Sequencing of 58 STRs using the Illumina® ForenSeq™ workflow and analysis of the data with the STRinNGS v.1.0 software

C. Hussing, C. Huber, R. Bytyci, H.S. Mogensen, C. Børsting, and N. Morling

Section of Forensic Genetics, Department of Forensic Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Denmark

Why do STR sequencing?

Example:

Among 120 Danes, 48 individuals (20.16 %) had allele 14 in the locus D8S1179:

Capillary elec-trophoresis (CE)		Sequence	Frequency in 120 Danes
D8S1179[14]	→	TCTA[14]	2.5 % (6 individuals)
		TCTA[1]TCTG[1]TCTA[12]	11.8 % (28 individuals)
		TCTA[1]TCTG[1]TGTA[1]TCTA[11]	1.7 % (4 individuals)
		TCTA[2]TCTG[1]TCTA[11]	4.2 % (10 individuals)
			20.2 % (48 individuals)

One CE allele can correspond to several sequence alleles
→ higher diversity → higher possible discrimination power

Why do STR sequencing?

Higher discrimination power can be valuable in cases of

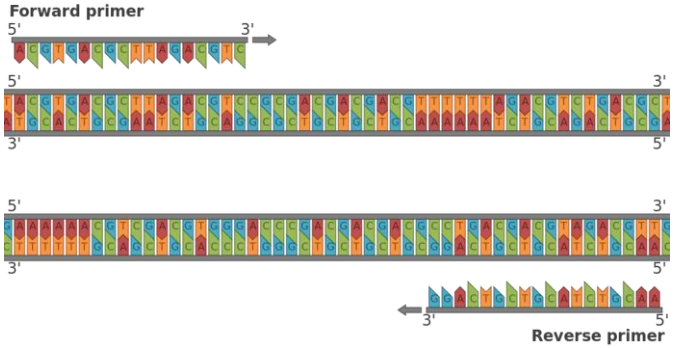
Mixtures: Separate DNA from two individuals, if possible

Partly degraded samples: Need to get as much information as possible from the few markers that can be typed

Complex relationship cases: Need to find genetic differences among closely related individuals

Sequencing method: The Illumina® ForenSeq™ workflow

1) ForenSeq™ library build
   - 58 STRs (27 autosomal, 24 Y-STRs, 7 X-STRs)
   - 94 identity-informative SNPs,
   - 56 ancestry-informative SNPs,
   - 24 phenotypic-informative SNPs
   - Amelogenin

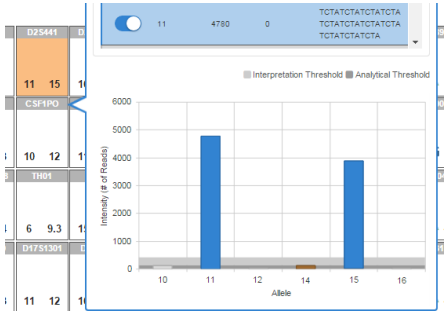


Source: www.wikipedia.org

2) Sequencing on the MiSeq FGx instrument

Source: www.illumina.com

3) Data analysis in ForenSeq™ Universal Analysis Software (FUAS)



Analysis of DNA sequences

Data analysis using FUAS

Reproducibility
CE-MPS concordance
Sensitivity / low template samples
DNA mixtures

Data analysis using STRinNGS v.1.0

Locus balance
Allele balances
SNPs in STR flanking regions

Reproducibility of DNA sequencing of STRs

30 individuals (12 males, 18 females, 1,308 genotypes):
Concordant: **98.6 %**, discordant: **1.4 %**

Marker	Duplicate 1 genotype	Duplicate 2 genotype	Sex
D9S1122	11, 13	11, 12, 13	-
	11, 12, 14	12, 13, 14	-
	11, 13	11, 12, 13	-
D17S1301	11, 13	11, 12, 13	-
	11, 13	11, 12, 13	-
D20S482	14, 16	14, 15, 16	-
	14, 16	14, 15, 16	-
	13, 15	13, 14, 15	-
D21S11	29, 32	29, 31, 32	-
	32, 32.2	31, 32, 32.2	-
DYS392	12, 13	10, 13	M
	14	13, 14	M
DXS10135	25, 34	25, 33, 34	F
DXS10103	17, 17	17, 19	F
	16, 16	17, 17	F
	15, 15	15, 20	F
	18, 19	-	F
	19, 20	-	F

Reproducibility of DNA sequencing of STRs

30 individuals (12 males, 18 females, 1,308 genotypes):
Concordant: **98.6 %**, discordant: **1.4 %**

Marker	Duplicate 1 genotype	Duplicate 2 genotype	Sex	
D9S1122	11, 13	11, 12, 13	-	
	11, 12, 14	12, 13, 14	-	*Called stutter*
	11, 13	11, 12, 13	-	
D17S1301	11, 13	11, 12, 13	-	
	11, 13	11, 12, 13	-	*Called stutter*
D20S482	14, 16	14, 15, 16	-	
	14, 16	14, 15, 16	-	*Called stutter*
	13, 15	13, 14, 15	-	
D21S11	29, 32	29, 31, 32	-	
	32, 32.2	31, 32, 32.2	-	*Called stutter*
DYS392	12, 13	10, 13	M	
	14	13, 14	M	*Called stutter*
DXS10135	25, 34	25, 33, 34	F	*Called stutter*
DXS10103	17, 17	17, 19	F	
	16, 16	17, 17	F	
	15, 15	15, 20	F	
	18, 19	-	F	
	19, 20	-	F	*Locus drop-outs*

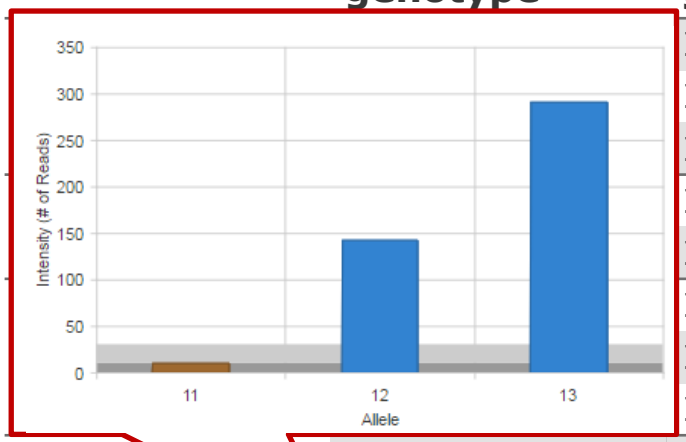
Reproducibility of DNA sequencing of STRs

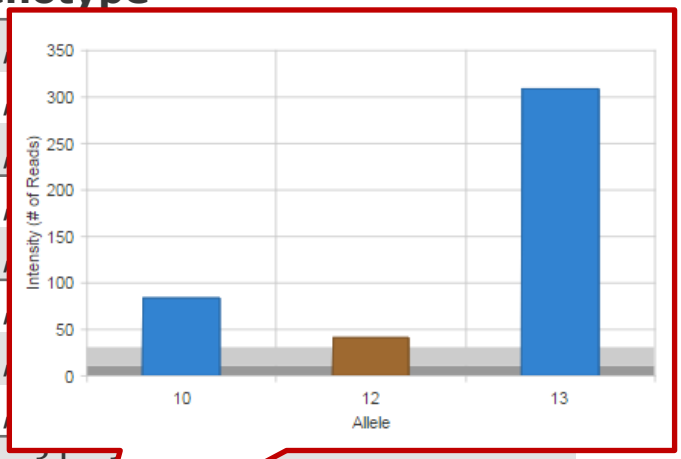
30 individuals (12 males, 18 females, 1,308 genotypes):
Concordant: **98.6 %**, discordant: **1.4 %**

