Bioinformatics for 9/11

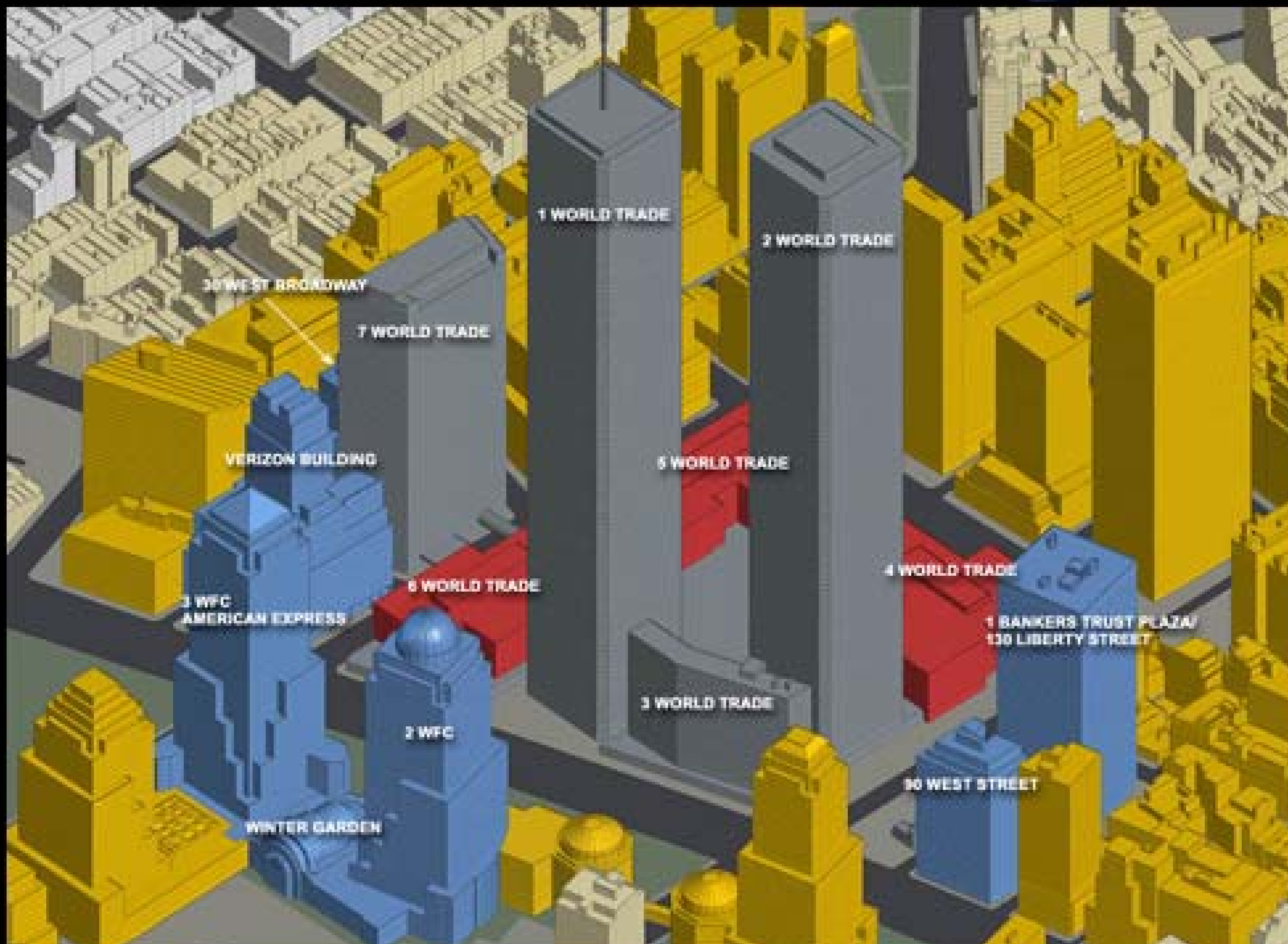
Simon Mercer
Director of Software Engineering
Gene Codes Corporation

World Trade Center

- Two 110 story towers
- 50,000 workers average week day
- Transportation hub - four subway lines
- Six basement levels – stores and plaza

- Collapse estimated at 130mph
- 24,000 gallons of jet fuel
- Fires burned at 1400 Fahrenheit for over three months
- 2bn pounds of rubble
- 2,749 victims, 1549 identified (Feb 2004)

