ISHI Workshop on New Loci and Kits
October 10, 2013 (Atlanta, GA)
New Autosomal and Y-STR Loci and Kits: Making Data Driven Decisions

NIST Studies with New Y-STR Loci and Kits

Michael D. Coble
NIST Applied Genetics Group

NIST Applied Genetics

NIST

Product Disclaimer

- **I will mention commercial STR kit names and information, but I am in no way attempting to endorse any specific products.**

- **NIST Disclaimer**: 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 it imply that any of the materials, instruments or equipment identified are necessarily the best available for the purpose.

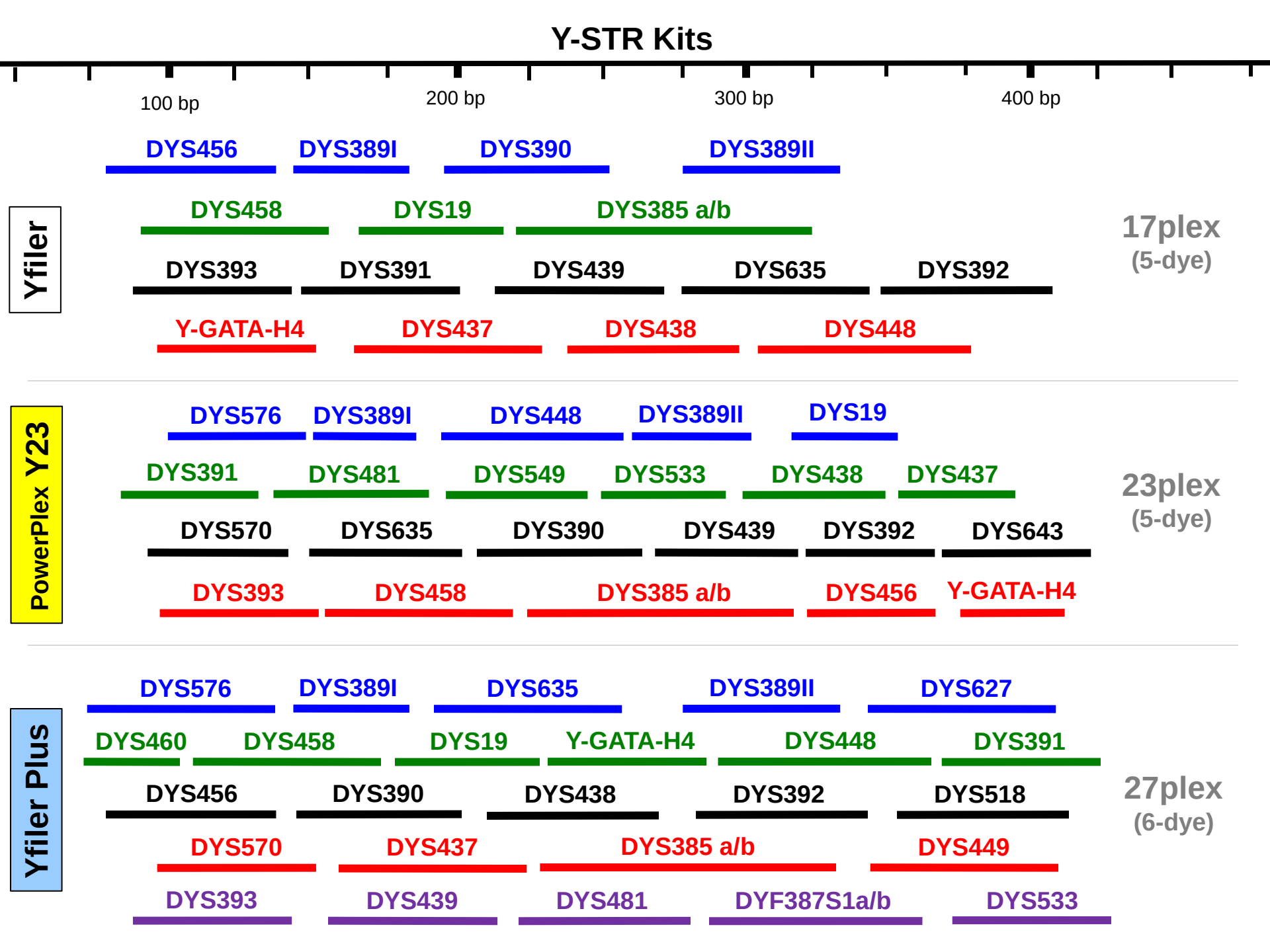
- **Points of view are mine** and do not necessarily represent the official position of the National Institute of Standards and Technology or the U.S. Department of Justice. **Our group receives or has received funding from the FBI Laboratory and the National Institute of Justice.**

Presentation Outline

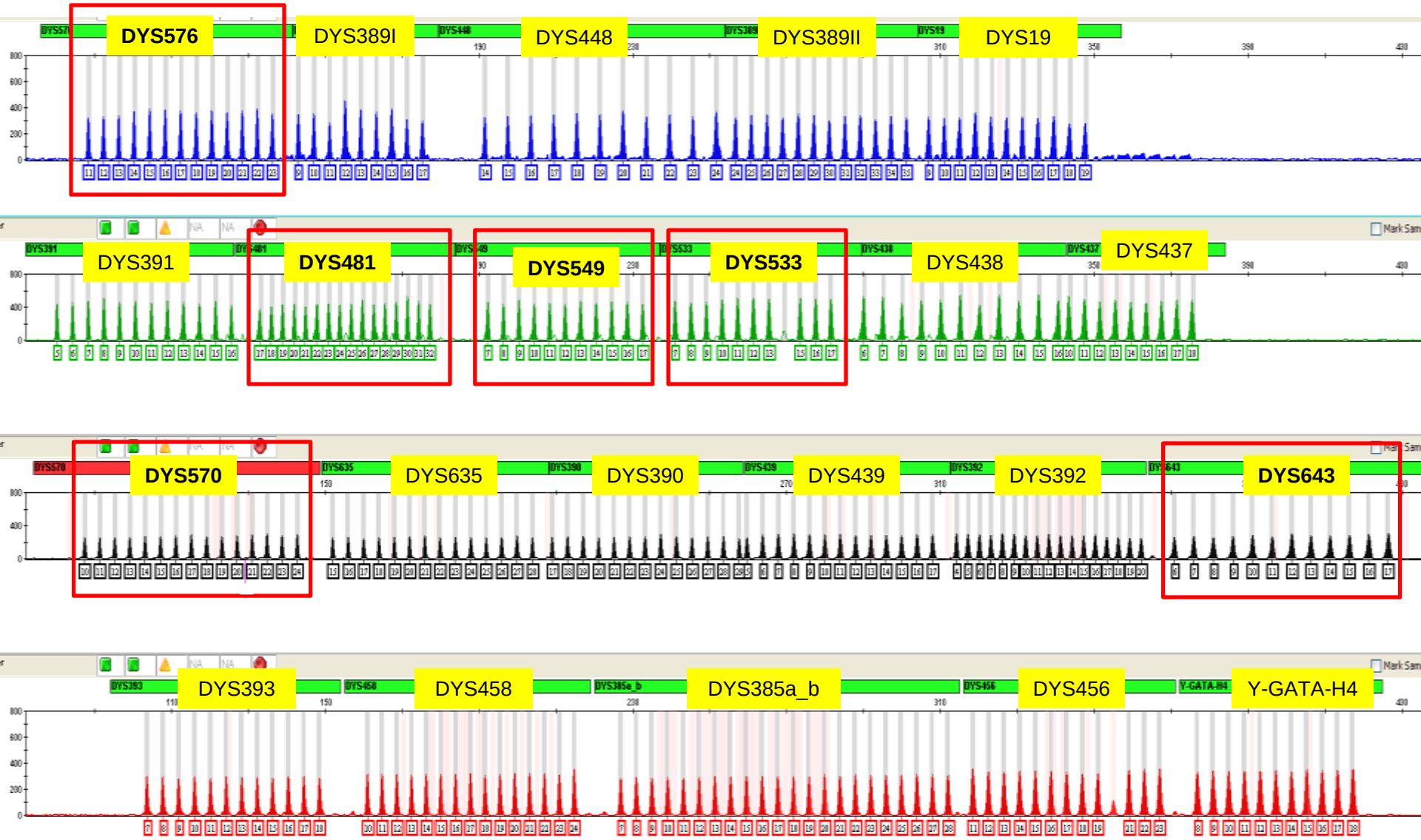
- Brief Y-chromosome review
- Y-STR kits and loci
- U.S. population haplotype data – *FSI Genetics* article and Yfiler Plus loci studies
- Rapidly mutating (RM) Y-STRs
  - Utility for common Y-STR haplotypes
  - Utility for close relatives

