



Heriot-Watt University  
Research Gateway

## The effectiveness of forensic genealogy techniques in the United Kingdom – an experimental assessment

### Citation for published version:

Thomson, J, Clayton, T, Cleary, J, Gleeson, M, Kennett, D, Leonard, M & Rutherford, D 2019, 'The effectiveness of forensic genealogy techniques in the United Kingdom – an experimental assessment', *Forensic Science International: Genetics Supplement Series*, vol. 7, no. 1, pp. 765-767.  
<https://doi.org/10.1016/j.fsigss.2019.10.169>

### Digital Object Identifier (DOI):

[10.1016/j.fsigss.2019.10.169](https://doi.org/10.1016/j.fsigss.2019.10.169)

### Link:

[Link to publication record in Heriot-Watt Research Portal](#)

### Document Version:

Peer reviewed version

### Published In:

Forensic Science International: Genetics Supplement Series

### Publisher Rights Statement:

© 2019 Elsevier B.V.

### General rights

Copyright for the publications made accessible via Heriot-Watt Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

### Take down policy

Heriot-Watt University has made every reasonable effort to ensure that the content in Heriot-Watt Research Portal complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [open.access@hw.ac.uk](mailto:open.access@hw.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.

# Journal Pre-proof

## THE EFFECTIVENESS OF FORENSIC GENEALOGY TECHNIQUES IN THE UNITED KINGDOM – AN EXPERIMENTAL ASSESSMENT

Jim Thomson, Tim Clayton, John Cleary, Maurice Gleeson, Debbie Kennett, Michelle Leonard, Donna Rutherford



PII: S1875-1768(19)30244-6

DOI: <https://doi.org/10.1016/j.fsigss.2019.10.169>

Reference: FSISS 1802

To appear in: *Forensic Science International: Genetics Supplement Series*

Received Date: 24 September 2019

Accepted Date: 16 October 2019

Please cite this article as: Thomson J, Clayton T, Cleary J, Gleeson M, Kennett D, Leonard M, Rutherford D, THE EFFECTIVENESS OF FORENSIC GENEALOGY TECHNIQUES IN THE UNITED KINGDOM – AN EXPERIMENTAL ASSESSMENT, *Forensic Science International: Genetics Supplement Series* (2019), doi: <https://doi.org/10.1016/j.fsigss.2019.10.169>

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2019 Published by Elsevier.

## THE EFFECTIVENESS OF FORENSIC GENEALOGY TECHNIQUES IN THE UNITED KINGDOM – AN EXPERIMENTAL ASSESSMENT

Jim Thomson<sup>1</sup>, Tim Clayton<sup>1</sup>, John Cleary<sup>2,3</sup>, Maurice Gleeson<sup>2</sup>, Debbie Kennett<sup>2,4</sup>, Michelle Leonard<sup>2</sup>, Donna Rutherford<sup>2</sup>.

<sup>1</sup> Eurofins Forensic Services, Teddington, UK

<sup>2</sup> ISOGG UK and Ireland

<sup>3</sup> Heriot-Watt University, Glasgow, Scotland, UK

<sup>4</sup> University College London, UK

### Abstract

The use of forensic genealogy techniques to identify Joseph James DeAngelo as the prime suspect in the Golden State Killer case in 2018 has opened up a new approach to investigation of cold cases. Since that breakthrough, these methods have been reportedly applied to more than 50 investigations. To date, all of these relate to investigations in the US, where the high uptake of “direct-to-consumer” (DTC) genetic testing by individuals conducting private ancestral research has provided the necessary publicly available data for successful forensic investigations. We have conducted a study to assess the likely effectiveness of forensic genealogy techniques if applied to investigations in the UK. Volunteers provided their own SNP array data, downloaded from a DTC provider of their choice. These data sets were anonymised and uploaded to the GEDmatch Genesis genealogy website, mimicking data sets from unsourced crime samples. A team of experienced genealogists then attempted to identify the donors of the anonymised data sets by identifying relatives on the database and triangulating to determine their shared family lineages which were further investigated using traditional resources (such as birth, marriage, death and census records). Four individuals out of a convenience sample of ten we correctly identified, at least to the level of one of a set of siblings.

