Promega

Y-STR Analysis Seminar

Variability of Y-STR Marker Sets in the NIST 1036 U.S. Population Samples

John M. Butler, Carolyn R. (Becky) Hill, and Michael D. Coble

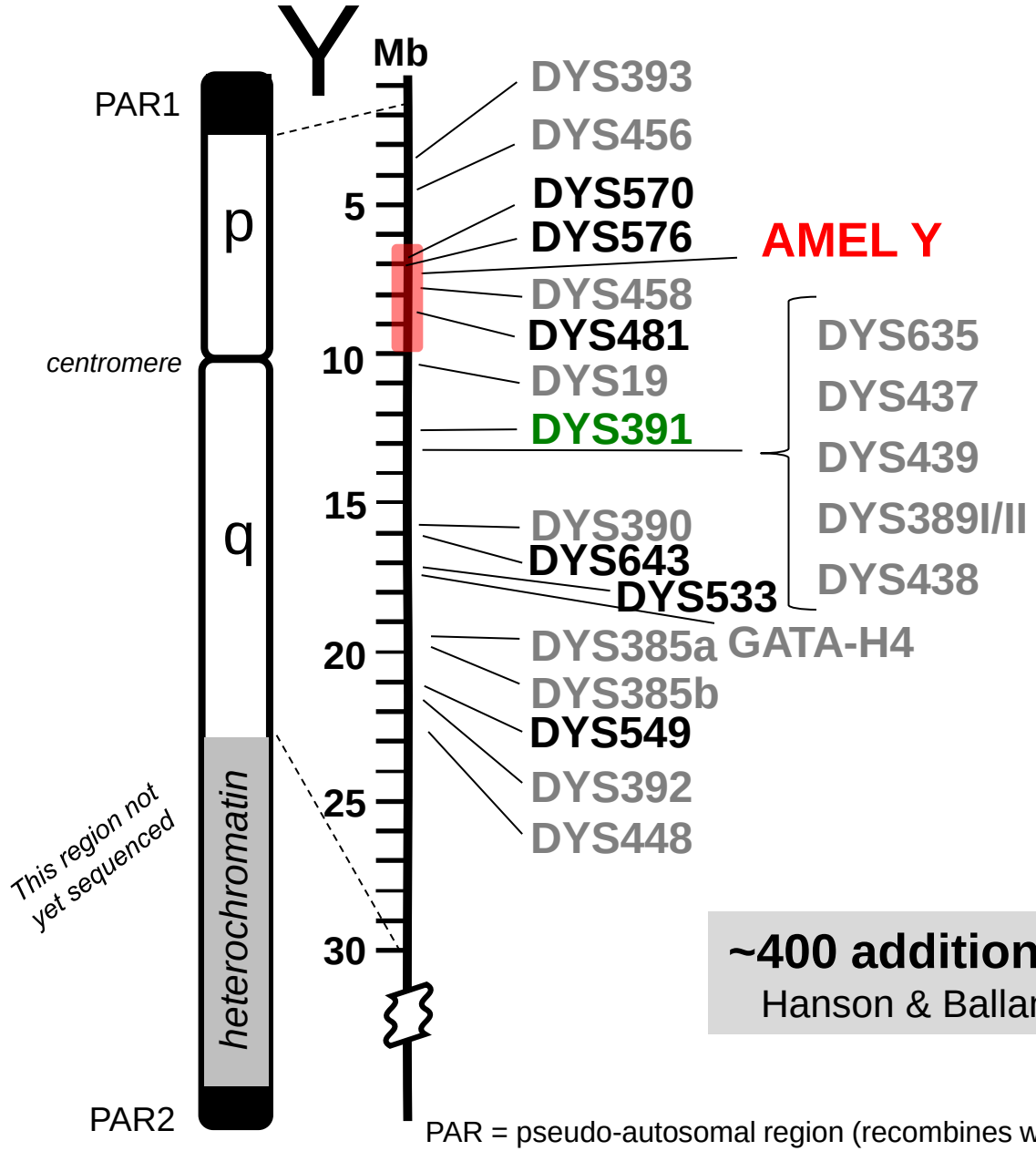
September 24, 2012
Milwaukee, WI

NIST

Presentation Outline

- John (Loci)
  - Information on 23 Y-STR loci
  - Variation observed in NIST 1036 samples
  - Haplotype comparisons (PPY, Yfiler, PPY23)
- Mike (PowerPlex Y23 Kit Experience)
  - Sensitivity studies
  - Stutter results
  - Mutation rates with father/son samples

Relative positions of 23 Y-STR loci available in kits for ChrY testing



Deletions of the Y-chromosome can encompass >1 Mb around the AMEL Y region
(DYS458 is often lost in these situations)

DYS391 has been proposed for inclusion as a future CODIS core locus to address males with AMEL Y deletions
(D.R. Hares, *FSI Genetics*, 2012, 6, e52-e54)

~400 additional Y-STRs currently known
Hanson & Ballantyne, *Legal Med* 2006;8(2):110-20

PAR = pseudo-autosomal region (recombines with X-chromosome)

NIST U.S. Samples (>1450)

- **NIST U.S. population samples**
  - 260 African American, 260 Caucasian, 140 Hispanic, 3 Asian
- **U.S. father/son paired samples**
  - ~**100 fathers/100 sons for each group**: 200 African American, 200 Caucasian, 200 Hispanic, 200 Asian
- **NIST SRM 2391b**, PCR-based DNA Profiling Standard (highly characterized)
  - 10 genomic DNA samples, 2 cell line samples
  - Includes 9947A and 9948
- **NIST SRM 2391c**, PCR-based DNA Profiling Standard
  - 4 genomic DNA (one mixture)
  - 2 cell lines (903 and FTA paper)

Hill, C.R., et al. (2011) Concordance and population studies along with stutter and peak height ratio analysis for the PowerPlex® ESX 17 and ESI 17 Systems. *Forensic Sci. Int. Genet.* 5(4): 269-275.

Publications using NIST Population Samples

Data available at
http://www.cstl.nist.gov/strbase/NISTpop.htm

1. Butler et al. (2003) *J. Forensic Sci.* – Identifiler allele frequencies
2. Butler et al. (2003) *J. Forensic Sci.* – miniSTR assay development
3. Drabek et al. (2004) *J. Forensic Sci.* – miniSTR concordance
4. Schoske et al. (2004) *Forensic Sci. Int.* – Y-STR 20plex & 11plex
5. Vallone et al. (2004) *J. Forensic Sci.* – 50 Y-SNPs
6. Coble & Butler (2005) *J. Forensic Sci.* – NC01 & NC02 assay development
7. Butler et al. (2005) *J. Forensic Sci.* – PowerPlex Y with Y-STR duplications & triplications
8. Vallone et al. (2005) *Forensic Sci. Int.* – 70 autosomal SNPs
9. Butler et al. (2006) *Forensic Sci. Int.* – 27 Y-STR additional loci
10. Hill et al. (2007) *J. Forensic Sci.* – MiniFiler concordance
11. Decker et al. (2008) *FSI Genetics* - Yfiler mutation rates
12. Saunier et al. (2008) *FSI Genetics* – mtDNA control region sequencing (AFDIL)
13. Just et al. (2008) *FSI Genetics* – mtGenome analysis (AFDIL)
14. Hill et al. (2008) *J. Forensic Sci.* – NC01-NC09 miniSTR loci
15. Diegoli et al. (2009) *FSI Genetics* – mtDNA control region sequencing (AFDIL)
16. Hill et al. (2009) *J. Forensic Sci.* – NIST 26plex
17. Lao et al. (2010) *Human Mutation* – 24 ancestry SNPs, Y-SNPs, mtDNA
18. Hill et al. (2011) *FSI Genetics* – ESI 17 & ESX 17 concordance
19. Diegoli et al. (2011) *FSI Genetics Suppl. Ser.* – Argus X-12 X-STR loci
20. Fondevila et al. (2012) *Int. J. Legal Med*. – 68 InDel loci
21. Fondevila et al. (2012) *FSI Genetics* – 34 ancestry SNPs
22. Butler et al. (2012) *Profiles in DNA* – introduces NIST 1036 data set
23. Hill et al. (2012) *FSI Genetics (forthcoming)* – 29 autosomal STRs in PowerPlex CS7 and other kits
24. Coble et al. (2012) *FSI Genetics (forthcoming)* – 23 Y-STRs in PowerPlex Y23

Testing also completed with 16 X-STR loci and 14 rapidly mutating (RM) Y-STRs

NIST 1036 U.S. Population Samples

- 1032 males + 4 females
  - 361 Caucasians (2 female)
  - 342 African Americans (1 female)
  - 236 Hispanics
  - 97 Asians (1 female)

***Unrelated* samples**
All known or potential related individuals (based on autosomal & lineage marker testing) have been removed from the 1036 data set (e.g., only sons were used from father-son samples)

- Anonymous donors with self-identified ancestry
  - Interstate Blood Bank (Memphis, TN) – obtained in 2002
  - Millennium Biotech, Inc. (Ft. Lauderdale, FL) – obtained in 2001
  - DNA Diagnostics Center (Fairfield, OH) – obtained in 2007

- **Complete profiles with 29 autosomal STRs + PowerPlex Y23**
  - **Examined with multiple kits** and in-house primer sets enabling concordance

- Additional DNA results available on subsets of these samples
  - mtDNA control region/whole genome (AFDIL)
  - >100 SNPs (AIMs), 68 InDel markers, X-STRs (AFDIL)
  - NIST assays: miniSTRs, 26plex, >100 Y-STRs, 50 Y-SNPs

Data available on STRBase: **http://www.cstl.nist.gov/biotech/strbase/NISTpop.htm**

Benefits of NIST 1036 Data Set

- **Elimination of potential null alleles due to primer binding site mutations** through extensive concordance testing performed with different PCR primer sets from all available commercial STR kits

- **Ancestry testing performed** on DNA samples with autosomal SNPs, Y-SNPs, and mtDNA sequencing to verify self-declared ancestry categorization

- **Related individuals removed** based on Y-STR and mtDNA results

Example of Related Individuals in Original NIST Data Set

- **Hispanic samples ZT79994 and ZT79995**

- Out of 24 autosomal STR loci, these samples **share a total of 22 alleles at 22 loci** (only D12S391 and Penta D have non-overlapping heterozygous alleles)
- **Full 23 Y-STR match** with PowerPlex Y23
- **Same mtDNA control region sequences**
- Kinship calculations
  - LR = 0 for parent-child
  - **LR = 56,300 for full-siblings (brothers)**
  - LR = 5,690 for half-siblings (or uncle-nephew, grandfather-grandson)
  - LR = 264 for first cousins

- ***Decision***: **Remove ZT79995 from final data set**
  - ZT79994 represents this individual's family in NIST 1036

Used PPY23 Information to Sort Out Potential Relatives in Our Data Set

Zero Differences in PPY23

Y27 and Y28 are most likely Father and Son (PI = 254 million). With no additional knowledge of the samples (e.g. ages of contributors) we may assumed that Y27 is the Father and Y28 is the son - pulled Y27 from the dataset.

Y16 and Y17 are most likely full-sibs (KI = 155 thousand, same mtDNA base composition) – pulled Y17

ZT79994 and ZT79995 are most likely full-sibs (KI = 56 thousand, same mtDNA haplotype) – pulled ZT79995

One mismatch in PPY-23

ZT79338 and ZT79339 are most likely full-sibs (KI = 33.5 trillion, same mtDNA haplotype) – pulled ZT79339

Other samples…

Analyzed the other 0 and 1 difference sample matches and found no other substantial kinship values. A few of the samples had "1st cousin" KIs up to 3.9, but this is weak evidence.

Also compared the amount of allele sharing in 2 difference matches, and found that most of the time there were about 10-12 (range = 6-15) markers that one or two shared alleles out of the 24 markers examined, so it is very unlikely that these would give significant KIs for close relatives.

Characteristics for PowerPlex Y23 Y-STR Loci

STR Marker	Position (Mb) GRCh37 (Feb 2009)	Repeat Motif	PPY23 Ladder Allele Range	# Observed Alleles in NIST 1036	Probability of Identity in NIST 1036
DYS393	3.13	AGAT	7 - 18	6	0.5018
DYS456	4.27	AGAT	11 - 23	8	0.3098
DYS570	6.86	TTTC	10 - 25	12	0.2161
DYS576	7.05	AAAG	11 - 23	9	0.1922
DYS458	7.87	GAAA	10 - 24	15	0.2166
DYS481	8.43	CTT	17 - 32	16	0.1670
DYS19	9.52	TAGA	9 - 19	10	0.2996
DYS391	14.10	TCTA	5 - 16	7	0.4758
DYS635	14.38	TSTA	15 - 28	11	0.2421
DYS437	14.47	TCTR	11 - 18	8	0.3800
DYS439	14.51	AGAT	6 - 17	11	0.3460
DYS389 I/II	14.61	TCTR	9 - 17 / 24 - 35	8 / 12	0.4203 / 0.2541
DYS438	14.94	TTTTC	6 - 16	7	0.2861
DYS390	17.27	TCTR	17 - 29	8	0.2242
DYS643	17.43	CTTTT	6 - 17	10	0.2373
DYS533	18.39	ATCT	7 - 17	7	0.3654
GATA-H4	18.74	TAGA	8 - 18	6	0.3957
DYS385 a/b	20.80, 20.84	GAAA	7 - 28	17 (69 combinations)	0.0700
DYS549	21.52	GATA	7 - 17	7	0.3095
DYS392	22.63	TAT	4 - 20	11	0.3752
DYS448	24.36	AGAGAT	14 - 24	16	0.2650

Some information from J.M. Butler (2012) *Advanced Topics in Forensic DNA Typing: Methodology*, Table 13.2

Probability of Identity

- The probability of identity (P_I), also referred to as the matching probability, is **the chance that two unrelated people selected at random will have the same genotype** (first described by George Sensabaugh in 1982). The P_I value of a single locus is determined by summing the square of the observed genotype frequencies.

- **Lower P_I values indicate more variability** with the genetic marker in the measured population because there are more genotypes occurring at a lower frequency.

- P_I values from <u>independently inherited loci can be multiplied together</u> to produce an expected profile P_I

Sensabaugh, G.F. (1982) Biochemical markers of individuality. In Saferstein, R. (ed.) *Forensic Science Handbook,* Prentice-Hall: New York, pp. 338-415.

Y-STR Locus Variability

(Probability of Identity values from NIST 1036)

Y-STR locus variability across 1032 males in our data set (loci are rank ordered by their probability of identity values). There are 17 different alleles seen in our data set for DYS385 in 69 combinations of the “a” and “b” amplicons. Results from the six new loci present in PowerPlex Y23 are highlighted.

Locus	Alleles Observed	P_I (total)	P_I (Cauc)	rank	P_I (AfAm)	rank	P_I (Hisp)	rank	P_I (Asian)	rank
DYS385a/b	17 (69)	0.0700	0.1482	(1)	0.0608	(1)	0.0815	(1)	0.0549	(1)
DYS481	16	0.1670	0.2653	(6)	0.1401	(2)	0.2033	(4)	0.1765	(2)
DYS576	9	0.1922	0.2292	(2)	0.1917	(3)	0.2020	(3)	0.2392	(7)
DYS570	12	0.2161	0.2575	(4)	0.2085	(4)	0.2008	(2)	0.1895	(4)
DYS458	15	0.2166	0.2352	(3)	0.2410	(6)	0.2144	(5)	0.1825	(3)
DYS390	8	0.2242	0.2966	(7)	0.3233	(11)	0.3605	(16)	0.2480	(8)
DYS643	10	0.2373	0.3790	(12)	0.2130	(5)	0.3292	(13)	0.2639	(9)
DYS635	11	0.2421	0.3313	(9)	0.2778	(9)	0.2604	(6)	0.2769	(11)
DYS389II	12	0.2541	0.3157	(8)	0.2623	(7)	0.2711	(7)	0.2054	(5)
DYS448	16	0.2650	0.3904	(13)	0.2789	(10)	0.2982	(8)	0.2739	(10)
DYS438	7	0.2861	0.3942	(14)	0.4311	(18)	0.3046	(9)	0.4201	(16)
DYS19	10	0.2996	0.4860	(21)	0.2746	(8)	0.3308	(14)	0.2384	(6)
DYS549	7	0.3095	0.3656	(11)	0.3325	(12)	0.3354	(15)	0.3540	(14)
DYS456	8	0.3098	0.2629	(5)	0.3571	(14)	0.3072	(10)	0.4322	(20)
DYS439	11	0.3460	0.3627	(10)	0.3502	(13)	0.3124	(11)	0.4004	(15)
DYS533	7	0.3654	0.4119	(16)	0.3775	(15)	0.3857	(19)	0.4220	(18)
DYS392	11	0.3752	0.4024	(15)	0.5546	(21)	0.3170	(12)	0.3445	(13)
DYS437	8	0.3800	0.4289	(18)	0.4950	(20)	0.3778	(17)	0.6248	(21)
Y-GATA-H4	6	0.3957	0.4286	(17)	0.3842	(16)	0.3983	(20)	0.4203	(17)
DYS389I	8	0.4203	0.4546	(20)	0.4808	(19)	0.3805	(18)	0.3232	(12)
DYS391	7	0.4758	0.4430	(19)	0.5603	(22)	0.4411	(21)	0.6992	(22)
DYS393	6	0.5018	0.6020	(22)	0.4169	(17)	0.5582	(22)	0.4231	(19)

From forthcoming *Profiles in DNA* article

N = 1032 males

	PowerPlex Y	Yfiler	PowerPlex Y23
# haplotypes	**891**	**1013**	**1029**
discrimination capacity	0.863	0.982	0.997
# times haplotype observed	PPY (12 loci)	Yfiler (17 loci)	PPY23 (23 loci)
1	821	998	1026
2	41	12	3
3	16	2	.
4	6	1	.
5	2	.	.
6	2	.	.
7	1	.	.
8	.	.	.
9	1	.	.
10	.	.	.
11	.	.	.
12	.	.	.
13	.	.	.
14	.	.	.
15	.	.	.
16	.	.	.
17	.	.	.
18	.	.	.
19	1	.	.

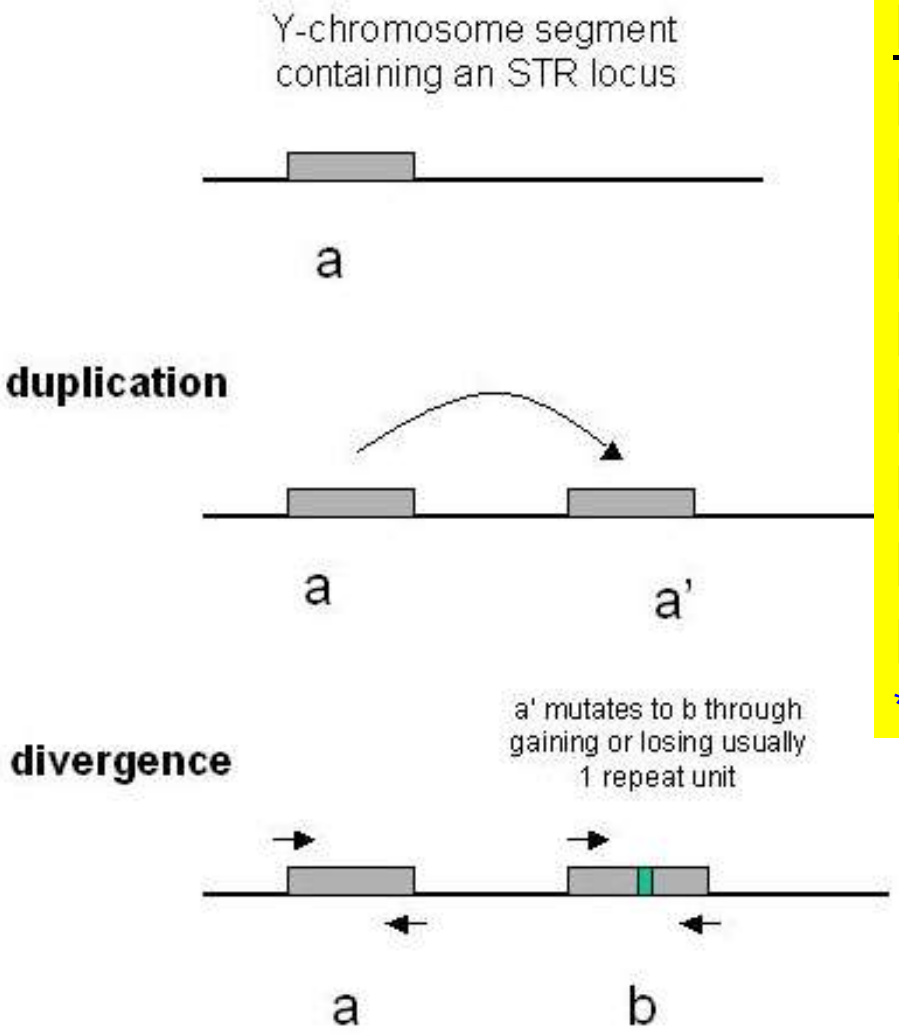
Number of unique and shared haplotypes observed with various combinations of Y-STR loci across 1032 U.S. population samples

1026 PPY23 haplotypes occur once; and
3 sets of sample pairs cannot be resolved from one another

From forthcoming *Profiles in DNA* article

Y-STR Locus Duplication, Triplication, or Deletion

Duplication and Divergence Model



Locus	# dup*	>1 repeat
DYS19	23	2
DYS389I	5	0
DYS389II	9	2
DYS390	1	0
DYS391	3	1
DYS392	0	0
DYS393	3	0
DYS385a/b	17	0