Marker	Duplicate 1 genotype	Duplicate 2 genotype	Sex	
D9S1122	11, 13	11, 12, 13	-	
	11, 12, 14	12, 13, 14	-	*Called stutter*
	11, 13	11, 12, 13	-	
D17S1301	11, 13	11, 12, 13	-	
	11, 13	11, 12, 13	-	*Called stutter*
D20S482	14, 16	14, 15, 16	-	
	14, 16	14, 15, 16	-	*Called stutter*
	13, 15	13, 14, 15	-	
D21S11	29, 32	29, 31, 32	-	
	32, 32.2	31, 32, 32.2	-	*Called stutter*
DYS392	12, 13	10, 13	M	
	14	13, 14	M	*Called stutter*
DXS10135	25, 34	25, 33, 34	F	*Called stutter*
DXS10103	17, 17	17, 19	F	
	16, 16	17, 17	F	
	15, 15	15, 20	F	
	18, 19	-	F	*Locus drop-outs*
	19, 20	-	F	

Reproducibility of DNA sequencing of STRs

30 individuals (12 males, 18 females, 1,308 genotypes):
Concordant: **98.6 %**, discordant: **1.4 %**





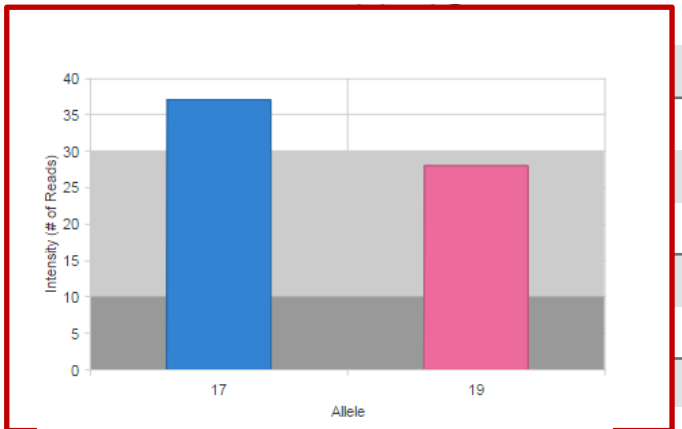
Marker	Duplicate 1 genotype	Duplicate 2 genotype	Sex
		11,	
		12,	
		11,	
		11,	
		11,	
		14,	
		14,	
		13,	
D21S11	29, 32	29, 31,	-
	32, 32.2	31, 32,	-
DYS392	12, 13	10, 13	M
	14	13, 14	M
DXS10135	25, 34	25, 33, 34	F
DXS10103	17, 17	17, 19	F
	16, 16	17, 17	F
	15, 15	15, 20	F
	18, 19	-	F
	19, 20	-	F

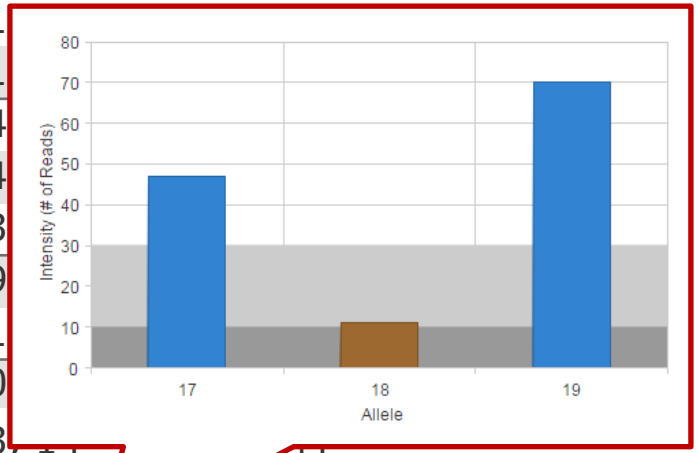
Called stutter and allele drop-in

Reproducibility of DNA sequencing of STRs

30 individuals (12 males, 18 females, 1,308 genotypes):
Concordant: **98.6 %**, discordant: **1.4 %**

Marker	Duplicate 1 genotype	Duplicate 2 genotype	Sex
	11, 13	11, 12, 13	-
D9S1122	11, 12, 14	12, 13, 14	-
	11, 13	11, 12, 13	-
		11	
		11	
		14	
		14	
		13	
		29	
		31	
		10	
	14	13, 14	
DXS10135	25, 34	25, 33,	F
	17, 17	17, 19	F
	16, 16	17, 17	F
DXS10103	15, 15	15, 20	F
	18, 19	-	F
	19, 20	-	F



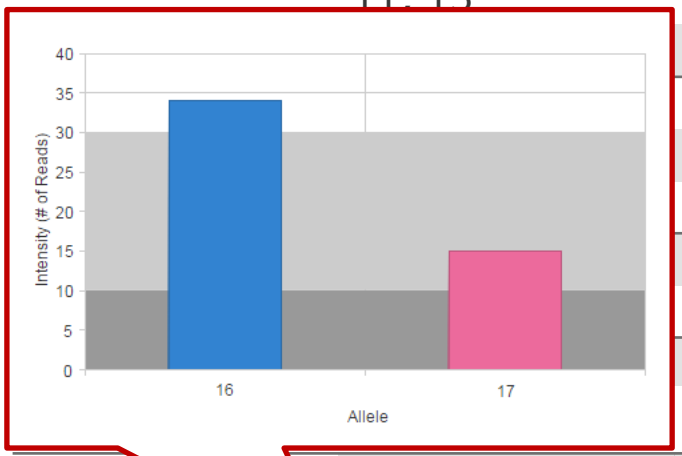


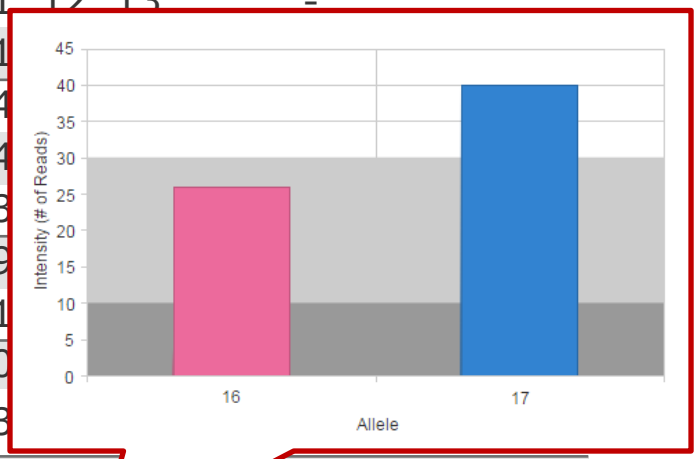
Allele drop-out (low coverage)

Reproducibility of DNA sequencing of STRs

30 individuals (12 males, 18 females, 1,308 genotypes):
Concordant: **98.6 %**, discordant: **1.4 %**

Marker	Duplicate 1 genotype	Duplicate 2 genotype	Sex
D9S1122	11, 13	11, 12, 13	-
	11, 12, 14	12, 13, 14	-
	11, 13	11, 12, 13	-
	11, 13	11, 12, 13	-
		11	
		14	
		14	
		13	
		29	
		31	
		10	
		13	
DXS101	25, 34	25, 33,	F
DXS10103	17, 17	17, 19	F
	16, 16	17, 17	F
	15, 15	15, 20	F
	18, 19	-	F
	19, 20	-	F



