDPR – farewell to World Families Network, Ysearch and Mitosearch, Cruwys News. (2018). https://cruwys.blogspot.com/2018/05/genealogy-and-dna-casualties-of-gdpr.html (accessed November 6, 2018).
[59] E. Shoup, Ancestry.com focuses on core offerings, Ancestry Blog. (2014). https://blogs.ancestry.com/ancestry/2014/06/04/ancestry-com-focuses-on-core-offerings/ (accessed November 6, 2018).
[60] FamilyTreeDNA, About the FamilyTreeDNA database, FamilyTreeDNA. (2019). https://www.familytreedna.com/why-ftdna.aspx (accessed April 7, 2019).
[61] S. Willuweit, L. Roewer, The new Y chromosome haplotype reference database, Forensic Sci Int Genet. 15 (2015) 43–48. doi:10.1016/j.fsigen.2014.11.024.
[62] M.M. Andersen, D.J. Balding, How convincing is a matching Y-chromosome profile?, PLOS Genetics. 13 (2017) e1007028. doi:10.1371/journal.pgen.1007028.
[63] Royal Society, Forensic DNA analysis: a primer for courts, Royal Society, London, 2017. https://royalsociety.org/about-us/programmes/science-and-law/.
[64] FBI, Combined DNA Index System (CODIS), Federal Bureau of Investigation. (n.d.). https://www.fbi.gov/services/laboratory/biometric-analysis/codis (accessed November 6, 2018).
[65] Home Office (UK), National DNA Database statistics, Q2 2018 to 2019, GOV.UK. (n.d.). https://www.gov.uk/government/statistics/national-dna-database-statistics (accessed November 8, 2018).
[66] FBI, CODIS - NDIS Statistics, Federal Bureau of Investigation. (n.d.). https://www.fbi.gov/services/laboratory/biometric-analysis/codis/ndis-statistics (accessed November 11, 2018).
[67] E. Murphy, Law and policy oversight of familial searches in recreational genealogy databases, Forensic Science International. (2018). doi:10.1016/j.forsciint.2018.08.027.
[68] C.D. Huff, D.J. Witherspoon, T.S. Simonson, J. Xing, W.S. Watkins, Y. Zhang, T.M. Tuohy, D.W. Neklason, R.W. Burt, S.L. Guthery, others, Maximum-likelihood estimation of recent shared ancestry (ERSA), Genome Research. 21 (2011) 768–774. http://genome.cshlp.org/content/21/5/768.short.
[69] RootsTech 2019 address From Ancestry CEO Margo Georgiadis, n.d. https://www.youtube.com/watch?v=K7xsTFNcvCs (accessed April 7, 2019).
[70] 23andMe, 23andMe: about us, 23andMe Media Center. (n.d.). https://mediacenter.23andme.com/company/about-us/ (accessed April 7, 2019).
[71] B. Rae-Venter, Using DNA to identify offenders and victims of violent crime. Presentation given at the Genetic Genealogy Ireland conference on 19th October, 2018, in: n.d. https://youtu.be/6J9RITSw8QE (accessed November 6, 2018).
[72] totheletter DNA, Genetic genealogy, Totheletter DNA. (n.d.). https://www.totheletterdna.com/genetic-genealogy/ (accessed November 6, 2018).
[73] D. Syndercombe Court, Forensic genealogy: some serious concerns, Forensic Science International: Genetics. 36 (2018) 203–204. doi:10.1016/j.fsigen.2018.07.011.
[74] DNA Doe Project, Press Release for Buckskin Girl, DNA Doe Project. (2018). http://dnadoeproject.org/press-release-for-buckskin-girl/ (accessed November 6, 2018).
[75] DNA Doe Project, Frequently Asked Questions, DNA Doe Project. (n.d.). http://dnadoeproject.org/faq/ (accessed November 6, 2018).
[76] L. DeFrancesco, DNA makes an appearance, Nature Biotechnology. (2018). doi:10.1038/nbt.4057.