*from www.yhrd.org, literature, and our work

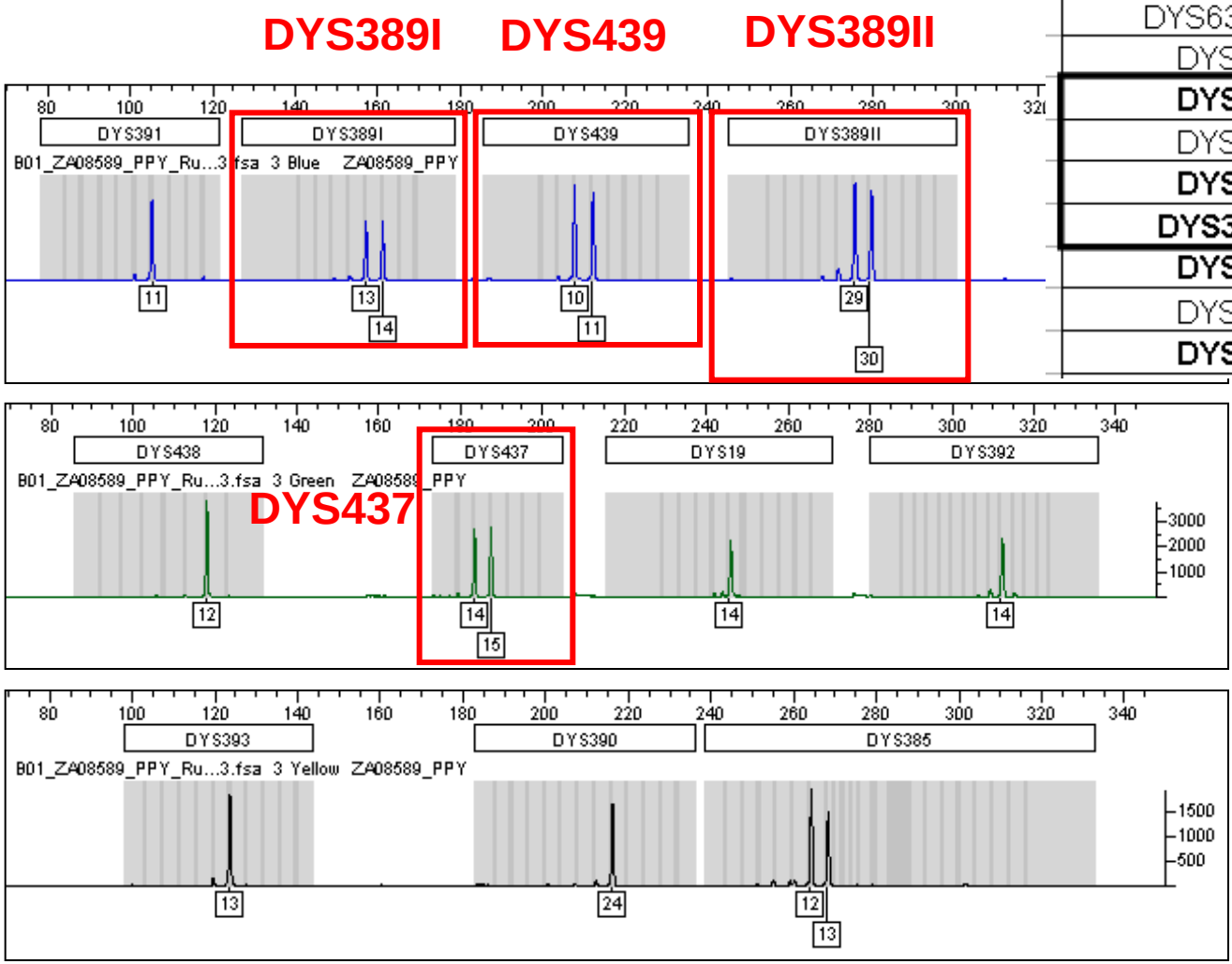
92% have single repeat difference

Since single-step mutations are most common, then single repeat spacing in duplicated alleles is expected

Butler et al. (2005) Chromosomal duplications along the Y-chromosome and their potential impact on Y-STR interpretation *J. Forensic Sci.* 50(4): 853-859

Duplication at Multiple Loci with Single-Source Sample

PowerPlex Y data



Y-chromosome mapping

q-arm

Y STR Marker	Position (Mb)
DYS391	13.413
DYS635 (C4)	13.690
DYS434	13.777
DYS437	13.778
DYS435	13.807
DYS439	13.826
DYS389 I/II	13.923
DYS388	14.057
DYS442	14.071
DYS438	14.248

Entire region of Y-chromosome has likely been duplicated and then diverged

Most duplications have a single repeat spread in allele patterns

Butler et al. (2005) Chromosomal duplications along the Y-chromosome and their potential impact on Y-STR interpretation *J. Forensic Sci.* 50(4): 853-859

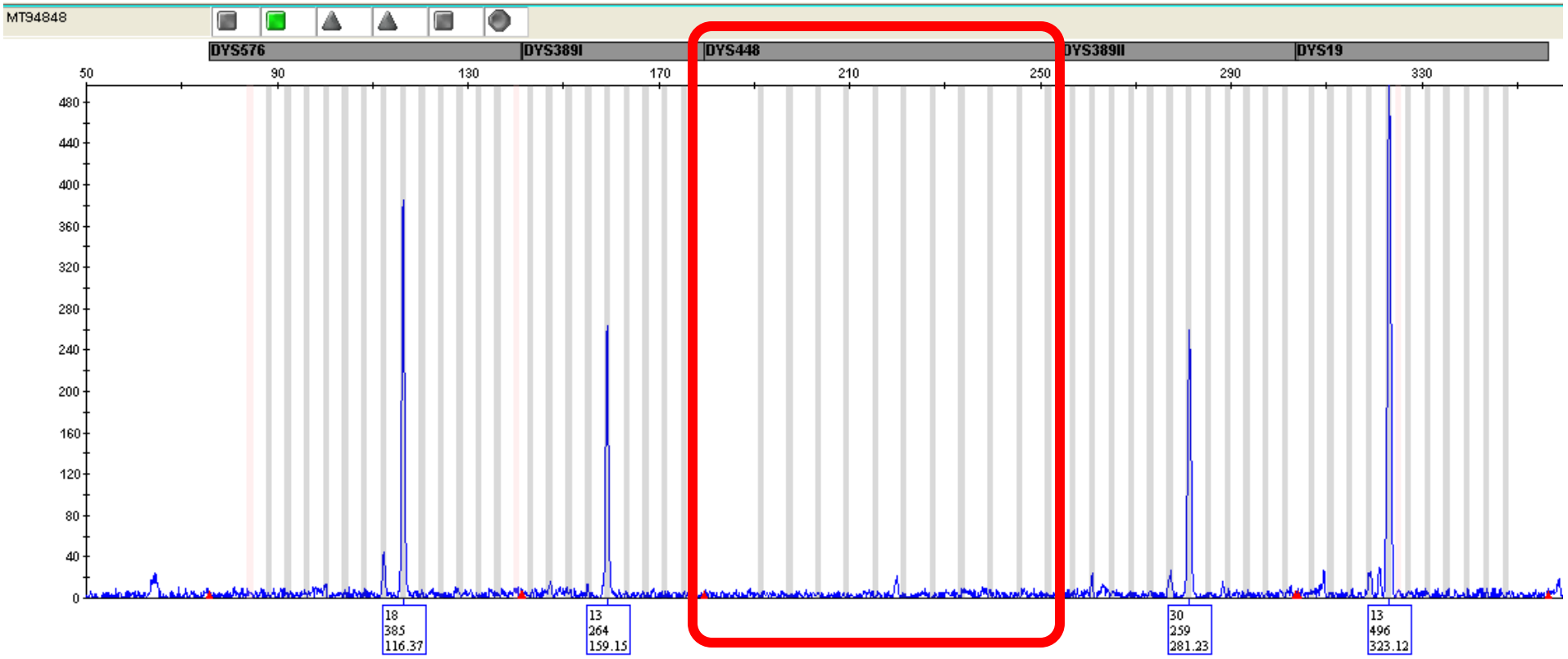
Y-STR Deletions, Duplications and Triplications
in Our NIST 1036 Sample Set using PowerPlex Y23

Sample	Pop	DYS19	DYS385a/b	DYS389I	DYS389II	DYS390	DYS391	DYS392	DYS393	DYS437	DYS438	DYS439	DYS448	DYS456	DYS458	DYS481	DYS533	DYS549	DYS570	DYS576	DYS635	DYS643	GATA-H4
C22B	AfAm	15	14,15	12	28	22	10	11	12	16	9	12	**19,20**	12	15	24	12	12	13	15	21	12	11
C33B	AfAm	14	11,14	N	N	23	11	13	14	15	12	N	19	15	17	22	12	14	17	17	23	10	12
C56B	AfAm	**14,15**	12,17	13	29	22	10	11	12	14	10	12	20	15	21	26	10	13	16	16	21	9	11
C97B	AfAm	15	15,16	12	29	22	10	11	13	17	10	13	**17.2,19,20**	14	12	25	12	12	18	17	21	12	12
MT95371	AfAm	15	13,15	12	29	22	11	11	13	17	10	11	**19,20**	15	16	27	14	14	16	18	21	12	12
OT05599	AfAm	15	13,15	12	29	22	10	12	13	16	9	12	**18,20**	15	16	27	11	12	17	19	19	13	11
PT83899	AfAm	**14,15,17**	13,14	14	30	22	10	11	**13,14**	17	10	10	21	14	19	22	10	11	18	18	22	11	11
PT83913	AfAm	15	15,16	12	29	22	10	11	12	18	10	12	**20,21**	16	15	25	12	13	17	16	19	12	11
PT83978	AfAm	15	14,17	13	30	22	10	11	13	16	9	12	**19,20**	15	15	26	14	12	19	14	20	13	11
PT84178	AfAm	17	14,15	12	29	23	10	12	13	16	8	11	**19,20**	15	19	27	12	13	15	19	20	12	12
C84A	Asian	13	13,15	13	29	25	10	11	14	14	10	13	N	15	18	25	12	12	20	20	21	11	11
MT94848	Cauc	13	16,16	13	30	24	10	11	13	14	10	12	N	17	15	22	12	14	20	18	21	11	13
MT94869	Cauc	14	11,14	13	30	24	10	13	13	15	12	12	19	15	18	**22,23**	13	13	17	18	24	10	13
MT97196	Cauc	14	11,14	13	29	23	11	13	13	15	12	12	21	17	17	**22,23**	12	13	15	17	24	10	11
UT57318	Cauc	14	11,14	14	30	24	11	13	13	15	12	**12,13**	18	16	17	22	14	13	18	17	23	10	12
WT51355	Cauc	15	11,14	14	30	24	11	13	13	15	12	13	19	16	17	**22,23**	13	12	18	18	23	10	12
WT52477	Cauc	14	18,18	13	**30,31**	24	10	11	13	14	10	**11,12**	20	15	17	22	12	12	19	17	21	11	13
Y4	Cauc	14	11,14	13	30	24	11	13	13	15	12	12	19	17	16	22	**12,13**	12	18	18	23	9	12
C53H	Hisp	12	12,14	13	29	24	10	13	13	**15,16**	12	11	20	16	16	23	12	13	18	17	23	9	12
OT07280	Hisp	14	11,14	12	29	24	11	**11,12**	13	15	12	13	19	15	19	22	12	13	18	16	24	10	12
ZA08589	Hisp	14	12,13	**13,14**	**29,30**	24	11	14	13	**14,15**	12	**10,11**	20	15	18	23	12	12	18	18	23	9	12
ZT80682	Hisp	14	12,13	13	**29,30**	23	11	13	13	16	12	11	19	15	16	24	12	12	18	19	23	10	11

Null Allele at DYS448

(Caucasian sample MT94848)

