Green Mountain DNA Conference July 24, 2024

Long Range mtDNA Sequencing for mtDNA Mixture Deconvolution

Presenters

Brian Young and Scott Tighe

Brian Young[1,2,3], Sophie Hadley[4], Jeff Smith[1], Bhupinder Singh[1],
George Duncan[5], Scott Tighe[4]
[1]NicheVision Forensics, [2]Syracuse University, [3]Florida International University, [4]University of Vermont, [5]Nova Southeastern University

1

Overview

We present the use of two technologies not yet widely used in mitochondrial DNA analysis:

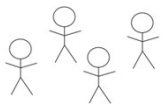
1. Long-range mtDNA PCR combined with Oxford Nanopore sequencing
2. Deconvolution of mixed mitochondrial samples

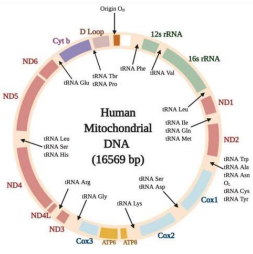
Through the combination of these technologies, we can derive mtDNA haplotypes for each individual contributor in a mixed sample.

2

Human Mitochondrial DNA 16,571 BP

Resolving mixed individuals with large overlapping sequences





3

PCR Amplification and ONT Sequencing

4

Multiplex Long Range PCR

PCR Amplifications of Mitochondria with 1500 bp overlap
- Takara Biosciences GXL master mix
- 6 Primer Pools 1500bp overlaps
- 9 cell line mixes

Long Range PCR Takara GXL PCR
- 25 ul GXL
- 6 ul primer multiplex
- 18 ul H20
- 1 ul DNA (1-3ng)

	Type	17257 (1:15)	17115 (1:50)	17066 (1:50)	Barcode
1	Single Source	2ul	—	—	1_single_source_17257
2	Single Source	—	2.8ul	—	2_single_source_17115
3	Single Source	—	—	2.2ul	3_single_source_17066
4	Mixed 3:1	3ul	1.4ul	—	4_mixed_3:1_17257v17115
5	Mixed 3:1	3ul	—	1.1u	5_mixed_3:1_17257v17066
6	Mixed 3:1	—	4.2ul	1.1ul	6_mixed_3:1_17115v17066
7	Mixed 1:3	1ul	4.2ul	—	7_mixed_1:3_17257v17115
8	Mixed 1:1	1ul	1.4ul	—	8_mixed_1:1_17257v17115
9	Mixed 10:1	10ul	1.4ul	—	9_mixed_10:1_17257v17115

Oxford Nanopore Sequencing
- Ligation NDB114 barcoded Library
- Multiplex PCR Product
- GridION with R10.4 flow cells

-

The goal is even PCR amplification efficiency for effective secondary analysis and sequencing

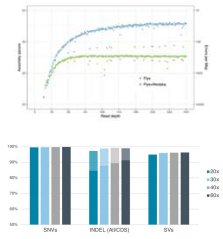
5

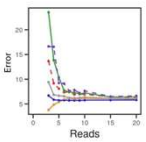
Nanopore Sequencing

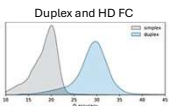
- 40-100 Gb per flow cell P2
- Detect base modifications on native DNA
- Reads up to 2 megabases
  - Amplicon
  - Native DNA
  - DNA base modification detection
  - Direct RNA sequencing
  - RNA base modifications
  - Low input PCR barcoding
  - Short read sequencing
  - Adaptive Sampling Sequencing

6

Increased Accuracy at 40x Coverage and New Duplex chemistry HD







MacKenzie M, Argyropoulos C. An Introduction to Nanopore Sequencing: Past, Present, and Future Considerations. Micromachines (Basel). 2023 Feb 16;14(2):459. doi: 10.3390/mi14020459. PMID: 36838159; PMCID: PMC9966803.

