# Crowdsourced and crowdfunded: The future of forensic DNA?

Nathan Scudder [a, c]

James Robertson [a]

Sally F. Kelty [b]

Simon J. Walsh [c]

Dennis McNevin [d]

*a* *National Centre for Forensic Studies, Faculty of Science and Technology, University of Canberra, ACT 2617, Australia*

*b* *Centre for Applied Psychology, Faculty of Health, University of Canberra, ACT 2617, Australia*

*c* *Australian Federal Police, GPO Box 401, Canberra ACT 2601, Australia*

*d* *Centre for Forensic Science, School of Mathematical and Physical Sciences, Faculty of Science, University of Technology Sydney, Broadway, NSW, 2007, Australia*

Corresponding author: Nathan.Scudder@canberra.edu.au

Word Count: 2,956 (with references)

# Crowdsourced and crowdfunded: The future of forensic DNA?

Forensic DNA analysis is dependent on comparing the known and the unknown. Expand the number of known profiles, and the likelihood of a successful match increases. Forensic use of DNA is moving towards comparing samples of unknown origin with publicly available genetic data, such as the records held by genetic genealogy providers. Use of forensic genetic genealogy has yielded a number of recent high-profile successes but has raised ethical and privacy concerns. Navigating family trees is complex, even more so when combined with a comparison of genetic relationships. This intelligence-gathering process has led to occasional false leads, and its use also risks a public backlash, similar to concerns over Cambridge Analytica. A cautious approach to use of this technique is therefore warranted.

Keywords: forensic genetic genealogy, privacy, familial DNA, forensic DNA analysis

*"Data! Data! Data!" he cried impatiently. "I can't make bricks without clay."*
- *Sherlock Holmes* (A. C. Doyle, The Adventures of Sherlock Holmes, 1892)

Forensic DNA analysis has, for 30 years, provided an invaluable tool for law enforcement. The ability to compare DNA from a crime scene with a suspect, or the DNA from recovered human remains with that of a close family member, has revolutionised forensic science [1]. Traditional forensic DNA analysis is dependent on the match – the known and the unknown. Expand the number of known profiles, and the likelihood of a match with an unknown increases.

As law enforcement moves beyond its own data holdings to publicly available genetic information, this is where forensic use of DNA has entered a larger discussion around 'Big Data' [2]. We have entered an era where there is a significant repository of publicly-accessible genetic data. The usefulness of these data is further increased by family links and the ability to overlay family genealogy records. Some commentators have compared this DNA data mining potential to the ethical issues raised by the recent trawling of Facebook data by political consulting firm Cambridge Analytica [3].

## Suspect identification

When Joseph James DeAngelo was arrested in California in April 2018 over a series of 30-year-old murders and assaults, attention quickly focused on how the suspect was found. In their search for the so-called 'Golden State Killer', police searched a public database that people use to compare sections of DNA with other individuals. This approach reportedly took months, involved police and consultant genealogists tracing records back to the suspect's great-great-great grandparents, and then building 25 family trees forward to the present day, eventually narrowing in on a single suspect [4-6].

However - as anyone who has attempted to trace their family tree would know – such a process is not for the faint-hearted. It is complex and difficult, prone to error and misinterpretation. Family trees have been described as more like 'entangled meshes' [7]. The use of public genealogy records adds an extra dimension to familial DNA matching, a technique used as early as 2003 in the United Kingdom and which attracted public attention in the 'Grim Sleeper' case in California in 2010. In that case, police had

a partial match to a DNA profile in a law enforcement database, identifying an immediate family member as the suspect [8, 9].

The potential to traverse genetic profiles from a cross-section of the population certainly recasts some of the criticism of familial DNA matching as being restricted to law enforcement databases which – for socio-economic reasons – generally have a higher representation from minority groups [10]. Forensic genetic genealogy, with its reliance on fee-for-service analysis, potentially skews in the opposite direction – towards families with a higher disposable income.

## Identifying human remains

Marcia King was murdered in 1981 before the advent of forensic DNA analysis. Her unidentified body was buried in a 'Jane Doe' grave, with only exhibits – including a vial of blood – retained by police. Over nearly four decades investigators exhausted all leads in attempting to identify the 'Buck Skin Girl', named for the type of jacket she was wearing when found dumped near a road in Ohio in the United States. Police had successfully developed a DNA profile for her but there was no match with law enforcement or missing person DNA holdings [11].

