

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

I'm calling from my new Android smartphone

Yes, it even includes unlimited Data

2

Mitochondrial DNA

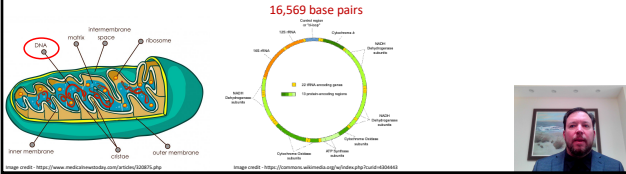
- Circular – 16.5 Kb
- Many copies – 100(0)s of mtGenome molecules per cell
- Maternally inherited – lineage marker, non-recombining

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...

NATURE COMMUNICATIONS

ARTICLE
 Received 5 Aug 2014 | Accepted 2 Oct 2014 | Published 2 Dec 2014
 DOI: 10.1038/ncomms6311 OPEN

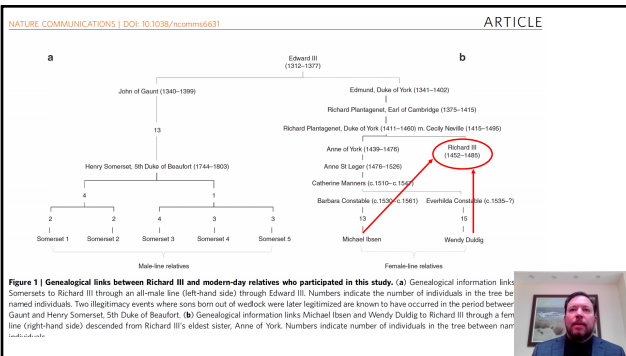
Identification of the remains of King Richard III

Tari E. King^{1,2}, Gloria Gonzalez Fortea^{3,4,5}, Patricia Balareca^{6,7}, Mark G. Thomas⁸, David Balding⁹, Parag Mehta¹⁰, Maura D'Elia¹¹, Rita Neumann¹², Heather Parson¹³, Michael Knapp¹⁴, Susan Wasie¹⁵, Laura Tomlin¹⁶, John Hall¹⁷, Marked Kowal¹⁸, Jo Appleby¹⁹, Peter Forster^{20,21}, David Davern²², Michael Holmstedt²³ & Keith Schumacher²⁴

Photo: <https://www.mirror.co.uk/news/uk-news/king-richard-iii-descendant-makes-5379382>



5



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

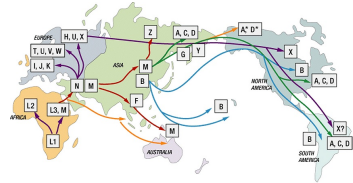
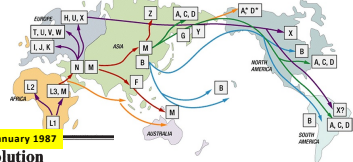


Image credit: www.familytreeDNA.com

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

Origin 16,569 → 1

Control Region 16,024 576

Mitochondrial Genome 16,569 bp

10

METHODOLOGY ARTICLE Open Access

A high-throughput Sanger strategy for human mitochondrial genome sequencing

Elizabeth A Lyons^{1,2*}, Melissa K Scheble^{1,2}, Kimberly Stork-Andreaga^{1,2}, Joel A Irwin^{1,2,4} and Rebecca S Just^{1,2,4*}

8 PCR Amplifications
127 Sequencing Primers

Lyons et al. BMC Genomics 2013, 14:881
<http://www.biomedcentral.com/1471-2164/14/881>

11

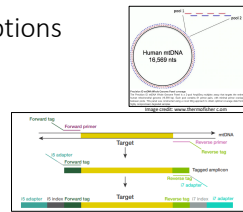
How do we get mtDNA sequence?

- Sanger sequencing
 - ~ 800 bases per reaction
- Next generation sequencing
 - ~ 5 x 10⁹ 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)

The screenshot shows the software interface for the Precision ID mtDNA Control Region Panel. It includes a 'Catalog number: A31443' and a 'Description' section. A green box highlights '14,000 bp per reaction' and '14,112 primer pairs'. A chromatogram displays signal intensity across the mtDNA sequence. A circular diagram of Human mtDNA (16,569 nts) is shown with primer pairs 'Fwd 1' and 'Rev 1'.



