Combined Extraction Method for Mitochondrial DNA and Proteins from Hair for Human Identification

How to get the most out of your sample

Craig Bartling, Ph.D.
Battelle Memorial Institute

When nuclear DNA is lacking, Alternatives are needed

Nuclear DNA (STRs/ SNPs) is the gold standard

When nuclear DNA is lacking, Alternatives are needed

Nuclear DNA (STRs/ SNPs) is the gold standard

mtDNA (SNPs, maternal lineage) is a good but limited alternative

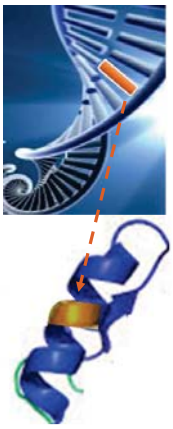
When nuclear DNA is lacking, Alternatives are needed

Nuclear DNA (STRs/ SNPs) is the gold standard

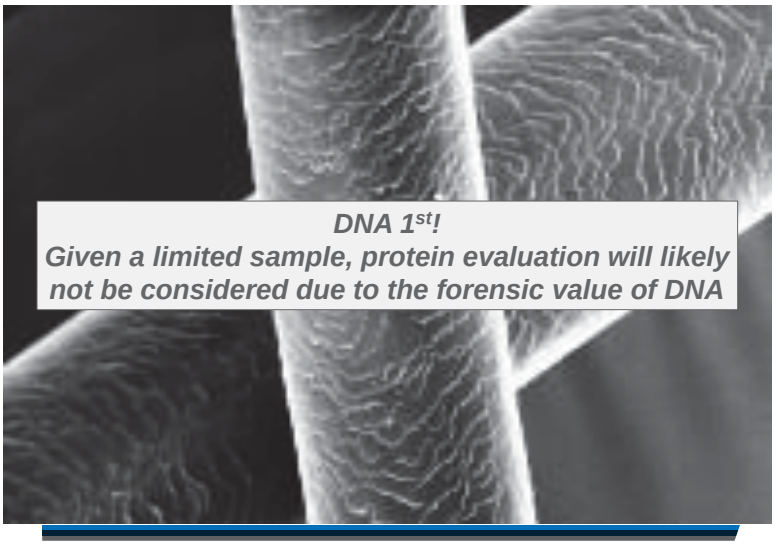
mtDNA (SNPs, maternal lineage) is a good but limited alternative

nsSNPs lead to SAPs/ GVPs

- **Discriminatory power: 1 in 12,500***
- ***Can mtDNA and protein signatures provide complete discrimination?***



*Parker et al 2016.



DNA 1st!
Given a limited sample, protein evaluation will likely not be considered due to the forensic value of DNA

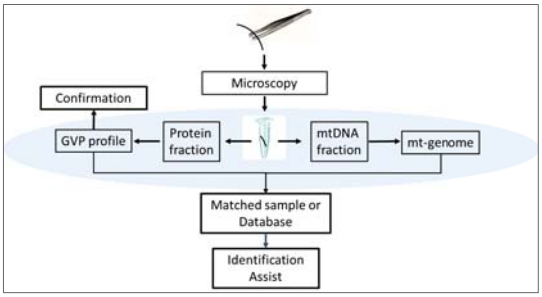
BATTELLE 5

Hair: DNA "needle" in a protein "haystack"

- *Hair is ~95% protein by mass*
- *Proteins can be more stable than DNA*
- *Nuclear DNA from hair is difficult to recover*
- *State of the art limited to mtDNA (limited) and microscopy (subjective)*

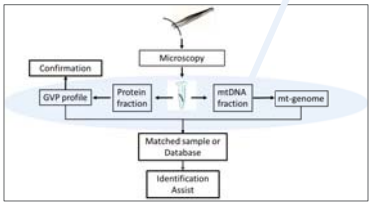
BATTELLE 6

Goal
Develop a workflow that enables the analysis of proteins from hair samples without sacrificing DNA analysis