What has happened in the past decade…

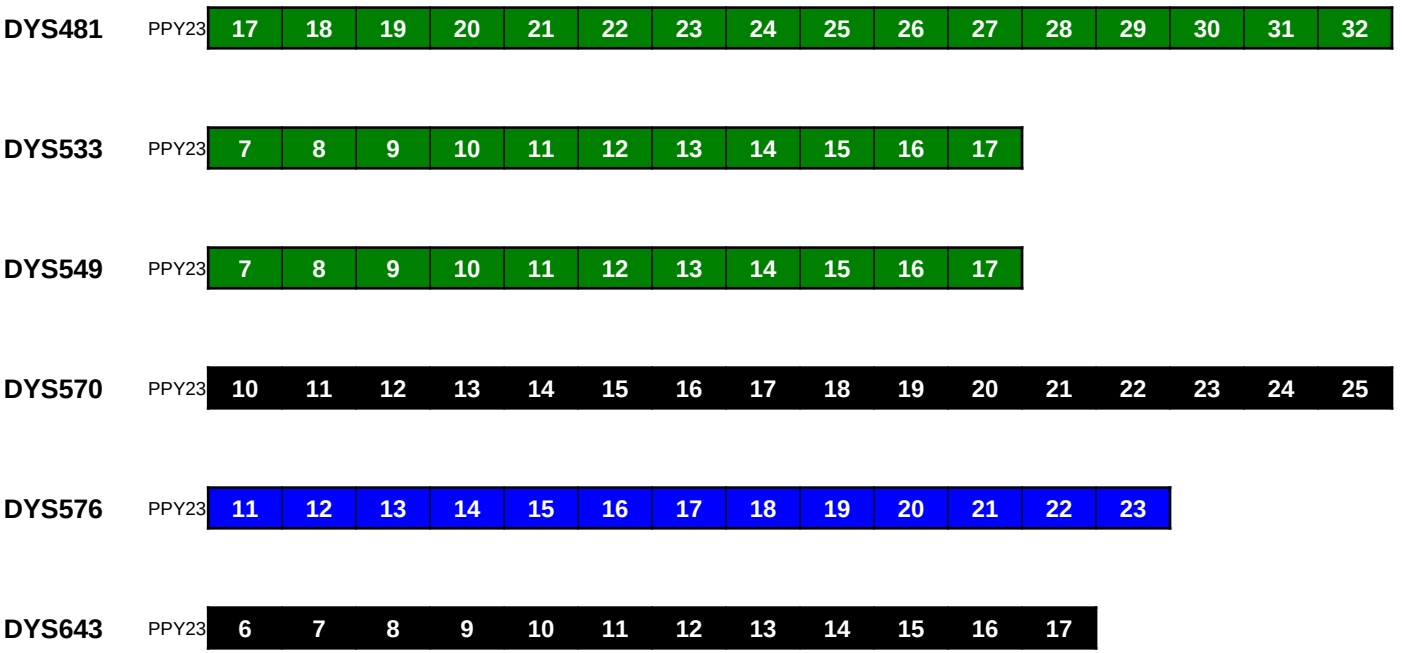
- **Selection of core Y-STR loci** (SWGDAM Jan 2003)
- "Full" Y-chromosome sequence became available in June 2003; over 700 Y-STR loci identified (only ~20 in 2000)
- **Commercial Y-STR kits released**
  - ~~Y-PLEX 6,5,12 (2001-03)~~, **PowerPlex Y** (9/03), **Yfiler** (12/04), **PPY23** (6/12)
    Yfiler Plus (coming soon)
- Many population studies performed and online databases generated with thousands of Y-STR haplotypes
- Forensic casework demonstrations showing value of Y-STR testing along with court acceptance
- Some renewed interest in Y-STRs to aid familial searching


Y-STR Kits
100 bp
200 bp
300 bp
400 bp
Yfiler
DYS456
DYS389I
DYS390
DYS389II
DYS458
DYS19
DYS385 a/b
DYS393
DYS391
DYS439
DYS635
DYS392
Y-GATA-H4
DYS437
DYS438
DYS448
17plex
(5-dye)
PowerPlex Y23
DYS576
DYS389I
DYS448
DYS389II
DYS19
DYS391
DYS481
DYS549
DYS533
DYS438
DYS437
DYS570
DYS635
DYS390
DYS439
DYS392
DYS643
DYS393
DYS458
DYS385 a/b
DYS456
Y-GATA-H4
23plex
(5-dye)
Yfiler Plus
DYS576
DYS389I
DYS635
DYS389II
DYS627
DYS460
DYS458
DYS19
Y-GATA-H4
DYS448
DYS391
DYS456
DYS390
DYS438
DYS392
DYS518
DYS570
DYS437
DYS385 a/b
DYS449
DYS393
DYS439
DYS481
DYF387S1a/b
DYS533
27plex
(6-dye)

PowerPlex Y23 Allelic Ladders



Allelic Ladder Alleles for Six New Loci



No alleles in the NIST 1032 males fell outside of these allele ranges

Variant alleles observed

DYS481 25.1

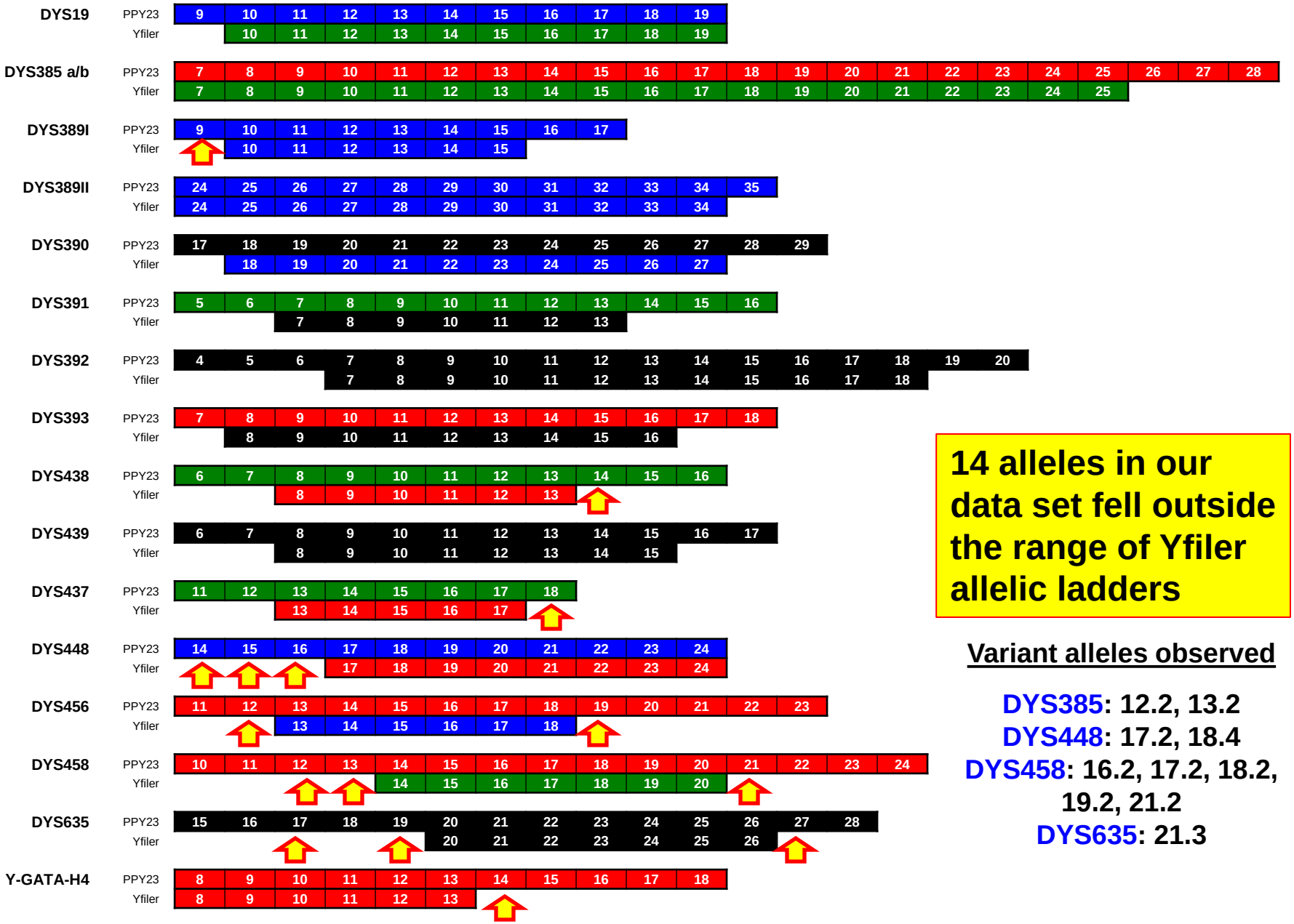
DYS643 11.1

Number of Alleles Present in Y-STR Kit Allelic Ladders

Y-STR	PowerPlex Y23	Yfiler	Difference
DYS19	11	10	+1
DYS385 a/b	22	19	+3
DYS389I	9	6	+3
DYS389II	12	11	+1
DYS390	13	10	+3
DYS391	12	7	+5
DYS392	17	12	+5
DYS393	12	9	+3
DYS438	11	6	+5
DYS439	12	8	+4
DYS437	8	5	+3
DYS448	11	8	+3
DYS456	13	6	+7
DYS458	15	7	+8
DYS635	14	7	+7
GATA-H4	11	6	+5
TOTAL	**203**	**137**	**+66**

203 + 79 (in 6 additional loci) = 282 alleles represented in PowerPlex Y23 ladders

