Warming Up Cold Cases:
An Introduction to Investigative Genetic Genealogy

Allison L Hade, DO
University of New Mexico
Department of Pathology
PGY4

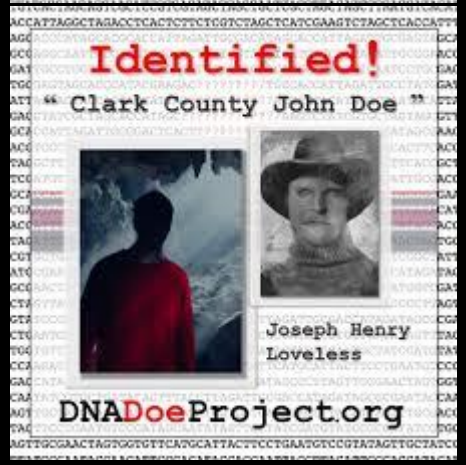
Official Disclosures

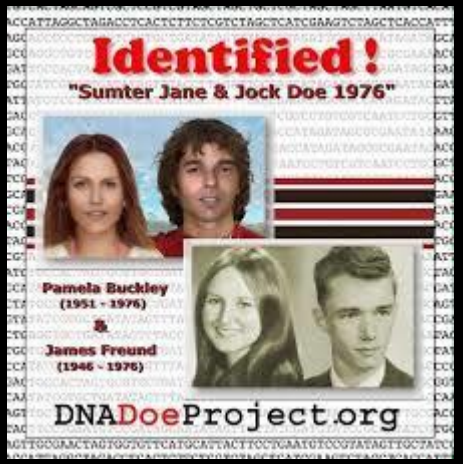
- Relevant Financial Relationships:
  - None
- Relevant Non-Financial Relationships:
  - None

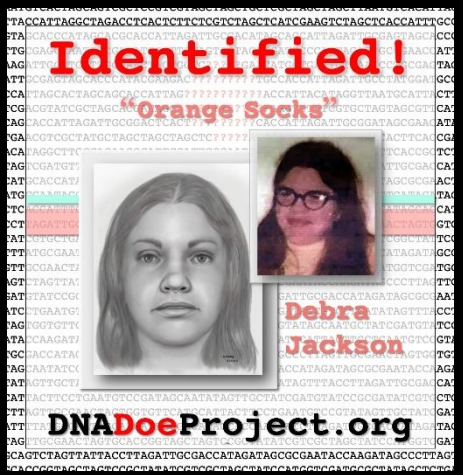
Special Thanks

- Dr. Daniel Babu, MD, UNM Department of Pathology, Associate Program Director
- Dr. Ryan Braunberger, MD, PGY4, UNM Department of Pathology
- Dr. Devon Chabot-Richards, MD, UNM Department of Pathology, Director of the Molecular Diagnostics and Oncology Laboratory
- Ainsley Cotter, National Center for Missing and Exploited Children, Forensic Case Manager
- Dr. Lauren Decker, MD, New Mexico Office of the Medical Investigator, Medical Investigator
- Dr. Lauren Dvorscak, MD, New Mexico Office of the Medical Investigator, Medical Investigator
- Dr. Heather Jarrell, MD, New Mexico Office of the Medical Investigator, Chief Medical Investigator

Joseph James DeAngelo:
Arrested April 24, 2018
The Golden State Killer







The DNA Doe Project:
70+ identifications as of 2021

Outline

- Introduction to Investigative Genetic Genealogy
- Challenges Involved in Discussing Investigative Genetic Genealogy
- Case Study: Dona Ana County Jane Doe
- Investigative Genetic Genealogy and Ethics
- The Legality of Investigative Genetic Genealogy
- Conclusions

Uses of DNA in Forensic Investigation

- DNA has become an increasingly important part of forensic investigations
- It is used to:
  - Identify suspects
  - Identify unknown remains
  - Separate unknown remains during mass casualty events

CODIS

- CODIS
  - Combined DNA Indexing System
  - Run and maintained by the FBI
    - Profiles uploaded include convicted offenders, arrestees, detainees, DNA profiles obtained from crime scenes, unidentified human remains, missing persons, and relatives of missing persons
  - Limitations:
    - 13 STRs for every profile
      - Only a preliminary match
        - Must be confirmed by a proper laboratory genetic match
    - Only a select number of profiles of only specific subsets of people

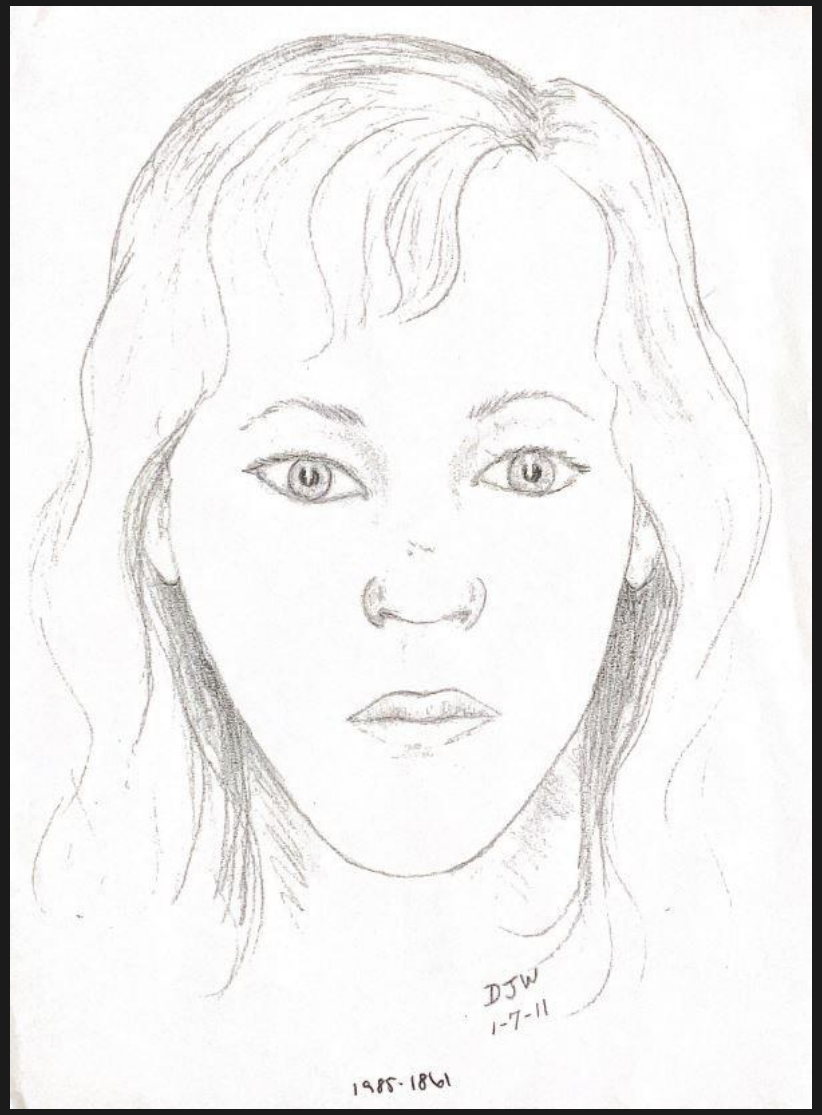
Investigative Genetic Genealogy: What Even Is It?

