

mtGenome Sequencing of NIST Reference Materials and Population Samples

Kevin Kiesler, M.S.

Research Biologist - NIST Applied Genetics Group

ISHI Workshop: The Future is Now for MPS mtDNA Analysis

September 24, 2018



Outline

- Reference materials and mtDNA sequencing
- Population scale sequencing
- Informatics

First – the Disclaimer

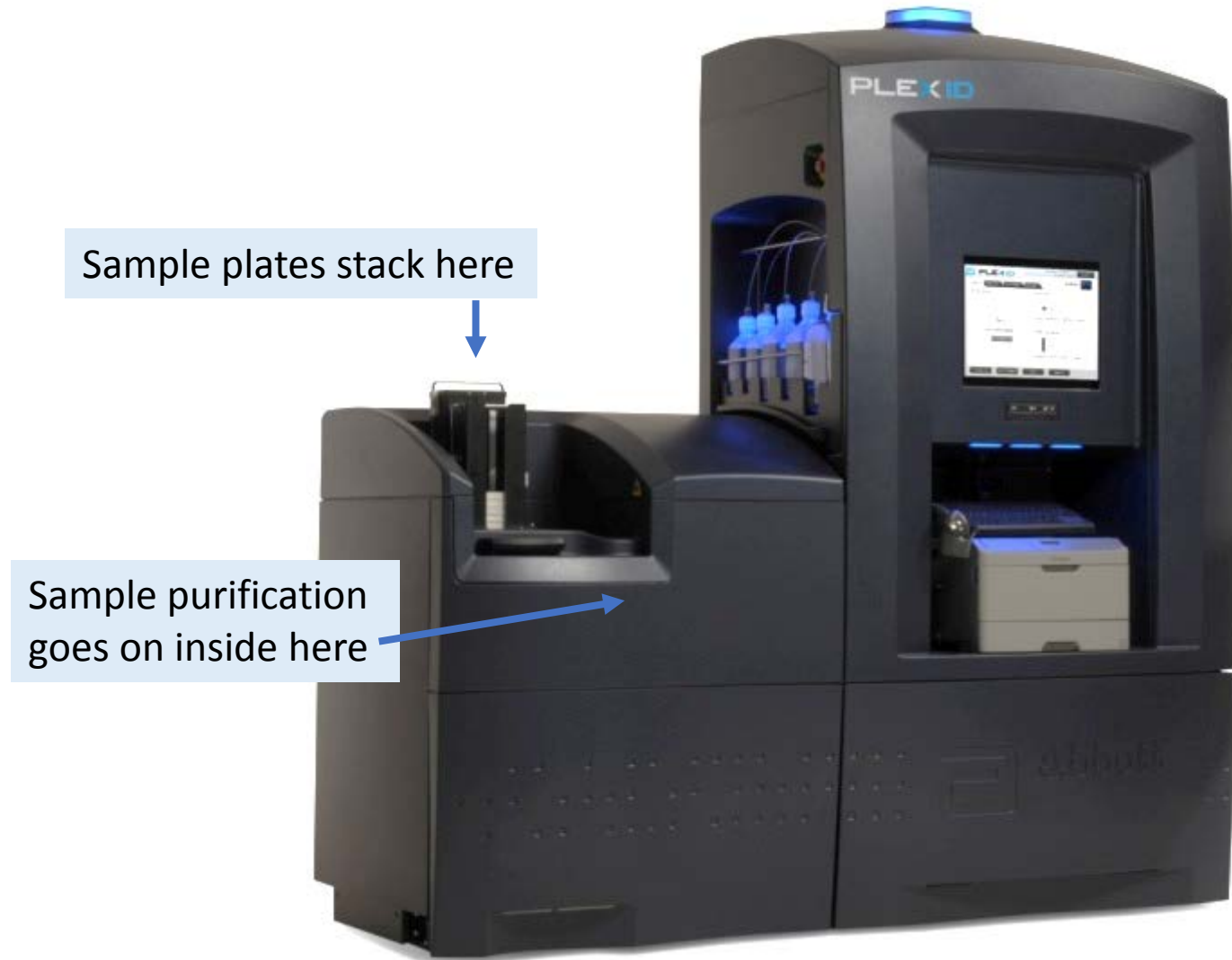
- **Disclaimer** - Points of view in this document are those of the authors and do not necessarily represent the official position or policies of the U.S. Department of Commerce or the Department of Justice. Certain commercial equipment, instruments, and materials are identified in order to specify experimental 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.

What is this?



- A) Time machine
- B) Water purifier
- C) Steam locomotive
- D) Mass spectrometer

What is this?



Sample plates stack here

Sample purification goes on inside here

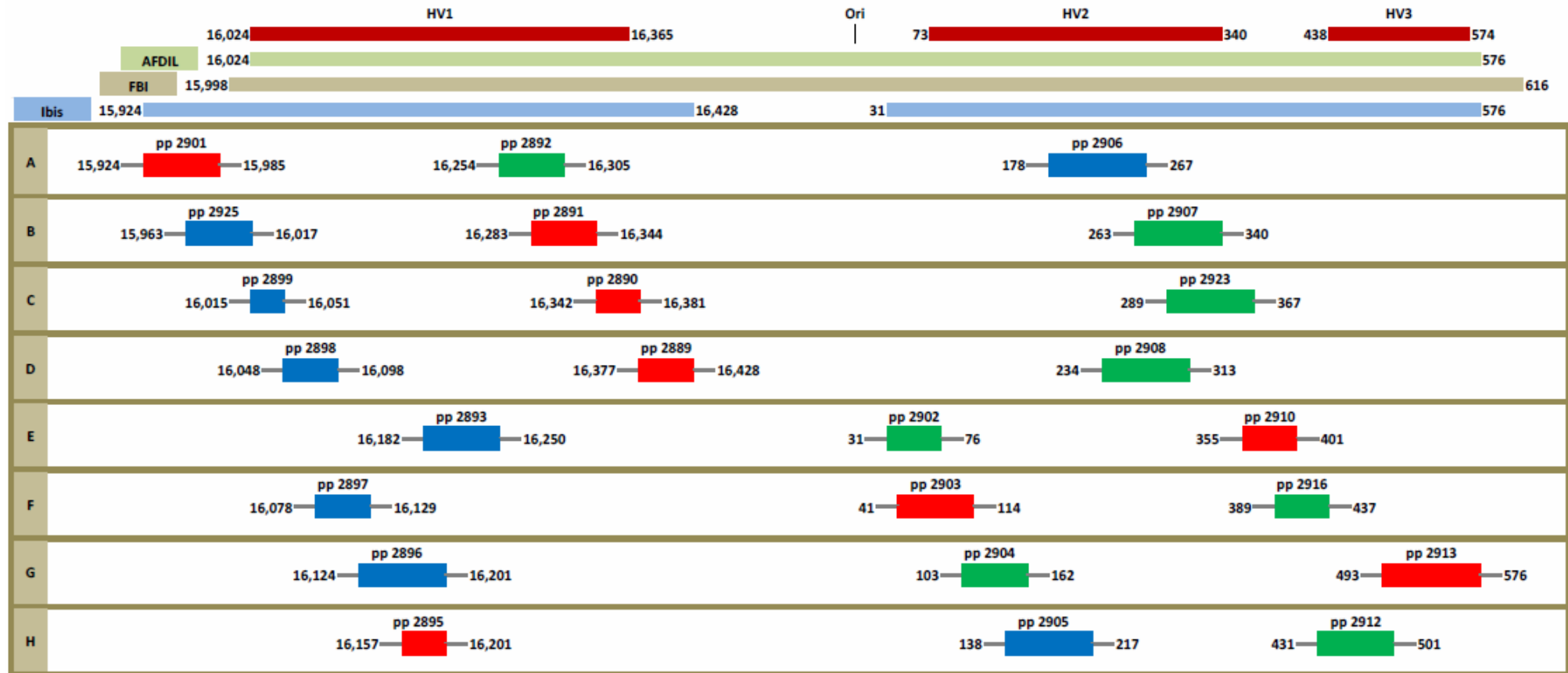
The **PLEX-ID**
From Abbott Molecular
& Ibis Biosciences

The mass spectrometer sits inside here

NIST evaluated this system in 2012 for potential “rapid, simple” mtDNA analysis

mtDNA by Mass Spectrometry

- PCR across HV1/HV2/HV3
 - Eight tri-plex reactions = 24 amplicons





Comparison of base composition analysis and Sanger sequencing of mitochondrial DNA for four U.S. population groups



Kevin M. Kiesler^{a,*}, Michael D. Coble^a, Thomas A. Hall^b, Peter M. Vallone^a

^a National Institute of Standards and Technology, 100 Bureau Drive, Gaithersburg, MD 20899, USA

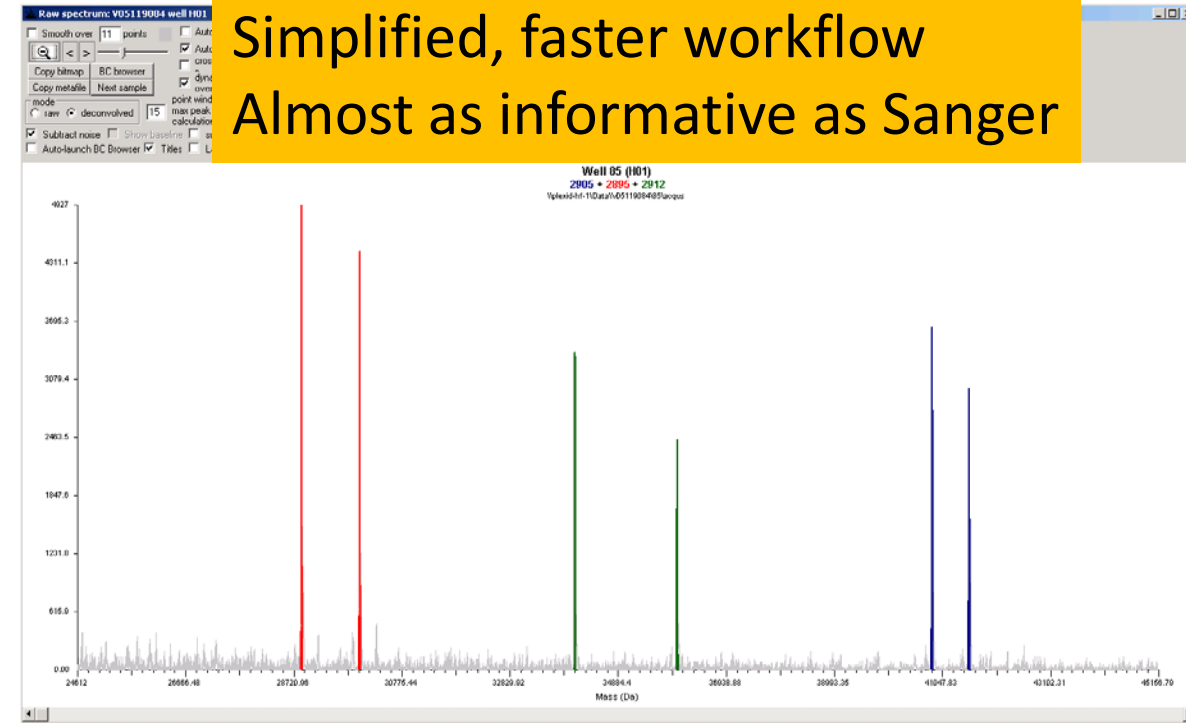
^b Ibis Biosciences, A Division of Abbott, 2251 Faraday Avenue, Suite 150, Carlsbad, CA 92008, USA

mtDNA by Mass Spec

- Desalt PCR products (automated)
- Electrospray injected Time-of-Flight (ESI-TOF) mass spectrometer
 - Review data for “trouble”
 - Sodium adducts, overlapping peaks, etc.
 - Compare to database
 - Known mtDNA types
 - Masses of PCR products
 - Predict DNA sequence
- Paper: compared with Sanger
 - n = 711 samples
 - 99.97 % concordant

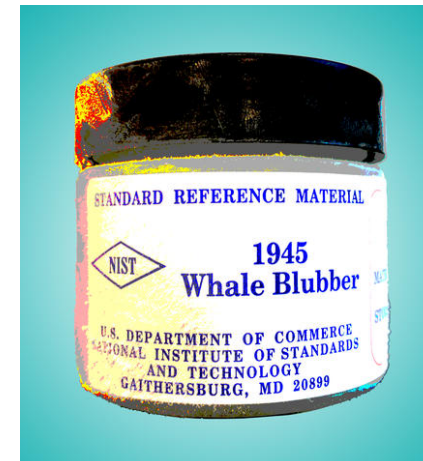
Outcome:

Simplified, faster workflow
Almost as informative as Sanger



NIST Standard Reference Materials (SRMs)

- ISO/Guide 30:2015
 - Reference materials (RMs) and certified reference materials (CRMs) (defined in 2.1 and 2.2) are widely used for the [calibration of measuring apparatus](#), for the [evaluation of measurement procedures](#) and for the internal or external [quality control of measurements](#) and laboratories.
- Assess a new technology - calibration
- Evaluate a method - implementation



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 from HV1 region of CHR containing C-stretch