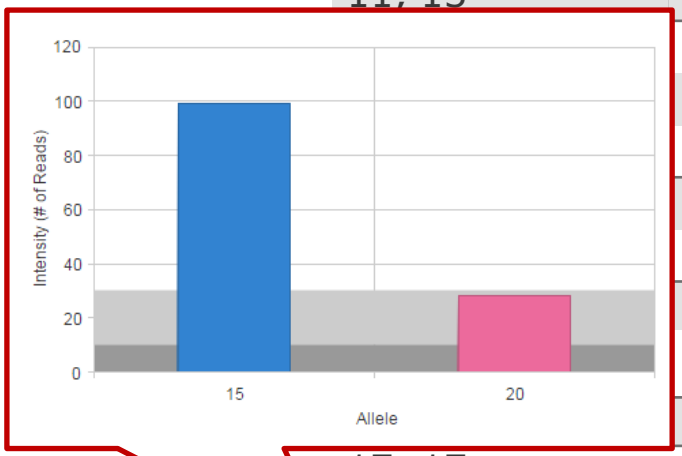


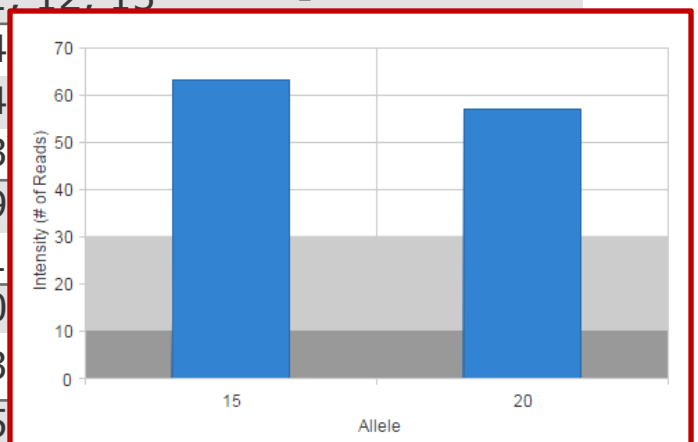
Allele drop-outs (low coverage)

Reproducibility of DNA sequencing of STRs

30 individuals (12 males, 18 females, 1,308 genotypes):
Concordant: **98.6 %**, discordant: **1.4 %**

Marker	Duplicate 1 genotype	Duplicate 2 genotype	Sex
D9S1122	11, 13	11, 12, 13	-
	11, 12, 14	12, 13, 14	-
	11, 13	11, 12, 13	-
D17S1301	11, 13	11, 12, 13	-
	11, 13	11, 12, 13	-
	14, 14	13, 29	-
	13, 31	10, 13	-
	25, 25	17, 19	F
DXS10103	16, 16	17, 17	F
	15, 15	15, 20	F
	18, 19	-	F
	19, 20	-	F





Allele drop-outs (low coverage)

Reproducibility of DNA sequencing of STRs

30 individuals (12 males, 18 females, 1,308 genotypes):
Concordant: **98.6 %**, discordant: **1.4 %**

Marker	Duplicate 1 genotype	Duplicate 2 genotype	Sex	
D9S1122	11, 13	11, 12, 13	-	
	11, 12, 14	12, 13, 14	-	*Called stutter*
	11, 13	11, 12, 13	-	
D17S1301	11, 13	11, 12, 13	-	*Called stutter*
	11, 13	11, 12, 13	-	
D20S482	14, 16	14, 15, 16	-	
	14, 16	14, 15, 16	-	*Called stutter*
	13, 15	13, 14, 15	-	
D21S11	29, 32	29, 31, 32	-	*Called stutter*
	32, 32.2	31, 32, 32.2	-	
DYS392	12, 13	10, 13	M	*Called stutter and allele drop-in*
	14	13, 14	M	
DXS10135	25, 34	25, 33, 34	F	*Called stutter*
DXS10103	17, 17	17, 19	F	
	16, 16	17, 17	F	*Locus and allele drop-outs (low coverage)*
	15, 15	15, 20	F	
	18, 19	-	F	
	19, 20	-	F	

CE-MPS concordance (15 autosomal markers)

30 individuals (12 males, 18 females, 450 genotypes):
Concordant: **99.1 %**, discordant: **0.9 %**

Marker	CE genotype	ForenSeq genotype
D2S1338	LDO	19, 25
FGA	LDO	19, 20
D21S11	29, 32	29, 31, 32
	32, 32.2	31, 32, 32.2

CE-MPS concordance (15 autosomal markers)

30 individuals (12 males, 18 females, 450 genotypes):
Concordant: **99.1 %**, discordant: **0.9 %**

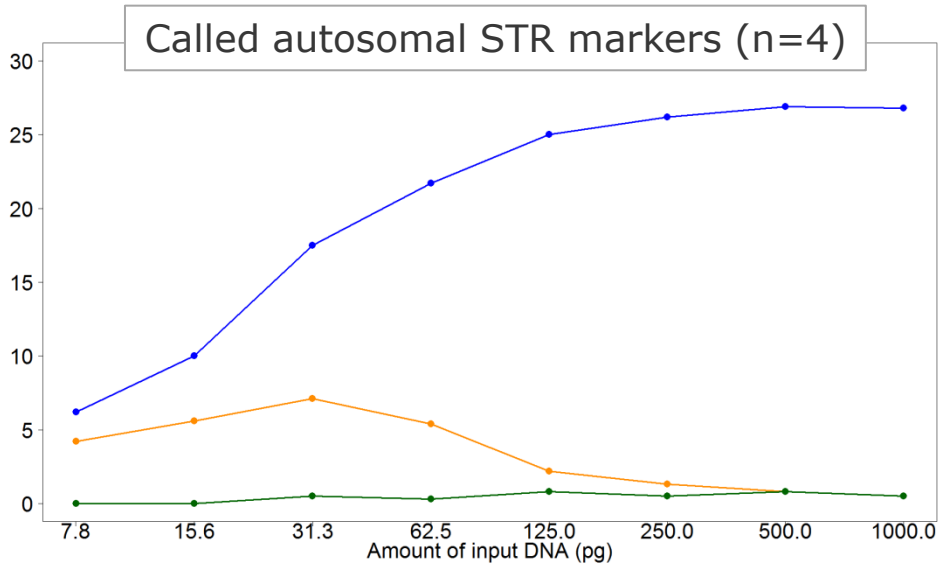
Marker	CE genotype	ForenSeq genotype	
D2S1338	LDO	19, 25	*Locus drop-out*
FGA	LDO	19, 20	*Locus drop-out*
D21S11	29, 32	29, 31, 32	*Called stutters*
	32, 32.2	31, 32, 32.2	