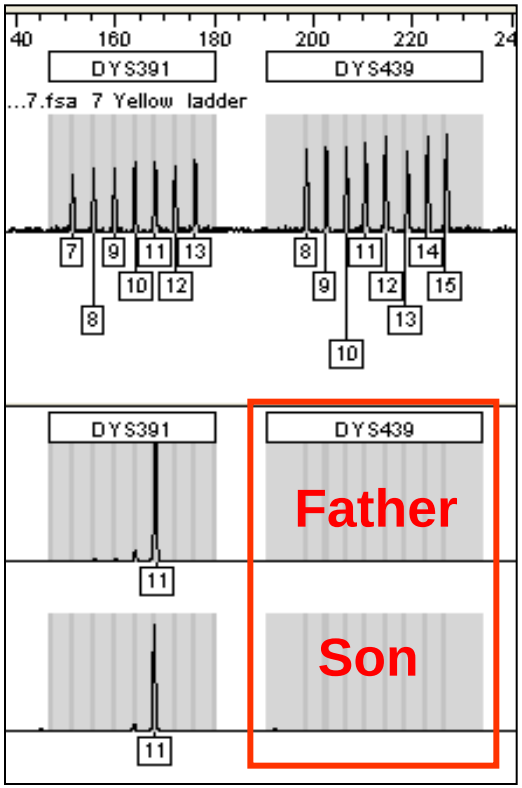
PowerPlex Y23 Blue Channel



Inheritance of DYS439 Deletion

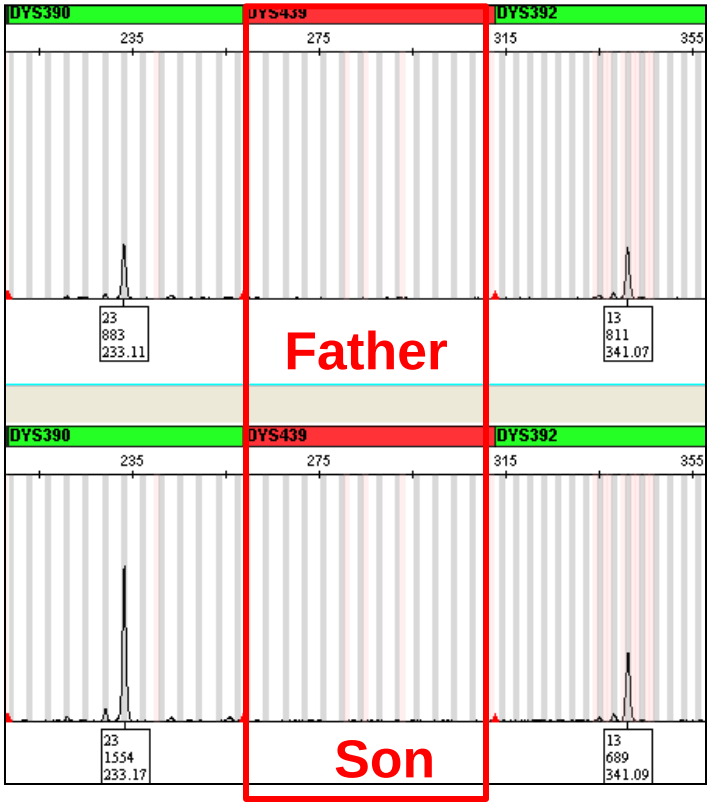
Seen in Both Father and Son

Yfiler data



DYS439

PowerPlex Y23 data

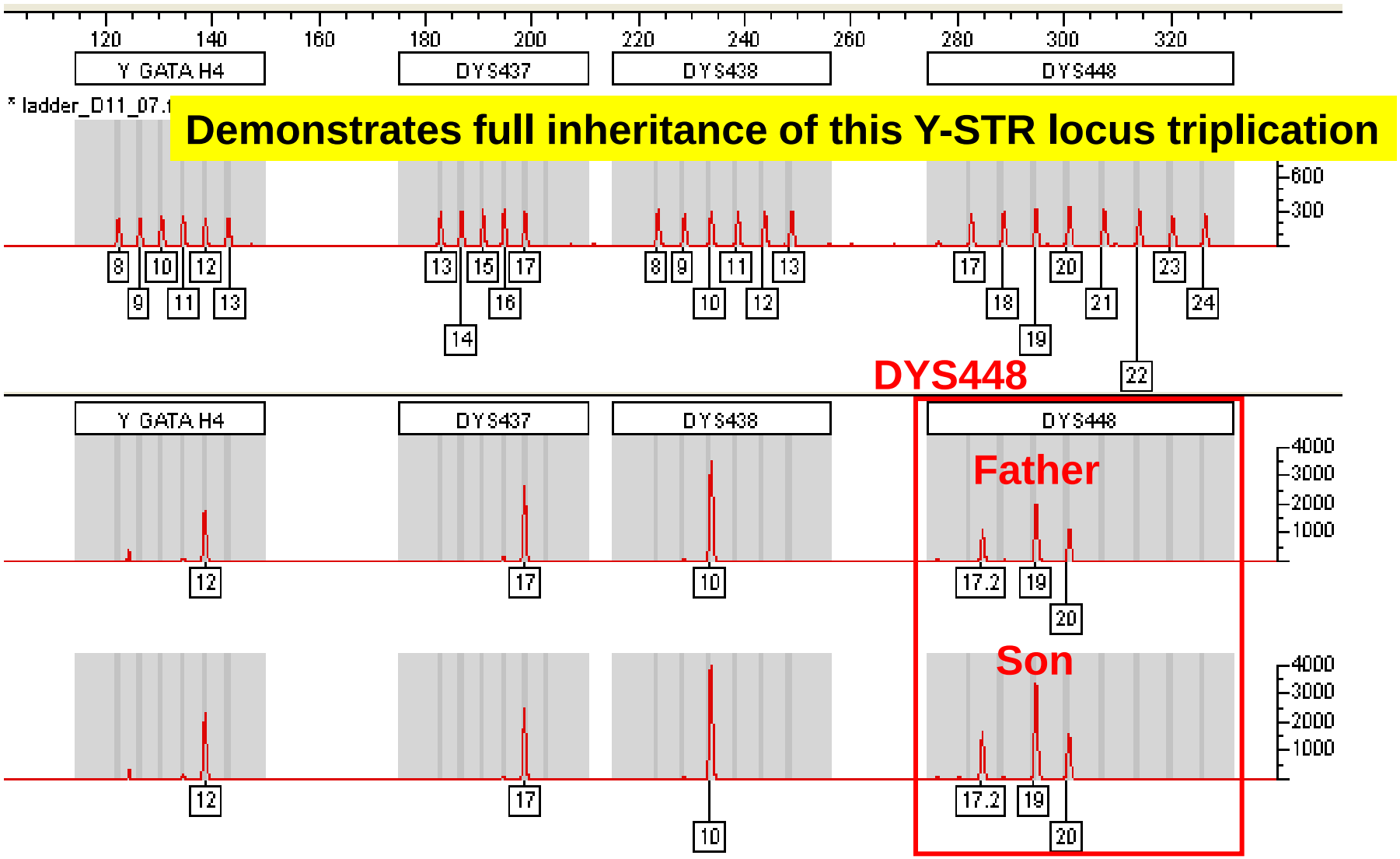


DYS439

Yfiler result

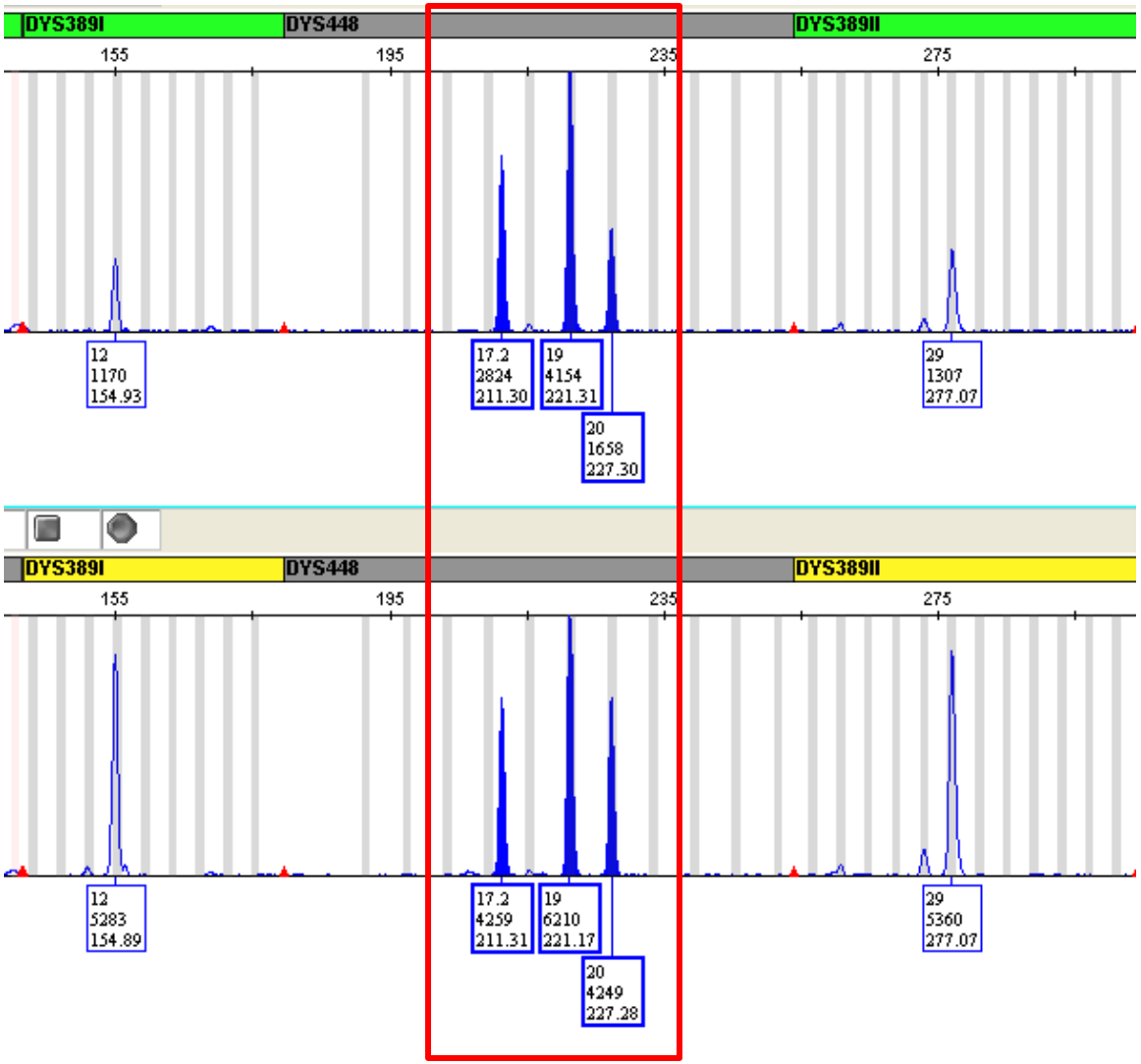
DYS448 Triplication

Seen in Both Father and Son



DYS448 Triplication

Seen in Both Father and Son



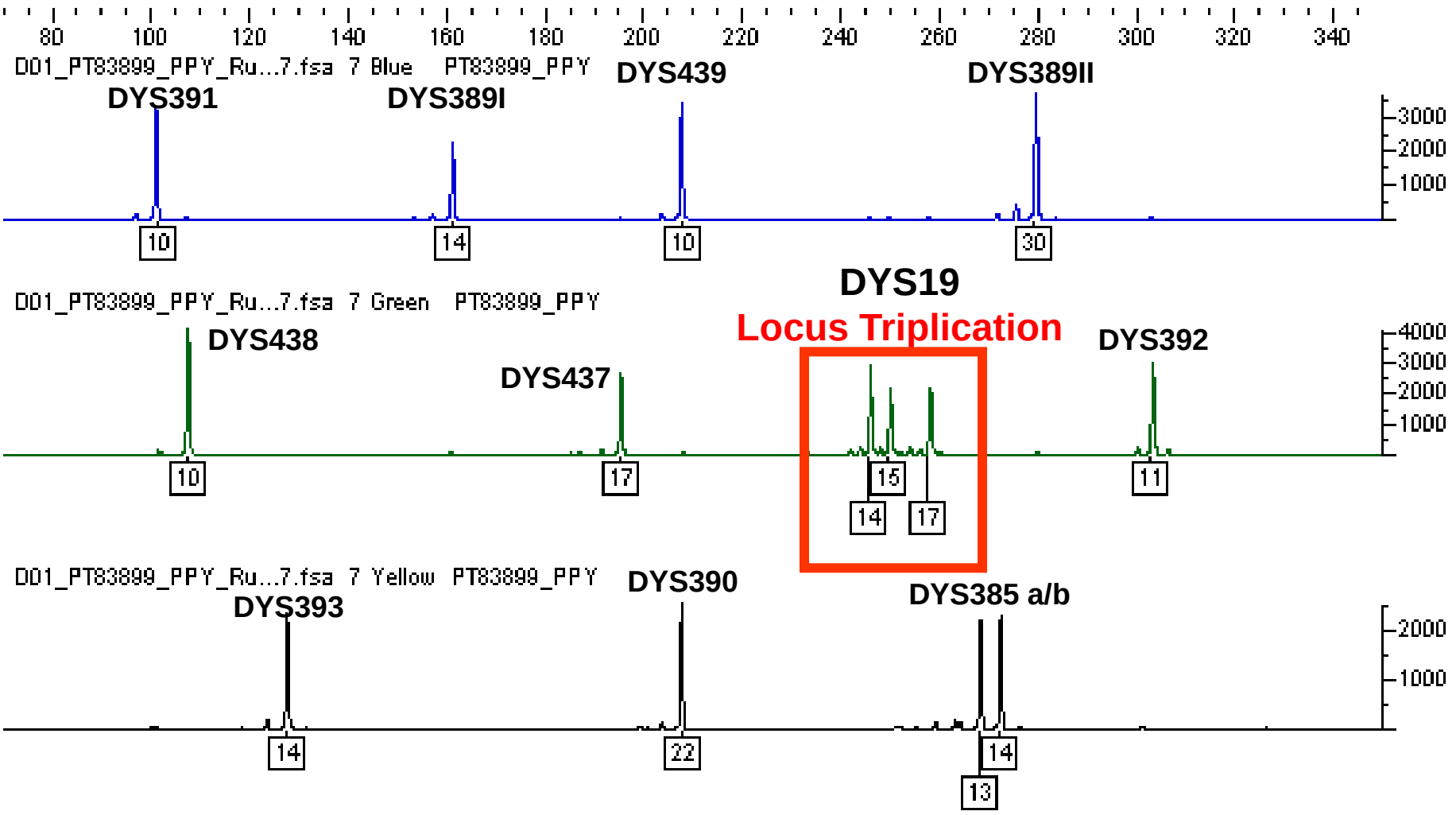
Father

Son

African American samples AF97B and C97B

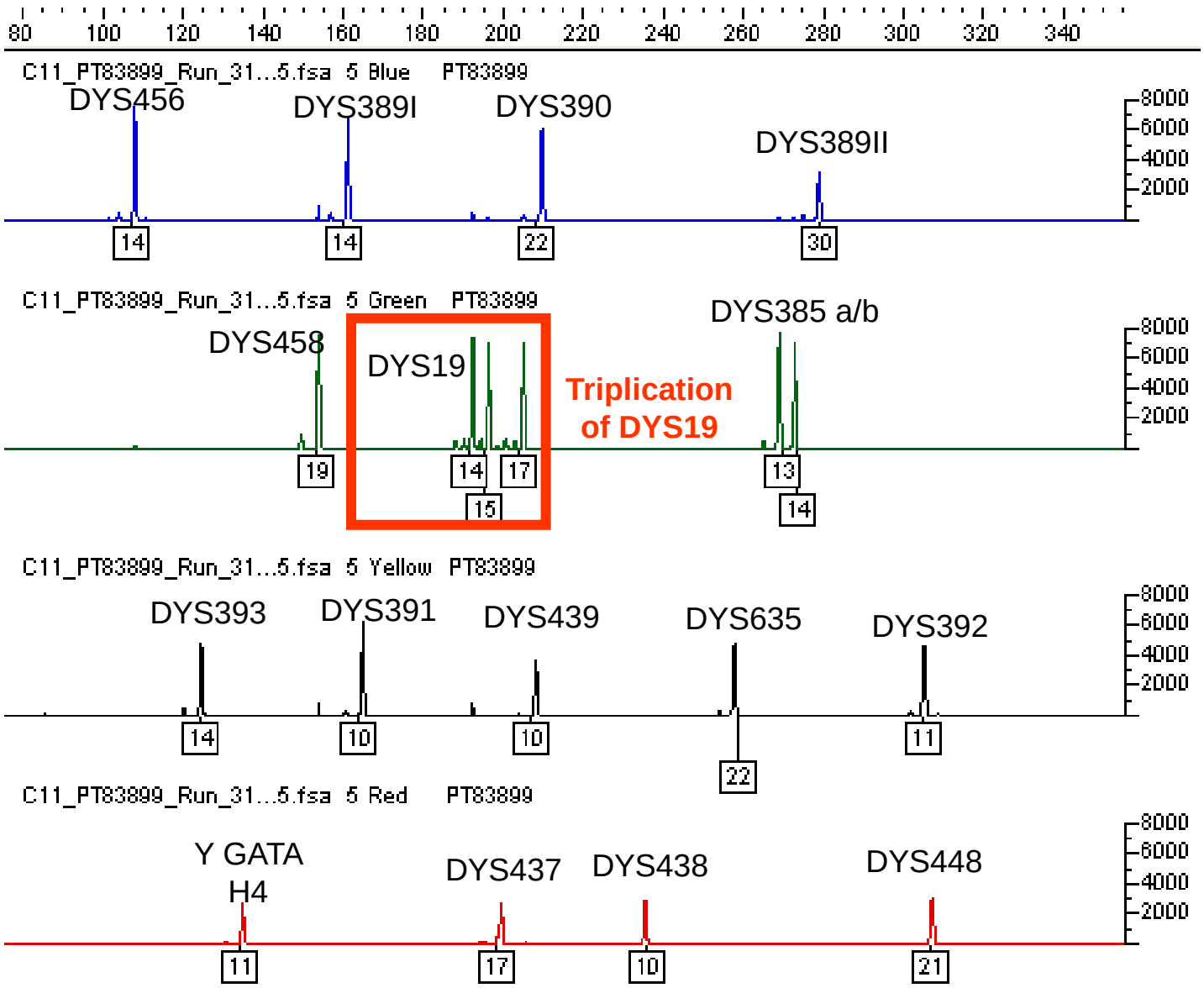
PowerPlex Y

DYS19 Locus Triplication



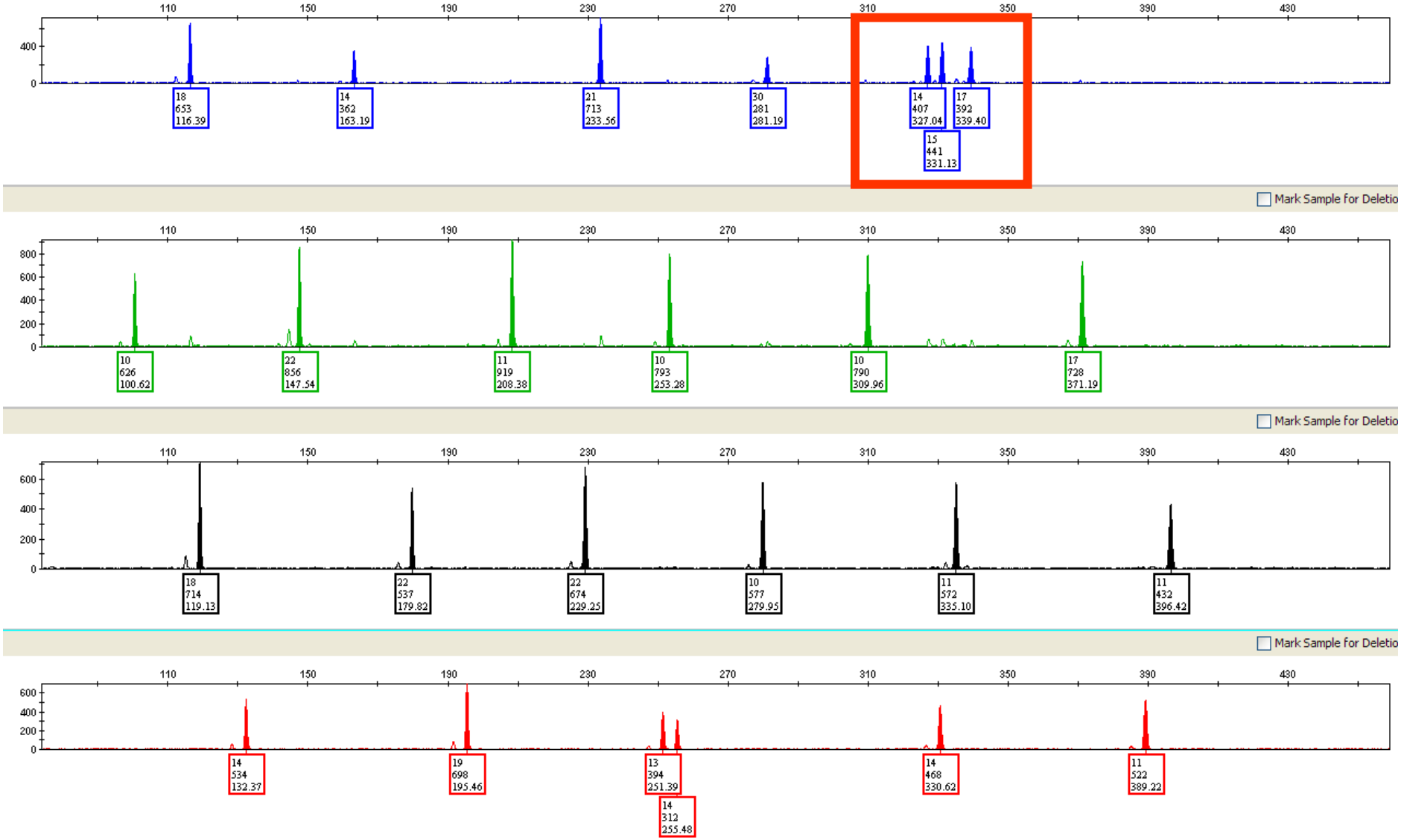
Yfiler

DYS19 Locus Triplication

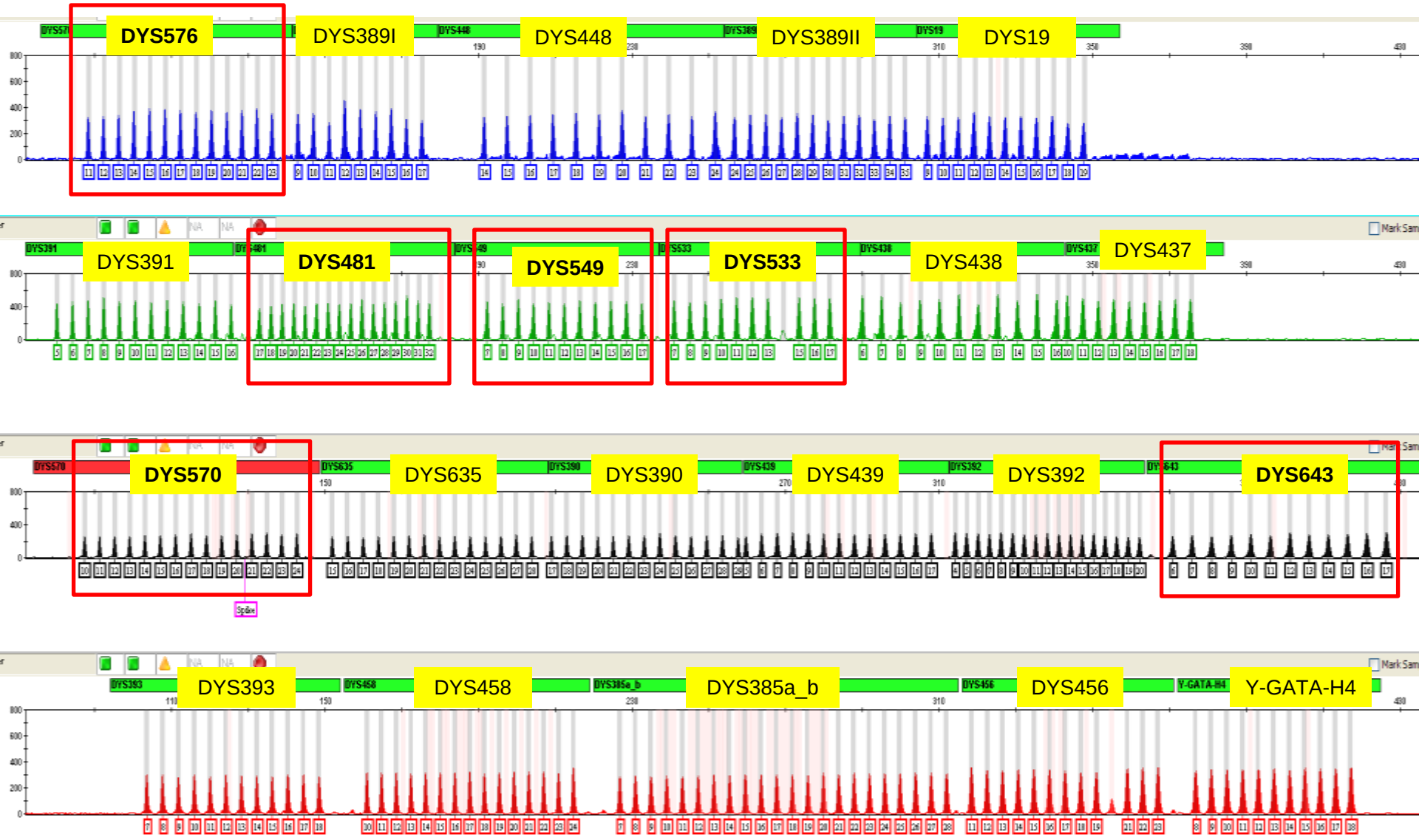


PowerPlex Y23

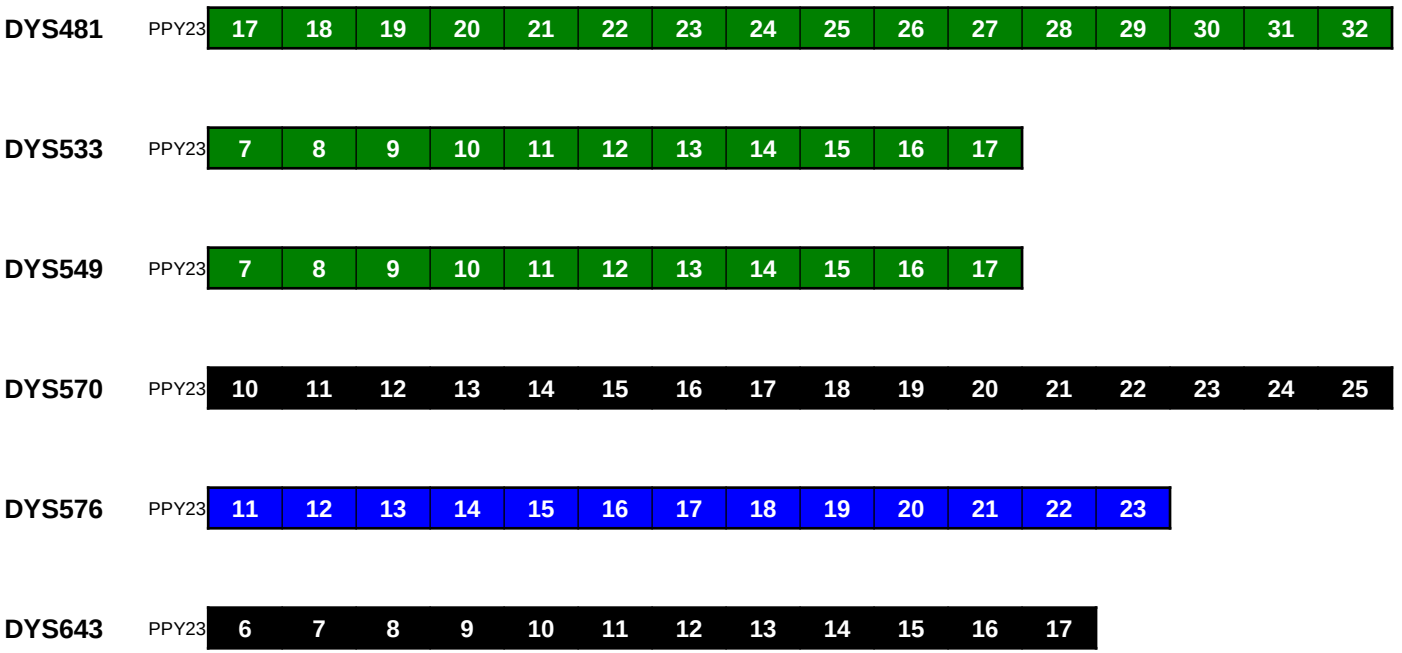
DYS19 Locus Triplication



PowerPlex Y23 Allelic Ladders



Allelic Ladder Alleles for Six New Loci



