NIST
National Institute of Standards and Technology
U.S. Department of Commerce

FORENSICS @ NIST

#NISTForensics

Sequencing and Standards for Characterization of the Mitochondrial Genome

Kevin Kiesler, M.S.

Research Biologist - NIST Applied Genetics Group

Forensics @NIST

November 7, 2018

Outline

- Sanger sequencing and the “next generation”
- Reference materials and mtDNA sequencing
- Population sequencing project

NIST Disclaimer

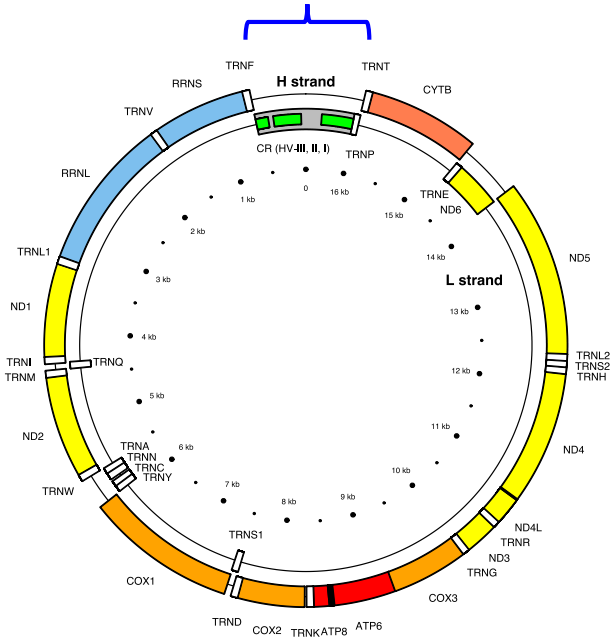
- **Disclaimer** - Points of view in this document are those of the authors and do not necessarily represent the official position or policies of the U.S. Department of Commerce or the Department of Justice. Certain commercial equipment, instruments, and materials are identified in order to specify experimental procedures as completely as possible. In no case does such identification imply a recommendation or endorsement by NIST, nor does it imply that any of the materials, instruments, or equipment identified are necessarily the best available for the purpose.

- All work presented has been reviewed and approved by the NIST Human Subjects Protections Office.

Mitochondrial DNA Sequencing - Introduction

- Mitochondrial DNA advantages
  - High copy number
    - Very small quantity of evidence required
  - Single nucleotide variants (SNPs or SNVs) = profile
    - Small region of DNA analyzed
    - Short PCR amplicons
  - Applicable to challenging/degraded/limited samples
    - Mass disaster, missing persons
      - Example: World Trade Center victim identification – 44,000 mtDNA profiles
- Disadvantage
  - Low power of discrimination using “Control Region”
    - 26,127 CR genotypes in EMPOP database for matching
  - Can be improved with whole mtGenome analysis

Control Region



Mitochondrial Genome Diagram

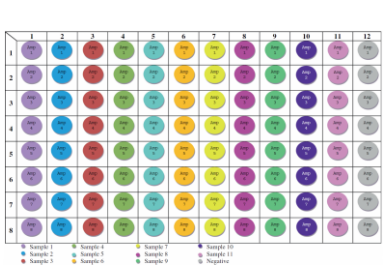
Image credit: Emmanuel Douzery

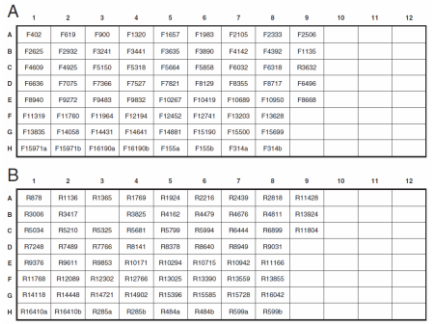
Sanger Sequencing Workflow - mtGenome

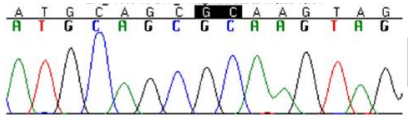
1 Extract DNA → 2 Quantitate DNA → 3 PCR Template for Sequencing → 4 Purify → 5 Sequencing Reactions → 6 Purify → 7 Size Separation (CE) → 8 Data Analysis → 9 Variant List

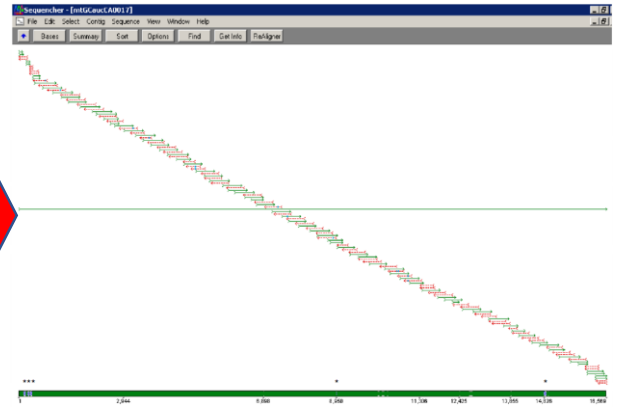
8 Reactions Per Sample

> 100 Reactions Per Sample









Lyons et al. BMC Genomics 2013, 14:881

Next Generation Sequencing Workflow

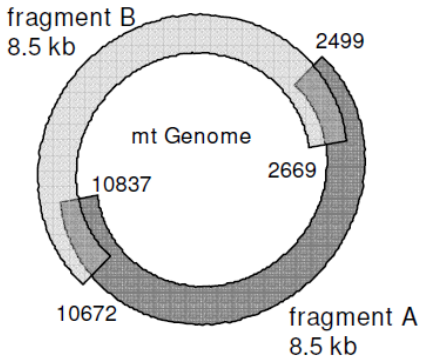
Manual Procedure

1	2	3	4	5	6	7	8
Extract DNA	Quantitate DNA	PCR Template for Sequencing	Library Preparation	QC & Quantify	Sequence	Data Analysis	Variant List

Two reactions per Sample

Several steps, one well per sample

Step 3: Amplify mtGenome with long PCR (2 reactions)



