



### B105 Forensic Genetic Genealogy for Law Enforcement

*Ellen M. Greytak, PhD\*, Parabon NanoLabs, Inc, Reston, VA 20190; CeCe Moore, Parabon NanoLabs, Reston, VA 20190; Steven Armentrout, PhD, Parabon NanoLabs, Inc., Reston, VA 20190*

**Learning Overview:** After attending this presentation, attendees will understand the science of genetic genealogy and how it can be applied to forensic samples to aid in law enforcement cases. The entire process, from evidence to conclusion, will be discussed, including how the data is generated, how searches are performed, and how genealogy research is used to narrow down the unknown individual's possible identities.

**Impact on the Forensic Science Community:** This presentation will impact the forensic science community by increasing the general level of knowledge about genetic genealogy, which is actively being used to solve decades-old cold cases. Attendees will learn about this new technology's investigative potential by seeing how it is used on simulated case studies modeled from actual casework.

Genetic genealogy (GG) is the attempt to identify an unknown individual by finding genetic relatives using a public genetic genealogy database and combining the results with traditional genealogical research of public records. Originally developed to help individuals who have lost their biological identities (e.g., adoptees), the same techniques are now generating leads for law enforcement. GG is distinct from familial searching of law enforcement DNA databases in terms of the type of genetic data used, the source of the comparison samples, and the identification process.

Unlike traditional STR analysis, genetic genealogy begins with the generation of dense autosomal single nucleotide polymorphism (SNP) data from a forensic DNA sample using SNP microarray genotyping or whole-genome sequencing. The Y-chromosome and/or mitochondrial DNA can also be analyzed to determine the individual's haplogroups. GG cannot use a short tandem repeat (STR) profile, as significantly more genetic information is needed to detect distant relatives when comparing data from multiple samples.

The SNP data is uploaded to a public GG database that allows law enforcement use, such as GEDmatch. Software algorithms compare the uploaded data to that from everyone in the database who has allowed such comparisons, reporting back the total amount of shared DNA (in centimorgans, cM) in the autosome and on the X-chromosome. The amount of shared autosomal DNA between two individuals correlates strongly with the degree of biological relatedness. However, for a given degree of relatedness, the amount of DNA sharing observed varies and, moreover, high levels of background relatedness in endogamous populations can inflate observed cM values.

The genetic relatives ("matches") found in the database serve as clues from which traditional genealogy research can proceed. First, each closely matching individual (typically, 3<sup>rd</sup> cousins or closer) must be identified and their family tree constructed using public records, going back to all possible common ancestors with the unidentified person. Once the possible common ancestors have been identified, descendency research is performed to identify all their descendants at the appropriate genetic distance to the match. Finally, the possible identity of the unknown individual must be narrowed down using a variety of sources of information, including triangulation among matches, the unknown individual's known sex and possible age, the geographical location of the crime, and/or predictions of ancestry and phenotype. Depending on the distance and the number of matches, GG can narrow the possible identities of the unknown individual to one, a few, or many. These leads are then investigated by law enforcement, and final identification is performed using direct STR matching.

Simulated case studies will be used to demonstrate how genetic genealogy is used in forensic investigations and the kind of information it can generate. The presenters have evaluated hundreds of forensic samples for their potential to be solved using GG, and statistics on the expected resolution rate for different types of samples will be given. This presentation will provide attendees with sufficient information about genetic genealogy to explain the technology to their agency and determine which cases would be good candidates for GG analysis. All attendees will gain the ability to thoroughly evaluate the accuracy of the many journal articles and media pieces published about GG.

#### Genetic Genealogy, DNA, Single Nucleotide Polymorphism