

# Development of New miniSTR Loci for Improved Analysis of Degraded DNA Samples

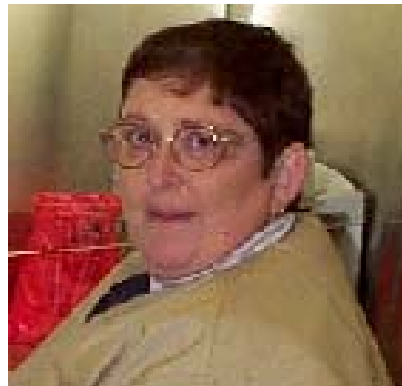
Michael D. Coble,

Peter M. Vallone and John M. Butler  
National Institute of Standards and Technology

# NIST Human Identity Project Team



John Butler



Margaret Kline



Jan Redman



Pete Vallone



Dave Duewer



Amy Decker



Mike Coble

## Former (Honorary) Project Team Members

Rich  
Schoske



Christian  
Ruitberg



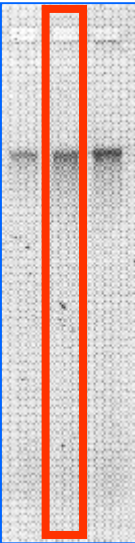
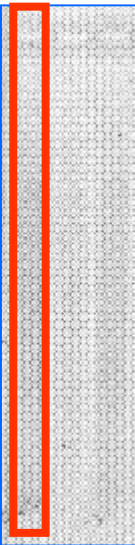
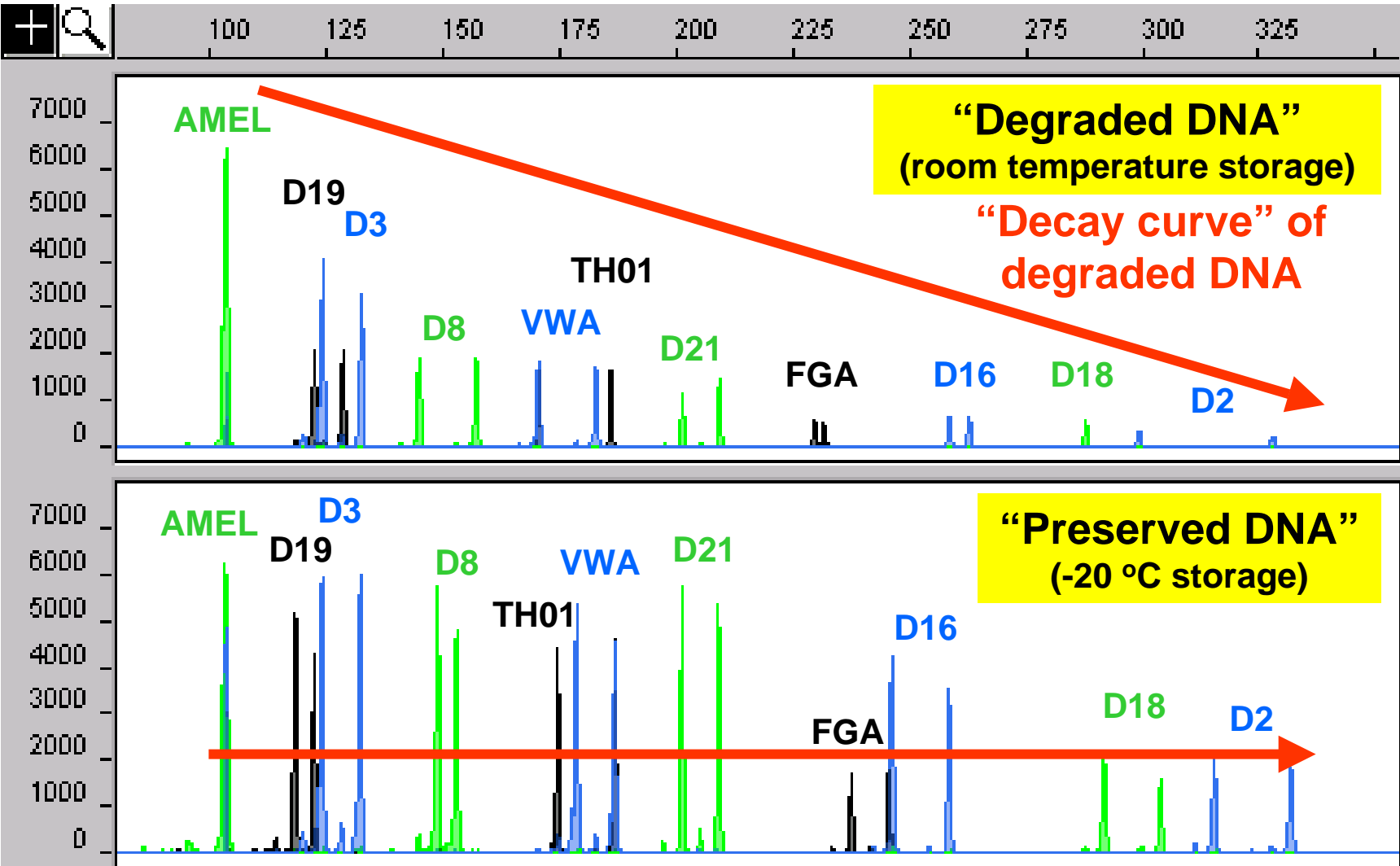
Jill  
Appleby



# Degraded DNA

*Loss of Signal for Larger PCR Products*

Yield gel results

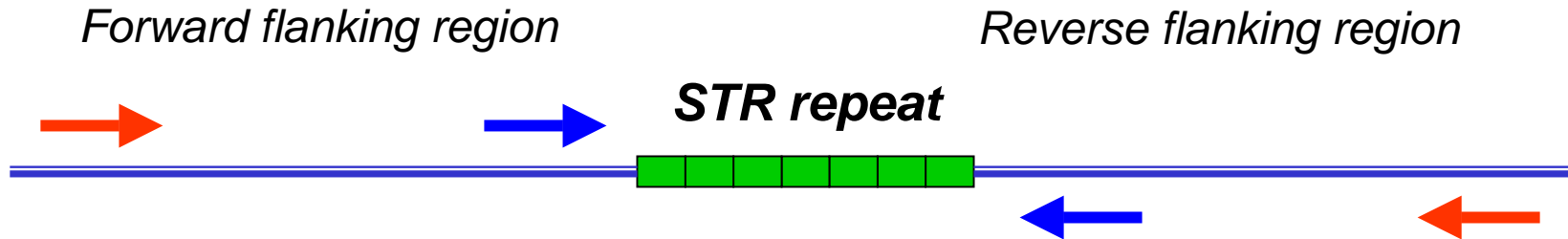


Results with SGM Plus STR kit (Applied Biosystems)

Data from a study done at NIST in May 2001

# STR Size Reduction

Through Moving Primer Positions Closer to Repeat



*Primer positions define PCR product size*

*Repeat information is independent of amplicon size*

## **Advantages of Approach:**

Size reduction enhances success rate with degraded DNA

Retains same marker information (database compatibility)

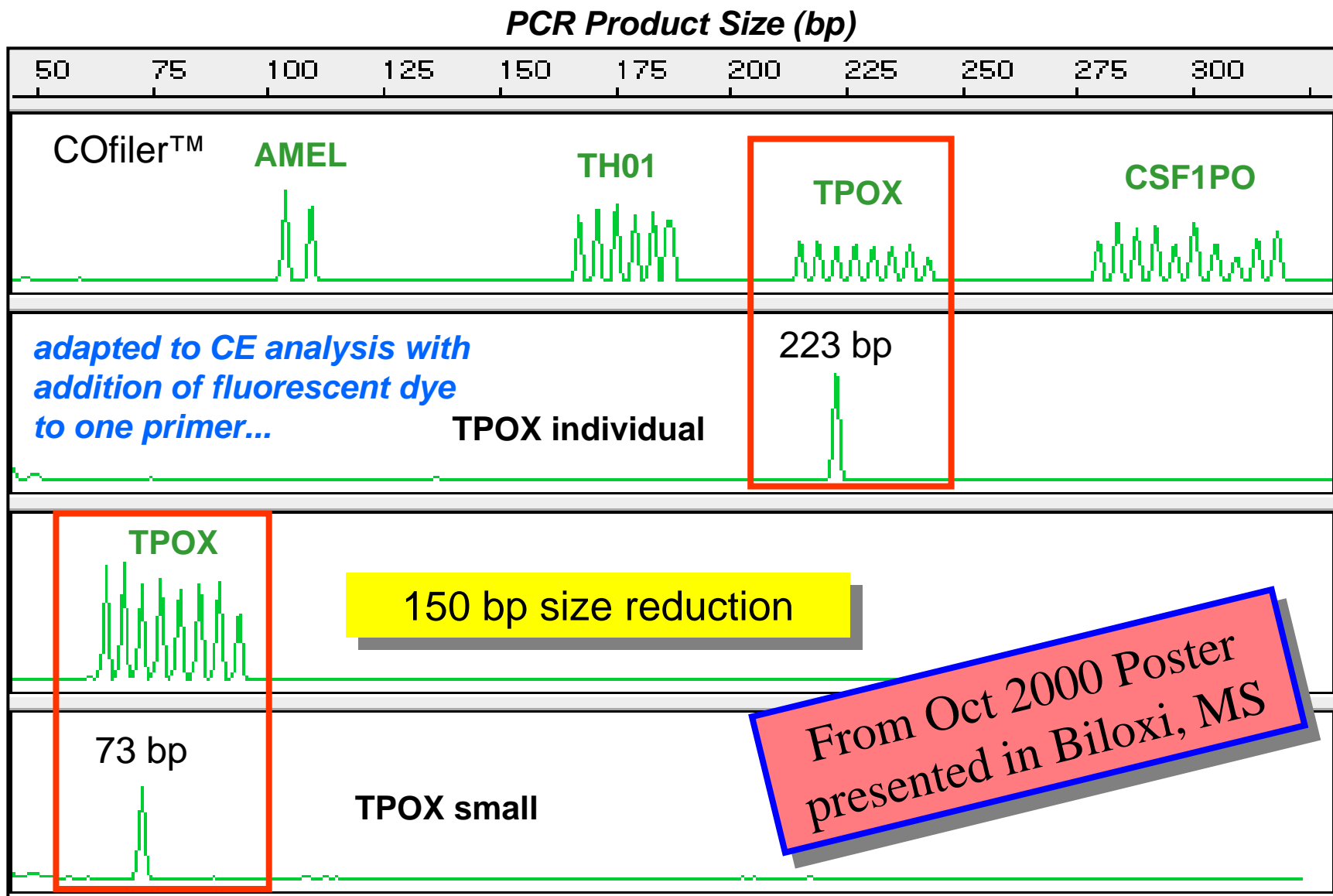
Uses highly polymorphic STR loci (high discriminatory power)

## Selected References on STRs with Degraded DNA

- Whitaker, J.P., et al. (1995) Short tandem repeat typing of bodies from a mass disaster: high success rate and characteristic amplification patterns in highly degraded samples. *BioTechniques* 18: 670-677
- Clayton, T.M., et al. (1995) Further validation of a quadruplex STR DNA typing system: a collaborative effort to identify victims of a mass disaster. *Forensic Sci.Int.* 76: 17-25
- Yoshida, K., et al. (1997) Evaluation of new primers for CSF1PO. *Int.J.Legal Med.* 110: 36-38
- Schmerer, W.M., et al. (1999) Optimized DNA extraction to improve reproducibility of short tandem repeat genotyping with highly degraded DNA as target. *Electrophoresis* 20: 1712-1716
- Wiegand, P. and Kleiber, M. (2001) Less is more--length reduction of STR amplicons using redesigned primers. *Int.J.Legal Med.* 114: 285-287

Smaller PCR product size improves success rates with degraded DNA

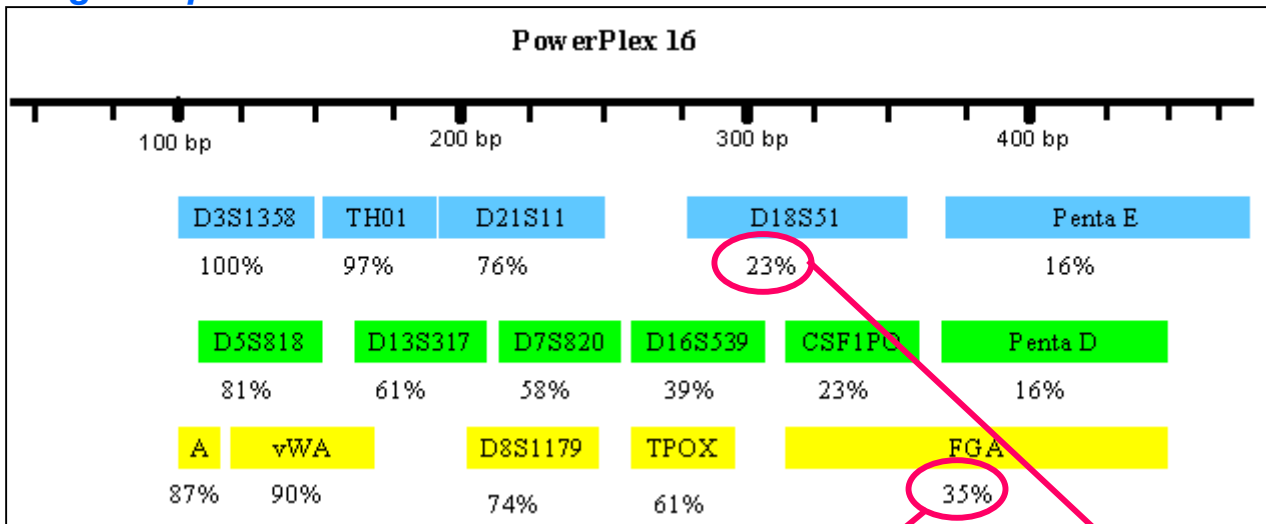
# Development of miniSTRs: Past Work



New primer sets are intended *to aid with typing degraded DNA samples* as well as future microchip CE and mass spectrometry applications...