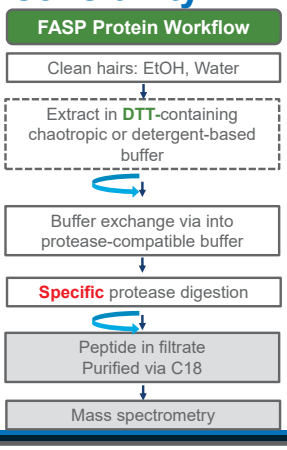
7 BATTELLE

1. *Optimize protein protocol for ease-of-use and sensitivity*
2. *Ensure compatibility with mtDNA forensic sample processing*
3. *Extract and analyze DNA and protein from the same sample*



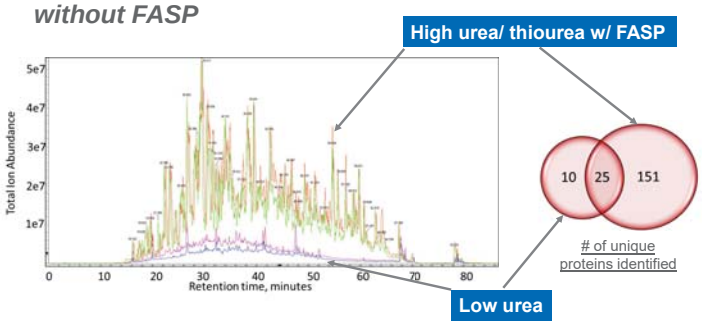
8 BATTELLE

1. Optimize protein protocol for ease-of-use and sensitivity



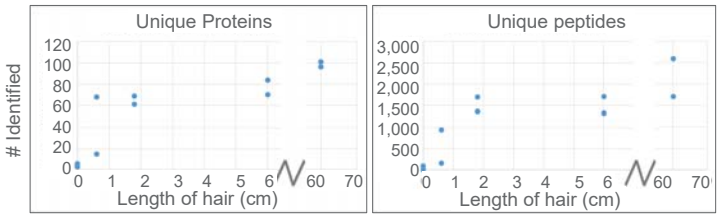
1. Optimize

- *High urea/ thiourea extraction buffer paired with FASP found to be superior to SDS/FASP or low urea/thiourea without FASP*



1. Optimize

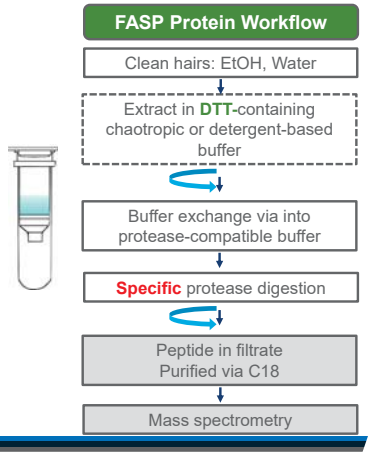
- *Rough working limit of >0.6 cm of hair established to identify unique peptides and proteins*



1. Optimize

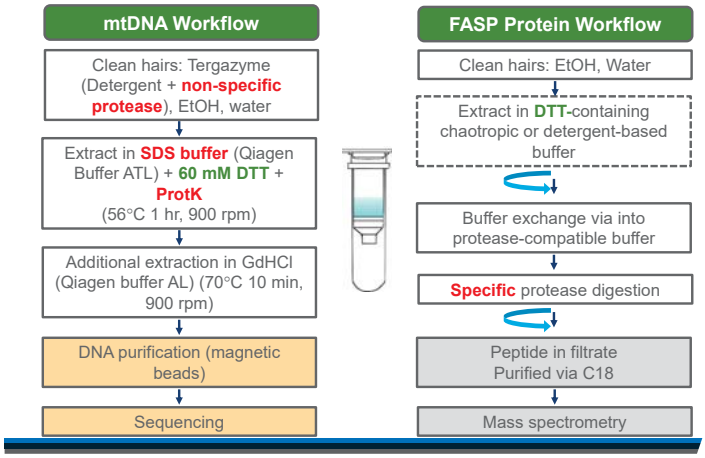
- *Rapid trypsin + lysC (45 minutes, 70 ºC) can be used in place of traditional trypsin + lysC (O/N, RT)*

Gold standard spin-column mtDNA workflows differ from spin-column protein workflows