7

Can sequence just about anything…

8

NicheVision
Danakil Depression-
Ethiopia Toxic Saline
Hot Springs

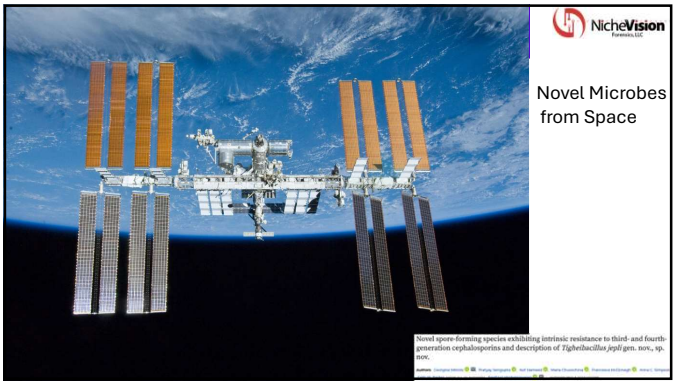
9

NicheVision
Limestone
Caves
Romanian
ISME
New Biofilm microbes
Cave Thiovulum (Candidatus Thiovulum stygium) differs
metabolically and genomically from marine species

10

NicheVision
Exoplanetary
Ecosystems
Evolution-Origins of Life
Large sulfur oxidizing bacteria of the
Thiovulaceae (Campylobacterota) thriving in the
sulfidic groundwater of Movile Cave, in Romania

11


NicheVision
Novel Microbes
from Space
Novel spore-forming species exhibiting intrinsic resistance to third- and fourth-
generation cephalosporins and description of Tigheibacillus jepli gen. nov., sp.
nov.

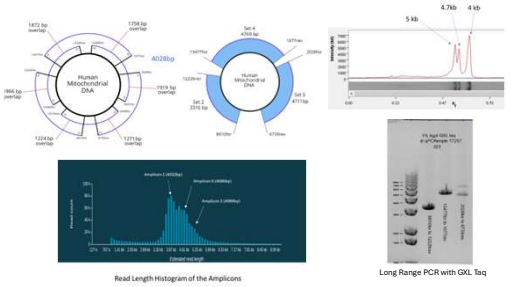
12

Unknown Infections

Organism	Read Count
Escherichia coli	11314 (42.07%)
Klebsiella pneumoniae	2163 (8.04%)
Enterococcus faecium	2056 (7.65%)
Enterococcus casseliflavus	1977 (7.35%)
Morganella morganii	1673 (6.22%)
Enterococcus sp. 6D12_DIV0197	816 (3.03%)
Enterococcus faecalis	469 (1.74%)
Enterococcus sp. 306_DIV0642	279 (1.04%)
Clostridium sp. CAG:299	264 (0.98%)
Enterococcus innesii	255 (0.95%)
Klebsiella aerogenes	223 (0.83%)
Klebsiella oxytoca	134 (0.5%)
Clostridium perfringens	133 (0.49%)
Morganella sp.	130 (0.48%)

13

Long Amplicon Mitochondrial Reassembly



Read Length Histogram of the Amplicons

Long Range PCR with GXL Taq

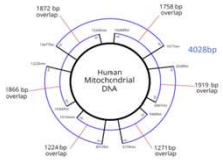
Brian Young NicheVision

14

PCR of Mitochondria

Long Range PCR with Takara GXL taq

- Set 1: 16488for to 3947rev
- Set 2: 8910for to 12226rev
- Set 3: 10360for to 15349rev
- Set 4: 13477for to 1677rev
- Set 5: 2028for to 6739rev
- Set 6: 5468for to 10154rev



Primers : Amanda Ramos, Cristina Santos, Luis Alvarez, Ramon Nogue, Maria Pilar Aluja. Human mitochondrial DNA complete amplification and sequencing: A new validated primer set that prevents nuclear DNA sequences of mitochondrial origin co-amplification. Electrophoresis 2009, 30, 1587–1593.