- SRM 2392-I

- One component

- DNA from cell line HL60

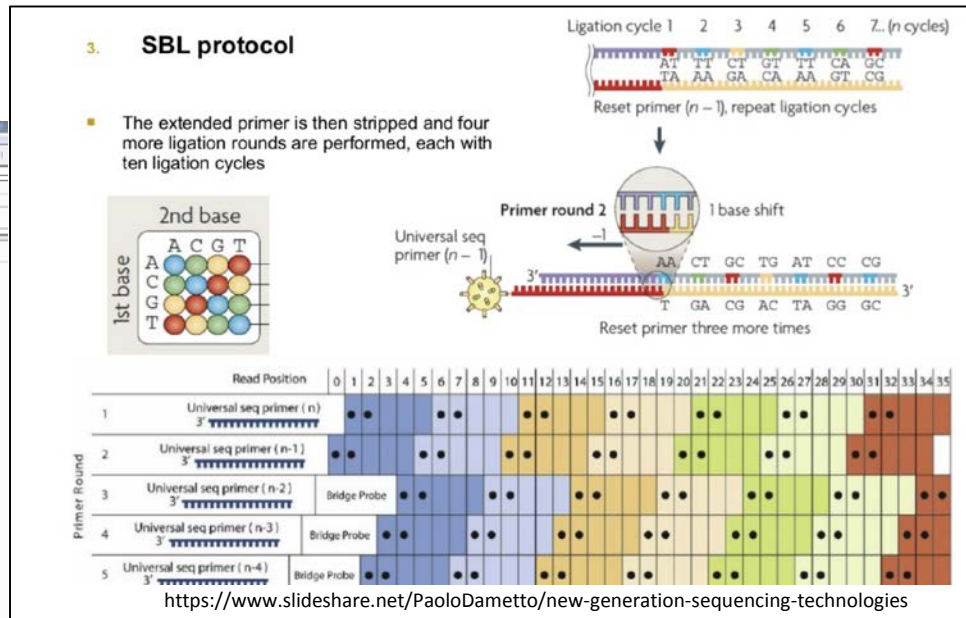
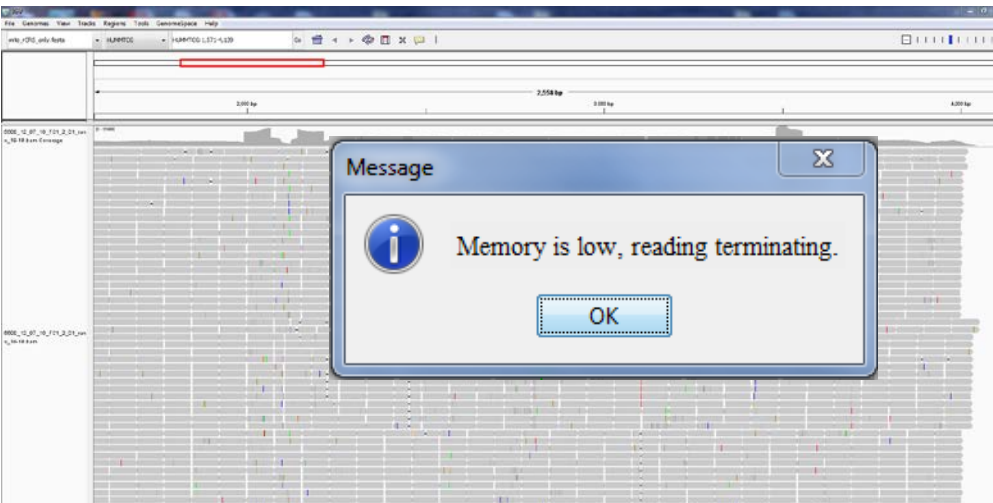
- Characterized with Sanger methods

- Released in 2001



Assessing NGS Using Reference Materials

- In 2012, all we had was a SOLiD system...
 - Sequencing by Oligonucleotide Ligation and Detection (SOLiD)
 - Ligation chemistry based (it's complicated) system for whole human genomes
- Very high coverage mtGenome sequence
 - ~ 45,000 X

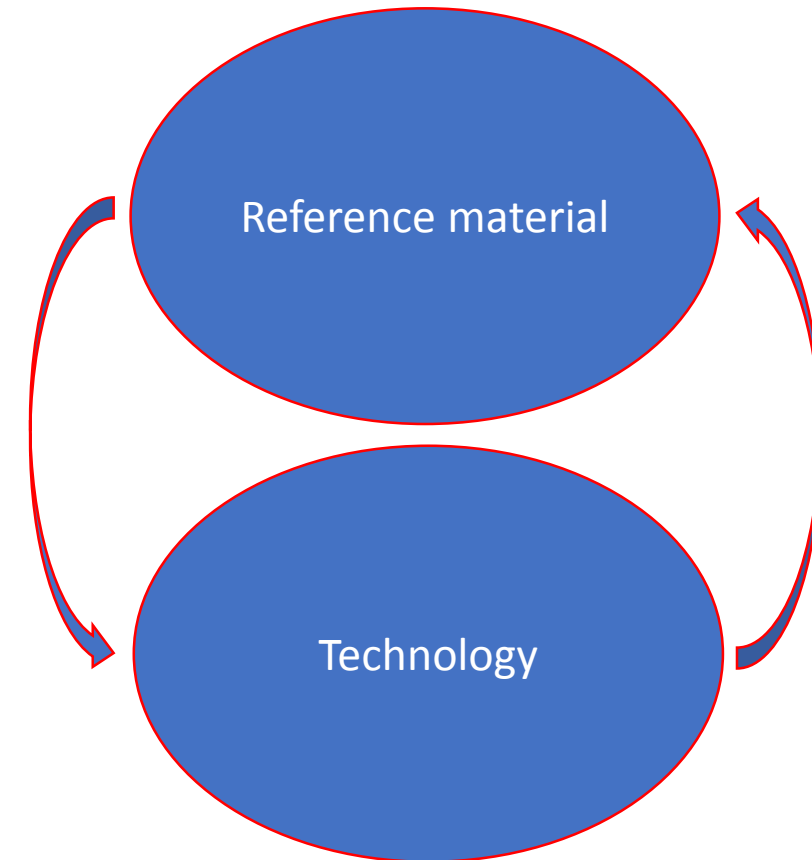


All the Sanger-Based Information Checked Out

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

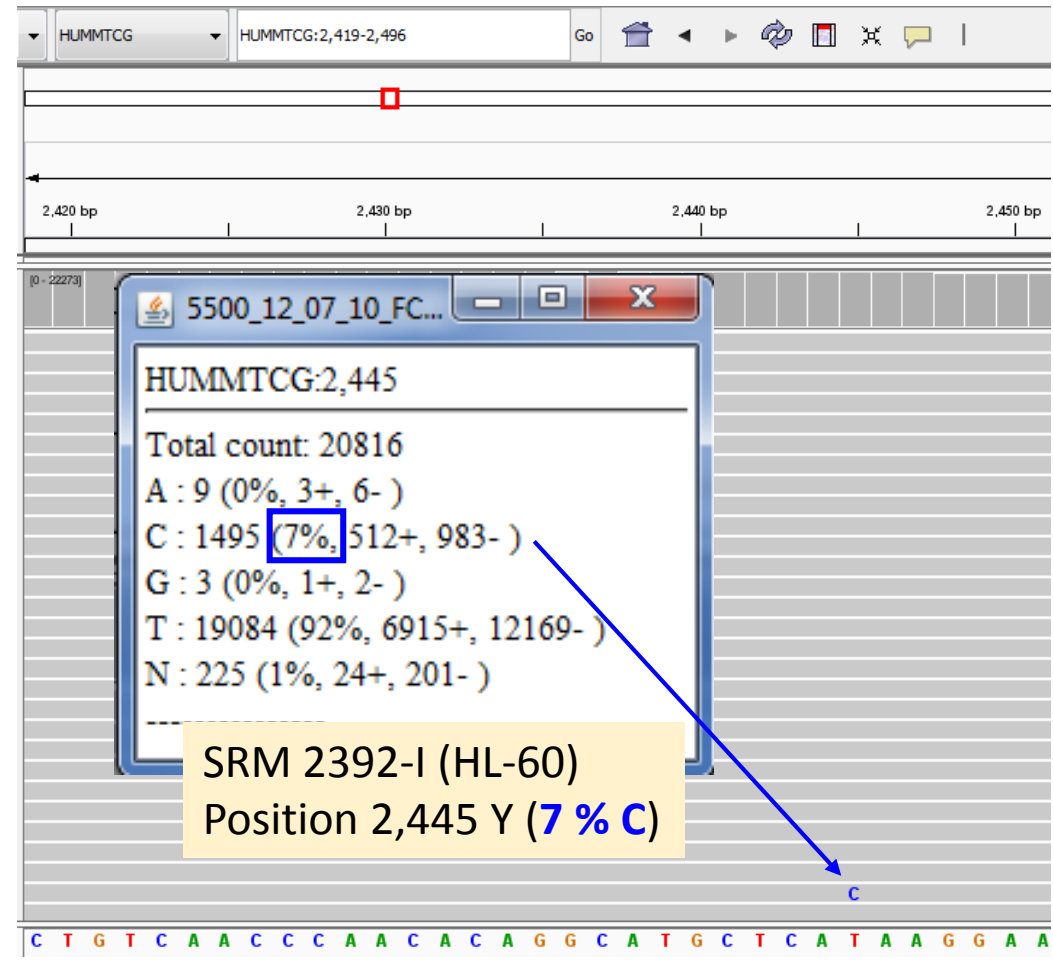
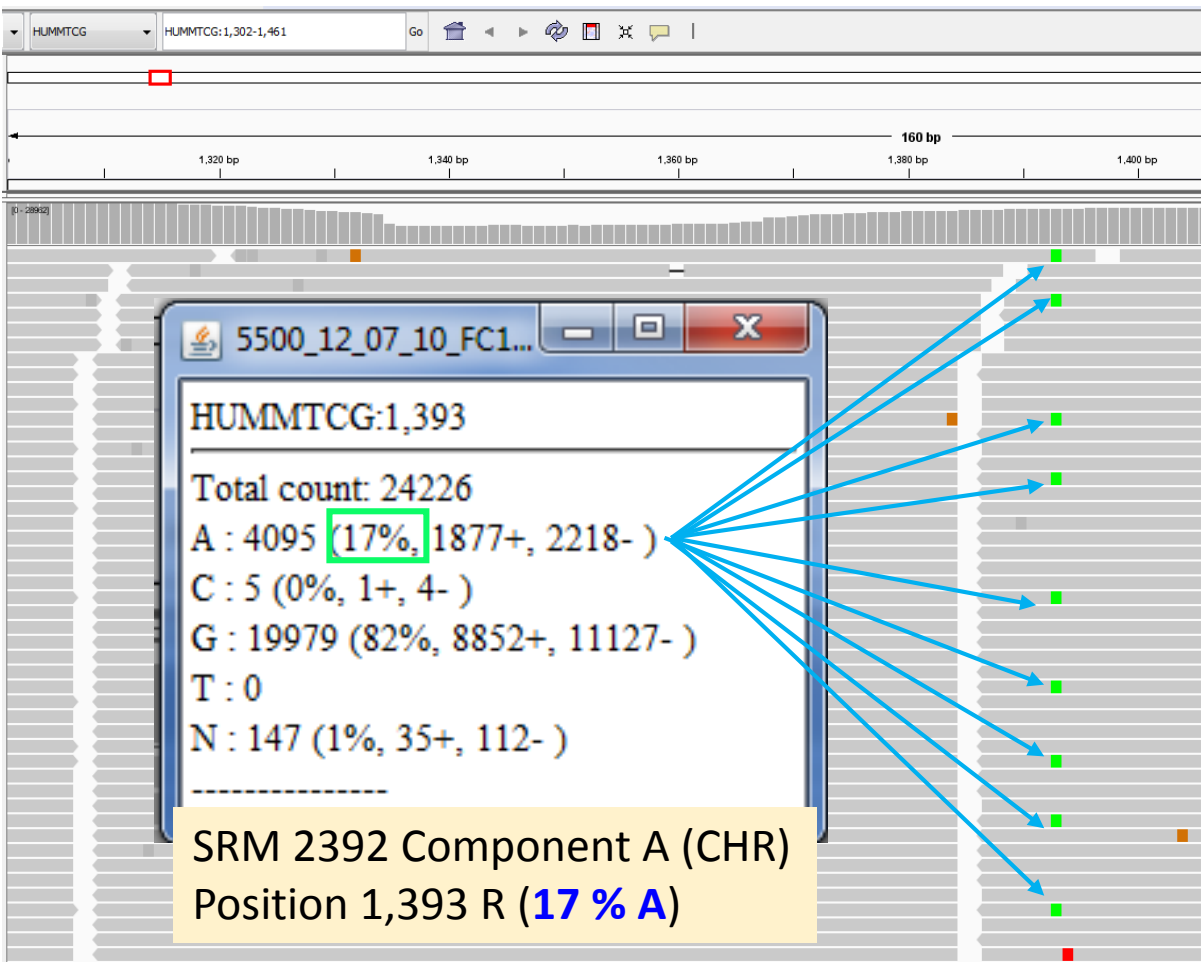
Site	rCRS	GM09947A	Comments
93	A	G ✓	
195	T	C ✓	
214	A	G ✓	
263	A	G ✓	
309.1		C ✗	insertion
309.2		C ✗	insertion
315.1		C ✗	insertion
750	A	G ✓	
1438	A	G ✓	
3107	C		deletion
4135	T	C ✓	
4769	A	G ✓	
7645	T	C ✓	
7861	T	C ✓	
8448	T	C ✓	
8860	A	G ✓	
9315	T	C ✓	
13572	T	C ✓	
13759	G	A ✓	
15326	A	G ✓	
16311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

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



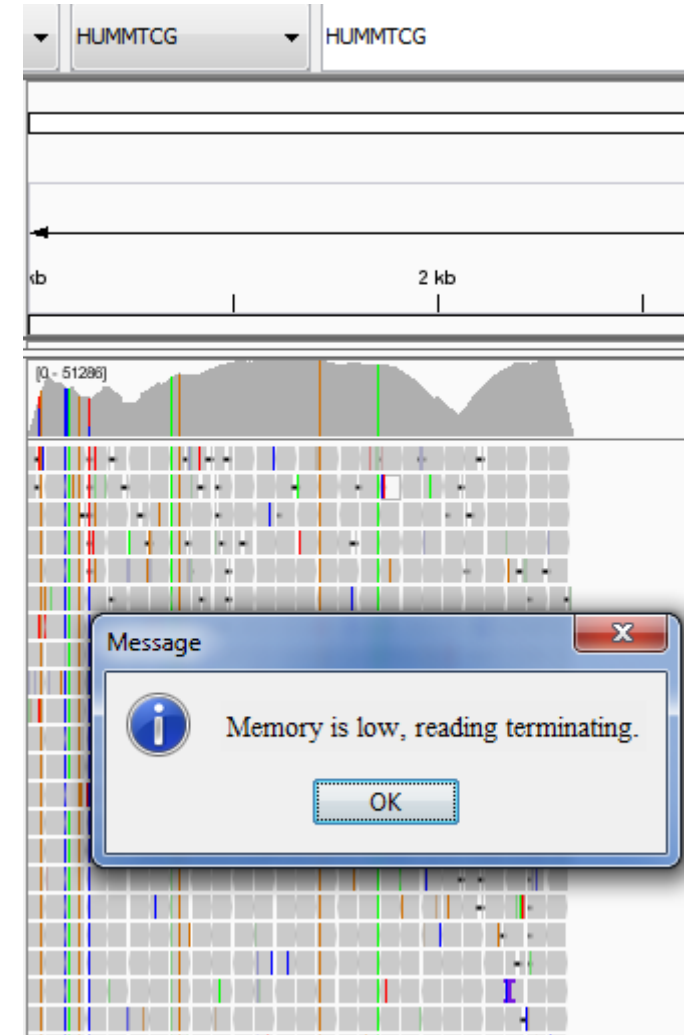
We Could Detect Some New Features!

