# Mitochondrial DNA Sequencing: the Next Generation

Forensics @ NIST
November 9, 2020
**Kevin Kiesler**
Research Biologist
NIST Applied Genetics Group

1

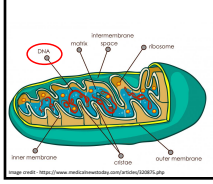
## Outline

- mtDNA background
- Applications
- Commercial kits
- Home brew methods
  - Degraded DNA
  - Reference samples
- Analysis software

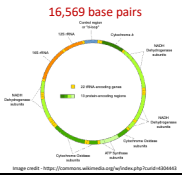

2

## Mitochondrial DNA

- Circular – 16.5 Kb
- Many copies – 100(0)s of mtGenome molecules per cell
- Maternally inherited – lineage marker, non-recombining



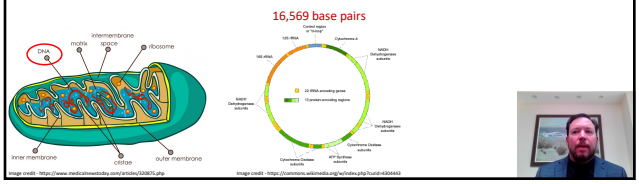

16,569 base pairs



3

## Mitochondrial DNA - Disadvantages

- Single marker – not as statistically powerful as nuclear markers
  - E.G. match probability with Short Tandem Repeats (STRs)
- Siblings have same mtGenome, some cousins/2nd cousins/3rd cousins...






4

## Mitochondrial DNA - Disadvantages???

- Single marker – not as powerful as nuclear markers (STRs)
  - E.G. match probability with Short Tandem Repeats (STRs)
- Siblings have same mtGenome, some cousins/2nd cousins/3rd cousins...


Photo: https://www.mirror.co.uk/news/uk-news/king-richard-iiis-descendant-makes-5378182

5

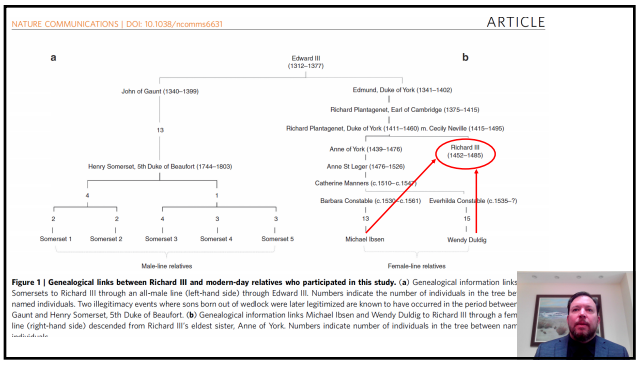



**Figure 1 | Genealogical links between Richard III and modern-day relatives who participated in this study.** (**a**) Genealogical information links Somersets to Richard III through an all-male line (left-hand side) through Edward III. Numbers indicate the number of individuals in the tree be named individuals. Two illegitimacy events where sons born out of wedlock were later legitimized are known to have occurred in the period between Gaunt and Henry Somerset, 5th Duke of Beaufort. (**b**) Genealogical information links Michael Ibsen and Wendy Duldig to Richard III through a fem line (right-hand side) descended from Richard III's eldest sister, Anne of York. Numbers indicate number of individuals in the tree between nam



6

## Applications of mtDNA Sequencing

- Missing persons
  - Bones, teeth, hair
- Mass disasters
  - World trade center 2001
    - 44,000 + mtDNA profiles generated
  - Tsunami in Thailand 2004
    - 507 teeth recovered
    - 258 identified (51 %)
- War graves

7

## Ancestry Prediction with mtDNA

- Genealogical relationships
  - Many generations
  - Haplogroups
- Geographical association
  - Migration