- Genetic information is utilized to infer distant relationships, then genealogic research creates a family tree to generate leads in law enforcement investigations
  - Unknown suspects
  - Unidentified persons/remains
- Utilizes forensically-collected DNA
  - Often damaged, degraded, or mixed
- Utilizes genetic testing and relatedness matching services through direct-to-consumer and third-party genetic databases

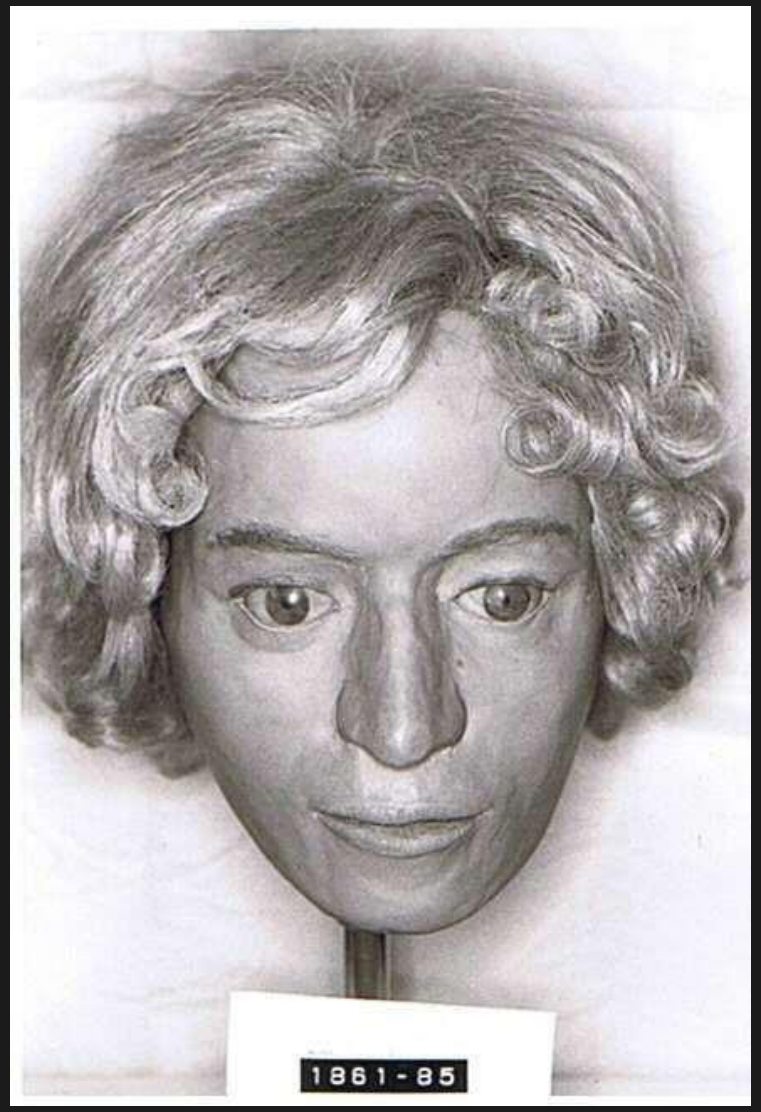
Challenges in Addressing Methods and Knowledge in Investigative Genetic Genealogy

- Fast-moving field
  - New companies and researchers pushing in new directions constantly
  - Most of the leading researchers and figures in the field are not scientists by training
- Proprietary concerns
  - All research is being conducted by private companies
  - Reluctant to share their methods and findings
    - Field has many black boxes