[77] H. Murphy, How an unlikely family history website transformed cold case investigations, The New York Times. (2018). https://www.nytimes.com/2018/10/15/science/gedmatch-genealogy-cold-cases.html (accessed November 6, 2018).
[78] K. Brown, DNA website had unwitting role in Golden State manhunt, Bloomberg News. (2018). https://www.bloomberg.com/news/articles/2018-05-29/killer-app-dna-site-had-unwitting-role-in-golden-state-manhunt (accessed November 6, 2018).
[79] S. Zhang, How a tiny website became the police's go-to genealogy database, The Atlantic. (2018). https://www.theatlantic.com/science/archive/2018/06/gedmatch-police-genealogy-database/561695/ (accessed November 6, 2018).
[80] D. Moberly, Protecting your privacy on GEDmatch, The Handwritten Past. (2018). https://thehandwrittenpast.com/2018/05/05/gedmatch-privacy/ (accessed November 6, 2018).
[81] K. Cooper, Time to move to GENESIS!, Kitty Cooper's Blog. (2018). https://blog.kittycooper.com/2018/12/time-to-move-to-genesis/ (accessed April 8, 2019).
[82] D. Kennett, The brave new world of genetic genealogy, MIT Technology Review. (2018). https://www.technologyreview.com/s/611048/the-brave-new-world-of-genetic-genealogy/ (accessed November 6, 2018).
[83] Legacy Tree Genealogists, The secrets in your spit: using genetic genealogy to solve cold cases, Legacy Tree. (2018). https://www.legacytree.com/blog/the-secrets-in-your-spit-using-genetic-genealogy-to-solve-cold-cases (accessed November 6, 2018).
[84] J.G. Russell, The bull in the DNA china shop, The Legal Genealogist. (2018). https://www.legalgenealogist.com/2018/04/29/the-bull-in-the-dna-china-shop/ (accessed November 6, 2018).
[85] S. Zhang, How a genealogy website led to the alleged Golden State Killer, The Atlantic. (2018). https://www.theatlantic.com/science/archive/2018/04/golden-state-killer-east-area-rapist-dna-genealogy/559070/ (accessed November 6, 2018).
[86] GEDmatch, GEDmatch site policy introduced on 17 August 2017. Retrieved from the Internet Archive on 25 April 2018., (n.d.). https://web.archive.org/web/20180427152614/https://www.gedmatch.com/policy.php.
[87] M. Feeney, "Genetic Informants" and the hunt for the Golden State Killer, Cato Institute. (2018). https://www.cato.org/blog/genetic-informants-hunt-golden-state-killer (accessed November 6, 2018).
[88] D. Kennett, Updates to the terms of service and privacy policy at GEDmatch, Cruwys News. (2018). https://cruwys.blogspot.com/2018/05/updates-to-terms-of-service-and-privacy.html (accessed November 6, 2018).
[89] D.F. Maron, Cold cases heat up as law enforcement uses genetics to solve past crimes, Scientific American. (2018). https://www.scientificamerican.com/article/cold-cases-heat-up-as-law-enforcement-uses-genetics-to-solve-past-crimes/ (accessed November 6, 2018).
[90] C. Whyte, Police can now track killers using relatives' DNA – but should they?, New Scientist. (2018). https://www.newscientist.com/article/mg23931901-500-police-can-now-track-killers-using-relatives-dna-but-should-they/ (accessed November 6, 2018).
[91] N. Moray, K.E. Pink, P. Borry, M.H. Larmuseau, Paternity testing under the cloak of recreational genetics, Eur. J. Hum. Genet. 25 (2017) 768–770. doi:10.1038/ejhg.2017.31.
[92] B.D.C. Curtis, How online genealogy created an unregulated forensic database for police, ABC News. (2018). https://www.abc.net.au/news/science/2018-08-13/online-

genealogy-police-dna-databases-golden-state-killer/10094220 (accessed November 6, 2018).