Fendt et al. BMC Genomics 2009, 10:139

Step 4: Library preparation

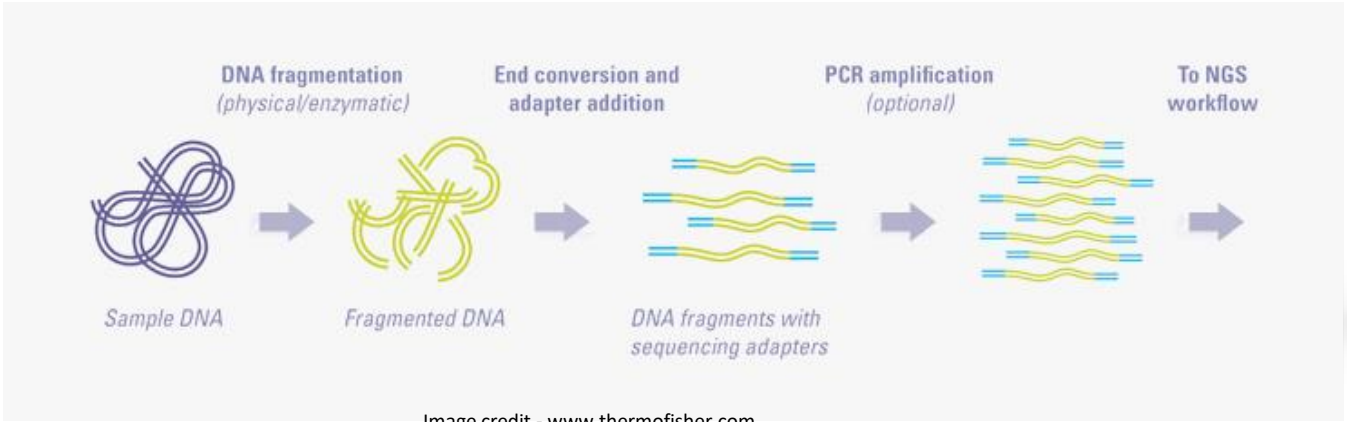


Image credit - www.thermofisher.com

Step 6: Sequence on MiSeq

Image credit:
www.Illumina.com

Next Generation Sequencing Workflow

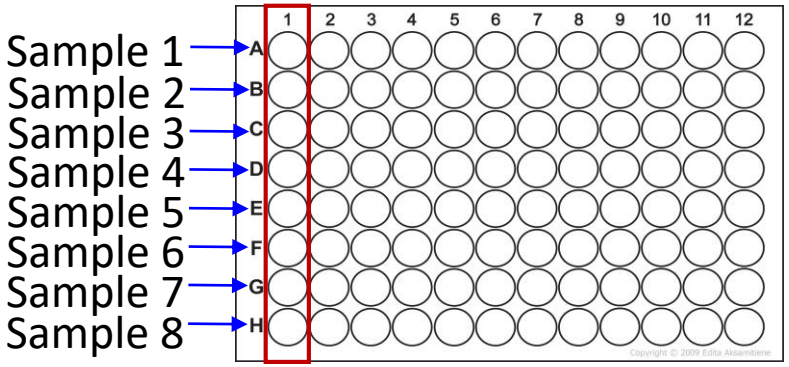
Automated Procedure

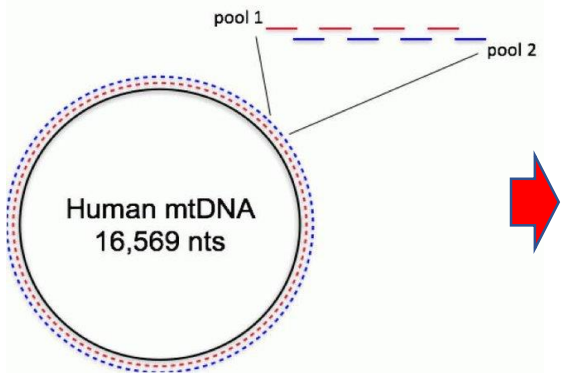
(Applied Biosystems Precision ID mtDNA Whole Genome Panel)

1 Extract DNA → 2 Quantitate DNA → 3 Automated Library Preparation → 4 Library Quantitation (manual) → 5 Automated Template Preparation → 6 Sequencing → 7 Data Analysis → 8 Variant List

Automated Reaction Setup

Step 3: Amplify mtGenome with 162 primer pairs in 2 reactions eight samples at a time on Ion Chef





Step 4: Prepare templated beads

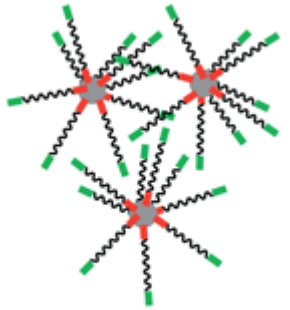
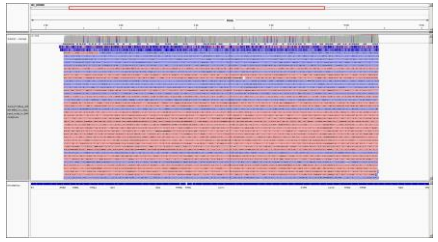


Image credit: www.wright.edu/~oleg.paliy/NGS.html

Step 5: Sequence on Ion Torrent S5 or PGM

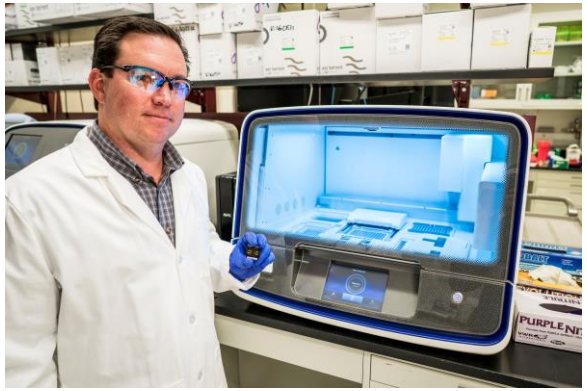
Image credit: www.thermofisher.com

Step 6: Align NGS reads to reference genome (rCRS)



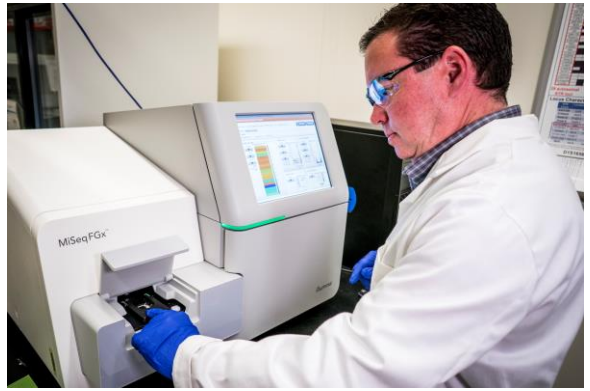
Sequencing Instruments at NIST

- Multiple platforms
- Orthogonal measurements
- Characterize Standard Reference Materials