BATTELLE 13

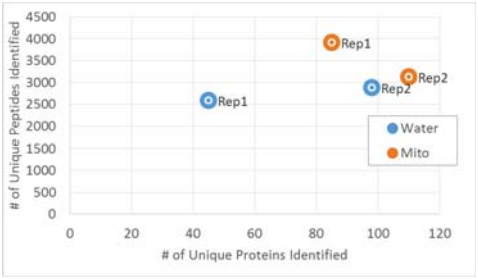
Gold standard spin-column mtDNA workflows differ from spin-column protein workflows



BATTELLE 14

2. Ensure compatibility with mtDNA forensic sample processing

- ***Rigorous washing protocols do not impact peptide yields***



15 BATTELLE

2. Ensure compatibility

- ***Rigorous washing procedures reduces background contamination***

Sample	# Unique Proteins	# Unique Peptides
EtOH, Water washing	8	50
LC/MS system blank	2	2
Tergazyme, EtOH, water	**2**	**2**

16 BATTELLE

3. Extract and analyze DNA and protein from the same sample: Truth Data

- ***Haplotypes and “Saplotypes” from 2 hair donors***

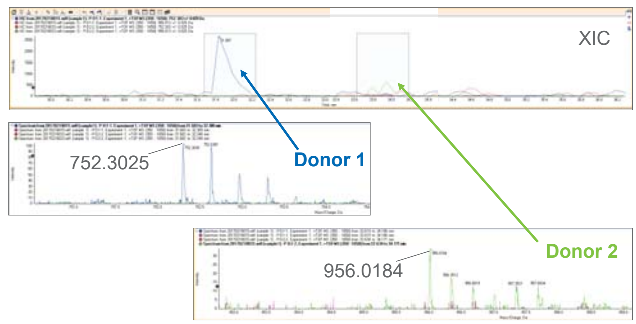
Position	Donor 1	Donor 2 (new)
	Haplotype	Haplotype
73	None	G
152	None	C
195	None	C
263	G	G
709	None	A
750	G	G
1438	G	G
1888	None	A
1896	C	N/A
2706	None	G
3010	A	None
3795	T	None
4216	None	C
4769	G	G
4917	None	G
7028	None	T
8697	None	A
8860	G	G
9899	None	C
10463	None	C
11251	None	G
11719	None	A
12633	None	A
13368	None	A
14766	None	T
14905	None	A
15047	A	None
15326	G	G
15452	None	A
15607	None	G
15928	None	A
16126	None	C
16163	None	G
16186	None	T
16189	C	C
16294	None	T
16519	C	C
Haplogroup	H1ab	T1a1

