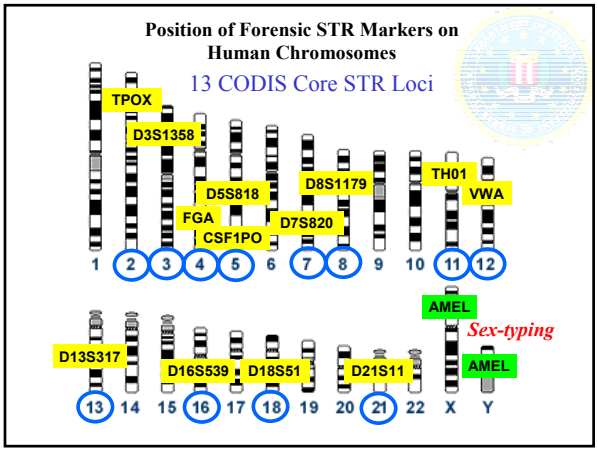
Presentation at 53rd American Academy of Forensic Sciences
February 22, 2001

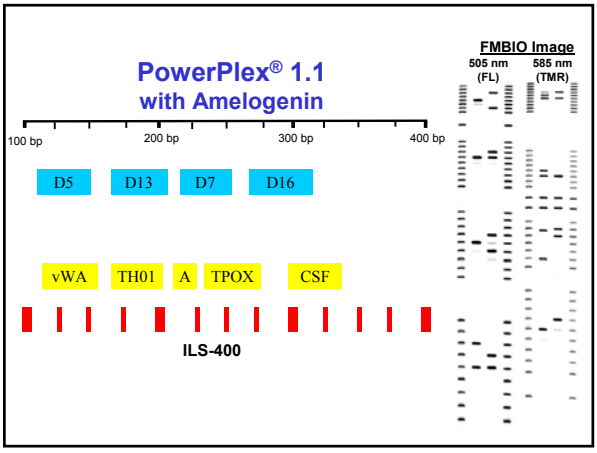
Comparison of Primer Sequences Used in Commercial STR Kits

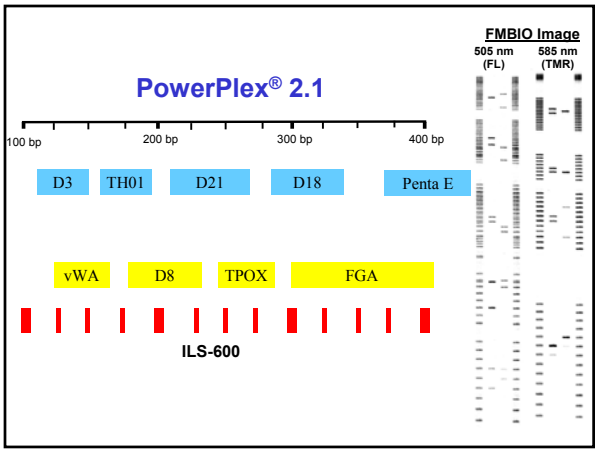
John M. Butler and Peter M. Vallone
National Institute of Standards and Technology
Joseph M. Devaney and Michael A. Marino
Transgenomic Inc.

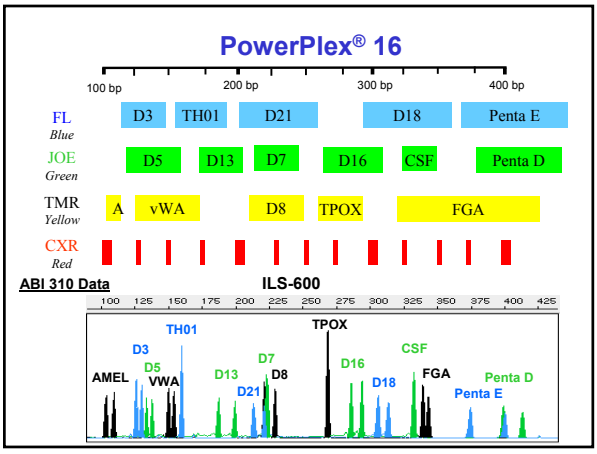
Presentation Overview

- Promega STR kits and primer sequence release
- Primer differences between kits – areas for potential null alleles
- Basics of primer design
- Analysis of parameters that impact multiplex primer design









news

July 27, 2000 Nature vol. 406, p. 336

Legal protests prompt DNA primer release

Paul Smaglik, Washington

A leading supplier of DNA testing kits widely used in legal actions has promised to make public its closely held proprietary information on the sequence of the primer used in the kits.