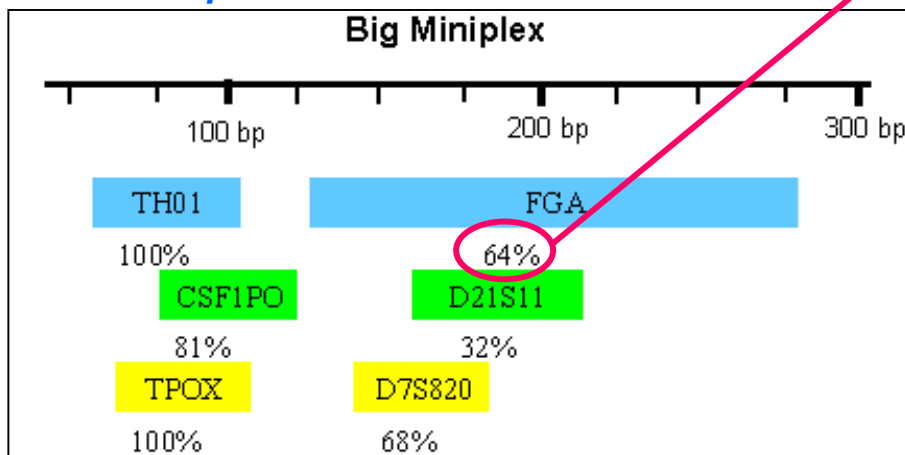
# Comparison of PCR Amplification Success Rates with Commercial Kit vs. miniSTR Assays

## Single amp for 15 STR loci



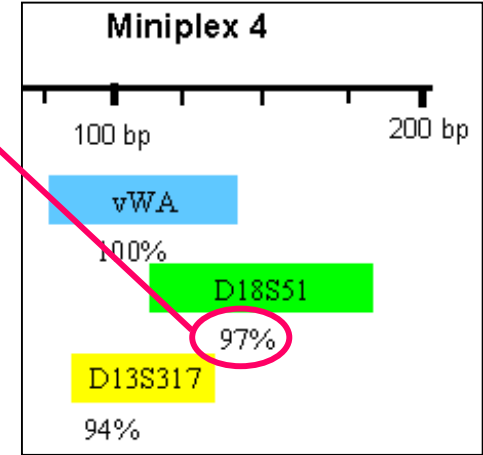
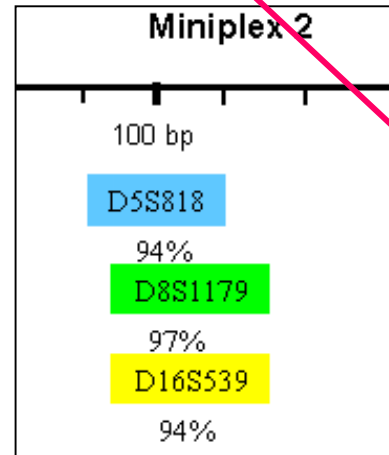
Study with 31 bones from the “Body Farm” (Knoxville, TN) and Franklin County Coroner’s Office (OH)

## Three amps for 12 STR loci



-183 bp

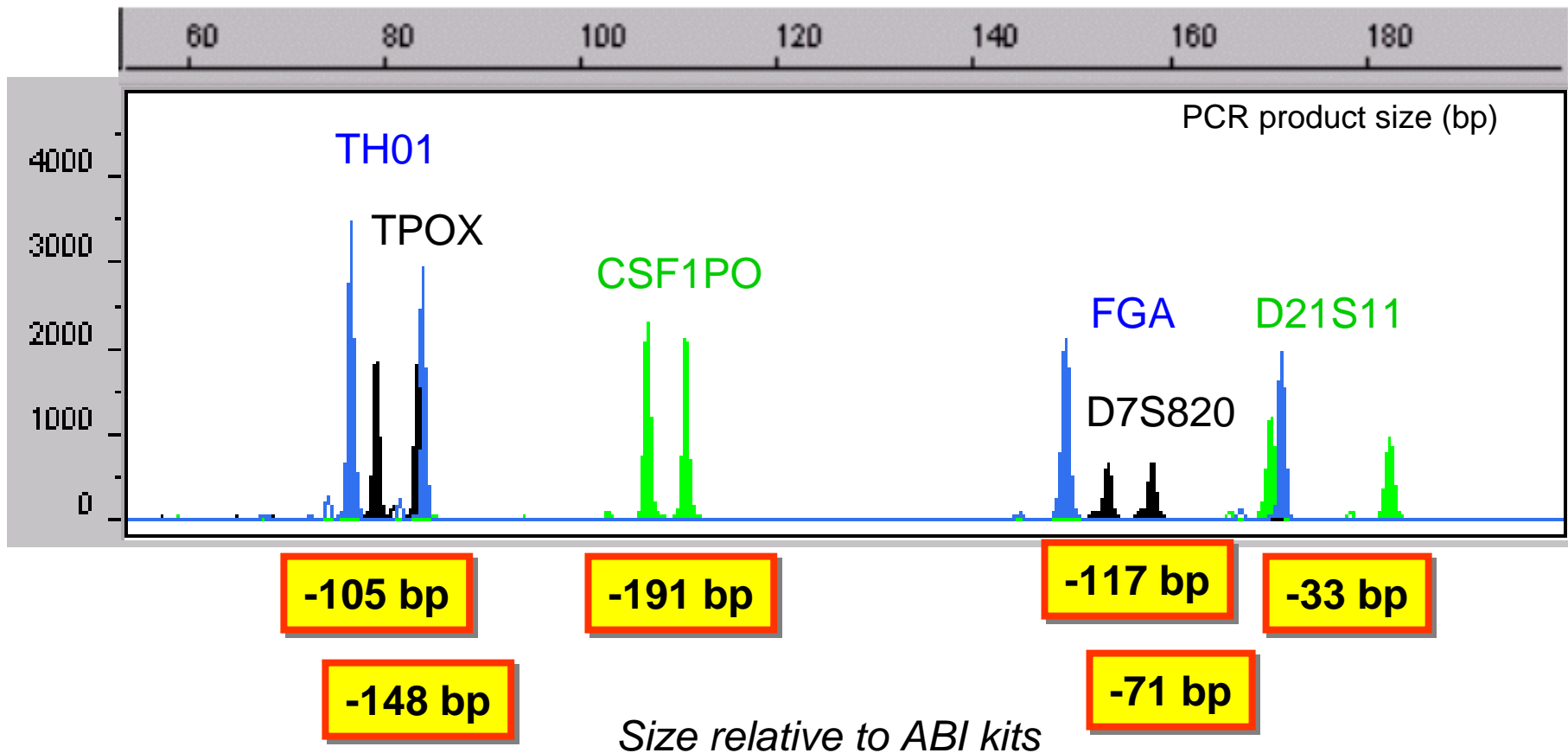
-173 bp



John M. Butler,<sup>1</sup> Ph.D.; Yin Shen,<sup>2,3</sup> Ph.D.; and Bruce R. McCord Ph.D.<sup>2</sup>

# The Development of Reduced Size STR Amplicons as Tools for Analysis of Degraded DNA\*

Describes new primer sequences for all CODIS loci and initial assays developed





# Reduction in PCR Product Size

<u>Locus</u>	<u>Size Difference (relative to ABI kits)</u>
TH01	-105 bp
FGA	-71 bp
CSF1PO	-191 bp
D21S11	-33 bp
TPOX	-148 bp
D7S820	-117 bp

**Not as much size reduction  
as other STR loci...**

# How Close Can a Stable Primer be Designed to the STR Repeat Region?

<u>Locus</u>		<u>Distance 3'end from Repeat</u>	<u>Comment</u>
<b>CSF1PO</b>	F	<b>14</b>	<i>partial repeat just 5' of repeat</i>
	R	6	
<b>FGA</b>	F	3	
	R	<b>23</b>	<i>partial repeat just 3' of repeat</i>
<b>TH01</b>	F	0	
	R	1	
<b>TPOX</b>	F	-4	
	R	5	
<b>VWA</b>	F	0	
	R	0	
<b>D3S1358</b>	F	-1	
	R	-1	
<b>D5S818</b>	F	4	
	R	-5	
<b>D7S820</b>	F	4	
	R	<b>65</b>	<i>polyA stretch just 3' of repeat</i>

# Problems with Large Allele Spreads

STR Locus	GenBank Accession	GenBank Allele	Allele Range	Allele Spread
CSF1PO	X14720	12	6–16	40 bp
FGA	M64982	21	12.2–51.2	156 bp
TH01	D00269	9	3–14	44 bp
TPOX	M68651	11	5–14	36 bp
vWA	M25858	18	10–25	60 bp
D3S1358	NT_005997	18	8–20	48 bp
D5S818	AC008512	11	7–16	36 bp
D7S820	AC004848	13	5–15	40 bp
D8S1179	AF216671	13	7–19	48 bp
D13S317	AL353628	11	5–16	44 bp
D16S539	AC024591	11	5–15	40 bp
D18S51	AP001534	18	7–27	80 bp
D21S11	AP000433	29	24–38.2	58 bp
Penta D	AP001752	13	2.2–17	73 bp
Penta E	AC027004	5	5–24	95 bp
D2S1338	AC010136	20	15–28	52 bp

# Why go beyond CODIS loci?

“STRs have proven to be highly successful [for mass disasters] in the past e.g. Waco disaster and various air disasters. However, even if the DNA is high quality there are occasions when there are insufficient family members available to achieve a high level of confidence with an association.”

Gill, P., Werrett, D.J., Budowle, B. and Guerrieri, R. (2004) **An assessment of whether SNPs will replace STRs in national DNA databases-Joint considerations of the DNA working group of the European Network of Forensic Science Institutes (ENFSI) and the Scientific Working Group on DNA Analysis Methods (SWGDM).** *Science&Justice*, 44(1): 51-53.

# Why go beyond CODIS loci?

“To achieve this purpose, either **new STRs** could be developed, or alternatively, existing STRs could be supplemented with a **SNP panel.**”

“There also efforts for modifying existing STR panels by decreasing the size amplicons by designing new primers.”

Gill, P., Werrett, D.J., Budowle, B. and Guerrieri, R. (2004) **An assessment of whether SNPs will replace STRs in national DNA databases-Joint considerations of the DNA working group of the European Network of Forensic Science Institutes (ENFSI) and the Scientific Working Group on DNA Analysis Methods (SWGDM).** *Science&Justice*, 44(1), *in press*.

# Why go beyond CODIS loci?

- Desirable to have markers unlinked from CODIS loci (different chromosomes) for some applications
- Small size ranges to aid amplification from degraded DNA samples

# Characterization of New miniSTR Loci

- Candidate STR marker selection
- Chromosomal locations and marker characteristics
- PCR primer design
- Initial testing results
- Population testing
- Allelic ladder construction
- Miniplex assay performance

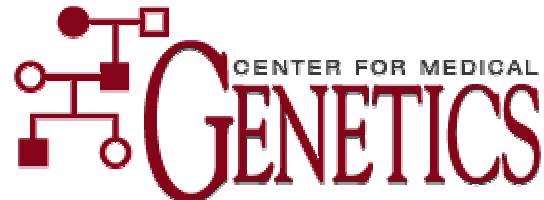
# Characterization of New miniSTR Loci

- Candidate STR marker selection
- Chromosomal locations and marker characteristics
- PCR primer design
- Initial testing results
- Population testing
- Allelic ladder construction
- Miniplex assay performance



# Candidate STR marker selection

<http://research.marshfieldclinic.org/genetics/>



1000 North Oak Avenue | Marshfield, WI 54449-5790  
Telephone : ( 715 ) 387-9150 | Fax : ( 715 ) 389-5757

<http://www.decode.com/>



<http://www.cidr.jhmi.edu/>



*Center for Inherited  
Disease Research*

# Candidate STR marker selection

SCIENCE VOL 298 20 DECEMBER 2002

## Genetic Structure of Human Populations

Noah A. Rosenberg,<sup>1\*</sup> Jonathan K. Pritchard,<sup>2</sup> James L. Weber,<sup>3</sup>  
Howard M. Cann,<sup>4</sup> Kenneth K. Kidd,<sup>5</sup> Lev A. Zhivotovsky,<sup>6</sup>  
Marcus W. Feldman<sup>7</sup>

1: Am J Phys Anthropol. 2003 Nov;122(3):259-68.

[Related Articles, Links](#)



Reconstruction of human evolutionary tree using polymorphic autosomal microsatellites.

Ayub Q, Mansoor A, Ismail M, Khaliq S, Mohyuddin A, Hameed A, Mazhar K, Rehman S, Siddiqi S, Papaioannou M, Piazza A, Cavalli-Sforza LL, Mehdi SQ.

# Characterization of New miniSTR Loci

Rosenberg *et al.* 2002 – 1062 samples; 377 STRs; diverse populations

Locus name	Alternate name	Heterozygosity	Number of alleles	Chromosome
D6S1017	GGAT3H10	0.748	9	6
D6S1009	GATA32B03	0.748	13	6
D8S261	AFM123XG5	0.747	17	8
D18S1390	18QTEL11	0.747	15	18
D13S1807	GATA11C08	0.747	9	13
D12S1052	GATA26D02	0.747	8	12
D16S2616	ATA41E04	0.746	11	16
D2S1780	GATA72G11	0.746	13	2
NA-D1S-3	GATA133A08	0.745	12	1
NA-D8S-2	GAAT1A4	0.745	9	8
D4S1625	GATA107	0.745	10	4
D18S851	GATA6D09	0.745	12	18
D8S1136	GATA41A01	0.745	11	8
NA-D5S-1	ATA52D02	0.744	27	5
D3S3038	GATA73D01	0.743	14	3
D15S1515	GATA197B10	0.743	9	15
D16S2624	GATA81D12	0.742	8	16
D2S2972	GATA176C01	0.741	14	2
D13S895	GGAA22G01	0.740	11	13
D11S1998	GATA23E06	0.740	9	11

Focus on:

High Heterozygosity

Small # of Alleles

Tetranucleotide Repeats

# Characterization of New miniSTR Loci

Rosenberg *et al.* 2002 – 1062 samples; 377 STRs; diverse populations

Locus name	Alternate name	Heterozygosity	Number of alleles	Chromosome
D6S1017	GGAT3H10	0.748	9	6
D6S1009	GATA32B03	0.748	13	6
D8S261	AFM123XG5	0.747	17	8
D18S1390	18QTEL11	0.747	15	18
D12S1052	GATA26D02	0.747	8	12
D16S2616	ATA41E04	0.746	11	16
D2S1780	GATA72G11	0.746	13	2
NA-D18-3	GATA133A08	0.745	12	1
NA-D8S-2	GAAT1A4	0.745	9	8
D4S1625	GATA107	0.745	10	4
D18S851	GATA6D09	0.745	12	18
D8S1136	GATA41A01	0.745	11	8
NA-D5S-1	ATA52D02	0.744	27	5
D3S3038	GATA73D01	0.743	14	3
D15S1515	GATA197B10	0.743	9	15
D16S2624	GATA81D12	0.742	8	16
D2S2972	GATA176C01	0.741	14	2
D13S895	GGAA22G01	0.740	11	13
D11S1998	GATA23E06	0.740	9	11

Focus on:

High Heterozygosity

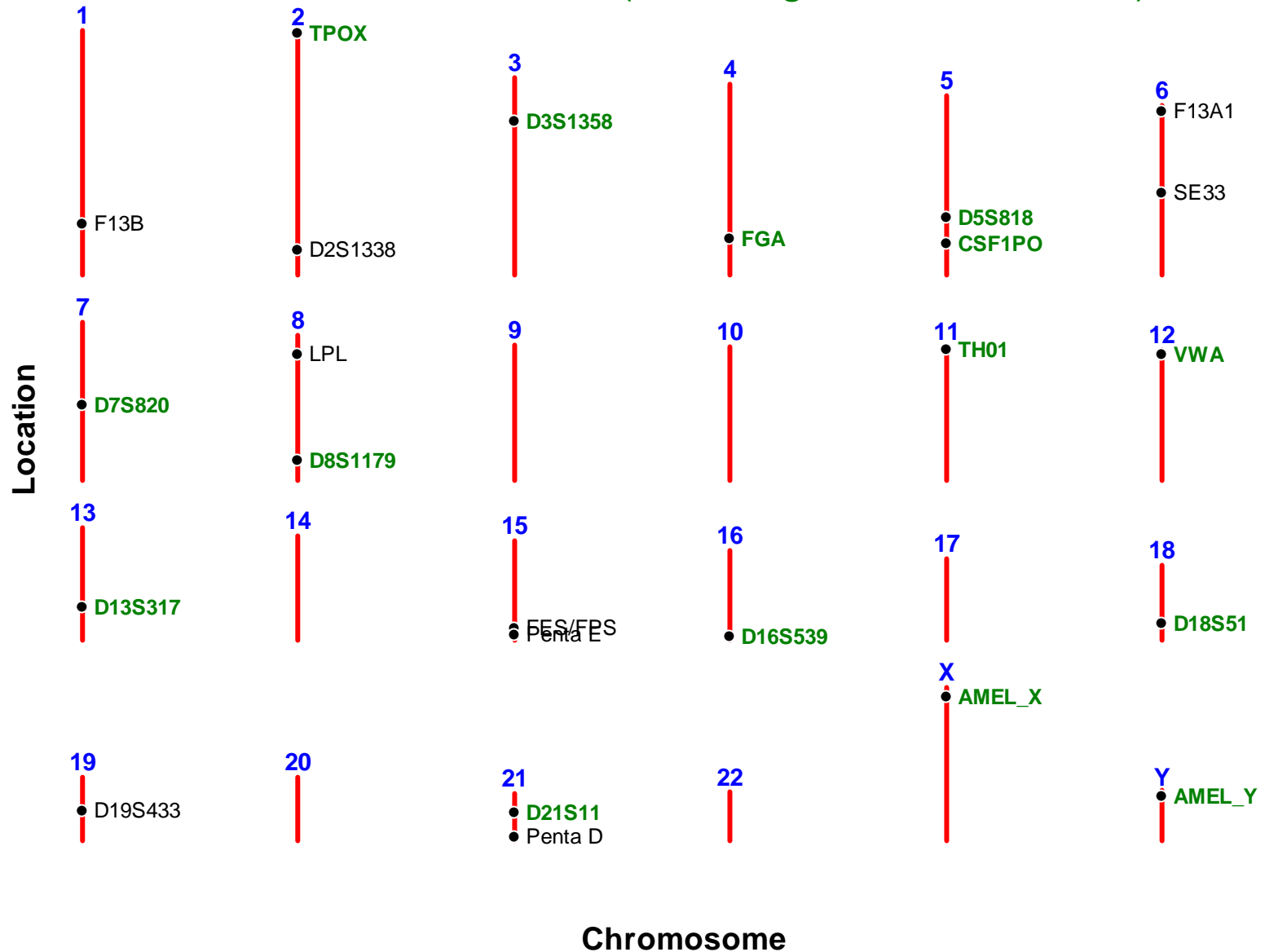
Small # of Alleles

Tetranucleotide Repeats

# Characterization of New miniSTR Loci

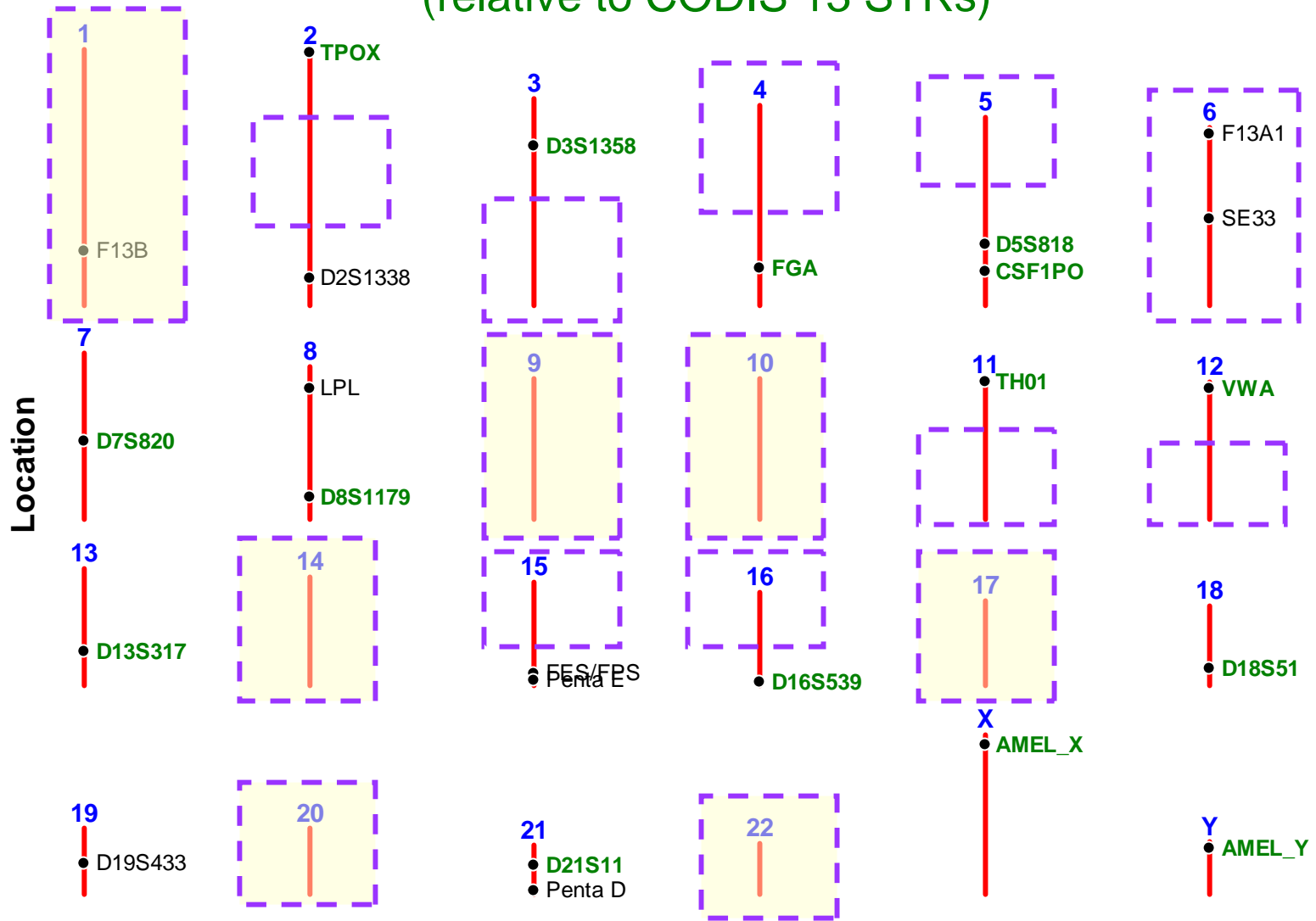
- Candidate STR marker selection
- Chromosomal locations and marker characteristics
- PCR primer design
- Initial testing results
- Population testing
- Allelic ladder construction
- Miniplex assay performance

# Commercial STR Kit Loci Positions (including CODIS 13 STRs)



*Positions determined along July 2003 Human Genome Reference Sequence (NCBI Build 34)*

# Locations of Focus for New miniSTR Loci (relative to CODIS 13 STRs)



Chromosome

# Characterization of New miniSTR Loci

- Candidate STR marker selection
- Chromosomal locations and marker characteristics
- PCR primer design
- Initial testing results
- Population testing
- Allelic ladder construction
- Miniplex assay performance



# PCR Primer Design

<http://frodo.wi.mit.edu/cgi-bin/primer3/primer3.cgi/>

## Primer3

[disclaimer](#)

[source code](#)

pick primers from a DNA sequence (see [NEW](#))

[cautions](#)

Paste source sequence below (5'->3', string of ACGTNacgtn -- other letters treated as N -- numbers and blanks ignored). FASTA format ok. Please N-out undesirable sequence (vector, ALUs, LINEs, etc.) or use a [Mispriming Library \(repeat library\)](#):

Pick left primer or use left primer below.

Pick hybridization probe (internal oligo) or use oligo below.

Pick right primer or use right primer below (5'->3' on opposite strand).

Pick Primers

Reset Form

**Sequence Id:**  A string to identify your output.

**Targets:**  E.g. 50,2 requires primers to surround the 2 bases at positions 50 and 51. Or mark the [source sequence](#) with [ and ]: e.g. ...ATCT[CCCC]TCAT.. means that primers must flank the central CCCC.

**Excluded Regions:**  E.g. 401,7 68,3 forbids selection of primers in the 7 bases starting at 401 and the 3 bases at 68. Or mark the [source sequence](#) with < and >: e.g. ...ATCT<CCCC>TCAT.. forbids primers in the central CCCC.

[NEW](#) **Product Size Ranges**

[Click here to specify the min, opt, and max product sizes only if you absolutely must. Using them is too slow \(and too computationally intensive for our server\).](#)

**Number To Return:**  **Max 3' Stability:**

**Max Mispriming:**  **Pair Max Mispriming:**

# PCR Primer Design

NCBI Entrez Nucleotide

Erbez PubMed Nucleotide Probe Genome Structure PMC

Search Nucleotide for [Go] [Clear]

Limits Preview/Index History Clipboard

Display FASTA Show: 20 Send to File Get Subsequence

1: [NT\\_086775](#). Reports Homo sapiens chro...[gi:51468451]

```
>gi|51468451|ref|NT_086775.1|Hs10_86451 Homo sapiens chromosome 10 genomic contig, alternate assembly
CGCCACTGCACTCCAGCCTGGGTGACAGAGCGAGACTCTGTCTCAAAAAAAAAAAAAAAAAAAAAAAAAA
GAGTATTGATAGTTTAAACCTACTACAATTTTGCATCTAGTTTCTCAAGTAGTTCTGTGGATTTAGAACC
AACATTTGCTGAGGGGAAAGAATTACCAACATCAAAGTCTTAGGATGTGTGAGGGCACCAGAAGAGAAGT
TTTGTCACTGTAAAATAACTCTCTTGGCAGAGCATGATAGAGACTGTAAACAGCAGGTTACTTCAATGGT
TATGTGCGATTCAAGTGAAGATCTCTACATTAATTGAAGCACAACTCAAGCAATTCATGTAGCTCTGATA
TGCTGGAGAACAATCTAGCAGTTAGCTATCAGCGGTGGTGTGGTTCCCTTTTAAAGCAAAGGTAGCTGTAG
TCTTCAGACTTGTGACCTGTGCAGCCTCCACATTCAAGTGAGGTTAAAGGGACTGCTGTTGGAACATGGA
TCACTCCAGGCCTATCTGTACACATGGACCCCTCTTCTGTGGCTTCCACTGCAACAAAAAGCCAATAAA
CACTAAATCAACATCTGGCAACTGTAAAAGAGGAATGATGCAACAGCAGGGCAGAAAAGATAAGAAAAGCT
CAAACCTTTCCTGGCATAAGCCACAAAAGTTCGTTGCTTAATGTAGTTATGAGATAATAAATATCTATGAAA
ACTATTTGTCAGTTTATTTAAGAGAGACGTGATAATAAGAAACCTTTAATTCTGCTTAATAAAAAGCAAC
TGATATTTATTTGGGGCAGAAAATATATAAAGTCTTTATATAAGCTGTTTCGTTTAAATCCTCTCAACAG
TCTCATGGCTAGGTATTATTATCCCACTTTACAATGAAAAAGAATGGTGGCAAAAATAATTTAATTTTTT
AAATATTTACAGGTAATAAGAGATGAAGCTGGAATCTGAAAAACAGAAAGTCTAACTTCTGAGACCATG
TTCTTAGTATTGTATATGCTTACATAGAGAGCCTTGTCTGTTCCCTACAGAATTTAGCTGTTTCCCTTTC
TTTATTGTTTCATCCCAGAAAATATCTTAGTAACTCTTTAAAAGTAAAGTAGATGGCTGTTAACTTA
AAATACTATATATGACACCATGTAATCTTTTTTTTTCTTTTTGAGATGGAGTTTTACTCTTGTGCCCAG
GCTGGAGTGCAATGGCGTGATCTCAGCTTACTGCAACCTCCACCTCCCAAGTTCAGCAATTCCTCTGCC
TCAGCCTCCCAAGTAGCTGGTATTACAGGCGTCCACCACACCCAGCTAATTTTTTTGTACTTTTAGT
AGAGATGGGGTTTCAACATGTTGGCCAGGCTGTCTCGAACTCCTGACCTCAGGTGATCCACGCACCTCA
GCTTCCCGAAGTGCTGAGATTACAGGCATGAGCCACTGTGCTGAGCTGACACCATGCAATCTTAAAAGCT
GAATGACTATTTTTGAAATCCTATGGATAATCTGGCAAAAATAATTGAGCTTATCCTTTCGTATAGTGA
```

Copy sequence data  
around the repeat from  
GenBank

<http://www.ncbi.nlm.nih.gov/>

# PCR Primer Design

<http://frodo.wi.mit.edu/cgi-bin/primer3/primer3.cgi/>

## Primer3

pick primers from a DNA sequence

[disclaimer](#)

[source](#)

[cautions](#)

[FAQ](#)

Paste source sequence below (5'→3', string of ACGTNacgtn -- other letters treated as N -- numbers and blanks ignored). FASTA format ok. Please N-out v  
ALUs, LINEs, etc.) or use a [Mispriming Library \(repeat library\)](#):

```
AACAAGATTTTCAGTAGTACTGGACATAGAGGACTTGATTCTTCCAATTCAGAACCCCTGATACAGGTCAACTACTGACAGTTCTGTTTTGCGGT
GGTTCCTAGTATGGTACCTGGCCAAGGGCACACTAGATCTTTGTCAAGGTAATGACTACTTTTTATTAAATGCTTTCCATGTATCAAGTTCTGT
GCCAAGCACTTGACATATATCATTATTTTATCCCGTGAAAGTAGTTATTGGTATCTTCATTTACAAAATAAAAAAAAAACAAGCTTAGTACTTAACT
CACTGCCTTGAACATAATTATTGCTTTAAAAGGTAGCTAGGATTCTTAATAGCTATTATTACAAAAGCATGAACAATCAGTAAAAAGCAAACTG
AGCATTAGCCCCAGGACCAATCTGGTCACAAAACATATTAATGAATTGAACAAATGAGTGAGTGGAAGGAAGGAAGGAAGGAAGGAAGGAAGGAA
GGAAGGAAGGAAGGAAGGAAATGAAGACAATACAACCAGAGTTGTTCTTTAATAACAAGACAAGGGAAAAAGAGAAGTGTCAAGAATAAGTGTT
```



Pick  
left



Pick  
hybridization



Pick  
right

Paste into Primer3

# PCR Primer Design

<http://frodo.wi.mit.edu/cgi-bin/primer3/primer3.cgi/>

[Sequence Id:](#)  A string to identify your output.

[Targets:](#)  E.g. 50,2 requires primers to surround the 2 bases at positions 50 and 51. Or TCAT.. means that primers must flank the central CCCC.

[Excluded Regions:](#)  E.g. 401,7 68,3 forbids selection of primers in the 7 bases starting at 401 and ...ATCT<CCCC>TCAT.. forbids primers in the central CCCC.