[93] The Local, Swedish police mull using DNA family tree websites to catch killers, The Local. (2018). https://www.thelocal.se/20180507/swedish-police-mull-using-dna-family-tree-websites (accessed November 6, 2018).
[94] S. Augenstein, Exclusive: The FBI had already accessed Family Tree DNA's database before cooperation, Forensic Magazine. (2019). https://www.forensicmag.com/news/2019/03/exclusive-fbi-had-already-accessed-family-tree-dnas-database-cooperation (accessed April 8, 2019).
[95] A. Vaughan, Home DNA-testing firm will let users block FBI access to their data, New Scientist. (n.d.). /article/2196433-home-dna-testing-firm-will-let-users-block-fbi-access-to-their-data/ (accessed April 8, 2019).
[96] S. Zhang, A DNA Company Wants You to Help Catch Criminals, The Atlantic. (2019). https://www.theatlantic.com/science/archive/2019/03/a-dna-company-wants-your-dna-to-catch-criminals/586120/ (accessed April 8, 2019).
[97] J.G. Russell, A good start by FTDNA, The Legal Genealogist. (2019). https://www.legalgenealogist.com/2019/03/13/a-good-start-by-ftdna/ (accessed April 8, 2019).
[98] M. Gleeson, My input to FTDNA's Citizen Panel, DNA and Family Tree Research. (2019). https://dnaandfamilytreeresearch.blogspot.com/2019/03/my-input-to-ftdnas-citizen-panel.html (accessed April 8, 2019).
[99] FamilyTreeDNA, FamilyTreeDNA terms of service, (n.d.). https://www.familytreedna.com/legal/terms-of-service (accessed April 8, 2019).
[100] FamilyTreeDNA, FamilyTreeDNA law enforcement guide, (n.d.). https://www.familytreedna.com/legal/law-enforcement-guide (accessed April 8, 2019).
[101] FTDNA Learning Center – Law Enforcement Matching – Frequently Asked Questions, FamilyTreeDNA Learning Center. (n.d.). https://www.familytreedna.com/learn/ftdna/law-enforcement-faq/ (accessed April 8, 2019).
[102] FamilyTreeDNA, FamilyTreeDNA transparency report, FamilyTreeDNA Learning Center. (n.d.). https://web.archive.org/web/20190204191230/https://www.familytreedna.com/learn/ftdna/transparency-report/ (accessed April 8, 2019).
[103] J. Petrone, Genetic genealogy firms reassure clients about data privacy After FTDNA divulges FBI partnership, GenomeWeb. (2019). https://www.genomeweb.com/applied-markets/genetic-genealogy-firms-reassure-clients-about-data-privacy-after-ftdna-divulges-fbi#.XKsmxphKjIV (accessed April 8, 2019).
[104] MyHeritage, MyHeritage terms and conditions, (n.d.). https://www.myheritage.com/FP/Company/popup.php?p=terms_conditions (accessed November 7, 2018).
[105] 23andMe, 23andMe terms of service, (n.d.). https://www.23andme.com/en-gb/about/tos/ (accessed November 7, 2018).
[106] Ancestry.com, Ancestry terms and conditions, (n.d.). https://www.ancestry.com/cs/legal/termsandconditions (accessed November 7, 2018).
[107] Samantha, UPDATE re: how to take an AncestryDNA test without saliva, Another Teen Mom. (2017). http://anotherteenmom.com/reviews/update-re-how-to-take-an-ancestrydna-test-without-saliva/ (accessed November 7, 2018).
[108] 23andMe, 23andMe transparency report, 23andMe. (2019). https://www.23andme.com/en-gb/transparency-report/ (accessed April 8, 2019).

[109] 23andMe, 23and Me Guide for Law Enforcement, 23andMe. (n.d.). https://www.23andme.com/law-enforcement-guide/ (accessed April 8, 2019).
[110] Ancestry.com, Ancestry transparency report, Ancestry.Com. (2018). https://www.ancestry.co.uk/cs/transparency (accessed April 8, 2019).