### Key Words

Genetic genealogy, forensic genealogy, ancestry, GEDmatch.

### Introduction

The rapid rise in prominence of genetic genealogy methods applied to forensic investigations in 2018 and 2019 has had a transformative impact on investigation of cold cases for which unsourced DNA evidence is available [1]. Since the identification of Joseph James DeAngelo as the prime suspect in the Golden State Killer case in April 2018, the methods have been applied to more than 50 cases, all related to crimes committed in the USA where the uptake of genetic genealogy DNA tests by the public has been highest. At present, the question of whether such investigative methods would yield successful results if applied to searches outside the USA (relating to (a) unidentified human remains or (b) the victims or perpetrators of crimes) remains untested. This study seeks to provide preliminary information to address this question based on a small convenience sample of ten individuals to assess the potential effectiveness of genetic genealogy methods in identifying UK residents.

### Materials and Methods

#### *Selection of subjects*

Ten volunteer subjects who had privately completed a DNA ancestry/genealogy test from one of the major direct-to-consumer (DTC) providers were recruited from Eurofins Forensic Services staff. All volunteers were fully briefed on the project aims and potential implications of third party scrutiny of

their DNA records and associated genealogy and provided full consent. All subjects were UK residents. The recruitment process was not intended to provide a representative sample of the UK population, but was a convenience sample suitable for an indicative feasibility assessment. For each subject, sex and date of birth were recorded. Subjects also provided information on their country of birth and self-defined ethnicity, and those of their parents and grandparents.

#### *Upload of data to GEDmatch Genesis*

All subjects downloaded their own raw DNA data files from their DTC provider. For each subject, a new account was created on the GEDmatch website using a random user name unattributable to the subject. DNA data was uploaded to the GEDmatch Genesis database and all uploads were changed to a privacy status of “*Research*” which ensured that the subject’s DNA was not included in match results of other users. The GEDmatch kit numbers were provided to the genealogy team, enabling them to conduct the required searches.

#### *GEDmatch searching*

All GEDmatch searches and subsequent genealogical investigations were conducted by a team of five highly-experienced genetic genealogists.

“One-to-many” searches were carried out for each subject to provide the list of closest matches in the GEDmatch database (estimated size at time of study: 1.2 million).

#### *Genetic genealogy investigation strategy*

Investigation of the identity of each subject progressed in several distinct stages:

1. Assessment of the amount of shared DNA with the top matches and the degrees of relationship predicted from them.
2. Identification of clusters of Shared Matches that might indicate a common ancestor.
3. Construction and/or examination of family trees of the top matches to attempt triangulation on a common ancestor.
4. Following successful triangulation, trees were built forward in time to identify candidates for the person of interest. Some candidates could be eliminated on the basis of gender, age, circumstantial evidence, and other information.

## **Results**

The ten subjects included six males and four females aged between 22 and 56. Seven were born in England or Wales with parents and grandparents also declared as White British. One was born in England to a Chilean father (of White British ancestry) and English mother. One was born in England to parents and grandparents of Indian ethnicity who had been resident in St Vincent and the Grenadines. One was born in Romania to Romanian parents and grandparents.

Outcomes of one-to-many comparison searches are summarised in table 1, with matches categorised by the total shared cM of DNA between the subject and the nearest matches. A greater amount of shared DNA is indicative of a closer relationship between the subject and the matched sample.

#### *Successful identifications*

Four of the ten subjects were identified by the genealogy investigation at least to the level of sibling (e.g. “Subject 10’ is most likely one of three siblings, X, Y or Z”).

#### Subject 1

One parent/child match and one sibling match to “subject 1” were identified on GEDmatch making this the most straightforward identification case. Cross-referencing of recent electoral registers and

births and marriage registers identified a candidate family group. "Subject 1" was identified as one of twin brothers, BB or DB, born in Leicester in 1977. The investigation did not identify whether these twins were monozygotic or dizygotic which would have implications for any further DNA testing were this a real case.