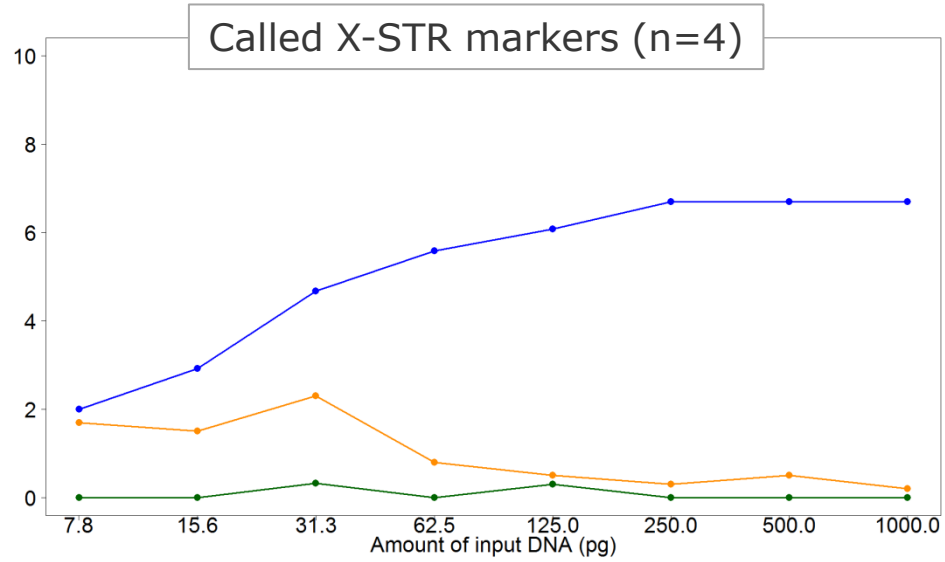
Low template DNA samples

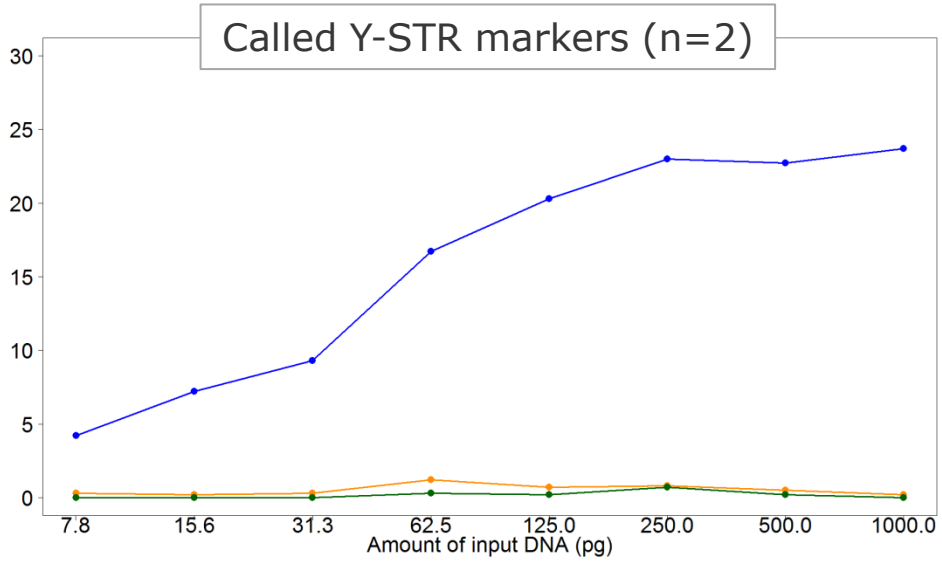
of typed markers

# of markers with allele drop-out

# of markers with allele drop-in







Low template DNA samples

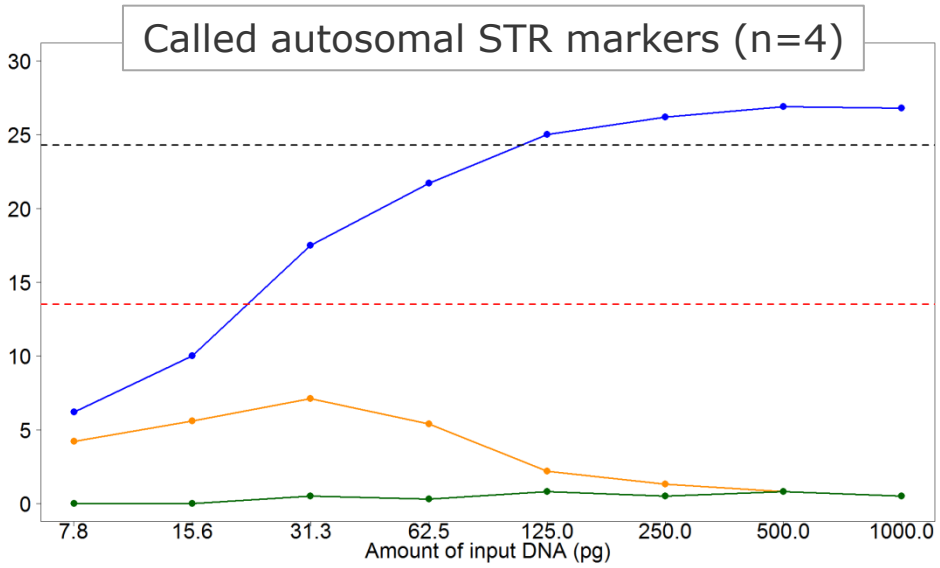
of typed markers

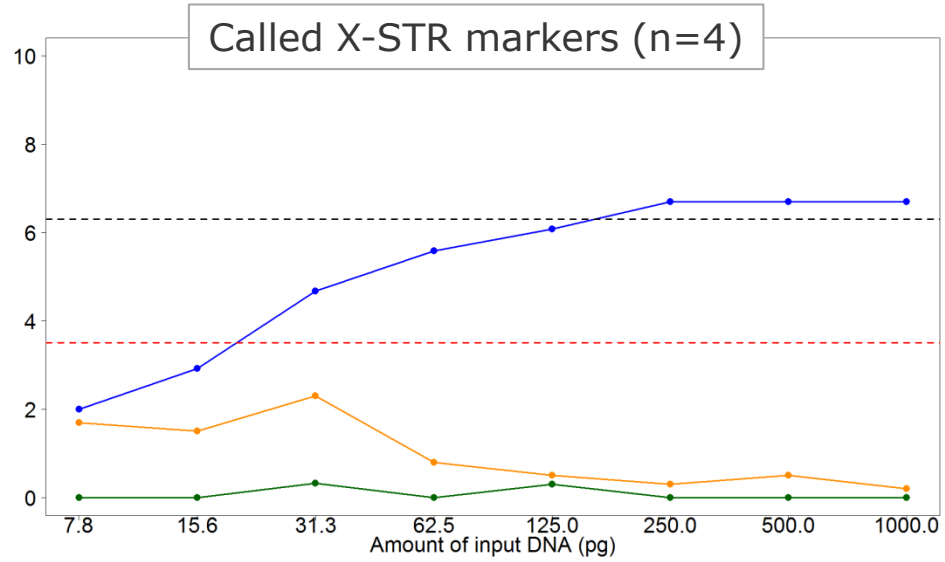
# of markers with allele drop-out

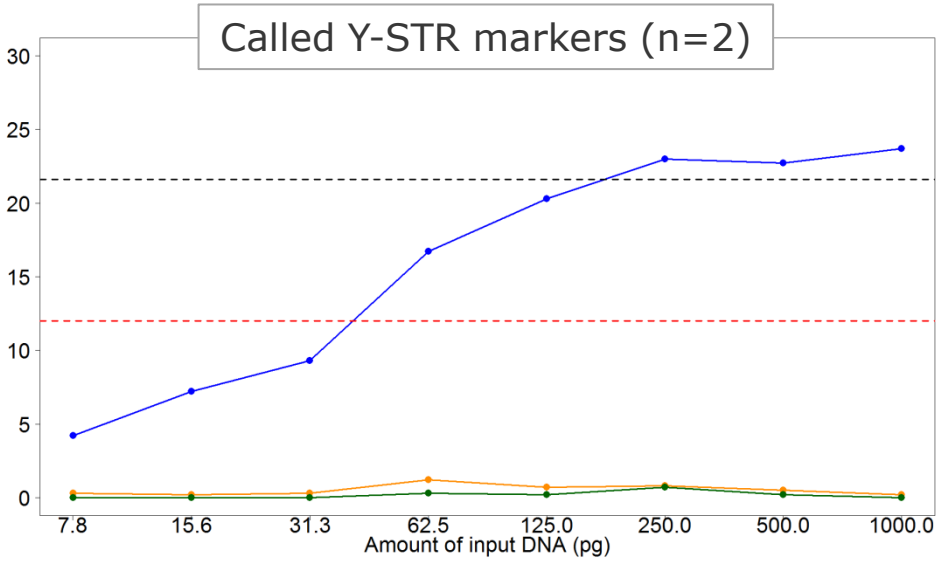
# of markers with allele drop-in

90 % of markers typed

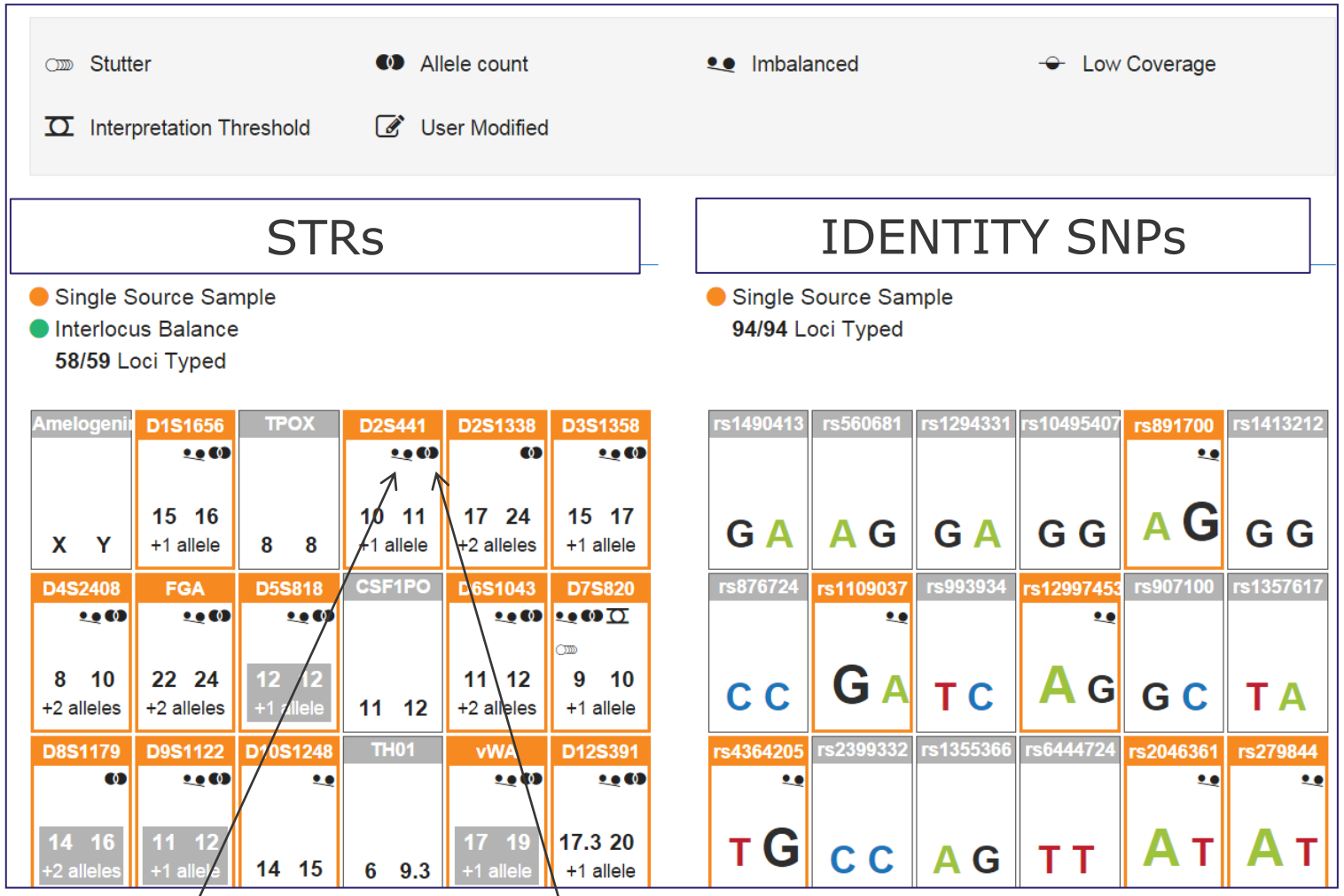
50 % of markers typed







DNA mixtures – STRs & SNPs - FUAS screenshot



”Imbalanced alleles” warning

”Multiple alleles” warning

DNA mixtures – autosomal STRs

Male:female mixture ratio	Average # of "Multiple alleles" warnings	Average # of "Imbalanced" warnings
1,000:1	1.5	0.5
100:1	3	3
50:1	8.5	3
25:1	9.5	4.5
12:1	16	9
6:1	19	19
3:1	19	20
1:1	20	17.5
1:3	19	21
1:6	17	19.5
1:12	14.5	13.5
1:25	10	5
1:50	4.5	2.5
1:100	1	1
1:1,000	1	2.5
30 single contributor samples	0.5	2.2

DNA mixtures – autosomal STRs

Male:female mixture ratio	Average # of "Multiple alleles" warnings	Average # of "Imbalanced" warnings
1,000:1	1.5	0.5
100:1	3	3
50:1	8.5	3
25:1	9.5	4.5
12:1	16	9
6:1	19	19
3:1	19	20
1:1	20	17.5
1:3	19	21
1:6	17	19.5
1:12	14.5	13.5
1:25	10	5
1:50	4.5	2.5
1:100	1	1
1:1,000	1	2.5
30 single contributor samples	0.5	2.2

Notable difference to single contributor samples

Calling STR alleles with STRinNGS v.1.0

In-house software for calling STR sequence alleles[1]

Predefined STR repeat structures and start and stop positions of flanking regions

Pipeline:

1) Identifies STR flanking regions

2) Counts STR repeat numbers and identifies SNPs in flanks

3) Provides detailed information for each unique sequence, e.g. locus coverage, allele coverage, and full STR and flanking sequences

[1]Friis SL et al.: Introduction of the Python script STRinNGS for analysis of STR regions in FASTQ or BAM files and expansion of the Danish STR sequence database to 11 STRs. Forensic Sci Int Genet. 2016; 21: 68-75.