In a number of recent cases, defence attorneys have successfully challenged the validity of DNA testing on the grounds of lack of access to such information. The companies that make the tests have withheld the information through commercial confidentiality. However, the market for such tests could diminish if they are rendered legally invalid.

On 18 April, for example, a state court in Vermont ruled that DNA results were inad-

Mozer: revealing sequences will remove doubt.

amplified, invalidating the test. Without knowing the sequence of the primers, however, courts cannot determine whether the primer sequence is present in the samples.

Promega researcher Tom Mozer says the chances of contamination are extremely remote. But defence attorneys have used the lack of disclosure to introduce doubt into the proceedings. Releasing the primer sequences will help dispel that doubt, Mozer says.

Arthur Eisenberg, director of the DNA Identity Laboratory at the University of North Texas Health Science Center in Fort Worth, agrees that contamination from primers is rare. "It's an argument that the defence has used to try to keep this testimony

Masibay, A., Mozer, T. J., and Sprecher, C. (2000) Promega Corporation reveals primer sequences in its testing kits [letter]. *J. Forensic Sci.* 45(6): 1360-1362

scientists analyse DNA from points in the genome at which the letters in the genetic code repeat, focusing on 13 sites where the number of repeats varies highly between individuals. The primers are designed to

regions matches, it is highly likely that the two samples came from the same individual.

Some defence attorneys have recently argued successfully that, in principle, the primers could contaminate the DNA being

able intellectual property, Eisenberg doubts that many scientists will bother to copy them all. Although it is feasible to create one or two primers by hand, a complete pre-prepared set is far more useful.

Eudora by QUALCOMM - [Cindy Sprecher, 03:53 PM 7/31/00 -0500, RE: primer sequences]

File Edit Mailbox Message Transfer Special Tools Window Help

Subject: RE: primer sequences

From: Cindy Sprecher <CSpreche@promega.com>
To: "John Butler" <john.butler@nist.gov>, Tom Mozer <TMozer@promega.com>
Cc: Arni Masibay <amasibay@promega.com>,
Cindy Sprecher
<CSpreche@promega.com>
Subject: RE: primer sequences
Date: Mon, 31 Jul 2000 15:53:06 -0500
X-Mailer: Internet Mail Service (5.5.2650.21)

John,

The attached file contains the primer sequence information for PowerPlex 16. Some of the unlabeled oligos have bases added onto the 5' end. We found that the addition of these bases reduce or eliminate the amount of split peaks due to incomplete non-template A addition by the Taq DNA polymerase.

Cindy

GenePrint® PowerPlex® 1.1

Locus	Dye	Primer Sequence
CSF1PO-F	TMR	AACCTGAGTCTGCCAAGGACTAGC
CS		
TP		
TP		
TH		
TH		
VW		
VW		
D1		
D1		
D7		
D7		
D1		
D1		
D5		
D5S818-R	FL	AGCCACAGTTTACAACATTTGTATCT

16 primers
391 bases
8 dye labels

GenePrint® PowerPlex® 2.1

Locus	Dye	Primer Sequence
D3S1358-F		ACTGCAGTCCAATCTGGGT
D3S1358-R	FL	ATGAAATCAACAGAGGCTTGC
TH		
TH		
D2		
D2		
D1		
D1		
Pe		
Pe		
W		
W		
D8		
D8		
TP		
TP		
FGA-F	TMR	GGCTGCAGGGCATAACATTA
FGA-R		ATTCTATGACTTTGCGCTTCAGGA