14

Precision ID mtDNA Whole Genome Panel

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

The screenshot shows the software interface for the Precision ID mtDNA Whole Genome Panel. It includes a 'Catalog number: A31442' and a 'Description' section. A green box highlights '14,000 bp per reaction' and '14,112 primer pairs'. A chromatogram displays signal intensity across the mtDNA sequence. A circular diagram of Human mtDNA (16,569 nts) is shown with primer pairs 'Fwd 1' and 'Rev 1'.



15

Verogen ForenSeq mtDNA Control Region Kit

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

VEROGEN

ForenSeq™ mtDNA Control Region Solution
Mitochondrial DNA analysis for operational laboratories

Forward tag
Target
Reverse tag
Target
Reverse tag
Target
Reverse tag
Target
Reverse tag

16

Verogen ForenSeq mtDNA Whole Genome Kit

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

VEROGEN

ForenSeq mtDNA Whole Genome Solution
A fully integrated, Sample-to-Result mitochondrial solution

Keyed Max Information

Most Accurate Sequencing Technology

Thoughtful Software

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

Outer Primer Forward
Inner Primer Forward
Outer Primer Reverse
Inner Primer Reverse

18

Hybridization Capture



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

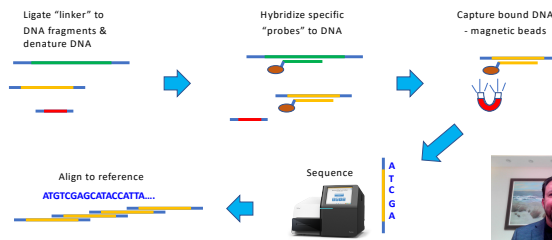
Charla Marshall^{a,b,*}, Kimberly Sturk-Andreaggi^{a,b}, Jennifer Daniels-Higginbotham^{a,b}, Robert Sean Oliver^{a,b}, Suzanne Barritt-Ross^{a,b}, Timothy P. McMahon^b

^a Armed Forces Medical Examiner System's Armed Forces DNA Identification Laboratory (AFMES-AFDIL), Department of Defense DNA Operations, 111 Pueblo Vista Dr., Dover AFB, DE 19902, United States
^b ADF Science, LLC, 9219 Corporate Blvd., Rockville, MD 20850, United States



19

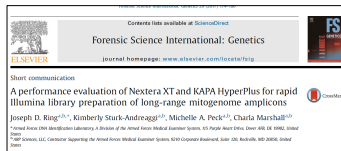
Hybridization Capture



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



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

Joseph D. King^{a,b,*}, Kimberly Sturk-Andreaggi^{a,b}, Michelle A. Peck^{a,b}, Charla Marshall^{a,b}

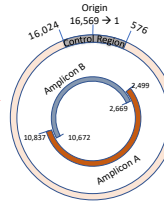

^a Armed Force DNA Identification Laboratory, A Division of the Armed Force Medical Examiner System, AFMES-AFDIL, Department of Defense DNA Operations, 111 Pueblo Vista Dr., Dover AFB, DE 19902, United States
^b ADF Science, LLC, Corporate Square, 9219 Corporate Blvd., Rockville, MD 20850, United States



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

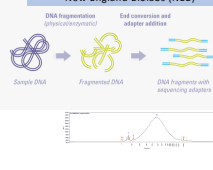

22

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:

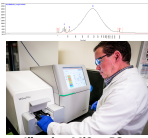
- Illumina
- Kapa Biosystems
- New England Biolabs (NEB)


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




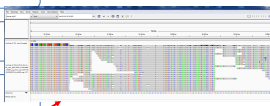
Illumina MiSeq FGx



25

Sequencing Workflow

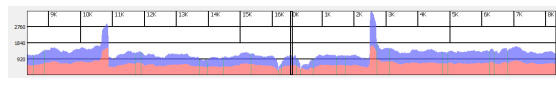
- PCR**
 - Two long amplicons ~ 8.5 kb each
 - QC – check on AATI Fragment Analyzer
- Library Preparation**
 - Fragmentation (enzymatic) of PCR products
 - Ligate adaptors/barcodes
- Sequencing**
 - QC & Quantitate libraries on AATI F/A
 - Run on MiSeq FGx
- Data Analysis**
 - Align to reference genome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