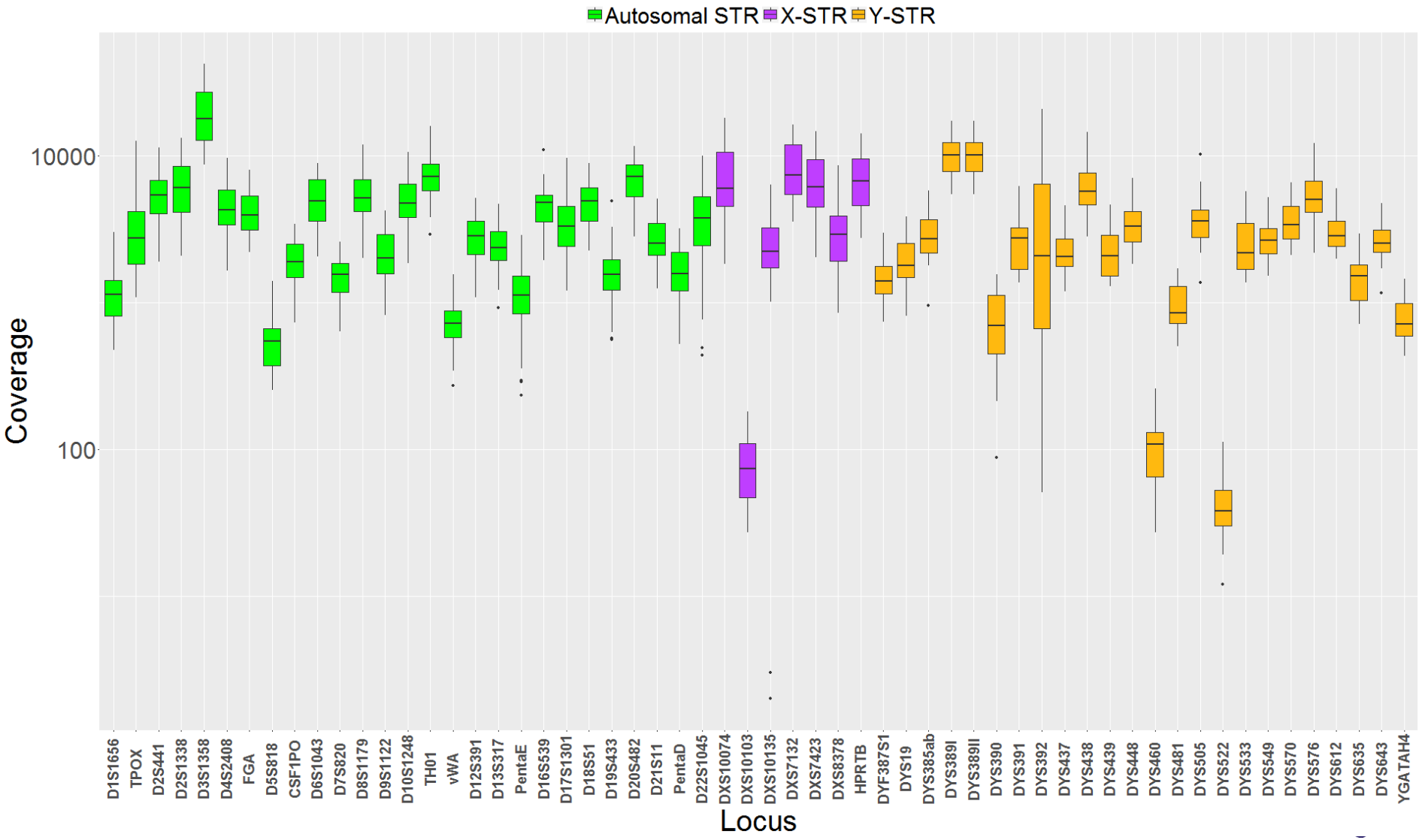
Calling STR alleles with STRinNGS v.1.0

STRinNGS v.1.0 (in-house) – features not in FUAS:

- Analyses of STR flanking regions
- Automatic naming of SNP-STR alleles
- Simultaneous obtainment of allele and locus coverage of all unique sequences from all samples in a run

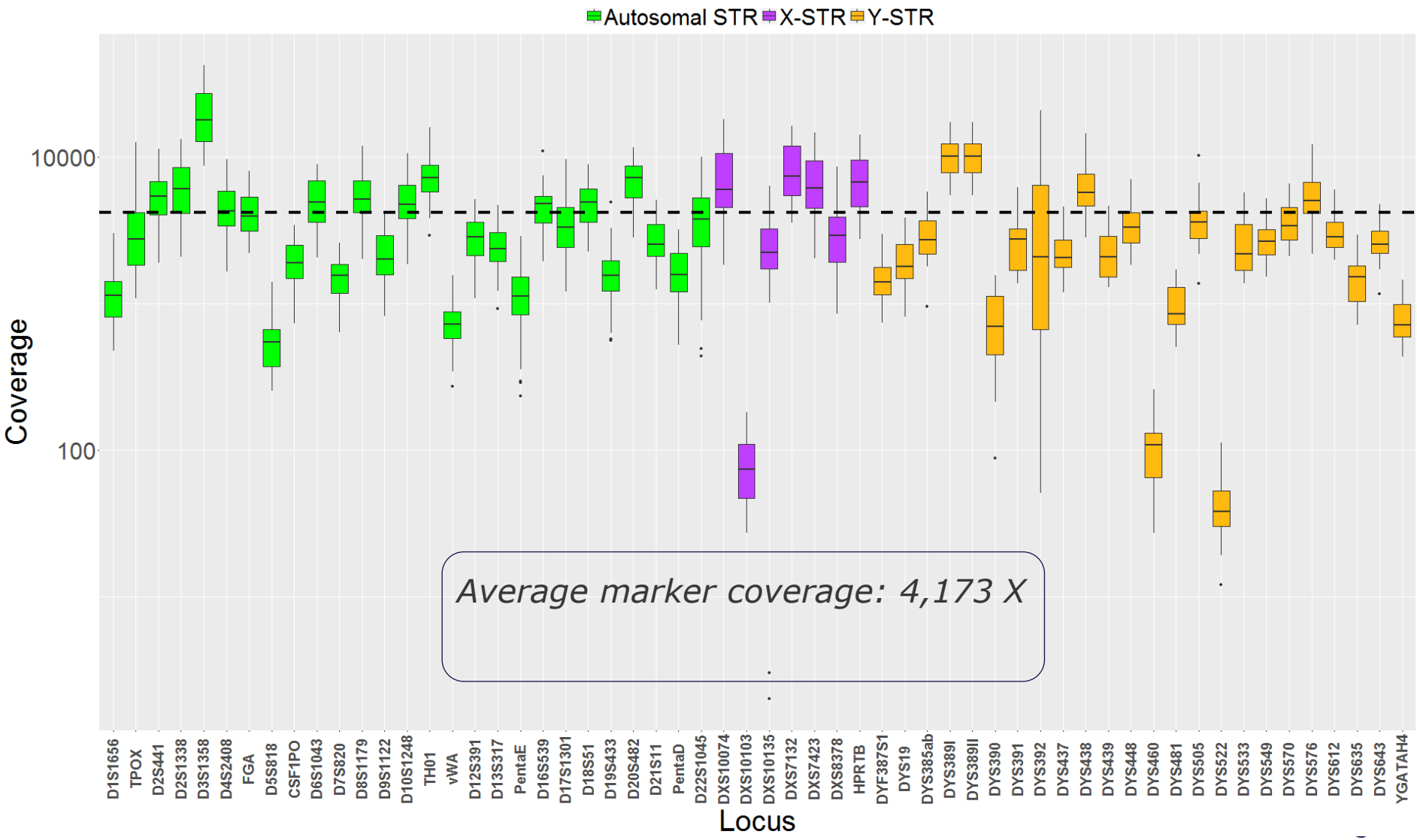
Coverage – numbers of relevant DNA sequences

30 individuals (12 males, 18 females)