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

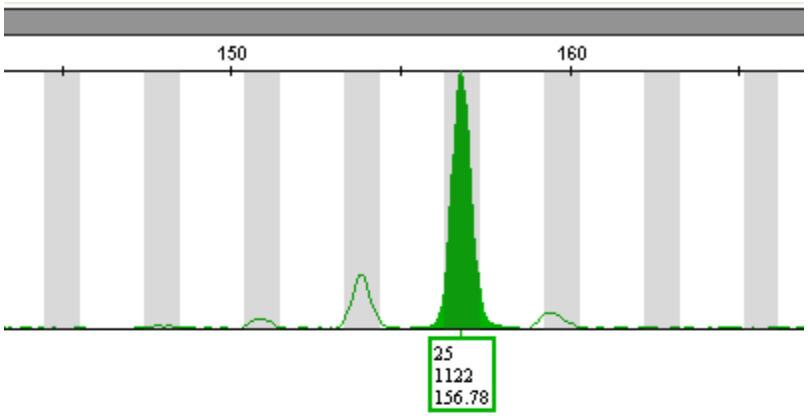
Variant alleles observed

DYS481 25.1

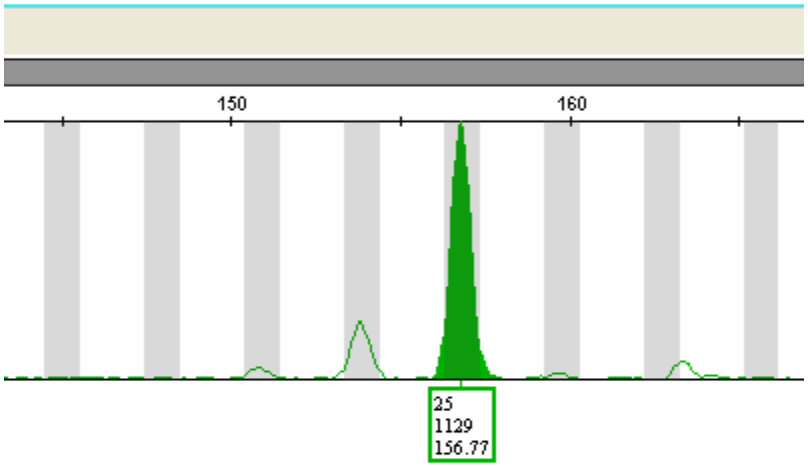
DYS643 11.1

Off-Ladder Allele at DYS481

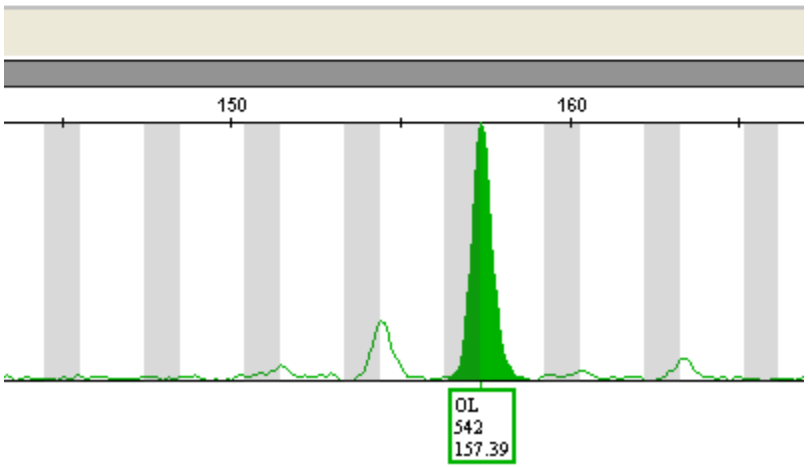
GT37765



GT37812



GT37420



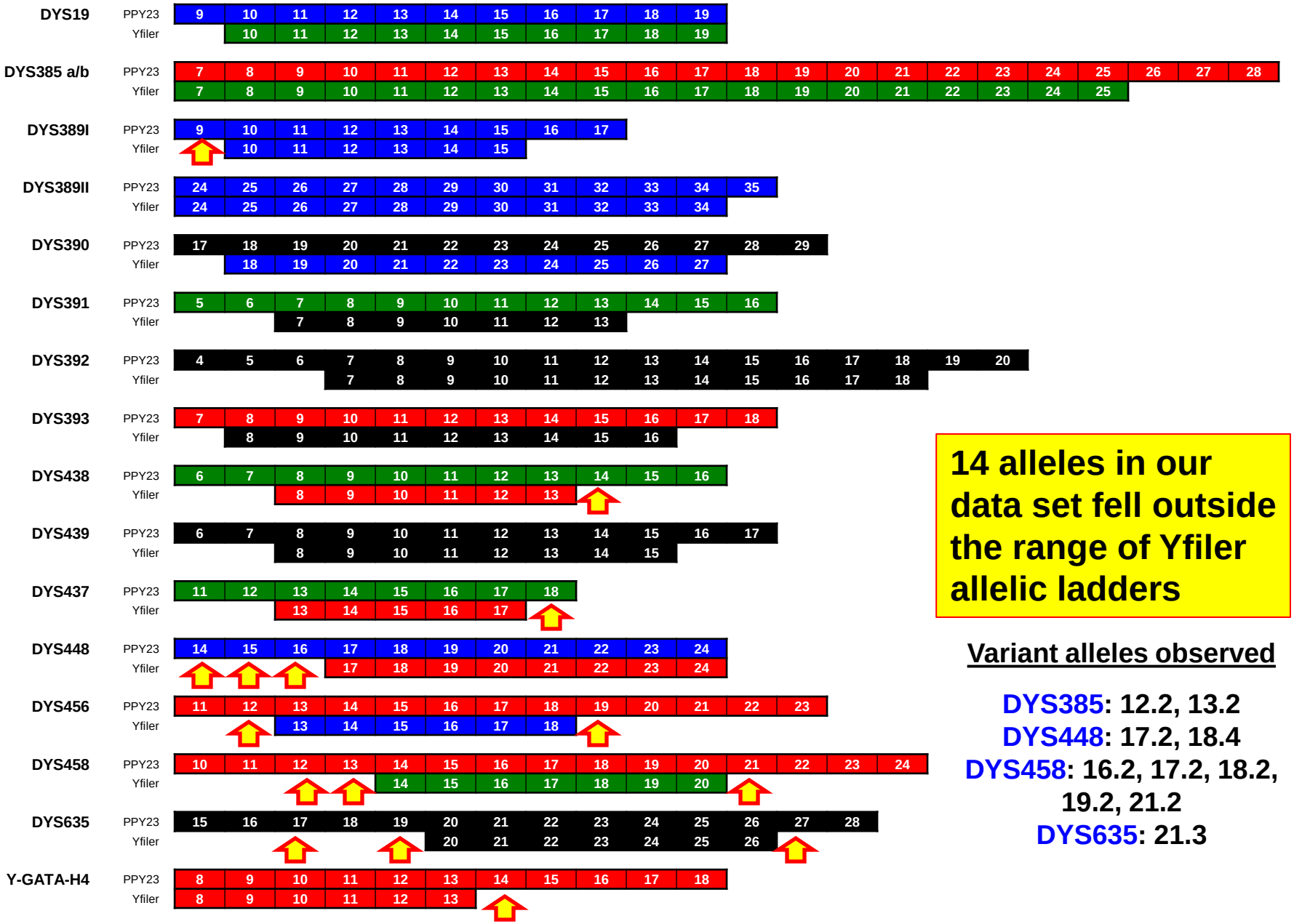
Labeled “25.1”
(sequencing planned to confirm reason)

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



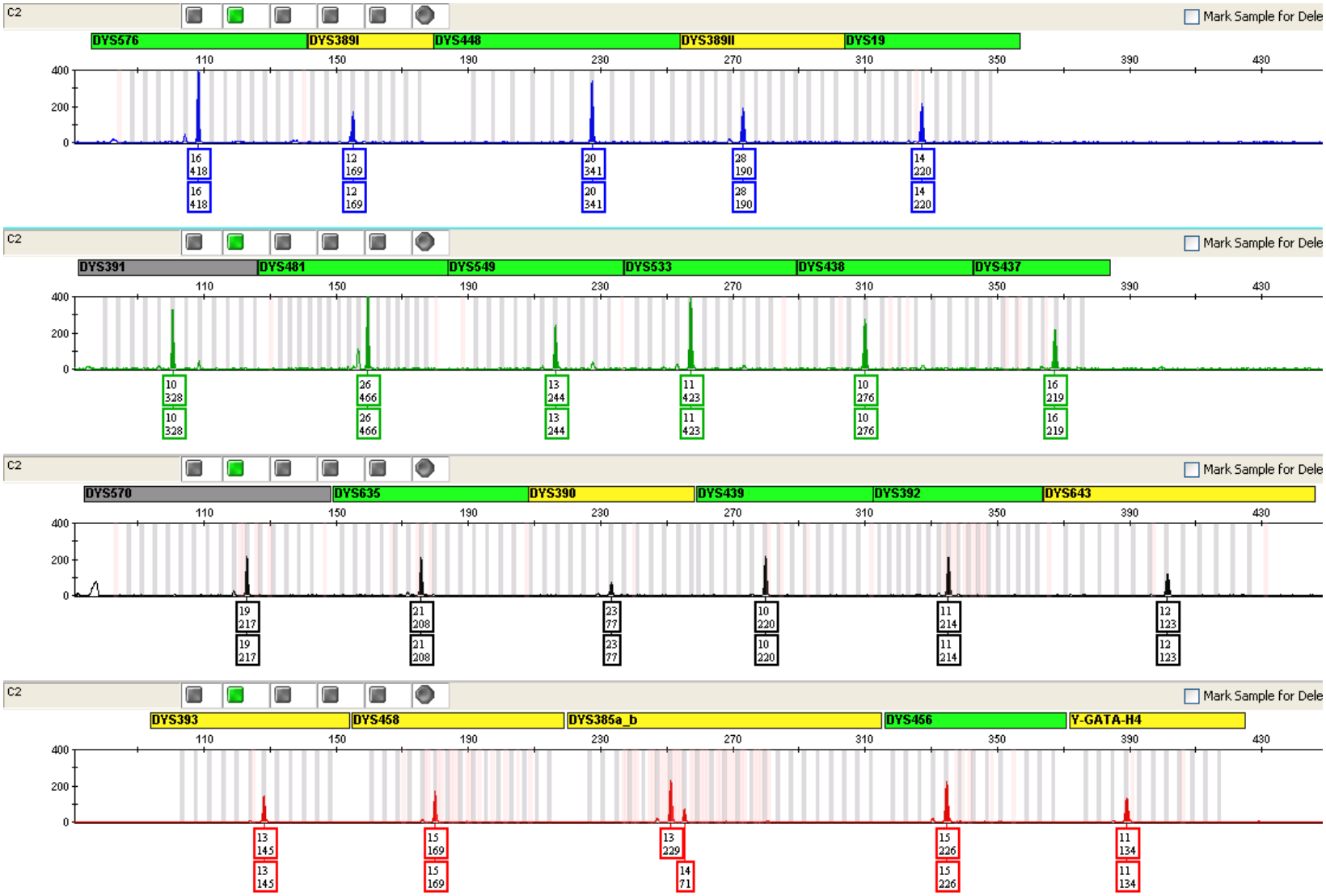
14 alleles in our data set fell outside the range of Yfiler allelic ladders

Variant alleles observed

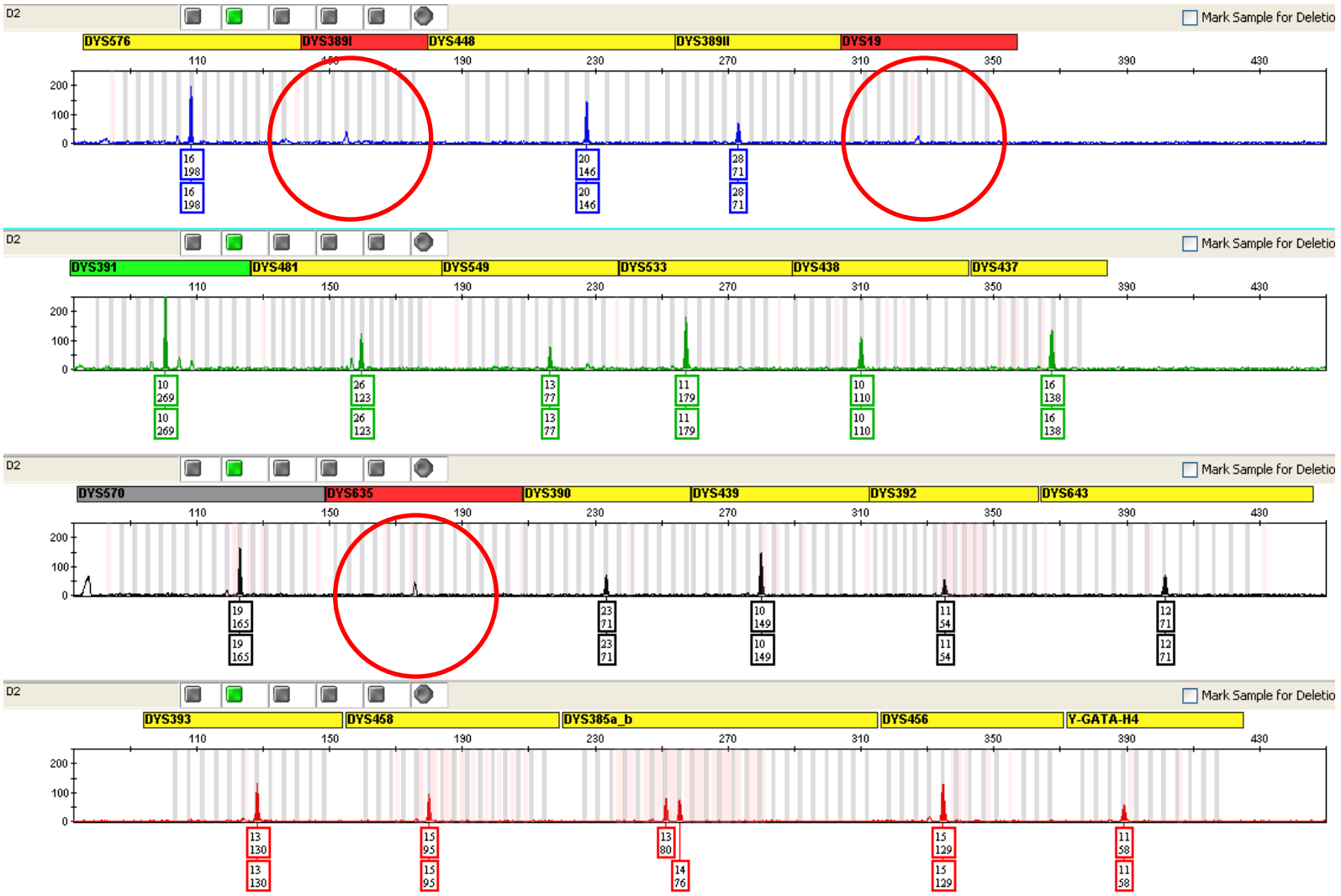
DYS385: 12.2, 13.2
DYS448: 17.2, 18.4
DYS458: 16.2, 17.2, 18.2, 19.2, 21.2
DYS635: 21.3

Sensitivity

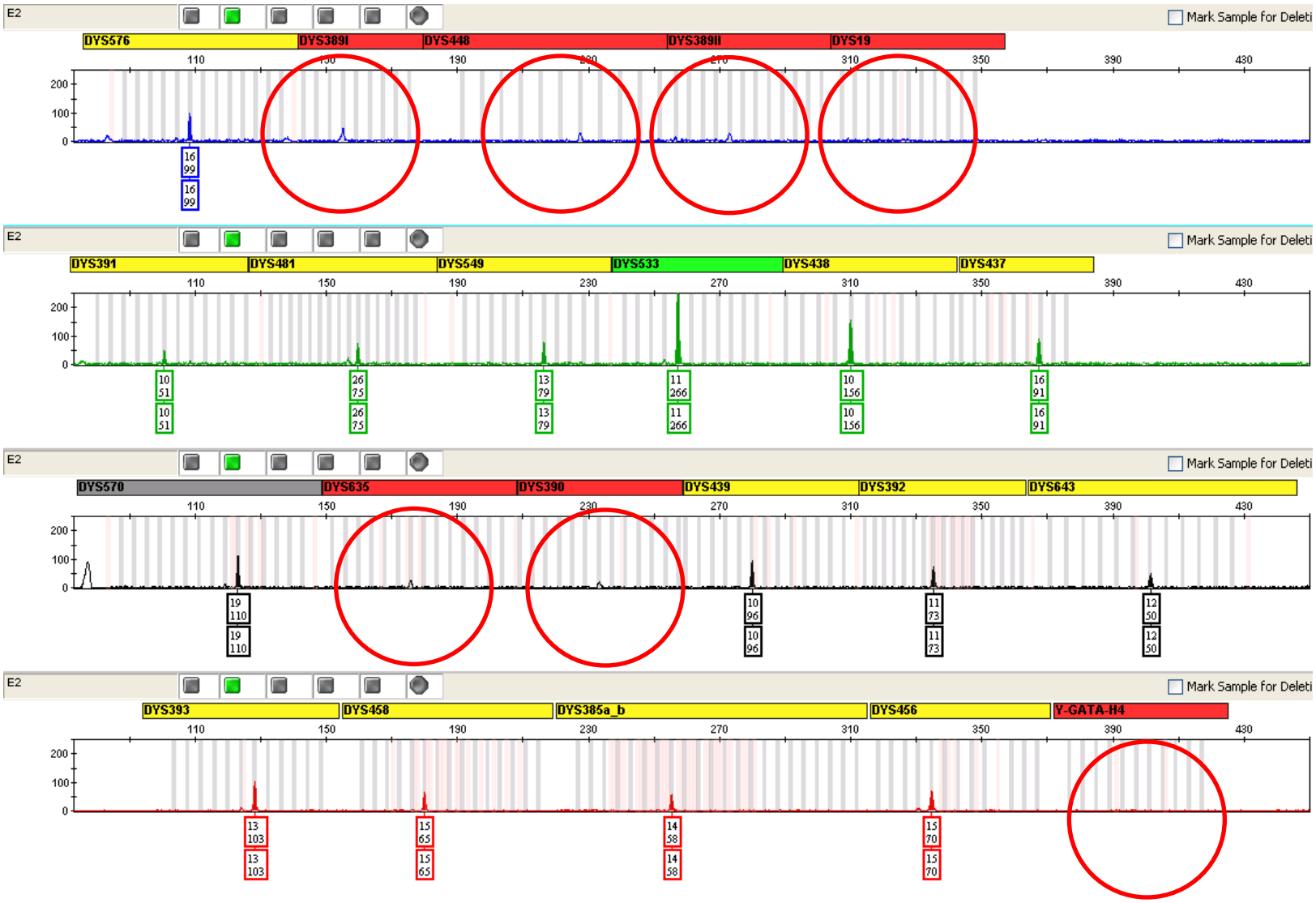
Sensitivity Study – Sample C (125pg)



Sensitivity Study – Sample D (62.5pg)

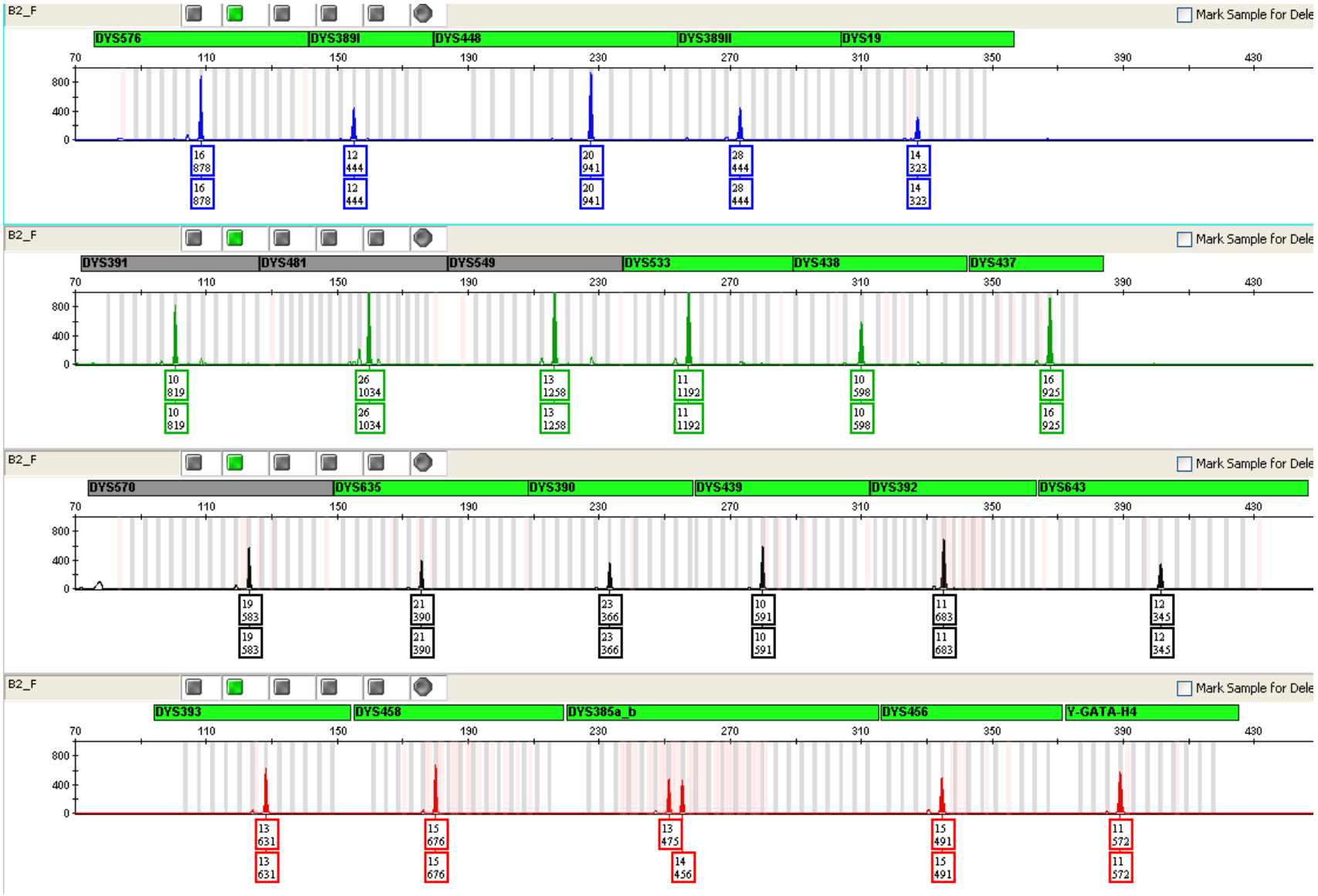


Sensitivity Study – Sample E (31.2pg)