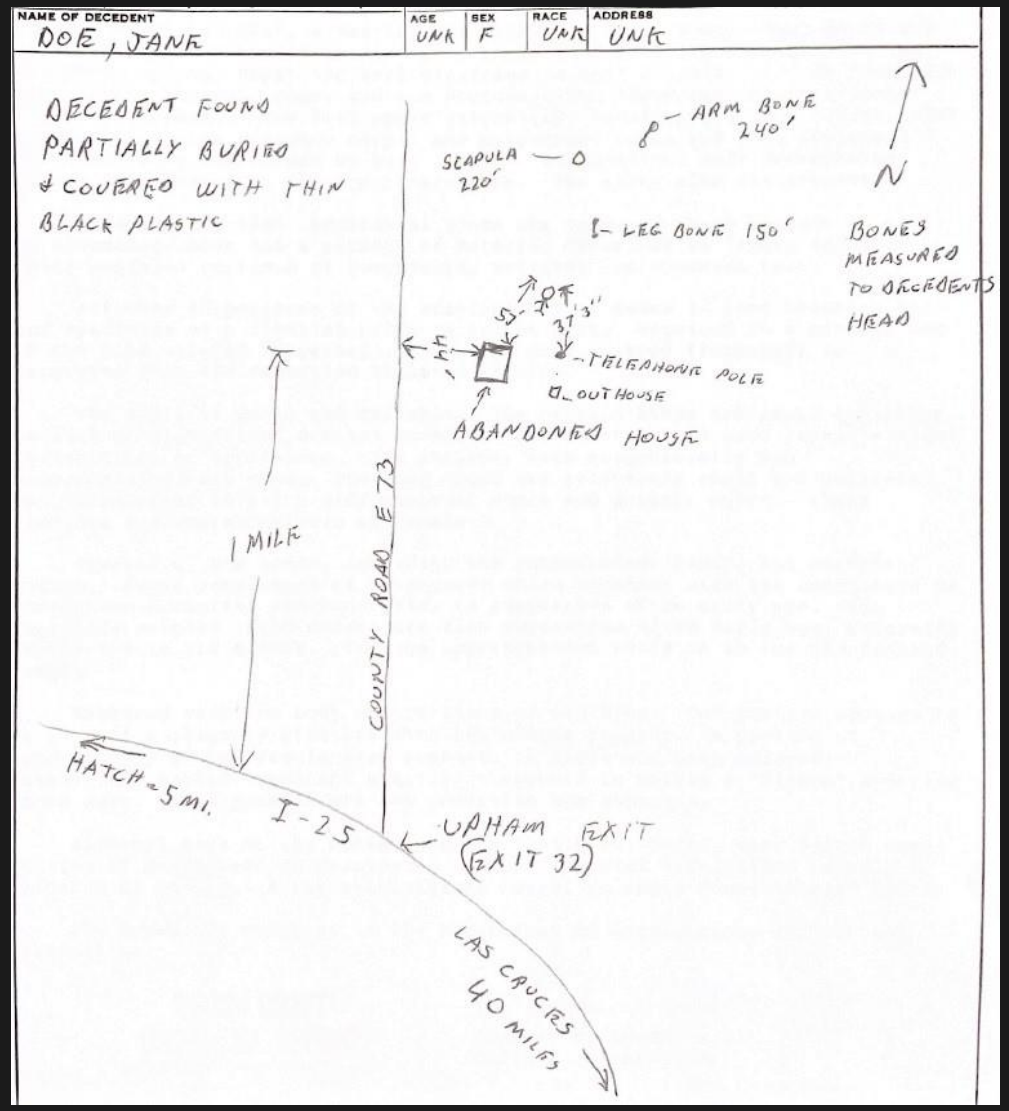
Case Study:
Dona Ana County Jane Doe



On March 10, 1985, human remains were discovered by hunters north of Hatch, NM off a well-traveled road by hunters and ranchers and near an abandoned house.



Dona Ana County Jane Doe: Crime Scene



CRIME SCENE PHOTO REMOVED FOR HANDOUT

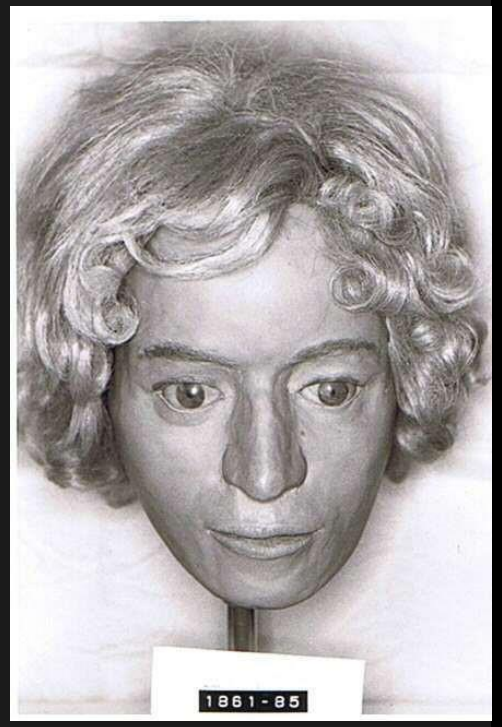
Dona Ana County Jane Doe: Initial Investigation

- Initial investigation included forensic anthropology by Dr. Stephen Adams, MD
- Skull was small and delicate, without closure of the cranial sutures.
  - Inominates, femur, and vertebral bodies also lacked full closure at the epiphyses
- Teeth were in good condition
  - Third molars partially erupted
- Inominates had a wide subpubic angle and sciatic notch
- Clothing included portions of bra, panties, and sweatshirt

SKELETAL REMAINS PHOTO REMOVED FOR HANDOUT

Dona Ana County Jane Doe: Continuing Investigation

- No dental fillings or alterations
  - No dental records for comparison
- Made a facial reconstruction by skull
- DNA was first entered in CODIS in 2012
  - Profile did not have all 13 STRs collected for CODIS
    - Limited utility
- Multiple attempts to ID by DNA through the years
  - Multiple possible relatives came forward
    - Direct comparisons of skeletal DNA to possible family member DNA, all ruled out
  - No hits, and no further evidence



Genetic Investigations: NAMUS

- NAMUS
  - National Missing and Unidentified Persons System
    - Free DNA, odontology, fingerprint examination, and anthropology tools
    - Best and most complete resource for biometrics for all missing persons cases
      - Up to individual LE departments and agencies to keep these profiles and biometrics current
  - Overseen by the National Institute of Justice (NIJ)
    - Currently run by RTI International, a nonprofit research institute
  - Attempting to develop a unit to assist with genealogy work
    - Long time coming and still in the beginning stages

NCMEC:
Forensic Service Unit

- National Center for Missing and Exploited Children
  - Missing Children Unit
    - Active missing persons cases
  - Child Sex Trafficking Analytical Team and Child Sex Trafficking Recovery Services Team
  - Forensic Service Unit
    - New team focusing on unidentified juvenile remains
    - Currently 2 case managers: Ainsley Cotter and Julia Vekasy-Quillin
    - About 400 cases concurrently
    - Offer services to LE and MEs:
      - Originating team retains lead investigator status unless requested
      - Offers partnership with the anthropology resources through the Smithsonian
      - FBI and Secret Service labs
      - Investigation and case management resources for departments that do not have budget or personnel to dedicate to cold cases

Dona Ana County Jane Doe: NCMEC Involvement

- Dona Ana County put the case onto NCMEC's waiting list in 2011
  - Dona Ana County Sherriff's department initially not responsive to NCMEC's requests for biometrics
  - NCMEC did not move on the case for this reason
- In 2021, Astrea Forensics, a laboratory partner of NCMEC, received grant money to do genetic genealogy work with specific materials
  - Approached Ainsley Cotter, forensic case manager with the Forensic Services Unit
  - Allowed her to select a case
- Ms. Cotter reached out to Dona Ana County Sherriff's Office and the New Mexico Office of the Medical Investigator in 2021
  - Dona Ana County Sherriff had a new officer assigned to the case who was very responsive with consent and cooperation
  - OMI proceeded with sample collection of hair and the femur