18 primers
442 bases
9 dye labels

Primer Sequences

GenePrint® PowerPlex® 16

Locus	Dye	Primer Sequence
AMEL-F	TMR	CCCTGGGCTCTGTAAAGAA
AMEL-R		ATCAGAGCTTAAACTGGGAAGCTG
CSF1PO-F	JOE	CCGGAGGTAAAGGTGTCTTAAAGT
CSF1PO-R		ATTTCCTGTGTCAGACCCTGTT
D13S317-F		ATTACAGAAGTCTGGGATGTGGAGGA
D13S317-R	JOE	GGCAGCCCAAAAAGACAGA
D16S539-F		GGGGGTCTAAGAGCTTGTAAAAAG
D16		
D18		
D18		
Per		
Per		
D21		
D21		
Per		
Per		
D3S		
D3S		
D5S		
D5S		
D7S		
D7S820-R		GATTCCACATTTATCCTCATTGAC
D8S1179-F	TMR	ATTGCAACTTATATGTATTTTTGTATTTCATG
D8S1179-R		ACCAAATTGTGTTCATGAGTATAGTTTC
FGA-F	TMR	GGCTGCAGGGCATAACATTA
FGA-R		ATTCTATGACTTTGCGCTTCAGGA
TH01-F	FL	GTGATTCCCATTGGCCTGTTC
TH01-R		ATTCCTGTGGGCTGAAAAGCTC
TPOX-F		GCACAGAACAGGCACTTAGG
TPOX-R	TMR	CGCTCAAACGTGAGGTTG
vWA-F		GCCCTAGTGGATGATAAGAATAATCAGTATGTG
vWA-R	TMR	GGACAGATGATAAATACATAGGATGGATGG

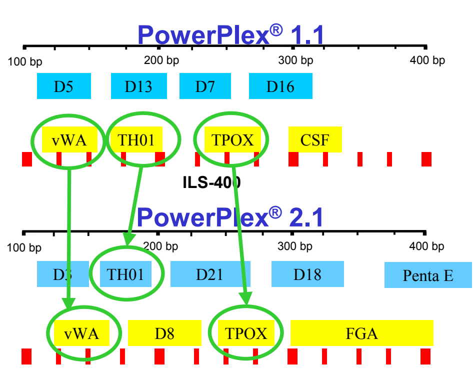
32 primers
769 bases
16 dye labels

Analysis of Released Promega Primer Sequences

A basic list of sequences is difficult to fully comprehend...

1. Examine primer position changes with same loci between kits
2. Analyze primer sequences with respect to primer design characteristics

Can we gain a better understanding of multiplex PCR primer design characteristics based on successful working STR kits?



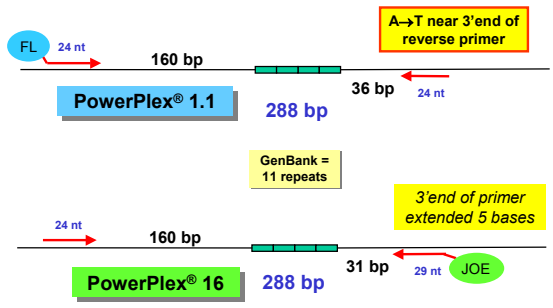
Overlapping Loci between Promega PowerPlex® 1.1 and 2.1 STR Kits

- PP1.1 and PP2.1 have TH01, TPOX, and VWA in common, but product sizes and primer positions differ for TH01 and TPOX
- Summary:
  - **TH01** allele 9 goes from 195 bp (1.1) to 176 bp (2.1) change of –19 bp
  - **TPOX** allele 11 goes from 244 bp (1.1) to 282 bp (2.1) change of +38 bp
  - **VWA** allele 18 stays consistent in size at 155 bp (1.1/2.1) **but monoplex primers differ for VWA**

Significant Primer Changes for Same Loci between PP1.1 and PP16

- D16S539: amplicon size remains constant
- D13S317: amplicon size remains constant
- CSF1PO: amplicon size increases +30 bp (PP16)

D16S539 Primer Changes



Same forward sequence; reverse sequence lengthened by 5 bases at 3'end; dye label changed from FL to JOE and labeled strand changed from forward to reverse; PCR product size remains unchanged