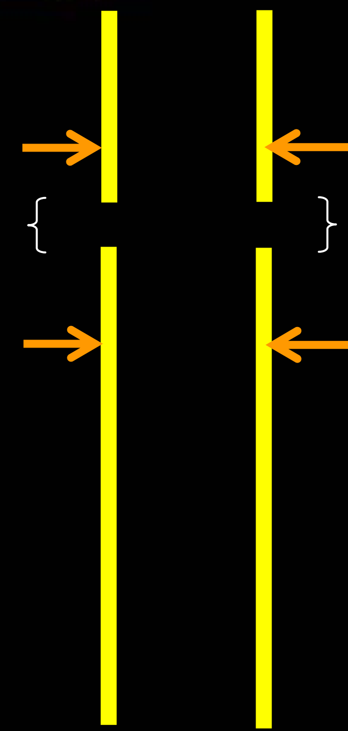
TCAG GENE
AGT CODES
FORENSICS

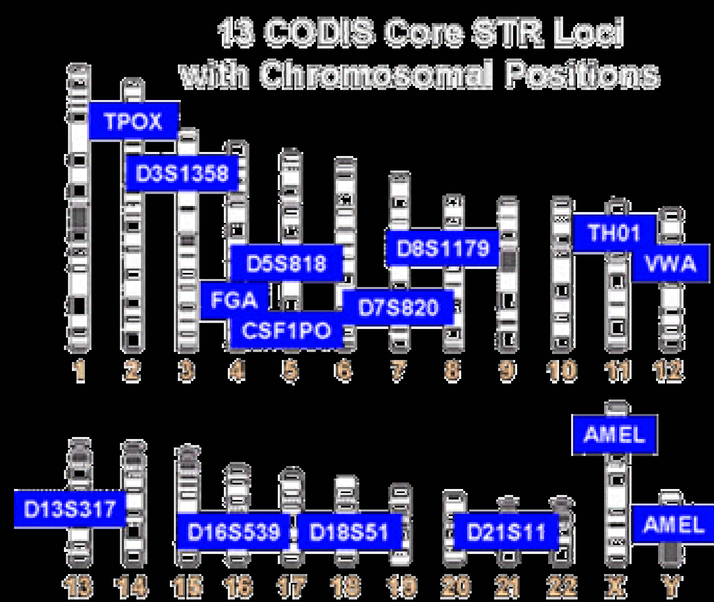


Human DNA Identification

Human Identification using DNA

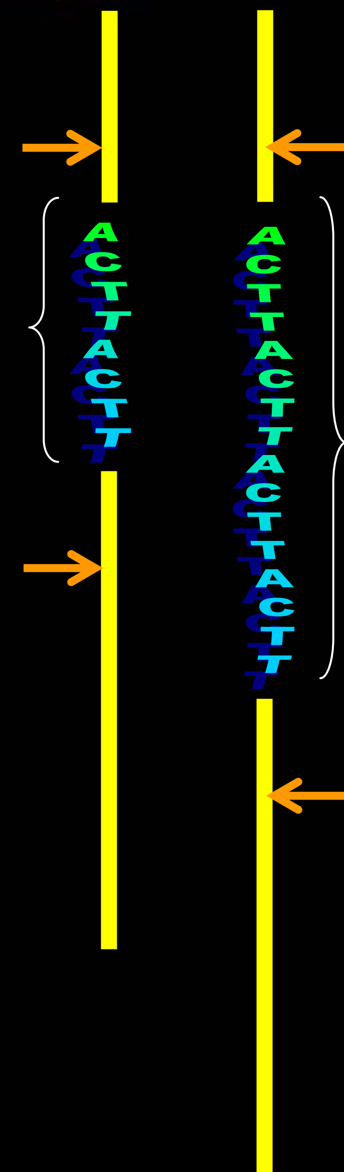
- Simple Tandem Repeats (STRs)
- 13 core loci where the lengths vary based on the number of STRs

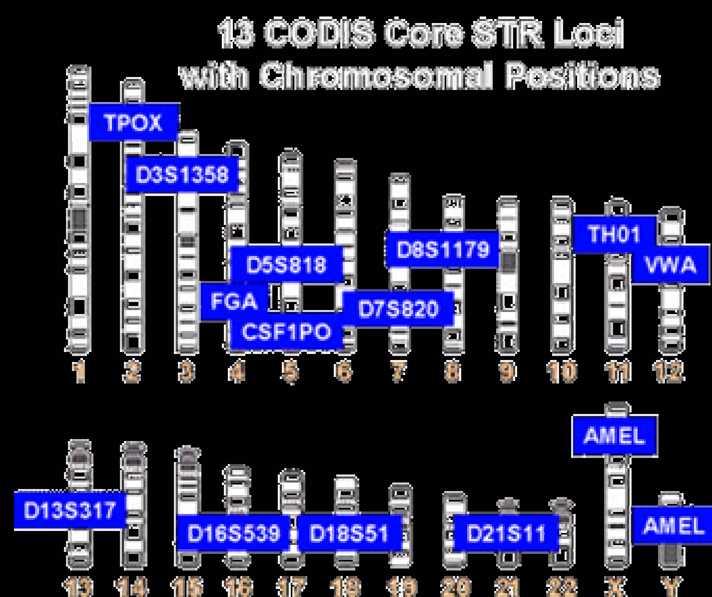




Human Identification using DNA

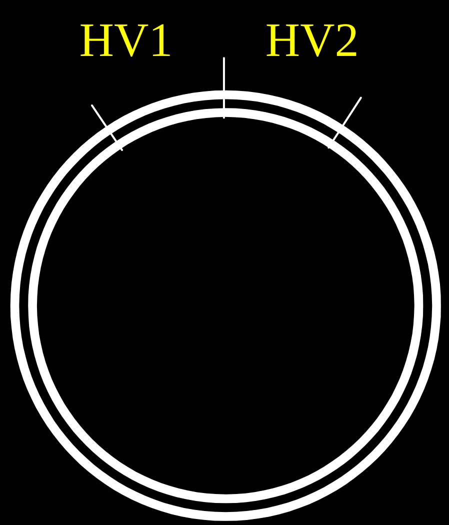
- Most people think of “Fingerprinting”
- 13 core loci where the lengths vary based on the number of STRs





Human Identification using DNA

- Mitochondrial DNA
- Matrilineal inheritance
- Can also sequence two, *hypervariable* regions on the mitochondrial genome
  - Not as specific as STRs but likely to be more available



Human Identification using DNA

- Single Nucleotide Polymorphisms (SNPs)
- Approx. 4M SNPs in Human nuclear DNA
- Panel of 70 SNPs from Orchid Cellmark
- Not (yet) validated for forensic casework

TC

TT

Kinship Analysis calculates the likelihood ratio that indicates if two people are related. In this example, as Full Siblings in the Hispanic population. (attr. Dr. George Carmody)

26-Oct-01 | KinTest© - CODIS 13 Core Loci

ID #1: | ID #2:

R. O. :

Full sibling

Test

		1	2	1	2	Fsib
1	D3S1358	14	16	16	15	0.72
2	vWA	14	13	15	13	10.40
3	FGA	18	26	18	26	133.10
4	D8S1179	12	18	18	18	20.55
5	D21S11	36	28	28	26	2.06
6	D18S51	22	22	18	17	0.25
7	D5S818	10	12	11	12	0.68
8	D13S317	10	10	12	10	2.73
9	D7S820	9	7	9	12	2.86
10	D16S539	10	12	12	12	1.12
11	THO1	8	6	6	7	0.79
12	TPOX	12	12	12	12	34.33
13	CSF1PO	12	10	10	10	1.24