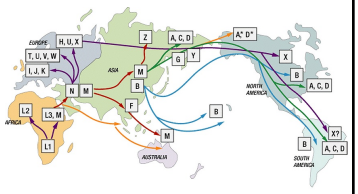
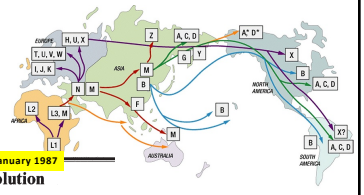



Image credit: www.familytreeDNA.com

8

## Ancestry Prediction with mtDNA

- Genealogical relationships
  - Many generations
  - Haplogroups
- Geographical association
  - Migration




Nature Vol 325 January 1987

**Mitochondrial DNA and human evolution**

Rebecca L. Cann*, Mark Stoneking & Allan C. Wilson

Department of Biochemistry, University of California, Berkeley, California 94720, USA

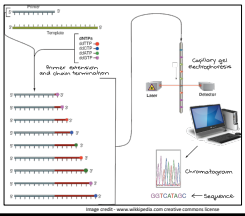
*Mitochondrial DNAs from 147 people, drawn from five geographic populations have been analysed by restriction mapping. All these mitochondrial DNAs stem from one woman who is postulated to have lived about 200,000 years ago, probably in Africa. All the populations examined except the African population have multiple origins, implying that each area was colonised repeatedly.*

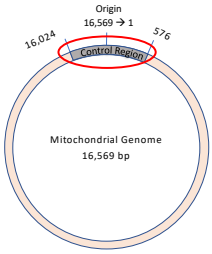
Image credit: www.familytreeDNA.com

9

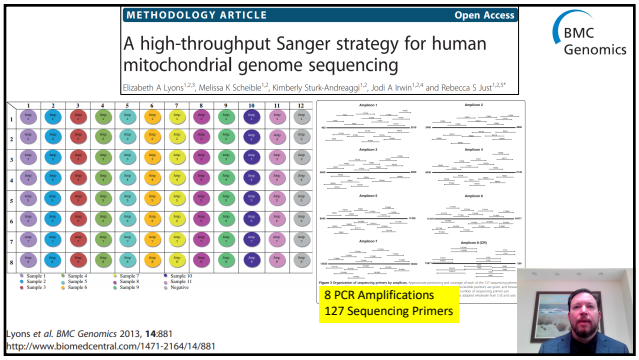
## How do we get mtDNA sequence?

- Sanger sequencing
  - ~ 800 bases per reaction



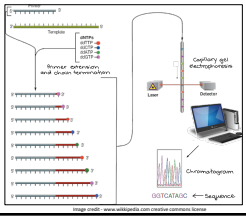

10

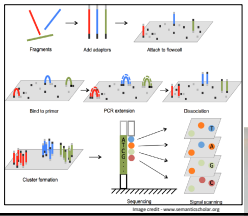



11

## How do we get mtDNA sequence?

- Sanger sequencing
  - ~ 800 bases per reaction
- Next generation sequencing
  - ~ 5 x $10^9$ bases per reaction

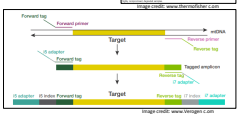



12

## Commercial Kits – A Few Options

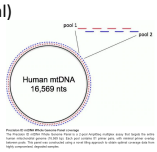
- Short amplicon methods
  - Thermo Fisher Scientific –
    - Precision ID mtDNA Control Region Panel
    - Precision ID mtDNA Whole Genome Panel
  - Verogen
    - ForenSeq mtDNA Control Region Kit
    - ForenSeq mtDNA Whole Genome Kit
  - Promega
    - PowerSeq CRM Nested System


13

## Precision ID mtDNA Control Region Panel

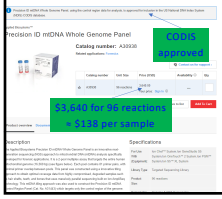
- Two PCR reactions, 7 primer pairs per reaction (14 total)
- Sensitivity to 2 pg DNA input (nuclear)