D16S539 Primer Changes

PowerPlex® 1.1 on top and PowerPlex® 16 on bottom

5′-FL-GGGGGTCTAAGAGCTTGTAAAAAG-3′

5′-GGGGGTCTAAGAGCTTGTAAAAAG-3′

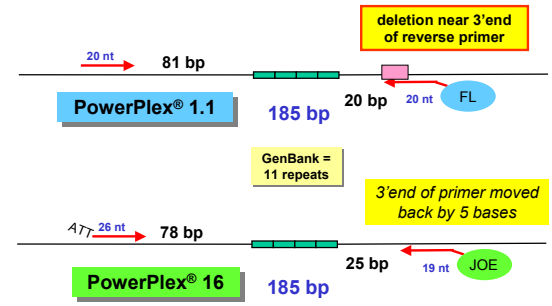
A→T in black population

3′-TACGAATGTCTACGTGTGTGTTTG-5′

3′-CTATGTACGAATGTCTACGTGTGTGTTTG-JOE-5′

Extra bases enhance primer binding and prevent allele drop-out from base change in DNA template

D13S317 Primer Changes



Forward primer lengthened by 6 bases, contains 5' tag, and is 3 bases closer to repeat region; reverse primer shortened by 1 base and moved 5 bases further from the repeat region; dye label changed from FL to JOE; labeled strand remains the same; PCR product size remains unchanged

D13S317 Primer Changes

PowerPlex® 1.1 on top and PowerPlex® 16 on bottom

5′-ACAGAAGTCTGGGATGTGGA-3′

5′-ATTACAGAAGTCTGGGATGTGGAGGA-3′

Added sequence to promote non-template addition

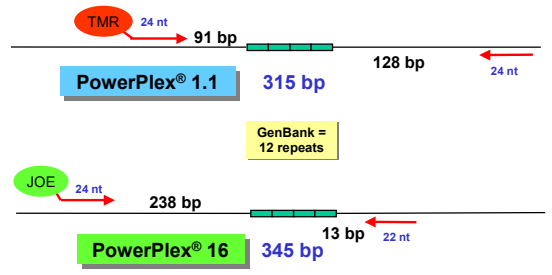
Deletion in black population

3′-AAGACAGACAGAAAAACCCG-FL-5′

3′-AGACAGAAAAACCCGACGG-JOE-5′

Shortening primer avoids 4 base deletion region that can cause allele dropout in some DNA templates

CSF1PO Primer Changes



Both forward and reverse primers are completely non-overlapping…

Relative Primer Positions

Null Alleles

- Allele is present in the DNA sample but fails to be amplified due to a nucleotide change in a primer binding site
- Allele dropout is a problem because a heterozygous sample appears falsely as a homozygote
- Two PCR primer sets can yield different results on samples originating from the same source
- This phenomenon impacts DNA databases
- Large concordance studies are typically performed prior to use of new STR kits

Apparent Null Alleles Observed During Concordance Studies

7/13 CODIS loci affected so far

D16S539 PP1.1 vs PP16 vs COfiler

Loss of alleles with PP1.1 in Black population samples; fine with PP16 and COfiler

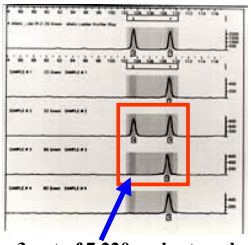
D8S1179 PP16 vs Profiler Plus or SGM Plus

Loss of alleles with Profiler Plus/SGM Plus in Asian samples; fine with PP16

Locus	Kits Compared	Results	Reference
D13	PP1.1 vs PP16 vs ProPlus	Loss of alleles 9,10, and 11 with PP1.1; fine with PP16 and ProPlus	Promega meeting Oct 2000
FGA	SGM vs SGM Plus	Loss of allele 26 with SGM Plus; weak amp of same allele with SGM	Cotton 2000
CSF	PP16 vs COfiler	Weak amp on allele 14 with COfiler; fine with PP16	Promega meeting Oct 2000
CSF	PP16 vs Profiler	Weak amp on allele 8 with PP16; fine with Profiler	Promega meeting Oct 2000
TPOX	PP16 vs Profiler	Weak amp on allele 9 with PP16; fine with Profiler	Promega meeting Oct 2000

Amelogenin Null Allele

http://www.fbi.gov/programs/lab/fsc/backissu/oct2000/shewale.htm



3 out of 7,220 males typed exhibited loss of X-allele

"The most probable explanation for this anomalous phenomenon is that these **samples had a mutation** on the X chromosome **within the primer-binding site** for the specific primer provided in the AmpFLSTR® Profiler Plus™ amplification kit."