Total: 2.1E+06

Fsib
Hsib
Pchild
Cousin

AfAm
Cauc
Hisp
Xxxx

Population
Hisp

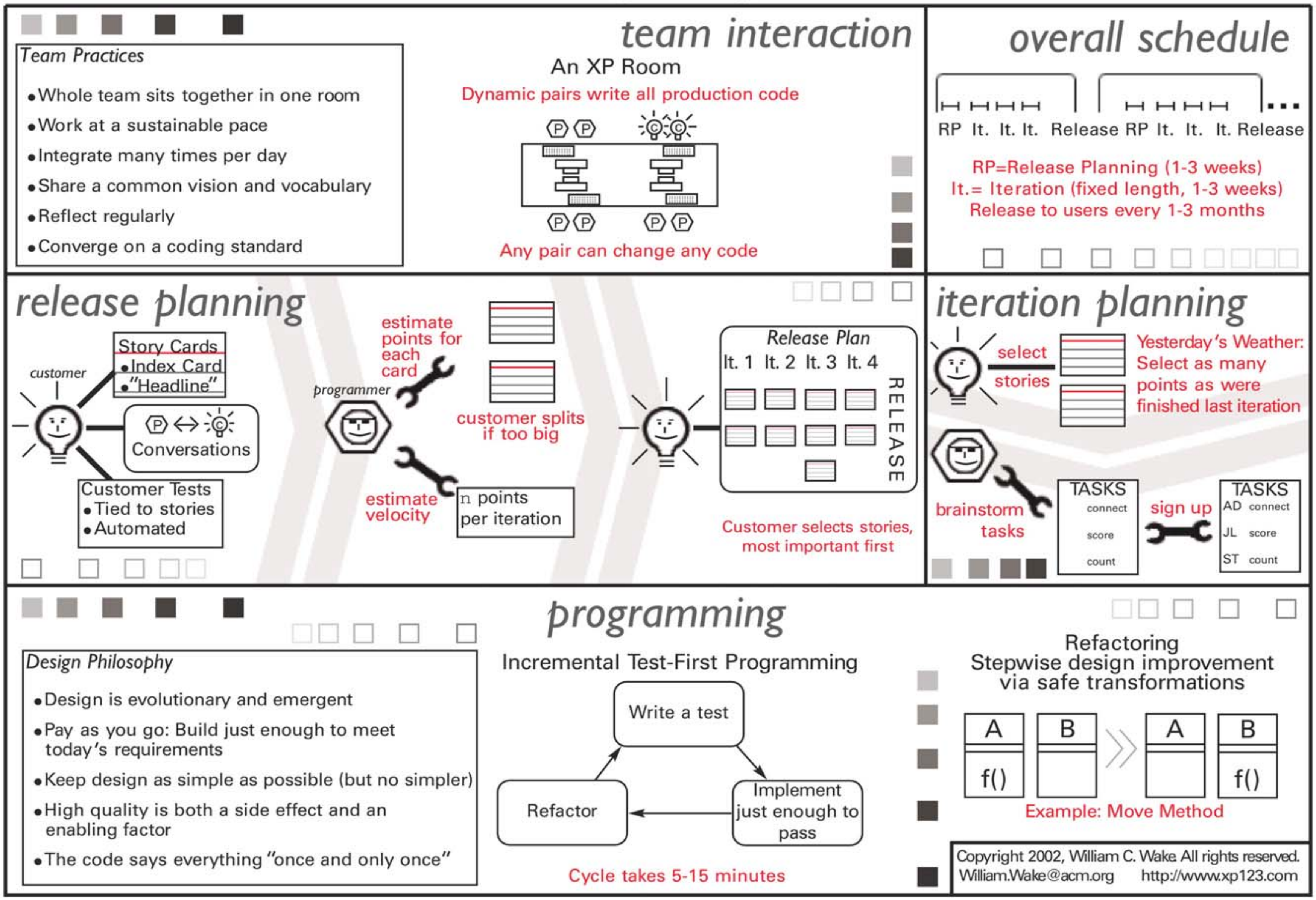
Hispanic pop stats

- Print -
Erase
STR 1 STR 2

Probability that these are full sibs, divided by probability that two unrelated people in this population would share the alleles

Extreme Programming

Extreme Programming Overview

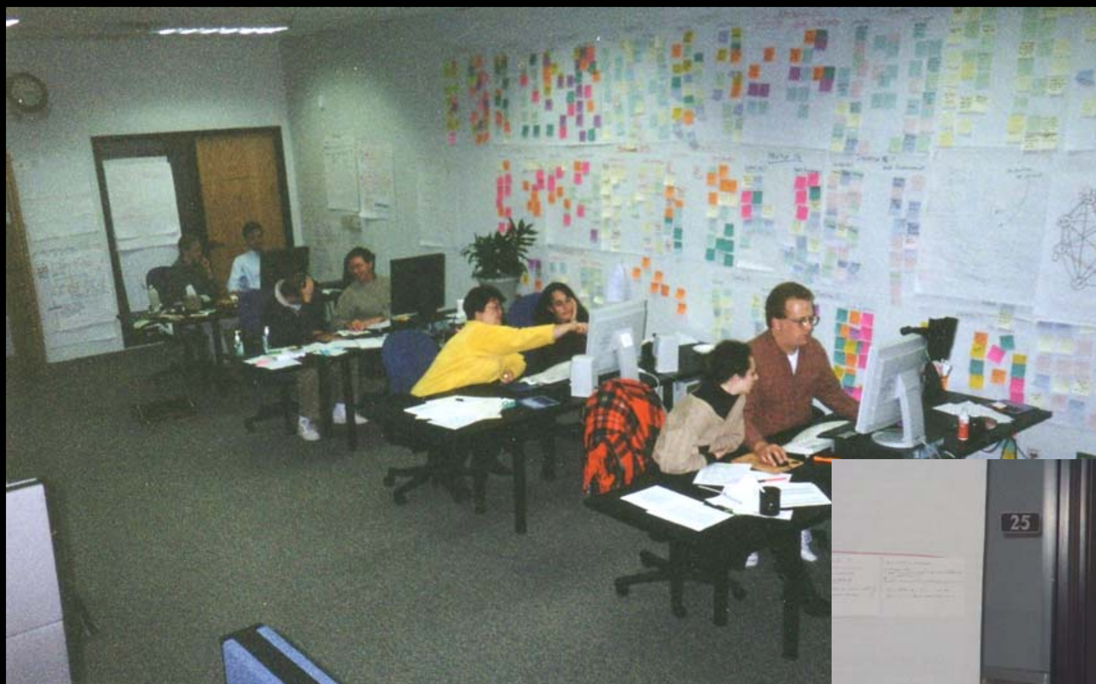


XP Timeline

- Coding started – November 5th 2001
- First demo to OCME – December 6th 2001
- First release delivered – December 13th 2001
- 78 releases since then (to Feb 2004)