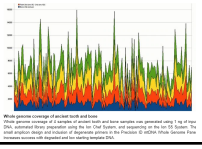



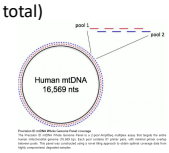

14

## Precision ID mtDNA Whole Genome Panel

- Two PCR reactions, 81 primer pairs per reaction (162 total)
- Sensitivity to 2 pg DNA input (nuclear)



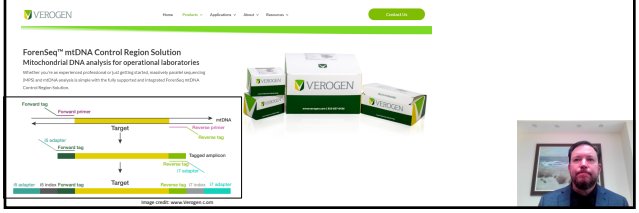






15

## Verogen ForenSeq mtDNA Control Region Kit

- Two PCR reactions, 18 amplicons
- List price $8,200 for 48 samples ($170 / sample)

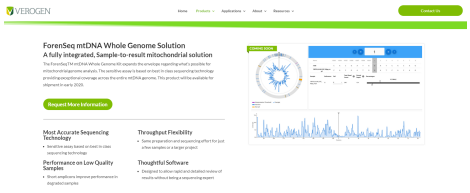





16

## Verogen ForenSeq mtDNA Whole Genome Kit

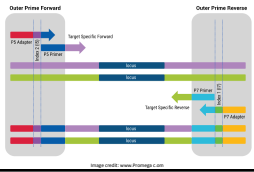
- Two PCR reactions, 245 amplicons
- List price $11,500 for 48 samples ($240 / sample)




17

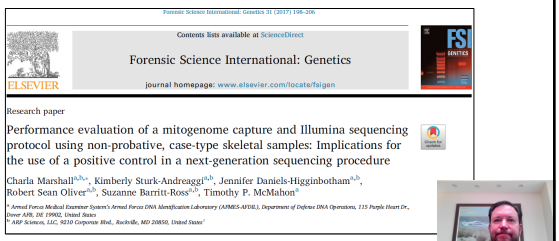
## Promega PowerSeq CRM Nested System

- One PCR reactions, 10 amplicons
- Approximate price $1,900 for 100 samples ($19 / sample)
- “Custom” product – not fully commercialized yet




18

## Hybridization Capture


Forensic Science International: Genetics 31 (2017) 198–206

Contents lists available at ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsigen

Research paper

Performance evaluation of a mitogenome capture and Illumina sequencing protocol using non-probative, case-type skeletal samples: Implications for the use of a positive control in a next-generation sequencing procedure

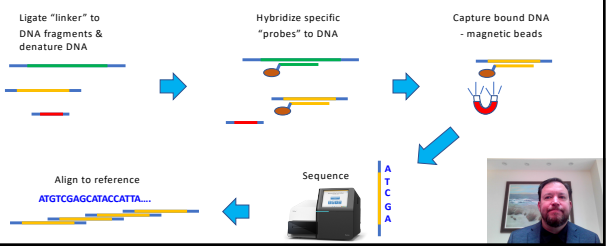
Charla Marshall[a,b,*], Kimberly Sturk-Andreaggi[a,b], Jennifer Daniels-Higginbotham[a,b], Robert Sean Oliver[a,b], Suzanne Barritt-Ross[a,b], Timothy P. McMahon[b]

[a] Armed Forces Medical Examiner System's Armed Forces DNA Identification Laboratory (AFMES-AFDIL), Department of Defense DNA Operations, 115 Purple Heart Dr., Dover AFB, DE 19902, United States
[b] ARP Sciences, LLC, 9210 Corporate Blvd., Rockville, MD 20850, United States