Paying for IGG

- Genetic testing is expensive, especially sequencing degraded forensic samples
- Some organizations provide funding for specific demographics
  - NCMEC provides services for unidentified remains suspected to be around 21 years or younger
- Some labs do pro bono work
  - Charley Project is a platform where missing persons cases can be publicized
    - Some labs monitor their cases and offer pro bono services
- Biggest source of funding is crowdsourcing
  - DNASolves has media outreach, crowdsourcing assistance
    - Considered the best resource for agencies to crowdsource funding for IGG at this point
    - Attempting to start their own competitor to GEDMatch
  - True crime community can be an asset in this area, though complicated

IGG Genetic Analysis: Whole Genome Sequencing

- Random fragmentation to create many fragments with overlapping sequences
- Read the various sequences millions of times after adapters are attached to allow for reading
- These reads of individual fragments are used to reconstruct larger and larger segments until the entire genome is read in usable data
  - Compared to other fragments in the sample (de novo)
  - Compared to a known ‘template’ genome (reference-based assembly)

IGG Genetic Analysis: DNA Extraction/Purification

- Sample Collection
  - Hair or bone powder
- DNA Extraction/Purification
  - Cells are lysed
  - Cellular debris is removed via centrifugation
  - Protease and RNAse treatments
  - DNA can be isolated with cold ethanol treatment

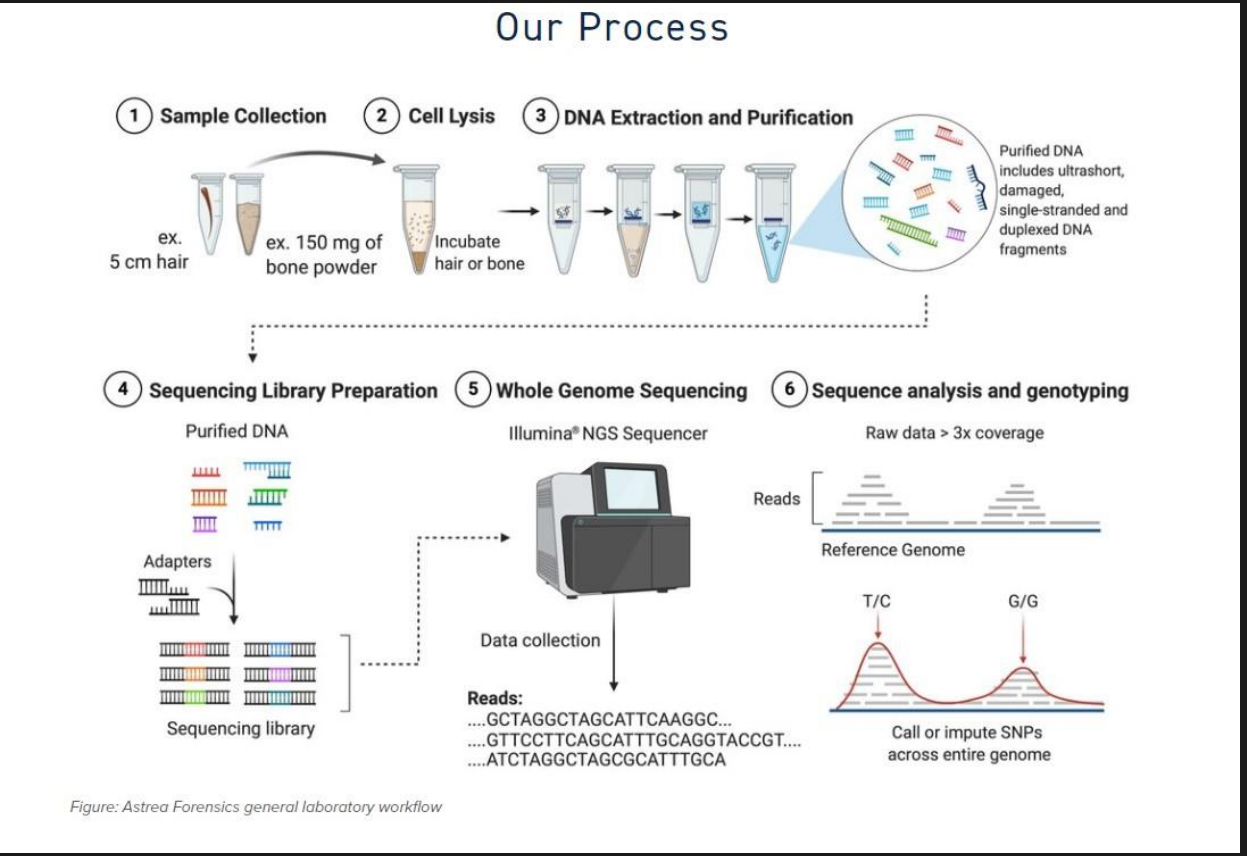
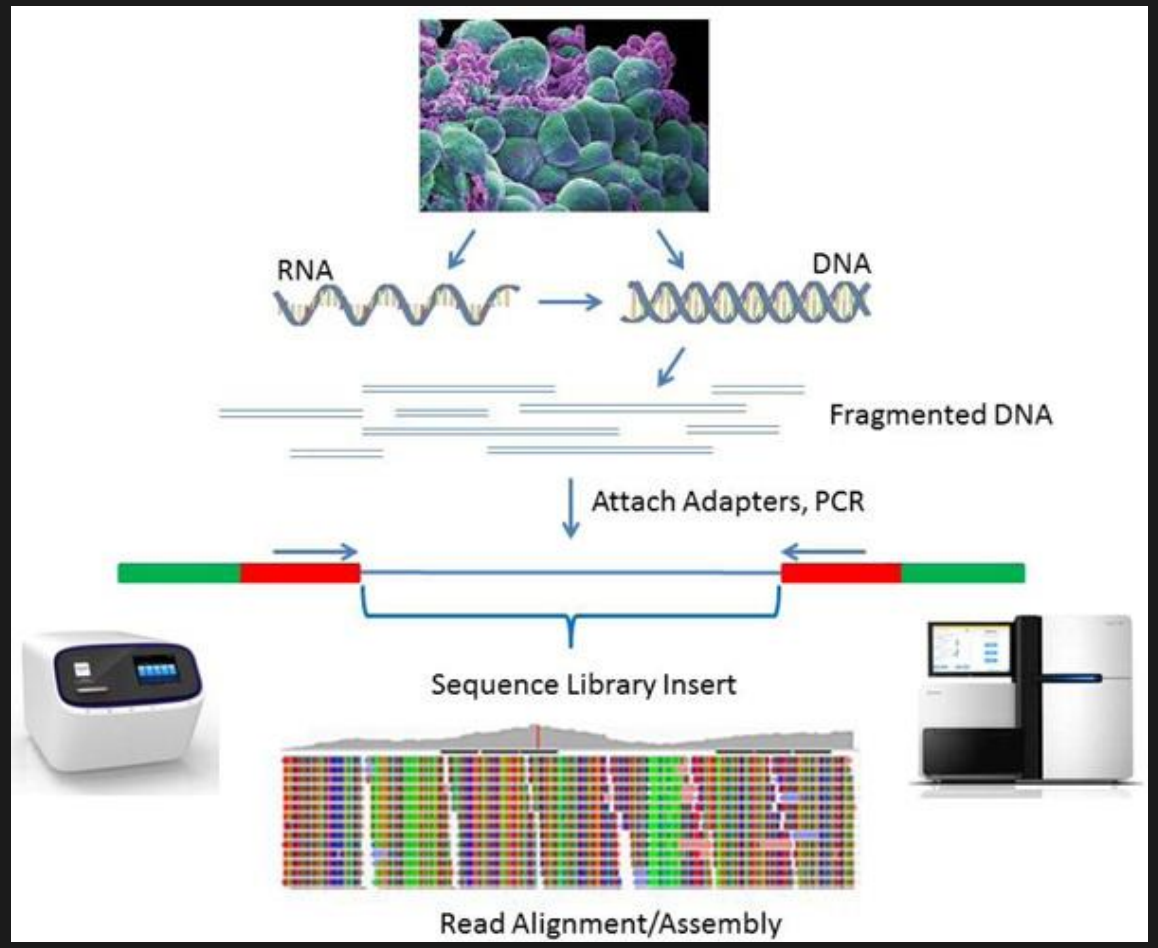