In 2018, the DNA Doe Project – a charity group formed in 2017 to apply forensic genetic genealogy to unsolved missing person cases – agreed to work with law enforcement on the case. Applying a 'crowd-funding' approach, where the team appealed for public donations, the charity funded whole genome sequencing of a sample from the remaining blood. While the genetic analysis was only partially successful, it produced significant amounts of genetic data consistent with the markers used by direct-to-consumer genetic providers.

The project team uploaded and compared the genetic data with publicly accessible genetic profiles and identified an individual who was a possible first cousin, once removed. Then, by searching that cousin's own shared family tree through a major genealogical website, they came to a presumptive identification – that cousin had flagged a relative in their family tree as 'Death-Unknown Missing-Presumed Dead' [11].

In a matter of hours, genealogists had provided a solid lead in a 37-year-old case which, with confirmatory DNA testing, led to the identification of the victim as Marcia King.

There are more than 500 unidentified human remains in Australia today [12]. Given the success of the DNA Doe Project's team of expert genealogists to date, applying such an approach could help bring closure to missing persons' families.

## Where things can go wrong

Law enforcement use of genealogical and DNA databases has not always yielded such results. In 1996, Angie Dodge was murdered in Idaho in the United States. DNA, believed to be from the suspect, was recovered from the crime scene. Nearly 20 years later, police obtained a warrant to search a specific database managed by genealogy provider AncestryDNA. This search yielded 41 partial matches, with one individual matching at 34 out of 35 Y-STR markers. Using a similar forensic genealogy approach, investigators reviewed close family members of that individual and settled on Michael Usry, Jr. as a suspect. Mr Usry, who later provided a DNA sample and was excluded as a suspect, was a young adult at the time of the murder and, coincidentally, had vacationed in Idaho around that time [13, 14].

While forensic genealogy is a useful intelligence tool, there are consequences for individuals if the tool incorrectly identifies a suspect. Mr Usry notes that it took a month to conduct the required DNA testing to clear his name. Online search engines still return a high number of results linking his name to the murder investigation [14]. While most of these links make it clear he was excluded as a suspect through further DNA testing, one asks 'Do you think Michael Usry Jr. could be involved in Angie's murder?'.

## Will people be put off genetic testing?

The potential for online genetic databases to be used to assist law enforcement is ever increasing. In each of the above cases, investigators uploaded some form of genetic data, of unknown origin, to a public database. This could amount to a breach of a provider's terms and conditions, but there may be little the company can do to prevent such use [15]. The direct-to-consumer testing market is expected to more than triple by 2022, to $A388 million. In 2017, AncestryDNA - the largest of the providers - reportedly sold 1.5 million testing kits over the 'Black Friday' sales weekend alone [16].

But use of forensic genealogy also has the potential to undermine consumer trust in genetic testing and online genealogy. Genetic providers may be more susceptible to consumer backlash about privacy concerns than social media companies such as Facebook which has continued to grow in spite of recent privacy concerns [17]. Many users do not find the need to engage with genetic providers on an ongoing basis, like they do with Facebook. After initial testing, users wishing to minimise privacy risks could potentially download their data and then delete their accounts, limiting further use of their data.

Genetic providers are also limited in their ability to implement privacy safeguards, such as identity verification, due to the very nature of their products. Individuals may legitimately use the tool without knowing their true birth name or names of family members.

## We should proceed with caution

Forensic genetic genealogy is one example of a trend in the intelligence value of publicly accessible data. A coroner in Ada Country in the United States noted that social media was also being used more frequently to assist in identification: 'Facebook is not something we thought we'd be using to find next of kin...We use it every single week' [18].

Publicly accessible data can even predict family relationships. Data scientists recently used next of kin volunteered by two million patients to assemble 223,000 family trees, the largest containing 100 relatives [19, 20].