[Product Size Ranges:](#)

[Click here to specify the min, opt, and max product sizes only if you absolutely must. Using them is too slow \(and too c](#)

[Number To Return:](#)  [Max 3' Stability:](#)

[Max Mispriming:](#)  [Pair Max Mispriming:](#)

## General Primer Picking Conditions

[Primer Size](#) Min:  Opt:  Max:

[Primer Tm](#) Min:  Opt:  Max:  [Max Tm Difference:](#)

[Product Tm](#) Min:  Opt:  Max:

[Primer GC%](#) Min:  Opt:  Max:

[Max Self Complementarity:](#)  [Max 3' Self Complementarity:](#)



# PCR Primer Design

5'	11	21	31	41	51	61	71	81	91
1 TAAC TTCTGC ATTGAAGACG	ACTCATAAAT TGAGTATTTA	ATTATTTCCC TAATAAAGGG	TGCTTTTGCT ACAGAAACGA	TAAGCTATTG ATTCGATAAC	TCAGTCACAG AGTCAGTGTC	AAGCTCCATC TTCGAGGTAG	TTTTCATATG AAAAGTATAC	TGGGAGAACA ACCTCTTGT	ACAAAATCAG TGTTTTAGTC
101 GAAGAAGTTT CTTCTTCAA	CTTCCACCTA GAAGGTGGAT	CTCTATCTAT GAGATAGATA	CTATCTATCT GATAGATAGA	ATCCATCCAT TAGGTAGGTA	CCATTCTATC GGTAAGATAG	TCTCTCTCTC AGAGAGAGAG	TATCCATCCA ATAGGTAGGT	TTCTATCTAT AAGATAGATA	CTATCTATCT GATAGATAGA
201 ATCTGTCTAT TAGACAGATA	CTATCTAATC GATAGATTAG	TAATCTATCT ATTAGATAGA	ATCTATCCAT TAGATAGGTA	CCACCCATCC GGTGGGTAGG	ATCCACCCTA TAGGTGGGAT	TCCATCTATC AGGTAGATAG	CACCCATCCA GTGGGTAGGT	TCCATCCACC AGGTAGGTGG	CACCCATCTC GTGGGATAGA
301 CTCTGTCTCT GAGAC	CCATCCATCC GATAGATTAG	ATCCATCCAT ATTAGATAGA	CCATCCATCC TAGGTAGGTA	ATCCATCCAT TAGGTAGGTA	CCATCCATCT GGTAGGTAGA	GCCTATCTAT CGGATAGATA	ATACCCACCC TATGGGTGGG	TATCTATCTA ATAGATAGAT	TCTATCTATC AGATAGATAG
401 TATCTATCTA ATAGATAGAT	TCTATCTATC AGATAGATAG	TATCTATCTA ATAGATAGAT	TCTATCATCT AGATAGTAGA	ACCAGTTATC TGGTCAATAG	AGGGAGGGAG TCCCTCCCTC	ACAAGTGGGG TGTTCACCCC	AGGAGAGAAC TCCTCTCTTG	AGACTTGCCC TCTGAACGGG	TGTCTGAATT ACAGACTTAA
501 GTACTGGAAA CATGACCTTT	CTTAGTCTGT GAATCAGACA	ACTTGCTCTT TGAACGAGAA	GTTCTATAAC CAAGATATTG	TCTACTCAAG AGATGAGTTC	AGTACAATAT TCATGTTATA	TGTTAGAACC ACAATCTTGG	AAATGAATTT TTTACTTAAA	TTTTAATTC A AAAATTAAGT	ACTTGGCAGG TGAACCGTCC

8 GATA repeats

# PCR Primer Design

	5'	11	21	31	41	51	61	71	81	91		
1	TAAC TTCTGC ATTGAAGACG	ACTCATAAAT TGAGTATTTA	5 GATA repeats				AG doublet repeats				GGGAGAACCA CCCTCTTGT	ACAAAATCAG TGTTTTAGTC
101	GAAGAAGTTT CTTCTTCAAA	CTTCCACCTA GAAGGTGGAT	CTCTATCTAT GAGATAGATA	CTATCTATCT GATAGATAG	ATCCATCCAT TAGGTAGGTA	CCATTCTATC GGTAAGATAG	TCTCTCTCTC AGAGAGAGAG	TATCCATCCA ATAGGTAGGT	TTCTATCTAT AAGATAGATA	CTATCTATCT GATAGATAGA		
201	ATCTGTCTAT TAGACAGATA	CTATCTAATC GATAGATTAG	TAATCTATCT ATTAGATAGA	ATCTATCCAT TAGATAGGTA	CCACCCATCC GGTGGGTAGG	ATCCACCCTA TAGGTGGGAT	TCCATCTATC AGGTAGATAG	CACCCATCCA GTGGGTAGGT	TCCATCCACC AGGTAGGTGG	CACCCATCTC GTGGGATAGA		
301	CTCTGTCTCT GAGAC	CCATCCATCC GATAGATTAG	ATCCATCCAT GATAGATTAG	CCATCCATCC TAGGTAGGTA	ATCCATCCAT TAGGTAGGTA	CCATCCATCT GGTAGGTAGA	GCCTATCTAT CGGATAGATA	ATACCCACCC TATGGGTGGG	TATCTATCTA ATAGATAGAT	TCTATCTATC AGATAGATAG		
401	TATCTATCTA ATAGATAGAT	TCTATCTATC AGATAGATAG	TATCTATCTA ATAGATAGAT	TCTATCATCT AGATAGTAGA	ACCAGTTATC TGGTCAATAG	AGGGAGGGAG TCCCTCCCTC	ACAAGTGGGG TGTTCAACCC	AGGAGAGAAC TCCTCTCTTG	AGACTTGCCC TCTGAACGGG	TGTCTGAATT ACAGACTTAA		
501	GTACTGGAAA CATGACCTTT	CTTAGTCTGT GAATCAGACA	ACTTGCTCTT TGAACGAGAA	GTTCTATAAC CAAGATATTG	TCTACTCAAG AGATGAGTTC	AGTACAATAT TCATGTTATA	TGTTAGAACC ACAATCTTGG	AAATGAATTT TTTACTTAAA	TTTTAATTC AAAATTAAGT	ACTTGGCAGG TGAACCGTCC		

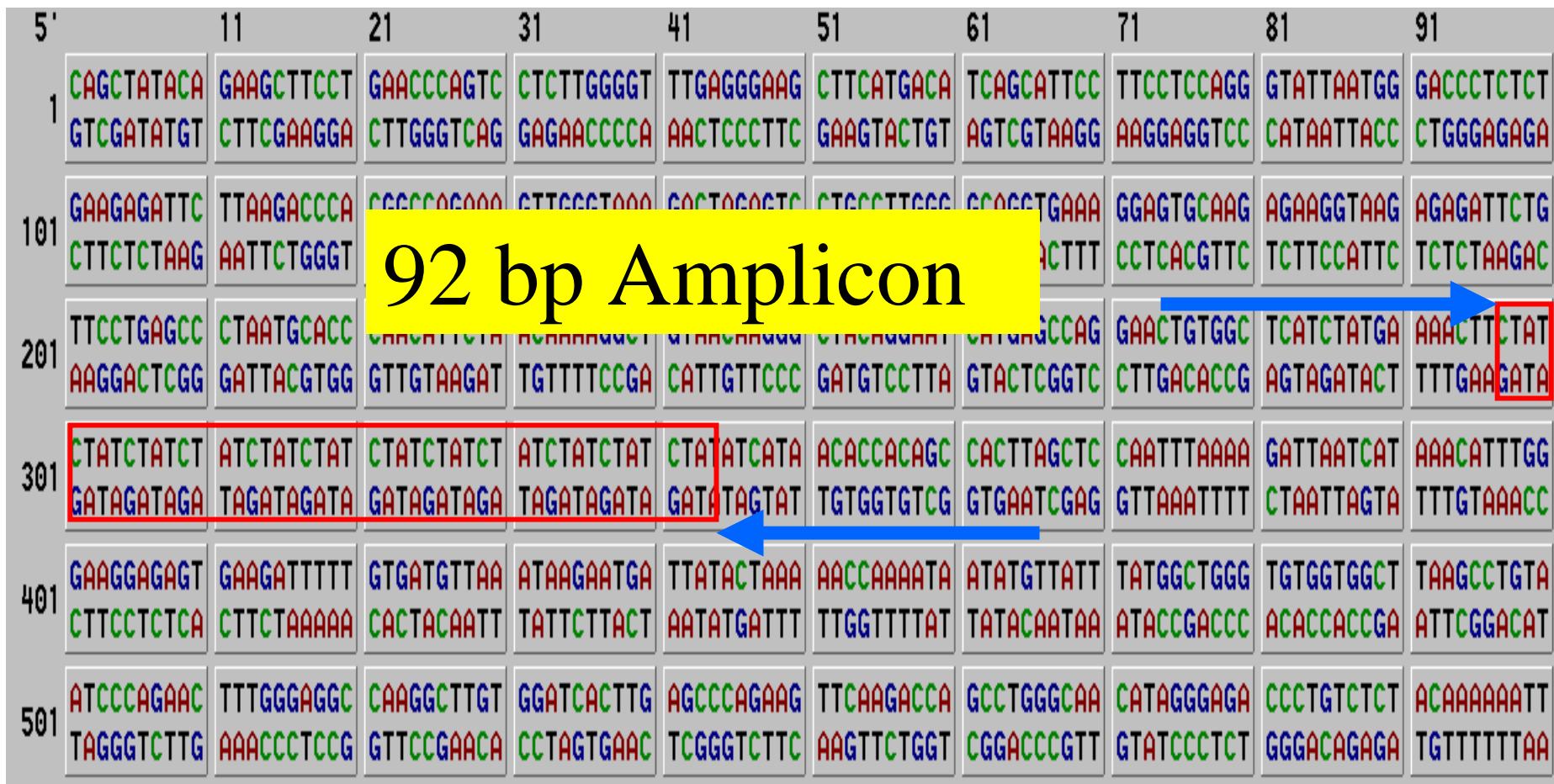
# PCR Primer Design

	5'	11	21	31	41	51	61	71	81	91
1	CAGCTATACA GTCGATATGT	GAAGCTTCCT CTTCGAAGGA	GAACCCAGTC CTTGGGTCAG	CTCTTGGGGT GAGAACCCCA	TTGAGGGAAG AACTCCCTTC	CTTCATGACA GAAGTACTGT	TCAGCATTCC AGTCGTAAAG	TTCCTCCAGG AAGGAGGTCC	GTATTAATGG CATAATTACC	GACCCCTCTCT CTGGGAGAGA
101	GAAGAGATTC CTTCTCTAAG	TTAAGACCCA AATTCTGGGT	CGGCCAGAAA GCCGGTCTTT	GTTGGGTAAA CAACCCATTT	GACTAGAGTC CTGATCTCAG	CTGCCTTGGG GACGGAACCC	GCAGGTGAAA CGTCCACTTT	GGAGTGCAAG CCTCACGTTC	AGAAGGTAAG TCTTCCATTC	AGAGATTCTG TCTCTAAGAC
201	TTCTGAGCC AAGGACTCGG	CTAATGCACC GATTACGTGG	CAACATTCTA GTTGTAAAGT	ACAAAAGGCT TGTTTTCCGA	GTAAACAAGGG CATTGTTCCC	CTACAGGAAT GATGTCCTTA	CATGAGCCAG GTACTCGGTC	GAACTGTGGC CTTGACACCG	TCATCTATGA AGTAGATACT	AAACTTCTAT TTTGAAGATA
301	CTATCTATCT GATAGATAGA	ATCTATCTAT TAGATAGATA	CTATCTATCT GATAGATAGA	ATCTATCTAT TAGATAGATA	CTATATCATA GATATAGTAT	ACACCACAGC TGTGGTGTCC	CACTTAGCTC GTGAATCGAG	CAATTTAAAA GTTAAATTTT	GATTAATCAT CTAATTAGTA	AAACATTTGG TTTGTAAACC
401	GAAGGAGAGT CTTCTCTCA	GAAGATTTTT CTTCTAAAAA	GTGATGTTAA CACTACAATT	ATAAGAATGA TATTCCTACT	TTATACTAAA AATATGATTT	AACCAAAATA TTGGTTTTAT	ATATGTTATT TATACAATAA	TATGGCTGGG ATACCGACCC	TGTGGTGGCT ACACCACCGA	TAAGCCTGTA ATTCGGACAT
501	ATCCAGAAC TAGGGTCTTG	TTTGGGAGGC AAACCCCTCCG	CAAGGCTTGT GTTCCGAACA	GGATCACTTG CCTAGTGAAC	AGCCCAGAA TCGGGTCTTC	TTCAAGACCA AAGTTCGGT	GCCTGGGCAA CGGACCCGTT	CATAGGGAGA GTATCCCTCT	CCCTGTCTCT GGGACAGAGA	ACAAAAAATT TGTTTTTTAA

D2S441

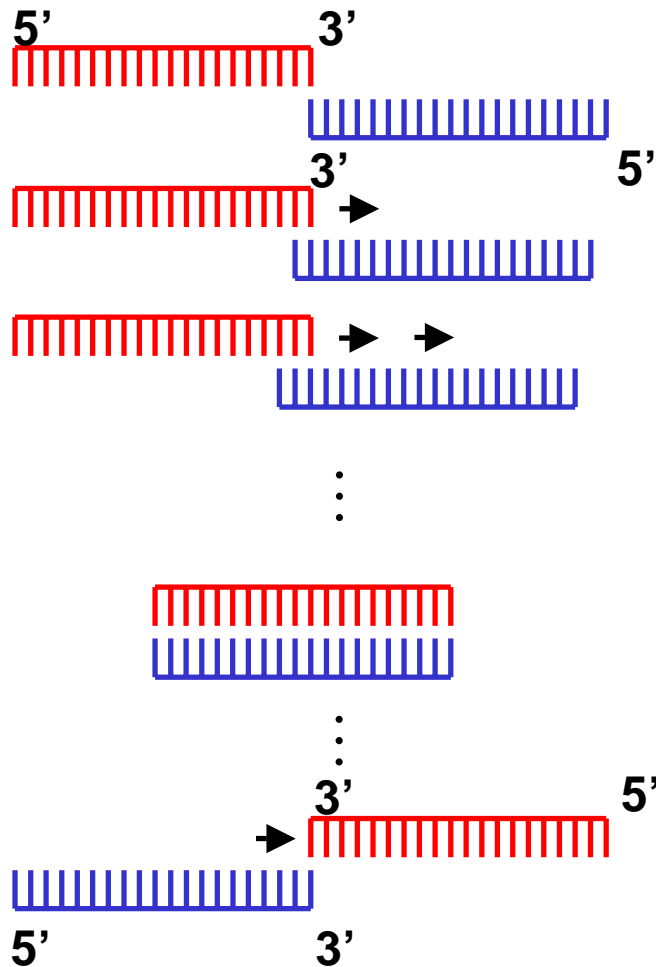


# PCR Primer Design



D2S441

# AutoDimer – primer screening software is now freely available



Auto Dimer Check

File About

Primer Dimer Screen

Hairpin Screen

Minimum SCORE Requirement

6

SAVE DATA

Cancel

# of Sequences

# of Hits

2

2

3

Total Number of Primer-Primer Comparisons

Na+ (Molar)

0.085

Temp for dG calc

37

Total Strand Conc (micromolar)

1.0

C:\Documents and Settings\petev\My Documents\Programming\AutoDimer paper for Biotechniques\Files for program testing\n=1.txt

```
testA ACGTATGCGGTATGCGAGCTA versus testA ACGTATGCGGTATGCGAGCTA
Matches = 8
Score = 6
ACGNATNCGT
est. tm = less than 0 oC
DeltaG 37 degrees = -2.36 kcal/mole

                    5'-ACGTATGCGGTATGCGAGCTA-3'
                      |||x||x|||
                    3'-ATCGAGCGGTATGCGTATGCA-5'

testB ACGCGATCGATGATCGACTG versus testB ACGCGATCGATGATCGACTG
Matches = 12
Score = 7
CNNTCGATNATCGANNG
est. tm = 7.5 oC
DeltaG 37 degrees = -2.42 kcal/mole

                    3'-GTCAGCTAGTAGCTAGCGCA-5'
                      |xx||||x||||xx|
                    5'-ACGCGATCGATGATCGACTG-3'
```

<http://www.cstl.nist.gov/biotech/strbase/AutoDimerHomepage/AutoDimerProgramHomepage.htm> BioTechniques (2004) 37: 226-231

# Characterization of New miniSTR Loci

- Candidate STR marker selection
- Chromosomal locations and marker characteristics
- PCR primer design
- **Initial testing results**
- Population testing
- Allelic ladder construction
- Miniplex assay performance

