# DNA Doe Project

### INTRODUCTION

DNA Doe Project is a 501(c)(3) non-profit dedicated to identifying unidentified remains using investigative genetic genealogy. Since 2017 we have identified 64 Jane & John Does. As the organization works exclusively on unidentified remains cases, we regularly encounter very challenging samples that are both relatively old, and have often experienced extreme environmental conditions. For these reasons, the samples often contain DNA that is quite degraded, and highly contaminated with bacterial and other exogenous DNA. To overcome these hurdles, we rely on whole genome sequencing, hybrid capture-based enrichment, plus a number of bioinformatics techniques to create usable data for upload to genealogical databases.

## CASE

In 2019 skeletonized remains were uncovered by workers in an open dirt lot. Based on the state of the remains, anthropologists estimated the post-mortem interval to be many decades. A femur femur and partial humerus shaft were provided for DNA extraction.

# DNA EXTRACTION

Astrea Forensics of Santa Cruz, CA performed a total of four DNA extractions, two from each bone. Fluorometry-based quants, including both endogenous DNA and exogenous contamination, are listed below.

Sample	Total DNA (ng)	
Humerus (1)	107.2	
Humerus (2)	126.8	
Femur (1)	412	
Femur (2)	369	

# Overcoming Challenging DNA Samples: Identifying a Pre-Civil War John Doe

KEVIN LORD - LAB LIAISON / BIOINFORMATICIAN DNA DOE PROJECT



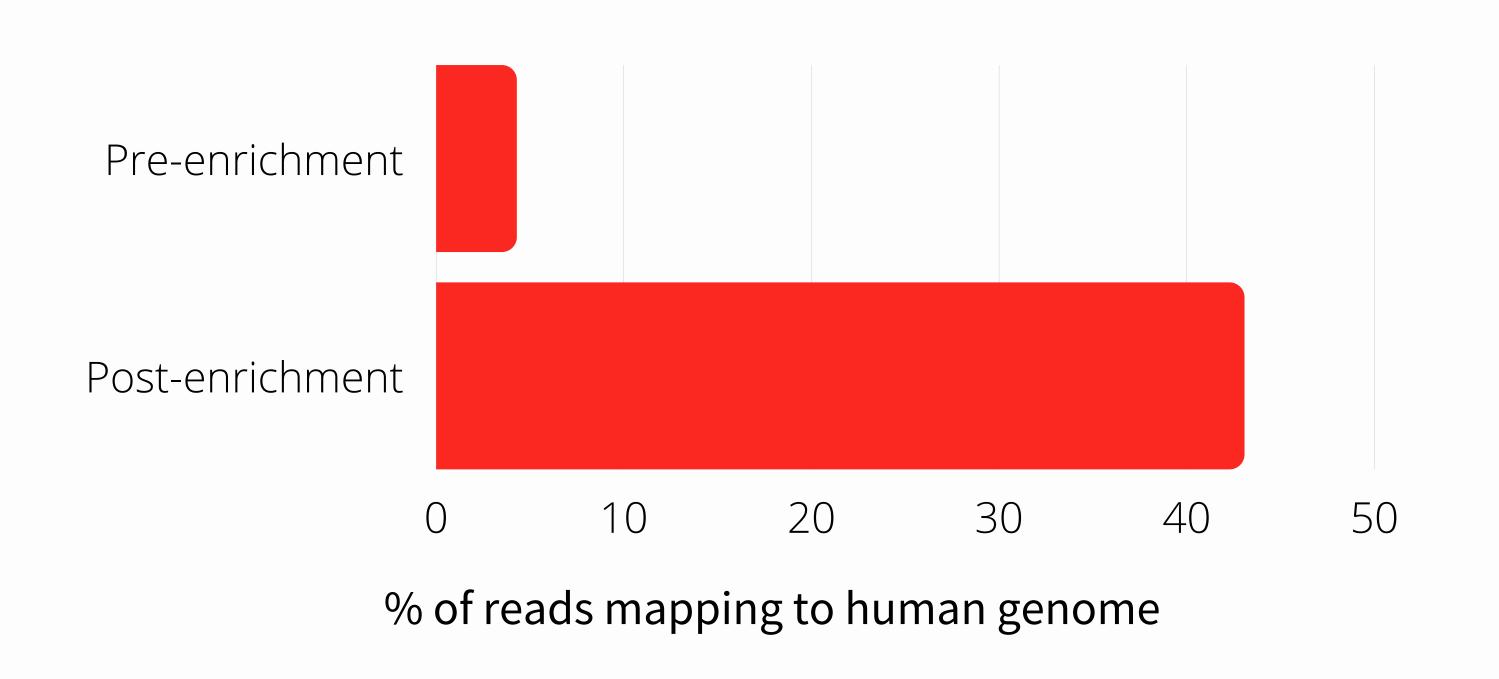
# HUMAN QUANT / QC

All four extracts were sent to HudsonAlpha Discovery of Huntsville, AL for library prep, QC, and sequencing. Unfortunately, despite the promising results from the fluorometric quant, the proportion of endogenous human DNA was determined to be quite low. This was reflected in the low ALU-based quant figures and mapping rates.

Sample	Human DNA (picograms)	Mapping rate (%)
Humerus (1)	14.25	4.14
Humerus (2)	18.65	4.29
Femur (1)	4.1	3.00
Femur (2)	3.85	2.92

### ENRICHMENT

One of the libraries with DNA extracted from the humerus was selected to undergo hybrid capture-based whole genome enrichment, which greatly improved the mapping rate.