Comparison of Alleles in Y-STR Kit Allelic Ladders



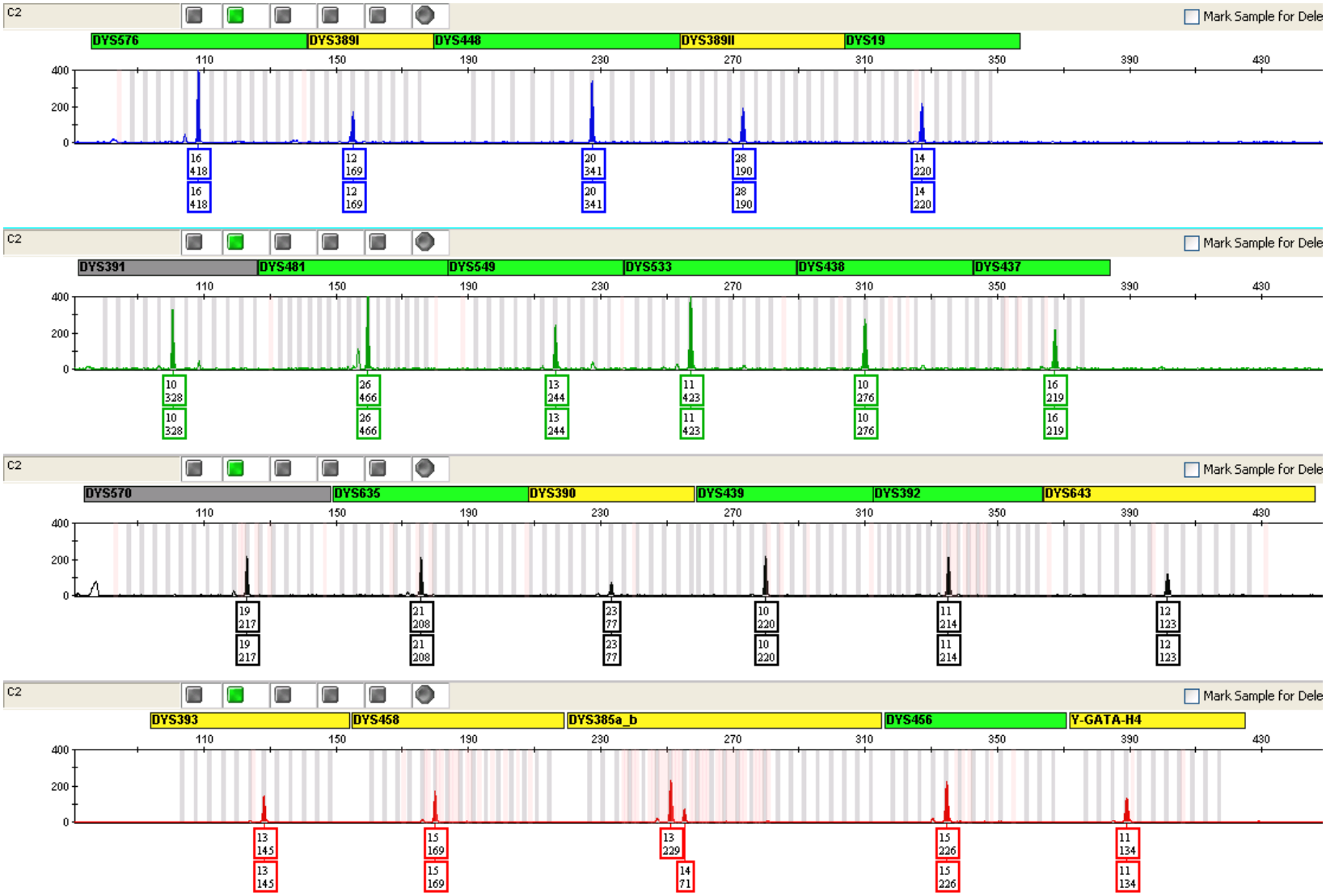
14 alleles in our data set fell outside the range of Yfiler allelic ladders

Variant alleles observed

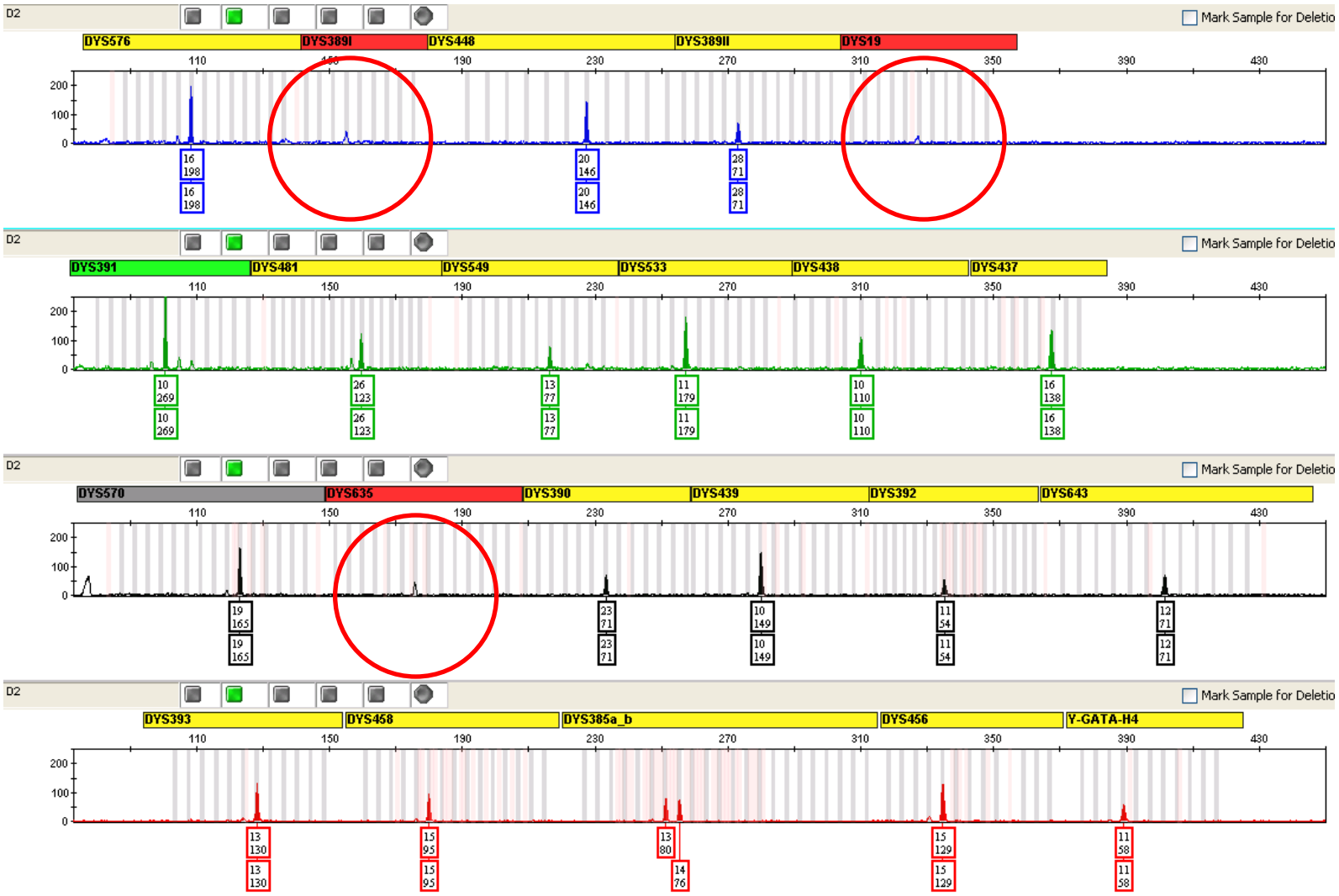
DYS385: 12.2, 13.2
DYS448: 17.2, 18.4
DYS458: 16.2, 17.2, 18.2, 19.2, 21.2
DYS635: 21.3

PowerPlex Y23 Sensitivity

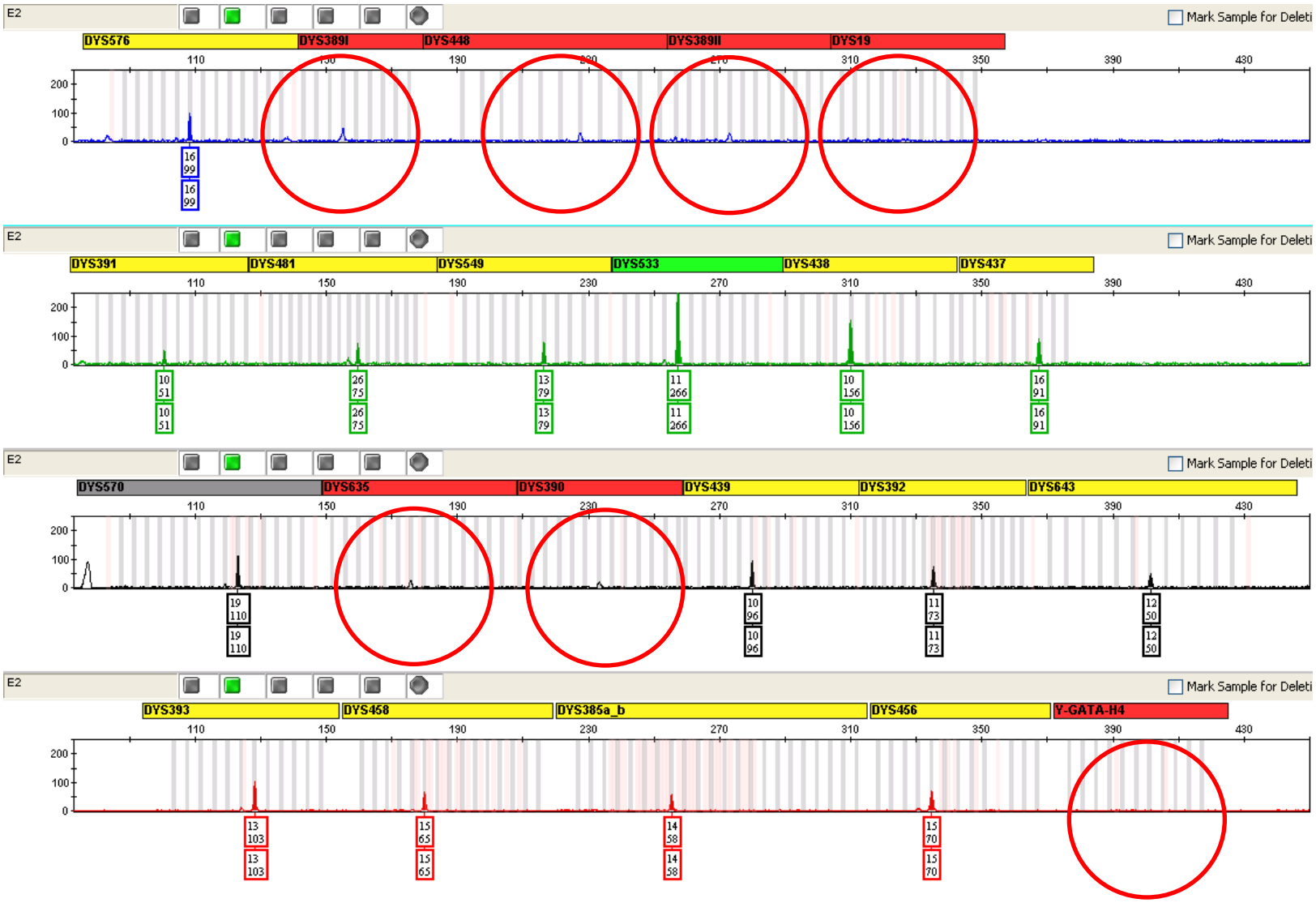
Sensitivity Study – Sample C (125pg)