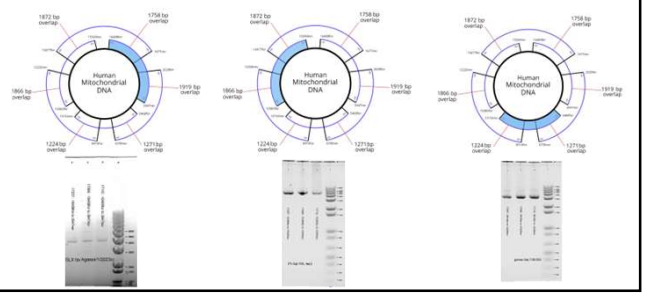
15

Run 1 Multiplex

Primer Set 1
4028bp

Primer Set 3
4989bp

Primer Set 6
4686bp



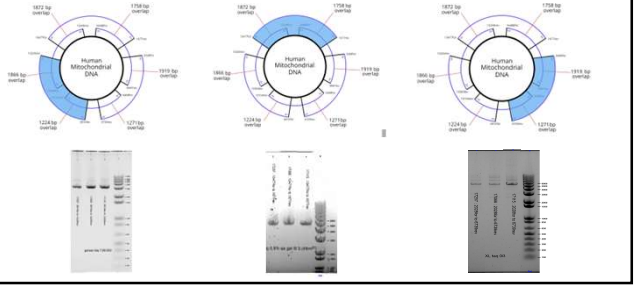
16

Run 2 Multiplex

Primer Set 2
3316bp

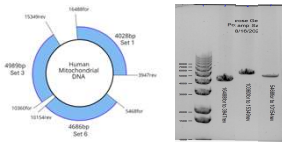
Primer Set 4
4769bp

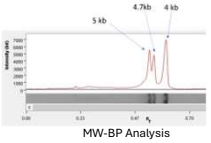
Primer Set 5
4711bp



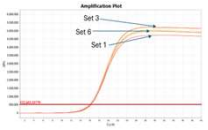
17

Validation of mixed amplicon pool by RTqPCR and ddPCR





MW-BP Analysis

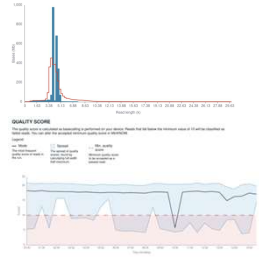


RTqPCR

18

Nanopore Results

READ LENGTHS - OUTLIERS REMOVED



QUALITY SCORE

Run summary

DATA OUTPUT

Estimated bases	Reads generated
3 Gb	690.89 k
Estimated N50	Total data produced (pass / fail)
4.21 kb	35.73 GB

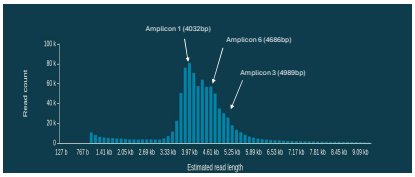
BASECALLING

Reads called	Bases called (min Q score: 10)	
100%	2.3 Gb	649.59 Mb
	Pass	Fail

19

Nanopore Read Length Results

*Interchangeable results from both runs

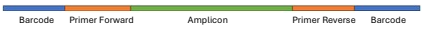


Read Length Histogram of the Amplicons

20

Nanopore Read Depth by Sample – Run 2

In-Vitro Mixture	Type	Nanopore Barcode	Total Bases (Mb)	Total Reads	Barcode label
1	Single Source	1	383.66	89, 680	1_single_source_17257
2	Single Source	2	328.94	78, 950	2_single_source_17115
3	Single Source	3	276.29	64,490	3_single_source_17066
4	Mixed 3:1	4	440.1	103,850	4_mixed_3:1_17257v17115
5	Mixed 3:1	5	411.09	96,800	5_mixed_3:1_17257v17066
6	Mixed 3:1	6	451.16	107,050	6_mixed_3:1_17115v17066
7	Mixed 1:3	7	435.01	103,130	7_mixed_1:3_17257v17115
8	Mixed 1:1	8	409.63	97,250	8_mixed_1:1_17257v17115
9	Mixed 10:1	9	429.52	100,750	9_mixed_10:1_17257v17115



21

Total Nanopore Output

Flow cells: 2

Total output: 6.92 Gbases

Total number of reads (> 1 Kb): 1.59 x 10^6

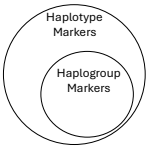
22

Mixed mtDNA Deconvolution and Individual Haplotyping

23

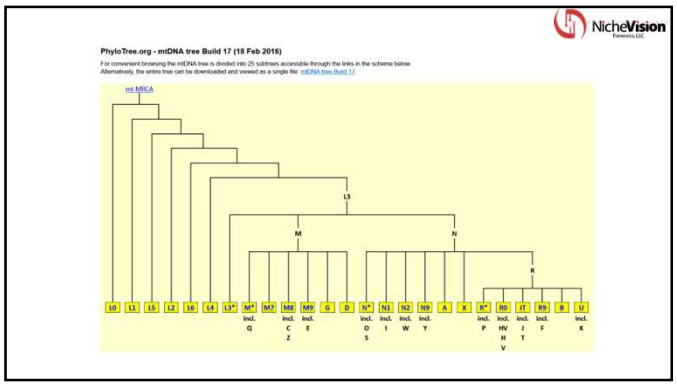
Definitions

- **Haplogroup:** a group of similar mtDNA sequences that share a common ancestor.
- **Haplotype:** the unique genetic sequence of an individual's mtDNA (i.e., their unique maternal genetic profile).



https://en.wikipedia.org/wiki/Human_mitochondrial_DNA_haplogroup

24



25

Current mtDNA Analysis Workflow

- Current practice for mtDNA haplotyping requires single-source samples due to inability to deconvolute mixtures.
- Being able to deconvolute mtDNA mixtures will enable exploitation of mtDNA from crime scenes.