GN	peptide	Donor 1	Donor 2	Associated rs#
JUP	SAIVHLINYQDDAELAT[R/h]	SAIVHLINYQDDAELATR	Neither identified	rs41283425
JUP	NEGTATYAAA[V/i]LFR	NEGTATYAAAVLFR	Neither identified	rs143043662
KRT31	DN[A/v]ELENLIR	DNAELENLIR	DNAELENLIR	rs6503627
KRT31	SQYE[A/v]LVETNR	SQYEALVETNR	SQYEALVETNR	rs112544857
KRT31	QNQEYQ[V/m]LLDVR	QNQEYQVLLDVR	QNQEYQVLLDVR	rs146247884
KRT32	LEGEIN[T/m]YR	Both identified	Both identified	rs2071563
KRT32	CQYEAMVEAN[R/h]R	CQYEAMVEANRR	Neither identified	rs72830046
KRT33A	QVVSSSEQLQSYQ[A/v]EIIELR	QVVSSSEQLQSYQAEIIELR	QVVSSSEQLQSYQAEIIELR	rs12937519
KRT33B	T[L/V]NALEIELQAQHNLR	Both identified	Both identified	rs11448884
KRT34	SQYEALVE[I/t]NR	Both identified	Both identified	rs2239710
KRT35	VSAMYSSS[S/p]CKLPSLSPVAR	VSAMYSSSSCKLPSLSPVAR	VSAMYSSSSCKLPSLSPVAR	rs743686
KRT35	YETEVSLR	YETEVSLR	Neither identified	rs138303882
KRT35	TN[C/y]S[P/a]RPICVPCPGGRF	TNCSPRPICVPCPGGRF	TNCSPRPICVPCPGGRF	rs2071601, rs12451652
KRT81	LYEEEI[L/r]ILQSHISDTSVVVK	ILQSHISDTSVVVK	Both identified	rs6580873
KRT81	GLTGGFGSHSVC[G/r]GFR	GLTGGFGSHSVCGGFR	Both identified	rs2071588
KRT83	DLNMDC[I/m]VAEIK	DLNMDCmVAEIK	DLNMDCIVAEIK	rs2852464
KRT83	LEAAVAQSEQQ[G/s]EAALSDAR	Both identified	LEAAVAQSEQQGEAALSDAR	rs140635030
KRT83	L[L/Q]EGEE[H/Q]R	All variants identified	All variants identified	
KRT83	[V/I]LQSHISDTSVVVK	Both identified	Both identified	
KRT85	EAECVEA[D/n]SGR	EAECVEADSGR	EAECVEADSGR	rs112554450
KRT86	EEINELN[R/c]MIQR	EEINELNR	EEINELNR	rs139895699
KRTAP3-2	[S/g]CSVPTGPATTICSSDK	Both identified	Both identified	
S100A3	SLACLCLYC[H/q]EYFK	SLACLCLYCHEYFK	SLACLCLYCHEYFK	rs41265164
KRT86	EEINELN[R/c]MIQR	EEINELNR	EEINELNR	
KRT86	L[Q/p]FYQNR	LQFYQNR	LQFYQNR	rs58717266
KRT31	LNVEVDAAPTVDLN[Q/R]VLNETR	Both identified	Both identified	
KRT85	DVD[C/g]AYLR	DVDCAYLR	DVDCAYLR	rs61740813
KRT33B	DNAEL[E/k]NLIR	DNAELENLIR	DNAELENLIR	
KRT33B	LNVEVDAAPAVDLN[Q/R]VLNETR	LNVEVDAAPAVDLNQVLNETR	LNVEVDAAPAVDLNQVLNETR	
KRT33B	[D/Y]SLENTLTESEAR	Both identified	Both identified	
KRT33B	EVEQWFATQTE[E/k]LNK	EVEQWFATQTEELNK	EVEQWFATQTEELNK	rs12450621
KRT39	QNQEYEIL[L/m]DVK	QNQEYEILLDVK	QNQEYEILLDVK	rs17843023
KRT32	MVVN[I/t]DNAK	MVVNIDNAK	MVVNIDNAK	rs2071560
KRT32	LNIEVD[A/t]APPVDLTR	LNIEVDAAPPVDLTR	LNIEVDAAPPVDLTR	rs146792525
KRT32	[T/C]VNTLEIELQAQHSLR	TVNTLEIELQAQHSLR	TVNTLEIELQAQHSLR	
KRT32	DSLEN[T/m]LTESEAR	DSLENTLTESEAR	DSLENTLTESEAR	rs16966929
KRT32	QNQEYQVLLD[V/i]R	QNQEYQVLLDVR	QNQEYQVLLDVR	rs149524933
KRT82	ELDVD[G/s]IIAEIK	ELDVDGIIAEIK	ELDVDGIIAEIK	rs61730589
KRTAP2-1	TGSCCGSTLSSLSYGGGCCQ[P/r]CCCR	TGSCCGSTLSSLSYGGGCCQr	TGSCCGSTLSSLSYGGGCCQPCCCR	rs9635728
PKP1	AVQYLSSQDE[K/n]YQAIGAYYIQHTCF(	AVQYLSSQDEK	Neither identified	
LGALS3	I[Q/H]VLVEPDHFK	IQVLVEPDHFK	IQVLVEPDHFK	