19

## Hybridization Capture

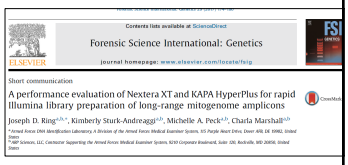





20

## In use at NIST for High-Quality DNA

- NIST/AFDIL population sequencing project
  - National Institute of Standards and Technology (NIST)
    - 659 samples attempted
    - African American, U.S. Caucasian, U.S. Hispanic
  - Armed Forces Medical Examiner System, Armed Forces DNA Identification Laboratory (AFMES-AFDIL)
    - 704 samples attempted
    - African American, U.S. Caucasian, U.S. Hispanic, Native American, Asian


Contents lists available at ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsig

Short communication

A performance evaluation of Nextera XT and KAPA HyperPlus for rapid Illumina library preparation of long-range mitogenome amplicons

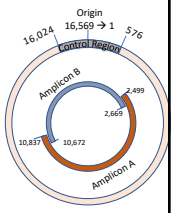
Joseph D. Ring[a,b,*], Kimberly Sturk-Andreaggi[a,c], Michelle A. Peck[a,c], Charla Marshall[a,b]

[a] Armed Forces DNA Identification Laboratory, A Division of the Armed Forces Medical Examiner System, 115 Purple Heart Drive, Dover AFB, DE 19902, United States
[b] ARP Sciences, LLC, Contractor Supporting the Armed Forces Medical Examiner System, 9210 Corporate Boulevard, Suite 330, Rockville, MD 20850, United States

21

## Method used at NIST

- PCR: • Two long amplicons ~ 8.5 kb each • QC – check on AATI Fragment Analyzer
- Library Preparation: • Fragmentation (enzymatic) of PCR products • Ligate adaptors/barcodes
- Sequencing: • QC & Quantitate libraries on AATI F/A • Run on MiSeq FGx
- Data Analysis: • Align to reference genome (rCRS) • Review/QC results

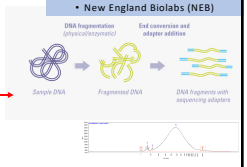



22

## Method used at NIST

- PCR: • Two long amplicons ~ 8.5 kb each • QC – check on AATI Fragment Analyzer
- Library Preparation: • Fragmentation (enzymatic) of PCR products • Ligate adaptors/barcodes
- Sequencing: • QC & Quantitate libraries on AATI F/A • Run on MiSeq FGx
- Data Analysis: • Align to reference genome (rCRS) • Review/QC results

Many kits available for library building.
Image credit: www.illumina.com
- Illumina
- Kapa Biosystems
- New England Biolabs (NEB)






23

## Sequencing Workflow

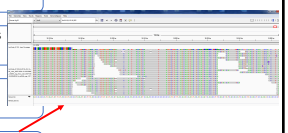
- PCR: • Two long amplicons ~ 8.5 kb each • QC – check on AATI Fragment Analyzer
- Library Preparation: • Fragmentation (enzymatic) of PCR products • Ligate adaptors/barcodes
- Sequencing: • QC & Quantitate libraries on Fragment Analyzer (Agilent) • Run on MiSeq FGx
- Data Analysis: • Align to reference genome (rCRS) • Review/QC results

Illumina MiSeq FGx

25

## Sequencing Workflow

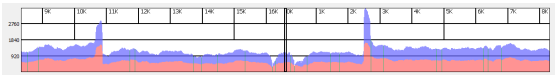
- PCR
  - Two long amplicons ~ 8.5 kb each
  - QC – check on AATI Fragment Analyzer
- Library Preparation
  - Fragmentation (enzymatic) of PCR products
  - Ligate adaptors/barcodes
- Sequencing
  - QC & Quantitate libraries on AATI F/A
  - Run on MiSeq FGx
- Data Analysis
  - Align to reference genome (rCRS)
  - Review/QC results