- Heteroplasmy below what we can see with Sanger methods



We Sent Some Samples to Service Labs, Friends

- Beckman Genomics sequencing service
 - Seven mtGenomes → One lane on a HiSeq
 - ≈ 40,000 X coverage!!! (≈ 3 to 4 million bases each)
 - Too much to display in IGV (not enough RAM) →
- Edge Bioserv (Gaithersburg, MD)
 - Ion Torrent Personal Genome Machine (PGM) in 2013
 - Seven mtGenomes on a 314 chip
 - ≈ 200 X to 400 X coverage
 - Illumina MiSeq in 2013
 - Seven mtGenomes on 2 x 150 v2 chemistry
 - ≈ 10,000 X to 20,000 X coverage
- Children's National Hospital (in collaboration)
 - Pacific Biosystems RS in 2014
 - Long-read technology, low coverage
 - Some variants missed



Then We Got Some Desktop Sequencers of Our Own

- Ion Torrent PGM
 - September 2012
 - Chemistry improvements over time
 - 200 bp, 300 bp, 400 bp, HiQ
- Illumina MiSeq
 - September 2013
 - Chemistry improvements over time
 - V2 2x150 bp, V2 2x250 bp, V3 2x300 bp



Multiple Orthogonal Measurements

- Great approach for certifying reference materials!

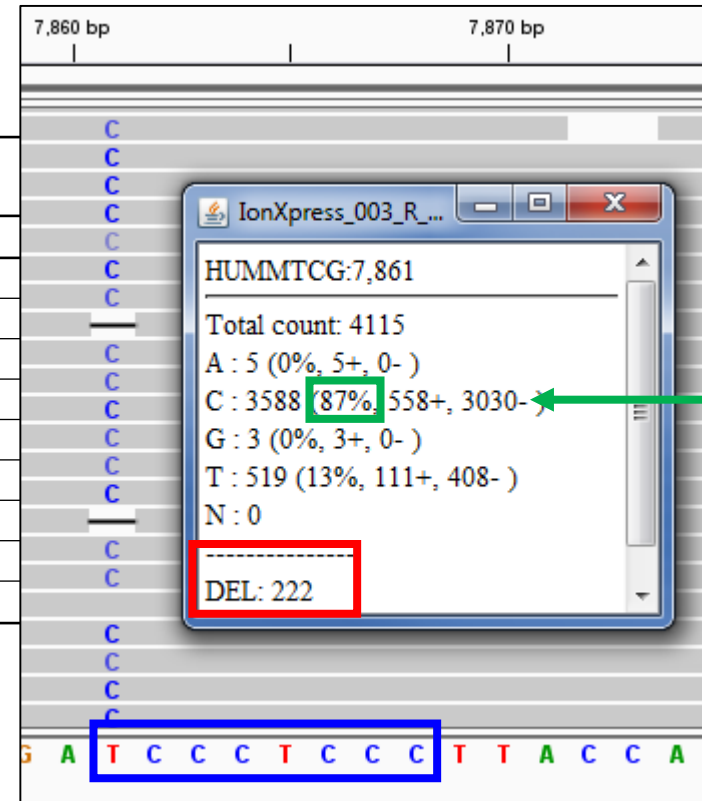
	SRM 2392			SRM 2392-I	
	Component A	Component B			
Nucleotide	64 T	1393 A	7861 C	2445	5149
PGM Edge	26.8	21.2	72.7	10.6	5.1
PGM NIST 1	24.3	15.6	45.0	7.5	9.1
PGM NIST 2	25.0	17.5	65.7	7.7	7.7
PGM NIST 3	29.7	16.5	59.4	7.7	7.7
PGM NIST HiQ	33.2	15.2	77.7	7.7	8.4
MiSeq Edge	33.0	19.0	87.4	10.7	6.8
MiSeq 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



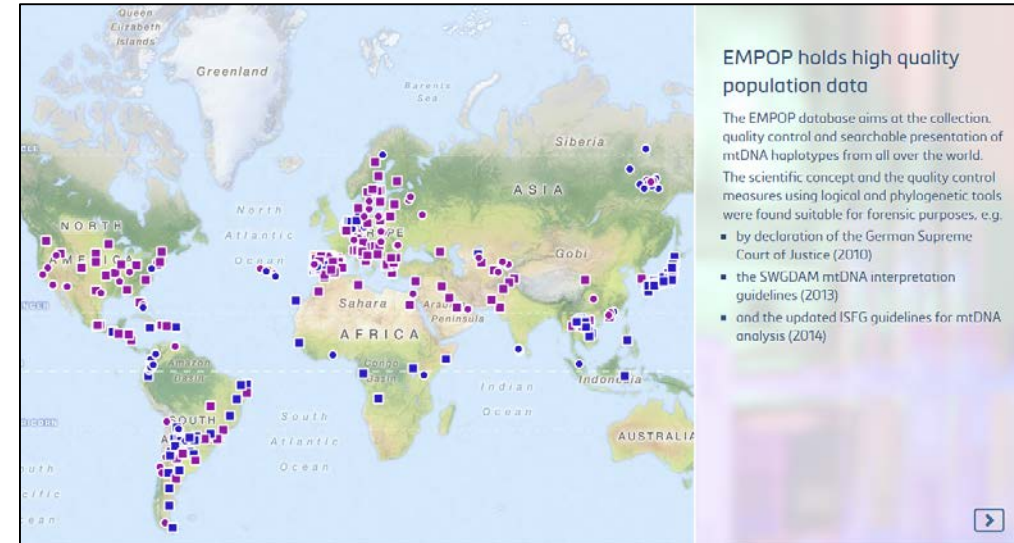
Conclusions

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

Population Scale Sequencing

Project Planning - Goals

- Submit forensic-quality **whole mtGenome** data to EMPOP
 - Current database (V3, 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 Planning

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



Illumina MiSeq FGx



Ion Torrent S5

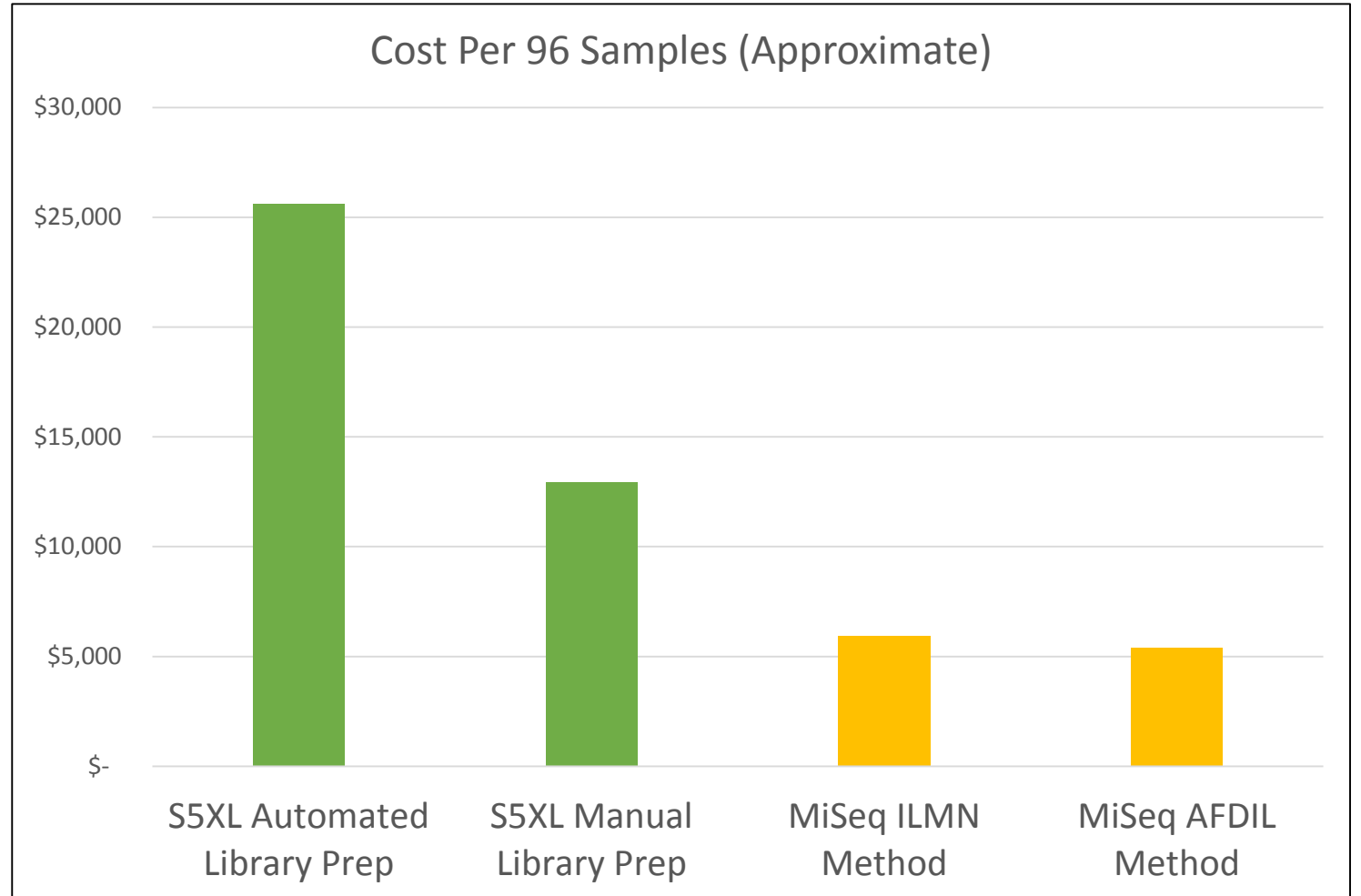


Ion Chef

Project Planning: Instrument Selection

- Considerations

- Cost
- Time/labor
 - Automation



Project Planning: Protocol Selection

- Options

- Illumina whole genome 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)

- AFDIL whole genome procedure

- 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

Obtain the following PCR Primers from a general oligo supplier:

Table 6 User-Supplied PCR Primers¹

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'

¹Stawski, 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

BMC Genomics



Methodology article

Open Access

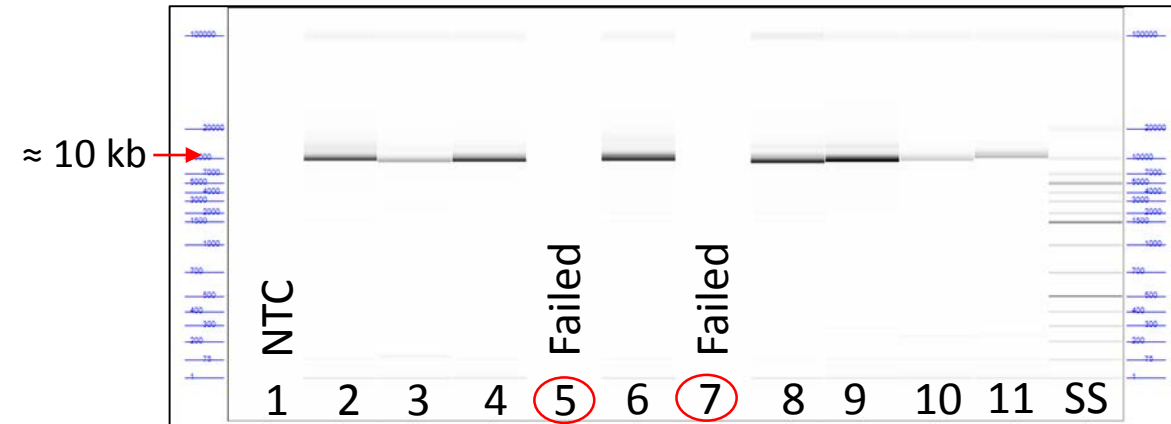
Sequencing strategy for the whole mitochondrial genome resulting in high quality sequences