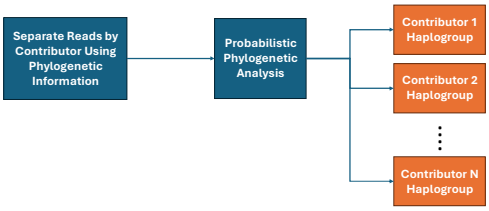
- Single hair
- Single bone/tooth

26

Workflow for Mixed mtDNA Samples >Classification of Reads by Maximum Likelihood

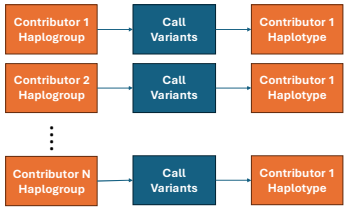
27

Workflow for Mixed mtDNA Samples >Haplogrouping by Phylogenetic Analysis



28

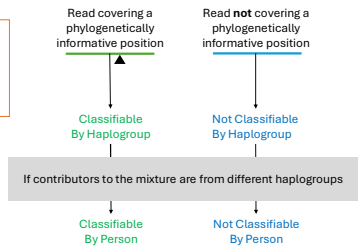
Workflow for Mixed mtDNA Samples >Variant Calling



29

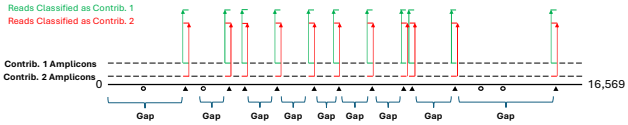
An Information Gap

- Not all short reads can be classified by contributor leaving large gaps in chromosome coverage.
- Hence, the isolated haplotypes are incomplete.



30

Illustration of Deconvolution Using Short Reads >Many Reads Cannot be Classified

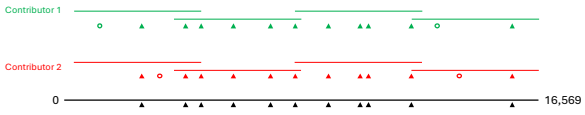


▲ Denotes a diagnostic SNP that can be used to classify a read by contributor.

○ Denotes a SNP that is not "diagnostic" but part of the haplotype of an individual.

31

Illustration of Deconvolution Using Long Reads >All Reads are Classifiable



32

Samples Used for Creating Mixtures

Accession	Coriell Population	Haplogroup (Estimated by MixtureAceMT)	Expected Haplotype[1,2]	
			Tree Span	Variant Calls
NA17257	Caucasian	U5b2b1b	U5 -> Leaf	73G 150T 263G 309.1C 315.1C 750G 1438G 12917 2706G 3197C 4769G 5460A 7028T 7768G 8860G 9477A 10664C 11487G 11963G 11719A 11823G 12309G 13372A 13634G 13663T 13417C 13639G 13827G 14182C 14764T 15229G 15487A 16192T 16270T 16526A
NA17115	African American	L1b1a3	L1b -> Leaf	73G 152C 182T 195T 199G 199C 247A 263G 309.1C 315.1C 357G 523- 524- 709A 719C 750G 769A 825A 1018A 1709C 2352C 2706G 2758A 3746G 3828C 3956C 3964T 3968A 4080A 4529G 4716G 4769G 5036G 5046A 5300C 5655C 5749T 5827C 5969G 7028T 7055G 7146G 7256T 7389C 7521A 7827T 8248G 8468T 8655T 8750G 8860G 9540C 10366G 10668A 10815C 11673C 11775A 12574C 12765T 13105G 13506T 13650T 13798C 13860A 13958A 14174C 14203G 14580A 14764T 14766G 15115C 15320G 16126C 16182C 16187T 16189C 16223T 16264T 16270T 16274T 16293G 16311C 16519C
NA17066	Central American	C1c6	C -> Leaf	73G 249- 263G 295- 301G 309.1C 315.1C 489C 750G 1438G 1888A 2706G 3592A 3698A 4715G 4769G 7028T 7196A 8584A 8701G 8860G 9230C 9540C 9599G 10398G 10600T 10673C 11775A 11914A 12414C 12705T 13105G 13263G 14316C 14764T 14783C 15043A 15301A 15320G 15487T 15808A 16153A 16223T 16298C 16325C 16327T

[1]Sources: Juat R, Parson W, Irwin J. 2019 Maximizing mtDNA Testing Potential with the Generation of High-Quality mtGenome Reference Data. NIJ Report 2011-MU-MU-K402; familytreedna.com/mtdna-haplogroup-mutations.aspx; Phylotree.com
[2]Required mutations defining the outer branches of the phylogenetic tree are highlighted in green.