The question of law enforcement use of social media has been raised in criminal cases in the United States. In *USA v. Daniel Gatson, et al.* the US District Court in New Jersey ruled that law enforcement could create fake social media accounts to entice suspects to engage with them online. However, the judgement specifically notes that the defendant consented to this by accepting the online friend request. Consent for use of online genealogical databases is broader, but arguably consent has been given for a narrow purpose [21].

In *Arquiette v. United States of America et al*, a US District Court case that was ultimately settled, an individual sued the United States Drug Enforcement

Administration (DEA) after it used her seized phone to create a fake social media account in her name. In that case, a copy of a letter from the social media platform to the DEA was publicly released informing them it had terminated the fake account and demanding that the DEA stop using any other fake accounts on its platform. The provider claimed that such profiles would 'threaten the integrity of [the Facebook] community' and that this would make users 'feel less safe and secure when using our service'.

Courts have continued to grapple with technological change and the impact of exploitation by law enforcement on individual rights and expectations of privacy [22, 23]. Similar arguments may arise with forensic genealogy. Courts may need to balance the benefits to society of solving crime with whether the user has given implied consent, both for themselves and their relatives.

Privacy legislation may also play a part. Europe and Australia, amongst many other countries, have strong privacy protections [24-26]. It is possible that privacy regulators may take an interest in this approach, although the regulatory focus tends to be on the holder of the personal information - in this case, the direct-to-consumer providers themselves. The application of specific health privacy laws, such as the United States HIPAA Privacy Rule, to DNA databases is less clear [27].

However, law enforcement is also holding genetic information. In almost all cases, a crime scene sample subjected to forensic genetic genealogy would be from an unknown source. If law enforcement already had a suspicion as to the identity of the donor, then uploading that genetic information to a public database could well amount to a breach of that individual's right to privacy over their genetic information.

For a sample of unknown origin, once uploaded and after an hypothesis of identity begins to form, that genetic data would begin to attract privacy protection. There would almost certainly be an obligation on law enforcement to remove the profile from any public database as soon as the donor of that genetic material was reasonably identifiable.

While the legal risk here would appear low, it must be remembered that the process can be two-way and there is potential for others to make a family connection to the uploaded crime scene profile.

It is possible that a suspect could upload their own genetic data (either because they are coincidentally an avid family historian, or for more sinister counter-intelligence purposes). In such a case, depending on how the site operates, they may receive e-mail notification of a new sibling – a twin, in fact - as law enforcement uploads the relevant crime scene data. Operational security is therefore a relevant factor.

Notwithstanding some degree of risk, recent successes in the application of forensic genetic genealogy will attract the attention of law enforcement [28]. In fact, a bill passed in May 2018 by the United States Congress specifically quarantines a portion of funding for so-called DNA cold case investigations [29]. It will be interesting to see for how long the capability can be readily exploited. The platform used in several of these case studies is a non-profit genetic database of fewer than a million profiles [30]. Open source and public genealogy platforms would struggle to exclude law enforcement (or anyone else with an interest in identifying someone for non-genealogy purposes) while still allowing their users the flexibility to transfer and freely upload their own genetic information.

But some commentators have noted that their very existence is fickle. One noted the compliance burden of the European Union's General Data Protection Regulation, with its commencement perhaps serving – at least in part – as a catalyst for several

smaller platforms to close down in recent months [31]. This may see the migration of genetic data to major commercial providers, less accessible to law enforcement.

The use of forensic genealogy brings us closer to a point where it may be possible – given enough data and resources – to identify any genetic sample. Crowdsourcing and crowdfunding means this technique is available to all.

Achieving an approach that is privacy compliant, balanced and effective is essential to maintaining public trust and minimising potential harm. Otherwise individuals who, having parted with $99 and a small vial of saliva, may suddenly find themselves part of a criminal investigation.

**Acknowledgements**

Research for this article was supported by an Australian Government Research Training Program Scholarship. This article also draws on research funded by the Endeavour Scholarships and Fellowships, a Department of Education and Training initiative.

The opinions expressed in this article are the authors and do not necessarily reflect the views of any associated agency or institution. This article has been updated and revised from a shortened version:

> Scudder, N., McNevin, D., 'Is your genome really your own? The public and forensic value of DNA', *The Conversation*, 2 May 2018.