Liane Fendt¹, Bettina Zimmermann¹, Martin Daniaux² and Walther Parson^{*1}

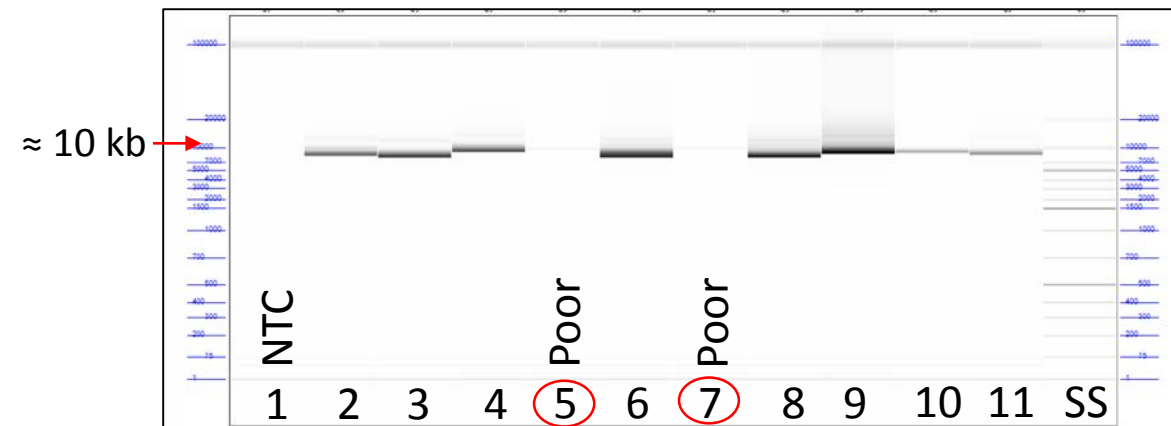
Pilot Data – Illumina Procedure Amplification

- PCR could be more robust
 - 18 of 95 ($\approx 18\%$) of samples - low yield
 - Reproducible
 - Only one of two amplicons affected
 - Not a sample quality issue

Amplicon 1 – First try



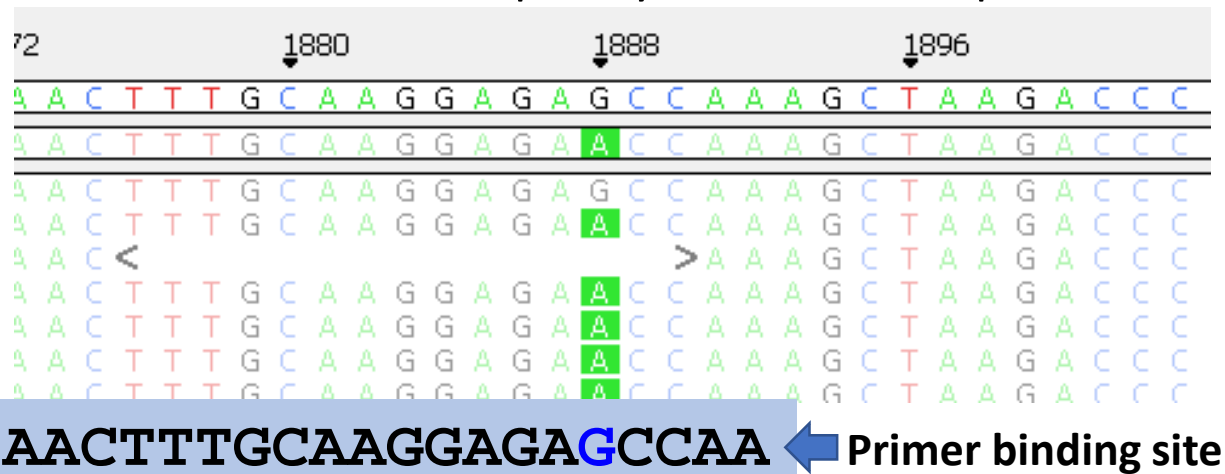
Amplicon 1 – Second try (re-PCR)



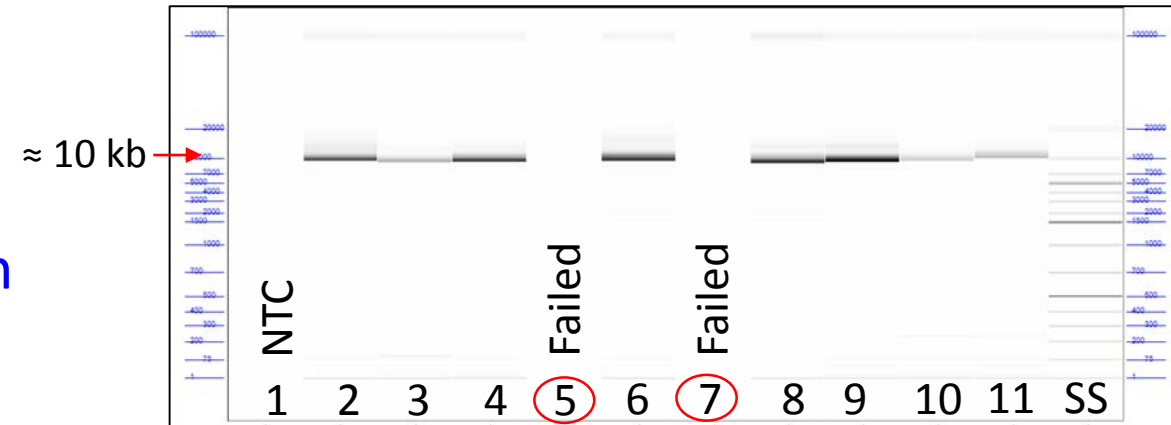
SS = Size standard

Pilot Data – Illumina Procedure Amplification

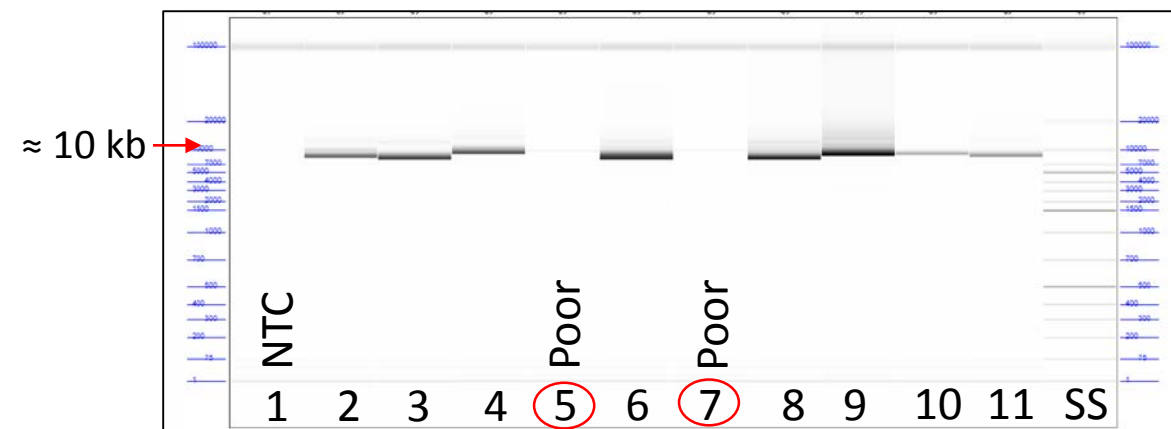
- PCR could be more robust
 - 18 of 95 ($\approx 18\%$) of samples - low yield
 - Reproducible
 - Amplicon 1 **primer binding site mutation**
 - 16 of 18 low/neg samples had this
 - All haplogroup T (G1,888A)
 - HG-T frequency $\approx 10\%$ in Europeans



Amplicon 1 – First try



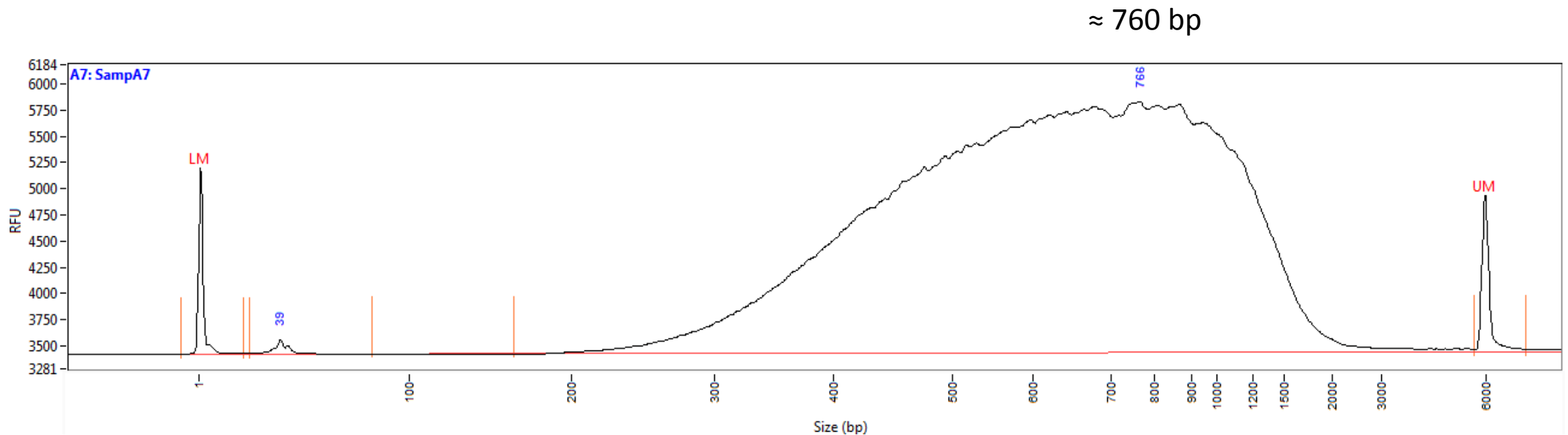
Amplicon 1 – Second try (re-PCR)



SS = Size standard

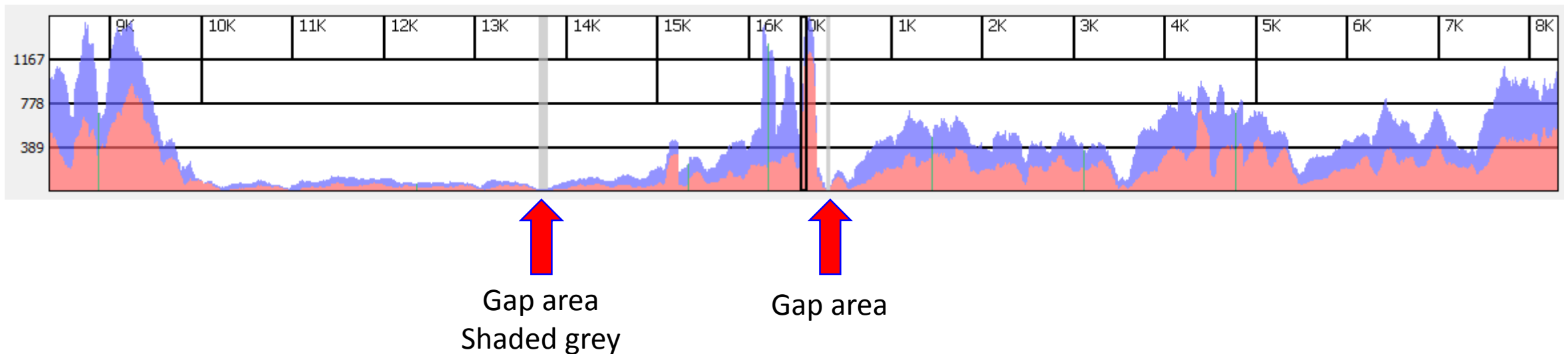
Library QC – Illumina Procedure

- Nextera XT library
 - Default parameters per protocol
 - 1 ng DNA
 - 5 minute fragmentation



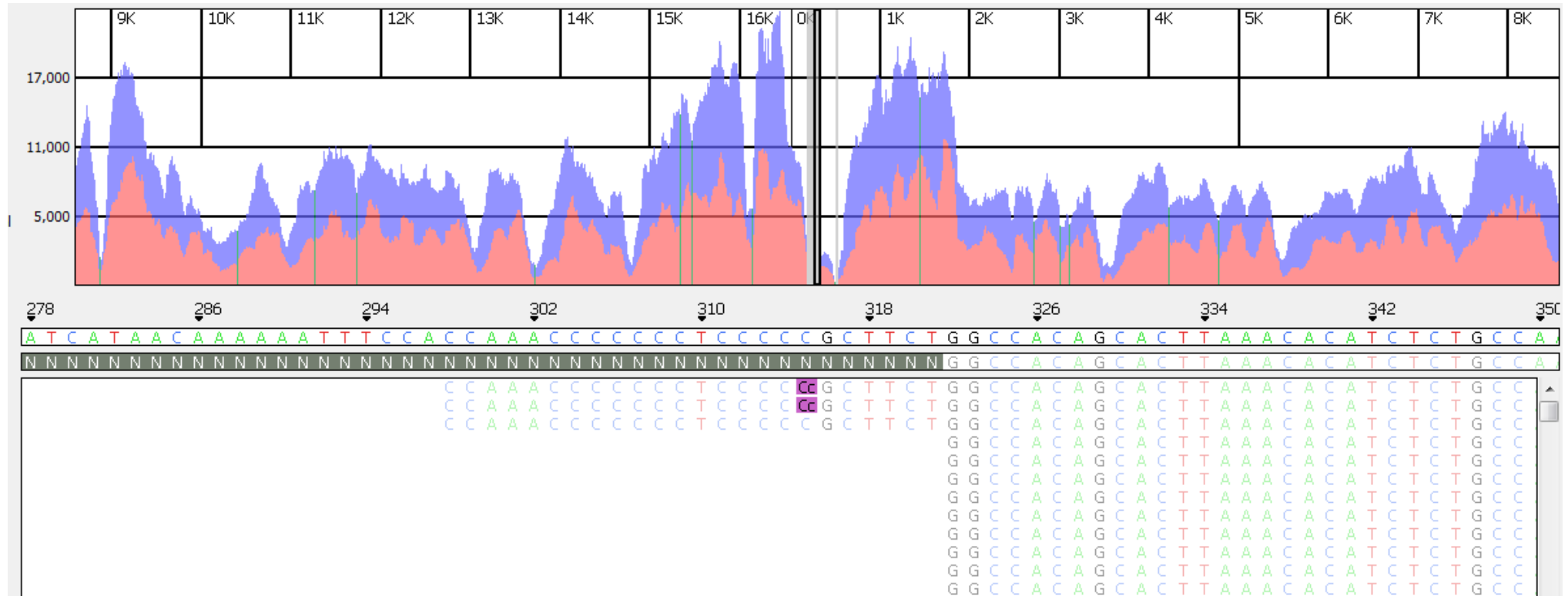
Sequencing Results - Illumina

- Most samples had > 99 % of mgGenome covered
 - Due to sensitivity of Nextera XT library prep
 - & Normalization of PCR input (amplicons 1 & 2 separately) ☹️
- Some gaps in coverage
 - Where amplification was poor