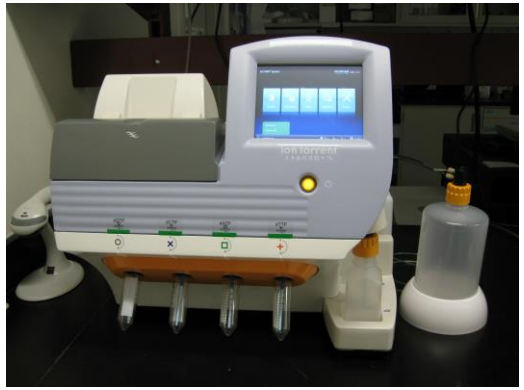


Ion Chef

Applied Biosystems SOLiD



Illumina MiSeq FGx



Ion Torrent PGM

Ion Torrent S5

NIST Mitochondrial Sequencing SRMs

National Institute of Standards & Technology

Certificate of Analysis

Standard Reference Material® 2392

Mitochondrial DNA Sequencing (Human)

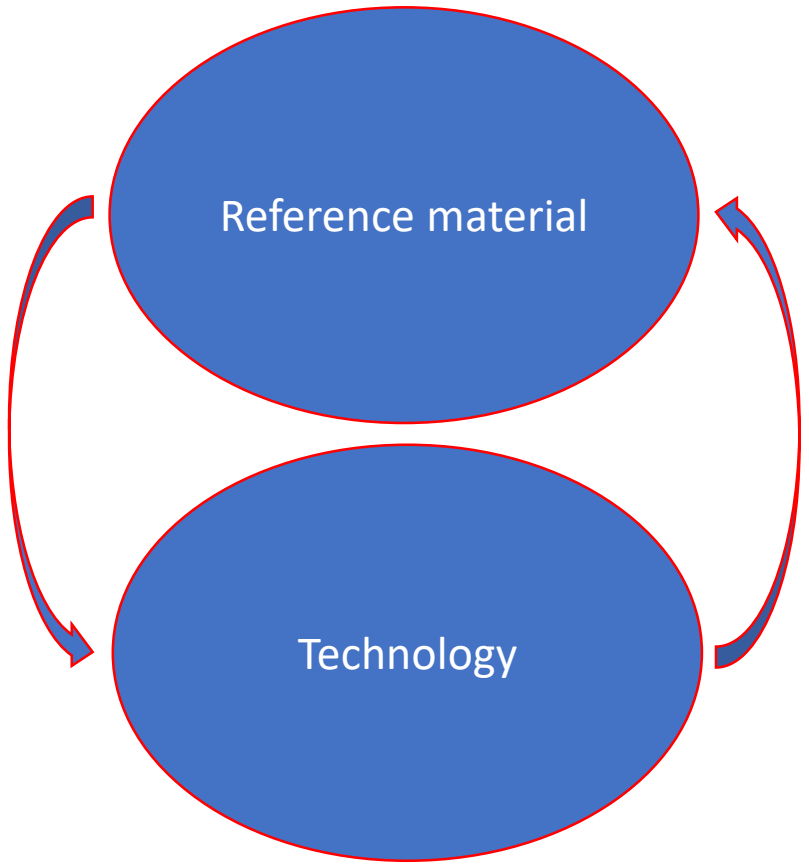
- SRM 2392
  - Three components
    - Component A: DNA from cell line CHR
    - Component B: DNA from cell line 9947A
    - Component C: Cloned fragment from HV1 region of CHR containing C-stretch
- SRM 2392-I
  - One component
    - DNA from cell line HL60
- Characterized with Sanger methods
  - Released in 2001

Sanger-Based Sequence Agrees With NGS Values

Table 2. Certified Human mtDNA Sequence Differences from the Revised Cambridge Reference Sequence for SRM 2392 Component GM09947A

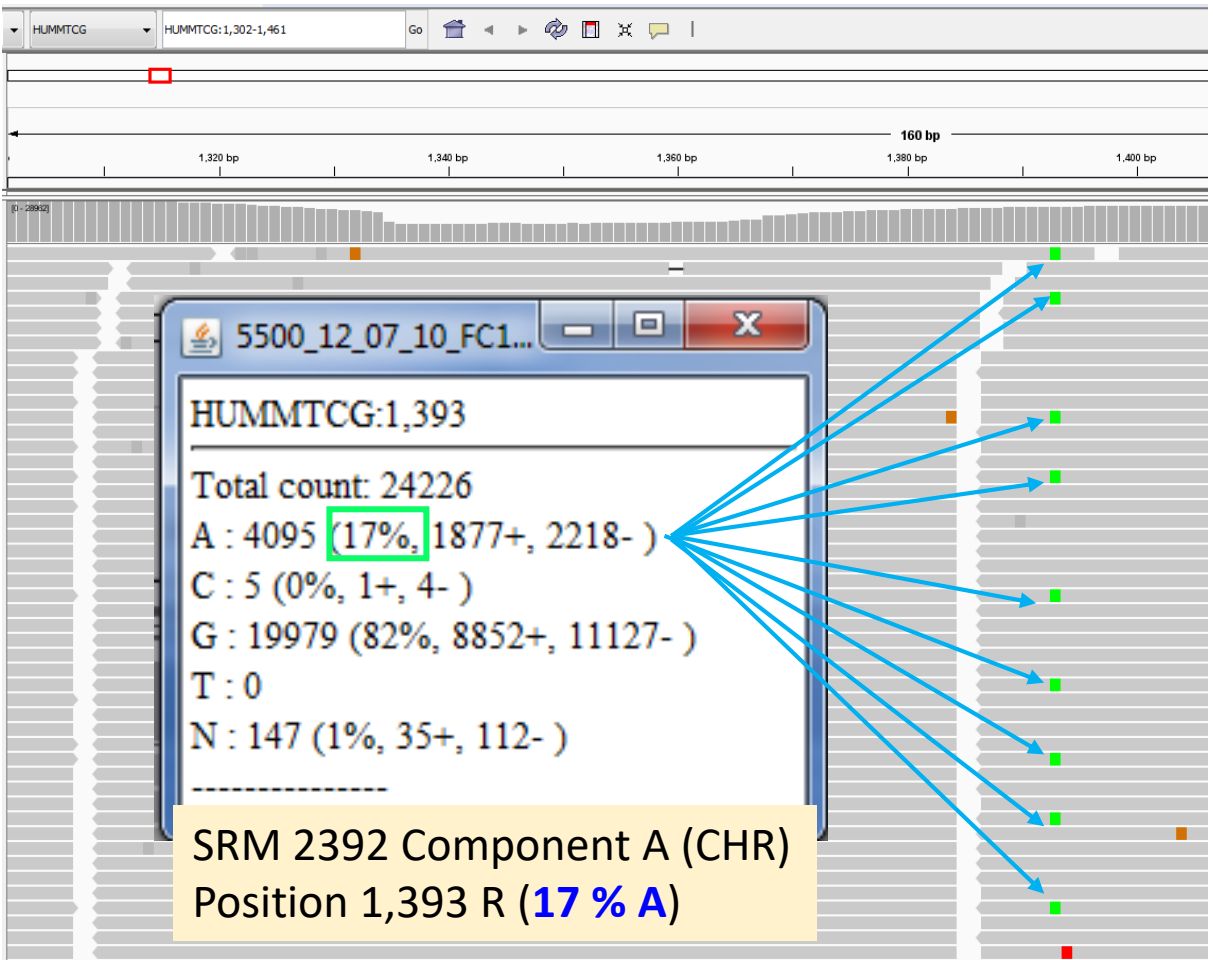
Site	rCRS	GM09947A	Comments
93	A	G √	
195	T	C √	
214	A	G √	
263	A	G √	
309.1		C X	insertion
309.2		C X	insertion
315.1		C X	insertion
750	A	G √	
1438	A	G √	
3107	C		deletion
4135	T	C √	
4769	A	G √	
7645	T	C √	
7861	T	C √	
8448	T	C √	
8860	A	G √	
9315	T	C √	
13572	T	C √	
13759	G	A √	
15326	A	G √	
16311	T	C √	
16519	T	C √	