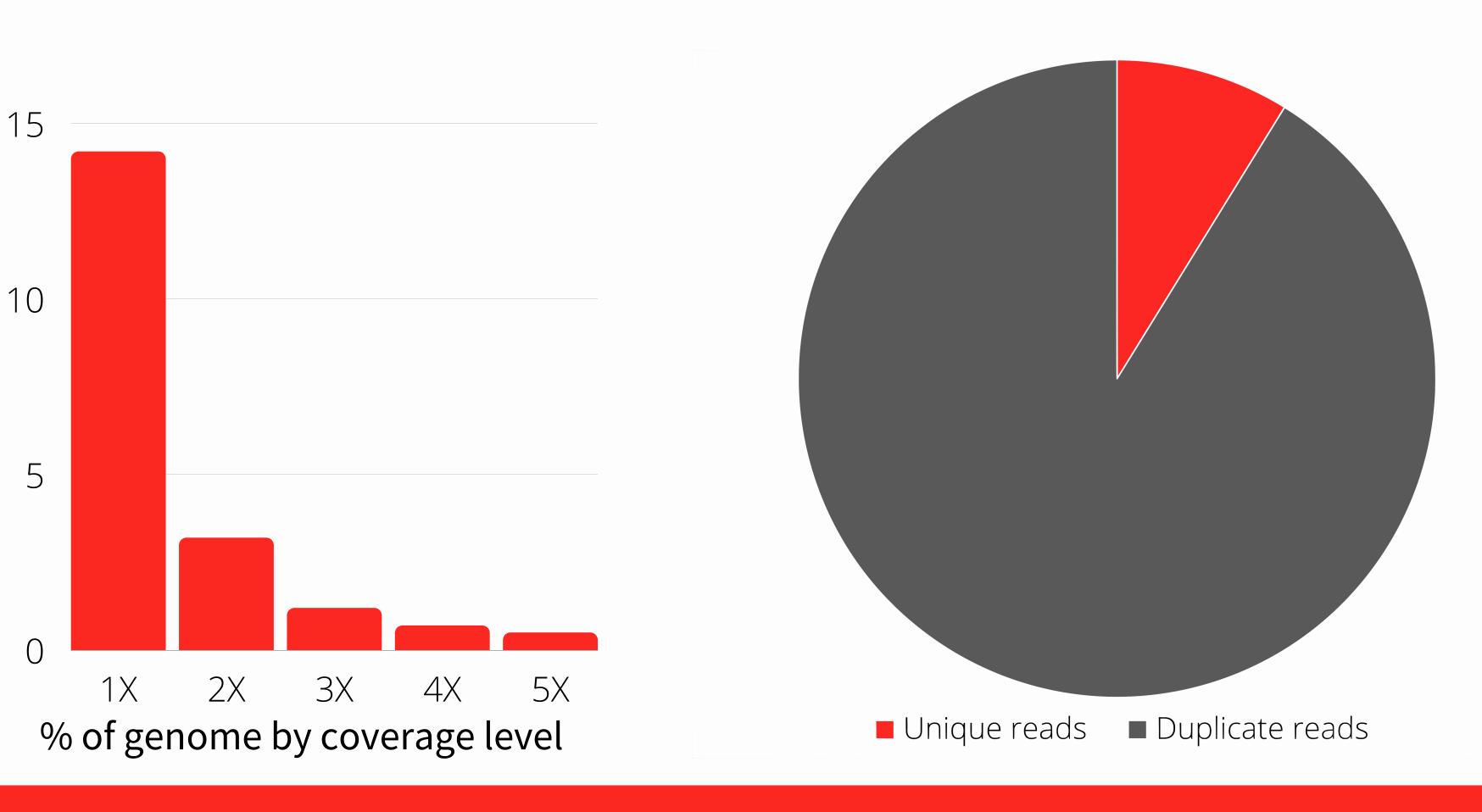


### BIOINFORMATICS

Saber Investigations of Marble Falls, TX handled the bioinformatics analysis. While enrichment greatly improved the mapping rate, sequencing results were still hampered by the low endogenous DNA content. This led to 91.1% of uninformative duplicate reads. Mean coverage was 0.3X and only 14.2% of the genome received at least 1X coverage (meaning >85% of the genome had no sequencing data at all). This necessitated a number of bioinformatics techniques to get the most out of the data.

- Mixture / contamination assessment
- Raw read pre-processing
- Imputation-based variant calling
- Filtering of variants to avoid "matchiness" in GEDmatch batching

A total of twelve different files using different filtering thresholds were created to ensure the best possible matching results were obtained.



## GENEALOGY RESEARCH

Once data was successfully uploaded to GEDmatch, a small team of volunteers began attempting to identify the John Doe. The top match shared approximately 169.1 cM (the amount of shared DNA, suggesting a relationship in the range of second to third cousins) with John Doe. After further research, it was determined that this person had a great-great-great grandfather that lived in the town in which he was discovered. Land survey maps from the 1850s showed he owned property in the approximate location where John Doe was discovered.

## TENTATIVE IDENTIFICATION

Based on the previously mentioned match and others, a tentative identification has been made in the case. He is believed to have been a man born in the late 1700s that died in the mid-1800s, and was likely buried on his family's property. The amount of DNA shared with the top match fits with what would be expected for a great-great grandchild relationship, though it's not usually encountered.