Figure: Astrea Forensics general laboratory workflow

IGG Genetic Analysis:
NGS Library Preparation and Initial Amplification

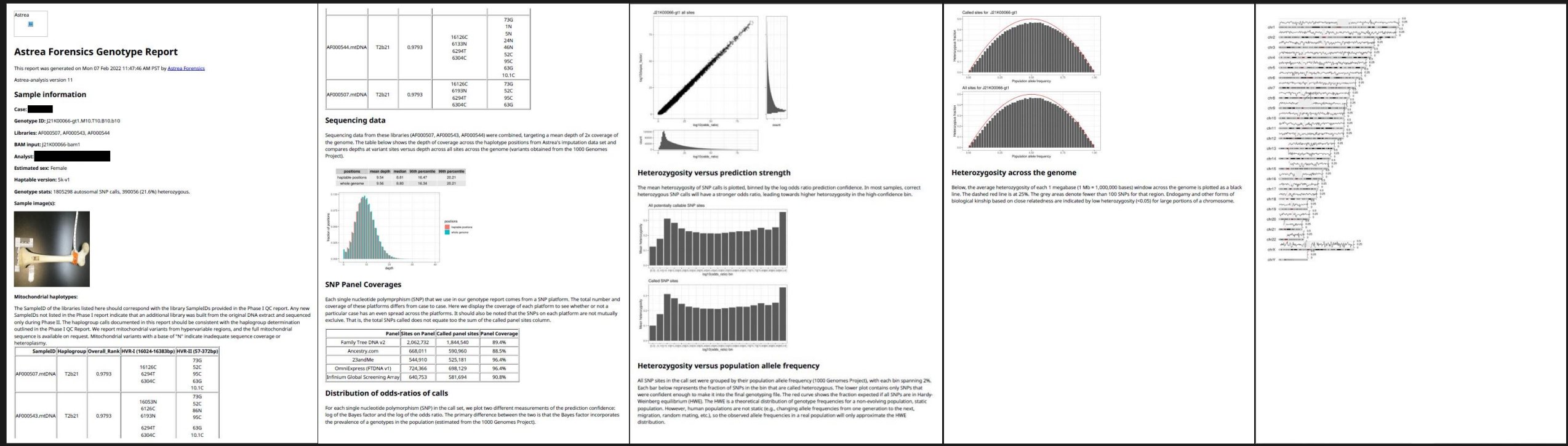
- DNA is fragmented
  - Physical or enzymatic techniques most common
- Ligation of adapters
- PCR cycles to enrich for product that has adapters ligated to both ends



IGG Genetic Analysis:
Phase 1 Completion QC Report

- *Phase I Deliverable*
- A summary QC report contains, whenever possible:
  1. **Sequencing metrics,**
  2. **Sex estimate**,
  3. **Mitochondrial lineage** determination,
  4. **Complete mitochondrial genome sequence**, if desired
  5. **Comparison** of nuclear DNA data between samples.
  6. **Recommendations** and information for next steps

Dona Ana County Jane Doe Astrea Report



Astrea

Astrea Forensics Genotype Report

This report was generated on Mon 07 Feb 2022 11:47:46 AM PST by Astrea Forensics

Astrea-analysis version 11

Sample information

Case: [redacted]

Genotype ID: J21K00066-gt1.M10.T10.B10.b10

Libraries: AF000507, AF000543, AF000544

BAM input: J21K00066-bam1

Analyst: [redacted]

Estimated sex: Female

Haptable version: 5k-v1

Genotype stats: 1805298 autosomal SNP calls, 390056 (21.6%) heterozygous.

Sample image(s):

Mitochondrial haplotypes:

The SampleID of the libraries listed here should correspond with the library SampleIDs provided in the Phase I QC report. Any new SampleIDs not listed in the Phase I report indicate that an additional library was built from the original DNA extract and sequenced only during Phase II. The haplogroup calls documented in this report should be consistent with the haplogroup determination outlined in the Phase I QC Report. We report mitochondrial variants from hypervariable regions, and the full mitochondrial sequence is available on request. Mitochondrial variants with a base of "N" indicate inadequate sequence coverage or heteroplasmy.