# Initial Testing Results

107 potential markers



61 markers with “clean” flanking regions



43 markers with amplicon size < 125bp



18 markers for initial testing



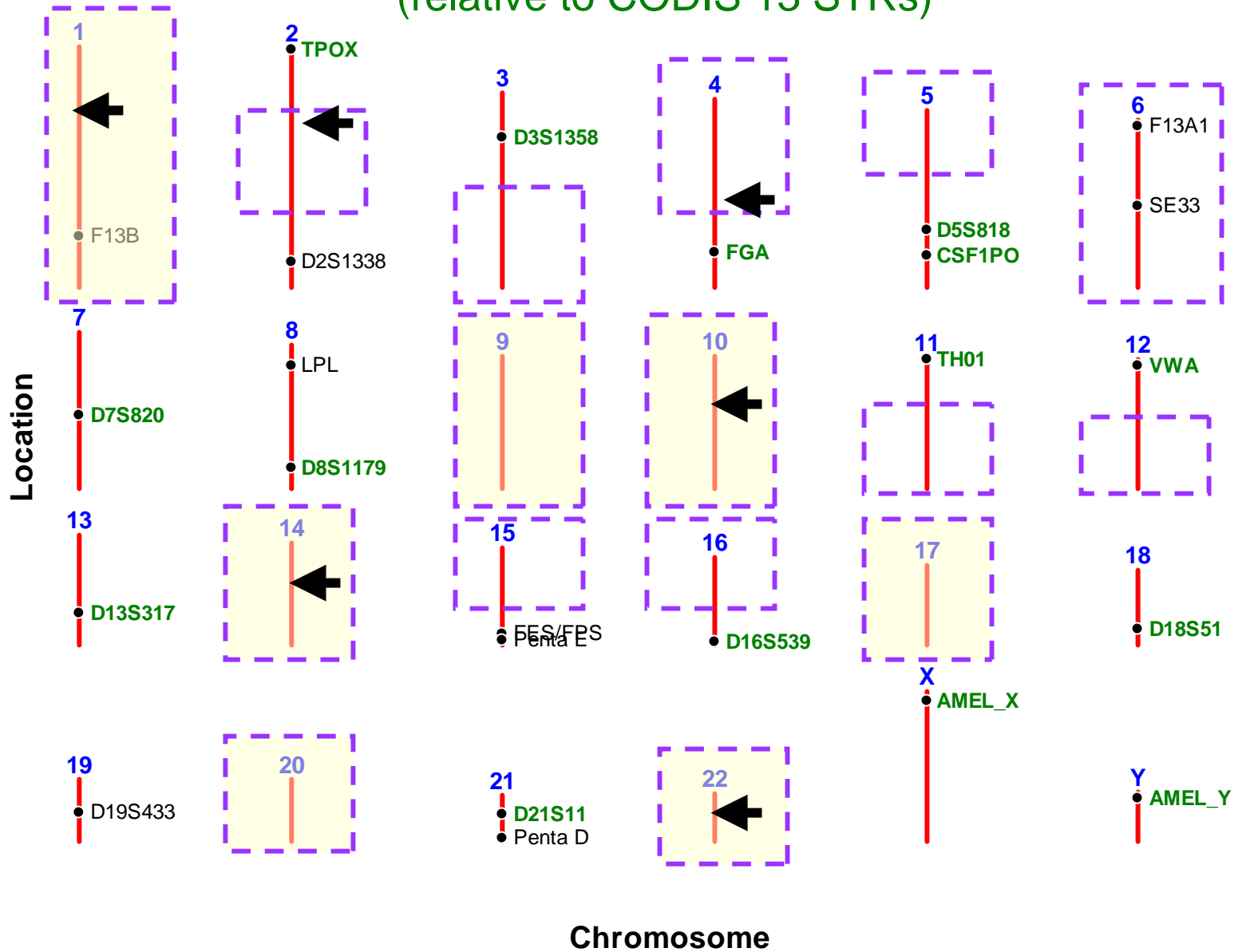
2 three loci miniplexes

# Initial Testing Results

Miniplex01- mD10S1248 - FAM  
mD14S1434 - VIC  
mD22S1045 - NED

Miniplex02- mD4S2364 - FAM  
mD2S441 - VIC  
mD1S1677 - NED

# Locations of Focus for New miniSTR Loci (relative to CODIS 13 STRs)



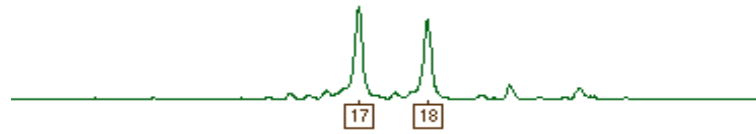


D10S1248



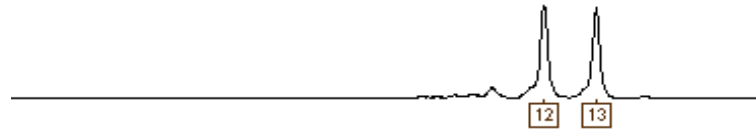
6-FAM

D14S1434



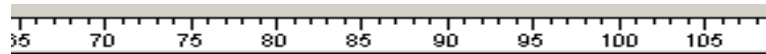
VIC

D22S1045



NED

Miniplex01

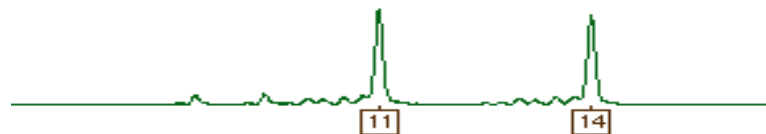


D4S2364



6-FAM

D2S441



VIC

D1S1677



NED

Miniplex02

# miniSTR characteristics

<b>STR Locus</b>	<b>Sequence Motif</b>	<b>Allele Range</b>	<b>Size Range (bp)</b>	<b>Observed Heterozygosity</b>
<b>D1S1677</b>	(GGAA) <sub>n</sub>	9-18	81-117	0.75
<b>D2S441</b>	(TCTA) <sub>n</sub>	9-17	78-110	0.76
<b>D4S2364</b>	(GAAT)(GGAT)(GAAT) <sub>n</sub>	8-12	67-83	0.53
<b>D10S1248</b>	(GGAA) <sub>n</sub>	10-20	83-123	0.78
<b>D14S1434</b>	(GATA) <sub>n</sub> (GACA) <sub>n</sub>	13-20	70-98	0.68
<b>D22S1045</b>	(TAA) <sub>n</sub>	5-16	76-109	0.77



# Characterization of New miniSTR Loci

- Candidate STR marker selection
- Chromosomal locations and marker characteristics
- PCR primer design
- Initial testing results
- **Population testing**
- Allelic ladder construction
- Miniplex assay performance

# NIST U.S. Population Samples

As of 06/2003 **666 males** (anonymous; self-identified ethnicities)

286 Caucasians  
252 African Americans  
128 Hispanics

Whole blood received from  
Interstate Blood Bank (Memphis, TN)

Working tubes/plates 1 ng/uL

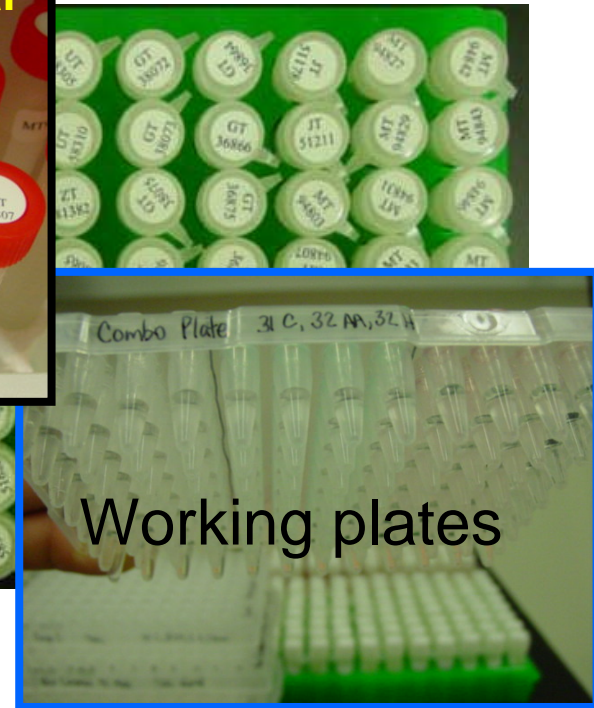
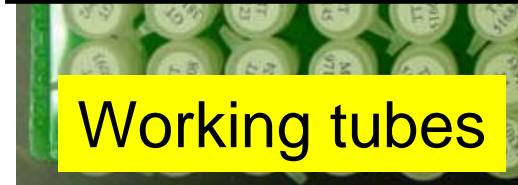
To date: (~50,000 allele calls)

Identifiler (15 autosomal markers + Amelogenin) (10,608)

Roche Linear Arrays (HV1/HV2 10 regions) (6,630)

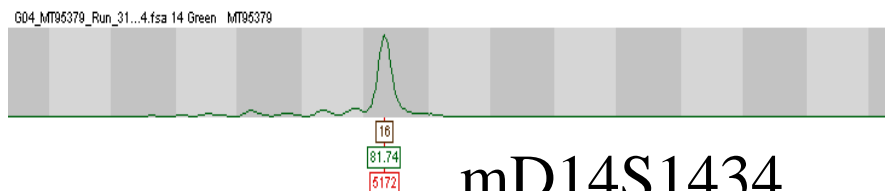
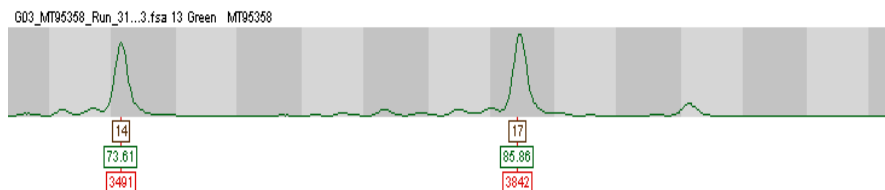
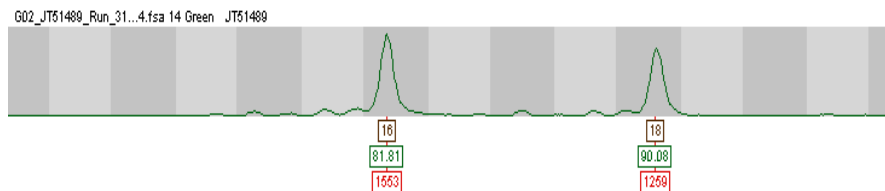
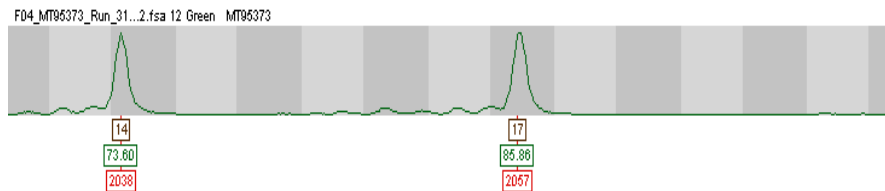
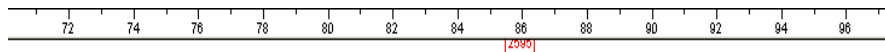
Y STRs 22 loci—27 amplicons (17,388)

Y SNPs 50 markers on sub-set of samples (11,498)

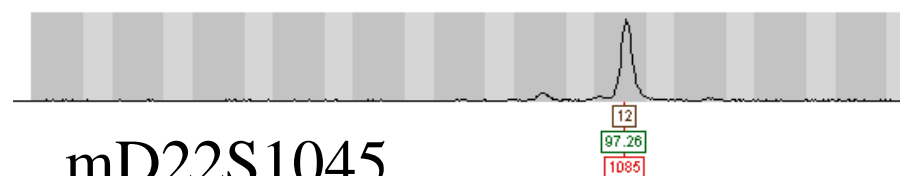
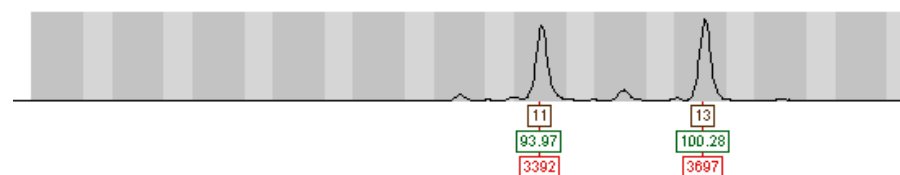
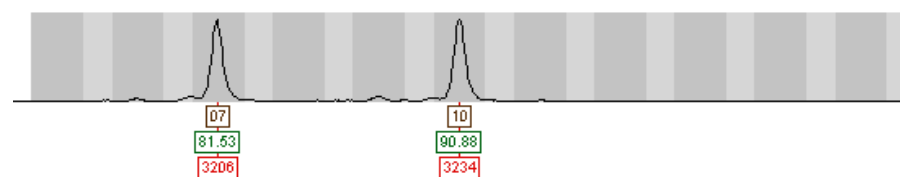
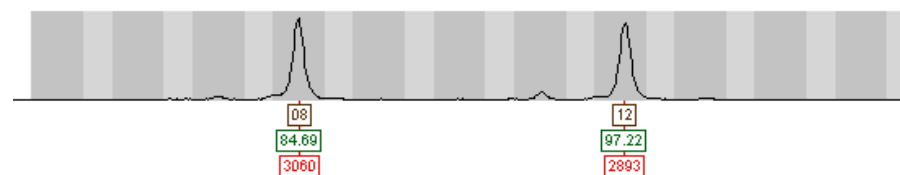
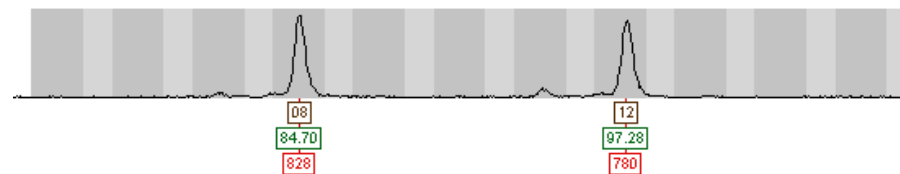
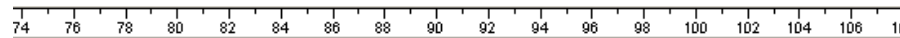


Samples supplied to **OhioU** for miniSTR typing and **AFDIL** for whole mtGenome sequencing

