B35 Whole Genome Sequencing (WGS) of Highly Degraded Samples for Forensic Genetic Genealogy and a Case Study

Kelly Harkins Kincaid, PhD, Astrea Forensics, Santa Cruz, CA 95063; Maria A. Nieves Colón, PhD, Astrea Forensics, Santa Cruz, CA 95063; Charles Vaske, PhD, Astrea Forensics, Santa Cruz, CA 95063; Cristina Verdugo, PhD, Astrea Forensics, Santa Cruz, CA 95063; Beth Nelson, BS, Agilent Technologies, Santa Clara, CA 95051-7201; Josh Kapp, BS, University of California Santa Cruz, Santa Cruz, CA 95064; Barbara Rae-Venter, PhD, Carmel, CA 93921; Kathryn Johnston, MD, Rolling Hills Estates, CA 90274; James Scharf, BS, Snohomish County Sheriff's Office, Everett, WA 98201; Beth Shapiro, PhD, University of California Santa Cruz, Santa Cruz, CA 95064; Richard Green, PhD, Santa Cruz, CA 95064*

Learning Overview: After attending this presentation, attendees will better understand the challenges for obtaining and analyzing DNA from rootless hair using WGS data. These WGS data are used to generate whole genome Single Nucleotide Polymorphism (SNP) profiles for purposes of genetic genealogy. The success and significance of this approach for Law Enforcement (LE) is outlined in a case study from Snohomish County, WA, which resulted in the successful identification of "Precious Jane Doe" after nearly 43 years.

Impact on the Forensic Science Community: This presentation will impact the forensic science community by demonstrating the value of rootless hair for use in DNA analyses, including investigative genetic genealogy. Rootless hair and other highly degraded samples are dominated by "ultrashort" DNA fragments (e.g., 30–60bp in length). These fragments are present in minute quantities and are often overwhelmed by the presence of exogenous, non-human DNA (e.g., microbial DNA). For these reasons, forensic samples, especially from cold cases, greatly benefit from the use of ancient DNA (aDNA) -based methods for DNA extraction and WGS library preparation designed to retain the shortest DNA fragments possible while maximizing the number of unique DNA molecules available for sequencing.

To date, DNA-based identification in forensics has typically relied on Polymerase Chain Reaction (PCR) -based assays targeting a handful of Short Tandem Repeat (STR) markers. While PCR is an extremely sensitive approach, the problem remains that the majority of STR amplicons are too long to amplify the ultrashort DNA contained in hair. A similar limitation is true for microarray technology, which can target millions of Single Nucleotide Polymorphisms (SNPs), and is the common method in Direct-to-Consumer (DTC) genetic tests. DTC tests as well as overall interest in genealogy have surged in popularity, accompanied by increasingly large databases containing human genotypes, or SNP profiles, used for finding relatives. Within the past few years, these databases have been repurposed for a new approach to human identification called forensic genetic genealogy, wherein relatives of unknown individuals are identified and used by LE as investigative leads. This approach has helped solve several high-profile cold cases such as the Golden State killer and "Buckskin Girl". However, forensic samples may not contain the quantity or quality of recoverable DNA necessary for array-based genotype analysis. The solution is a WGS approach that is capable of producing genotypes from highly degraded samples in a format similar to DTC tests and compatible with genetic databases.

By applying aDNA techniques for extraction and library preparation to samples such as rootless hair or bone, it becomes possible to capture and sequence short, degraded fragments of DNA from both the nuclear and mitochondrial genome. The resulting WGS data are combined with an imputation approach to generate genotype files suitable for genetic genealogy investigation.

This approach was applied to rootless hair provided by the Snohomish County Sheriff's Office to identify human remains belonging to an unknown female recovered in August 1977. The remains were exhumed in 2008 for DNA testing. A partial genetic profile from the femur bone failed to yield a Combined DNA Index System (CODIS) match. Since then, at least four additional attempts to generate DNA data were unsuccessful. In 2019, hair samples sent to the University of California Santa Cruz/Astrea Forensics underwent DNA extraction using 5cm of a single rootless hair followed by library preparation, sequencing, and genotyping by imputation.

Results demonstrate the successful application of aDNA-like techniques for DNA extraction and library preparation on highly degraded forensic samples. From the library, 346 million sequence reads from the Jane Doe hair were used as input to generate a genotype containing 1,437,366 full autosomal SNP calls. The genotype file was submitted to the law enforcement agency and their genetic genealogists, and uploaded to GEDmatch®. By combining DNA and genealogical analysis, "Precious Jane Doe" was positively identified in 2020 as Lisa Roberts, 43 years after her remains were found.

Genetic Genealogy, WGS, Rootless Hair

---

*Presenting Author