## References

1. Martin PD, Schmitter H, Schneider PM. A brief history of the formation of DNA databases in forensic science within Europe. Forensic Science International. 2001;119(2):225-31
2. Joh EE. Policing by numbers: big data and the Fourth Amendment. Wash L Rev. 2014;89:35
3. The ethics of catching criminals using their family's DNA [editorial]. Nature. 557, 5 (2018)
4. Pleasance C. Did home DNA testing kits on ancestry websites lead to the Golden State Killer? Daily Mail. 2018 April 28
5. Arango T. The Cold Case That Inspired the 'Golden State Killer' Detective to Try Genealogy. New York Times. 2018 May 3
6. Cooper K. How to find a killer using DNA and genealogy. Kitty Cooper Blog. 2018 April 29. Available from http://blog.kittycooper.com/2018/04/how-to-find-a-killer-using-dna-and-genealogy/
7. Rutherford A. You're Descended from Royalty and So Is Everybody Else. Nautilus. 2018 January 4
8. Selk A. The ingenious and 'dystopian' DNA technique police used to hunt the 'Golden State Killer' suspect. Washington Post. 2018 April 28
9. Bieber FR, Brenner CH, Lazer D. Finding criminals through DNA of their relatives. Science. 2006;312(5778)
10. Murphy E. Relative doubt: familial searches of DNA databases. Michigan Law Review. 2010:291-348

11. Augenstein S. 'Buck Skin Girl' Case Break Is Success of New DNA Doe Project. Forensic Magazine. 2018 April 16
12. Ward J. Australia has 2,000 missing persons and 500 unidentified human remains - a dedicated lab could find matches. The Conversation. 2018 January 31
13. Hall TO. The Y-Chromosome in Forensic and Public Health Genetics. University of Washington. 2016. Available from: http://hdl.handle.net/1773/38208
14. Clark B. Contradictory DNA results put focus on test methods. Post Register. 2017 July 29
15. Larkin L. Genealogy and the Golden State Killer. The DNA Geek Blog. 2018 April 26. Available from: http://thednageek.com/genealogy-and-the-golden-state-killer/
16. Brown KV. The Consumer DNA Testing Market Is Already Booming, But It's About To Explode. Gizmodo. 2018 January 21
17. Roettgers J. #DeleteFacebook Didn't Happen: Facebook Grows Users in Q1, Beats Earnings Estimates. Variety. 2018 April 25
18. Social Media Forensics: Coroners using Facebook to find victims' next of kin. KIVITV. 2018 February 22
19. Kaiser J. Family trees hidden in medical records could predict your disease risk. 2018 May 17. Available from: http://dx.doi.org/10.1126/science.aau2102
20. Murphy H. 'Will You Be My Emergency Contact?' Takes on a Whole New Meaning. New York Times. 2018 May 17
21. Kirkpatrick B. DNA security: my thoughts in the wake of the Golden State Killer case development. Watershed DNA Blog. 2018 May 5. Available from: https://www.watersheddna.com/blog-and-news/goldenstatekillercase
22. Joh EE. Bait, Mask, and Ruse: Technology and Police Deception. Harv L Rev F. 2014;128:246.
23. Ford M. How the Supreme Court Could Rewrite the Rules for DNA Searches. New Republic. 2018 April 30
24. Williams R, Wienroth M. Social and ethical aspects of Forensic Genetics: a critical review. Forensic Science Review. 2017;29(2)
25. DLA Piper. Data Protection Laws of the World: Full Handbook. DLA Piper. 2017
26. Office of the Australian Information Commissioner. Australian Privacy Principles Guidelines. 2015
27. Ram N. You Can't Hide Your Genes. Slate. 2018 May 4
28. Graham J. Boston police explore using commercial DNA databases. Boston Herald. 2018 May 6
29. Vergano D, Aldhous P. A New Law Would Give More Money For Prosecuting Cold Cases Like The Golden State Killer. Buzzfeed. 2018 May 16
30. Kerr O. Tentative Thoughts on the Use of Genealogy Sites to Solve Crimes. Reason. 2018 May 2
31. Estes R. World Families Network, Ysearch and Mitosearch Bite the Dust – Thanks So Much GDPR. DNAeXplained. 2018 May 15