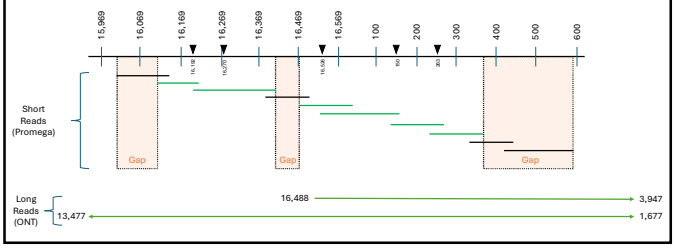
33

Sample	Maximum Likelihood Estimate				Number of Reads Classified	Total Reads	Percent Classified
	Haplogroup		Proportion				
	Expected	Called	Expected	Called			
CAU	--	U5b2b1b	1.00	1.00	25,991	89,680	29%
AA	--	L1b1a3	1.00	1.00	17,518	78,950	22%
MX	--	C1c6	1.00	1.00	19,332	64,490	30%
3:1 CAU:AA	U5b2b1b L1b1a3	U5b2b1b L1b1a3	0.75 0.25	0.82 0.18	21,953 5,096	103,850	26%
3:1 CAU:MX	U5b2b1b C1c6	U5b2b1b --	0.75 0.25	1.00 --	26,898 --	96,800	28%
3:1 AA:MX	L1b1a3 C1c6	L1b1a3 C1c6	0.75 0.25	0.61 0.39	14,536 9,418	107,050	22%
1:3 CAU:AA	U5b2b1b L1b1a3	U5b2b1b L1b1a3	0.25 0.75	0.27 0.73	17,013 6,549	103,130	23%
1:1 CAU:AA	U5b2b1b L1b1a3	U5b2b1b L1b1a3	0.50 0.50	0.56 0.44	14,095 11,214	97,250	26%
10:1 CAU:AA	U5b2b1b L1b1a3	U5b2b1b L1b1a3	0.91 0.09	0.94 0.06	27,255 1,868	100,750	29%

34

Illustration of Information Gains in the mtDNA CR Using Short or Long Reads

▼ = Diagnostic SNP for haplogroup U5b2b1b



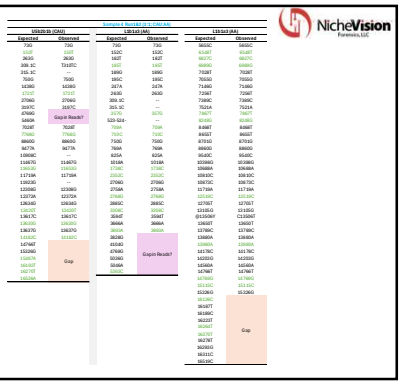
35

Haplotype Analysis

36

3:1 mixture
NA17257: NA17115
CAU:AA
Haplogroup-informative SNP calls Are Concordant With Expected

Additional private variant calls that define the haplotype are not shown.



37

Conclusions

- Long range sequencing by Oxford Nanopore Technologies (ONT) combined with probabilistic deconvolution holds promise for mtDNA analysis of mixed samples.
- Here, the mtDNA genome was covered by 6 amplicons to accommodate slightly degraded DNA. However, entire mtDNA chromosomes can possibly be sequenced. If done on mixed samples, the mixture will be "deconvoluted by sequencing", and the haplotypes will be "phased by sequencing".

38

Long Range mtDNA Sequencing for mtDNA Mixture Deconvolution

Brian Young[1,2,3], Sophia Hadley[4], Jeff Smith[1], Bhupinder Singh[1], George Duncan[5], Scott Tighe[4]

[1]NicheVision Forensics, [2]Syracuse University, [3]Florida International University, [4]University of Vermont, [5]Nova Southeastern University

Green Mountain DNA Conference 2024

Brian@nichevision.com
Scott.Tighe@uvm.edu

39