Summary of Amelogenin Primer Information from the Literature

- Amel-X allele dropout reported to occur in ABI kits (see Reliagene article in Oct 2000 *Forensic Science Comm.*)
- According to Cotton et al 2000, ABI primers are the same as Forensic Science Service (Sullivan 1993)
- AMEL-F is 5 bases shorter in PP16 at 3'end; AMEL-R is identical between ABI kits and PP16

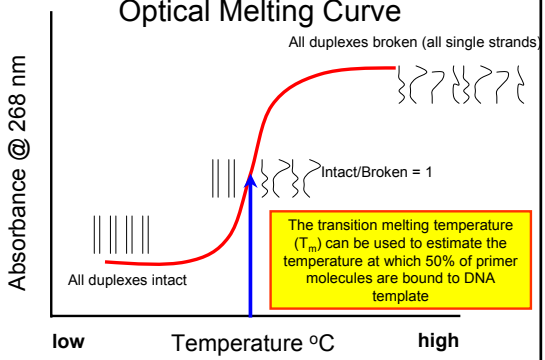
```
AMEL-F (PP16) 5'-CCCTGGGCTCTGTAAAGAA-3'
AMEL-F (ABI)  5'-CCCTGGGCTCTGTAAAGAATAGTG-3'

AMEL-R (PP16/ABI) 3'-GTCGAAGGGTCAAATTCGAGACTA-5'
```

Primer Design

- Typically performed with assistance of computer program to identify possible primer that are then tested empirically
- Various computer programs:
  - Gene Runner (PC), Oligo (PC/Mac), Primer Express (Mac)
  - Primer 3 (web based)
- Critical parameters examined:
  - Predicted T_m (melting temperature)
  - Primer dimer and hairpin formation
  - Contiguous base runs (usually <5 bases)
  - GC content (number of G and C nucleotides within primer)

Optical Melting Curve



Predicted T_m does **not** account for the issues of kinetics, mis-priming, and secondary structure formation associated with performing PCR.

How is T_m predicted?

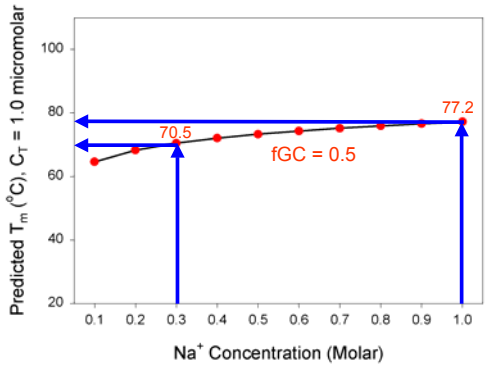
For non-self-complementary duplexes the "T_m" can be calculated from the equation

$$T_m = \frac{\Delta H}{\Delta S + R\,ln(C_T/4)}$$

- Values for ΔH and ΔS have been evaluated for duplexes with varying sequence content and context -- creating "nearest neighbor" data sets
- Evaluated parameters for ΔH and ΔS are used to estimate T_m in the above equation
- C_T represents the total strand concentration

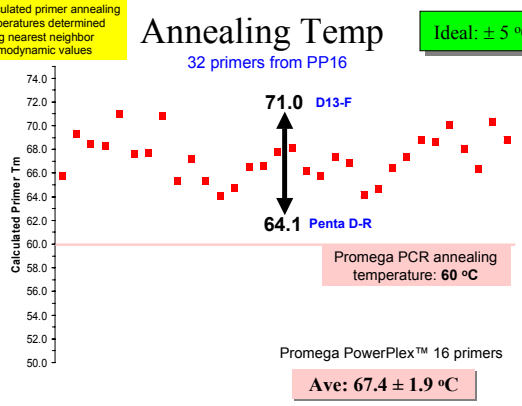
Owczarzy, R., **Vallone, P. M.**, Gallo, F. J., Paner, T. M., Lane, M. J., and Benight, A. S. (1997) Predicting sequence-dependent melting stability of short duplex DNA oligomers. *Biopolymers*. 44(3): 217-239.
SantaLucia, J. (1998) A unified view of polymer, dumbbell, and oligonucleotide DNA nearest- neighbor thermodynamics. *Proc.Natl.Acad.Sci.U.S.A.* 95(4): 1460-1465.