- 2500+ automated tests run on every code change
- 107 code development cycles

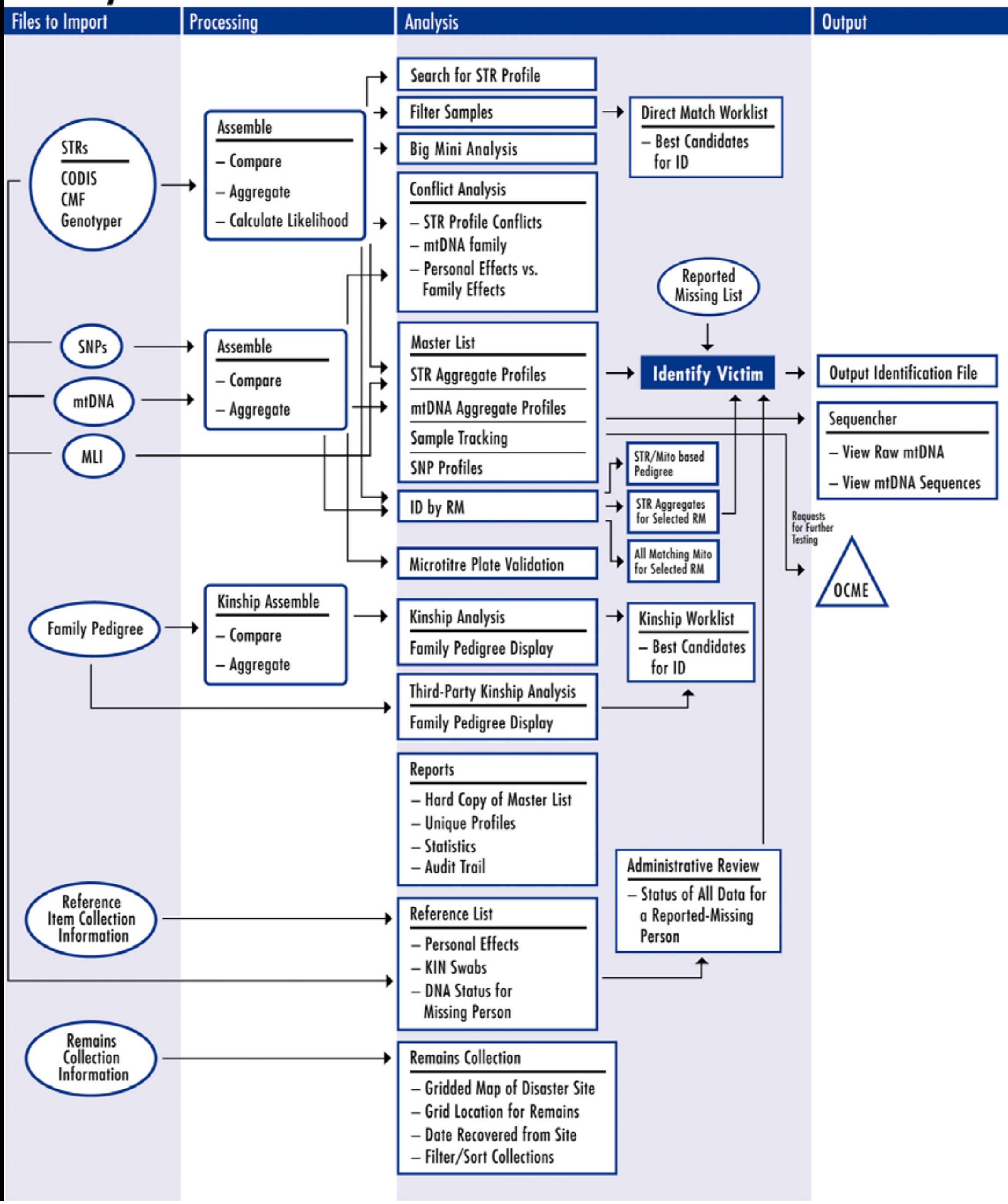
TCA GENE
AGT CODES
FORENSICS



M-FISys

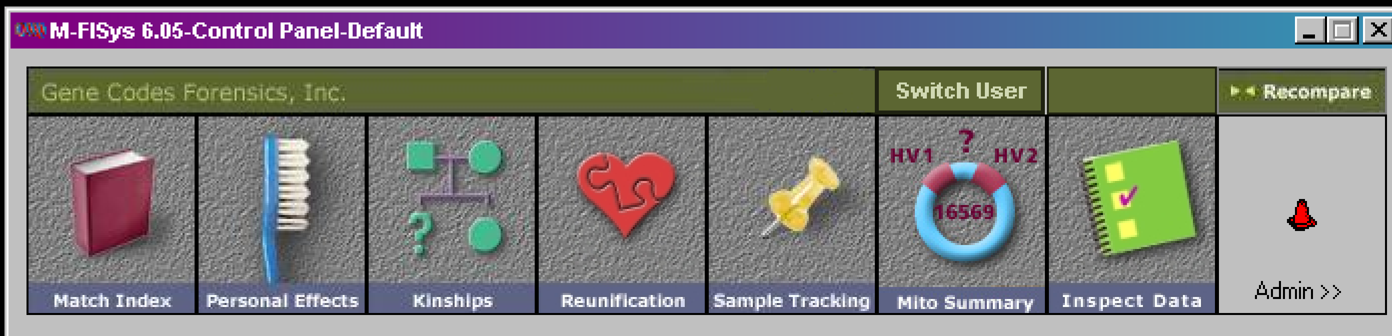
Mass Fatality Identification SYStem

M-FISys Identification Process



- 22 data sources
- Data curation, not victim identification
- Aggregate formation a critical first step

(Example data used here)



Samples found to match each other on the basis of DNA (or other) evidence

TCA GENE
AGT CODES
FORENSICS

M-FISys 6.05-Master List-Default

139476 | Locate | ID | Samples: 61 | Identified Aggregates: 0 | Identifiable Aggregates: 9 | Inidentified Aggregates: 1 | Compared by STRs

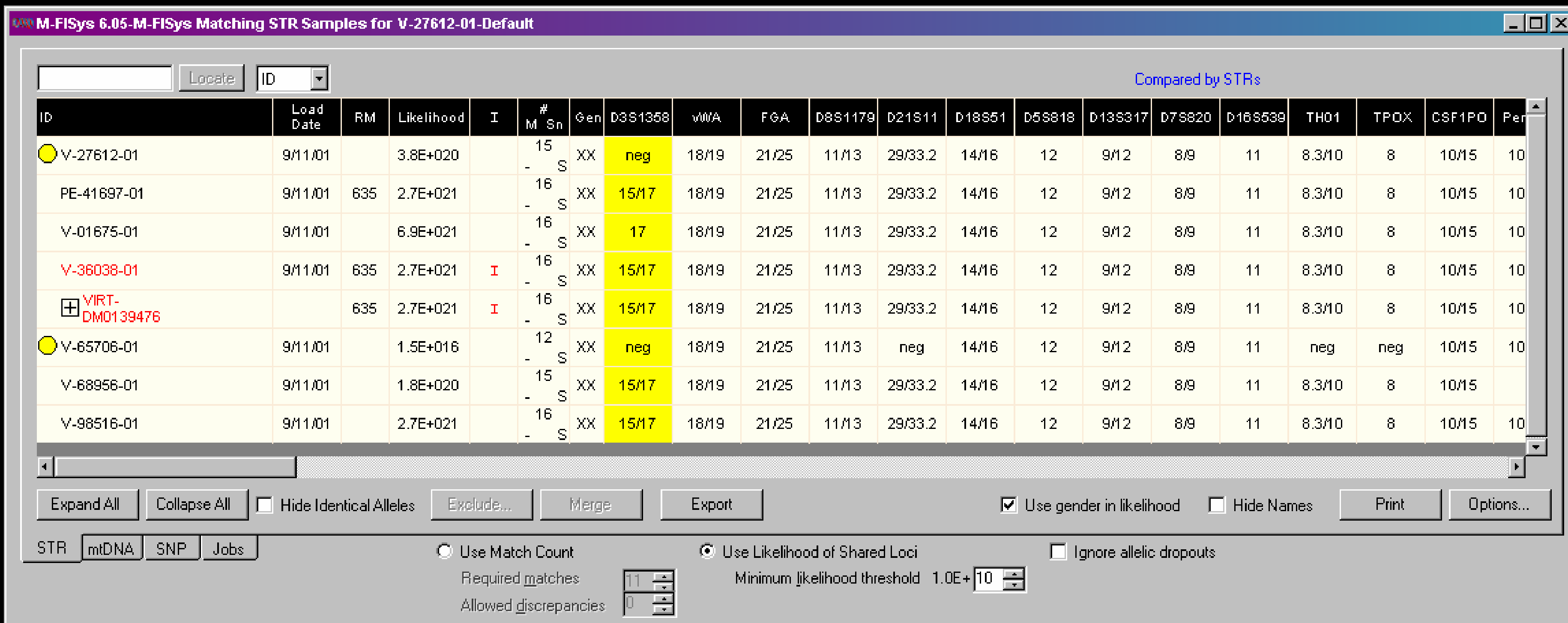
ID	Load Date	RM	Likelihood	I	# M Sn	Gen	D3S1358	vWA	FGA	D8S1179	D21S11	D18S51	D5S818	D13S317	D7S820	D16S539	TH01	TPOX	CSF1PO	Penta
⊟ RM# 635 (8)			*2.7E+021*	2		XX	15/17	18/19	21/25	11/13	29/33.2	14/16	12	9/12	8/9	11	8.3/10	8	10/15	10/13
PE-41697-01	9/11/01	635	2.7E+021		16 - S	XX	15/17	18/19	21/25	11/13	29/33.2	14/16	12	9/12	8/9	11	8.3/10	8	10/15	10/13
V-27612-01	9/11/01		3.8E+020 3.8E+020		15 - S	XX	neg	18/19	21/25	11/13	29/33.2	14/16	12	9/12	8/9	11	8.3/10	8	10/15	10/13
V-36038-01	9/11/01	635	2.7E+021 2.7E+021	I	16 - S	XX	15/17	18/19	21/25	11/13	29/33.2	14/16	12	9/12	8/9	11	8.3/10	8	10/15	10/13
⊞ VIRT-DM0139476		635	2.7E+021 2.7E+021	I	16 - S	XX	15/17	18/19	21/25	11/13	29/33.2	14/16	12	9/12	8/9	11	8.3/10	8	10/15	10/13
⊟ VIRT-DM0150349		[illegible]	[illegible]	[illegible]	[illegible] S	neg	15/17	neg	21/25	11/13	neg	14/16	12	neg	8/9	neg	8.3/10	neg	neg	10/13
V-50349-01	9/11/01		4.0E+006		5 - S	neg	15/17	neg	neg	11/13	neg	neg	12	neg	8/9	neg	8.3/10	neg	neg	-
V-50349-02	9/11/01		2.9E+009		7 - S	neg	15/17	neg	21/25	11/13	neg	14/16	12	neg	8/9	neg	8.3/10	neg	neg	-
V-50349-03	9/11/01		2.2E+006		5 - S	neg	15/17	neg	neg	neg	neg	14/16	12	neg	neg	neg	8.3/10	neg	neg	10/13
V-65706-01	9/11/01		1.5E+016 1.5E+016		12 - S	XX	neg	18/19	21/25	11/13	neg	14/16	12	9/12	8/9	11	neg	neg	10/15	10/13
V-68956-01	9/11/01		1.8E+020 1.8E+020		15 - S	XX	15/17	18/19	21/25	11/13	29/33.2	14/16	12	9/12	8/9	11	8.3/10	8	10/15	-
V-98516-01	9/11/01		2.7E+021 2.7E+021		16 - S	XX	15/17	18/19	21/25	11/13	29/33.2	14/16	12	9/12	8/9	11	8.3/10	8	10/15	10/13
⊞ RM# 243 (7)			*2.8E+018*	0		XY	15	17/19	23/24	13/14	30/32.2	9/17	12/13	11/13	8/10	12	8/9	8/11	10/11	9/10
⊞ RM# 967 (6)			*5.8E+024*	0		XX	12/16	18	15.2/23	15	28/28.2	12/14	13/14	11/12	9/14	10	7/9.3	8	6/10	9/13