Sensitivity Study – Sample D (62.5pg)

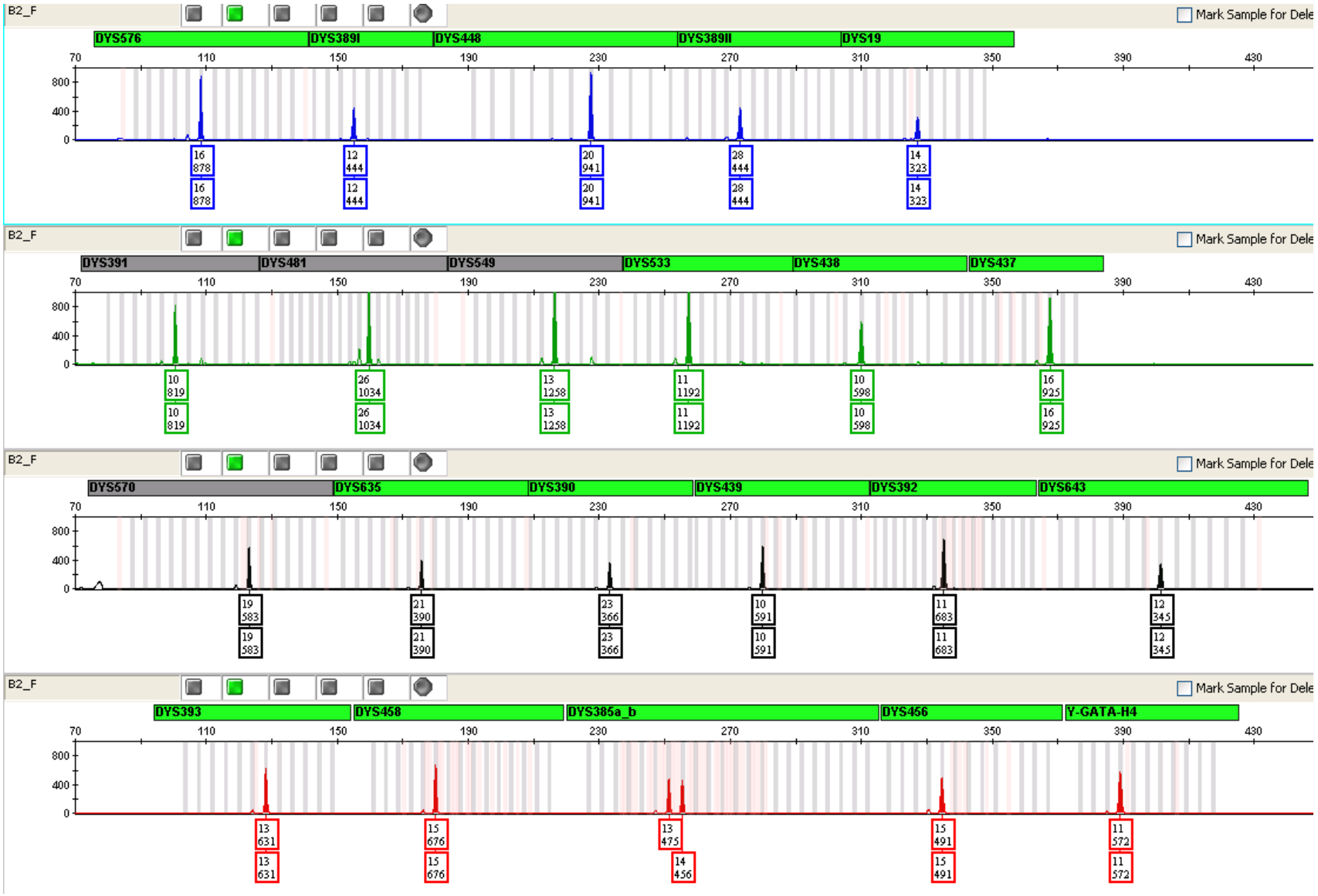


Sensitivity Study – Sample E (31.2pg)