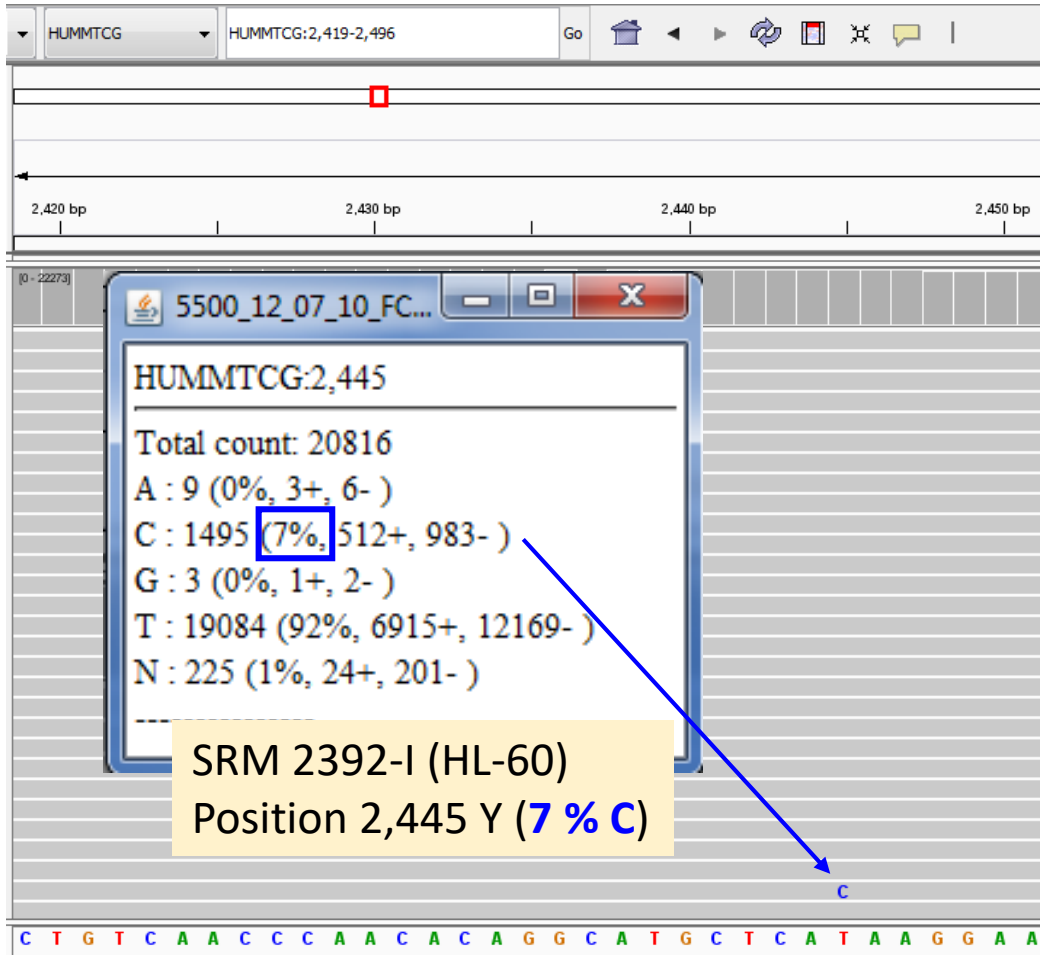
Previously no software available that correctly handled these C-stretch insertions with forensic nomenclature.



NGS Can Detect Low Level Heteroplasmy

- Level is below what we can reliably see with Sanger methods





Multiple Orthogonal Measurements

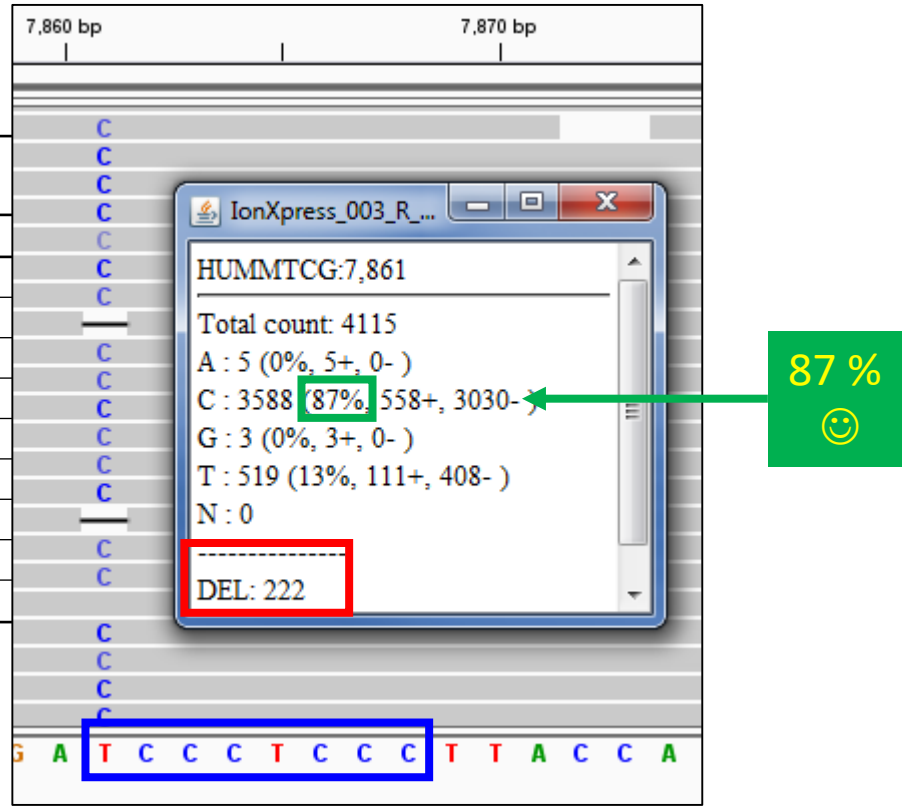
- Great approach for certifying reference materials!