RM# 635: Rhea Elza Leonardo (Chain G1)
Method: DNA match to Personal Effect

Expand All | Collapse All | Hide Identical Alleles | Exclude... | Merge | Export | ☑ Use gender in likelihood | ☐ Hide Names | Print | Options...

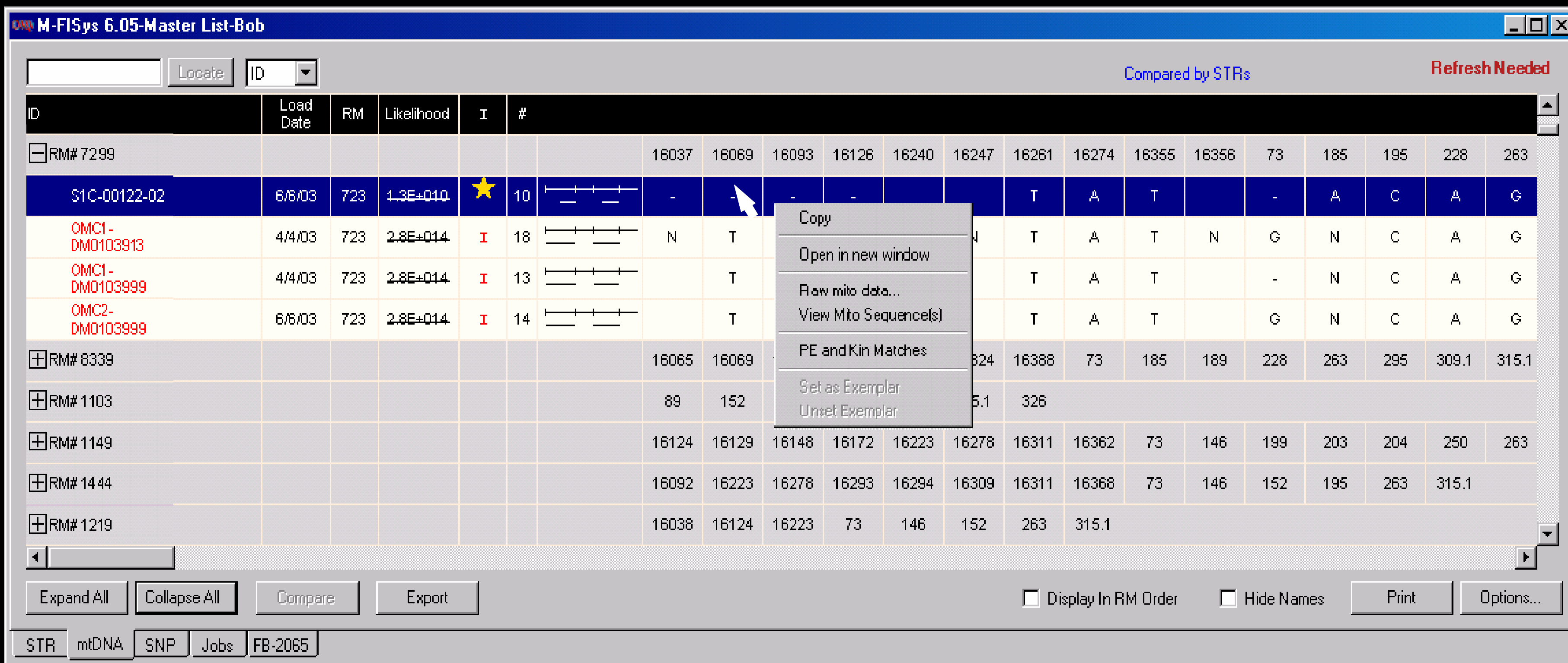
STR | mtDNA | SNP | Jobs | FB-2065

TCA GENE
AGT CODES
FORENSICS

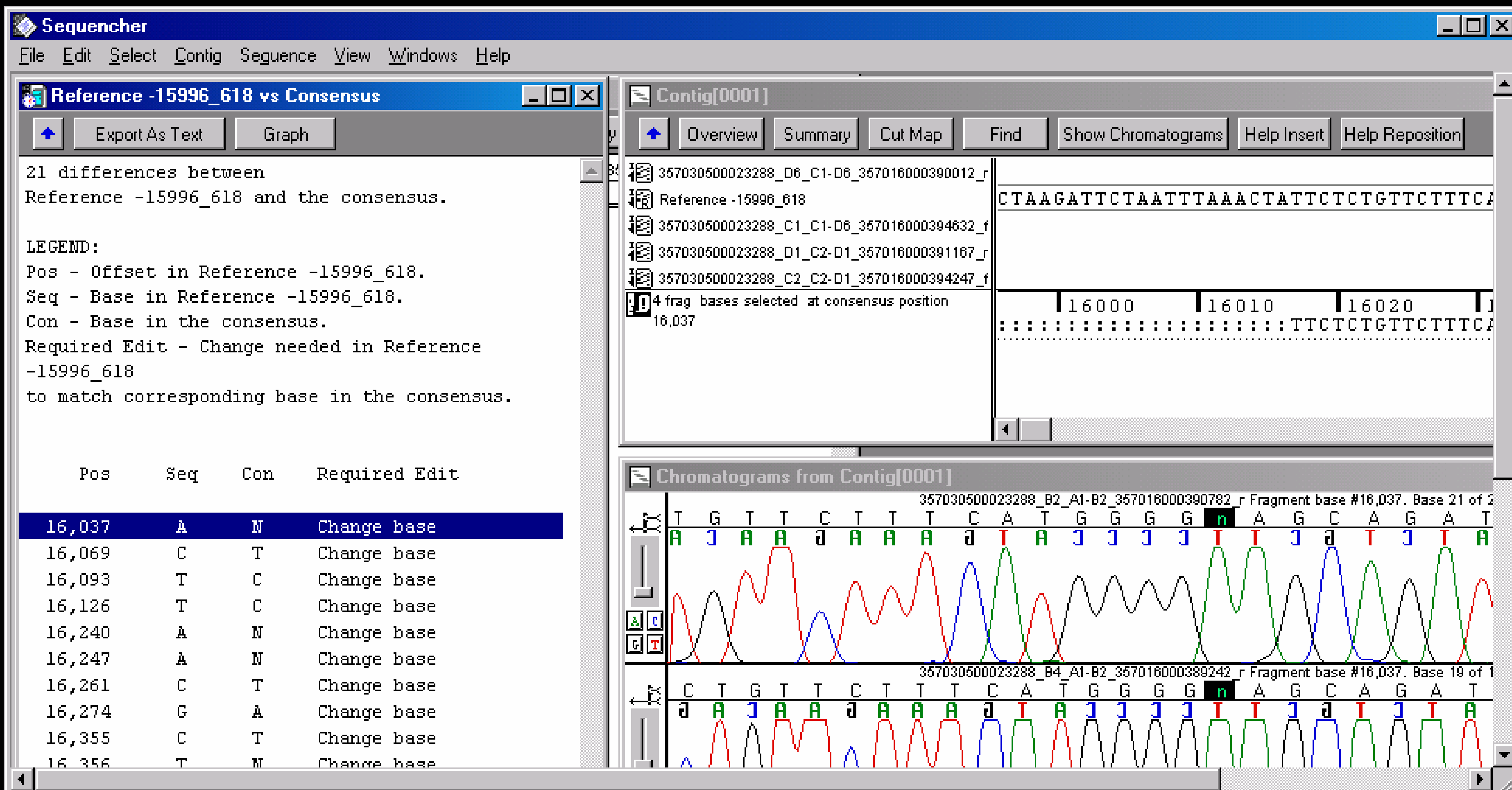


ID	Load Date	RM	Likelihood	I	# M Sn	Gen	D3S1358	vWA	FGA	D8S1179	D21S11	D18S51	D5S818	D13S317	D7S820	D16S539	TH01	TPOX	CSF1PO	Pen
V-27612-01	9/11/01		3.8E+020		15 - S	XX	neg	18/19	21/25	11/13	29/33.2	14/16	12	9/12	8/9	11	8.3/10	8	10/15	10
PE-41697-01	9/11/01	635	2.7E+021		16 - S	XX	15/17	18/19	21/25	11/13	29/33.2	14/16	12	9/12	8/9	11	8.3/10	8	10/15	10
V-01675-01	9/11/01		6.9E+021		16 - S	XX	17	18/19	21/25	11/13	29/33.2	14/16	12	9/12	8/9	11	8.3/10	8	10/15	10
V-36038-01	9/11/01	635	2.7E+021	I	16 - S	XX	15/17	18/19	21/25	11/13	29/33.2	14/16	12	9/12	8/9	11	8.3/10	8	10/15	10
VIRT-DM0139476		635	2.7E+021	I	16 - S	XX	15/17	18/19	21/25	11/13	29/33.2	14/16	12	9/12	8/9	11	8.3/10	8	10/15	10