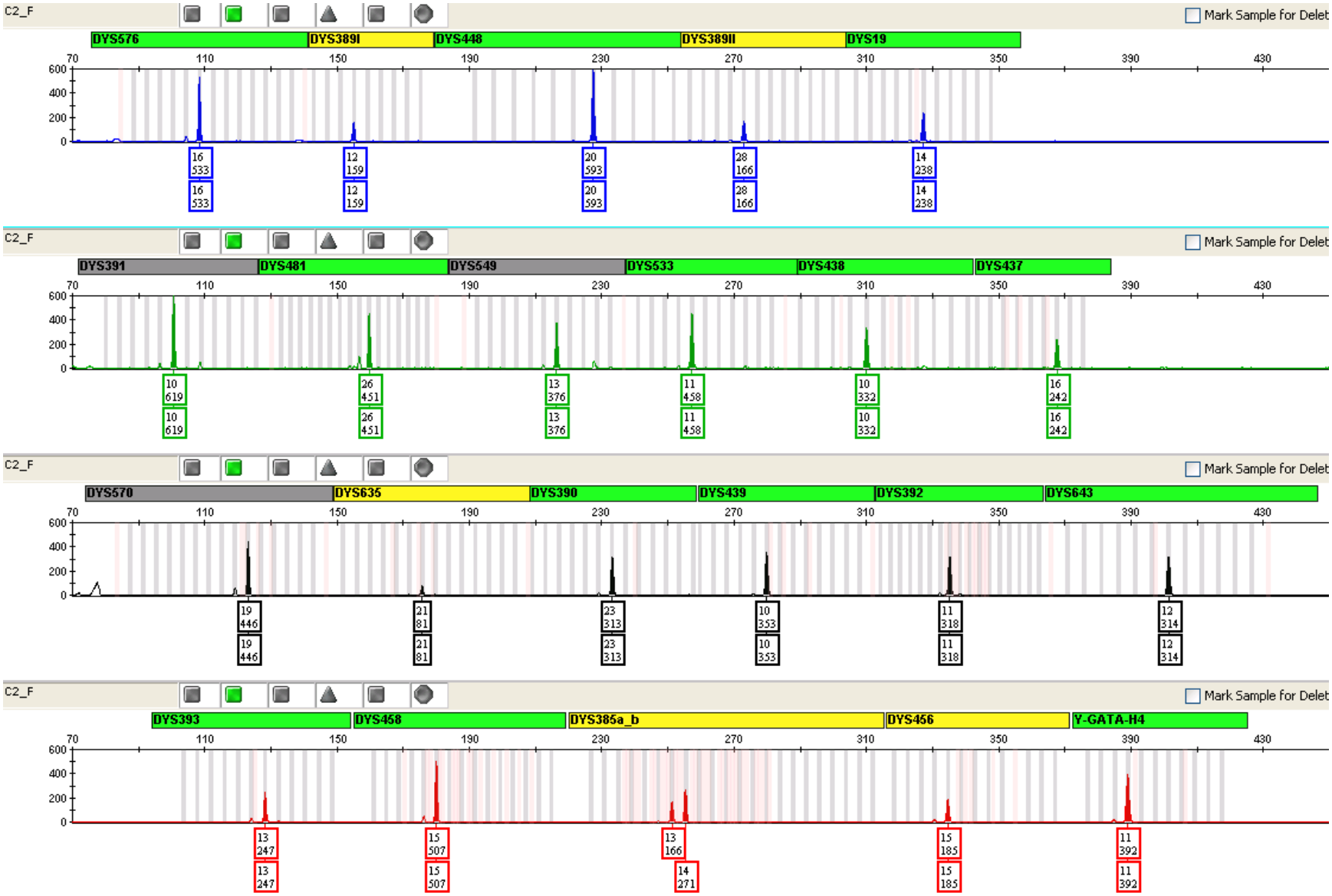


PowerPlex Y23
Mixtures

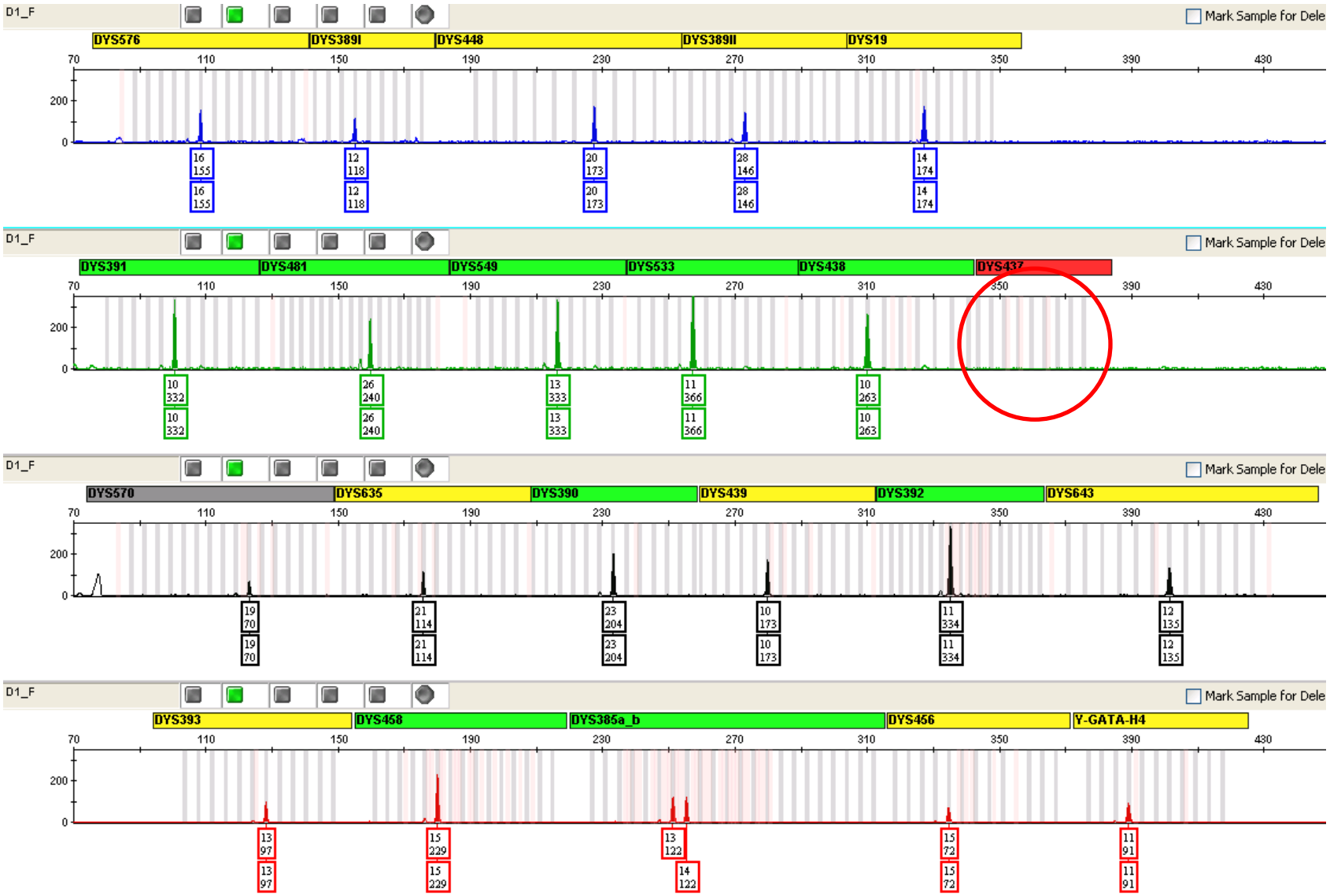
250pg male + 400ng female



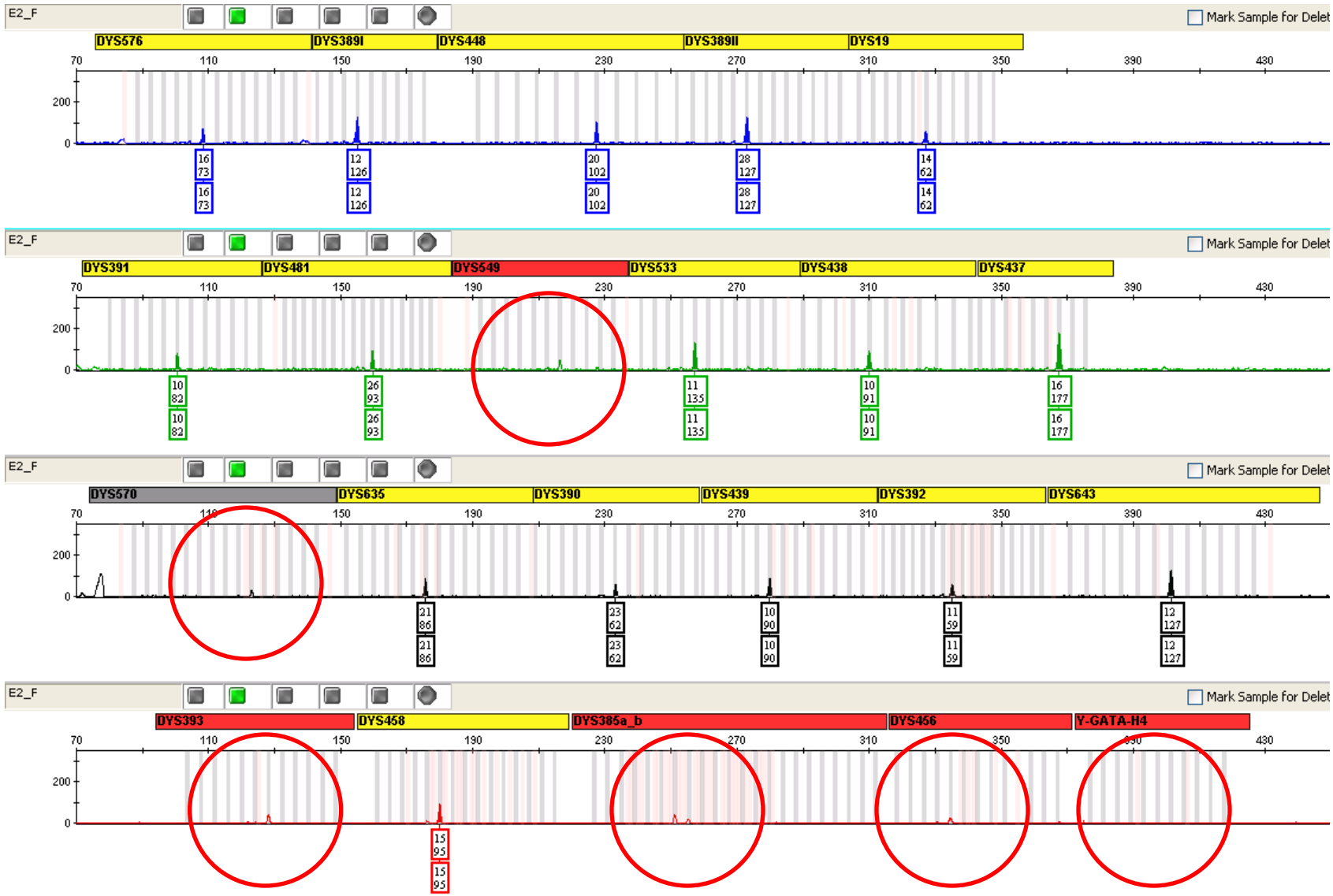
125pg male + 400ng female



62.5pg male + 400ng female

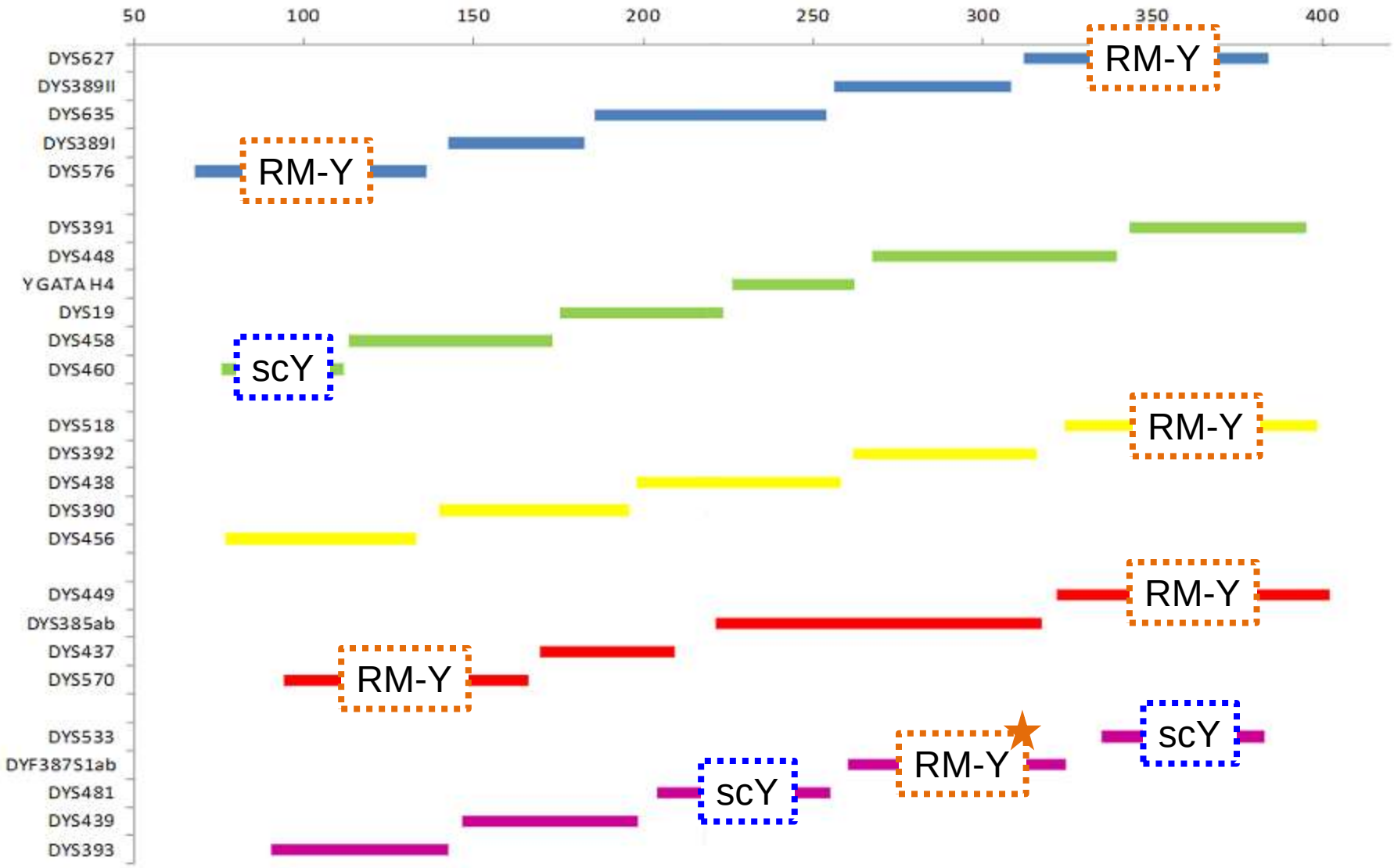


31.2pg male + 400ng female



16/23 loci amplified

Proposed Yfiler Plus Kit



RM = Rapidly Mutating, sc = single copy

Slide courtesy of Dr. Julio Mulero (Life Technologies)

Discrimination Capacity

- is a measure of the number of unique haplotypes in a given population

$$DC = \frac{\#H}{N}$$

# of Haplotypes

Population size

N = 948 males	PowerPlex Y	Yfiler	PowerPlex Y23	Yfiler Plus*
# haplotypes	**816**	**930**	**945**	**946**
discrimination capacity	0.8608	0.9810	0.9968	0.9979
# times haplotype observed	PPY (12 loci)	Yfiler (17 loci)	PPY23 (23 loci)	Yfiler Plus* (26 loci)*
1	751	916	942	944
2	42	11	3	2
3	12	2	.	.
4	4	1	.	.
5	2	.	.	.
6	2	.	.	.
7	.	.	.	.
8	1	.	.	.
9	.	.	.	.
10	.	.	.	.
11	1	.	.	.
12	.	.	.	.
13	.	.	.	.
14	.	.	.	.
15	.	.	.	.
16	.	.	.	.
17	.	.	.	.
18	.	.	.	.
19	.	.	.	.
20	1	.	.	.

Number of unique and shared haplotypes observed with various combinations of Y-STR loci across 948 U.S. population samples

944 haplotypes occur once; and 2 sets of sample pairs cannot be resolved from one another

*Note: Analysis does not include information from DYS460 – only 26 of the 27 markers in Yfiler Plus were examined in this study.

N = 948 males	Yfiler	New Loci*	Yfiler Plus*
# haplotypes	**930**	**945**	**946**
discrimination capacity	0.9810	0.9842	0.9979
# times haplotype observed	Yfiler (17 loci)	New Loci* (9 loci)	Yfiler Plus* (26 loci)
1	916	918	944
2	11	15	2
3	2	.	.
4	1	.	.
5	.	.	.
6	.	.	.
7	.	.	.
8	.	.	.
9	.	.	.
10	.	.	.
11	.	.	.
12	.	.	.
13	.	.	.
14	.	.	.
15	.	.	.
16	.	.	.
17	.	.	.
18	.	.	.
19	.	.	.
20	.	.	.

9 of the 10 new loci alone perform slightly better than Yfiler

*Note: Analysis does not include information from DYS460 in this study

Disadvantages of the Y-Chromosome

- Loci are not independent of one another and therefore rare random match probabilities cannot be generated with the product rule; must use haplotypes (combination of alleles observed at all tested loci)

