VEROGEN

Forensic Genetic Genealogy with GEDmatch

Realizing the collective power of community, DNA, and sequencing, genealogy offers a new method for creating investigative leads.

Highlights

- **Powerful genetic intelligence**
  Data from DNA samples and genealogical databases help identify people through relatives.

- **Closure for cold and contemporary cases**
  New investigative leads for missing persons identifications, innocence projects, and criminal investigations.

- **Turnkey genealogy solution**
  End-to-end solution packaged in a familiar workflow and optimized for genealogy applications.

Introduction

Genealogists, historians, and adoptees have traditionally used genealogy to find birth families and build family trees. The rise of direct-to-consumer (DTC) genetic tests, such as 23andMe, create the potential for law enforcement to partner with these communities and for forensic laboratories to solve cold cases. Forensic genetic genealogy (FGG), also called investigative genetic genealogy (IGG), combines genealogy with DNA analysis to produce investigative leads in cases of unidentified remains or unsolved crimes.

Verogen is developing an end-to-end, next-generation sequencing (NGS) solution to empower FGG by aiding identification in ways traditional methods cannot. This application note describes how FGG uses DNA data and the Verogen GEDmatch database to generate the genetic intelligence that can lead to an identification. It also presents real-world examples of FGG results linking perpetrators to crimes and exonerating the wrongly convicted with equal levels of assuredness.

Genetic genealogy as a forensic tool

FGG employs genealogical databases, such as GEDmatch, and family trees to develop intelligence in cases of missing persons or unsolved violent crimes. As an impartial, science-driven tool, FGG is particularly useful when traditional methods are inconclusive, or all other options are exhausted. The FGG process starts after results from a forensic database, such as CODIS, are inconclusive. The DNA profile from an unidentified sample is uploaded to GEDmatch for comparison against profiles compiled from known, voluntary contributors. GEDmatch results indicate potential relatives based on the amount of shared genetic material. Using these relatives, a genealogist constructs a family tree that extends over multiple generations to narrow results for the unidentified sample. In criminal investigations, FGG specifically compares crime-scene DNA to volunteer DNA in GEDmatch. Shared segments inform the creation of the multi-generation family tree that can lead investigators to a suspect (Figure 1). The workflow is similar for unidentified remains.[1]

1. Collect DNA from crime scene

2. Upload crime-scene DNA profile to GEDmatch

3. Review results for shared DNA

4. Build a family tree based on shared DNA

5. Identify a suspect through common ancestors

6. Collect DNA from the supsect

7. Compare supsect DNA to the crime-scene DNA

Figure 1: The FGG workflow for a criminal investigation

VEROGEN

Kinship insights

Identity by descent (IBD) describes a DNA segment that two people have inherited from a common ancestor—a concept that provides the framework for FGG work. Two samples sharing overlapping IBD segments across chromosomes allows genealogists to approximate the degree of kinship, with recombination events helping distinguish whether DNA is inherited or shared by chance. GEDmatch converts the SNP calls (A, G, C, or T) from a DTC genetic test into IBD segments (Figure 2). The length of an IBD segment is measured in centimorgans (cM), which measure genetic distance. The larger the cM value, the more DNA is shared between two people and the more likely they are to be related (Table 1).

Expanded database capability

Where permitted, forensic databases already enable short-range familial searches to find close relatives. However, they use only a small number of markers and restrict the pool of DNA profiles to people who have come into contact with law enforcement. When a forensic database fails to produce a hit, a genealogical database can broaden the search by increasing the DNA profile pool and expanding the search criteria.

GEDmatch is the only genealogical database that aggregates DNA profiles from all DTC genetic testing companies—an approach that allows it to compare data from unknown samples against a diverse and extensive set of volunteer data. A long-range familial search finds distant relatives or sets of relatives within this extended data set. The results yield a simple measure of kinship based on IBD segments that can, for example, identify second cousins of a sample contributor.[3]

Figure 3 illustrates an example GEDmatch query result of an association between two people on a segment of human chromosome 6. The horizontal dark blue bars indicate the two shared segments and the table displays information about each segment: start and end locations on the chromosome, total length in centimorgans, and the number of shared SNPs. SNP allele calls, which are not necessary for FGG, are excluded from GEDmatch results.
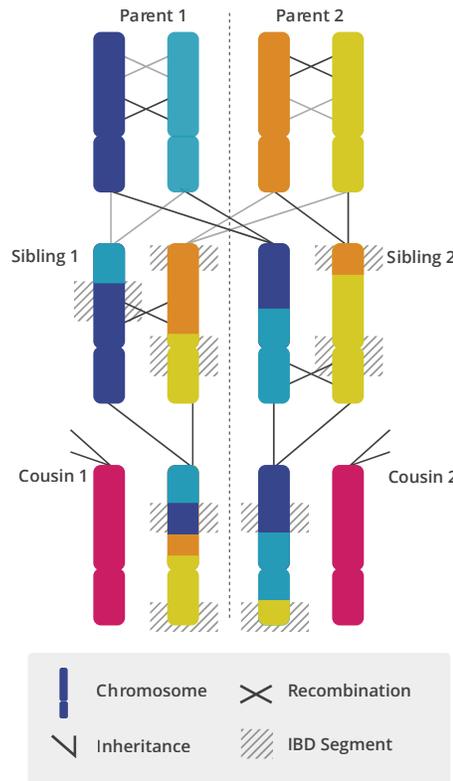