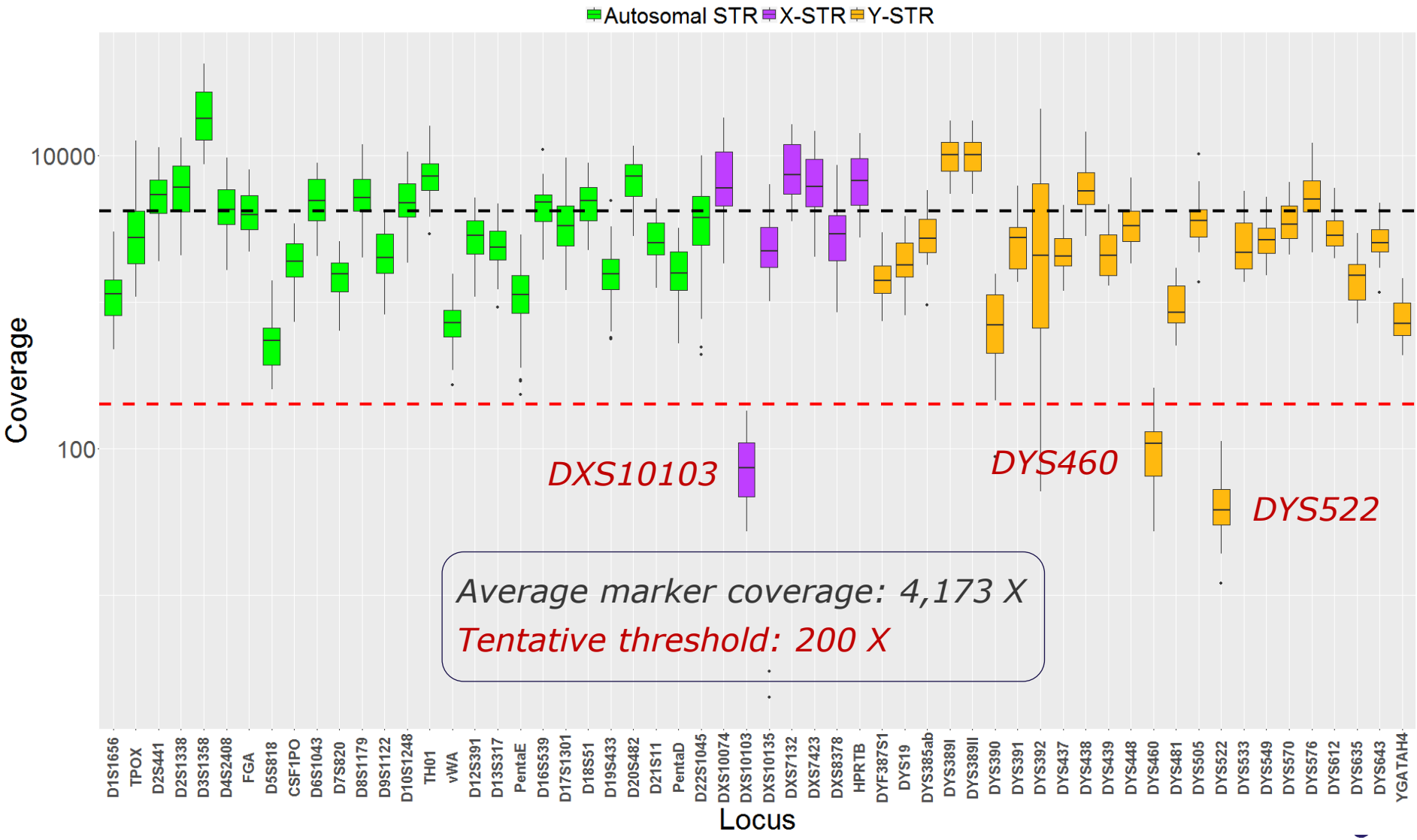
Coverage – numbers of relevant DNA sequences

30 individuals (12 males, 18 females)



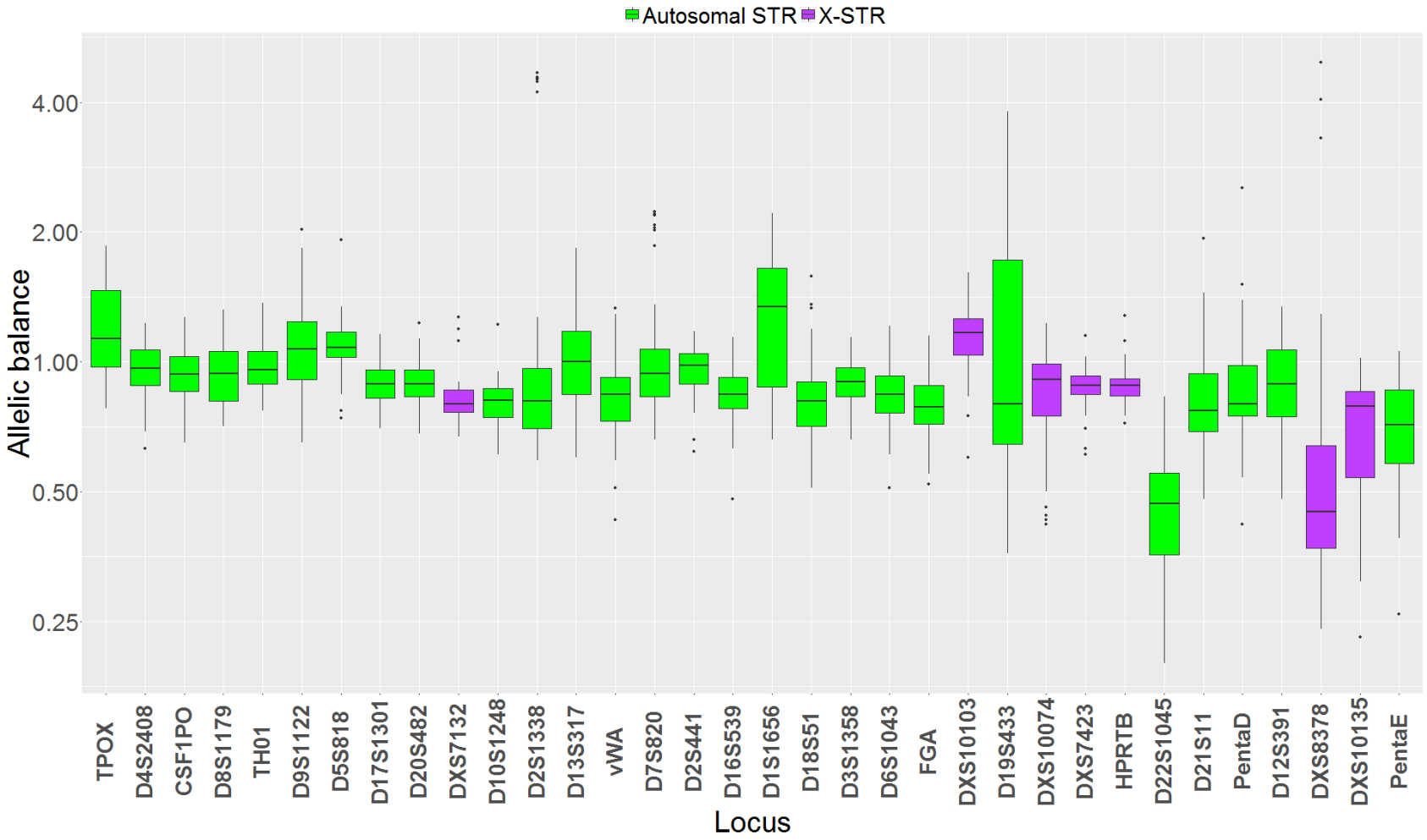
Coverage – numbers of relevant DNA sequences

30 individuals (12 males, 18 females)



Allele balances (longest allele / shortest allele)