- **Paternal lineages possess the same Y-STR haplotype** (barring mutation) and thus fathers, sons, brothers, uncles, and paternal cousins cannot be distinguished from one another

- Not as informative as autosomal STR results
  - **More like addition (10 + 10 + 10 = 30) than multiplication (10 x 10 x 10 = 1,000)**

Rapidly Mutating (RM) Y-STRs

Trying to separate

close male relatives

Rapidly Mutating Y-STRs

The American Journal of Human Genetics 87, 341–353, September 10, 2010

ARTICLE

Mutability of Y-Chromosomal Microsatellites: Rates, Characteristics, Molecular Bases, and Forensic Implications

Kaye N. Ballantyne,[1] Miriam Goedbloed,[1] Rixun Fang,[2] Onno Schaap,[1] Oscar Lao,[1] Andreas Wollstein,[1,3] Ying Choi,[1] Kate van Duijn,[1] Mark Vermeulen,[1] Silke Brauer,[1,4] Ronny Decorte,[5] Micaela Poetsch,[6] Nicole von Wurmb-Schwark,[7] Peter de Knijff,[8] Damian Labuda,[9] Hélène Vézina,[10] Hans Knoblauch,[11] Rüdiger Lessig,[12] Lutz Roewer,[13] Rafal Ploski,[14] Tadeusz Dobosz,[15] Lotte Henke,[16] Jürgen Henke,[16] Manohar R. Furtado,[2] and Manfred Kayser[1,*]

Manfred Kayser

13 markers evaluated

Forensic Science International: Genetics 6 (2012) 208–218

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

A new future of forensic Y-chromosome analysis: Rapidly mutating Y-STRs for differentiating male relatives and paternal lineages

Kaye N. Ballantyne [a,1,2], Victoria Keerl [a,1,3], Andreas Wollstein [a,b], Ying Choi [a], Sofia B. Zuniga [c], Arwin Ralf [a], Mark Vermeulen [a], Peter de Knijff [c], Manfred Kayser [a,*]

[a] Department of Forensic Molecular Biology, Erasmus MC University Medical Center Rotterdam, 3000 CA Rotterdam, The Netherlands
[b] Cologne Center for Genomics, University of Cologne, D-50674 Cologne, Germany
[c] Department of Human Genetics, Leiden University Medical Center, 2300 RC Leiden, The Netherlands

Using Y-STRs with a higher mutation rate, father-son and brother pairs can sometimes be distinguished

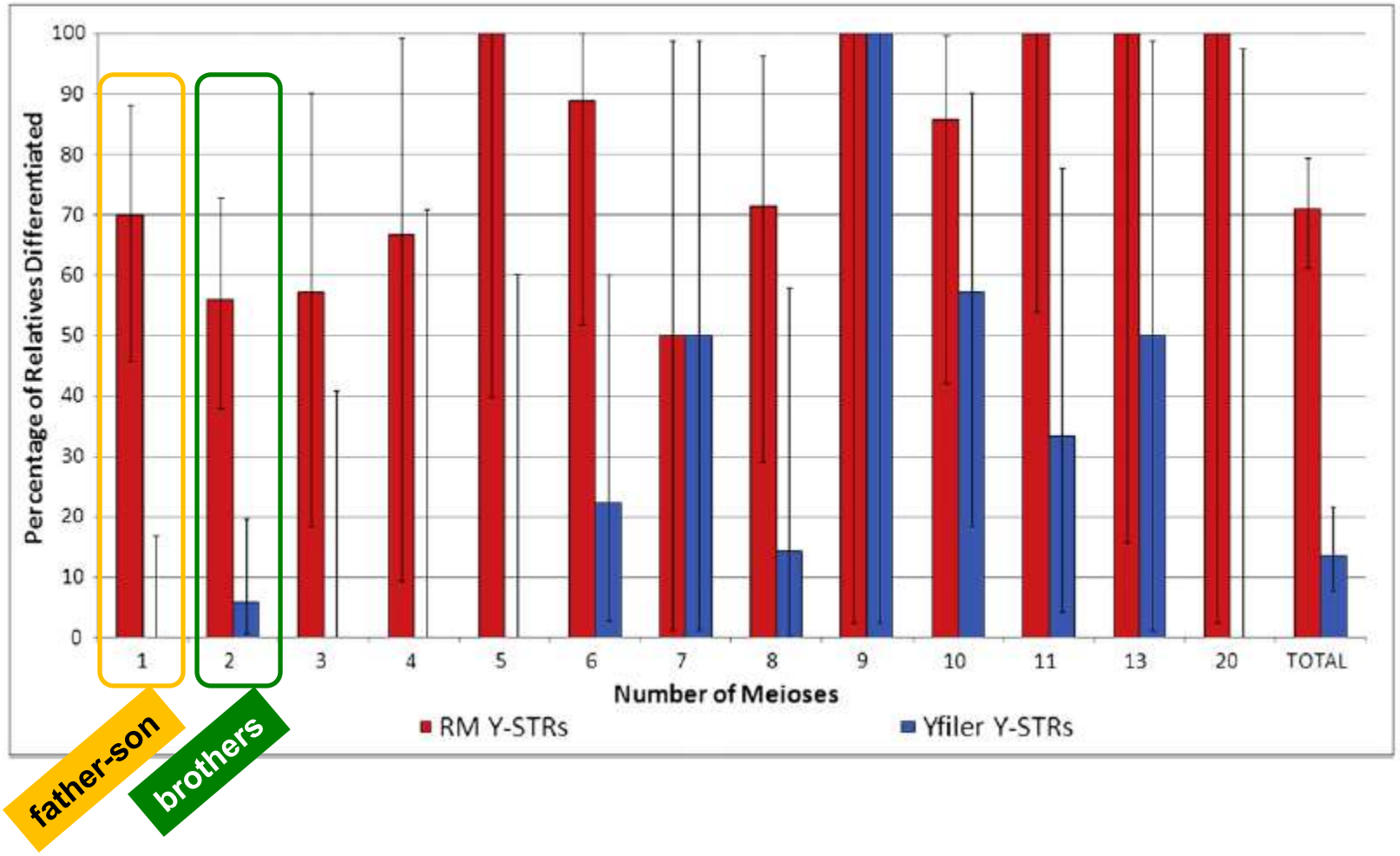


Figure 4. Male Relative Differentiation with Newly Identified 13 RM Y-STRs and Commonly Used 17 Yfiler Y-STRs