Figure 2: IBD segment lengths approximate the amount of shared DNA between two samples[2]

Targeted assays for challenging samples

Although the GEDmatch database is a powerful tool, the large-scale SNP array methods that generate the data it contains are incompatible with the challenging samples typical in forensic casework. Verogen leverages GEDmatch data to identify an optimum set of compatible markers to combine in a dense, NGS SNP assay designed for low-quantity, low-quality samples. The amplicons are small—ideal for degraded DNA—and selected to be the most informative. Sequencing these amplicons on the MiSeq FGx® Sequencing System targets the data relevant to FGG. In turn, GEDmatch uses only the critical markers to find relatives. The Verogen NGS workflow ultimately requires less sample than traditional methods and limits data to the specific set needed for FGG.

Chr	Start Location	End Location	Centimorgans	SNPs
6	35,996	4,942,186	15.7	2,810
6	106,070,220	186,072,685	56	12,372

Figure 3: A graphical representation of two shared segments on chromosome 6

Application Note

VEROGEN

Table 1: Centimorgan values as a measure of kinship

Relationship	IBD Segment Length (cM)	Matching Chromosomes (%)
Identical twins	3800	100
Parents	3400	50
Full siblings	2500	37.5
Grandparents	1700	25
Aunts and uncles	1700	25
Grandparents	850	12.5
First cousins	850	12.5
Second cousins	212.5	3.125
First cousins twice removed	212.5	3.125

Justice and public safety

FGG plays a growing role in correcting miscarriages of justice and identifying violent offenders who evaded detection. After decades of dead ends, GEDmatch assisted with the identification of Joseph James DeAngelo as the alleged Golden State Killer. His arrest in April 2018 cast a spotlight on FGG, highlighting its potential as a mainstream forensic tool. In June 2019, the conviction of William Earl Talbott II for the 1987 murder of a young Canadian couple, Tanya Van Cuylenborg and Jay Cook, marked the first-ever guilty verdict for a case that relied on FGG evidence. Data from DNA collected at the crime scene over 30 years ago were uploaded to GEDmatch, helping identify Talbott after decades of anonymity.[4]

In another cold case, FGG helped to not only identify the offender, but release the man wrongly convicted of the crime. Unsatisfied with the conviction of Christopher Tapp for the 1996 rape and murder of her daughter, Angie Dodge, Carol Dodge pushed for a proper resolution of the case. Fraught with inconsistencies, the original case included an alleged coerced confession that Tapp later recanted and a lack of direct evidence. Critically, none of the crime-scene DNA matched Tapp. In May 2019, an FGG breakthrough led to the arrest of Brian Leigh Dripps and the definitive clearing of Tapp. A DNA sample from the crime scene was processed and compared with profiles in GEDmatch, identifying Dripps as a possible suspect. Investigators then obtained a discarded cigarette butt for comparison to hair and semen samples from the 1996 crime scene, which matched. When questioned, Dripps admitted to the crime.[5]

Conclusion

Without an identification, cases can languish for decades, suspending survivors, families, and the wrongly convicted in a state of uncertainty while offenders go free. The ability of Verogen NGS technology to advance challenging samples, limit data exposure, and find relatives through shared DNA unlocks an opportunity to push genealogy further. FGG has already demonstrated the capability to generate investigative intelligence that can identify the anonymous and reanimate cold cases.

To learn more about the GEDmatch database, visit **www.verogen.com/a-message-to-verogen-customers-about-the-gedmatch-partnership**.

References

1. Callaghan, Thomas F., "Responsible genetic genealogy," *Science* 366, no. 6462 (October 2019): 155, 10.1126/science.aaz6578.
2. Greytak, Ellen M., CeCe Moore, Steven L. Armentrout, "Genetic genealogy for cold case and active investigations," *Forensic Science International* 299 (June 2019): 103–113, https://doi.org/10.1016/j.forsciint.2019.03.039.
3. Erlich, Yaniv, Tal Shor, Istik Pe'er, Shai Carmi, "Identity inference of genomic data using long-range familial searches," *Science* 362, no. 6416 (November 2018): 690–694, https://doi.org/10.1126/science.aau4832.
4. Heather Murphy, "Genealogy Sites Have Helped Identify Suspects. Now They've Helped Convict One." *New York Times*, July 1, 2019.
5. Shane Bishop, "Police arrest Idaho man in 23-year-old cold-case murder of Angie Dodge," *NBC News*, May 16, 2019.

Verogen · 1.833.837.6436 toll-free (North America) · +1.858.285.4101 tel · +44 (0) 208 054 8706 (United Kingdom) info@verogen.com · www.verogen.com
© 2020 Verogen Inc. All rights reserved. | Document # VD2020005 Rev. B | September 2020