SampleID	Haplogroup	Overall_Rank	HVR-I (16024-16383bp)	HVR-II (57-372bp)
AF000507.mtDNA	T2b21	0.9793	16126C 6294T 6304C	73G 52C 95C 63G 10.1C
AF000543.mtDNA	T2b21	0.9793	16053N 6126C 6193N 6294T 6304C	73G 52C 86N 95C 63G 10.1C
AF000544.mtDNA	T2b21	0.9793	16126C 6133N 6294T 6304C	73G 1N 5N 24N 46N 52C 95C 63G 10.1C
AF000507.mtDNA	T2b21	0.9793	16126C 6193N 6294T 6304C	73G 52C 95C 63G

Sequencing data

Sequencing data from these libraries (AF000507, AF000543, AF000544) were combined, targeting a mean depth of 2x coverage of the genome. The table below shows the depth of coverage across the haplotype positions from Astrea's imputation data set and compares depths at variant sites versus depth across all sites across the genome (variants obtained from the 1000 Genomes Project).

positions	mean depth	median	95th percentile	99th percentile
haptable positions	9.54	8.81	16.47	20.21
whole genome	9.56	8.80	16.34	20.21

SNP Panel Coverages

Each single nucleotide polymprphism (SNP) that we use in our genotype report comes from a SNP platform. The total number and coverage of these platforms differs from case to case. Here we display the coverage of each platform to see whether or not a particular case has an even spread across the platforms. It should also be noted that the SNPs on each platform are not mutually exclusive. That is, the total SNPs called does not equate too the sum of the called panel sites column.

Panel	Sites on Panel	Called panel sites	Panel Coverage
Family Tree DNA v2	2,062,732	1,844,540	89.4%
Ancestry.com	668,011	590,960	88.5%
23andMe	544,910	525,181	96.4%
OmniExpress (FTDNA v1)	724,366	698,129	96.4%
Infinium Global Screening Array	640,753	581,694	90.8%

Distribution of odds-ratios of calls

For each single nucleotide polymorphism (SNP) in the call set, we plot two different measurements of the prediction confidence: log of the Bayes factor and the log of the odds ratio. The primary difference between the two is that the Bayes factor incorporates the prevalence of a genotypes in the population (estimated from the 1000 Genomes Project).

Heterozygosity versus prediction strength

The mean heterozygosity of SNP calls is plotted, binned by the log odds ratio prediction confidence. In most samples, correct heterozygous SNP calls will have a stronger odds ratio, leading towards higher heterozygosity in the high-confidence bin.

Heterozygosity across the genome

Below, the average heterozygosity of each 1 megabase (1 Mb = 1,000,000 bases) window across the genome is plotted as a black line. The dashed red line is at 25%. The grey areas denote fewer than 100 SNPs for that region. Endogamy and other forms of biological kinship based on close relatedness are indicated by low heterozygosity (<0.05) for large portions of a chromosome.

Heterozygosity versus population allele frequency

All SNP sites in the call set were grouped by their population allele frequency (1000 Genomes Project), with each bin spanning 2%. Each bar below represents the fraction of SNPs in the bin that are called heterozygous. The lower plot contains only SNPs that were confident enough to make it into the final genotyping file. The red curve shows the fraction expected if all SNPs are in Hardy-Weinberg equilibrium (HWE). The HWE is a theoretical distribution of genotype frequencies for a non-evolving population, static population. However, human populations are not static (e.g., changing allele frequencies from one generation to the next, migration, random mating, etc.), so the observed allele frequencies in a real population will only approximate the HWE distribution.

Phase II Deliverable

Summary Report (see above)

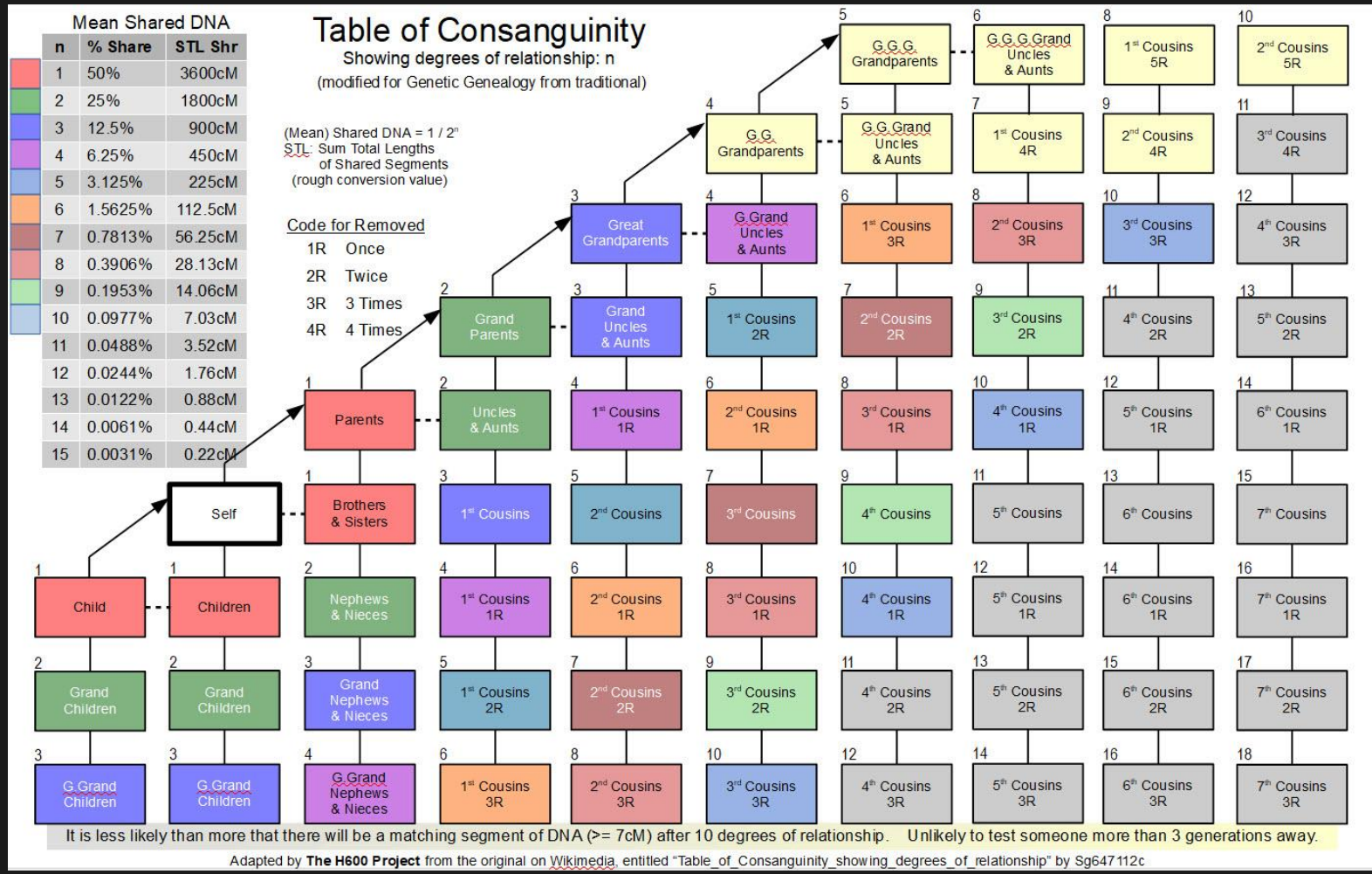
Genotype text file containing between 1-2 million SNPs in a format suitable for upload to genetic genealogy databases.