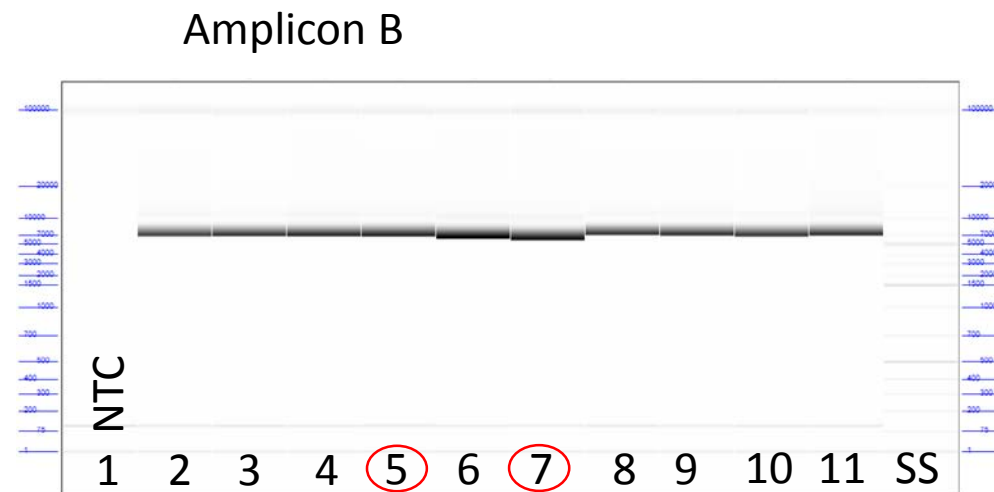
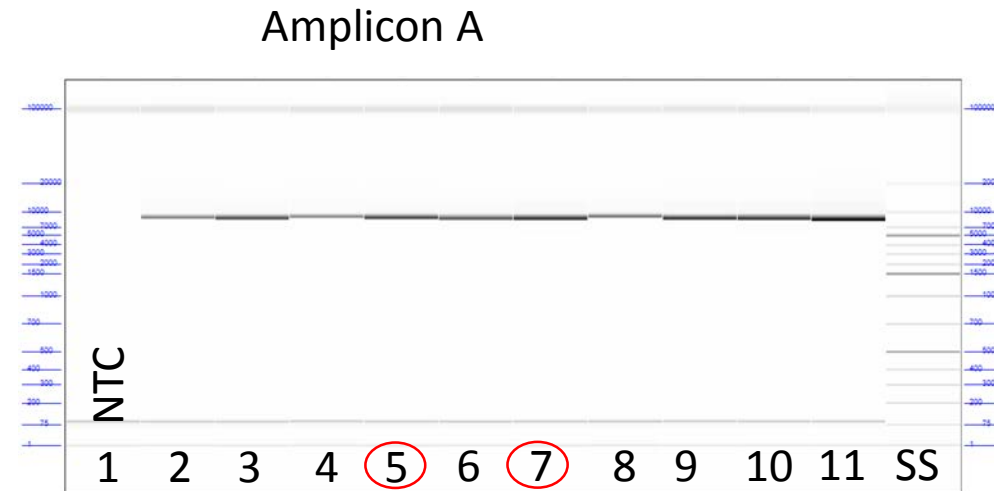
Sequencing Results - Illumina

- Some gaps in coverage
 - Frequent gaps in HV2 C-stretch
 - Even with good overall coverage



Pilot Data – AFDIL Procedure Amplification

- Primers from Fendt *et al.* 2009
- TaKaRa LA Taq w/GC Buffer & BSA
- 100 % of samples amplified
 - Amplicon B gives higher yield than Amp A
 - Amp B contains control region
 - Both amplicons \approx 8 kb
- Advantage of GC Buffer & BSA?



SS = Size standard

Protocol Optimization – Kapa HyperPlus

- What are the best parameters for library preparation?
 - Fragmentation time
 - 5 min to 30 min
 - Input DNA
 - 10 ng to 1000 ng (*PCR product*)
- Optimum library size
 - 2 x 300 bp cartridge planned
 - ≈ 450 bp library molecules
 - 300 bp insert + 150 bp adaptors
 - No more than ≈ 1500 bp → poor clustering

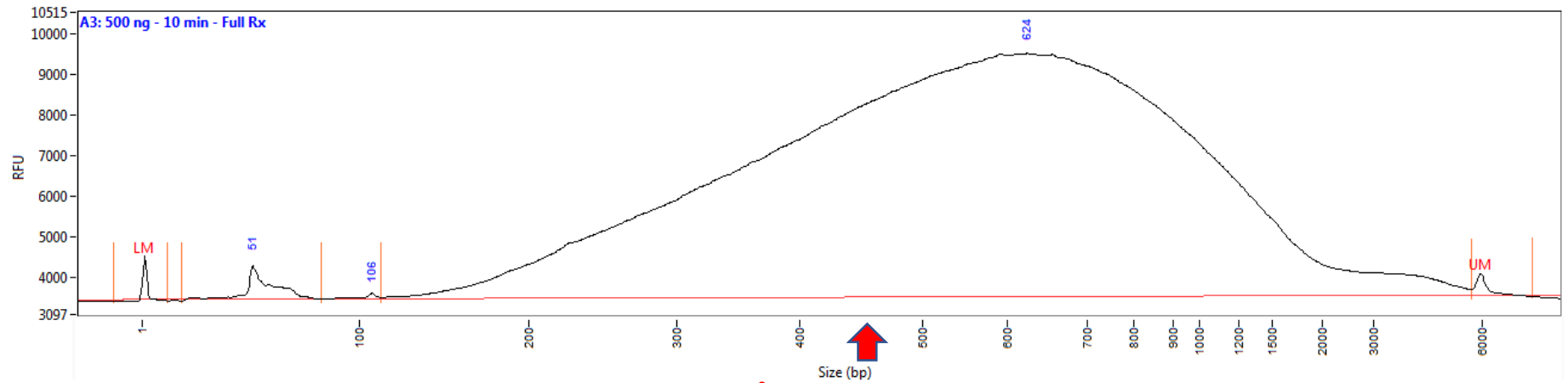
Mode fragment length	Incubation time at 37°C*	Optimization range
600 bp	5 min	3 – 10 min
350 bp	10 min	5 – 20 min
200 bp	20 min	10 – 25 min
150 bp	30 min	20 – 40 min

DNA input	Expected conversion rate
1 – 10 ng	5 – 20%
11 – 100 ng	10 – 50%
>100 ng	50 – 100%