# Population Sample Typing with Genotyper Macos



mD14S1434

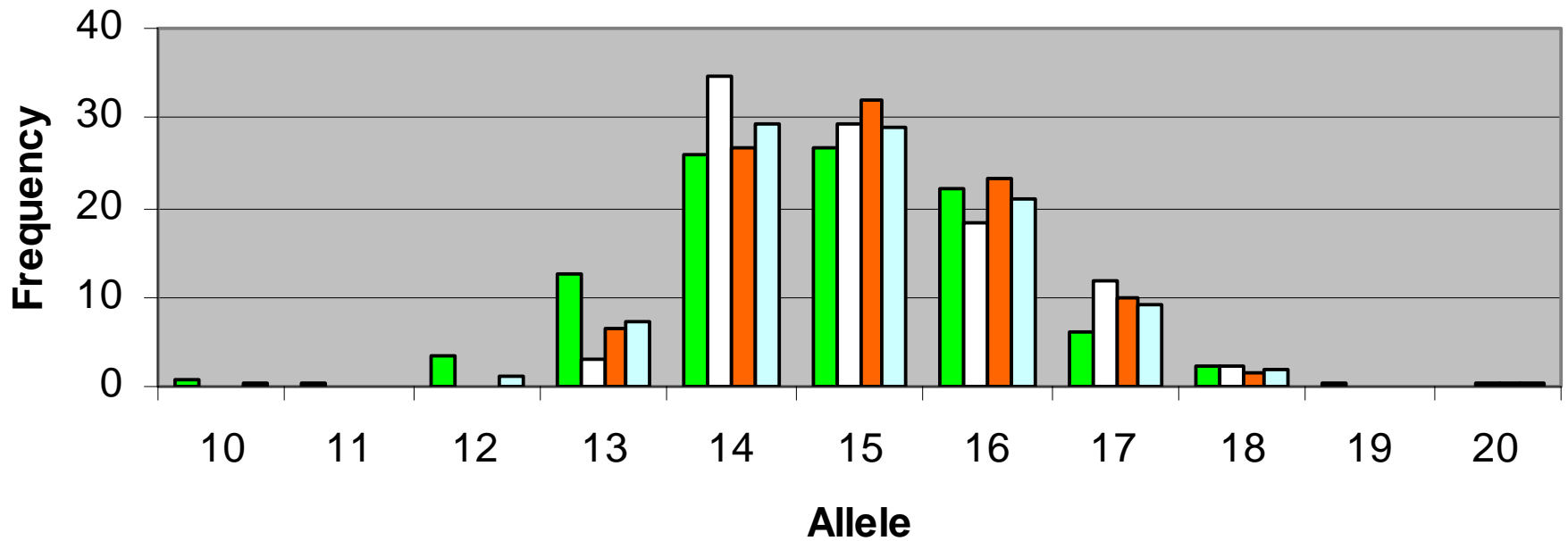


mD22S1045

# Population Data Analysis

**D10S1248**

11 alleles observed

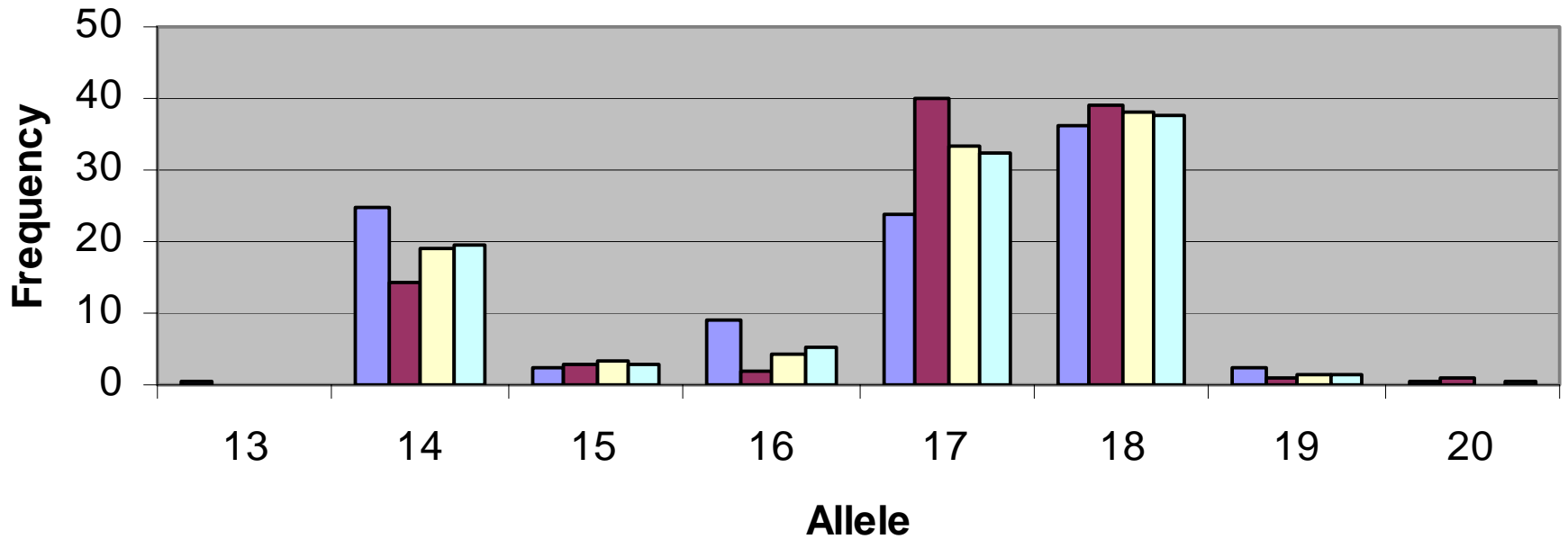


■ Afr. Am (N=166) □ Cau (N=164) ■ Hisp (N=135) □ Total (N=465)

# Population Data Analysis

**D14S1434**

**8 alleles observed**

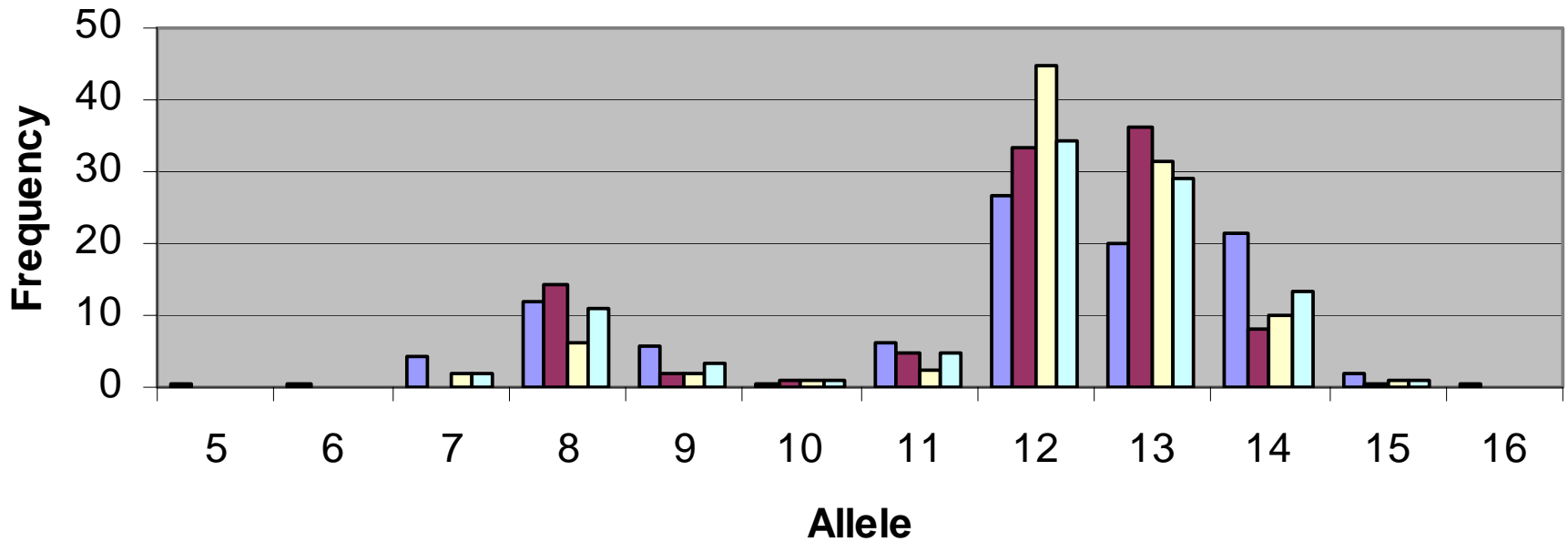


■ Afr. Am. (N=168) ■ Cau. (N=163) ■ Hisp. (N=133) ■ Total (N=464)

# Population Data Analysis

**D22S1045**

**12 alleles observed**



■ Afr. Am. (N=167) ■ Cau. (N=164) ■ Hisp. (N=136) ■ Total (N=467)

# Population Testing –Miniplexes vs. Identifiler

Heterozygosity    Marker

0.8784	D2S1338
0.8753	D18S51
0.8710	FGA
0.8393	D21S11
0.8245	vWA
0.8076	D7S820
0.7970	D19S433
<b>0.7759</b>	<b>mD10S1248 - mini01</b>
0.7759	D16S539
<b>0.7674</b>	<b>mD22S1045 - mini01</b>
0.7674	D8S1179
<b>0.7590</b>	<b>mD2S441 - mini02</b>
0.7548	D3S1358
0.7526	D13S317
<b>0.7463</b>	<b>mD1S1677 - mini02</b>
0.7378	CSF1PO
0.7378	TH01
0.7294	D5S818
0.7146	TPOX
<b>0.6765</b>	<b>mD14S1434 - mini01</b>
<b>0.5307</b>	<b>mD4S2364 - mini02</b>

N = 474 Individuals  
164 African Americans  
170 Caucasians  
140 Hispanics

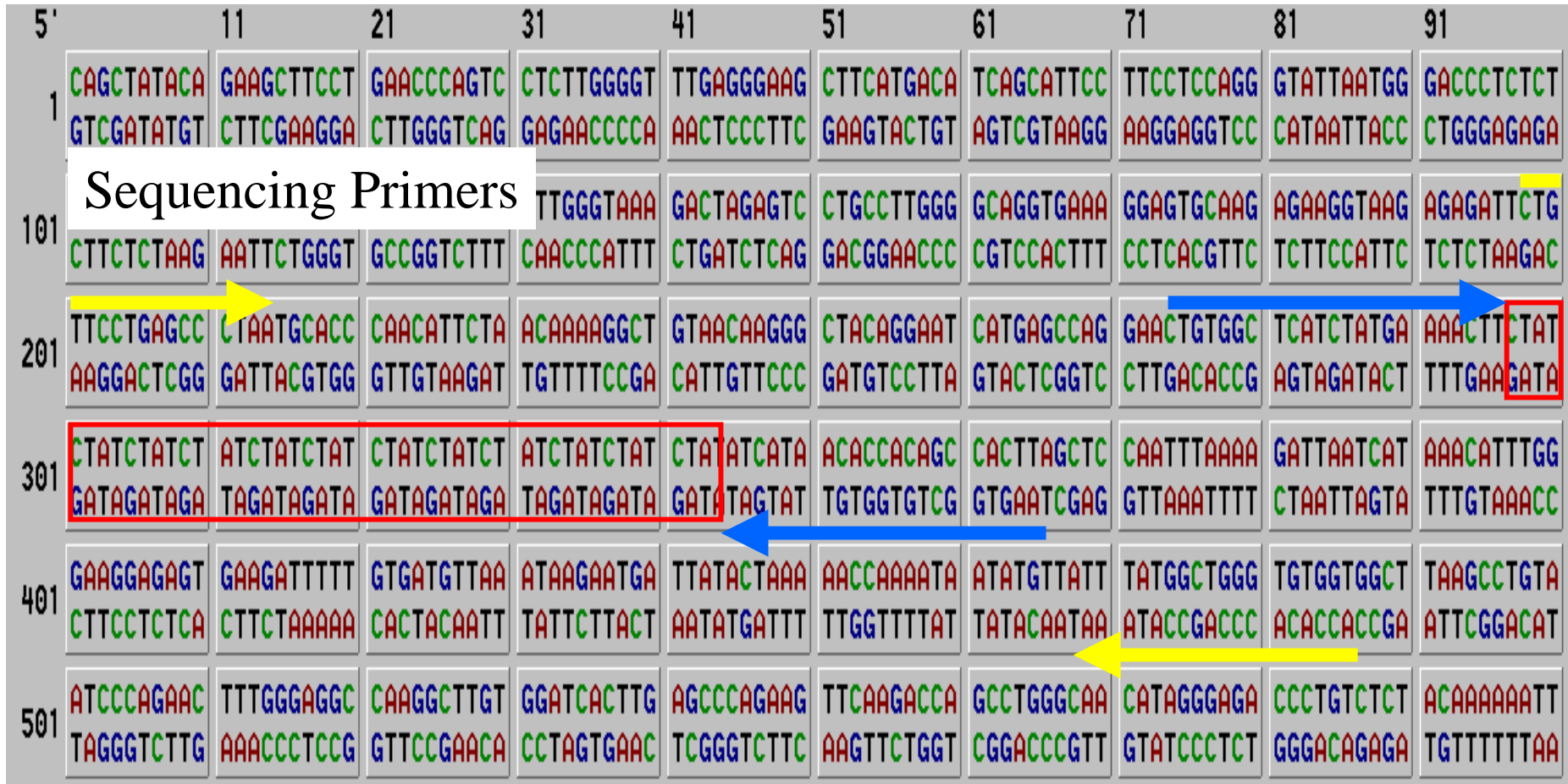
# Characterization of New miniSTR Loci

- Candidate STR marker selection
- Chromosomal locations and marker characteristics
- PCR primer design
- Initial testing results
- Population testing
- Allelic ladder construction
- Miniplex assay performance



# Characterization of New miniSTR Loci

D2S441



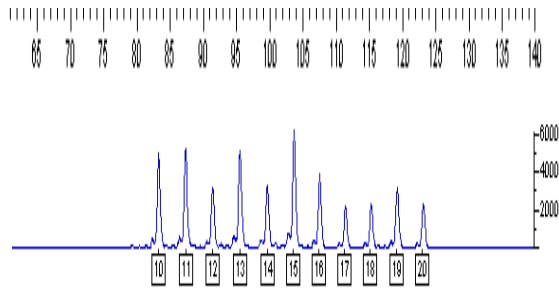


# miniSTR Assay

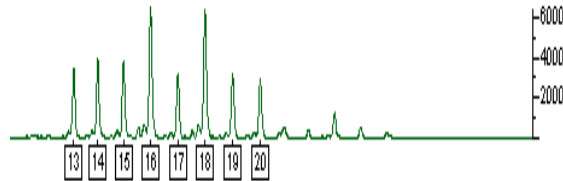
## Allelic Ladders

### Miniplex01

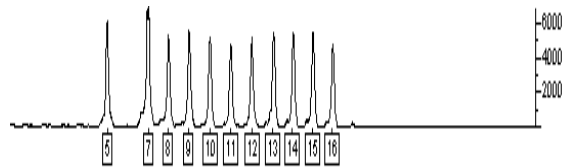
**D10S1248**



**D14S1434**

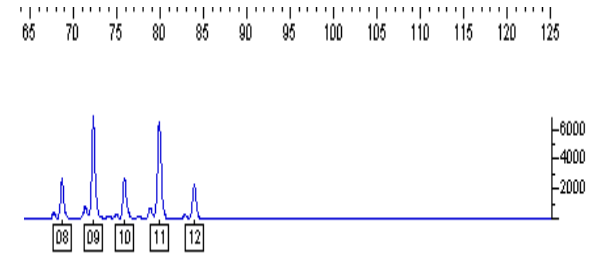


**D22S1045**

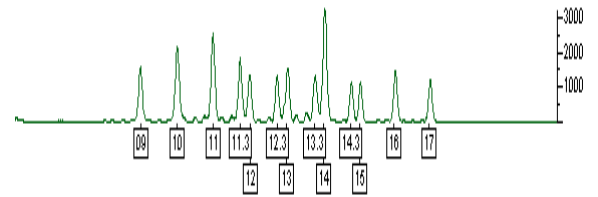


### Miniplex02

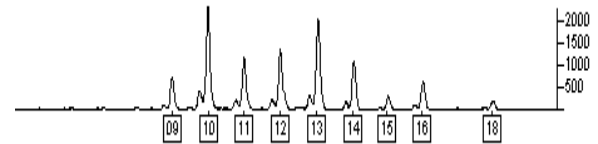
**D4S2364**



**D2S441**



**D1S1677**



D10S1248

09	Highest peak at	79.00 ±	0.50 bp in blue
10	Highest peak at	83.00 ±	0.50 bp in blue
11	Highest peak at	87.00 ±	0.50 bp in blue
12	Highest peak at	91.00 ±	0.50 bp in blue
13	Highest peak at	95.00 ±	0.50 bp in blue
14	Highest peak at	99.00 ±	0.50 bp in blue
15	Highest peak at	103.00 ±	0.50 bp in blue
16	Highest peak at	107.00 ±	0.50 bp in blue
17	Highest peak at	111.00 ±	0.50 bp in blue
18	Highest peak at	115.00 ±	0.50 bp in blue
19	Highest peak at	119.00 ±	0.50 bp in blue
20	Highest peak at	123.00 ±	0.50 bp in blue