26

## Uniform coverage across the mtGenome

More coverage where amplicons overlap

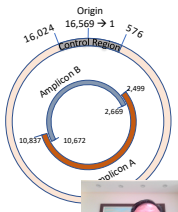


KAPA Hyper Plus Library Kit

27

## Quality Control

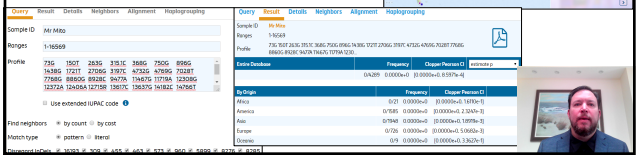
- PCR
  - Two long amplicons ~ 8.5 kb each
  - QC +/- on AATI Fragment Analyzer
- Library Preparation
  - Fragmentation (enzymatic) of PCR products
  - Ligate adaptors/barcodes
- Sequencing
  - QC & Quantitate libraries on AATI F/A
  - Run on MiSeq FGx
- Data Analysis
  - Align to reference genome (rCRS)
  - **Dual Review QC** of data




28

## EMPOP - the International mtDNA Database

- For estimating match statistics
  - Current database (V4, Release 12)
    - n = 46,963 mtDNA sequences (partial genome)
    - n = 4,289 whole genome sequences

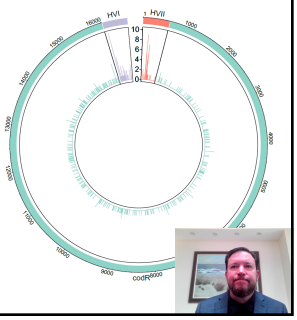




29

## Heteroplasmy Observed

| Dataset | Total Individuals | Total PHPs | Individuals with PHPs | Individuals with 1 PHP | Individuals with 2 PHPs | Individuals with 3 PHPs |
|---|---|---|---|---|---|---|
| COAF | 112 | 37 | 31 (28 %) | 26 | 4 | 1 |
| COCN | 112 | 41 | 30 (27 %) | 20 | 9 | 1 |
| COHS | 109 | 36 | 27 (25 %) | 20 | 5 | 2 |
| NTAF | 256 | 77 | 60 (23 %) | 43 | 17 | 0 |
| NTCN | 260 | 92 | 77 (30 %) | 65 | 10 | 2 |
| NTHS | 138 | 53 | 43 (31 %) | 34 | 8 | 1 |
| DSAS | 169 | 62 | 54 (32 %) | 49 | 2 | 3 |
| DSNA | 171 | 48 | 43 (25 %) | 38 | 5 | 0 |
| All | 1327 | 446 | 365 (28 %) | 295 | 60 | 10 |






30

## Analysis Software

| Software Name | Company Name | Forensic mtDNA Nomenclature Support | NGS or Sanger Support |
|---|---|---|---|
| CLCbio Genomics Workbench | Qiagen | Requires AQME plugin | Sanger and NGS |
| Converge | Thermo Fisher Scientific | Yes | NGS |
| DNASTAR LaserGene | DNASTAR, Inc. | No | Sanger and NGS |
| ForenSeq Universal Analysis Software | Verogen, Inc. | Yes | NGS |
| Geneious | Biomatters, Ltd. | No | Sanger and NGS |
| GeneMarker HTS | SoftGenetics, LLC. | Yes | NGS |
| Sequencher | Gene Codes Corporation | Yes | Sanger |