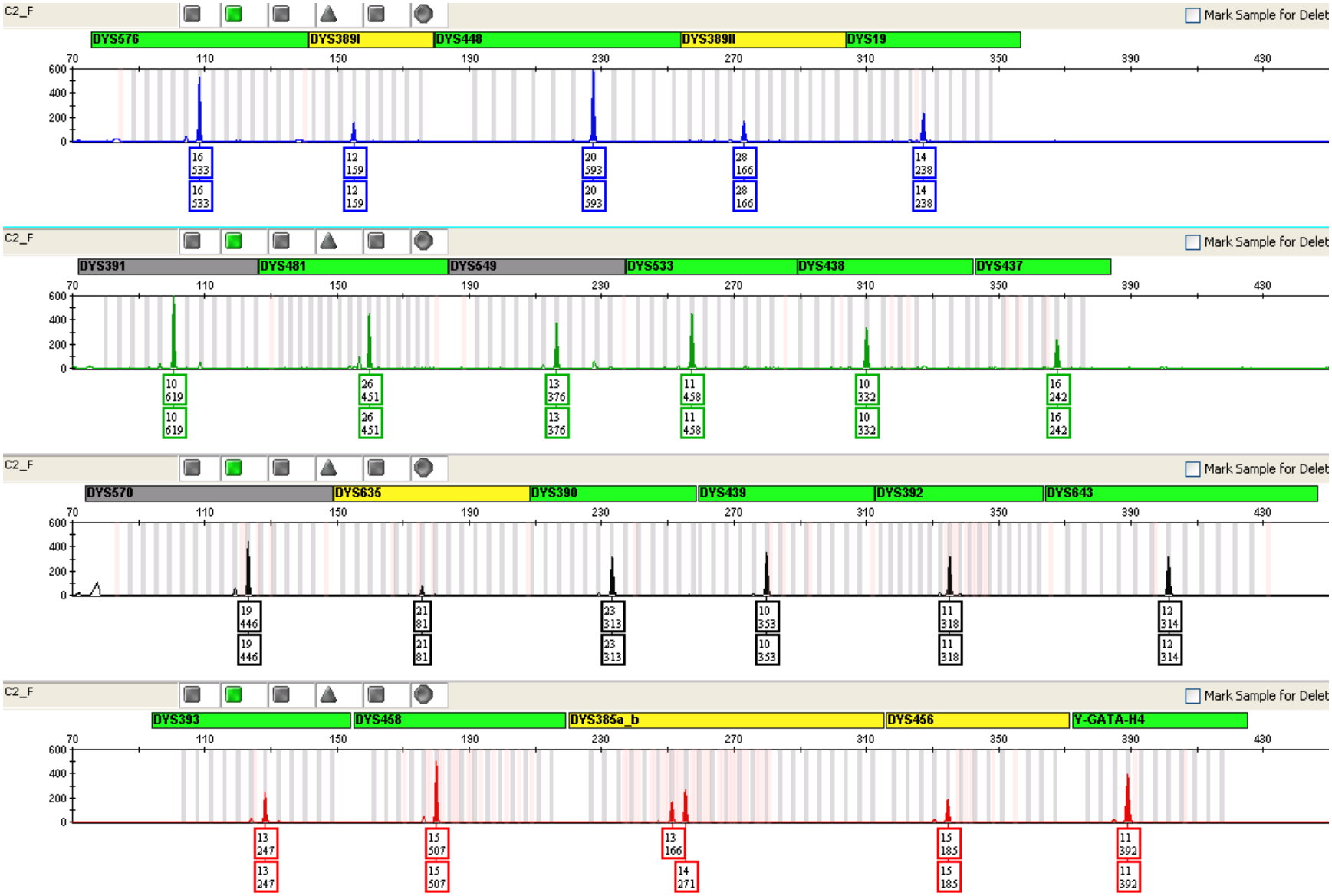


Mixtures

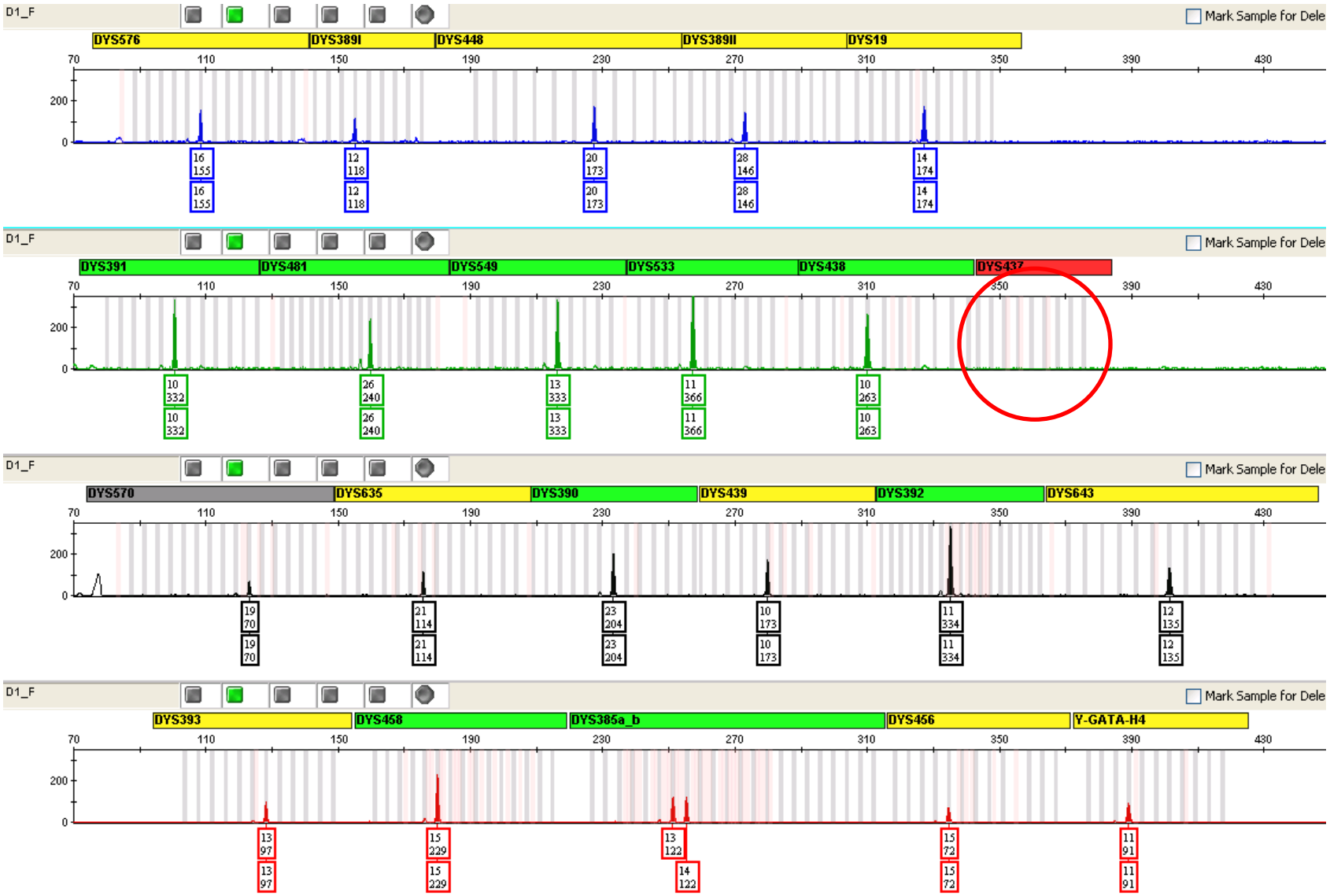
250pg male + 400ng female



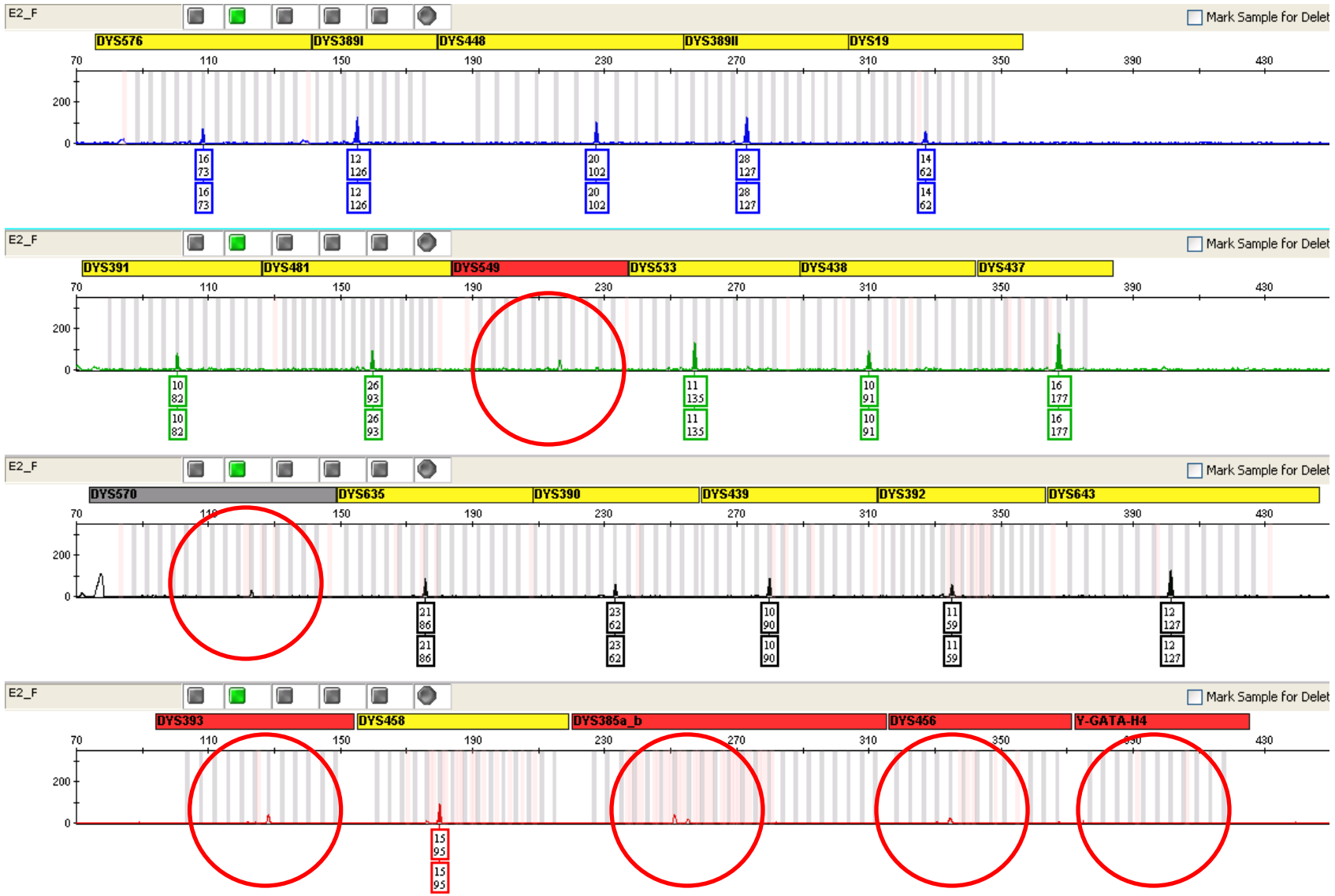
125pg male + 400ng female



62.5pg male + 400ng female



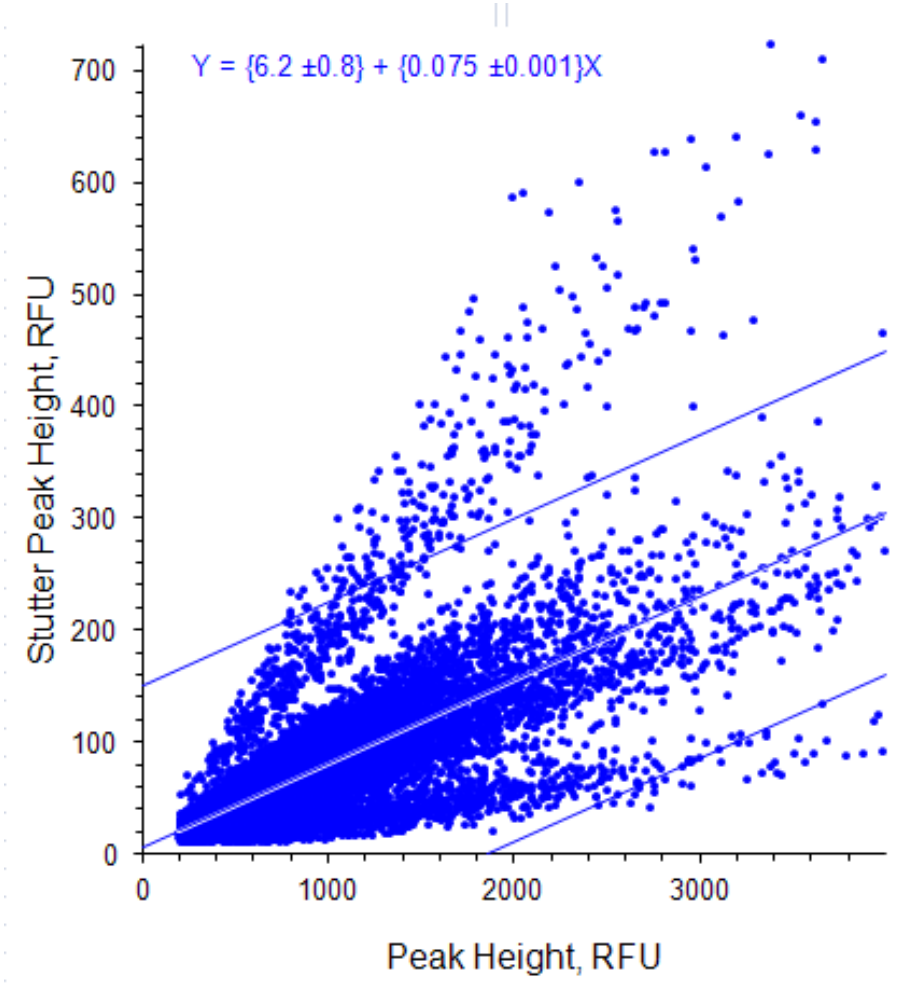
31.2pg male + 400ng female



16/23 loci amplified

Stutter Analysis

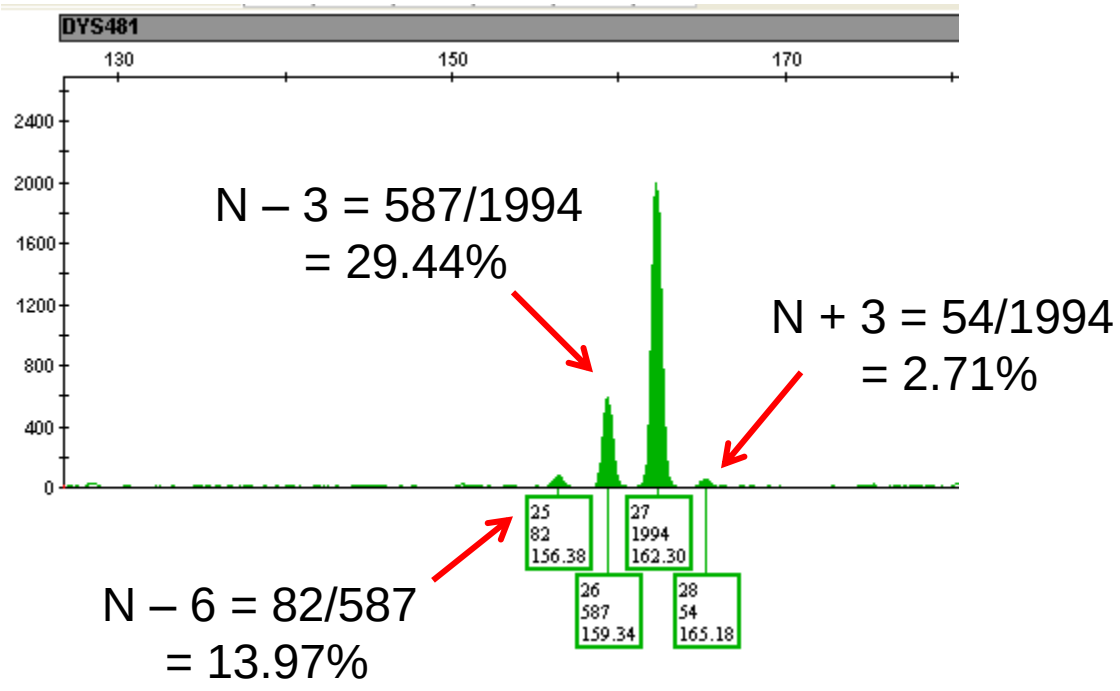
Stutter Observed
(Big Picture View)



Locus-Specific Stutter Values

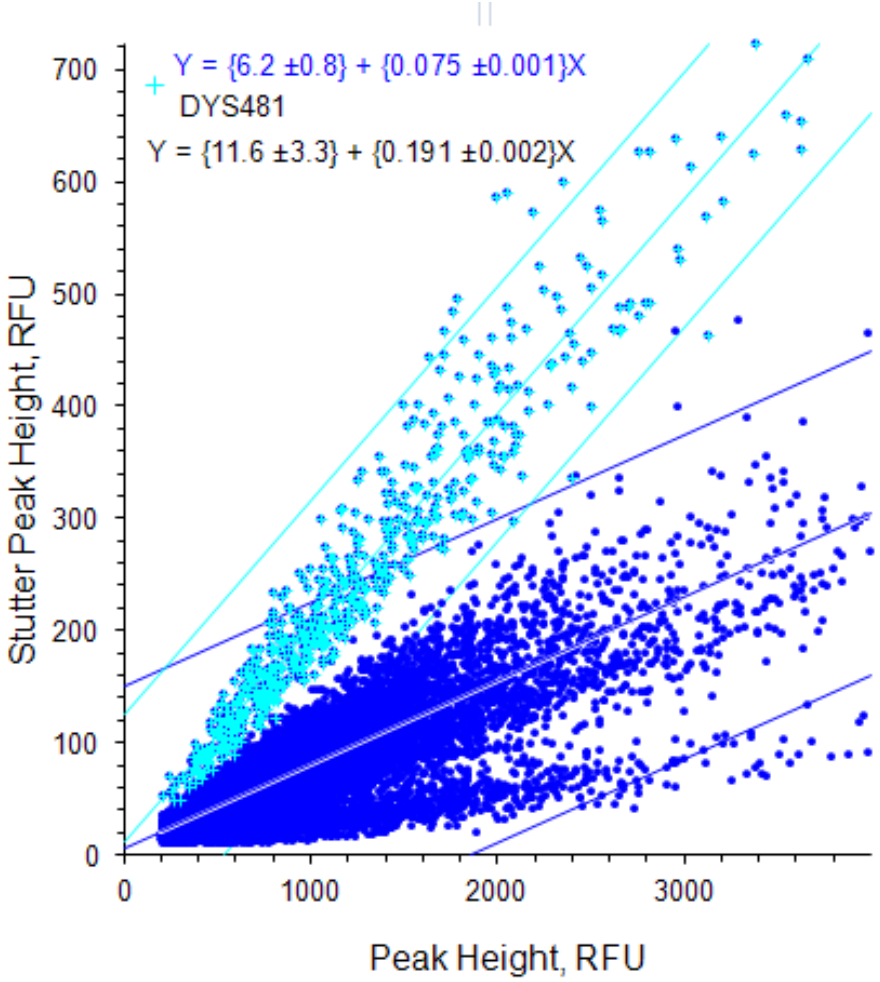
		Percentage Stutter															
Locus	#Ratio	Mean	SD	Median	MADe	Min	1%	2.5%	5%	10%	25%	75%	90%	95%	97.5%	99%	Max
DYS643	414	2.8	1.0	2.7	0.9	1.3	1.4	1.5	1.6	1.7	2.1	3.3	3.9	4.2	4.8	5.3	11.7
DYS448	559	3.0	1.0	2.8	0.6	1.5	1.8	1.9	2.0	2.1	2.4	3.2	4.0	4.9	5.9	7.1	8.3
DYS438	524	3.2	0.7	3.1	0.6	1.1	1.6	1.9	2.1	2.3	2.8	3.6	4.0	4.4	4.8	5.4	7.3
DYS437	444	5.0	1.2	4.9	1.0	2.5	3.1	3.4	3.6	3.8	4.3	5.6	6.4	7.0	7.6	9.3	12.3
DYS389I	528	5.9	1.7	5.7	1.2	2.0	3.0	3.4	3.8	4.4	4.9	6.5	7.3	8.4	9.9	12.1	19.3
DYS635	585	7.0	1.8	6.7	1.8	3.2	4.0	4.5	4.8	5.0	5.5	7.9	9.3	10.3	11.5	12.2	13.2
DYS439	591	6.7	1.1	6.7	1.1	4.0	4.4	4.8	5.2	5.4	5.9	7.3	8.0	8.5	9.0	9.8	10.2
DYS19	578	7.1	1.6	6.8	1.2	4.2	4.6	4.8	5.2	5.5	6.0	7.7	8.8	10.8	11.5	12.6	14.0
DYS533	581	7.0	1.1	6.9	1.0	2.5	4.2	4.7	5.4	5.7	6.3	7.6	8.2	8.7	9.4	10.0	11.3
DYS391	582	7.3	1.3	7.1	1.0	3.6	5.0	5.4	5.8	6.1	6.5	7.8	8.5	9.0	9.8	11.7	17.5
Y-GATA-H4	572	7.3	1.1	7.3	1.1	3.9	5.0	5.2	5.5	6.0	6.6	8.0	8.8	9.2	9.7	10.1	11.5
DYS549	588	7.4	1.2	7.4	1.2	3.0	4.8	5.3	5.6	5.8	6.6	8.2	8.9	9.4	10.0	10.3	11.1
DYS390	596	7.9	1.5	8.1	1.6	4.6	5.1	5.3	5.5	5.8	6.7	9.0	9.8	10.2	10.5	10.9	14.2
DYS385a_b	806	8.8	2.3	8.6	2.5	4.2	5.2	5.5	5.8	6.0	6.8	10.2	11.9	13.3	14.1	14.9	23.1
DYS392	592	9.8	2.3	9.3	2.7	2.4	6.4	6.8	7.0	7.2	7.8	11.6	12.9	13.5	14.0	14.9	17.9
DYS458	586	9.4	1.3	9.4	1.3	5.5	6.4	6.9	7.3	7.8	8.6	10.2	11.1	11.6	12.0	12.9	14.0
DYS393	594	9.9	1.3	9.7	0.9	7.6	7.9	8.1	8.5	8.8	9.2	10.4	11.3	12.0	12.7	13.2	23.3
DYS570	595	10.7	1.3	10.5	1.2	6.8	7.8	8.3	8.8	9.2	9.9	11.5	12.3	13.0	13.5	14.0	15.5
DYS576	600	10.7	1.5	10.7	1.3	5.9	7.3	7.9	8.3	8.9	9.8	11.6	12.5	13.2	13.9	14.5	17.5
DYS456	576	10.9	1.4	10.8	1.2	6.6	8.0	8.3	8.8	9.3	10.1	11.7	12.7	13.4	14.0	14.2	16.7
DYS389II	561	11.6	1.4	11.3	1.4	7.3	9.1	9.3	9.5	9.9	10.6	12.4	13.4	14.1	14.9	15.2	16.4
DYS481	586	20.3	2.9	19.7	2.7	14.0	15.5	16.1	16.5	17.3	18.1	22.2	24.2	26.1	26.9	27.5	29.4