GVP: Donor 1 versus Donor 2

- ***KRTAP3-2 SAP Evidence***

TGSCCGSTLSSLSYGGGCCQr: 752.3030 m/z
TGSCCGSTFSSLSYGGGCCQPCCCR: 956.0127 m/z

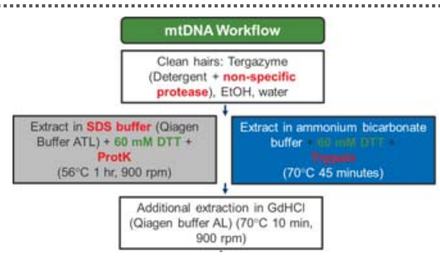


3. DNA and protein from One sample

- ***Can Rapid Trypsin can be substituted for Prot K in mtDNA workflow?***

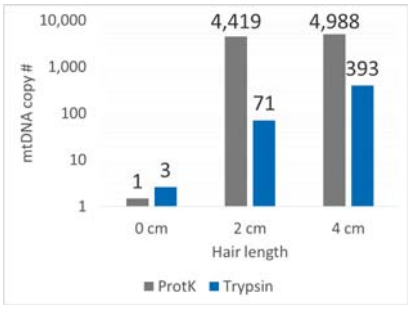
3. DNA and protein from One sample

- ***Can Rapid Trypsin can be substituted for Prot K in mtDNA workflow?***



3. DNA and protein from One sample

- *Rapid Trypsin can substitute for ProtK but mtDNA but yields are lower*



BATTELLE

3. DNA and protein from One sample

- *Mito profiles (Precision ID mtDNA whole genome panel) can be obtained from trypsinized hair*

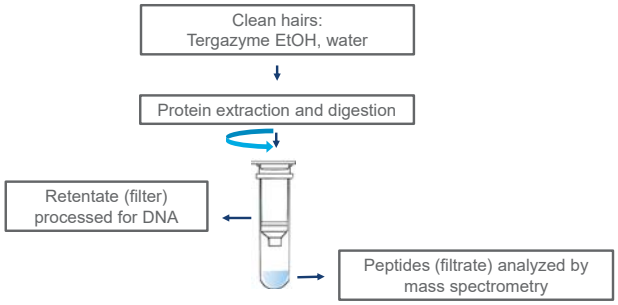
Sample	Donor	Enzyme	Hair length	# Mapped Reads	Haplogroup prediction* (Missing, private mutations)
BN1	#2	ProtK	4 cm	1,575,760	T1a1 (0, 2)
BN2	#2	ProtK	2 cm	1,767,023	T1a1 (0, 0)
BT1	#2	Trypsin	4 cm	1,454,222	T1a1 (0, 1)
BT2	#2	Trypsin	2 cm	1,135,412	T1a1 (0, 1)
B (saliva)	#2	ProtK	Saliva	1,738,314	**T1a1 (0, 1)**
2800M**	2800M	N/A	N/A	1,949,741	**H1c+T152C (0/0)**
NC	None	ProtK	N/A	54,491	--
TC	None	Trypsin	N/A	56,140	--

*https://empop.online/
**Corroborated by: http://forensic.yonsei.ac.kr/presentation/113.pdf ; https://www.promega.com/resources/profiles-in-dna/2016/analyzing-data-from-next-generation-sequencers-using-the-powerseq-automitoy-system/

BATTELLE

3. DNA and protein from One sample

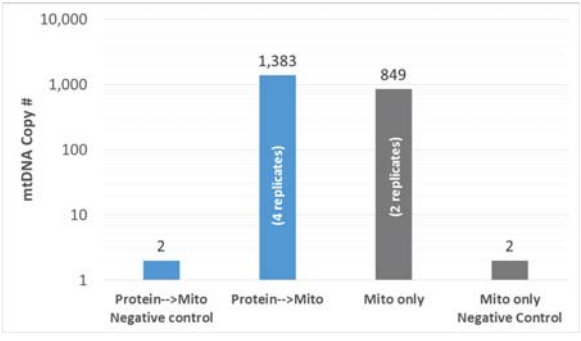
- *Can we extract proteins first then DNA without compromising DNA yields?*



BATTELLE

3. DNA and protein from One sample

- *mtDNA can be recovered following protein digestion*



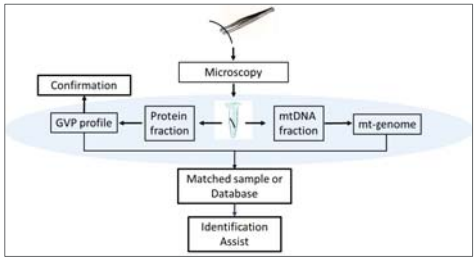
BATTELLE

Summary and Next Steps

- Summary
  - ✓Developed easy spin column-based format with rapid digestion
  - ✓Working limit approaching ~0.6 cm of hair
  - ✓Rigorous washing procedures reduces background contamination without impacting peptide yields
  - ✓Whole mito genome profiles can be obtained from trypsinized hair
  - ✓mtDNA can be recovered following protein digestion
- Next steps
  - Proteomic and mtDNA profiles of mito DNA following protein digestion
  - Sensitivity analysis, validation
  - Other relevant samples (e.g., touch)

BATTELLE

GVP Path to forensic laboratory implementation



BATTELLE

Acknowledgements

Rich Chou
Lindsay Catlin, M.S.
Larry Mullins
Mark Wilson, Ph.D.
Rich Guerrieri, M.S.