Results from differentiating between male relatives from analyzing 103 pairs from 80 male pedigrees, sorted according to the number of generations separating pedigree members, based on 13 RM Y-STRs (in red) and 17 Yfiler Y-STRs (in blue). Error bars represent 95% binomial confidence intervals. Note that these samples are independent from the father-son pairs initially used to establish the Y-STR mutation rates.

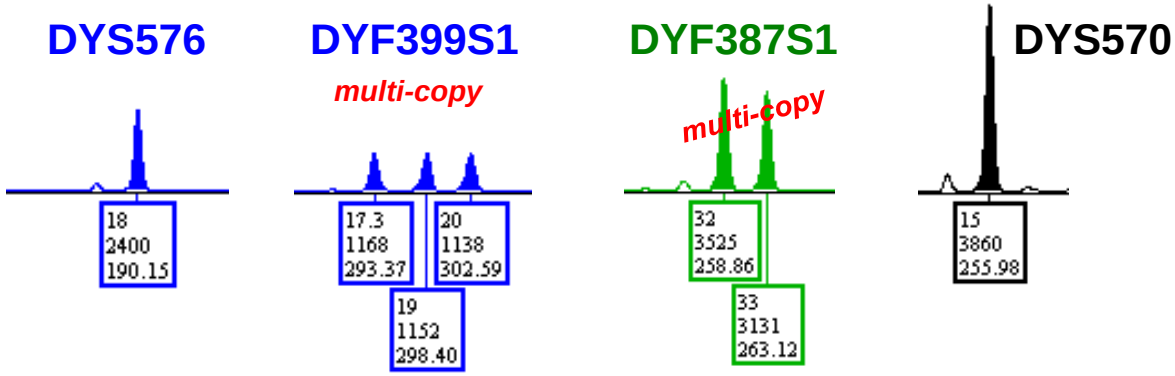
The American Journal of Human Genetics 87, 341–353, September 10, 2010

K. Ballantyne et al. 2010; K. Ballantyne et al. 2012

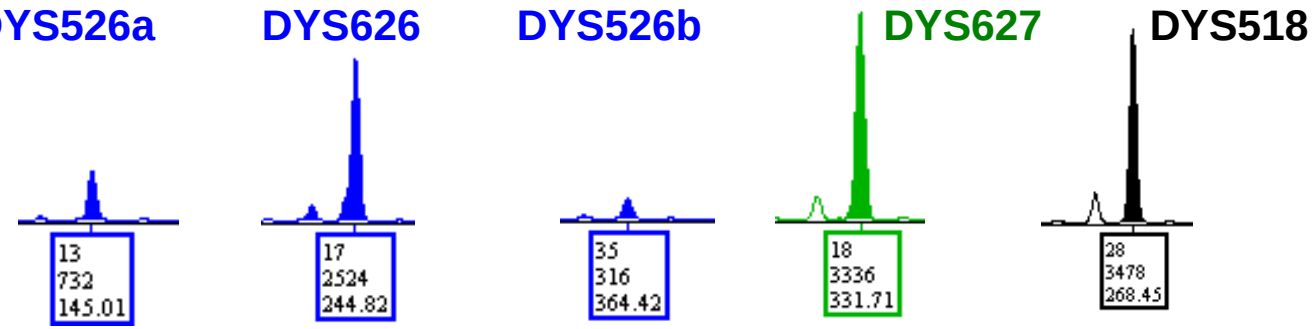
Rapidly Mutating (RM) Y-STRs

NIST supplied data from 1,296 U.S. samples (634 population + 331 father/son pairs)
to RM Y-STR Study Group led by Manfred Kayser

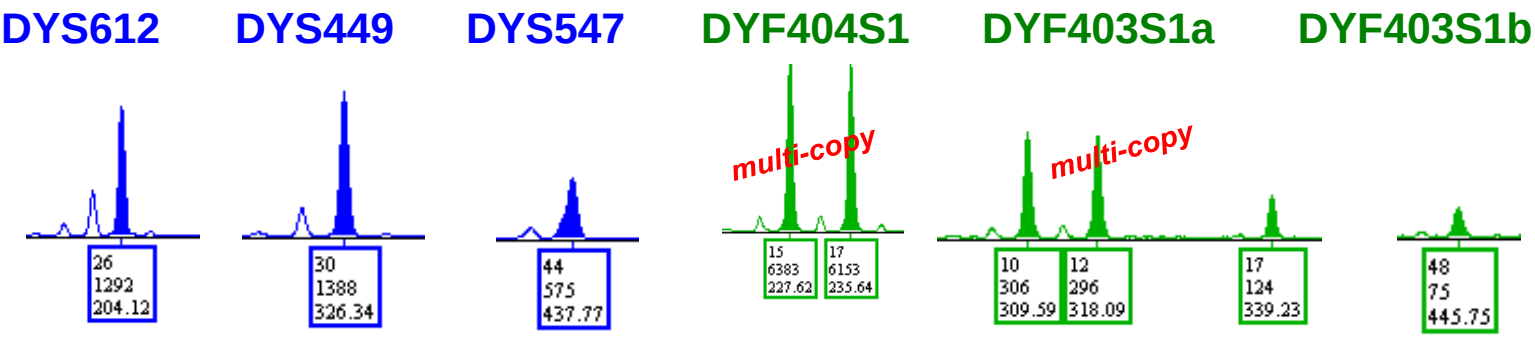
RM Y-STR
Multiplex 1



RM Y-STR
Multiplex 2



RM Y-STR
Multiplex 3

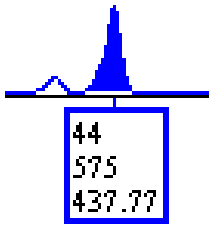


Why do these markers mutate “rapidly”?

Markers in Yfiler Plus

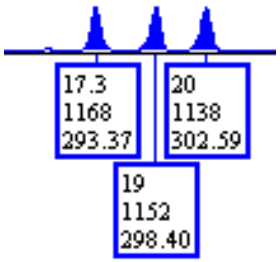
Locus	(average mutation rate)
DYS449	(1.2%)
DYS518	(1.8%)
DYS547	(2.4%)
DYS570	(1.2%)
DYS576	(1.4%)
DYS612	(1.4%)
DYS626	(1.2%)
DYS627	(1.2%)
DYF387S1	(1.6%)
DYF399S1	(7.7%)
DYF403S1 a/b	(3.1/1.2%)
DYF404S1	(1.3%)
DYS526 a/b	(1.3%)

“Large” number of repeats



DYS547

Multi-copy Markers



DYF399S1

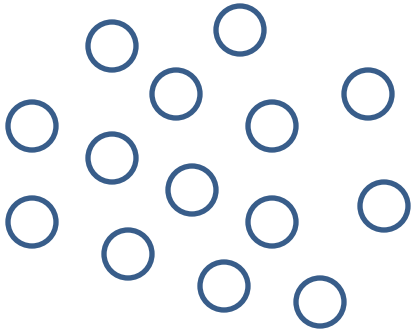
DYS458 (0.64%) is highest in Yfiler loci where average is ~0.2%

Gene Diversity

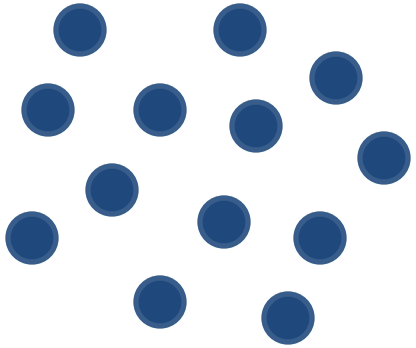
- is a measure of the uniqueness of a particular marker in a given population

$$GD = (1 - \sum_i x_i^2)$$

Relative frequency of each allele



Marker Y



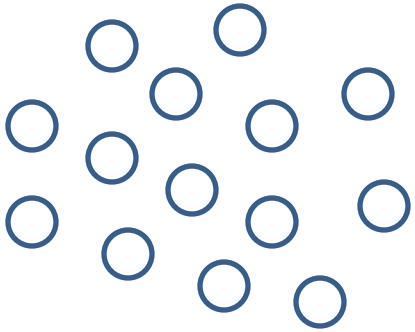
1 type = 100%

N = 100

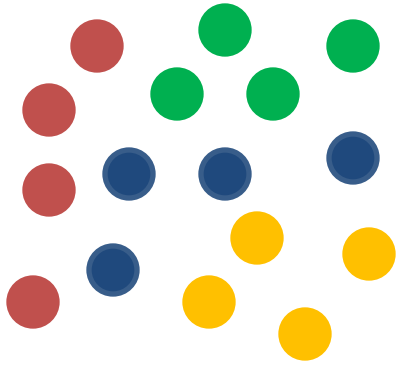
$$GD = (1 - \underbrace{\sum_i x_i^2}_{})$$