V-65706-01	9/11/01		1.5E+016		12 - S	XX	neg	18/19	21/25	11/13	neg	14/16	12	9/12	8/9	11	neg	neg	10/15	10
V-68956-01	9/11/01		1.8E+020		15 - S	XX	15/17	18/19	21/25	11/13	29/33.2	14/16	12	9/12	8/9	11	8.3/10	8	10/15	
V-98516-01	9/11/01		2.7E+021		16 - S	XX	15/17	18/19	21/25	11/13	29/33.2	14/16	12	9/12	8/9	11	8.3/10	8	10/15	10

TCA GENE
AGT CODES
FORENSICS



TCA GENE
AGT CODES
FORENSICS



TCA GENE AGT CODES FORENSICS

M-FISys 6.05-Master List-Default

Locate | ID

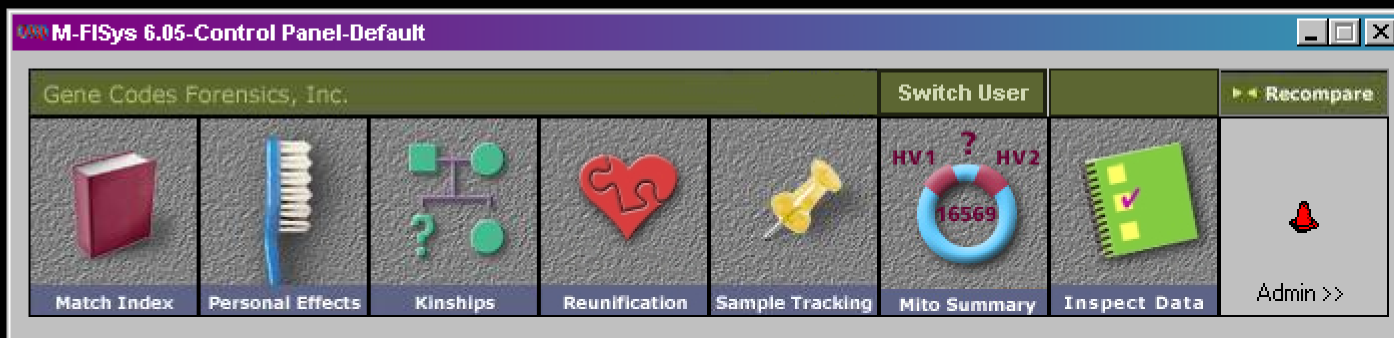
Compared by SNPs

ID	Load Date	RM	Likelihood	I	# STRs	Total	% Het	Homo-zygous	Failed T-SNP	Mis-matches
RM# 263			1.4E+032			71	51%	35	0	
RM# 635			2.2E+032			71	51%	35	0	
PE-41697-01	9/23/03	635	2.2E+032		16	71	51%	35	0	0/0
V-36038-01	9/23/03	635	2.8E+031	I	16	70	51%	34	0	0/0
V-49612-01	9/23/03		9.7E+029		0	65	52%	31	0	0/0
V-39476-01	9/23/03	635	1.7E+006	I	16	19	63%	7	4	0/0
RM# 702			2.3E+035			71	46%	38	0	
PE-96387-01	9/23/03	702	2.3E+035		16	71	46%	38	0	0/0
V-50084-01	9/23/03		2.3E+035		16	71	46%	38	0	0/0
V-68803-01	9/23/03		3.3E+027		16	56	46%	30	0	0/0
V-17499-01	9/23/03		1.5E+022		0	46	46%	25	0	0/0
V-85815-01	9/23/03		2.2E+022		4	45	40%	27	0	0/0