D14S1434

13	Highest peak at	70.40 ±	1.00 bp in green
14	Highest peak at	74.30 ±	1.00 bp in green
15	Highest peak at	78.20 ±	1.00 bp in green
16	Highest peak at	82.10 ±	1.00 bp in green
17	Highest peak at	86.00 ±	1.00 bp in green
18	Highest peak at	89.90 ±	1.00 bp in green
19	Highest peak at	93.80 ±	1.00 bp in green
20	Highest peak at	97.70 ±	1.00 bp in green

D22S1045

05	Highest peak at	75.40 ±	1.00 bp in yellow
06	Highest peak at	78.50 ±	1.00 bp in yellow
07	Highest peak at	81.60 ±	1.00 bp in yellow
08	Highest peak at	84.70 ±	1.00 bp in yellow
09	Highest peak at	87.80 ±	1.00 bp in yellow
10	Highest peak at	90.90 ±	1.00 bp in yellow
11	Highest peak at	94.00 ±	1.00 bp in yellow
12	Highest peak at	97.10 ±	1.00 bp in yellow
13	Highest peak at	100.20 ±	1.00 bp in yellow
14	Highest peak at	103.30 ±	1.00 bp in yellow
15	Highest peak at	106.40 ±	1.00 bp in yellow
16	Highest peak at	109.50 ±	1.00 bp in yellow

Macros for analysis  
have been developed

# Characterization of New miniSTR Loci

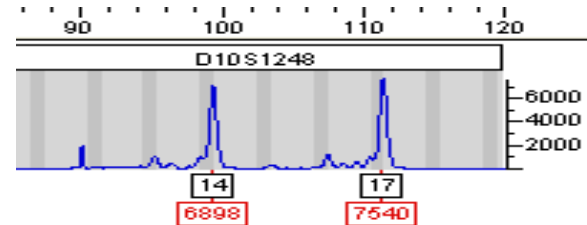
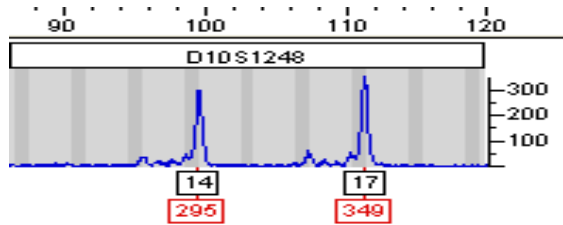
- Candidate STR marker selection
- Chromosomal locations and marker characteristics
- PCR primer design
- Initial testing results
- Population testing
- Allelic ladder construction
- **Miniplex assay performance**
  - Sensitivity
  - Inheritance with family samples
  - Allele sizing precision
  - Locus stutter percentage characterization
  - Analysis on real-world samples

# miniSTR Assay Sensitivity (D10S1248)

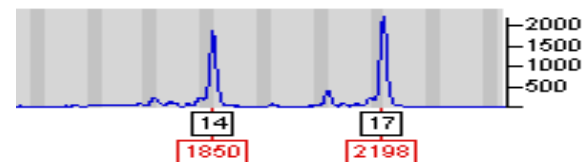
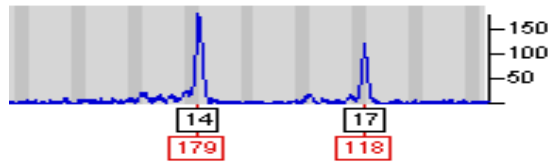
28 cycles – 1U Taq

32 cycles – 2U Taq

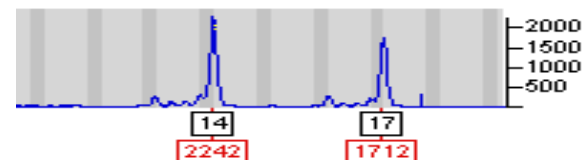
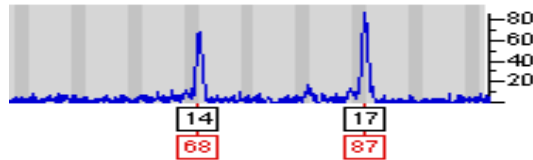
200 pg



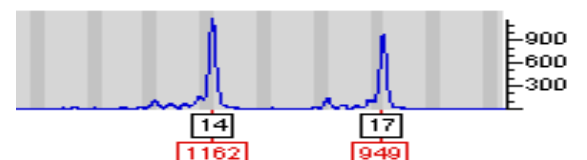
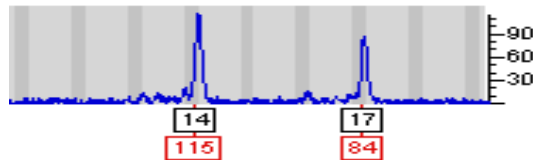
100 pg



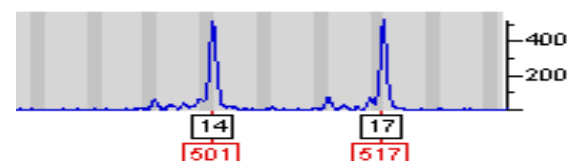
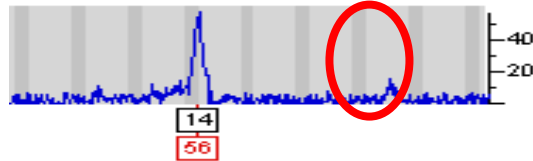
50 pg



20 pg



10 pg

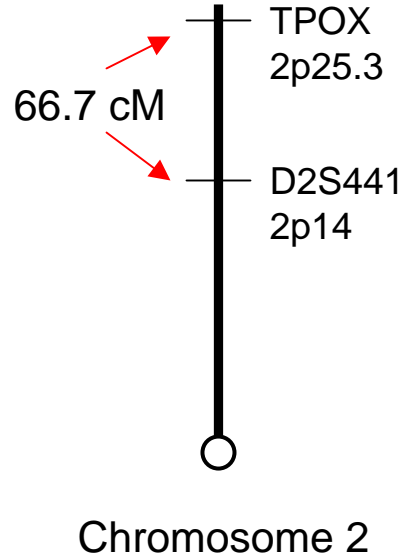
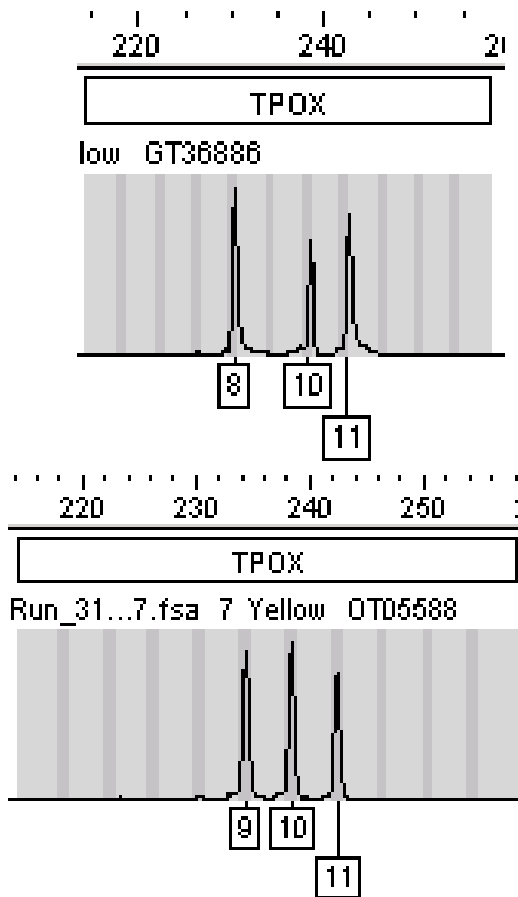


5 pg

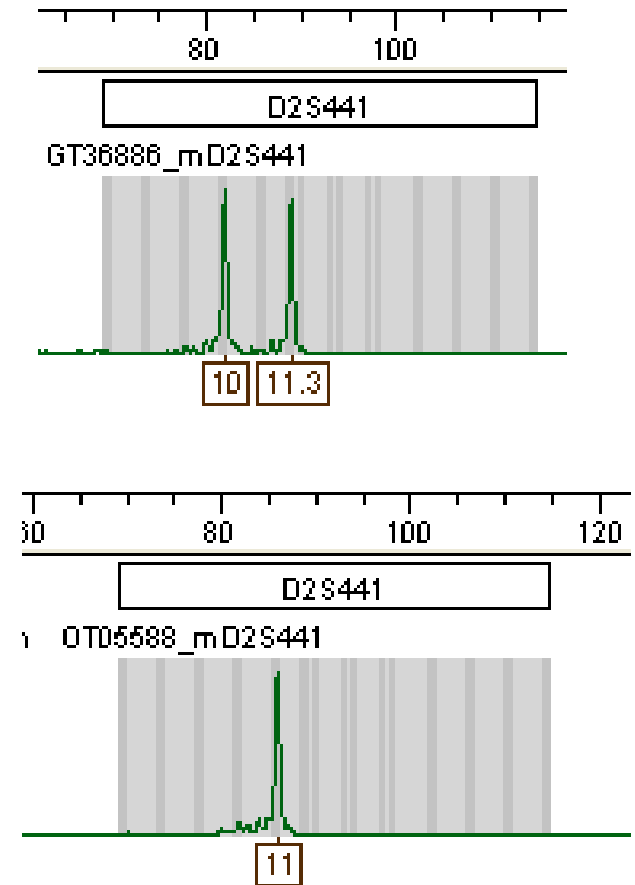


# Concordance on Tri-Allelic Patterns for TPOX

## Identifiler Data



## mD2S441 Data



miniSTR primers for TPOX also gave tri-allelic patterns

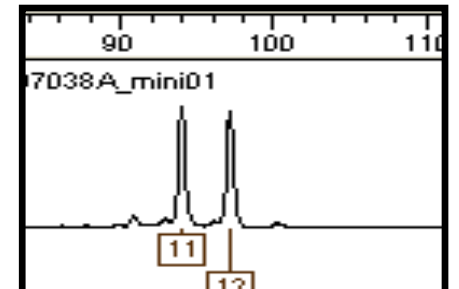
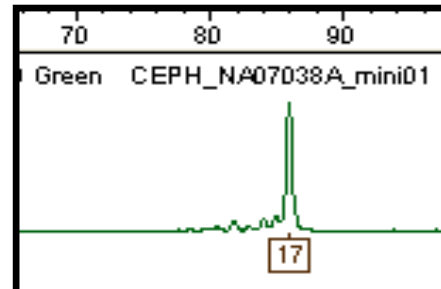
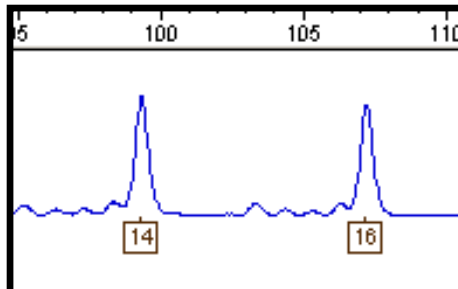
# CEPH Family miniSTR Results

**mD10S1248**

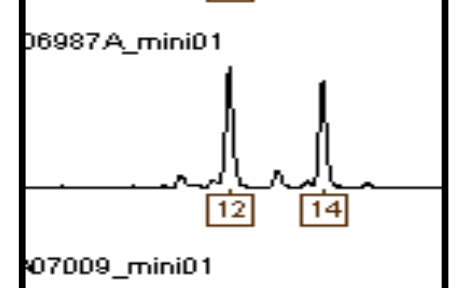
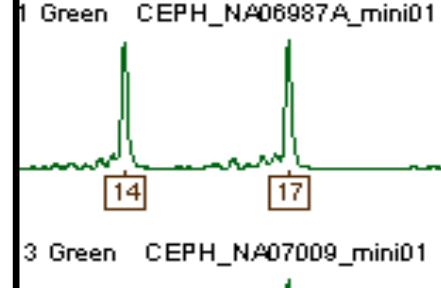
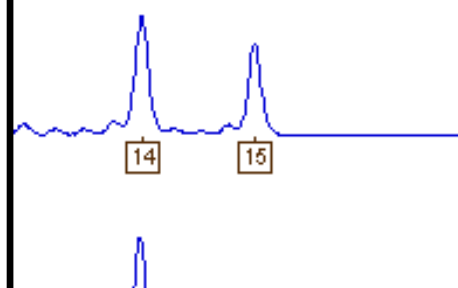
**mD14S2364**

**mD22S1045**

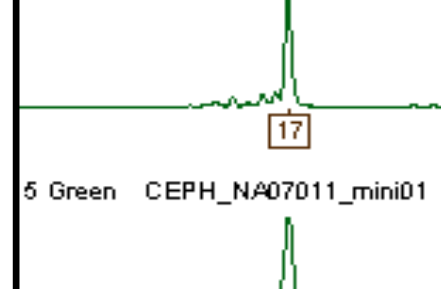
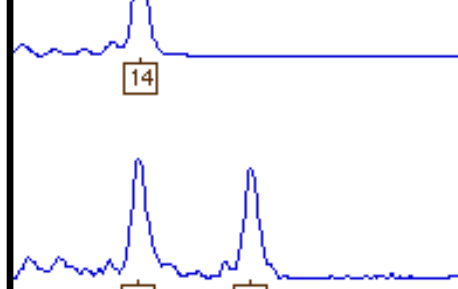
**Father**



