Capillary Electrophoresis in DNA Analysis

STR Analysis

NEAFS Workshop
Mystic, CT
September 29-30, 2004

Dr. John M. Butler
Dr. Bruce R. McCord

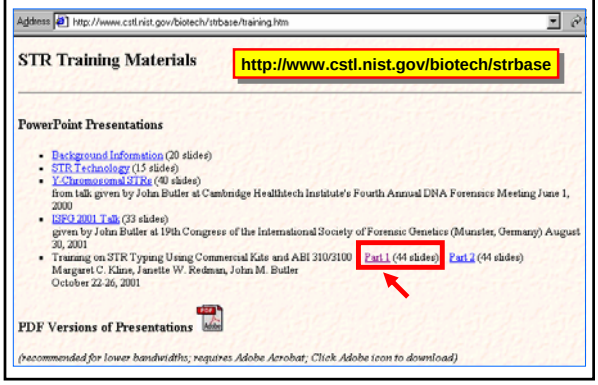
NIST National Institute of Standards and Technology
Technology Administration, U.S. Department of Commerce

FIU Florida International University
Miami's public research university

Outline for Workshop

- Introductions
- STR Analysis
- Introduction to CE and ABI 310
- Data Interpretation
- Additional Topics – Real-time PCR and miniSTRs
- Higher Throughput Approaches
- Troubleshooting the ABI 310 (Participant Roundtable)
- Additional Topics – Y-STRs, validation, accuracy
- Review and Test

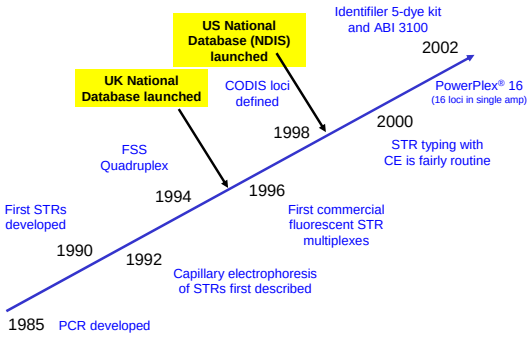
Training Materials on STRBase



Human Identity Testing

- Forensic cases -- **matching suspect with evidence**
- Paternity testing -- **identifying father**
- Historical investigations
- Missing persons investigations
- Mass disasters -- **putting pieces back together**
- Military DNA "dog tag"
- Convicted felon DNA databases

Historical Perspective on DNA Typing



National DNA Index System (NDIS)

FBI LABORATORY

CODIS COMBINED DNA INDEX SYSTEM

http://www.fbi.gov/hq/lab/codis/index1.htm

Combined DNA Index System (CODIS)

Launched in October 1998 and now links all 50 states

Used for linking serial crimes and unsolved cases with repeat offenders

Convicted offender and forensic case samples along with a missing persons index

Requires 13 core STR markers

>18,000 investigations aided nationwide as of June 2004

Contains more than 1.8 million DNA profiles

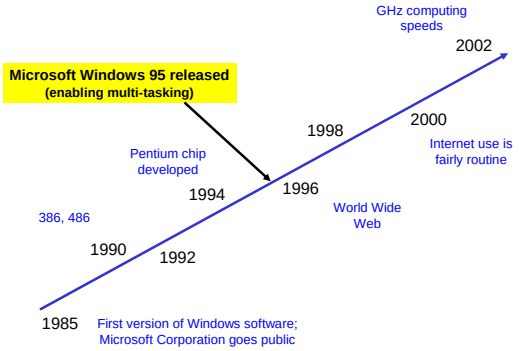
Analogy of Computers and DNA Typing

- Computers originated in 1946 and began as sophisticated devices that were appreciated and utilized by a select few (the "nerds").
- Computer use has now expanded to the layperson and general household use as computers have become easier to use.
- One of the greatest advances for the personal computer was the advent of multitasking (the ability to perform simultaneous applications).

- PCR and DNA typing methods first began as experimental exercises in 1984 but have progressed to become a general workhorse among forensic laboratories around the world.
- Multiplexing is a corresponding breakthrough for PCR as simultaneous amplification of different segments of DNA can be performed in a single reaction.
- STR typing is now widely used because of the availability of commercial kits that makes DNA testing easier to perform.

Adapted from John J. Sninsky, *PCR Applications: Protocols for Functional Genomics*, Academic Press, 1999.