26

Uniform coverage across the mtGenome

More coverage where amplicons overlap




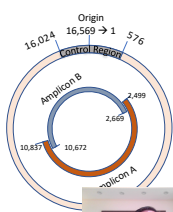
[KAPA Hyper Plus](#) Library Kit



27

Quality Control

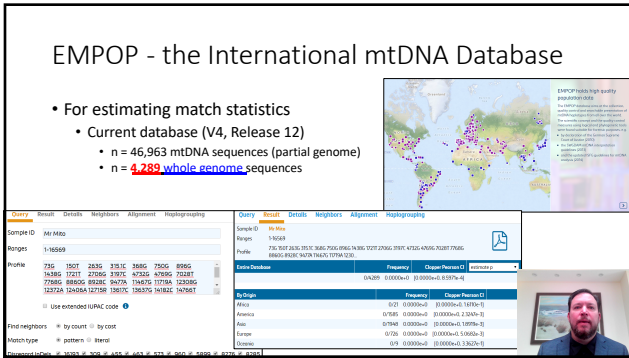
- PCR**
 - Two long amplicons ~ 8.5 kb each
 - QC +/- on AATI Fragment Analyzer
- Library Preparation**
 - Fragmentation (enzymatic) of PCR products
 - Ligate adaptors/barcodes
- Sequencing**
 - QC & Quantitate libraries on AATI F/A
 - Run on MiSeq FGx
- Data Analysis**
 - Align to reference genome (rCRS)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

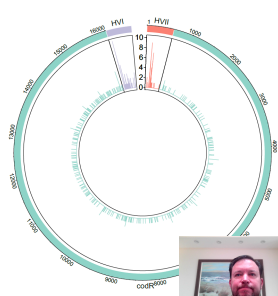


The screenshot shows the EMPOP website interface. At the top, there's a world map with colored dots representing mtDNA sequences. Below the map, there's a sample details section for 'Sample ID: Mt Mito' and 'Pop: 1-16569'. It includes a 'Profile' section with a sequence and a 'GenBank' section with a table of population frequencies for various regions like Africa, America, Asia, Europe, and Global.

29

Heteroplasmy Observed

Dataset	Total Individuals	Total Pkgs	Individuals with Pkgs	Individuals with 1 Pkg	Individuals with 2 Pkgs	Individuals with 3 Pkgs
COAF	112	37	31 (28%)	26	4	1
COFN	112	41	30 (27%)	20	9	1
COFS	109	36	27 (25%)	20	5	2
NTAF	256	77	60 (23%)	43	17	0
NTFN	260	92	77 (30%)	65	10	2
NTFS	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



The circular diagram shows the mtDNA genome with segments representing different regions. A legend indicates the number of individuals with heteroplasmy in each region: Africa (1), America (2), Asia (2), Europe (2), and Global (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 →

The screenshot displays the GeneMarker HTS interface. At the top, there's a 'Coverage Map' showing a blue line graph of signal intensity across a genomic region. Below that is a 'Sequence Pileup' showing individual sequencing reads aligned to a reference sequence. At the bottom, a 'Results Table' is visible, containing columns for various metrics like read counts and quality scores. A small inset photo of a man is in the bottom right corner.

32

Converge – for use with Ion Torrent Systems

Integrates with Torrent Server Suite (TSS) instrument control

The screenshot shows the 'converge' software interface. On the left is an image of an Ion Torrent SS sequencer. The main area displays a login screen with the text 'appliedbiosystems' and 'converge'. It includes a 'Sign in to this institution' section with a text input field and a 'Sign In' button. A small inset photo of a man is in the bottom right corner.

33

Linear Plot

Low coverage threshold (10x) | Show depth (coverage) | Show depth (coverage) | Other variants by (Quality) | (Align (PIL) | Print)

The screenshot shows a 'Linear Plot' visualization. The top part is a bar chart representing 'Low coverage threshold (10x)'. Below that is a large area plot showing 'Depth (coverage)' on the y-axis and genomic coordinates on the x-axis. The plot features a pink shaded area representing coverage and a blue line representing a signal. At the bottom, there are tracks for 'Variant Calls' and 'Other variants by (Quality)'. A small inset photo of a man is in the bottom right corner.