**Mother**



**Daughter**

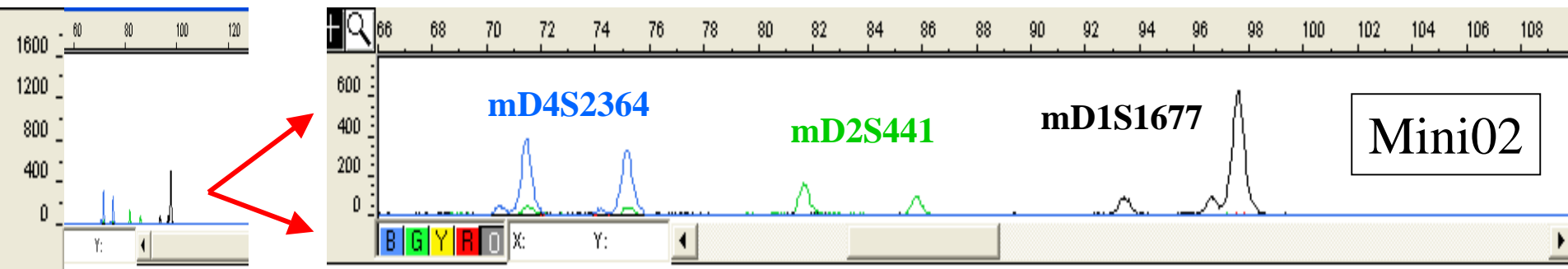
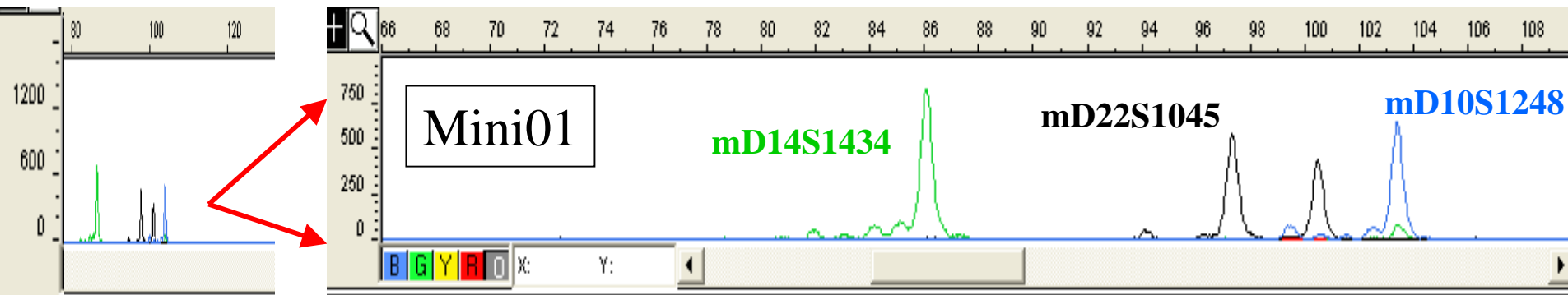
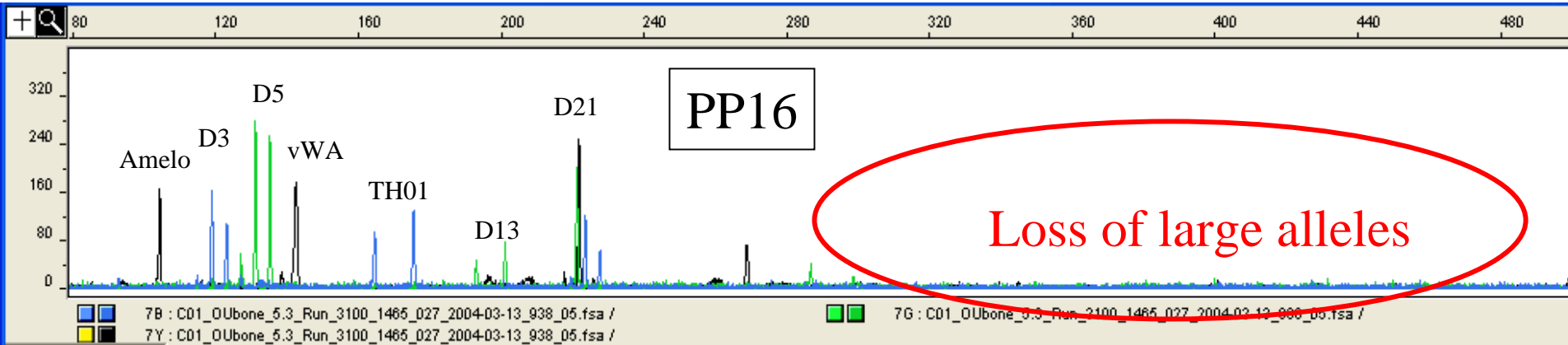


**Son**





# Sensitivity - Degraded DNA from an OU Bone Sample



*Michael D. Coble,<sup>1</sup> Ph.D. and John M. Butler,<sup>1</sup> Ph.D.*

## Characterization of New MiniSTR Loci to Aid Analysis of Degraded DNA\*

---

For more information... January issue of JFS

[http://www.cstl.nist.gov/biotech/strbase/pub\\_pres/Coble2005miniSTR.pdf](http://www.cstl.nist.gov/biotech/strbase/pub_pres/Coble2005miniSTR.pdf)

# “STR typing of human telogen hairs -- a new approach”

- Hellmann, *et al.* (2001) *Int. J. Legal Med.* 114(4-5): 269-273
- Primer pairs with **annealing positions close to the repeat units** of the STR loci FES/FPS, TPOX, and TH01 were used for amplification.
- Complete digestion of hair with increased  $\text{CaCl}_2$  buffer
- **Shed telogen hairs could be typed!**

Complete Digestion

		D10S	D14S	D22S
PV 2.5cm	10ul	-	-	08, 13
	5ul	-	-	08, 08
	2.5ul	-	-	-
PV 2.5cm	10ul	-	-	-
	5ul	-	-	-
	2.5ul	-	-	-
PV 2.7cm	10ul	-	-	-
	5ul	-	-	-
	2.5ul	-	-	-
PV 2.7cm	10ul	-	-	08, 13 (<50)
	5ul	-	-	13, 13
	2.5ul	-	-	-
PV 5 hairs	10ul	16, 16	17, 18	08, 13
	5ul	16, 16	17, 18	08, 08
	2.5ul	15, 15	-	08, 13
PV 5 hairs	10ul	-	-	08, 13
	5ul	-	-	08, 13
	2.5ul	-	-	13, 13

# miniSTR Typing of Hairs

## Complete Digestion Protocol

32 cycles; 2U Taq

KEC 20cm	10ul	16, 17	17	13
	5ul	16, 17	17	13
	2.5ul	16, 17	17	13
MK 18cm	10ul	14, 15	17, 18	08, 12
	5ul	14, 15	17, 18	08, 12
	2.5ul	14, 15	17, 18	08, 12

“Longer” Hairs – greater success

<u>Genotypes</u>	<u>D10S</u>	<u>D14S</u>	<u>D22S</u>
PV	15, 16	17, 18	08, 13
KEC	16, 17	17, 17	13, 13
MK	14, 15	17, 18	08, 12

		MicroTissue Grinding		
		D10S	D14S	D22S
PV 2.5cm	10ul	-	-	-
	5ul	-	-	-
	2.5ul	-	-	-
PV 2.5cm	10ul	-	-	-
	5ul	-	-	-
	2.5ul	-	-	-
PV 2.7cm	10ul	-	-	-
	5ul	-	-	-
	2.5ul	-	-	-
PV 2.7cm	10ul	-	-	-
	5ul	-	-	-
	2.5ul	-	-	-
PV 5 hairs	10ul	-	-	-
	5ul	-	-	-
	2.5ul	-	-	08, 08
PV 5 hairs	10ul	-	-	-
	5ul	-	-	-
	2.5ul	-	-	-
KEC 20cm	10ul	-	-	-
	5ul	-	-	-
	2.5ul	-	-	13, 13
MK 18cm	10ul	-	-	-
	5ul	-	-	-
	2.5ul	14, 15 (<50)	17, 18	08, 12

## miniSTR Typing of Hairs

### MicroTissue Grinding Protocol

32 cycles; 2U Taq

	32 cycles 5 ul	32 cycles 2.5 ul	36 cycles 2.5 ul (5U Taq)
Dark Hair (Phenol)			
JB01 (1.5cm)	18, 18	18, 18	-
JB02 (1.3cm)	-	-	13, 14
JB03 (1.5cm)	-	-	08, 08
JB04 (1.5cm)	17, 17	-	-
JB07 (1.3cm)	-	-	-
JB09 (1.0cm)	-	16, 16	-
			22, 22; 13, 13
JB10 (3.3cm)	-	16, 16; 17, 17; 13, 13	
Dark Hair (Qiagen)			
JB01 (1.1cm)	14, 14	-	-
JB02 (1.2cm)	14, 16	-	13, 13
JB03 (1.8cm)	-	13, 13	21, 21
JB04 (1.4cm)	-	-	-
JB05 (1.2cm)	-	-	16, 16; 13, 13
JB06 (1.5cm)	-	-	-
Gray Hair (Phenol)			
JB01 (1.8cm)	14, 16	14, 15; 17, 17	14, 16; 18, 18; 13, 13
JB02 (2.0cm)	16, 16	16, 16; 18, 18, 13, 13	16, 16; 13, 13
JB03 (1.1cm)	16, 16	17, 17	13, 13
JB04 (1.4cm)	-	-	16, 16; 18, 18
JB05 (1.7cm)	15, 15	16, 16	18, 18; 13, 13
JB06 (1.0cm)	-	18, 18	-
JB11 (1.0cm)	13, 13	-	-
			16, 16; 18, 18
Gray Hair (Qiagen)			
JB14 (1.5cm)	18, 18 13, 13	16, 16	
JB15 (1.7cm)	-	-	13, 13
JB16 (1.1cm)	-	13, 13	13, 13
JB17 (0.8cm)	18, 18	-	13, 13
JB19 (1.4cm)	16, 16	-	16, 16
			14, 14; 18, 18
JB20 (2.0cm)	-	-	

## miniSTR Typing of Hairs

26 hairs (0.8 cm – 3.3 cm)

← 1.8 cm hair

# miniSTR Typing of Hairs

	"Correct"	"Partial"	"Incorrect"	"Did Not Type"
Dark Phenol (7) %	4/63 0.06	2/63 0.03	5/63 0.08	52/63 0.83
Dark Qiagen (6) %	4/54 0.07	1/54 0.02	2/54 0.04	47/54 0.87
Gray Phenol (6) %	12/54 0.22	6/54 0.11	4/54 0.07	32/54 0.59
Gray Qiagen (7) %	9/63 0.14	4/63 0.06	2/63 0.03	48/63 0.76

Greater success with gray hairs

# Future Plans

Testing and characterization of more markers.

Population Databasing.

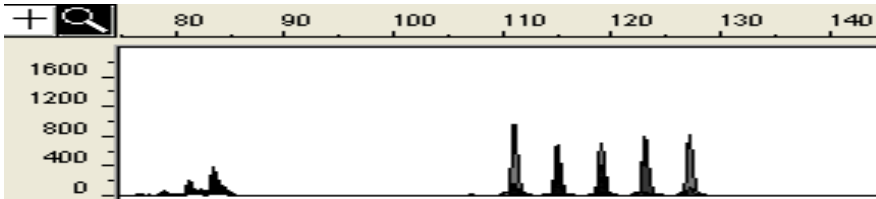
Testing on degraded materials.

Information will be posted on STRBase website and published as these loci are characterized



# Potential New Miniplex

## mD6S474



Tetranucleotide repeat – GATA

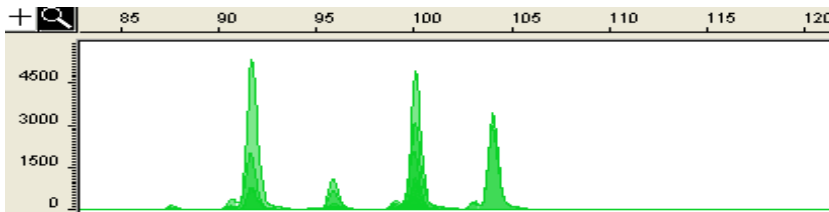
Exp. Het. – 0.77

Distance to Repeat

F – 0

R – 7

## mD3S3053



Tetranucleotide repeat – GATA

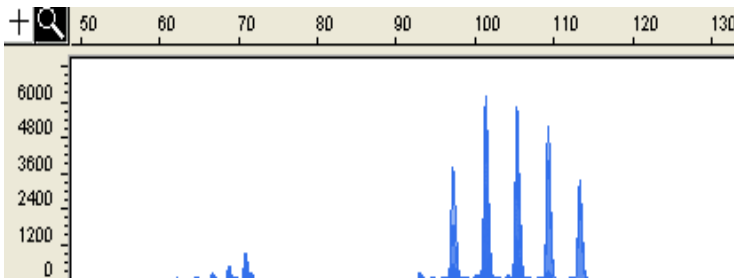
Exp. Het. – 0.74

Distance to Repeat

F – 14

R – 0

## mD20S482



Tetranucleotide repeat – GATA

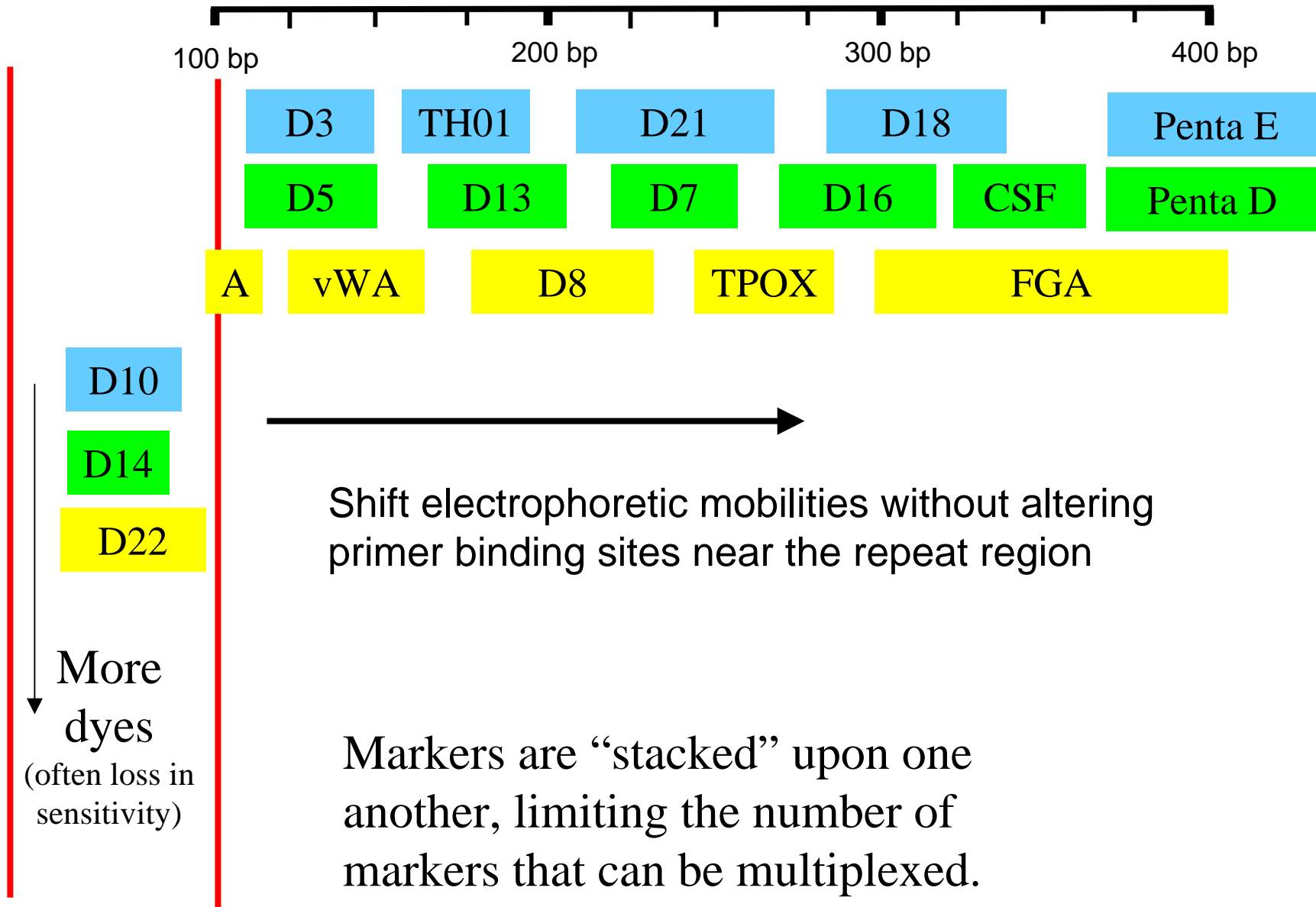
Exp. Het. – 0.85

Distance to Repeat

F – 3

R – 6

# Expanding Multiplex-ability



# Funding and Collaborations

We are funded by an Interagency Agreement between **National Institute of Justice** and NIST Office of Law Enforcement Standards

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

## Human Identity Project Team

John Butler (Project Leader)

Peter Vallone

Michael Coble

Margaret Kline

Jan Redman

Amy Decker

David Duewer

## Collaborators (also funded by NIJ):

**Bruce McCord** and students (Ohio U)  
for miniSTR work

<http://www.cstl.nist.gov/biotech/strbase/Coble.htm>  
[michael.coble@nist.gov](mailto:michael.coble@nist.gov)