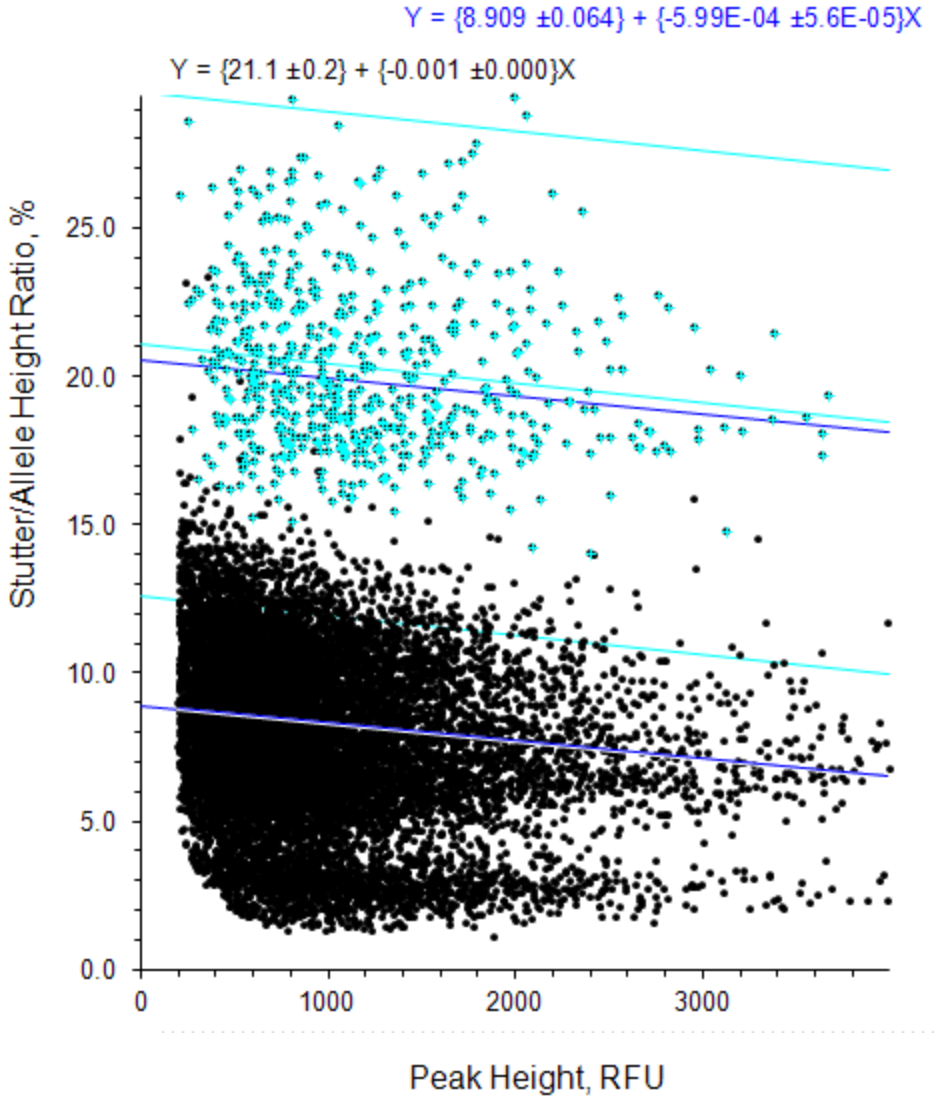
DYS481 – Highest Stutter Examined



Alleles above 50RFU

DYS481 (light blue)





DYS481 Allele-Specific Stutter

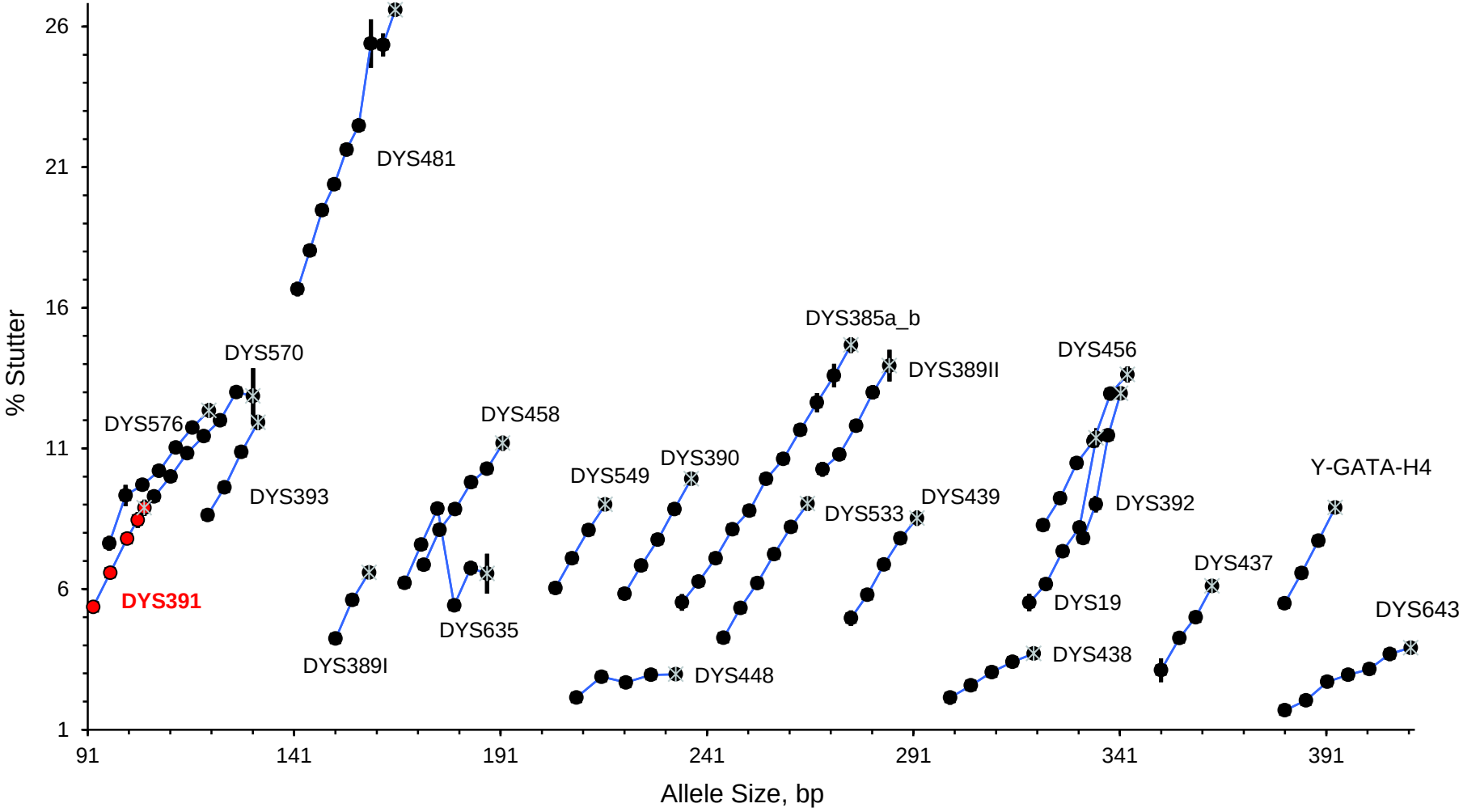
			Stutter		
Locus	Allele	Size	#	Median	MADe
DYS481	21	141.8	35	16.7	0.8
	22	144.8	204	18.0	0.8
	23	147.8	81	19.5	0.8
	24	150.7	52	20.4	0.7
	25	153.7	89	21.6	1.1
	26	156.7	52	22.5	0.9
	27	159.6	28	25.4	2.3
	28	162.6	29	25.3	1.1
	29	165.5	11	26.6	0.4
		Avg	581	21.8	1.1
		SD		3.5	

DYS392 and DYS481 Allele-Specific Stutter (Trinucleotide repeat markers)

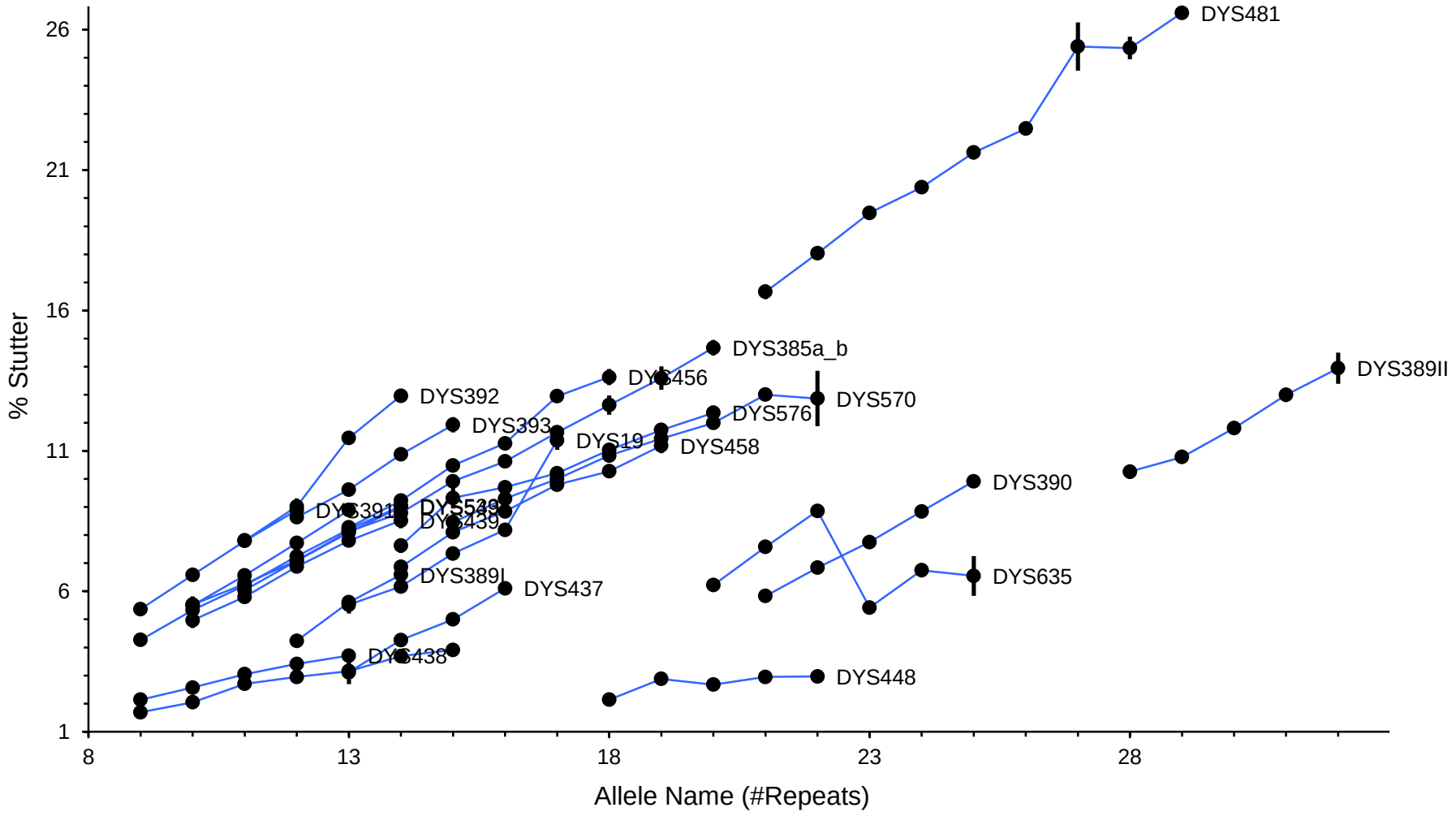
			Stutter		
Locus	Allele	Size	#	Median	MADe
DYS392	11	332.1	284	7.8	0.7
	12	335.2	31	9.0	0.8
	13	338.2	233	11.5	1.1
	14	341.2	31	13.0	0.5
		Avg	579	10.3	0.8
		SD		2.3	

			Stutter		
Locus	Allele	Size	#	Median	MADe
DYS481	21	141.8	35	16.7	0.8
	22	144.8	204	18.0	0.8
	23	147.8	81	19.5	0.8
	24	150.7	52	20.4	0.7
	25	153.7	89	21.6	1.1
	26	156.7	52	22.5	0.9
	27	159.6	28	25.4	2.3
	28	162.6	29	25.3	1.1
	29	165.5	11	26.6	0.4
		Avg	581	21.8	1.1
		SD		3.5	

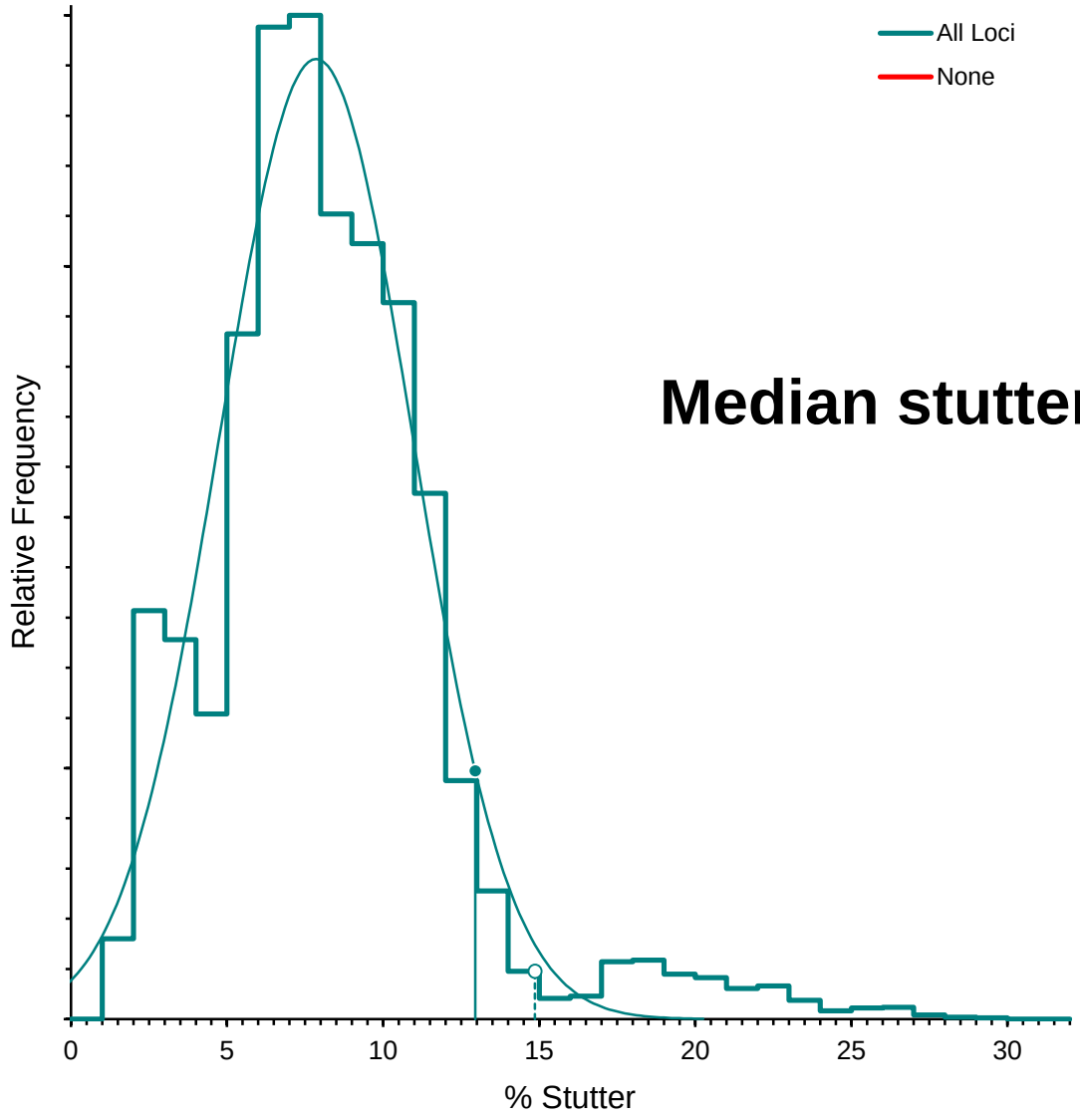
Stutter Trends (Size Scale)



Stutter Trends (Repeat Scale)



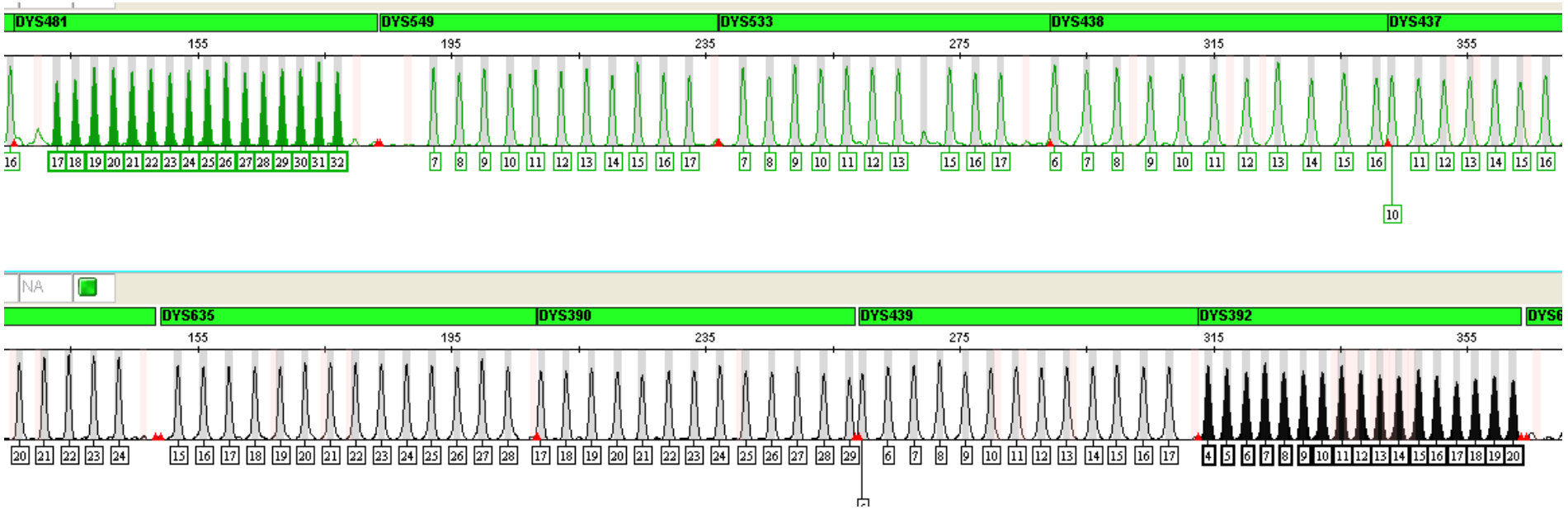
Stutter Density Plot



Median stutter = 7.85%

Forward Stutter Analysis

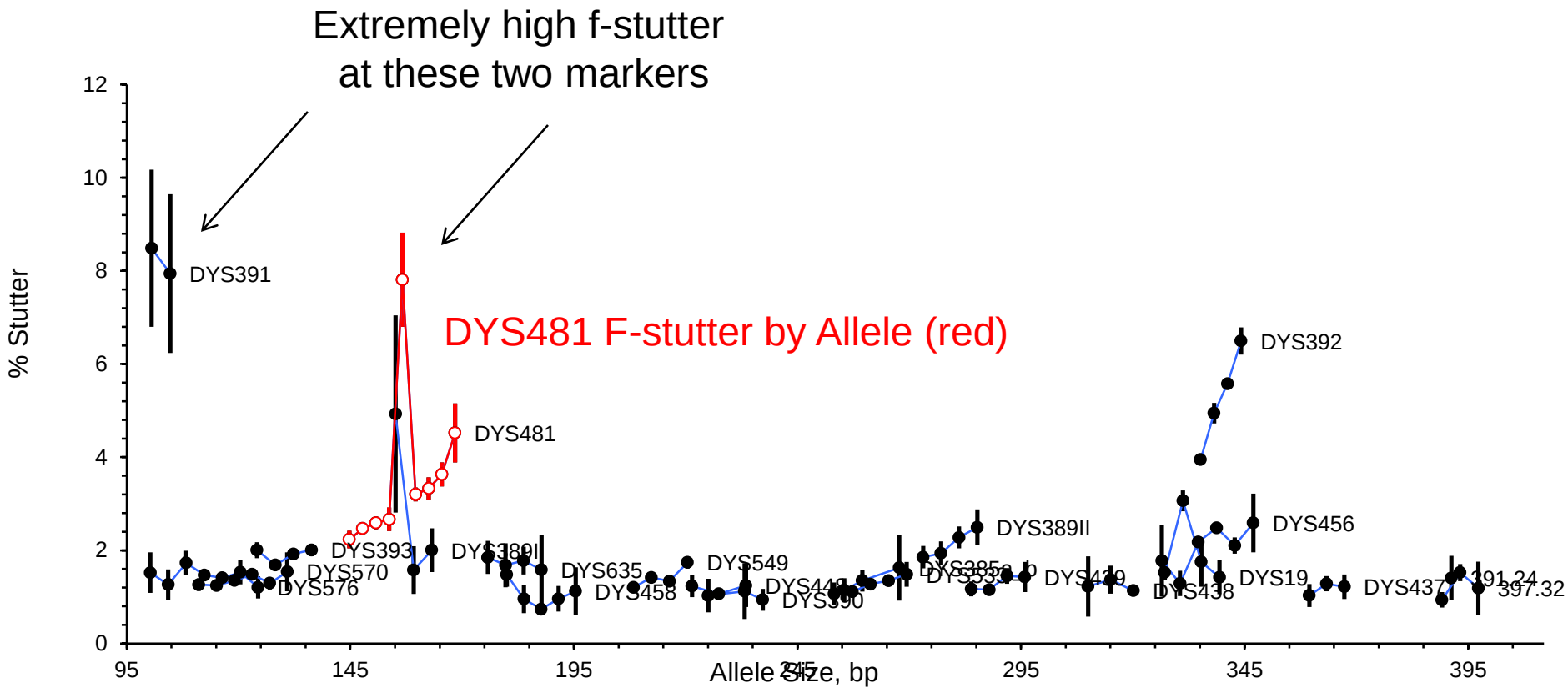
DYS481 (new) and DYS392 are trinucleotide repeat markers