ID	Amel	65882	68532	234217	231480	62059	56608	61955	220875	58388	63799	219561	60188	182622	85187	212605	58091	66026	63836	214373	238155	63979	235106	241554	66561	[illegible]
RM# 263	M	CC	TT	TC	CC	TT	TC	CC	TC	TC	TC	CC	TT	TC	TC	TT	TC	CC	CC	TC	TC	TC	TT	TT	TT	T
RM# 635	F	TC	TC	TC	TC	TT	TC	TT	TC	TC	TC	CC	TT	TC	TC	TC	TC	TC	TC	TT	TC	TT	TT	TC	TT	C
PE-41697-01	F	TC	TC	TC	TC	TT	TC	TT	TC	TC	TC	CC	TT	TC	TC	TC	TC	TC	TC	TT	TC	TT	TT	TC	TT	C
V-36038-01	F	TC	TC	TC	TC	TT	TC	TT	TC	TC	TC	CC	TT	TC	TC	TC	TC	TC	TC	TT	TC	TT	TT	TC	TT	C
V-49612-01	F	TC	TC	TC	TC	neg	TC	TT	TC	TC	TC	CC	TT	TC	TC	TC	TC	TC	TC	TT	TC	TT	TT	TC	TT	ne
V-39476-01	neg	neg	TC	TC	neg	neg	neg	neg	TC	TC	TC	CC	neg	neg	neg	neg	TC	neg	neg	neg	neg	neg	TT	TC	neg	ne
RM# 702	F	TC	TT	TC	TC	TT	CC	CC	TC	TC	TC	TT	CC	TC	CC	TC	TT	TT	CC	TT	TT	TC	TT	TC	TC	T
PE-96387-01	F	TC	TT	TC	TC	TT	CC	CC	TC	TC	TC	TT	CC	TC	CC	TC	TT	TT	CC	TT	TT	TC	TT	TC	TC	T
V-50084-01	F	TC	TT	TC	TC	TT	CC	CC	TC	TC	TC	TT	CC	TC	CC	TC	TT	TT	CC	TT	TT	TC	TT	TC	TC	T
V-68803-01	F	TC	TT	neg	neg	TT	CC	CC	TC	TC	TC	neg	CC	neg	neg	TC	TT	TT	CC	TT	neg	TC	neg	TC	TC	T
V-17499-01	neg	TC	TT	TC	TC	TT	CC	CC	TC	neg	neg	TT	CC	neg	neg	TC	neg	TT	CC	TT	TT	TC	TT	TC	TC	T
V-85815-01	F	neg	TT	neg	TC	TT	neg	CC	neg	neg	TC	TT	CC	neg	CC	neg	TT	TT	CC	TT	TT	TC	neg	TC	TC	ne

Expand All | Collapse All | Export | Hide SNP colors | Hide Names | Print | Options...

STR | mtDNA | SNP | Jobs | FB-2065



Samples available for DNA analysis

TCA GENE
AGT CODES
FORENSICS

M-FISys 7.00-Reference Samples by RM-Default

Reported Missing

15 | **Harold Norris Abhidhar**

Member of Service

Notification Status

Admin Review Status

Reference Samples DNA Status

Personal Effect Samples

Type	Source	Item #	Sample Name	# of Loci	mtDNA Length	# of SNPs
☐ Toothbrush: PE	Albany	1	CORE-PE-094070-01	16	-	71
☐			PE-94070-01	16	-	71

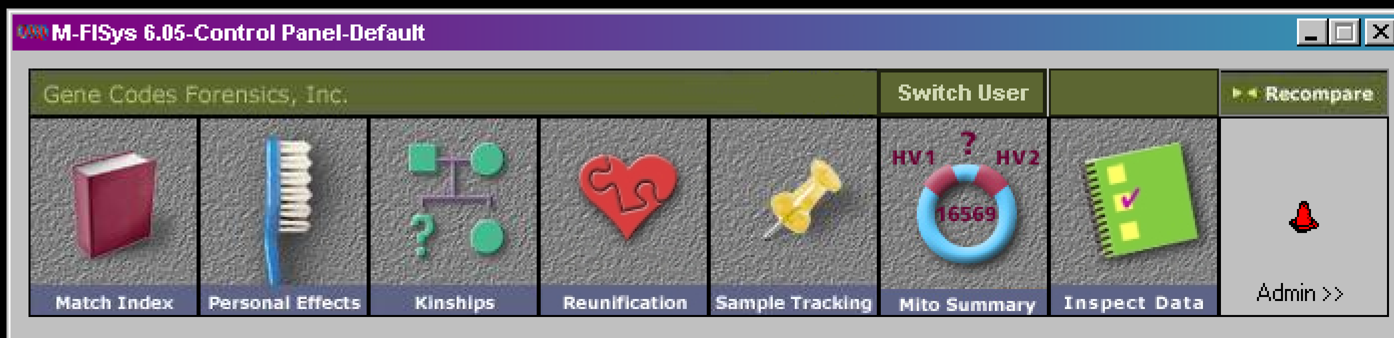
Kin Samples

Type	Source	Item #	Sample Name	# of Loci	mtDNA Length	# of SNPs
☐ Buccal: BM	Albany	0	CORE-BM-02017 #00	16	689	71
☐			BM-02017 #00	16	689	71
☐ Buccal: BD	Albany	2	CORE-BD-08707 #02	16	-	71
☐			BD-08707 #02	16	-	71
☐ Buccal: PR	Albany	3	CORE-PR-07192 #03	-	-	71
☐			PR-07192 #03	-	-	71

☐ Hide Names

01WDI List | Print

RM View | List View



Family relationships of victims

TCA GENE
AGT CODES
FORENSICS

M-FISys 7.00-Family Display-Administrator

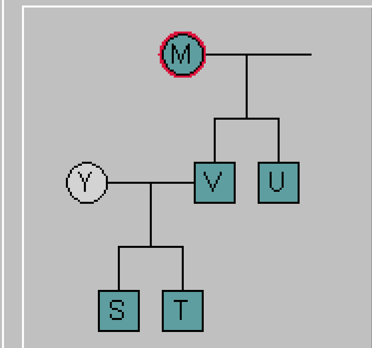
Family

#76

Victim

VIRT-DM0718273
VIRT-DM0318633
VIRT-DM0018303
VIRT-DM0018306
VIRT-DM0401141
VIRT-DM0714771
VIRT-DM0515755
VIRT-DM0000206
VIRT-DM0202227