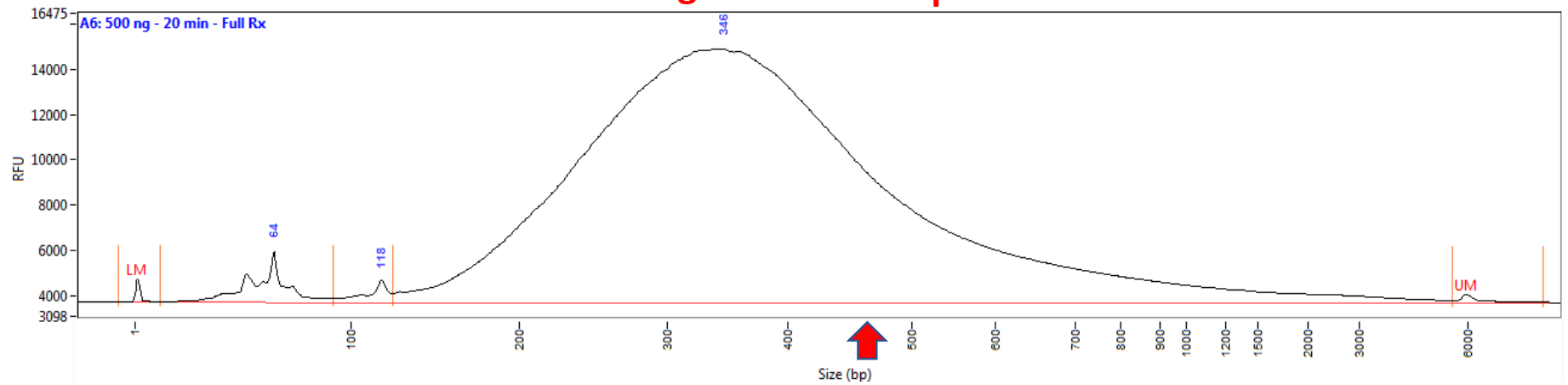
Fragmentation Time Optimization

All reactions
500 ng DNA
input

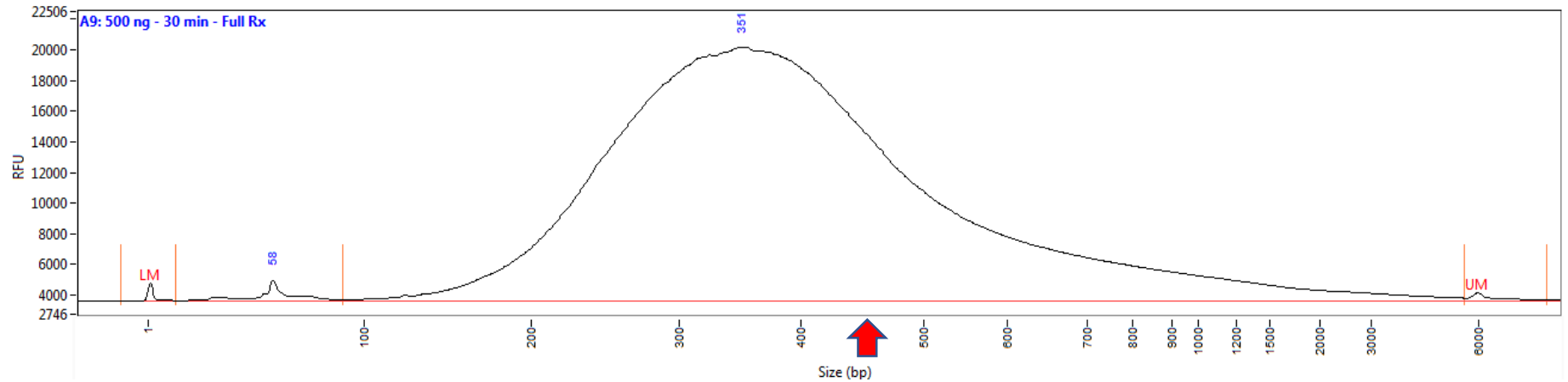
10 min



20 min



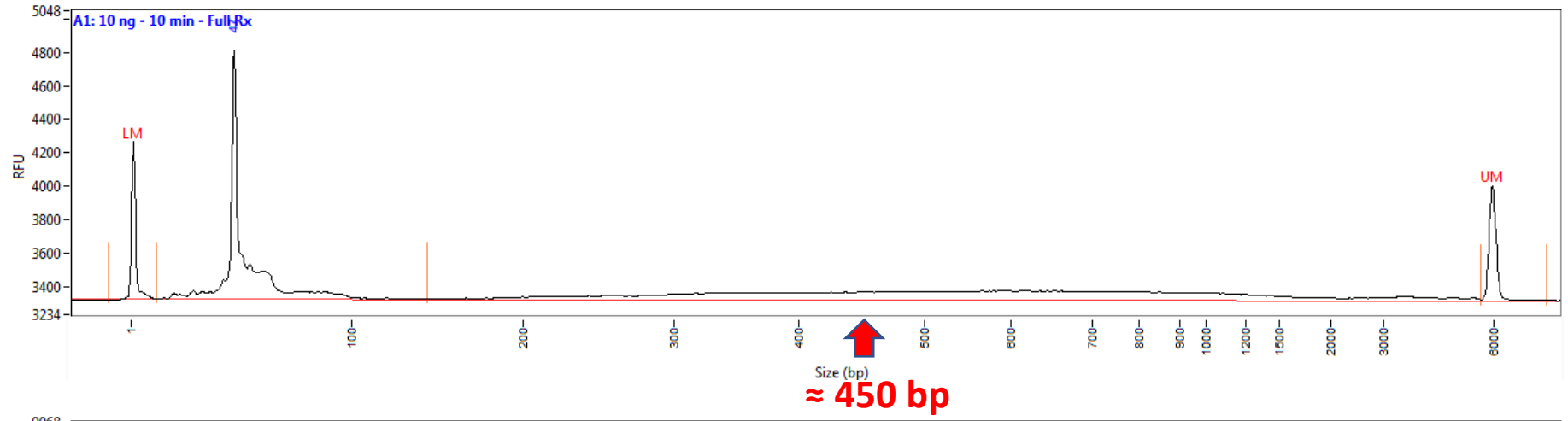
30 min



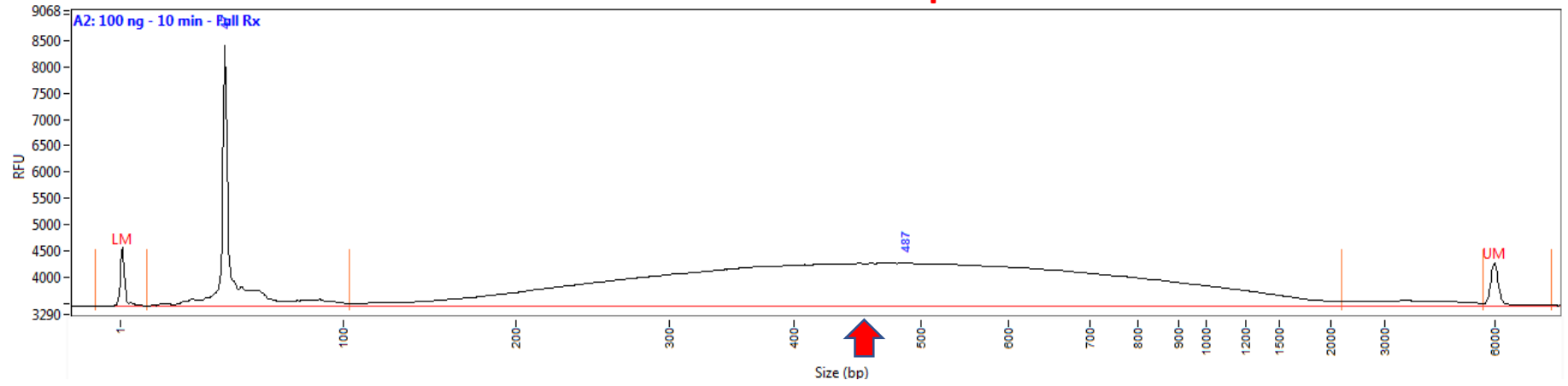
DNA Input Optimization

All reactions
10 min
fragmentation

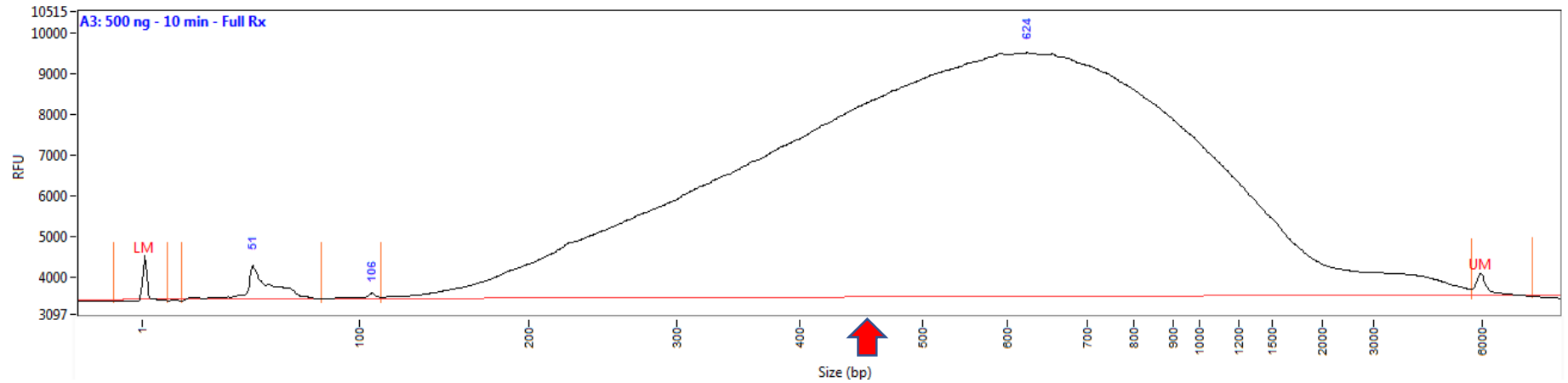
10 ng



100 ng

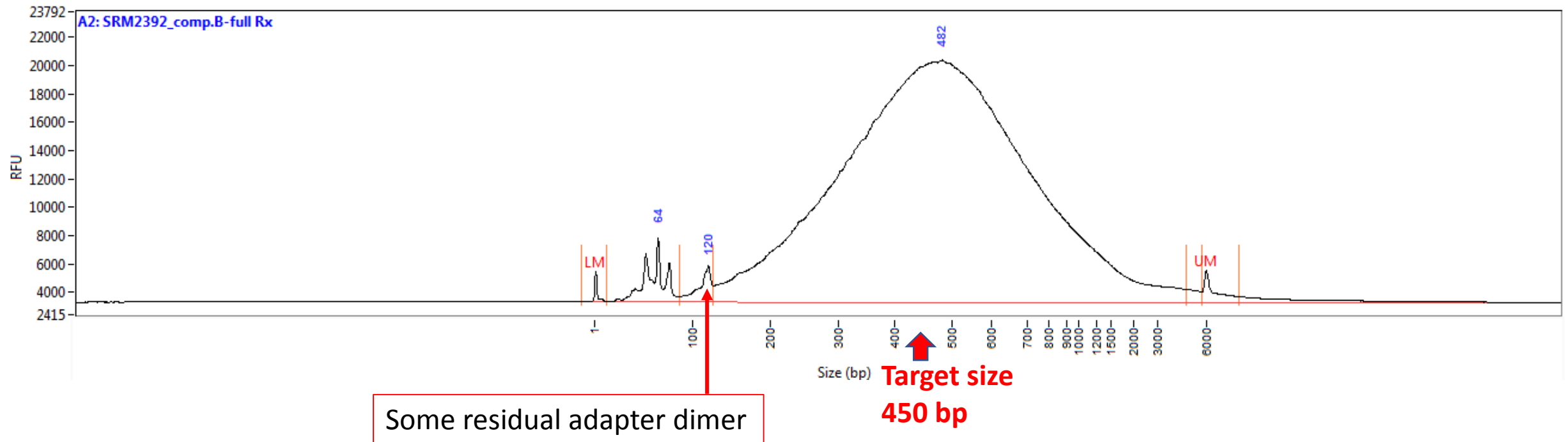


500 ng



Optimized Conditions

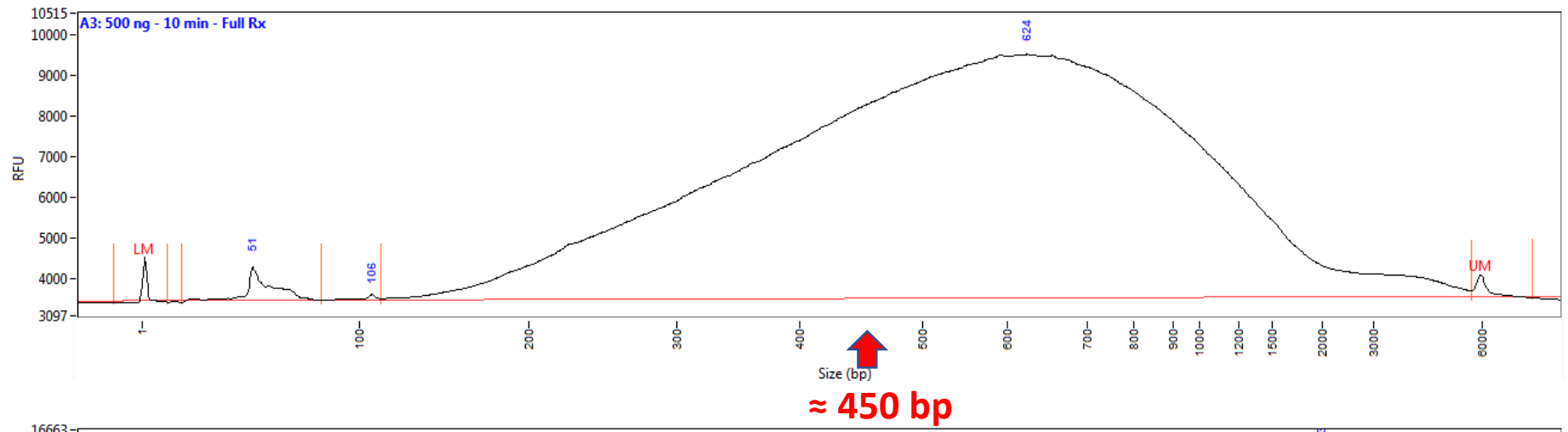
- Fragmentation time = 15 minutes
- DNA input \approx 300 – 400 ng



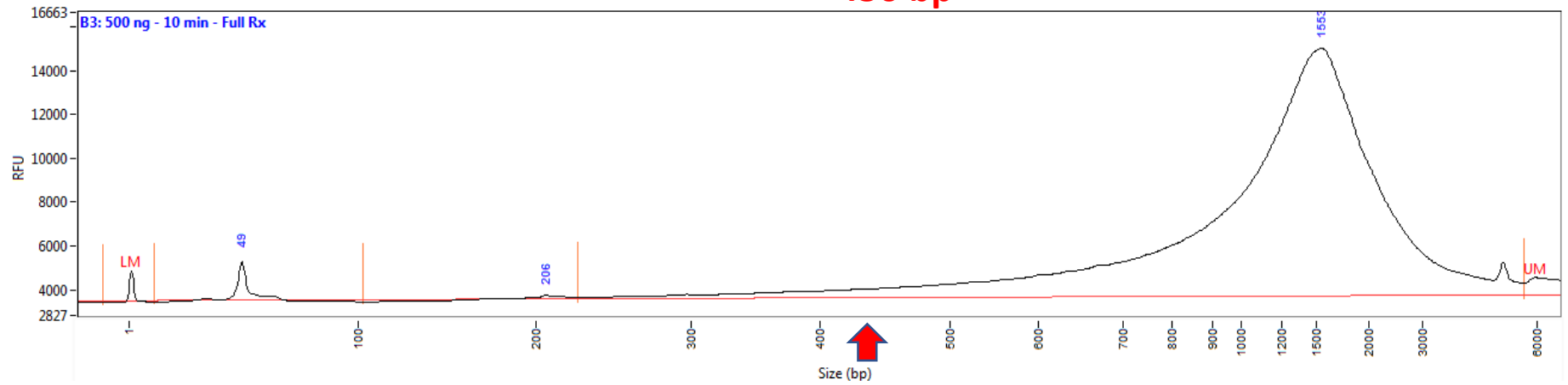
Half Reaction Trial

All reactions
10 min frag.,
500 ng DNA

Full
Rx



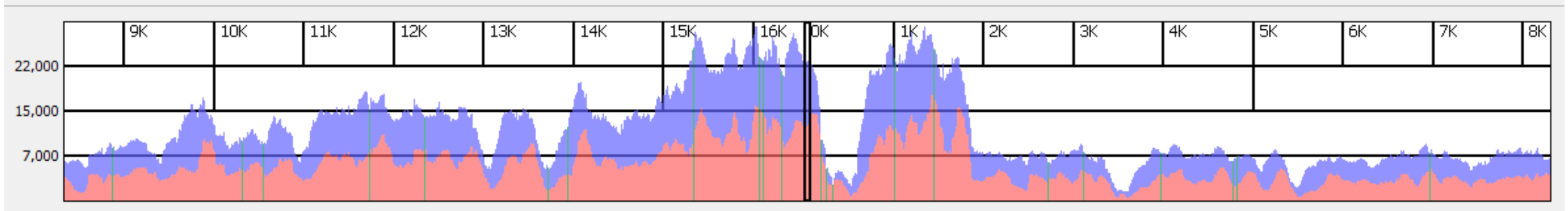
Half
Rx



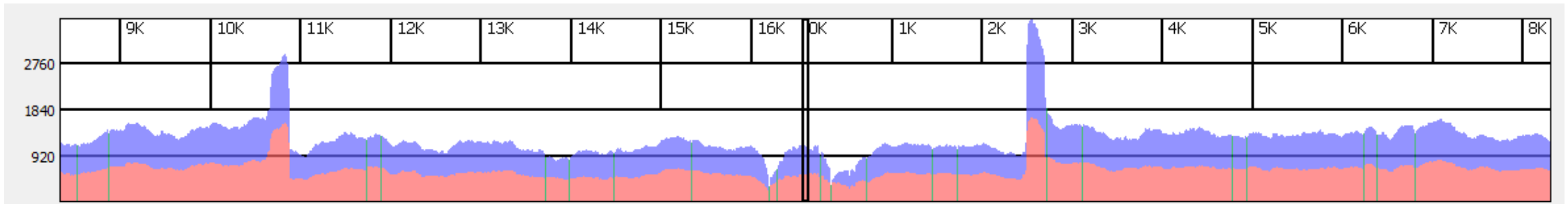
Reaction kinetics not equal at half volume library construction.