Genetic Genealogy Work

- With the genome sequenced, the case shifted from Astrea to Innovative Forensic Investigations
- Uploaded genome library to GEDMatch
  - Third-party hobbyist website
    - Profiles uploaded from users
      - Accepts profiles from 20+ Direct-to-Consumer companies
      - Also raw data from microarrays and whole-genome sequencing
    - Ancient DNA from artifacts or deceased people also accepted
  - LE uploads are only accepted on GEDMatch Pro
    - Allows GEDMatch to control how these datasets are used and what they can match against
      - Unidentified human remains may be compared against the entire database
      - Profiles to identify perpetrators can only be matched to an opt-in portion of the database

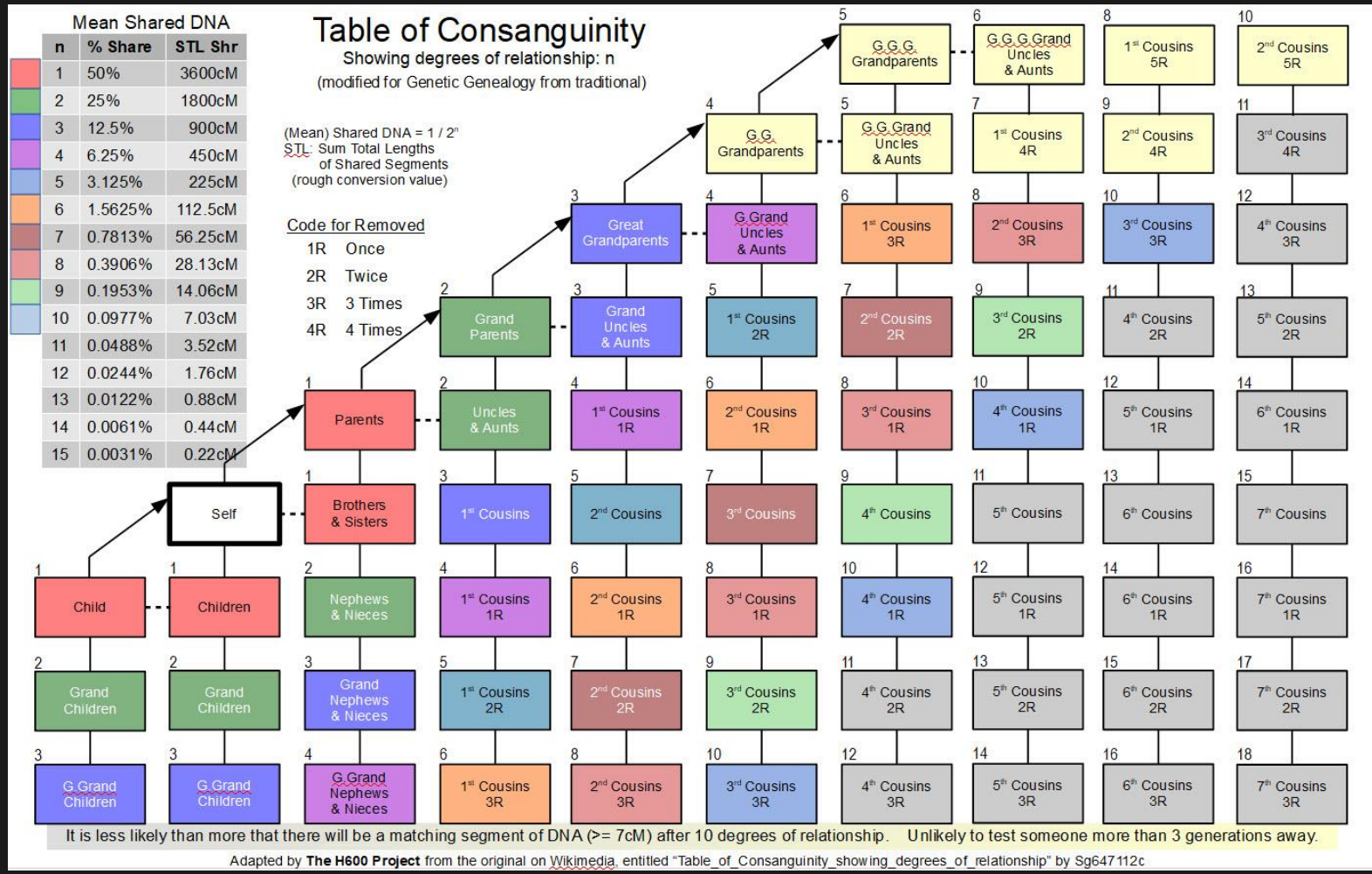
GEDMatch Process

- SNP profile is uploaded
- GEDMatch parses SNP data to ensure usability
- Profile is assigned a kit number and compressed to site-specific binary ‘token file’
- Uses their One-to-Many tool to look for all SNPs in common between any 2 kits