The Effect of Salt on Predicted T_m of a 20mer

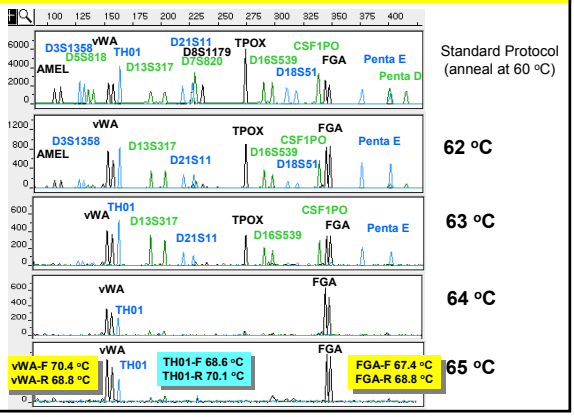


Primer Characteristics: Calc Tm

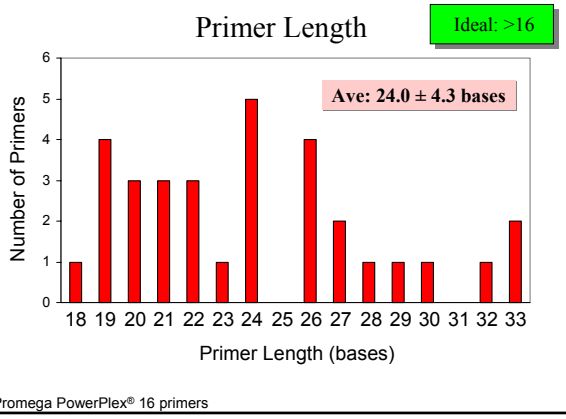
Calculated primer annealing temperatures determined using nearest neighbor thermodynamic values



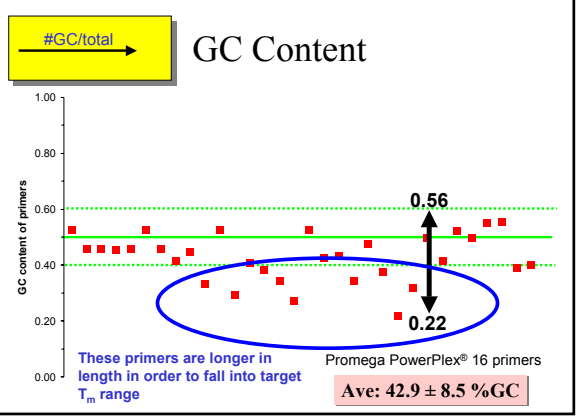
PP16 Results when Amplified with Higher Annealing Temperatures



Primer Characteristics: Length



Primer Characteristics: Fraction GC

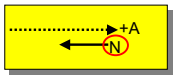


Primer Characteristics: Primer Ends

Primer Ends

- 5'-end of impacts non-template addition
  - Only primer opposite dye label impacts fluorescent data collected
- 3'-end critical to primer annealing
  - Where degenerate bases might be placed if a polymorphic nucleotide is known to occur in various DNA templates

Primer Characteristics: Primer Ends



5'-Nucleotide

Last Base for Primer Opposite Dye Label

Promega PP16

100% purines!

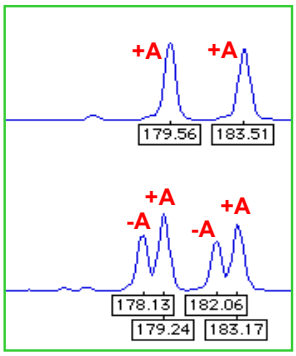
A: 11/16
G: 5/16
T: 0/16
C: 0/16

"ATT" added to 5'-end of 7/16 primers

Impacts non-template addition

placing the sequence **G**TTTCTT on the 5' end of reverse primers resulted in nearly 100% adenylation of the 3' end of the forward strand… (Brownstein,M.J. et al. 1996)

Impact of the 5' nucleotide on Non-Template Addition

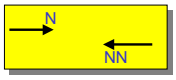


5'-ACAAG…

Last Base for Primer Opposite Dye Label

5'-CCAAG…

Primer Characteristics: Primer Ends



3'-Nucleotides

Promega PowerPlex 16

(out of 32 primers)