Pilot Data – More Consistent Coverage Depth

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)

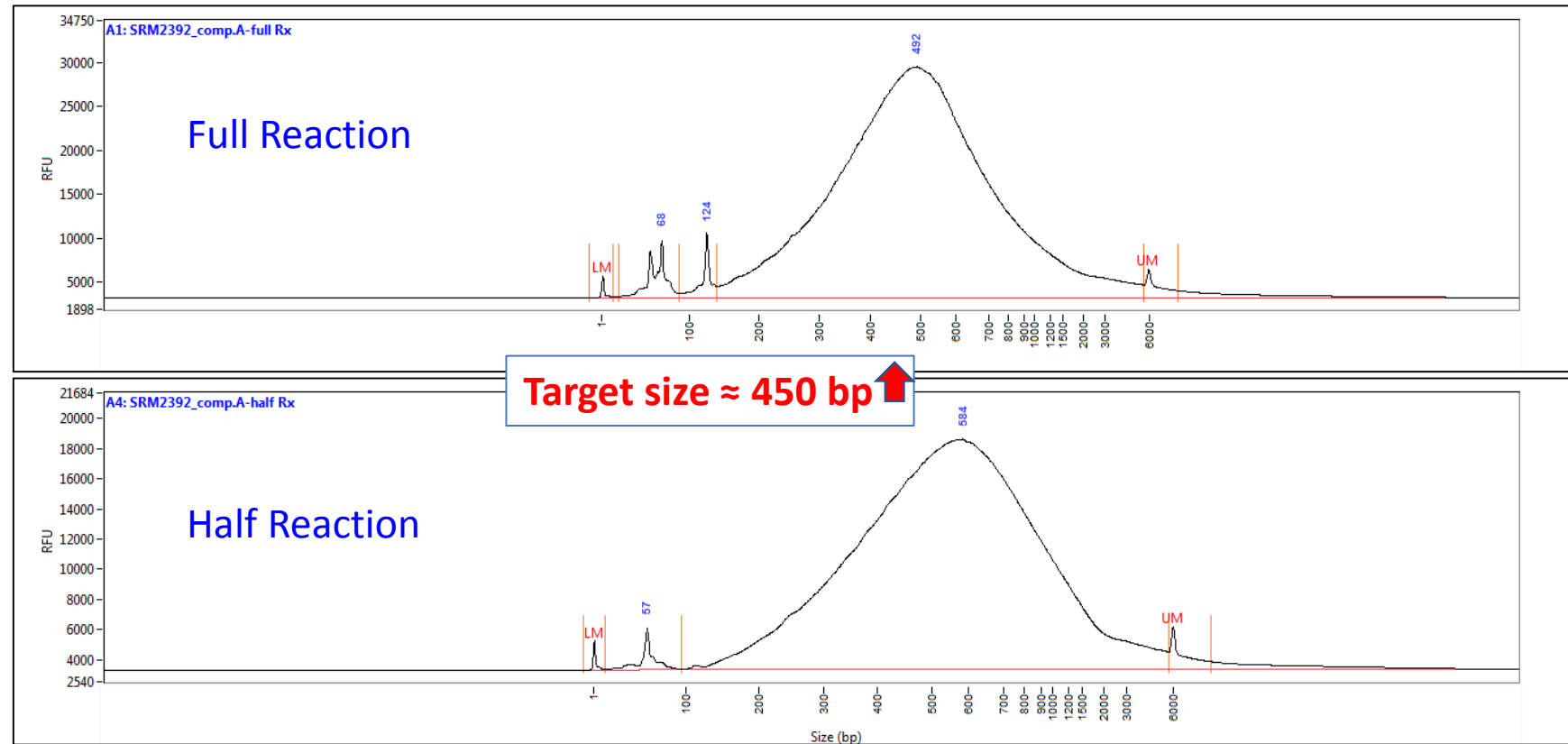
Pilot Data – Haplogroup Estimation

- 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)

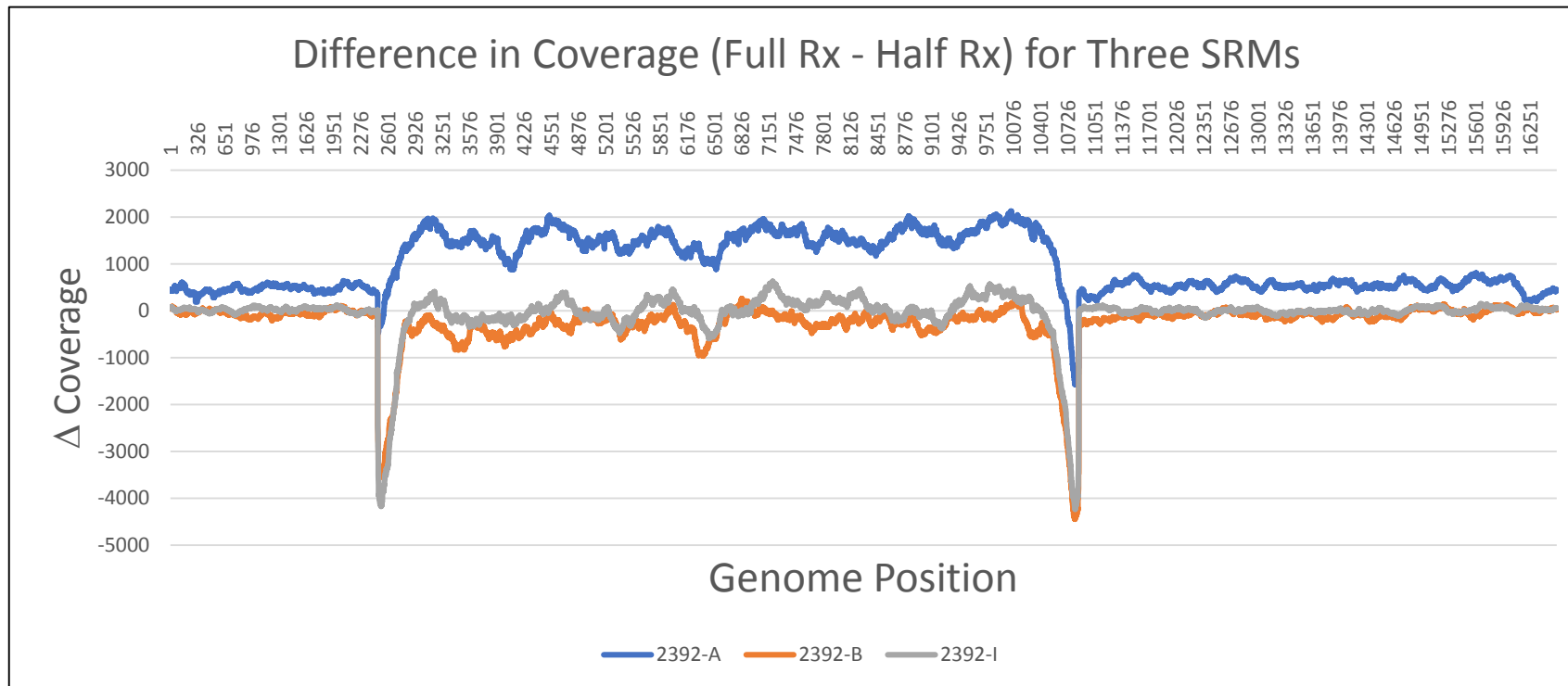
Half Reaction Library Prep - Save \$\$\$?

- Optimized fragmentation conditions
 - DNA input \approx 350 ng
 - Full reaction
 - 15 min
 - Half reaction
 - 20 min



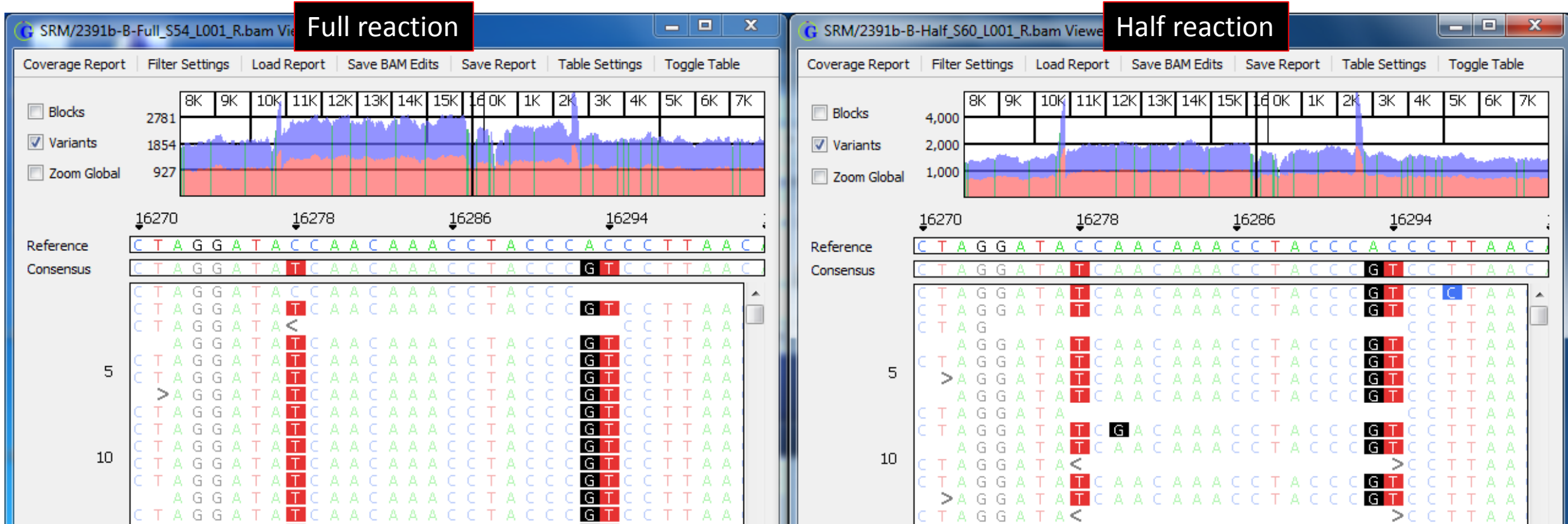
Compare Results – Full Reaction vs Half

- No major bias in coverage introduced
 - Half reaction always higher where amplicons overlap

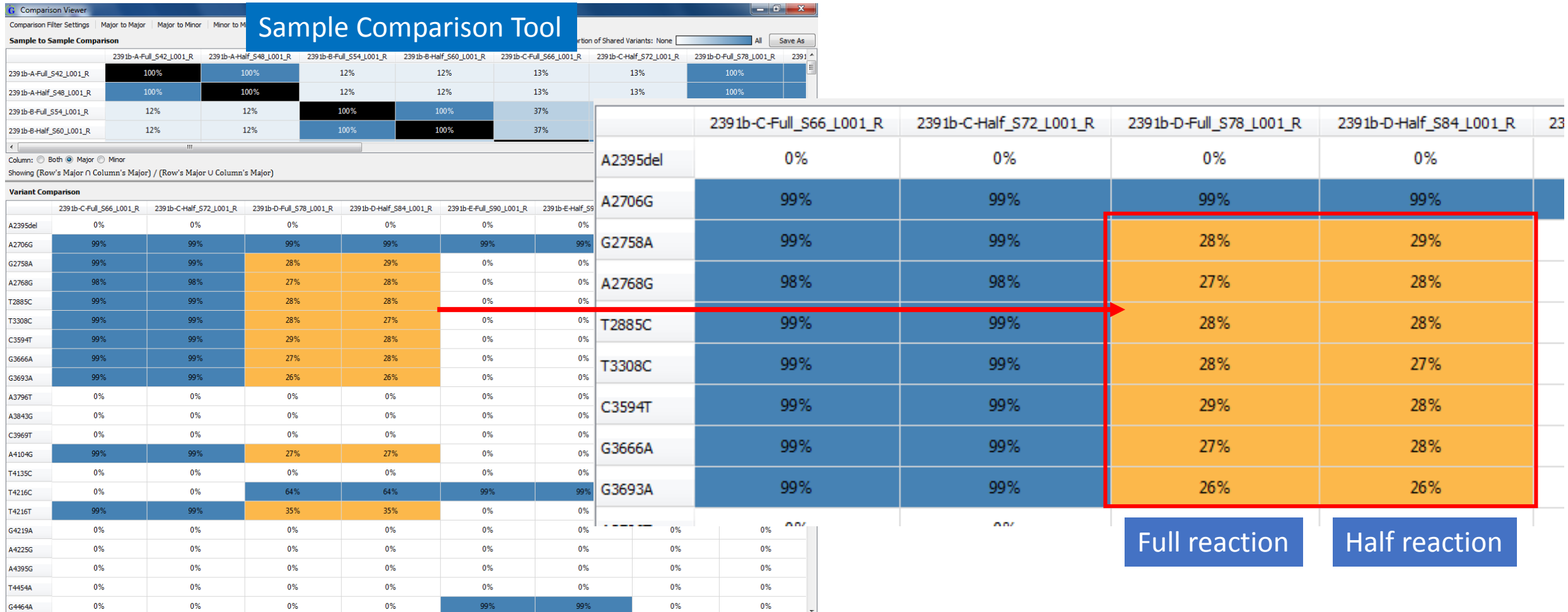


Compare Results – Full Reaction vs Half

- No major bias in coverage introduced
 - Half reaction always higher where amplicons overlap