	SRM 2392				
	Component A	Component B		SRM 2392-I	
Nucleotide	64 T	1393 A	7861 C	2445	5149
PGM Edge	26.8	21.2	72.7	10.6	5.1
PGM NIST 1	24.3	15.6	45.0	7.5	9.1
PGM NIST 2	25.0	17.5	65.7	7.7	7.7
PGM NIST 3	29.7	16.5	59.4	7.7	7.7
PGM NIST HiQ	33.2	15.2	77.7	7.7	8.4
MiSeq Edge	33.0	19.0	87.4	10.7	6.8
MiSeq NIST	31.6	17.9	88.4	9.1	6.4
HiSeq BC	30.6	16.9	88.3	7.4	7.1
SOLiD NIST	29.0	16.7	87.3	7.3	7.0
Average	29.2	17.4	74.6	8.4	7.3
St. Dev.	3.1	1.7	14.5	1.3	1.1

Multiple Orthogonal Measurements

- Great approach for characterizing reference materials!

	SRM 2392		
	Component A	Component B	
Nucleotide	64 T	1393 A	7861 C
PGM Edge	26.8	21.2	**72.7**
PGM NIST 1	24.3	15.6	**45.0**
PGM NIST 2	25.0	17.5	**65.7**
PGM NIST 3	29.7	16.5	**59.4**
PGM NIST HiQ	33.2	15.2	**77.7**
MiSeq Edge	33.0	19.0	87.4
MiSeq NIST	31.6	17.9	88.4
HiSeq BC	30.6	16.9	88.3
SOLiD NIST	29.0	16.7	87.3
Average	29.2	17.4	74.6
St. Dev.	3.1	1.7	**14.5**

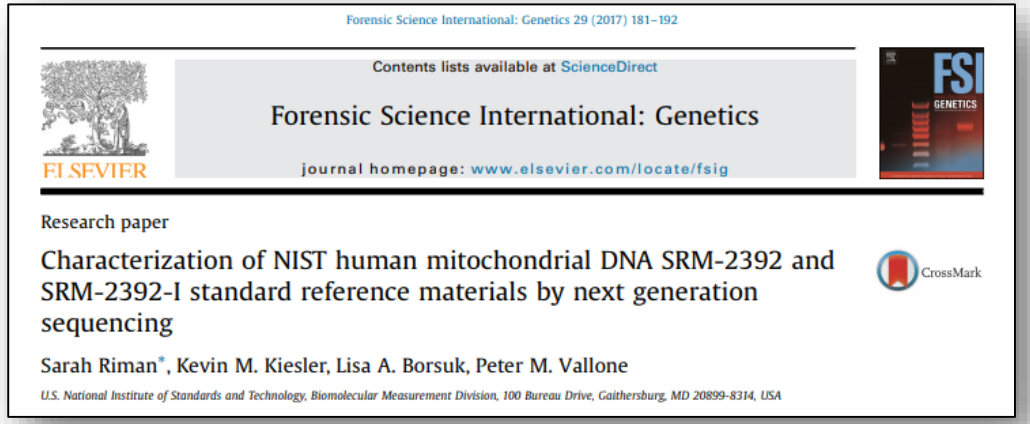


It can also educate you about your technology.

Multiple Orthogonal Measurements

- Great approach for characterizing ~~reference materials~~! SOFTWARE

	SRM-9947 A					
	Position 1393		Position 3242		Position 7861	
Platforms & Analysis	REF G%	VAR A%	REF G%	VAR A%	REF T%	VAR C%
PGM (CLC)	84.5	15.5	97	3	**31.5**	**68.5**
MiSeq (CLC)	82	18	96	4	12	88
PGM (Galaxy)	84.5	15.5	97	3	**21.5**	**78.5**
MiSeq (Galaxy)	82	18	95	5	11	89
PGM (GM-HTS)	85	15.0	97	3	**13**	**83**
MiSeq (GM-HTS)	83	17.0	97	3	12	88
PGM (STRait Razor)	84	16	97	3	12	88
MiSeq (STRait Razor)	83	17	96	4	11	89



Forensic Science International: Genetics 29 (2017) 181–192

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

Research paper

Characterization of NIST human mitochondrial DNA SRM-2392 and SRM-2392-I standard reference materials by next generation sequencing

Sarah Riman*, Kevin M. Kiesler, Lisa A. Borsuk, Peter M. Vallone

U.S. National Institute of Standards and Technology, Biomolecular Measurement Division, 100 Bureau Drive, Gaithersburg, MD 20899-8314, USA

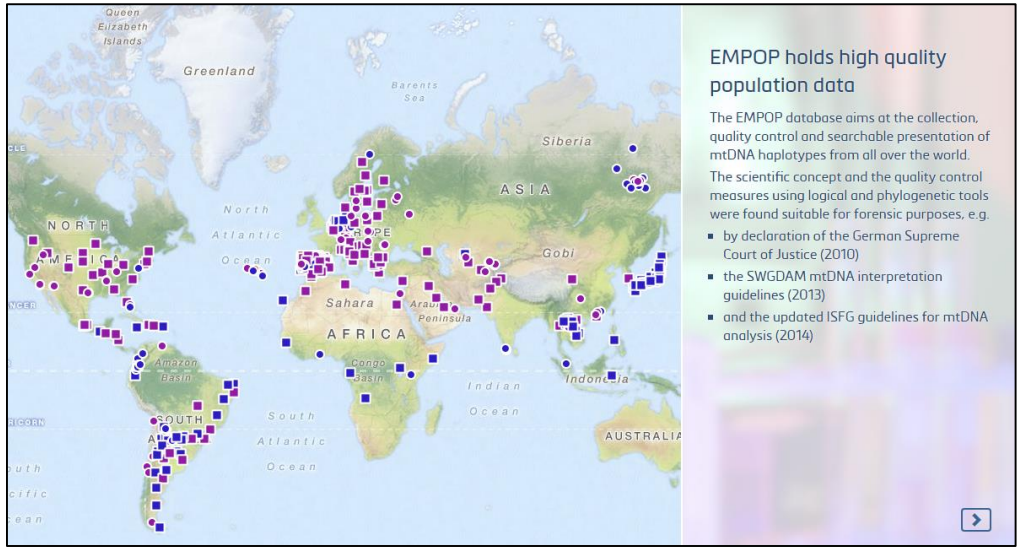
Conclusions

- Reference materials
  - Can identify technical limitations/bias
    - Often need multiple measurements
    - Orthogonal techniques
  - Help to select best procedures

