DNA DOE Project
Our First Two Years of Success

ISHI 30
24 Sep 2019

Colleen Fitzpatrick, PhD
Margaret Press, PhD
Co-Founders
DNA Doe Project
DNA DOE Project Chronology

- Feb 2017  First discussions
- Jul 10, 2017  First sample sent to lab for sequencing
- Sep 27, 2017  Joseph Chandler case uploaded to GEDmatch
- Oct 2017  Incorporated as 501(c)3
- Mar 6, 2017  Joseph Newton Chandler identified
- Mar 29, 2017  Buckskin Girl Identified
DNA Doe Project Stats
Sept 2019

Volunteers ~ 70

Shortest Solve Time 4 hrs

Longest Solve Time 19 ½ mos

Avg Solve Time 60 days (no outliers)
<table>
<thead>
<tr>
<th>Category</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Waiting shipment to lab</td>
<td>3</td>
</tr>
<tr>
<td>Bone extraction stage</td>
<td>6</td>
</tr>
<tr>
<td>DNA Processing</td>
<td>19</td>
</tr>
<tr>
<td>Bioinformatics</td>
<td>7</td>
</tr>
<tr>
<td>Genealogical Research</td>
<td>17</td>
</tr>
<tr>
<td>Success Stories</td>
<td>16</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>68</strong></td>
</tr>
</tbody>
</table>

- **~50% Success rate**
- **Only 1 slam dunk**
Initial Challenge of Using Degraded DNA

Question: Would genetic genealogy tools work for degraded samples?

Joseph Newton Chandler III

- DNA was degraded – 88% No calls
- Nonuniform SNP Distribution
**Initial Challenge - Degraded DNA**

- Experimented with Confidence Levels

High confidence: Few matches

We found a happy medium

Low confidence: Many matches

We still use a three-kit model
Initial Challenge - Degraded DNA

- Degraded Our Own 23&Me data to match Chandler’s, compared matches before & after

![Graph showing rank on GEDmatch before and after degradation]

Closest matches are the same rank
Initial Challenge - Degraded DNA

- Two 30x sequences ➔ One 60x sequence

⇒ Initial 88% decreased to 55% no calls
⇒ More matches
⇒ Solved the case
Lessons Learned

• Single most important decision: **Who will process the DNA sample?**
  - Your sample will be consumed
  - With wrong type of processing the evidence will be lost forever

• What is the quality/quantity of your DNA sample?
  - Is it degraded or contaminated?
  - Can the company handle compromised DNA?

• Genealogy can always be redone
Quantity vs Quality

- **Microarray**: ~70% Call Rate
- **Whole Genome Sequencing**:
  - > 1 ng: 70% Call Rate
  - > 4 ng: 10% Call Rate
- **Both**: ~1 ng
  - 10% Call Rate

DNA Project
Advantages of WGS over Chip-Based Technology

• You have the whole genome on your hard drive
  - No further degradation
  - Sample will not be exhausted

• Unaffected by changes in DTC SNP testing platforms

• More robust when applied to degraded DNA

• Sequence can be reused to create multiple GEDmatch kits

• WGS includes mtDNA, Y-DNA, CODIS markers, SNPs

• Sequences can be combined to improve data
What You Must Know about GEDmatch

• Terms of Service – Member default opt-out for law enforcement use

• Users must now opt-in to allow law enforcement to use their accounts

• Aggressive campaign to get users to opt-in

• August 2019 = 120,000 opt-ins = ~10%
What You Must Know about FTDNA

- Terms of Service – Member default opt-in for law enforcement use
- Law enforcement required to complete submission form, identify themselves
- Only top 20 to 40 matches are shown
- Law enforcement upload costs $700
- If you have a close match, you receive reports, not access
Three Main Challenges

- **Manpower limitations**
- **Condition of DNA**
- **Database composition**

70 Volunteers on waiting list
DNA Challenges & Success Stories

Heavy contamination
- 99% bacterial contamination
- Proprietary bioinformatics software

Degraded and contaminated DNA
- 99% no calls, improved with imputation
- Only 7.2% human DNA

Low input DNA
- WGS successful with 4 ng
- Highest match 1C1R despite database restrictions
Database Challenges & Success Stories

Endogamy
- Matches share multiple family lines
- Follow segments, find non-endogamous lines

High incidence of unknown parentage in database
- Matches who do not know their parents
- Need extra time to solve a mystery to solve the mystery

Recent immigration
- Father from another country
- Genetic networks, segment analysis, Y & mtDNA
Another Database Challenge & Success Story

GEDmatch opt-out default and FTDNA restrictions
- Solutions evolving, pushing tools
- New Jane Doe case solved in 10 days with only opt-ins

Special Mention

Ethnic composition of database
- Some ethnicities are under-represented
- Half of unsolved cases Native American/Hispanic women
- Prompted outreach through our NA/Hispanic Task Force
Range of Difficulty

Buckskin Girl
- Four clock hours
- ~ 20 Man-hours
- About 10 volunteers
- 1st cousin once removed

Mill Creek Shed Man
- 19 1/2 Months
- Estimated 15k Man-hours
- ~ 30 Volunteers
- Highly contaminated DNA
- African American ancestry
Acknowledgements

Dr. Gregory Magoon
Aerodyne Research

Justin Loe
Full Genomes Corporation

Dr. Weining Tang
Genologue

UNGHTSC
GEDmatch

FamilyTreeDNA

Lucigen Corp.
Our generous donors
NCMEC

HudsonAlpha Institute
for Biotechnology
DNA Solutions

Othram, Inc.

Fulgent Therapeutics, LLC
Bode Cellmark Forensics, Inc.

Our wonderful volunteers

DNA Doe Project
For Further Information

DNA Doe Project

www.DNADoeProject.org

www.facebook.com/DNADoeProject

admin@dnadoeproject.org

Forensic Genealogy Online Training

Forensicgenealogytraining.org