34

haplogroup prediction

Sample ID: S-2563-1 (C03)
 Sample Name: C03
 Profile Gender: N/A
 Sample Origin: N/A
 Sample Collection Date: N/A
 Analysis Parameters: Analyze Parameters
 Modified By: supervisor

Profile: 000

Var	Value	Frequency	Status	IMPTP Status	Var Shared With	Classification	Variant Coverage	Quality Score
780	89.82	100	Confirmed	Unshared	0.5	True Variant	3799	1
1302	92	100	Unshared	Unshared	0.55	True Variant	1844	0.887
1302	88.43	100	Confirmed	Unshared	0.5	True Variant	2140	1
2032	100	100	Confirmed	Unshared	0.5	True Variant	442	1
2032	99.10	100	Unshared	Confirmed	0.5	True Variant	259	0.403
2032	98.47	100	Unshared	Confirmed	0.55	True Variant	676	0.915
2032	85.58	100	Unshared	Unshared	0.41	True Variant	154	0.815
2500	87.83	100	Confirmed	Unshared	0.5	True Variant	1395	1
14000	88.22	100	Confirmed	Unshared	0.5	True Variant	1364	1
27002	95.8	100	Confirmed	Unshared	0.5	True Variant	2038	1
14800	99.93	100	Confirmed	Unshared	0.5	True Variant	2030	1
17000	99.99	100	Confirmed	Unshared	0.5	True Variant	1100	1
21007	99.7	100	Confirmed	Unshared	0.5	True Variant	1311	1
22007	99.73	100	Confirmed	Unshared	0.5	True Variant	1440	1
80000	99.22	100	Confirmed	Unshared	0.5	True Variant	1484	1
90000	99.89	100	Confirmed	Unshared	0.5	True Variant	872	1
90000	99.76	100	Confirmed	Unshared	0.5	True Variant	1303	1

35

Verogen Universal Analysis Software (UAS)

Sample 2, 888

Sample Details

mtDNA Navigator

Coverage Plot

Position Viewer

Position	Reference	Sample	Quality	Depth
16167	C	C	30	1
16168	T	T	30	1
16169	C	C	30	1
16170	T	T	30	1
16171	C	C	30	1
16172	T	T	30	1
16173	C	C	30	1
16174	T	T	30	1
16175	C	C	30	1
16176	T	T	30	1
16177	C	C	30	1
16178	T	T	30	1
16179	C	C	30	1
16180	T	T	30	1
16181	C	C	30	1
16182	T	T	30	1
16183	C	C	30	1
16184	T	T	30	1
16185	C	C	30	1
16186	T	T	30	1
16187	C	C	30	1
16188	T	T	30	1
16189	C	C	30	1
16190	T	T	30	1
16191	C	C	30	1
16192	T	T	30	1
16193	C	C	30	1
16194	T	T	30	1
16195	C	C	30	1
16196	T	T	30	1
16197	C	C	30	1
16198	T	T	30	1
16199	C	C	30	1
16200	T	T	30	1

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



The Standard Reference Material (SRM) 2391d is intended primarily for use in the identification of forensic and anthropological samples. It contains a mixture of 15 autosomal short tandem repeat (STR) loci, the mitochondrial DNA (mtDNA) control region, and a Y-chromosomal STR locus. The DNA is extracted from a human donor and is available in a liquid form. The SRM 2391d is composed of 1000 µg of DNA in 100 µL of water. The DNA is stable for at least 12 months at room temperature. The SRM 2391d is available in 100 µL aliquots in a 1 mL vial. The SRM 2391d is available in 100 µL aliquots in a 1 mL vial. The SRM 2391d is available in 100 µL aliquots in a 1 mL vial.

SRM 2391d Component A	SRM 2391d Component B	SRM 2391d Component C	SRM 2391d Component D
101	101	101	101
102	102	102	102
103	103	103	103
104	104	104	104
105	105	105	105
106	106	106	106
107	107	107	107
108	108	108	108
109	109	109	109
110	110	110	110
111	111	111	111
112	112	112	112
113	113	113	113
114	114	114	114
115	115	115	115
116	116	116	116
117	117	117	117
118	118	118	118
119	119	119	119
120	120	120	120
121	121	121	121
122	122	122	122
123	123	123	123
124	124	124	124
125	125	125	125
126	126	126	126
127	127	127	127
128	128	128	128
129	129	129	129
130	130	130	130
131	131	131	131
132	132	132	132
133	133	133	133
134	134	134	134
135	135	135	135
136	136	136	136
137	137	137	137
138	138	138	138
139	139	139	139
140	140	140	140
141	141	141	141
142	142	142	142
143	143	143	143
144	144	144	144
145	145	145	145
146	146	146	146
147	147	147	147
148	148	148	148
149	149	149	149
150	150	150	150



38

Thank you!

Questions?

Contact:
Kevin Kiesler
Kevin.Kiesler@nist.gov



Live long and prosper.
-Vulcan proverb



39