Population Scale Sequencing

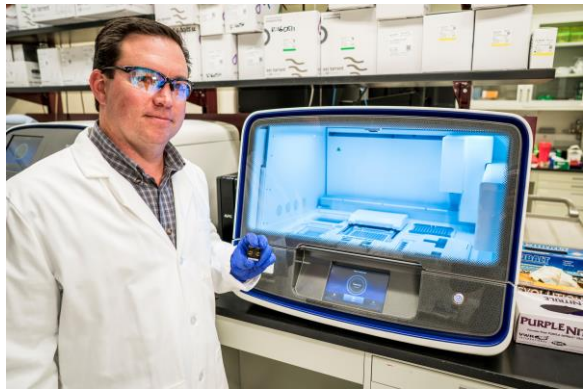
Project Goals

- Submit forensic-quality whole mtGenome data to EMPOP
  - Database used for match statistics
  - Current version (V4, Release 11)
    - n = 26,127 control region sequences
    - n = **256** whole genome sequences
- NIST population samples (n > 1,000)
  - African American, Asian, Caucasian, Hispanic
- Sequencing plan
  - Start with Caucasian population
  - ≈ 440 mtGenomes

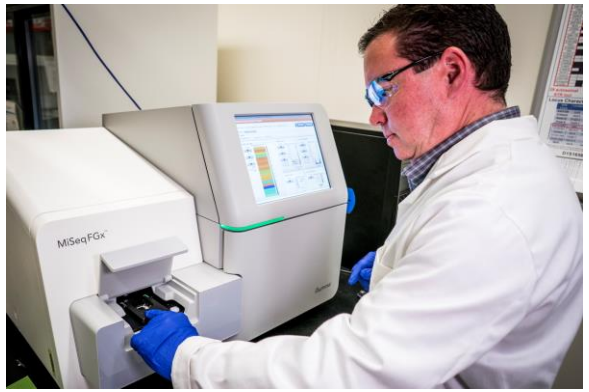


Project Plans

- What instrument do we use?
- What protocol/chemistry do we use?
- What analysis procedure do we use?
    - Software, data review, etc.



Ion Chef



Illumina MiSeq FGx

Ion Torrent S5

GeneMarker® HTS
Mitochondrial NGS Analysis Software

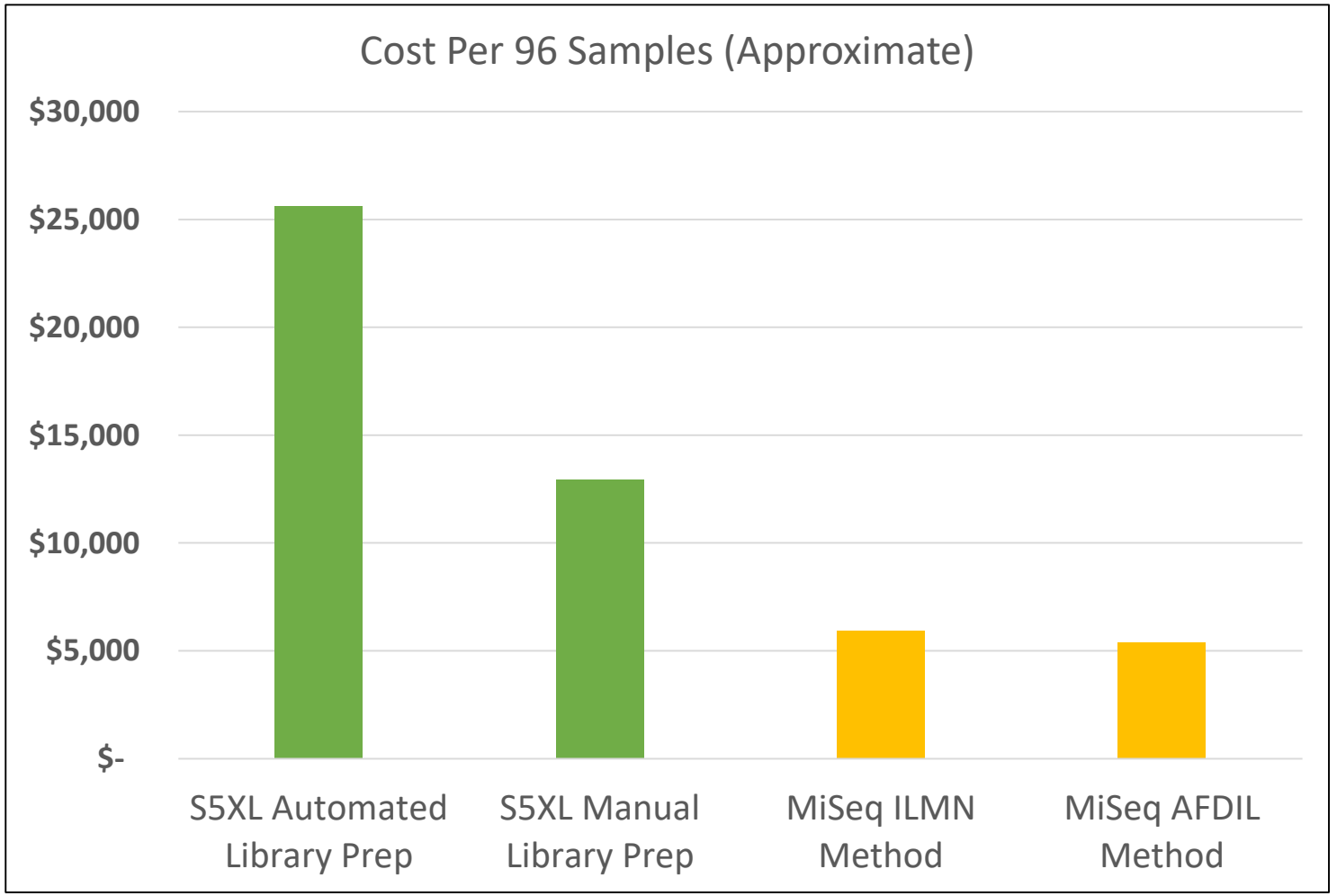
Galaxy

NextGENe®
2nd Generation Sequence Analysis Software

CLC bio
A QIAGEN® Company
Gx

Project Planning: Instrument Selection

- Considerations
  - Cost
  - Time/labor
    - Automation



Project Planning: Protocol Selection

- Options
  - Illumina mtGenome procedure
    - Long PCR primers developed by Dr. Mark Wilson's lab
    - TaKaRa LA Taq
    - Illumina Nextera XT library preparation
    - Illumina MiSeq v2 2x150 cartridge (per protocol)