Funded by a Battelle Internal Research and Development Award

BATTELLE

BATTELLE

It can be done

Craig Bartling, Ph.D.
bartlingc@battelle.org
614-424-5377

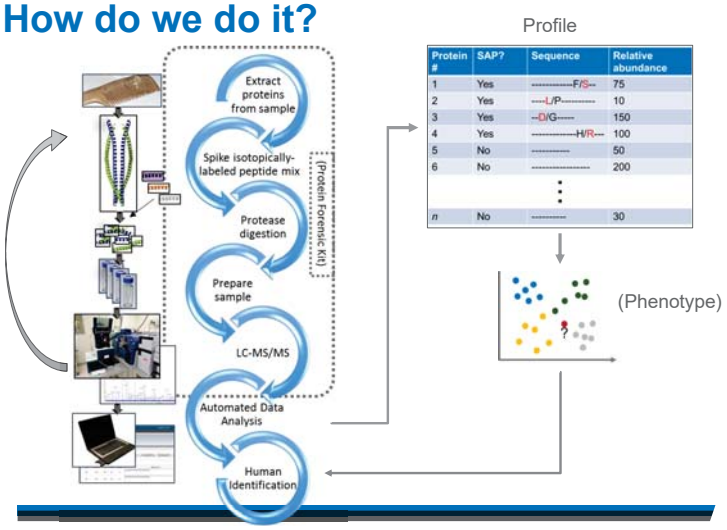
Backup slides

Can combined mtDNA-protein profiles be used to…

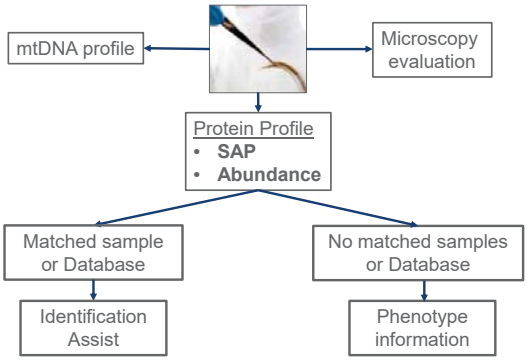
- *Discriminate one human from another?*
- *Distinguish ethnicity?*
- *Distinguish head hair from different body sites?*
- *Provide other identifying information?*
  - *Animal vs. human hair?*
  - *True hair color/ has hair been dyed/ bleached?*
  - *Exposure?*
  - *Other phenotypes?*

		Saliva				BN2				BT2		
Position		Variant	Frequency	Quality		Variant	Frequency	Quality		Variant	Frequency	Quality
73		G	100	37427.9		G	100	36669.1		G	100	36951.2
152		C	100	37697.9		C	100	37742.6		C	100	37515.7
195		C	100	37342.6		C	100	37484.8		C	100	37127.9
263		G	100	37769.1		G	100	37667.6		G	100	37051.6
311		C	99.9	14816.8		NO VARIANT CALLED				C	100	14878.4
709		A	100	37032.7		A	100	36486.8		A	95.1	32924.5
750		G	100	36453.5		G	100	37576.8		G	100	36047.4
1438		G	100	36388		G	100	37631.1		G	100	37600.1
1888		A	100	37666.1		A	96.3	34089.9		A	97.2	34928.3
2706		G	100	37382.9		G	100	35800.3		G	100	35649.7
4216		C	100	37181.3		C	100	35083.6		C	100	35475.2
4769		G	100	37866.8		G	100	37804.8		G	100	29467.4
4917		G	100	36673		G	100	36256.7		G	100	30627.2
7028		T	100	36871.2		T	100	36406.4		T	95.1	33025.5
8697		A	100	7671.85		A	100	9902.94		A	97.5	3479.31
8860		G	100	37747.9		G	100	37779.3		G	100	32373.7
9899		C	100	37660.2		C	100	35758.9		C	92.5	30794.8
10463		C	100	37243.4		C	100	29705.6		C	97	20907
11251		G	100	37007.7		G	100	35878		G	91.5	24316.8
11719		A	100	36965.1		A	100	36002.6		A	100	35998.9
12633		A	100	36970		A	100	36748.1		A	100	35306.8
13368		A	100	36800.1		A	100	35810.8		A	97	34718
14766		T	100	37534.2		T	100	35713.1		T	100	37662.9
14905		A	100	37729.3		A	100	36560.1		A	97.2	23896.4
15326		G	100	37592.8		G	100	37565.5		G	100	37748.2
15452		A	100	37060.4		A	100	34983.3		A	100	35915.8
15607		G	100	37223.4		G	100	36613.2		G	96.8	34545.7
15928		A	100	37052.9		A	100	36960.1		A	96.8	34572.4
16126		C	100	37177.5		C	100	37021.9		C	100	37271.4
16163		G	100	35235.6		G	100	35282.5		G	100	35870.7
16186		T	71.3	17318.3		T	71.8	17638.2		T	71.2	17309
16189		C	100	36132.2		C	100	36495.2		C	100	36907
16294		T	100	36586.2		T	100	36881.3		T	100	35892.7
16519		C	100	37431		C	100	37550.7		C	100	37492.4
16642		G	100	37660.1		G	100	36308		G	100	37645.8