#### Subject 6

"Subject 6" was identified following an investigation of the existing family tree of the closest match (an individual "AD" at 125cM predicting ~85% probability of being 3C or closer). Descendants of common great-great grandparents were targeted in the search leading to candidate family lines on the maternal side. This search strategy eventually identified "subject 6" as either TC or BC (brothers).

#### Subject 9

One of the top matches to "subject 9" (364cM – likely 2C-3C) was identifiable as "KW", related to members of a community of Indo-Vincentians (people from St Vincent and the Grenadines of Indian origin) resident in the UK. A public Facebook SVG Indian Heritage Group provided useful historical information enabling KW's tree to be extended back to migrants from India to St Vincent in 1860. Trees were then built forwards to find potential candidates for "subject 9", noting that high levels of endogamy in small communities may make people appear more closely related than they actually are. The process led to identifying JT1 as the most likely candidate to be "subject 9". JT1's true relationship to KW (3C1R) was more distant than predicted.

#### Subject 10

The top match ("AR") at 53 cM had only ~50-60% possibility of being related to "subject 10" as 4C or closer. "AR" was Scottish, with paternal ancestry from Shetland (an island group 170km north of the Scottish mainland). Several other close matches were also Shetlanders or had recent ancestors from Shetland. A 5000-person tree was constructed for the Shetland cluster, drawing upon some very well constructed trees made available publicly by family historians with a deep knowledge of Shetland communities. The presence of matches without Shetland ancestry indicated that "subject 10" was a Shetlander in just part of their tree, for example only on the maternal or paternal side

By this process a candidate marriage between Shetlander JohnT and SN from Yorkshire was identified, with three sons (JT2, DT, RT) as possible candidates to be "subject 10".

#### *Unsuccessful investigations*

The other six subjects in the study could not be identified. At least two of these provided some promising lines of investigation and further work could be done to narrow the search. Furthermore, in the context of forensic investigations these individuals could be monitored over time to check for new matches with higher shared DNA which may appear at a future date.

#### **Conclusions**

This study provides preliminary data to demonstrate the efficacy of genetic genealogy methods in identifying UK residents, The demonstration that four out of a convenience sample of ten volunteers could be identified either by name, or as one of a set of siblings is a useful indicator that the predominance of US individuals on GEDmatch and other genealogical databases does not preclude successful identification of individuals from outside the US.

Provision of such evidence is needed by regulatory bodies outside the US to prompt further consideration of the ethical and regulatory frameworks to enable safe and socially-acceptable introduction of genetic genealogy methods with appropriate regard for privacy and data security issues. This will facilitate the introduction of this valuable investigative approach to assist in the

identification of unidentified human remains as well as the victims and perpetrators of crime in the UK and elsewhere.

#### References

- [1] D. Kennett, Using genetic genealogy databases in missing persons cases and to develop suspect leads in violent crimes, *Forensic Sci. Int.* 301 (2019) 107–117.  
doi:10.1016/j.forsciint.2019.05.016.

Journal Pre-proof

Table 1. Summary of one-to-many searches. Subjects are identified by real initials with the numbers of close matches in each cM range shown

Range cM	Subject 1	Subject 2	Subject 3	Subject 4	Subject 5	Subject 6	Subject 7	Subject 8	Subject 9	Subject 10	Likely relationship
~35-70	1	0	0	0	0	0	0	0	0	0	Parent, child or duplicate of kit holder
220-330	1	0	0	0	0	0	0	0	0	0	Sibling (full)
120-220	0	0	0	0	0	0	0	0	0	0	Half siblings, aunts, uncles, grandparents/children
550-1200	0	0	0	0	0	0	0	0	0	0	1C ~ 1C removed
80-500	0	0	1	0	0	0	0	0	2	0	approx 2C~3C or 1C removed
50-80	2	0	0	0	3	0	0	2	4	1	approx 3C~4C
30-50	6	1	7	8	7	16	6	36	2	5	approx 4C or more distant (could be closer)
20-30	108	99	81	74	71	199	67	823	2	41	approx 4C or more distant (could be closer)