Degree of Relatedness

- The closer a person is related, the longer the segments of DNA they share
  - 50% of DNA should be shared by siblings
    - Long stretches of SNPs shared
  - More distant relations will have less DNA shared
    - Shorter segments shared
- 1st degree = siblings or parents; 2nd = 1st cousins, aunts, uncles; etc
  - Different average lengths of cM shared and different percent of SNPs shared, depending on degree
- For genetic genealogy, error rates increase the less DNA is shared
  - Calculated degrees of relatedness useful out to 3rd cousins
  - Closer relations are more useful, as there is less error and the degree calculation can be trusted more



Genealogy Research in IGG

- Algorithm generates match lists with shared cM length
  - Profiles with identifiable second or third cousins are considered viable for investigation
- Family trees of suspected matches are generated through genealogical techniques
  - Work back to find a common ancestor, then conduct descendancy research
  - Look for points of intersection between different reported relations in family trees
- Cannot determine individual identity, only degree of relation

Dona Ana County:
Social Media and Genealogy Research in IGG

- In this case, once suspected first-degree relatives were identified by genealogy Facebook was used to research the family
- One daughter of this family was mentioned but had no photos
  - Only referred to in the past tense
- This was enough evidence for Dona Ana County to contact the suspected mother and ask her if she had a daughter who went missing in the 1980s
- She had a missing daughter, and agreed to a DNA test

Dorothy Harrison

- Went missing at age 16
- Last seen July 25, 1984 when she left her home in Wichita, Kansas
- Got into a car with two other women and was not seen again
  - Frequently left home with friends
- Last heard from in August, 1984 when she contacted her family by hone
  - She was in Los Angeles for the Olympics
  - Called them again from El Paso
    - Claimed to be on her way home
- Murderer remains unidentified

Genetic Genealogy Research: Success Rates

- In databases with at least 1 million individuals from a similar genetic background to that of the uploaded SNP profile, >95% likelihood of a third cousin match
- Individuals of European ancestry much more likely to get a match
  - 75% of the large database uploads are of predominately Northern European genetic ancestry
  - 60% of all people of European genetic ancestry will have a third cousin or closer match
  - 15% will have second cousin or closer match
- But what about people of predominately non-European genetic ancestry?

Welcome to the Can of Worms: IGG and Ethics

- Database biases
  - Particularly affect those of non-European descent
- Surreptitious DNA testing
  - Collection, upload, and retention
- Lack of independent regulatory oversight
- Opt-in vs opt-out policies

IGG in the Courts

- IGG can only be used to generate leads
  - It cannot be the sole evidence on which arrests or identifications are made
  - Often only generates leads as far an extended family
- Corroborating evidence should include:
  - A confirmed match resulting from a direct comparison of the STR profile of the remains or evidence to a living suspect or relative
  - Even if close relative or suspect is directly in the genealogy database, corroborative testing is required
- IGG alone is not admissible in criminal cases, and can only be used as supporting evidence for other genetic testing
- Limited to the identification of human remains and investigation of violent crimes
  - Only after reasonable investigative leads have already been pursued

Conclusion

- Investigative genetic genealogy is a fast-moving and complex field
- It can be a useful tool for identifying remains in cold cases that have exhausted other avenues of investigation
- There are challenges in utilizing IGG in practice
  - There are funding sources and assistance available
- Making use of this new tool in forensics and medical examination could yield new movement in cold cases

Works Referenced

- CODIS and NDIS Fact Sheet. FBI Official Database. https://www.fbi.gov/how-we-can-help-you/dna-fingerprint-act-of-2005-expungement-policy/codis-and-ndis-fact-sheet. Accessed October 13, 2022.
- Cotter, Ainsley. National Center for Missing an Exploited Children, Forensic Case Manager. Interviewed by Allison Hade on September 22, 2022.
- Dawnay N, Flamson R, Hall MJR, Steadman DW. “Impact of sample degradation and inhibition on field-based DNA identification of human remains.” Forensic Sci Int Genet. 2018 Nov;37:46-53. doi: 10.1016/j.fsigen.2018.07.018. Epub 2018 Jul 25. PMID: 30077890.
- “Dorothy Harrison.” Unidentified Wiki. https://unidentified-awareness.fandom.com/wiki/Dorothy_Harrison. Accesed October 17, 2022.
- Guerrini CJ, Wickenheiser RA, Bettinger B, McGuire AL, Fullerton SM, “Four misconceptions about investigative genetic genealogy,” J Law and the Biosciences, January-June 2021; 8(1), lsab001, https://doi.org/10.1093/jlb/lsab001
- Gupta N. “DNA Extraction and Polymerase Chain Reaction.” J Cytol. 2019 Apr-Jun;36(2):116-117. doi: 10.4103/JOC.JOC_110_18. PMID: 30992648; PMCID: PMC6425773.
- H600 Project. “Cosanguinity.” Official Website. https://h600.org/wiki/Consanguinity. Accessed October 14, 2022.
- Head SR, Komori HK, LaMere SA, Whisenant T, Van Nieuwerburgh F, Salomon DR, Ordoukhanian P. “Library construction for next-generation sequencing: overviews and challenges.” Biotechniques. 2014 Feb 1;56(2):61-4, 66, 68, passim. doi: 10.2144/000114133. PMID: 24502796; PMCID: PMC4351865.
- Kling D, Phillips C, Kennett D, Tillmar A. “Investigative genetic genealogy: Current methods, knowledge and practice.” Forensic Sci Int Genet. 2021 May;52:102474. doi: 10.1016/j.fsigen.2021.102474. Epub 2021 Jan 30. PMID: 33592389.
- National Missing and Unidentified Persons System (NAMUS). Official Website. https://namus.nij.ojp.gov/about. Accessed Sept 14, 2022.
- Ng PC, Kirkness EF. Whole genome sequencing. Methods Mol Biol. 2010;628:215-26. doi: 10.1007/978-1-60327-367-1_12. PMID: 20238084.
- RTI International. “RTI Awarded National Missing and Unidentified Persons System Contract.” 2021, May 4. RTI International Website. https://www.rti.org/announcements/rti-awarded-national-missing-and-unidentified-persons-system-contract. Accessed Sept 14, 2022.