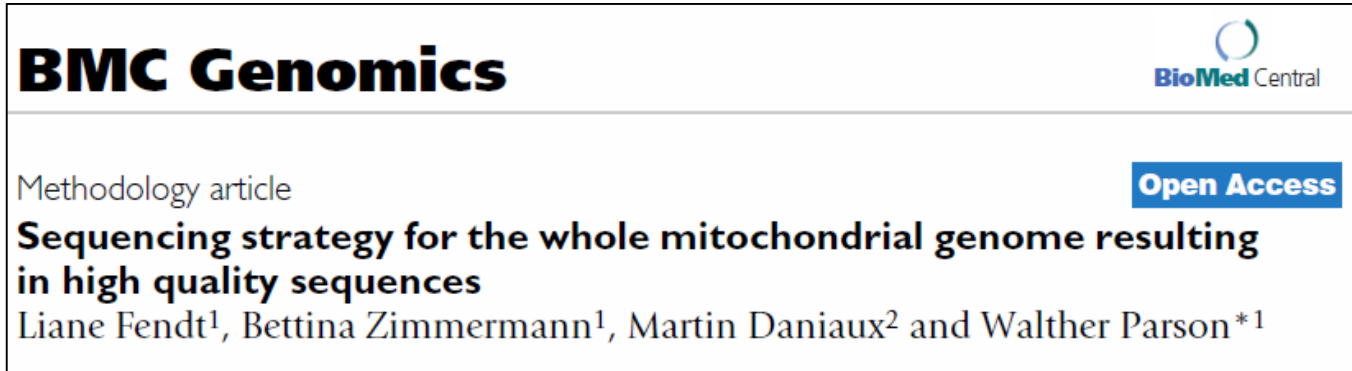
Obtain the following PCR Primers from a general oligo supplier:

Table 6 User-Supplied PCR Primers[1]

Primer	Sequence
MTL-F1	5′- AAA GCA CAT ACC AAG GCC AC -3′
MTL-F2	5′- TAT CCG CCA TCC CAT ACA TT -3′
MTL-R1	5′- TTG GCT CTC CTT GCA AAG TT -3′
MTL-R2	5′- AAT GTT GAG CCG TAG ATG CC -3′

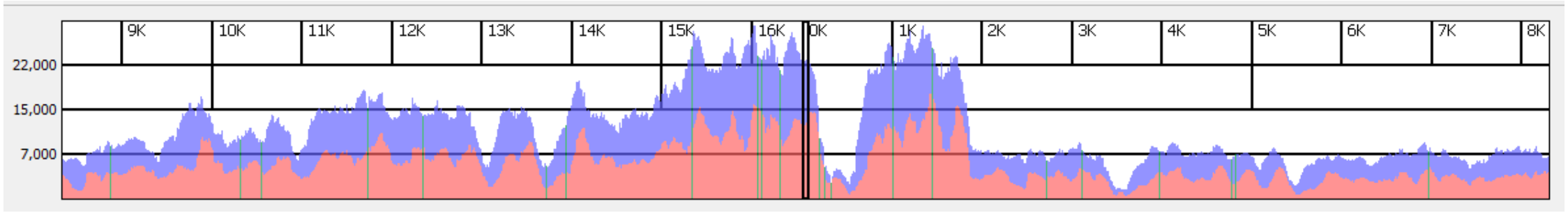
[1]Stawski, H., B. J. Bintz, E. S. Burnside, and M. Wilson. 2013. Preparing Whole Genome Human Mitochondrial DNA Libraries for Next Generation Sequencing (NGS) Using Illumina Nextera XT. Poster presentation at the 65th Annual American Academy of Forensic Sciences Conference. In: Proceedings of the American Academy of Forensic Sciences. Washington, D.C. www.aafs.org/sites/default/files/pdf/ProceedingsWashingtonDC2013.pdf

  - mtGenome procedure used by Armed Forces DNA Identification Lab (AFDIL)
    - Long PCR primers from Fendt *et al.,* BMC Genomics 2009, 10:139
    - TaKaRa LA Taq (GC Buffer & BSA)
    - Kapa HyperPlus Library Kit
    - Illumina V3 2x300 cartridge

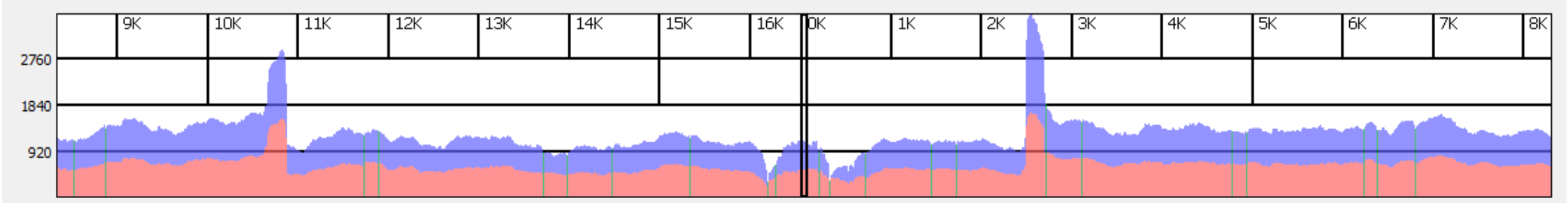


More Consistent Coverage Depth with Kapa HyperPlus

AFDIL method allows higher multiplexing with less likelihood of dropout sites



Illumina Whole mtGenome Method (Nextera Library Kit)



AFDIL Whole mtGenome Method (Kapa Hyper Plus Library Kit)