[111] Ancestry.com, Ancestry guide for law enforcement, Ancestry.Com. (n.d.). https://www.ancestry.co.uk/cs/legal/lawenforcement (accessed April 8, 2019).
[112] S. Nelson, S.M. Fullerton, "Bridge to the literature"? Third-party genetic interpretation tools and the views of tool developers, Journal of Genetic Counseling. 27 (2018) 770–781. https://link-springer-com.libproxy.ucl.ac.uk/article/10.1007/s10897-018-0217-9 (accessed October 31, 2018).
[113] B. Resnick, The fascinating, stunningly complicated new science of genetics, Vox. (2018). https://www.vox.com/science-and-health/2018/8/23/17527708/genetics-genome-sequencing-gwas-polygenic-risk-score (accessed August 29, 2018).
[114] L.L.P. Larkin, Cystic fibrosis: a case study in genetic privacy, The DNA Geek. (2017). https://thednageek.com/cystic-fibrosis-a-case-study-in-genetic-privacy/ (accessed November 7, 2018).
[115] K. Cooper, When the DNA says your parents are related, Kitty Cooper's Blog. (n.d.). https://blog.kittycooper.com/2018/07/when-the-dna-says-your-parents-are-related/ (accessed November 7, 2018).
[116] Y. Erlich, A. Narayanan, Routes for breaching and protecting genetic privacy, Nat Rev Genet. 15 (2014) 409–421. doi:10.1038/nrg3723.
[117] P.M. Ney, L. Ceze, T. Kohno, Computer security risks of distant relative matching in consumer genetic databases, ArXiv:1810.02895 [Cs]. (2018). http://arxiv.org/abs/1810.02895 (accessed October 31, 2018).
[118] B. Durie, What is genealogy? Philosophy, education, motivations and future prospects, Genealogy. 1 (2017) 4. doi:10.3390/genealogy1010004.
[119] J. Hovorka, The perceived value of professional genealogy, part one, Association of Professional Genealogists Quarterly. XXXIII (2018) 44–47.
[120] J. Hovorka, The perceived value of professional genealogy, part three, Association of Professional Genealogists Quarterly. XXXIII (2018) 44–47.
[121] M.H.D. Larmuseau, Growth of ancestry DNA testing risks huge increase in paternity issues, Nature Human Behaviour. 3 (2019) 5. doi:10.1038/s41562-018-0499-9.
[122] J.G. Russell, Opt-out is not informed consent, The Legal Genealogist. (2019). https://www.legalgenealogist.com/2019/03/31/opt-out-is-not-informed-consent/ (accessed April 9, 2019).
[123] E.E. Joh, DNA theft: recognizing the crime on nonconsensual genetic collection and testing, Boston University Law Review. 91 (2011) 665–700. http://www.bu.edu/law/journals-archive/bulr/documents/joh.pdf.
[124] Human Tissue Act 2004, Legislation.Gov.Uk. (n.d.). https://www.legislation.gov.uk/ukpga/2004/30/contents (accessed November 12, 2018).
[125] National Conference of State Legislatures, DNA arrestee laws, (n.d.). http://www.ncsl.org/Documents/cj/ArresteeDNALaws.pdf (accessed November 7, 2018).
[126] National Conference of State Legislatures, Convicted offenders required to submit DNA samples, (n.d.). http://www.ncsl.org/Documents/cj/ConvictedOffendersDNALaws.pdf.
[127] M. Schenk, Cassano case: cigarette butt sealed deal, The News-Gazette (Champaign, Illinois). (2018). http://www.news-gazette.com/news/local/2018-08-29/cassano-case-cigarette-butt-sealed-deal.html (accessed November 7, 2018).

[128] M. Schenk, Cassano case: suspect lived within blocks of victim, The News-Gazette (Champaign, Illinois). (2018). http://www.news-gazette.com/news/local/2018-08-29/cassano-case-suspect-lived-within-blocks-victim.html (accessed November 7, 2018).