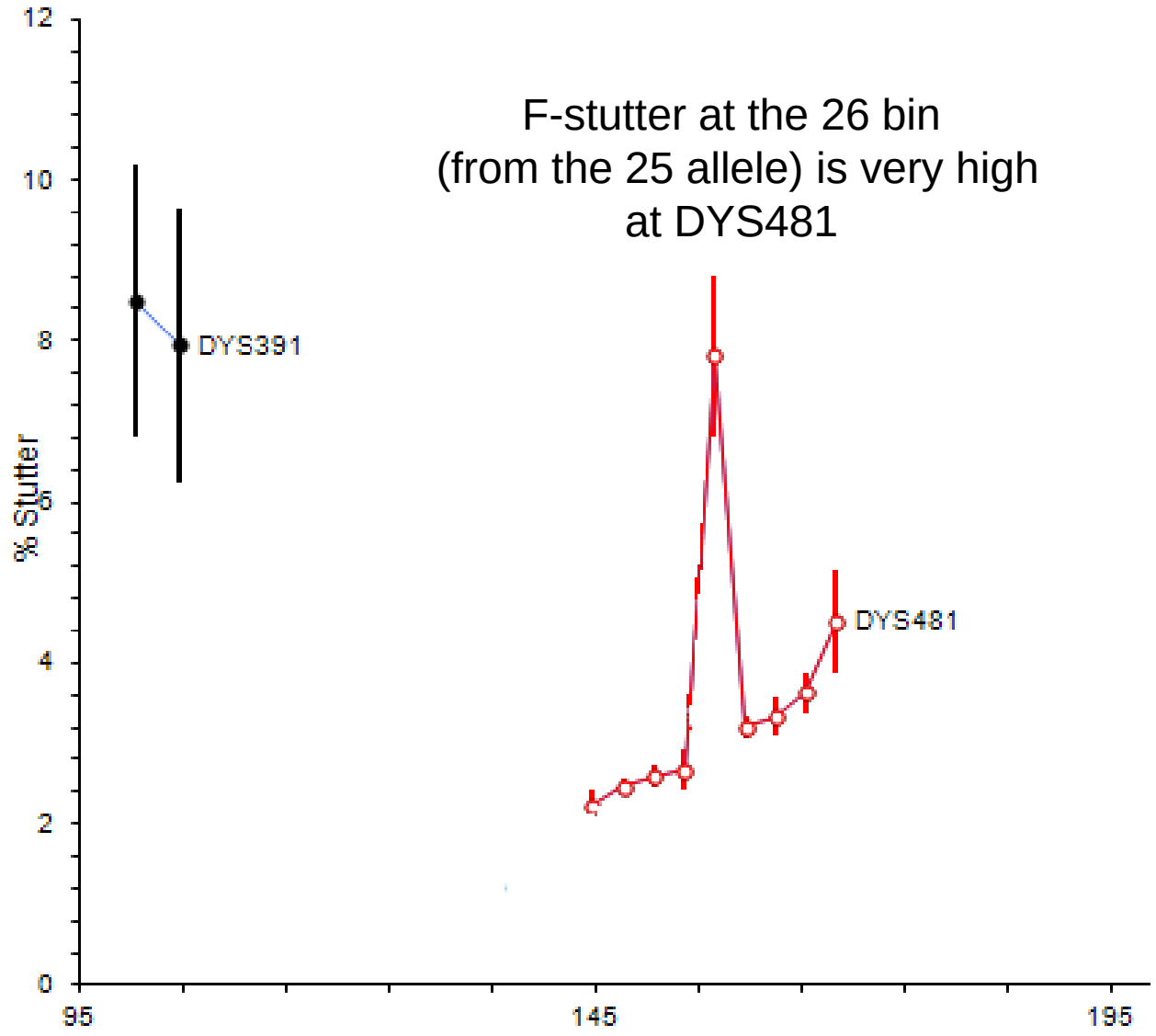


Locus-Specific F-Stutter Values

		Percentage Stutter															
Locus	#Ratio	Mean	SD	Median	MADe	Min	1%	2.5%	5%	10%	25%	75%	90%	95%	97.5%	99%	Max
DYS458	114	1.3	0.9	1.0	0.6	0.3	0.3	0.3	0.4	0.4	0.6	1.7	2.4	3.1	4.0	4.2	4.4
DYS390	30	1.4	1.1	1.0	0.4	0.4	0.4	0.4	0.5	0.7	0.8	1.6	2.3	3.2	4.4	5.3	5.9
DYS448	55	1.4	1.1	1.1	0.4	0.3	0.4	0.5	0.5	0.6	0.9	1.5	2.5	3.4	5.2	5.9	6.0
DYS437	80	1.4	0.7	1.1	0.5	0.4	0.4	0.5	0.6	0.7	0.9	1.7	2.2	2.6	2.9	3.6	4.2
DYS438	112	1.3	0.7	1.1	0.5	0.3	0.4	0.4	0.5	0.7	0.9	1.6	2.2	2.6	2.9	3.2	4.3
DYS439	207	1.4	1.7	1.2	0.6	0.5	0.6	0.6	0.6	0.7	0.9	1.6	2.1	2.4	2.8	3.4	24.8
DYS533	327	1.4	0.7	1.2	0.4	0.4	0.6	0.7	0.7	0.8	1.0	1.6	2.2	2.6	3.2	3.9	7.4
DYS385a_b	57	1.8	1.6	1.3	0.7	0.5	0.5	0.6	0.6	0.7	0.9	1.8	2.9	5.4	5.9	7.7	10.1
DYS570	265	1.5	0.6	1.3	0.4	0.6	0.7	0.8	0.8	0.9	1.1	1.7	2.3	2.7	3.1	3.7	5.4
Y-GATA-H4	79	1.5	0.7	1.4	0.7	0.4	0.4	0.6	0.6	0.7	0.9	2.0	2.4	2.7	3.0	3.4	3.7
DYS549	342	1.5	0.6	1.4	0.4	0.6	0.8	0.8	0.9	1.0	1.1	1.8	2.2	2.6	2.8	3.9	6.6
DYS576	267	1.7	0.8	1.4	0.6	0.5	0.6	0.6	0.7	0.9	1.2	2.0	2.6	3.1	3.7	4.8	5.2
DYS643	29	1.9	1.2	1.5	0.7	0.5	0.6	0.7	0.7	0.9	1.2	2.1	3.3	4.9	5.4	5.4	5.4
DYS635	89	1.9	1.1	1.6	0.9	0.4	0.5	0.6	0.6	0.8	1.1	2.3	3.2	4.3	4.5	5.0	5.4
DYS393	427	1.9	0.6	1.8	0.5	0.8	1.1	1.2	1.2	1.3	1.5	2.1	2.6	2.9	3.4	3.9	4.8
DYS389II	177	2.5	1.3	2.1	0.9	0.7	0.8	0.9	1.1	1.3	1.6	3.1	3.8	5.0	5.6	7.5	8.4
DYS456	385	2.4	0.8	2.3	0.7	0.7	0.9	1.1	1.3	1.5	1.9	2.8	3.5	3.8	4.4	4.8	6.2
DYS389I	67	3.4	3.2	2.4	1.8	0.4	0.5	0.6	0.7	0.8	1.2	3.5	7.9	11.1	12.3	13.5	14.9
DYS19	167	2.7	1.4	2.5	1.4	0.8	0.8	0.9	0.9	1.1	1.7	3.5	4.9	5.3	5.8	6.6	7.3
DYS481	492	3.5	2.2	2.8	0.9	1.1	1.6	1.8	1.9	2.0	2.3	3.5	6.6	9.1	10.4	11.8	13.1
DYS392	571	5.0	1.3	4.9	1.4	1.4	2.8	3.1	3.3	3.5	4.0	5.7	6.5	7.1	7.6	8.4	14.5
DYS391	139	6.9	5.4	8.3	7.4	0.3	0.3	0.4	0.5	0.7	1.2	11.0	13.0	15.2	17.6	19.5	21.1

F-stutter Trends (Size Scale)



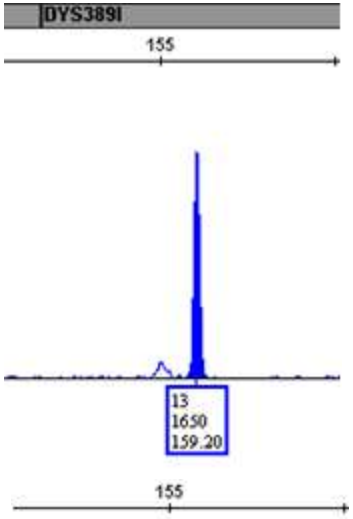

F-stutter at the 26 bin
(from the 25 allele) is very high
at DYS481
% Stutter
12
10
8
6
4
2
0
95
145
195
DYS391
DYS481

High f-stutter at DYS481

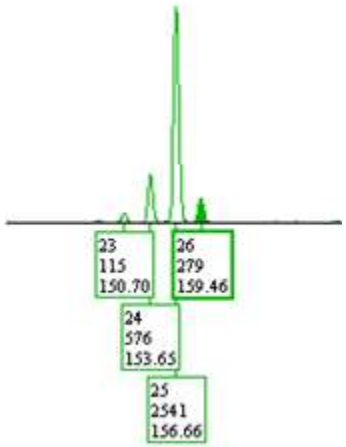
DYS481							
Sample	All S	All P	Sze S	Sze P	Hgt S	Hgt P	PHR
JT51484	26	25	159.40	156.57	216	1647	13.115
ZT79338	26	25	159.54	156.64	113	897	12.598
GT37019	26	25	159.33	156.47	26	207	12.560
GT37190	26	25	159.31	156.54	404	3375	11.970
PT84212	26	25	159.23	156.64	47	397	11.839
GT37170	26	25	159.37	156.60	510	4340	11.751
ZT79620	26	25	159.41	156.66	124	1126	11.012
JT51481	26	25	159.46	156.66	279	2541	10.980
GT37020	26	25	159.38	156.51	88	807	10.905
OT05896	26	25	159.50	156.76	51	471	10.828
PT83873	26	25	159.34	156.63	228	2155	10.580
OT05598	26	25	159.30	156.58	54	516	10.465
ZT79339	26	25	159.46	156.60	64	613	10.440
MT97122	26	25	159.36	156.59	56	547	10.238
GT36886	26	25	159.47	156.64	131	1310	10.000
PT83872	26	25	159.43	156.68	230	2320	9.914
MT97139	26	25	159.47	156.61	164	1671	9.814
PT83868	26	25	159.31	156.62	39	400	9.750
ZT79303	26	25	159.39	156.61	149	1545	9.644
MT94883	26	25	159.38	156.72	61	642	9.502
GT37169	26	25	159.36	156.59	131	1387	9.445

Top 21 samples with high f-stutter at DYS481

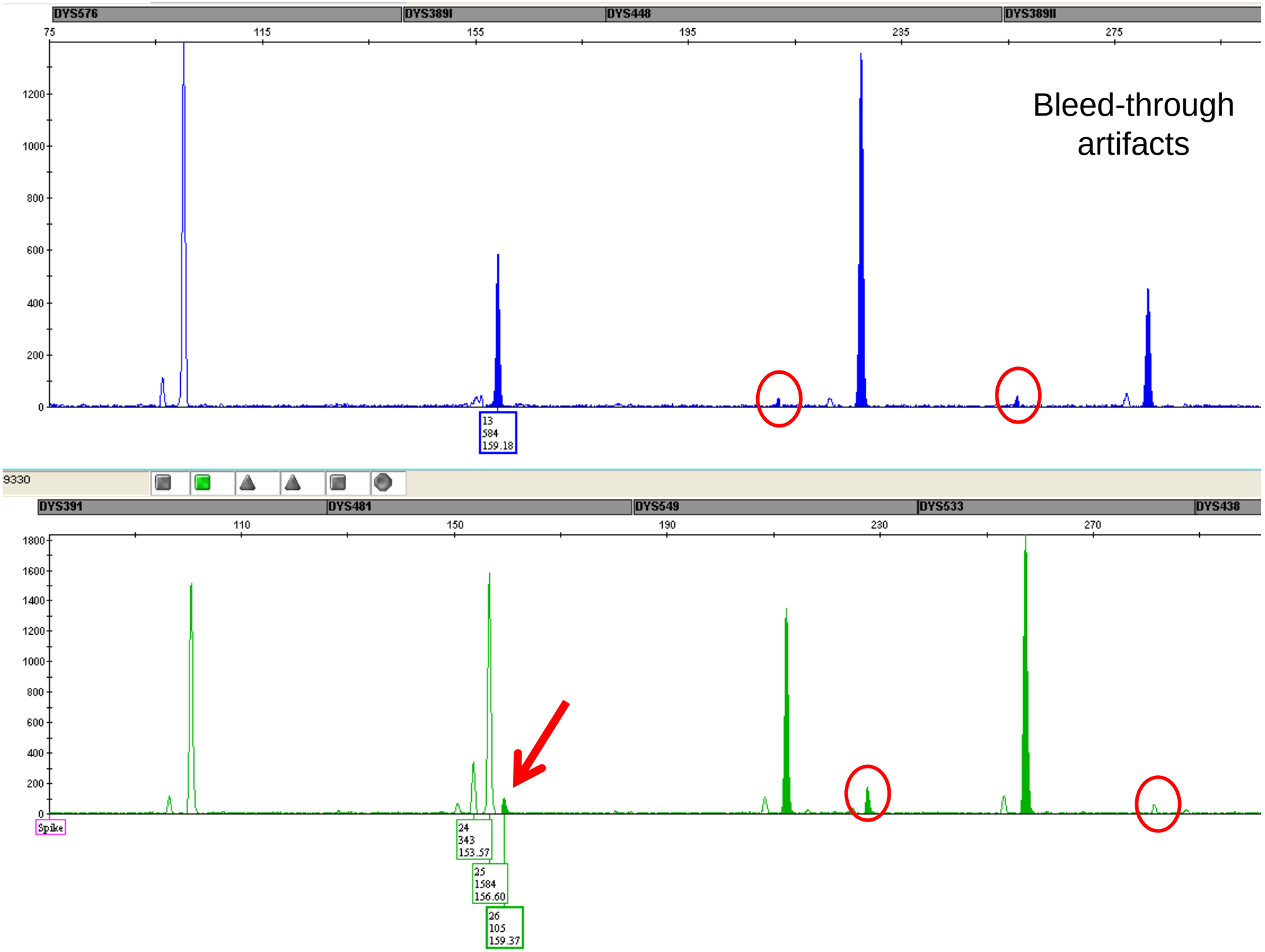
All of these samples have a 13 allele at DYS389-I



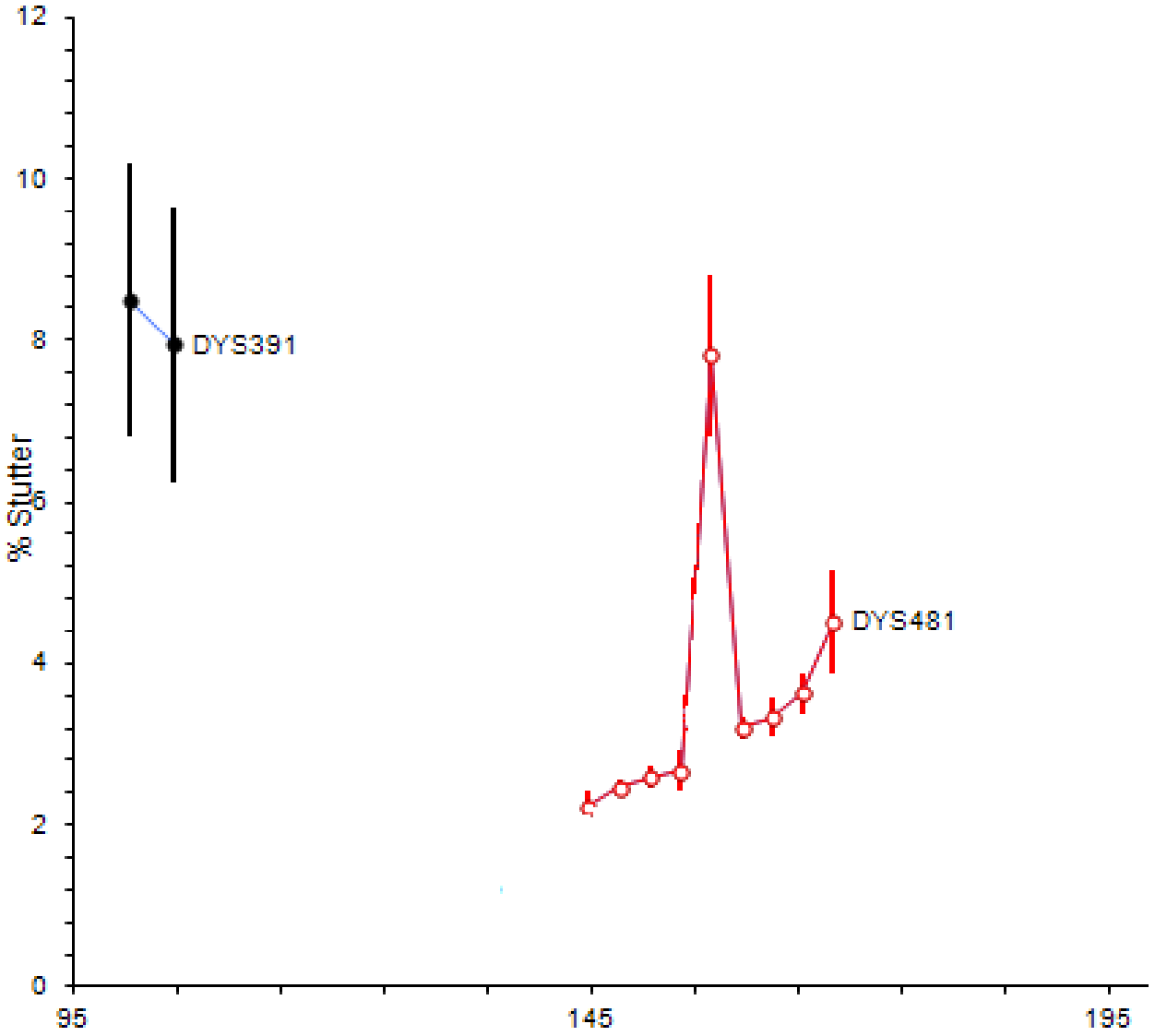
13 allele – 159.20 bp



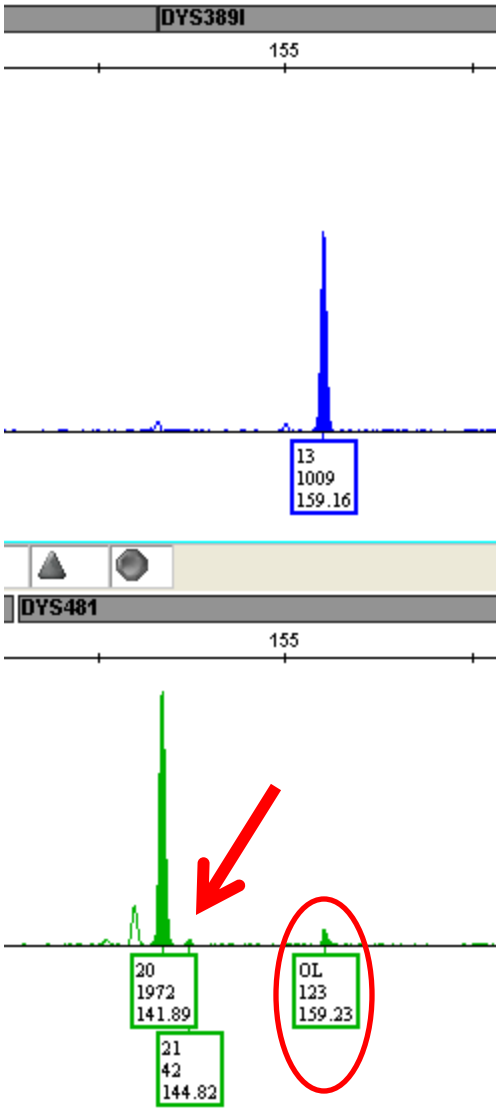
26 allele – 159.46 bp


DYS576
DYS389I
DYS448
DYS389II
75
115
155
195
235
275
1200
1000
800
600
400
200
0
13
584
159.18
Bleed-through artifacts
9330
DYS391
DYS481
DYS549
DYS533
DYS438
110
150
190
230
270
1800
1600
1400
1200
1000
800
600
400
200
0
Spike
24
343
153.57
25
1584
156.60
26
105
159.37

Pull-up from the 13 allele at DYS389-I is falling into the 26 bin of DYS481 and overestimating the true f-stutter percentage.
This is likely also happening with DYS391 (with pull-up from DYS576).