Pedigree



Reported | Adjusted

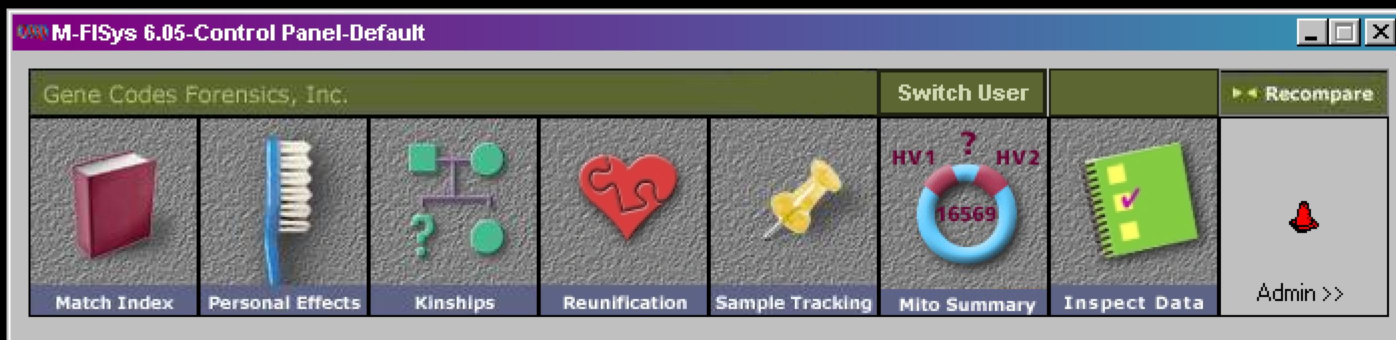
Profiles

	VIRT-DM0111673	BM-01651 #01	BU-51601 #01	BU-64642 #01	BS-51602 #02
Gen	XY	XX	XY	XY	XY
D3S1358	16	14/16	16	16	15/16
vWA	17/19	19/20	17/19	17/19	17
FGA	22/25	22/25	22/25	22/25	22
D8S1179	14/16	14	14/16	14/16	14/15
D21S11	32.2	28/32.2	32.2	32.2	31/32.2
D18S51	17/18	18	17/18	17/18	17/18
D5S818	12/13	8/13	12/13	12/13	12/13
D13S317	12/14	11/14	12/14	12/14	12/14
D7S820	8	8/9	8	8	8/11
D16S539	9/14	9/14	9/14	9/14	9/14
TH01	7	7	7	7	6/7
TPOX	6/8	6/9	6/8	6/8	6
CSF1PO	8/12	8/14	8/12	8/12	9/12
Penta D	6/8	neg	neg	neg	6/9
Penta E	11	neg	neg	neg	11/15
min LR to V	1.4E+023	2.7E+006	9.9E+008	9.9E+008	2.0E+007

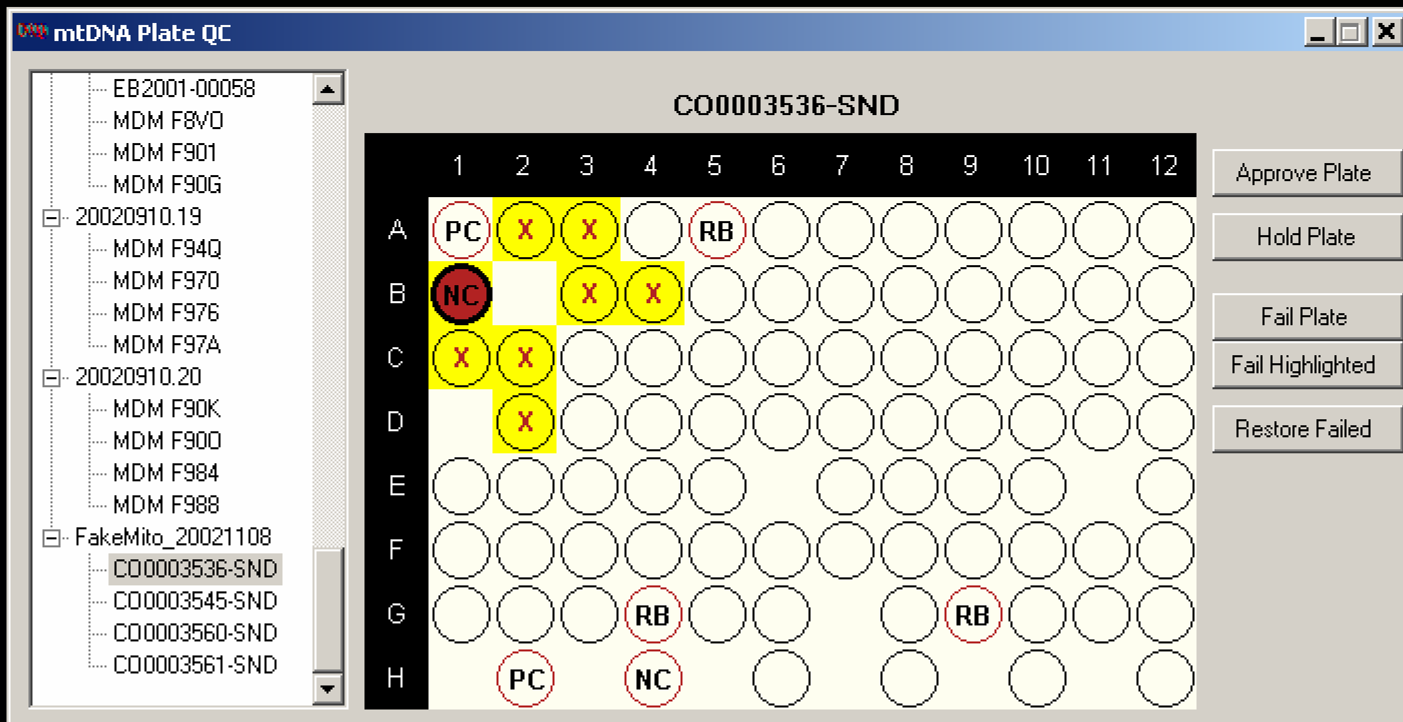
Identification Method

add'l pieces

Kinship Work List



Quality control and other data curation tools


mtDNA Plate QC
EB2001-00058
MDM F8VO
MDM F901
MDM F90G
20020910.19
MDM F94Q
MDM F970
MDM F976
MDM F97A
20020910.20
MDM F90K
MDM F90O
MDM F984
MDM F988
FakeMito_20021108
CO0003536-SND
CO0003545-SND
CO0003560-SND
CO0003561-SND
CO0003536-SND
1 2 3 4 5 6 7 8 9 10 11 12
A B C D E F G H
PC
X
RB
NC
Approve Plate
Hold Plate
Fail Plate
Fail Highlighted
Restore Failed

TCA GENE
AGT CODES
FORENSICS