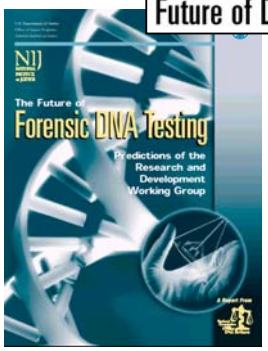
Historical Comparison with Computers



Importance of Standardized Data Formats

- DNA databases work because everyone is submitting data on the same genetic markers (6 SGM loci or 13 CODIS STR loci) to insure "communication" between all the samples
- Internet works because computers are networked with common protocols (TCP/IP) and "speak" a common language

National Commission on the Future of DNA Evidence



http://www.ojp.usdoj.gov/nij/pubs-sum/183697.htm

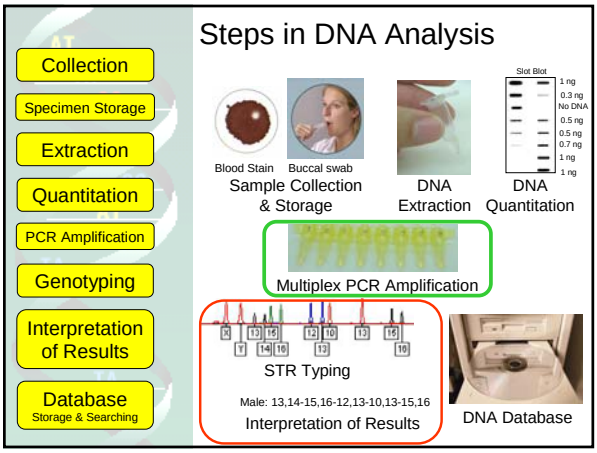
•Report published in Nov 2000

•Asked to estimate where DNA testing would be 2, 5, and 10 years into the future

Conclusions

STR typing is here to stay for a few years because of DNA databases

Steps in DNA Analysis



Basic Concepts

PCR (polymerase chain reaction) – method of amplifying a specific region of the genome – go from 1 to over a billion copies in 2-3 hours

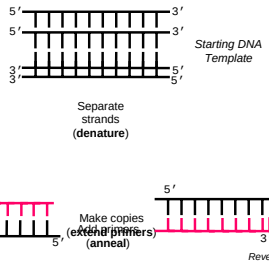
Locus region of the genome being examined

Allele the state of the genetic variation being examined
(STRs = number of repeat units)
(SNPs = base sequence at the site)

Chromosomes are paired so…
Homozygous – Alleles are identical on each chromosome
Heterozygous - Alleles differ on each on each chromosome

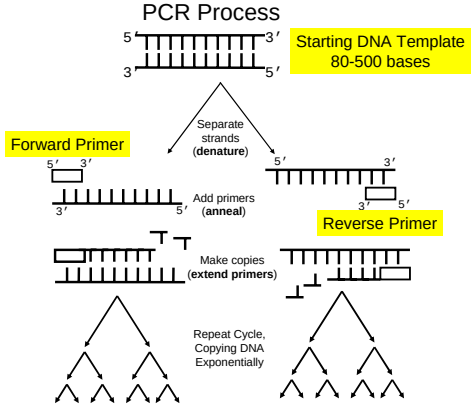
DNA Amplification with the Polymerase Chain Reaction (PCR)

Components of PCR
Thermal stable polymerase
dNTPs (building blocks)
$MgCl_2$
Buffer
Primers (specific to site)



In 32 cycles at 100% efficiency, 1.07 billion copies of targeted DNA region are created

PCR Process



Butler, J.M. (2001) *Forensic DNA Typing*, Figure 4.2, ©Academic Press

Thermal Cycling Parameters

Step in Protocol	AmpFlSTR® kits (Applied Biosystems)	PowerPlex® STR kits (Promega Corporation)
Initial Incubation	95 °C for 11 minutes	95 °C for 11 minutes
Thermal Cycling	**28 cycles**	**30 cycles**[a]
Denature	94 °C for 1 minute	94 °C for 30 seconds (cycle 1-10) 90 °C for 30 seconds (cycle 11-30)
Anneal	59 °C for 1 minute	60 °C for 30 seconds
Extend	72 °C for 1 minute	70 °C for 45 seconds
Final Extension	60 °C for 45 minutes	60 °C for 30 minutes
Final Soak	25 °C (until samples removed)	4 °C (until samples removed)

[a] The first 10 cycles are run with a denaturation temperature of 94 °C and the last 20 cycles are run at 90 °C instead. The Promega PowerPlex 1.1 and 2.1 kits also use specific ramp times between the different temperature steps that differ from the conventional 1 °C/second.