0

DC = 1/100 = 0.01



Marker Y



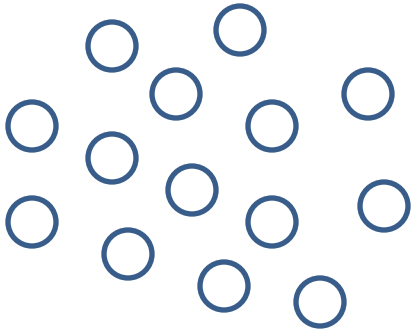
4 types = 25%

N = 100

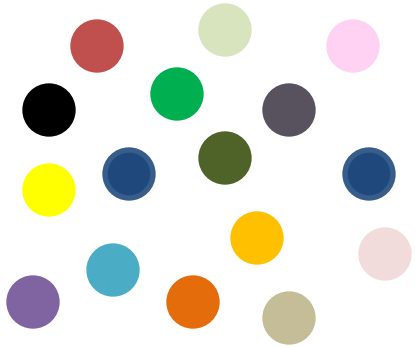
$$GD = (1 - \sum_i x_i^2)$$

0.75

DC = 4/100 = 0.04



Marker Y



100 types = 0%

N = 100

$$GD = (1 - \sum_i x_i^2)$$

0.99

DC = 100/100 = 1.0

Gene Diversity of the Markers

Marker	GD	DC
DYS576	0.766	0.035
DYF399S1	0.993	0.587
DYF387S1	0.870	0.098
DYS570	0.743	0.035
RM-01 (all)	**0.9998**	**0.9764**

Marker	GD	DC
DYS526a/b	0.923	0.138
DYS626	0.794	0.043
DYS627	0.848	0.043
DYS518	0.791	0.039
RM-02 (all)	**0.9985**	**0.8661**

DYS385a/b
GD = 0.929

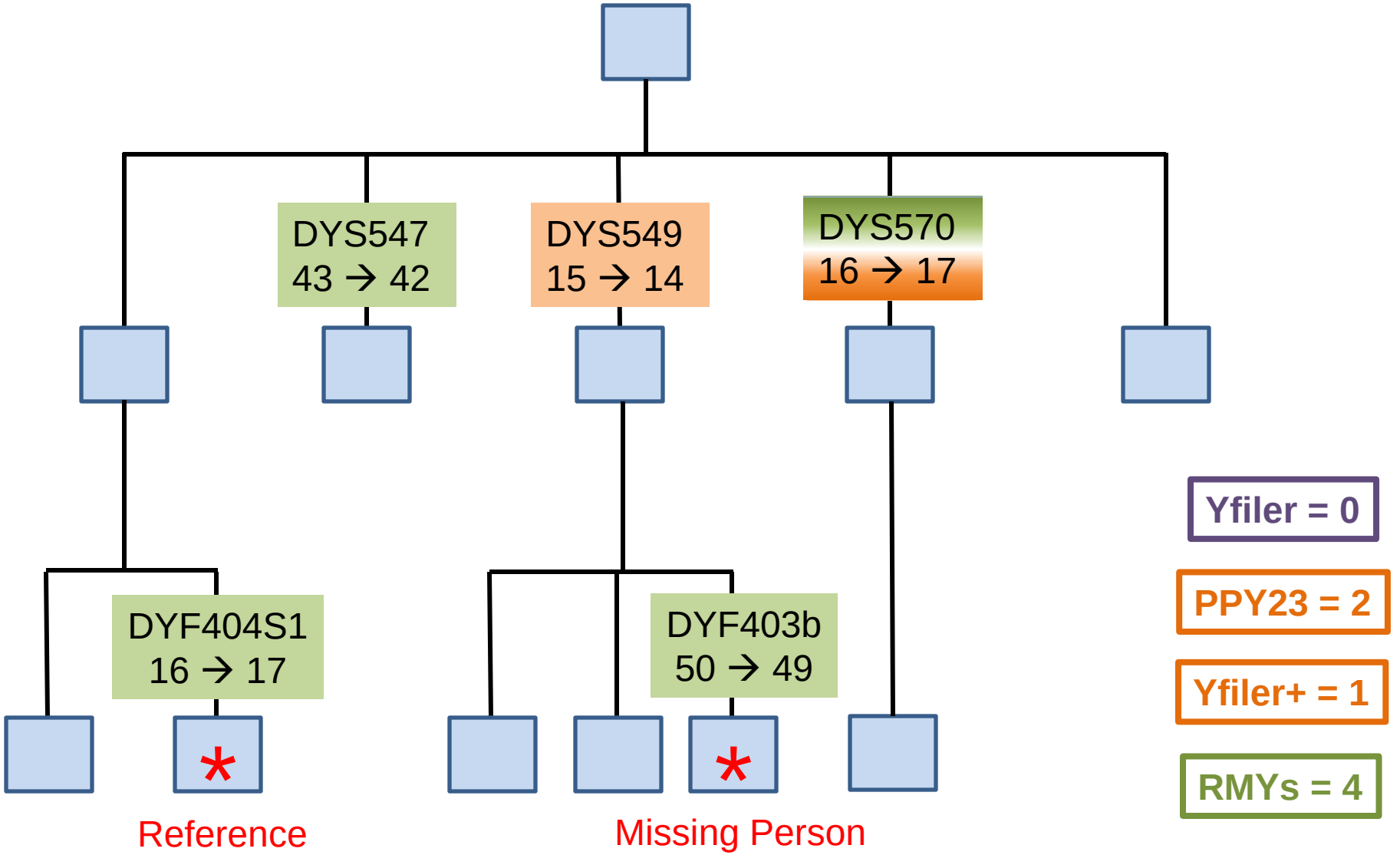
Marker	GD	DC
DYF403S1a/b	0.923	0.791
DYF404S1	0.902	0.110
DYS612	0.832	0.043
DYS449	0.796	0.043
DYS547	0.798	0.039
RM-03 (all)	**1.000**	**0.9984**

All 13 RM Y-STRs
resolved 948 males

Paternal Relatives in the Database

	PPY-23	mtDNA	Kinship Index	RM Mutations
Y27	match	n/a	Father-Son	0
Y28	match	n/a	**254,325,532**	
Y16	match	match	Full Sib	0
Y17	match	match	**155,463**	
ZT79994	match	match	Full Sib	1
ZT79995	match	match	**56,327**	
GT37828	match	C1 (Native)	Cousin	4
C87H	match	n/a	**0.228**	
PT84348	match	L1b (African)	Cousin	3
ZT80369	match	C1 (Native)	**0**	
ZT79304	match	L2a (African)	Cousin	3
PT84253	match	L1b (African)	**0.568**	

Y-STR mutations in a Paternal Lineage



Mutation Rate Information

Meioses	Mutations	Group
63	15	AfAm
89	25	Asian
91	11	Caucasian
88	20	Hispanic
331	71	total
		(21.4%)

+1 Repeat (Son)	-1 Repeat (Son)	
8	6	AfAm
11	13	Asian
5	6	Caucasian
8	12	Hispanic
+2 Repeat (Son)	-2 Repeat (Son)	
0	1	AfAm
1	0	Asian
0	0	Caucasian
0	0	Hispanic

Marker	# of Mutations
DYF399S1	**15**
DYF403S1a/b	**11**
DYS627	**7**
DYS612	**7**
DYS518	**6**
DYS570	**5**
DYS626	**5**
DYS547	**4**
DYS526a/b	**3**
DYS576	**3**
DYS449	**3**
DYF404S1	**1**
DYF387S1	**1**

Interpretational Issues

- We will need to move away from simply “excluding” based upon a set number of discordant markers.
- The Likelihood Ratio can provide weight to the evidence based upon competing propositions.
- This will require information on the ***haplotype frequency*** and ***mutation rate data***.

Relating two deep-rooted pedigrees from Central Germany by high-resolution Y-STR haplotyping

Manfred Kayser [a,*], Mark Vermeulen [a,b], Hans Knoblauch [c], Herbert Schuster [d], Michael Krawczak [e], Lutz Roewer [f]

Forensic Science International: Genetics 1 (2007) 125–128.

Summary

- Rapidly Mutating Y-STRs are highly diverse markers that can discriminate common haplotypes and close relatives.

- These markers may create interpretational issues for paternity/missing persons cases, but LRs can be useful for evaluating these situations.

- An international consortium is gathering frequency and mutation rate data.

Acknowledgments

NIST Team for This Work

John Butler

Becky Hill

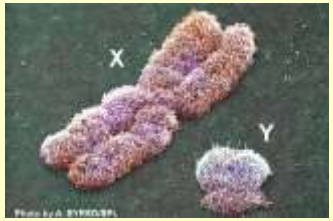
NIST Office of Special Programs

Contact Info: mcoble@nist.gov
301-975-4330

Funding from the **National Institute of Justice (NIJ)** through NIST Office of Law Enforcement Standards

RM-YSTR Consortium

Manfred Kayser



Arwin Ralf

Kaye Ballantyne

ISHI and the Promega Corporation