An example of “true f-stutter”



The bleed through at the 13 allele from DYS389-I does not affect the f-stutter at the 20 allele of DYS481. Note here that the f-stutter is 42/1972 = 2.1%.

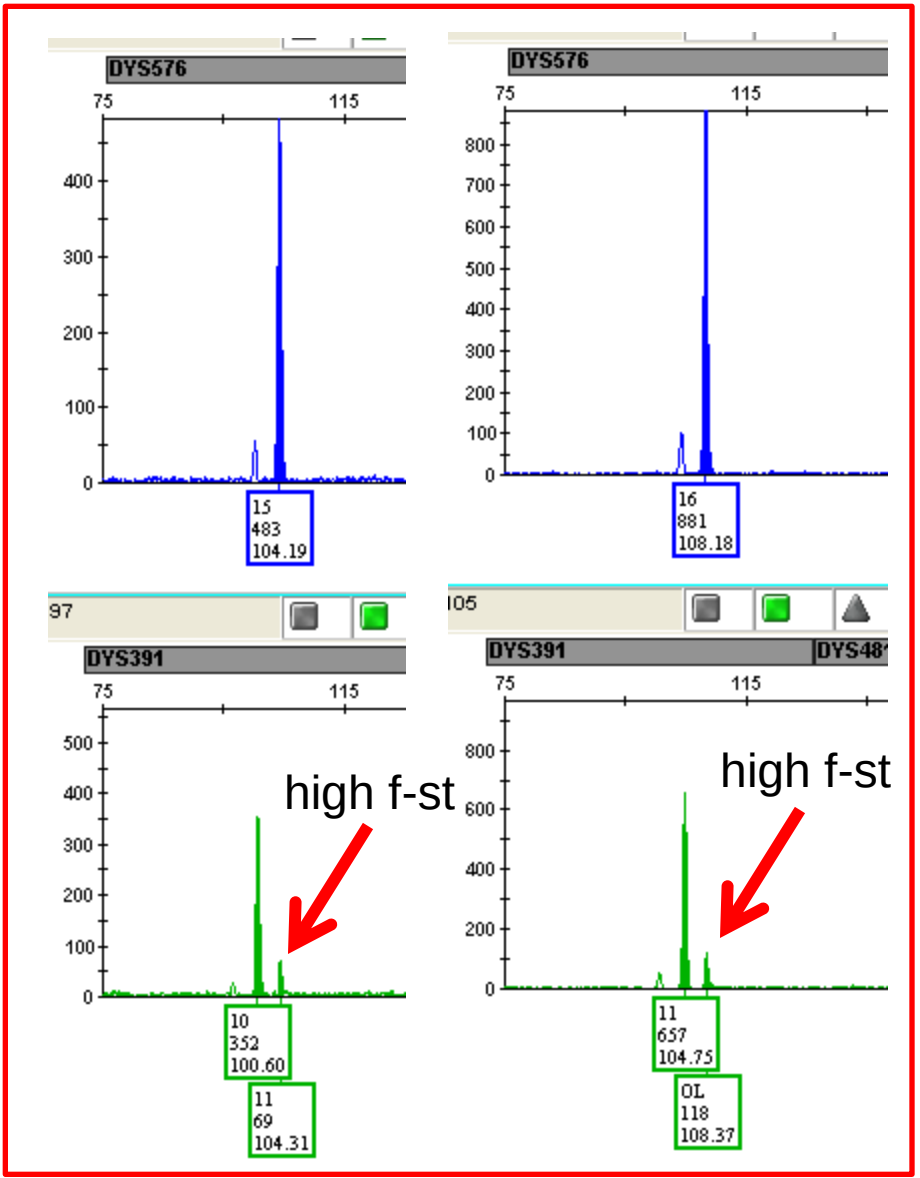
High f-stutter at DYS391

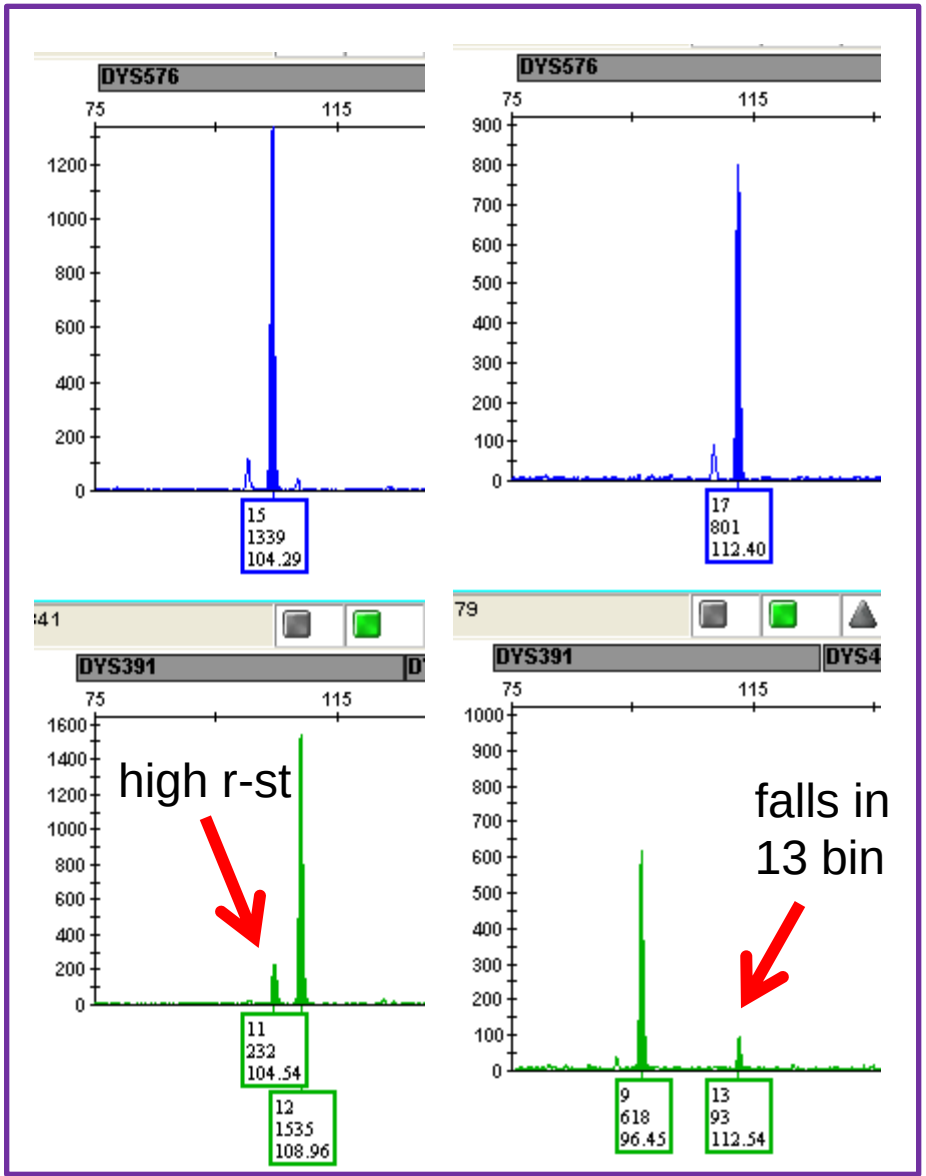
DYS391							
Sample	All S	All P	Sze S	Sze P	Hgt S	Hgt P	PHR
PT83907	11	10	104.52	100.63	175	830	21.084
GT38097	11	10	104.31	100.60	69	352	19.602
ZT79322	11	10	104.34	100.57	213	1099	19.381
MT95105	OL	11	108.37	104.75	118	657	17.960
PT84200	11	10	104.39	100.63	95	551	17.241
UT58298	11	10	104.31	100.51	205	1218	16.831
MT95095	11	10	104.43	100.68	47	299	15.719
WT51556	11	10	104.54	100.62	55	362	15.193
GT37032	11	10	104.34	100.58	52	351	14.815
GT38072	OL	11	108.47	104.76	96	670	14.328
PT83897	11	10	104.39	100.62	182	1307	13.925
WT52485	11	10	104.50	100.58	235	1755	13.390
ZT81372	12	11	108.50	104.75	120	916	13.100
ZT79976	11	10	104.39	100.55	325	2487	13.068
PT84226	11	10	104.41	100.48	169	1304	12.960
PT83886	11	10	104.40	100.58	131	1021	12.831
PT83916	11	10	104.38	100.57	131	1021	12.831
OT05594	OL	11	108.33	104.72	103	808	12.748
TT51435	11	10	104.48	100.55	89	718	12.396
JT51471	11	10	104.36	100.49	100	811	12.330
UT58318	12	11	108.41	104.77	259	2110	12.275
OT05592	11	10	104.32	100.53	90	741	12.146
MT97199	12	11	108.38	104.77	624	5322	11.725
JT51478	11	10	104.40	100.60	124	1058	11.720
PT84180	11	10	104.32	100.56	111	954	11.635
ZT79337	11	10	104.35	100.53	207	1788	11.577
PT83870	11	10	104.44	100.61	83	718	11.560
ZT79311	11	10	104.35	100.58	124	1073	11.556
ZT80696	11	10	104.42	100.53	113	979	11.542
OT05576	11	10	104.41	100.58	122	1067	11.434

Top 30 samples with high f-stutter at DYS391

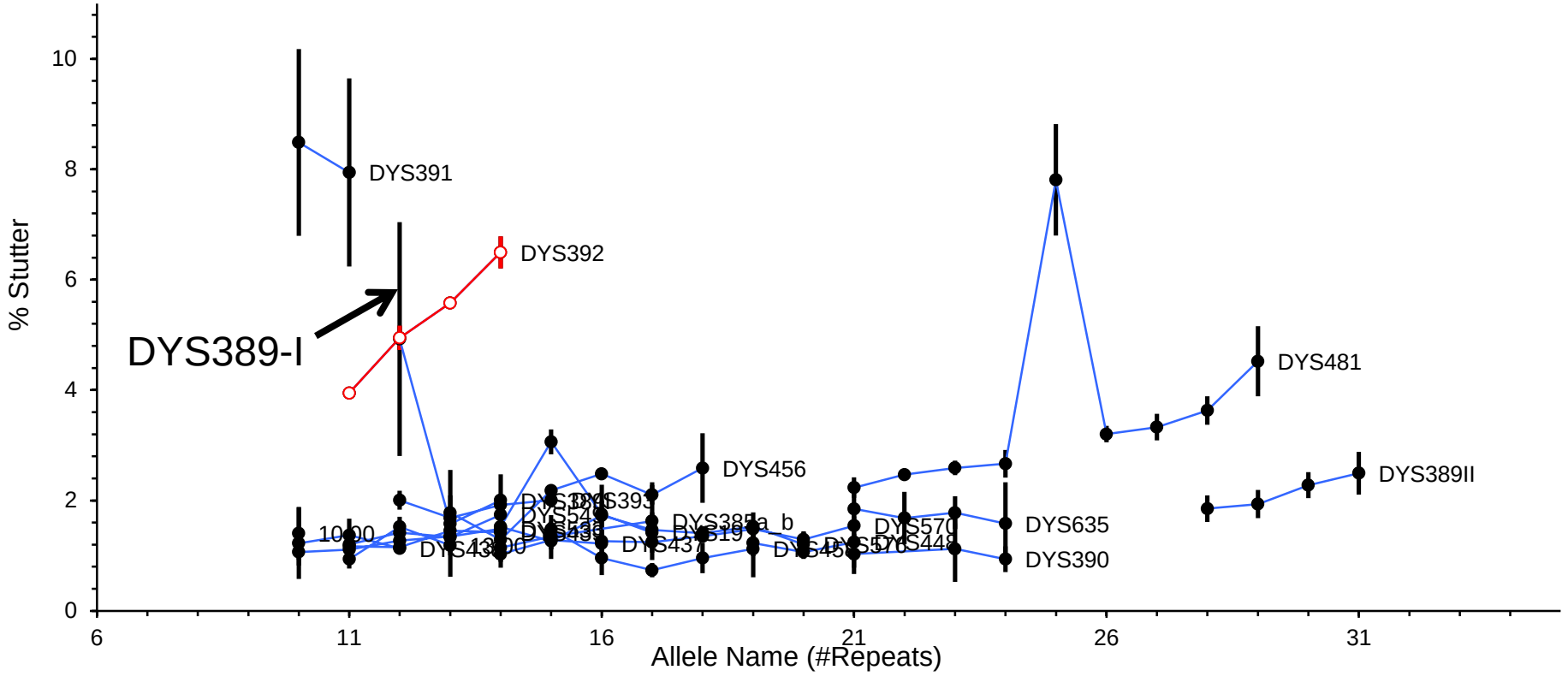
All of these samples have either a 15 or 16 allele at DYS576

DYS576 pull-up is observed in DYS391 bins



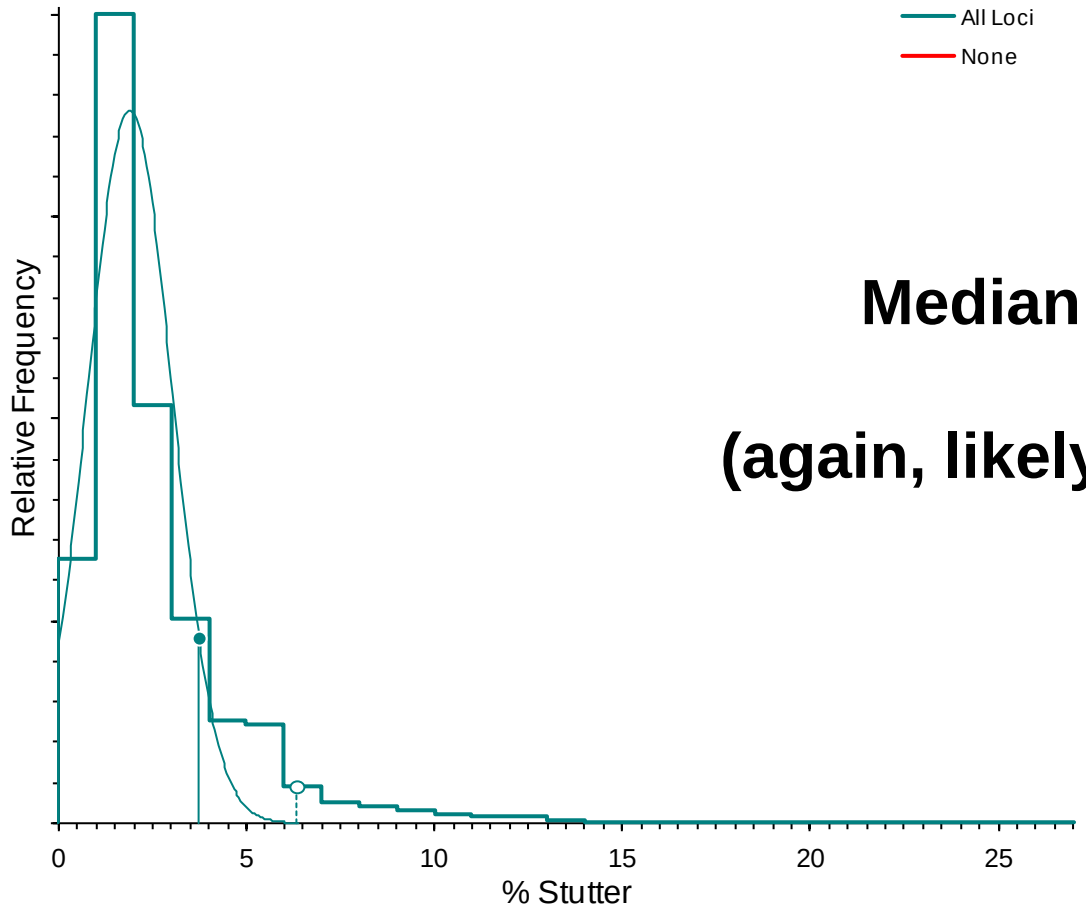


Stutter Trends (Repeat Scale)



Omitting the overestimation at DYS389-I, DYS391, and DYS481 – forward stutter is typically flat – DYS392 and DYS481 (trinucleotides) show an increasing trend.

Stutter Density Plot



Median stutter = 3.75%

(again, likely an overestimation)

Performance with related males

Father-Son Mutation Rates

Father-Son Mutations - Summary

	Meioses	Mutations	
AfAm	85	10	
Asian	101	16	
Caucasian	100	6	
Hispanic	100	12	
Sum	386	44	11.40%

	Father to Son		Father to Son	
	+1 repeat	-1 repeat	+2 repeat	-2 repeat
DYS439	2	4		
DYS389II	3	2		
DYS481	3	2		
DYS570	3	2		
DYS576	2	2		
DYS389I	2	2		
Y-GATA-H4	1	1	1	1
DYS458	2	1		
DYS549	1	1		
DYS635	1	1		
DYS19	1	0		
DYS390	0	1		
DYS643	1	0		
DYS385a	1	0		
DYS448	0	0		
DYS391	0	0		
DYS533	0	0		
DYS438	0	0		
DYS437	0	0		
DYS392	0	0		
DYS393	0	0		
DYS385b	0	0		
DYS456	0	0		
(sum)	23	23	1	1

	Mutations
DYS439	6
DYS389II	5
DYS481	5
→ **DYS570**	5
→ **DYS576**	4
DYS389I	4
Y-GATA-H4	4
DYS458	3
DYS549	2
DYS635	2
DYS19	1
DYS390	1
DYS643	1
DYS385a	1
DYS448	0
DYS391	0
DYS533	0
DYS438	0
DYS437	0
DYS392	0
DYS393	0
DYS385b	0
DYS456	0

G Model
FSIGEN-744; No. of Pages 11
ARTICLE IN PRESS
Forensic Science International: Genetics xxx (2011) xxx–xxx
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

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,*]

Summary

- Chromosomal positions were defined for the 23 regions of the Y-chromosome amplified by PowerPlex Y23
- The additional 6 Y-STRs with PowerPlex Y23 improve haplotype resolution
  - In our NIST 1032 males, haplotype discrimination capacity went from 98.2% (Yfiler) to 99.7% (PPY23)
- Excellent sensitivity was observed with PowerPlex Y23
- Extensive stutter calculations were performed
  - Forward stutter for the trinucleotide DYS481 was characterized

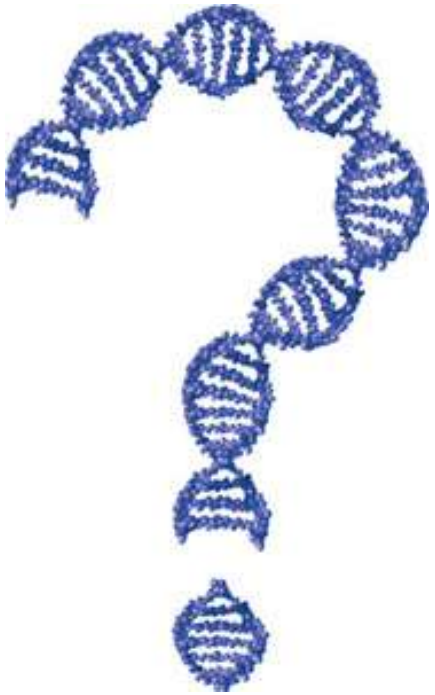
Thank You for Your Attention!

Our team publications and presentations are available at:
http://www.cstl.nist.gov/biotech/strbase/NISTpub.htm

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

Thanks to Promega for providing the PowerPlex Y23 kits

Questions?



john.butler@nist.gov
301-975-4049

michael.coble@nist.gov
301-975-4330