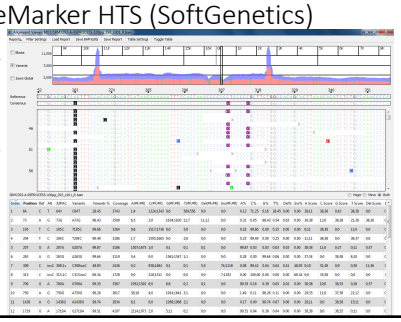
31

## GeneMarker HTS (SoftGenetics)

Coverage Map

Sequence Pileup

Results Table

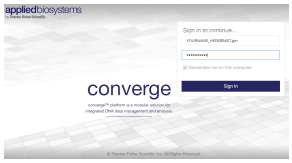


32

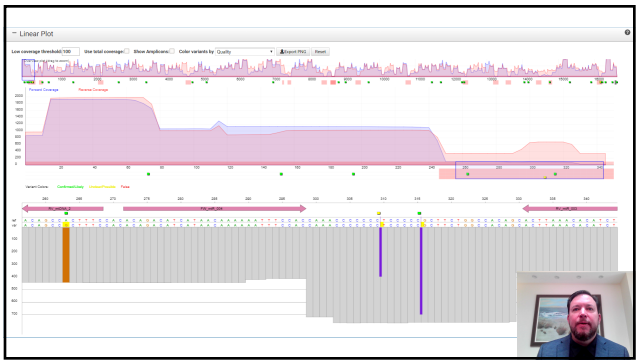
## Converge – for use with Ion Torrent Systems

Integrates with Torrent Server Suite (TSS) instrument control

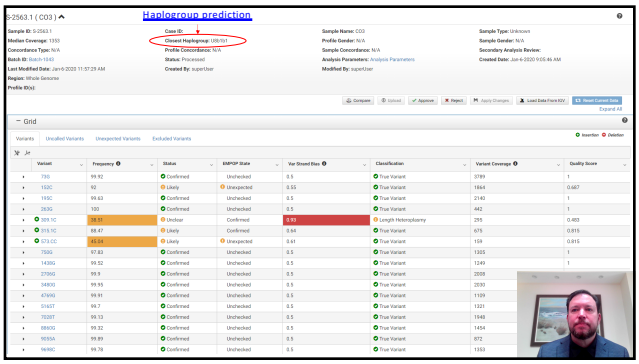
Ion Torrent S5




33

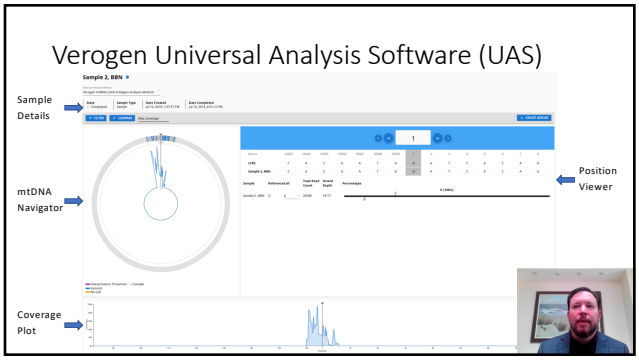



34




35

## Verogen Universal Analysis Software (UAS)




36

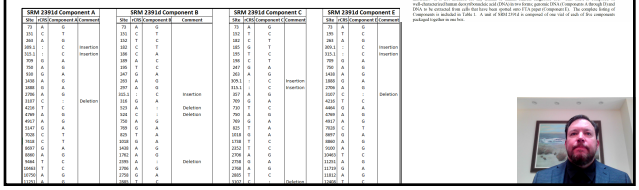
## NIST Mitochondrial Sequencing SRMs

- SRM 2392
  - Three components
    - Component A: DNA from cell line CHR
    - Component B: DNA from cell line 9947A
    - Component C: Cloned fragment containing C-stretch
- SRM 2392-I
  - One component
    - DNA from cell line HL60
- Characterized with Sanger methods
  - Released in 2001
  - Certified for mtDNA sequence


37

## SRM 2391d - PCR Based DNA Profiling Standard

- For STR testing
- Now contains mtDNA sequence
  - Information value



National Institute of Standards & Technology

Certificate of Analysis

Standard Reference Material® 2391d

PCR-Based DNA Profiling Standard



38

## Thank you!

## Questions?

Contact:
Kevin Kiesler
Kevin.Kiesler@nist.gov

Live long and prosper.

-Vulcan proverb

39