How do we do it?

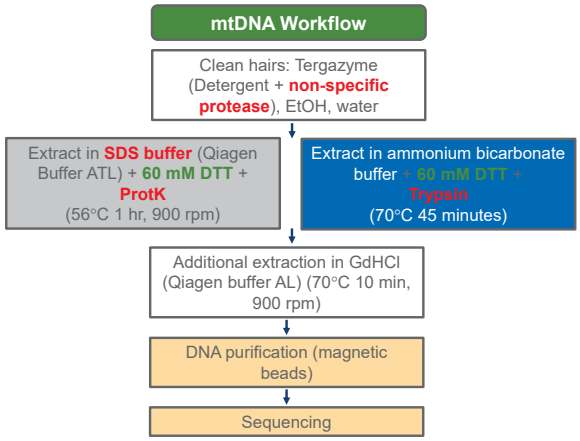


Using protein sequence *and* abundance provides unique and complementary information



Why success can be achieved: Key challenges can be overcome

- Sample preparation optimization
  - Currently investigating SDS vs urea-based extraction, detection limits, QC measures, degradation/ inhibition studies
  - Integrated sample processing
  - Mass spectrometry detection limits are improving
- Methodology can be standardized and independently verified in a manner similar to DNA-based forensics
  - Evaluated four NGS platforms, and based on sponsor criteria, down-selected one for implementing STR genotyping of human DNA samples
  - Currently evaluating a number of different RapidDNA and nanopore-based sequencing devices for field-based human identification applications



Step 3: Extract and analyze DNA and protein from the same sample

- *Mito profiles can be obtained from trypsinized hair*

Sample	Donor	Enzyme	Hair length	# Mapped Reads	Haplogroup prediction* (no mismatches)
BN1	#2	ProtK	4 cm	1,575,760	T1a1/ T1a1'3
BN2	#2	ProtK	2 cm	1,767,023	**T1a1/** T1a1'3
BT1	#2	Trypsin	4 cm	1,454,222	T1a1/ T1a1'3
BT2	#2	Trypsin	2 cm	1,135,412	T1a1/ T1a1'3
B (saliva)	#2	ProtK	Saliva	1,738,314	T1a1/ T1a1'3
2800M**	2800M	N/A	N/A	1,949,741	**H1c(T152C)**/ H1c/ H1(T152C)
NC	None	ProtK	N/A	54,491	H1c2/ H1c(T152C)/ H1c/ H1(T152C)/ H1(T16189C)
TC	None	Trypsin	N/A	56,140	H1c2/ H1aj/ H1c

*https://dna.jameslick.com/mthap/
**Corroborated by: http://forensic.yonsei.ac.kr/presentation/113.pdf ; https://www.promega.com/resources/profiles-in-dna/2016/analyzing-data-from-next-generation-sequencers-using-the-powerseq-automitoy-system/

Step 1: Optimize protein protocol for ease-of-use and sensitivity

- *Filter passivation*

BATTELLE