[129] C.N. Maguire, L.A. McCallum, C. Storey, J.P. Whitaker, Familial searching: a specialist forensic DNA profiling service utilising the National DNA Database® to identify unknown offenders via their relatives—the UK experience, Forensic Science International: Genetics. 8 (2014) 1–9. doi:10.1016/j.fsigen.2013.07.004.
[130] J. Kim, D. Mammo, M.B. Siegel, S.H. Katsanis, Policy implications for familial searching, Investig Genet. 2 (2011) 22. doi:10.1186/2041-2223-2-22.
[131] C.R. Munoz, Sheriff: new DNA technology solves 1999 murder of Sarasota woman, Herald-Tribune (Sarasota). (2018). https://www.heraldtribune.com/news/20180919/sheriff-new-dna-technology-solves-1999-murder-of-sarasota-woman?start=19 (accessed November 7, 2018).
[132] R. Leibrock, The conspiracy issue: the Golden State Killer had an accomplice!, Sacramento News & Review. (2018). https://www.newsreview.com/sacramento/conspiracy-issue-the-golden/content?oid=26507120 (accessed November 7, 2018).
[133] Ballotpedia, California Proposition 69, required collection of DNA samples from felons (2004), Ballotpedia. (n.d.). https://ballotpedia.org/California_Proposition_69,_Required_Collection_of_DNA_Samples_from_Felons_(2004) (accessed November 8, 2018).
[134] M. DeLisi, Forensic epidemiology: harnessing the power of public DNA sources to capture career criminals, Forensic Science International. 291 (2018) e20–e21. doi:10.1016/j.forsciint.2018.07.018.
[135] P.J. Speaker, The jurisdictional return on investment from processing the backlog of untested sexual assault kits, Forensic Science International: Synergy. 1 (2019) 18–23. doi:10.1016/j.fsisyn.2019.02.055.
[136] S.M. Fullerton, R. Rohlfs, Should police detectives have unrestricted access to public genetic databases?, Leapsmag. (2018). https://leapsmag.com/should-police-detectives-have-total-access-to-public-genetic-databases/ (accessed October 28, 2018).
[137] A. Cutway, DNA match from genealogy site leads to arrest in 2001 murder of UCF student, Click Orlando. (2018). https://www.clickorlando.com/news/dna-match-from-genealogy-site-leads-to-arrest-in-2001-murder-of-ucf-student (accessed November 7, 2018).
[138] E. Campanella, Government using ancestry websites to deport immigrants: lawyer, court documents, Global News. (2018). https://globalnews.ca/news/4616715/ancestry-dna-test-deportation-cbsa/ (accessed November 7, 2018).
[139] NZ Herald, 47 year old US murder mystery linked to NZ through victim DNA and Kiwi-made bra, New Zealand Herald. (2018). https://www.nzherald.co.nz/nz/news/article.cfm?c_id=1&objectid=12157532 (accessed November 9, 2018).
[140] M. (Doc) Edge, G. Coop, How lucky was the genetic investigation in the Golden State Killer case?, BioRxiv. (2019) 531384. doi:10.1101/531384.
[141] S. Zhang, V. Darbha, The Atlantic argument: you should be worried about your DNA privacy, (2018). https://www.theatlantic.com/video/index/574543/dna-privacy/ (accessed November 7, 2018).
[142] K. Dedrickson, Universal DNA databases: a way to improve privacy?, J Law Biosci. 4 (2017) 637–647. doi:10.1093/jlb/lsx041.

[143] J.W. Hazel, E.W. Clayton, B.A. Malin, C. Slobogin, Is it time for a universal genetic forensic database?, Science. 362 (2018) 898–900. doi:10.1126/science.aav5475.
[144] K.S. Moran, Damned by DNA — Balancing personal privacy with public safety, Forensic Science International. 292 (2018) e3–e4. doi:10.1016/j.forsciint.2018.09.011.
[145] Biometrics and Forensics Ethics Groups, Ethical principles, 2018. https://www.gov.uk/government/publications/ethical-principles-biometrics-and-forensics-ethics-group (accessed November 8, 2018).