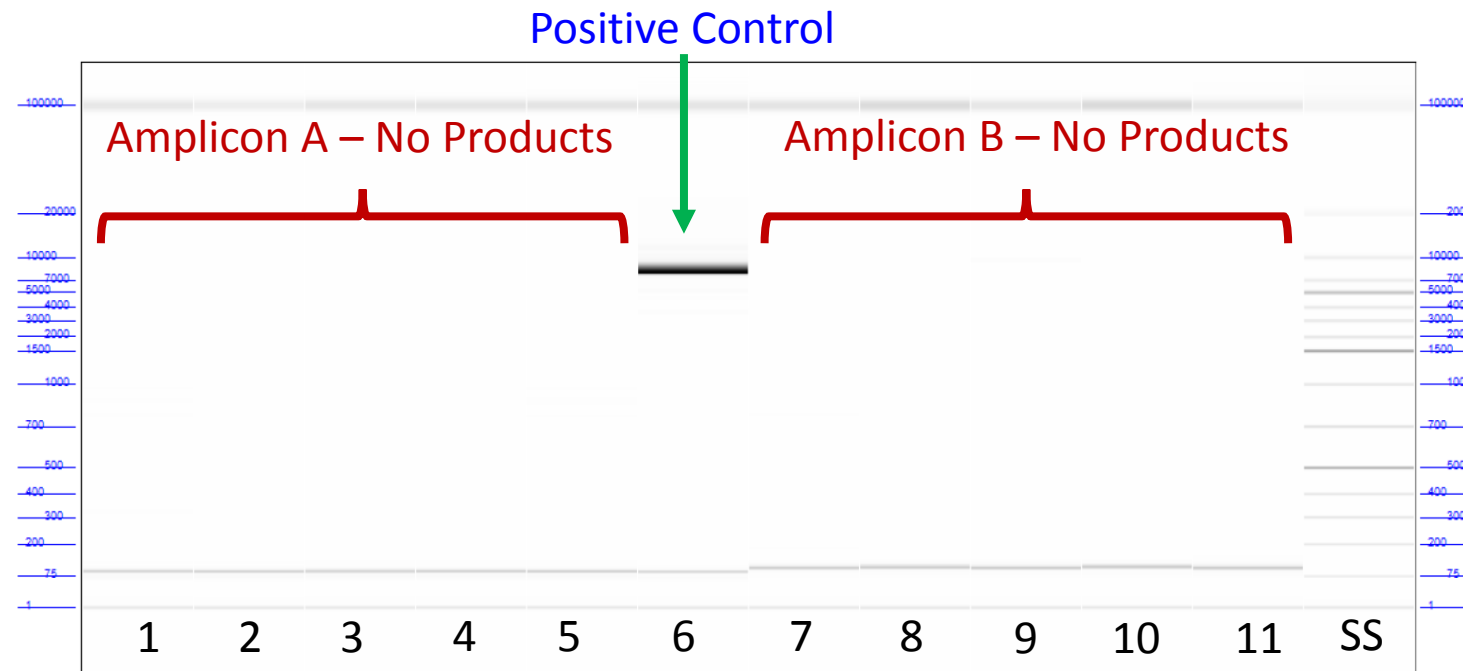


Mixture Sample (3:1) Matches Expected Ratio



Surprise! Some Degraded Samples

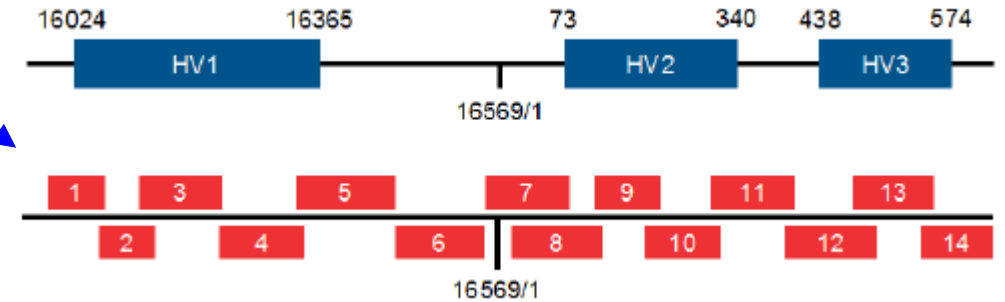
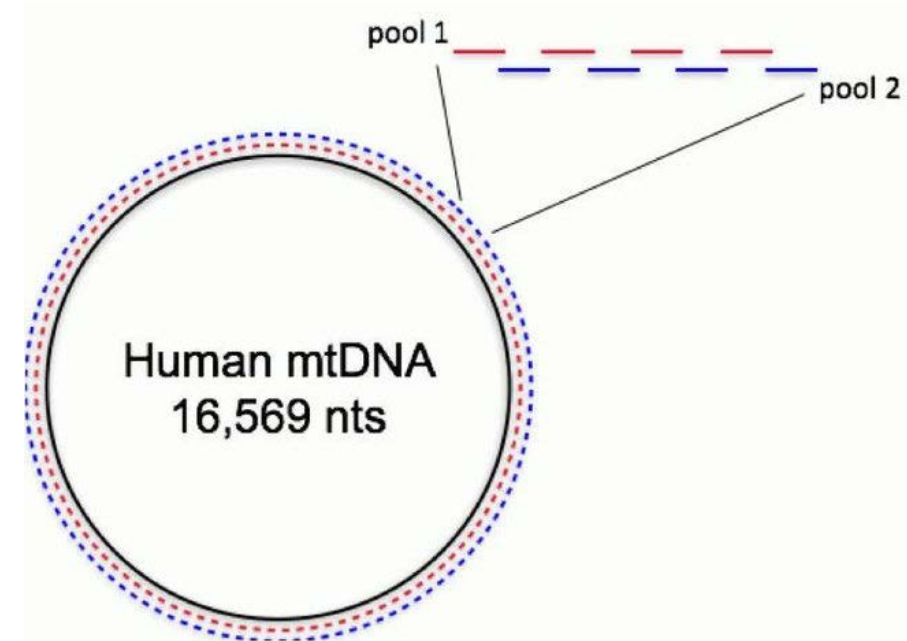
- Did not amplify with long PCR (≈ 8 kb)
 - Buccal swabs extracted ≈ 2005
 - Works great with STR multiplexes



SS = Size standard

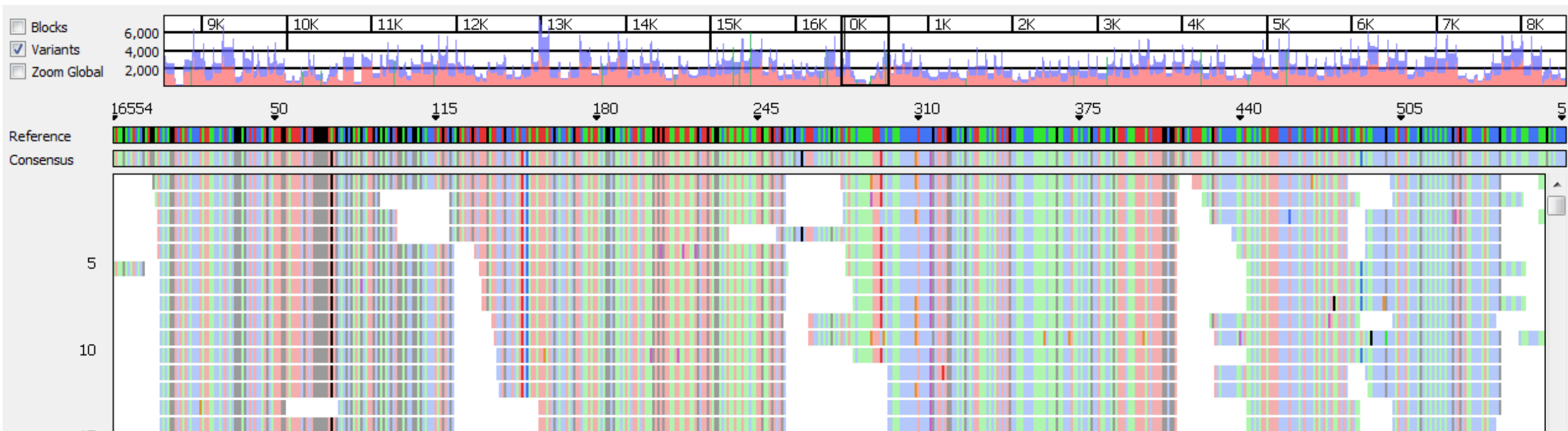
Thermo Fisher Precision ID Panels

- Two mtDNA products
 - Whole Genome Panel
 - Two primer pool multiplex
 - 81 primer pairs x 2 pools = 162 amplicons
 - \approx 160 bp amplicon size
 - Control Region Panel
 - Two primer pool multiplex
 - 7 primer pairs x 2 pools = 14 amplicons
 - 1.2 kb control region of mtGenome
 - \approx 150 bp amplicon size



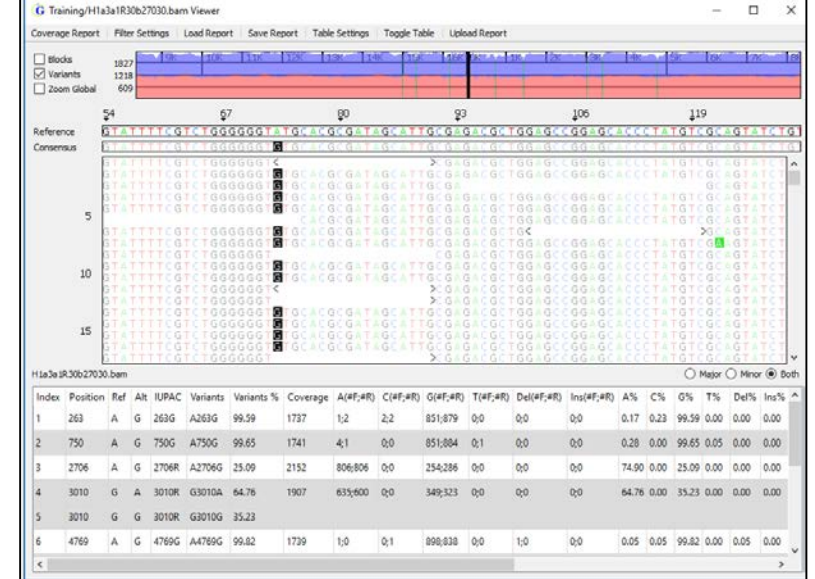
Sequencing Results – Whole Genome

- Full genome coverage
- All expected variants detected for SRM 2392 A & B, 2392-I
- Looking forward to using Thermo Fisher's analysis software



Informatics

- Forensic mtDNA nomenclature is challenging!
 - We should clone Dr. Parson
- Commercial software
 - Softgenetics GeneMarker HTS
 - Capability for forensic nomenclature
 - EMPOP compatibility
 - CLC Genomics Workbench
 - AFDIL / Qiagen – developed AQME Tool
 - ThermoFisher Scientific
 - Converge mtDNA Analysis (coming soon)



Forensic Science International: Genetics 31 (2017) 189–197

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

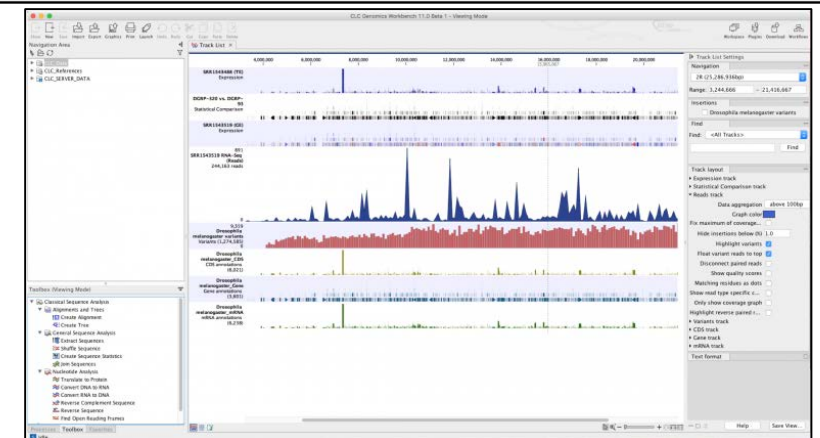
ELSEVIER

Short communication

AQME: A forensic mitochondrial DNA analysis tool for next-generation sequencing data

Kimberly Sturk-Andreaggi^{a,b,*}, Michelle A. Peck^{a,b}, Cecilie Boysen^{c,1}, Patrick Dekker^{c,2}, Timothy P. McMahon^a, Charla K. 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 ARF Sciences, LLC, Contractor Supporting the Armed Forces Medical Examiner System, 9210 Corporate Boulevard, Suite 120, Rockville, MD 20850, United States
^c QIAGEN Bioinformatics, Silkeborgvej 2, 8000 Aarhus C, Denmark



Phylogenetic Analysis – Additional QC Step

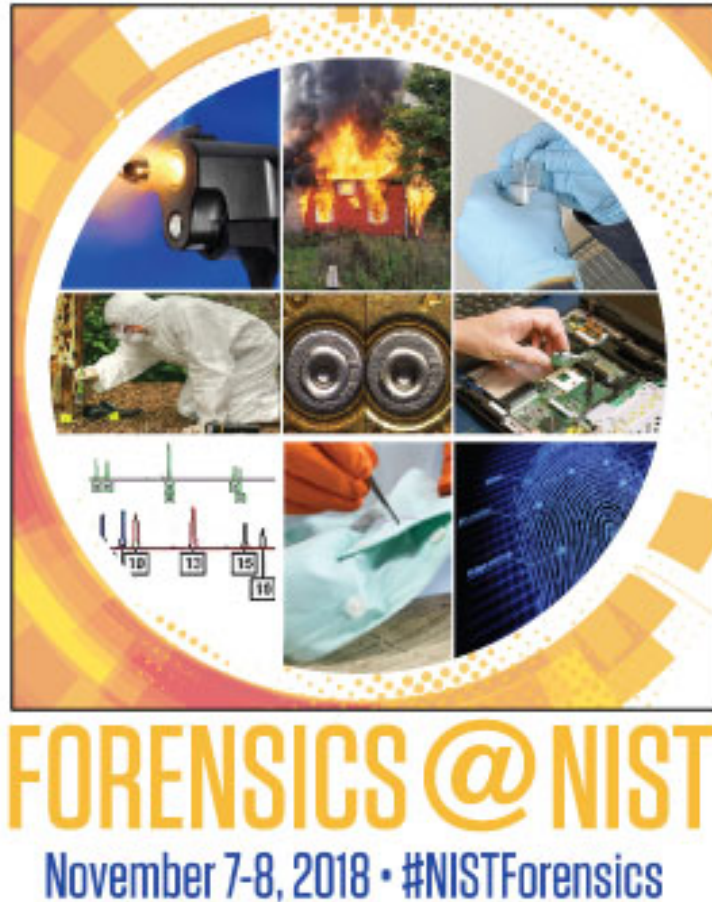
For high-throughput genotyping we will check these private mutation sites carefully.

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)

Conclusions

- Mitochondrial Genome Protocol
 - Multi-factorial decision process for selection
 - Selected AFDIL-developed procedure for reference quality samples
 - Even coverage
 - Higher multiplexing
 - Cost saving with $\frac{1}{2}$ reaction
 - Degraded samples will need a different procedure
 - Analysis method must be high-throughput
 - High accuracy required for EMPOP submission

Save the Date!



- November 7
 - Keynote: [John Butler](#)
 - [Forensic Genetics](#)
 - Fingerprints
 - Digital & Multimedia
 - Footwear Impression
- November 8
 - Keynote: Sheila Willis
 - Trace Evidence
 - Drugs and Toxins
 - Firearms and Tool Marks

Thank You! Questions?

Contact info:

Kevin.Kiesler@NIST.gov

A P P L I E D

G E N E T I C S

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

- Funding

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

√ Acknowledgements √

Armed Forces DNA Identification Laboratory (AFDIL)

Kim Andreaggi

Charla Marshall

Thermo Fisher Scientific

Matt Gabriel

NIST Applied Genetics Group NGS Team

Dr. Peter Vallone, Group leader

Lisa Borsuk

Sarah Riman

Becky Steffen

Katherine Gettings