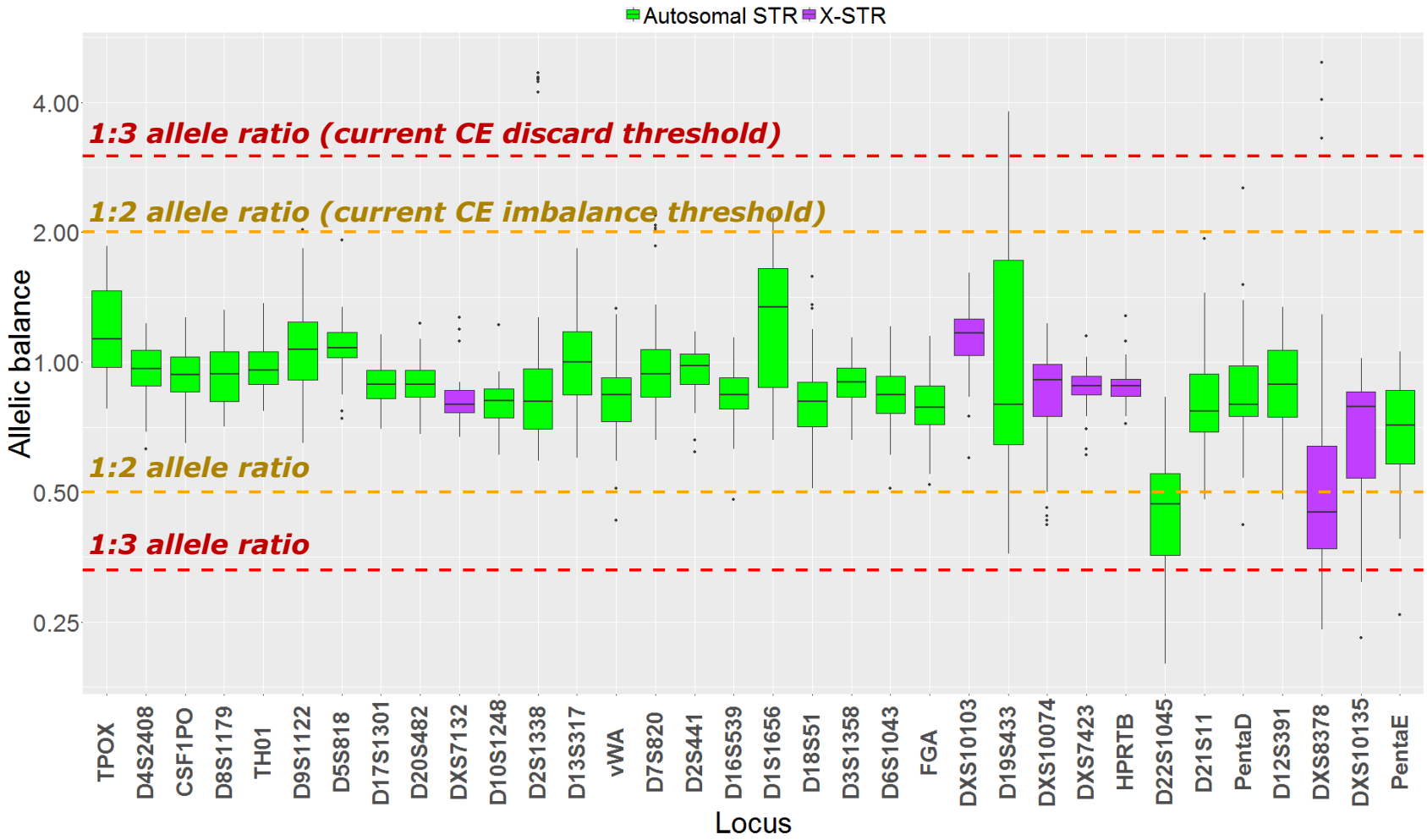
30 individuals (12 males, 18 females)



Increasing amplicon length

Allele balances (longest allele / shortest allele)

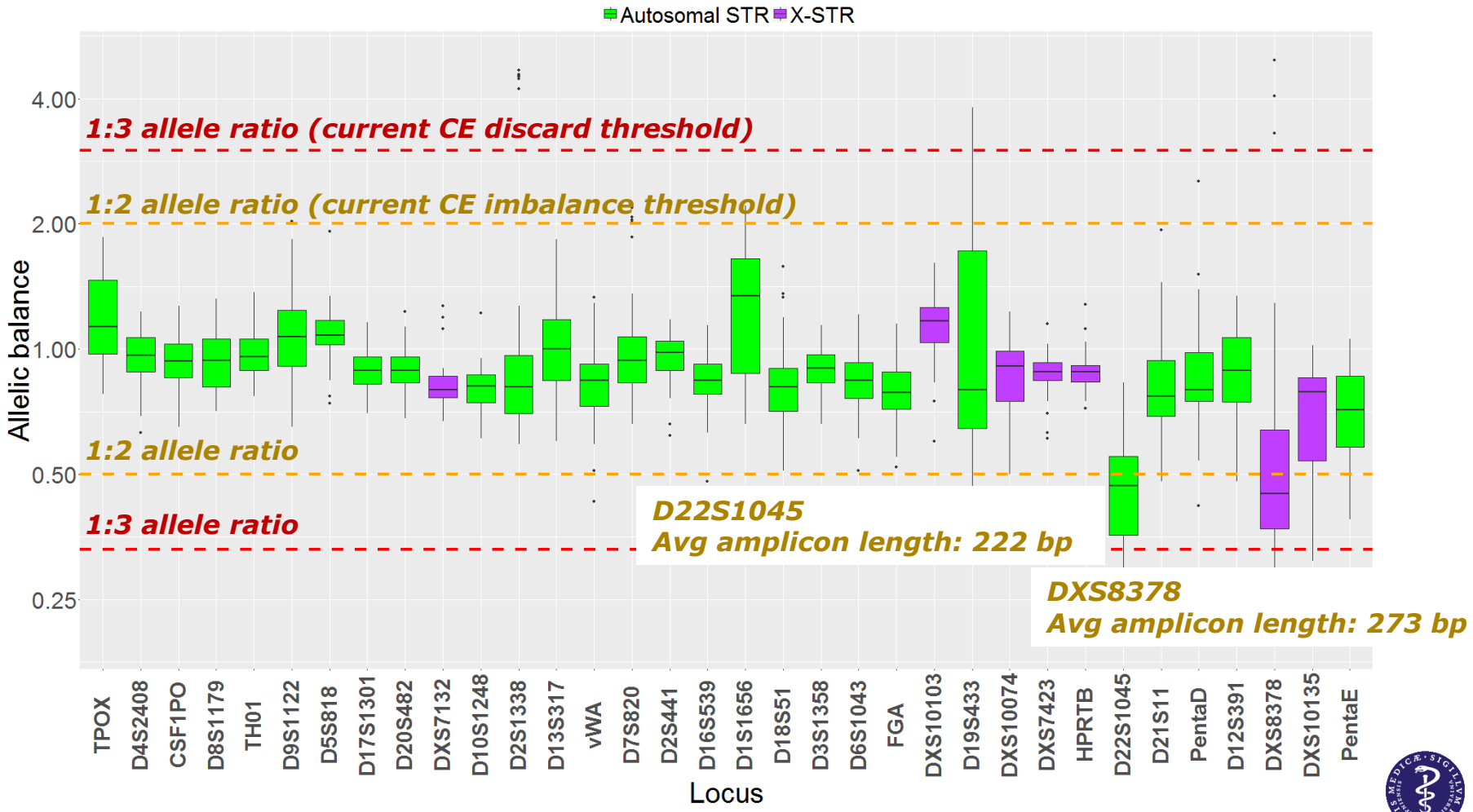
30 individuals (12 males, 18 females)



Increasing amplicon length

Allele balances (longest allele / shortest allele)

30 individuals (12 males, 18 females)



Conclusion

The ForenSeq™ kit gave good results in sequencing 58 STRs simultaneously

Lack of reproducibility was mainly due to calling of stutters as a true, third allele and low coverage in DXS10103

Few discordances to CE genotyping was observed (CE locus drop-out, stutters called as true alleles by ForenSeq™)

Above 90 % of markers were called at ≥250 pg input DNA, and above 50 % at ≥62.5 pg

The FUA software could effectively mark mixtures as imbalanced and/or containing more than two alleles in mixtures ranging from 1:1 to 1:25/1:50 ratios of DNA

Conclusion

STRinNGS v.1.0 proved to be efficient in analysing STRs

Acceptable locus and allele coverages were obtained except for the locus coverage of DXS10103, DYS460, and DYS522 (below 200 X)

A total 26 SNPs were observed in the flanking regions of 19 different STRs.

Acknowledgements

Members of the staff of the Section of Forensic Genetics, Faculty of Health and Medical Sciences, University of Copenhagen:

Technician Anja Ladegaard Jørgensen

Bioinformatician Carina Grøntved Jønck

Forensic Geneticist Eszter Rockenbauer, MSc, PhD

Forensic Geneticist Anders Buchard, MSc, PhD