Most important portion of primer for annealing…

Last Base	
A	7
T	5
G	11
C	9

%A: 22% **%G: 34%**

%T: 16% %C: 28%

4 out of 4 possibilities

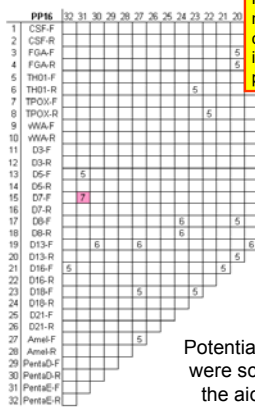
Last Two Bases	
AA	1
TA	3
GA	3
TT	2
GT	2
CT	1
AG	3
TG	6
GG	2
AC	3
TC	4
GC	1
CC	1

13 out of 16 possibilities

Last Three Bases			
GAA	1	AGG	1
ATA	1	TGG	1
TTA	2	AAC	1
AGA	1	GAC	2
GGA	2	ATC	1
TTT	1	TTC	2
GTT	1	CTC	1
AGT	1	TGC	1
GGT	1	TCC	1
TCT	1		
AAG	3		
ATG	2		
TTG	1		
GTG	2		
CTG	1		

24 out of 64 possibilities

Primer Characteristics: Dimers



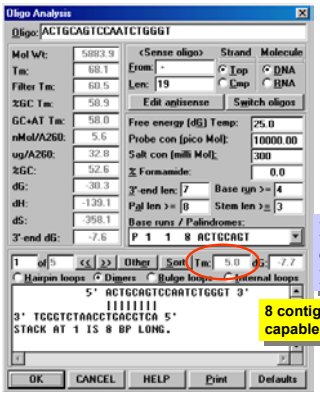
Numerical values indicate the number of bases that are complementary and can potentially interact between two different primers

D3S1358-F versus D3S1358-F
Blast = 8

5-ACTGCAGTCCAATCTGGGT-3
3-TGGGTCTAACCTGACGTCA-5

Potential primer dimer interactions were scanned for all primers with the aid of a computer program

Primer Characteristics: Dimers

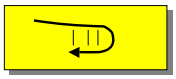


Analysis of D3-F Primer from PowerPlex® 16 Kit

Predicted to be stable at 5 °C or less (not significant under PCR conditions)

8 contiguous bases capable of dimer formation

Primer Characteristics: Hairpins



Hairpins

$T_m < T_{PCRanneal}$

Promega PowerPlex 16

AMEL-R
(same as FSS/ABI kits)

```
5' ATCAGAGCTTAAAC
          |||||     ]
3'      GTCGAAGGGT
```

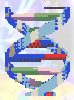
Stable to ~70.1 °C

Summary of Results

- Some primer positions have been moved between various Promega STR kits that amplify the same loci
- Concordance studies are important to see if typing results differ with new primer sets and can be greatly aided by knowledge of primer sequence positions
- There is nothing unusual about the primers included in commercial STR kits…they fall within typical primer design parameters
- Careful evaluation of released primer sequences that come from empirically balanced multiplexes can lead to a better understanding of optimal multiplex PCR primer design parameters

STRBase

Short Tandem Repeat DNA Internet Database



Published in *Nucleic Acids Research* Database Issue Jan 2001
(downloadable pdf on home page)

General Information	Forensic Interest Data	Supplemental Info
•Intro to STRs (downloadable PowerPoint)	•FBI CODIS Core Loci	•Reference List 1373
•STR Fact Sheets	•DAB Standards	•Technology Review
•Sequence Information	•NIST SRM 2391	•Addresses for Scientists
•Multiplex STR Kits	•Published PCR Primers	•Links to Other Web Sites
•Variant Allele Reports	•Y-Chromosome STRs	
	•Population Data	Released Promega sequences now listed
	•Validation Studies	

http://www.cstl.nist.gov/biotech/strbase

Acknowledgments

Funding:

National Institute of Justice Grant #97-LB-VX-0003

Interagency Agreement between NIJ and NIST Office of Law Enforcement Standards

Pete Vallone
Margaret Kline
Lisa Forman

Please contact NIJ if you would like to know more about this research

Collaborators

Joe Devaney (Transgenomic)
Mike Marino (Transgenomic)