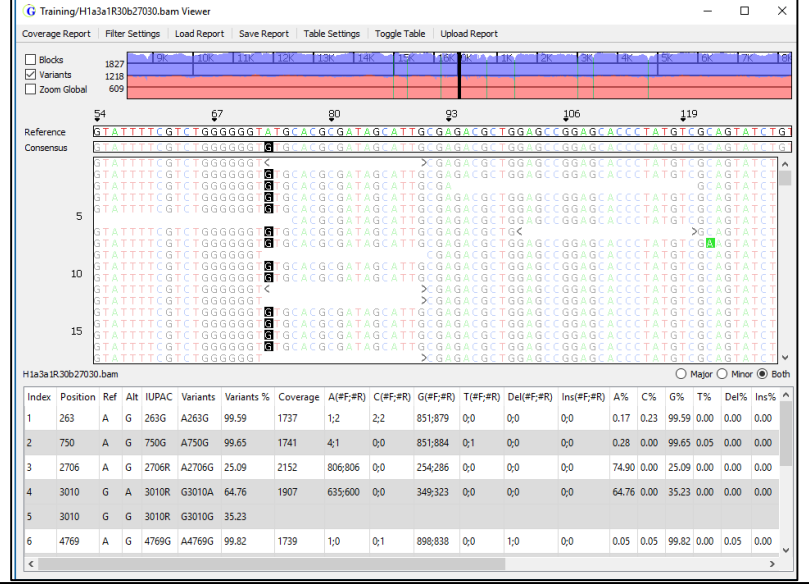
Haplogroup Estimation from EMPOP

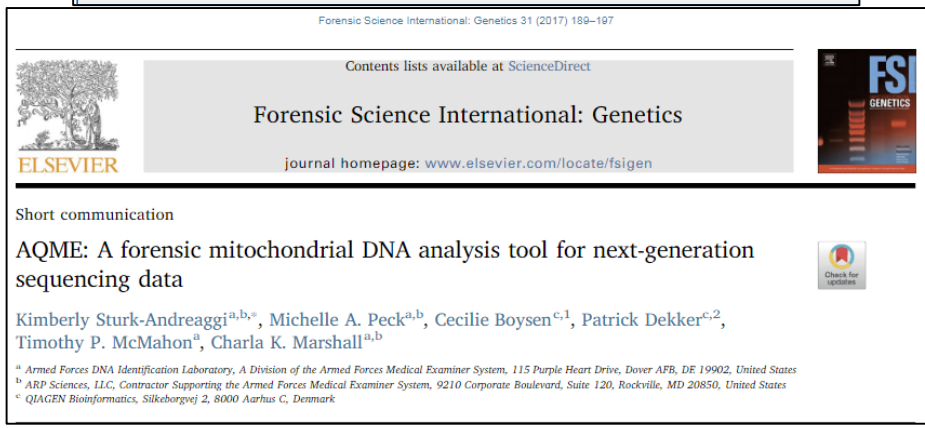
- No surprise haplogroups

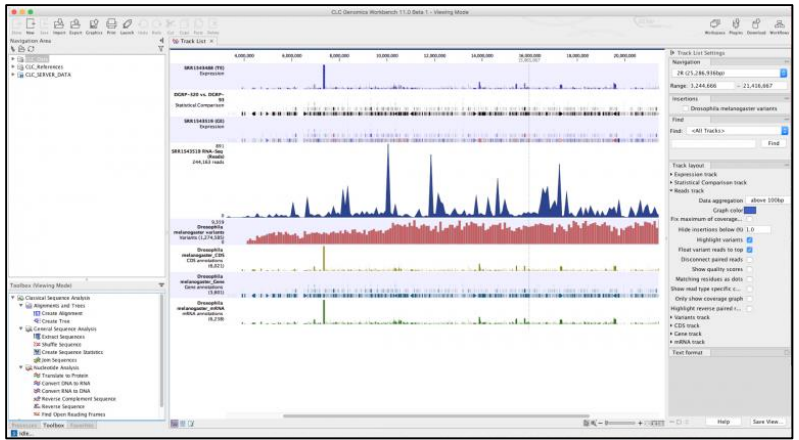
Sample	Missing Mutations	Private Mutations	Haplogroup	Continent
GT38086	none	T16189C	**H1c21**	Europe (H)
GT38087	none	none	**T2b6b**	Europe
GT38089	none	T16189C	**J1c8a**	Europe
GT38091	none	C10933T A15467G	**V**	Europe
GT38092	none	C198T A9327G A13801G T15670C	**H5e1a1**	Europe (H)
GT38093	none	none	**H1o**	Europe
GT38094	none	A11252G	**H65**	Europe (H)
GT38095	-573.1C	G709A C9727T	**I1a1b**	Europe (I1a1)
GT38097	none	G8027A G15301A	**K1a1b1**	Europe
GT38098	A16183C -309.1C -309.2C	C16111T T152C	**H1b1**	Europe
GT38100	none	-309.1C G4655A	**H1**	Europe
GT38106	none	T2416C A8817G	**H2a2a**	Europe
GT38107	none	C3388A C8788T	**U5a1a1**	Europe
GT38108	none	T4373C T15313C	**H1+16189**	Europe
JA44327	-573.1C	G16474T	**I2**	Europe (I)
JM28315	-309.1C	none	**K2a6**	Europe
JT52345	none	A16138G A73G	**H5a1c1a**	Europe
JT52346	none	none	**H5g**	Europe
MT97121	none	T16093C G7762A	**J1c2**	Europe
MT97122	none	none	**T2b2b**	Europe
MT97123	none	A16158G	**T1a1**	Europe (T1a)
MT97124	T16093C	-309.1C -524.3A -524.4C	**K1a**	Europe
MT97125	none	T152C	**U5a2d1a**	Europe (U5a)

Informatics

- Forensic mtDNA nomenclature is challenging!
- Commercial software now available
  - Softgenetics GeneMarker HTS
    - Compatible with forensic nomenclature
    - EMPOP formatted report
  - CLC Genomics Workbench
    - AFDIL / Qiagen – developed AQME Tool
  - ThermoFisher Scientific
    - Converge mtDNA Analysis (released October 2018)







Conclusions

- Mitochondrial Genome Protocol
  - Cost and data quality directed decision process
  - Selected AFDIL-developed procedure for reference-quality samples
    - Even coverage
    - Higher multiplexing
  - Degraded samples will need a different procedure
  - Analysis method must be high-throughput
    - High accuracy required for EMPOP submission

Thank You!
Questions?

Contact info:
Kevin.Kiesler@NIST.gov

√ Acknowledgements √

Armed Forces DNA Idenification Laboratory (AFDIL)
Kim Andreaggi
Charla Marshall

NIST Applied Genetics Group NGS Team
Dr. Peter Vallone, Group leader
Lisa Borsuk
Sarah Riman
Becky Steffen
Katherine Gettings

- Funding
  - NIST Special Programs Office: *Forensic DNA*
  - FBI Biometrics Center of Excellence: *Forensic DNA Typing as a Biometric tool*.

Q&A Session 10:55 – 11:05
Break 11:05 – 11:20

- Digital & Trace Tour signups please report to registration booth @ 11:00