Butler, J.M. (2001) *Forensic DNA Typing*, Table 4.3, ©Academic Press

Advantages of PCR

- Minute amounts of DNA template may be used from as little as a single cell.
- DNA degraded to fragments only a few hundred base pairs in length can serve as effective templates for amplification.
- Large numbers of copies of specific DNA sequences can be amplified simultaneously with multiplex PCR reactions.
- Contaminant DNA, such as fungal and bacterial sources, will not amplify because human-specific primers are used.
- Commercial kits are now available for easy PCR reaction setup and amplification.

Butler, J.M. (2001) *Forensic DNA Typing*, Chapter 4, p. 50, ©Academic Press

Potential Pitfalls of PCR

- The target DNA template may not amplify due to the presence of PCR inhibitors in the extracted DNA
- Amplification may fail due to sequence changes in the primer binding region of the genomic DNA template
- Contamination from other human DNA sources besides the forensic evidence at hand or previously amplified DNA samples is possible without careful laboratory technique and validated protocols

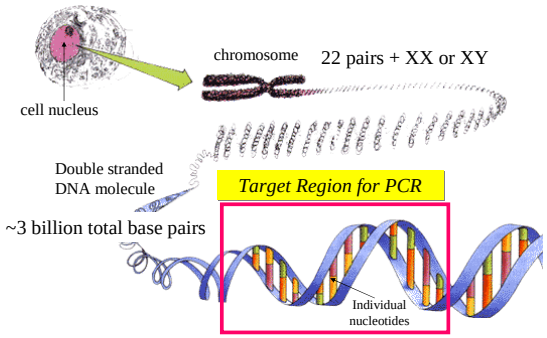
Butler, J.M. (2001) *Forensic DNA Typing*, Chapter 4, pp. 50-51, ©Academic Press

Tips for Avoiding Contamination

- Pre- and post-PCR sample processing areas should be physically separated.
- Equipment, such as pipettors, and reagents for setting up PCR should be kept separate from other lab supplies, especially those used for analysis of PCR products.
- Disposable gloves should be worn and changed frequently.
- Reactions may also be set up in a laminar flow hood, if available.
- Aerosol-resistant pipet tips should be used and changed on every new sample to prevent cross-contamination during liquid transfers.
- Reagents should be carefully prepared to avoid the presence of any contaminating DNA or nucleases.
- Ultraviolet irradiation of laboratory PCR set-up space when the area is not in use and cleaning workspaces and instruments with isopropanol and/or 10% bleach solutions help to insure that extraneous DNA molecules are destroyed prior to DNA extraction or PCR set-up

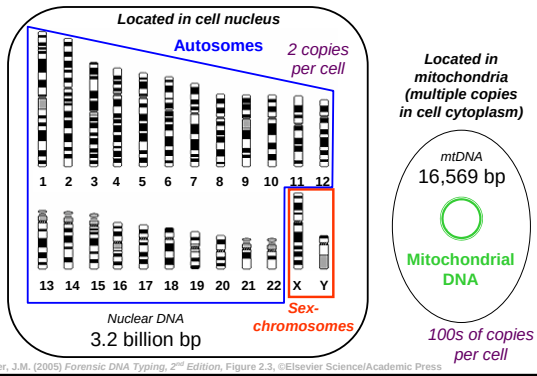
Butler, J.M. (2001) *Forensic DNA Typing*, Chapter 4, pp. 49-50, ©Academic Press

DNA in the Cell



Human Genome

23 Pairs of Chromosomes + mtDNA



Butler, J.M. (2005) *Forensic DNA Typing, 2nd Edition*, Figure 2.3, ©Elsevier Science/Academic Press

What Type of Genetic Variation?

•Length Variation

short tandem repeats (STRs)

CTAGTCGT**(GATA)(GATA)(GATA)**GCGATCGT

•Sequence Variation

single nucleotide polymorphisms (SNPs)

insertions/deletions

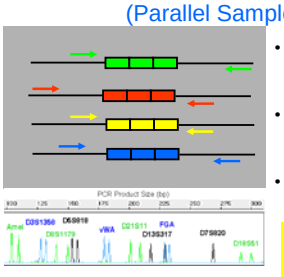
GCTAGTCGATGCTC**(G/A)**GCGTATGCTGTAGC

Why STRs are Preferred Genetic Markers

- Rapid processing is attainable
- Abundant throughout the genome
- Highly variable within various populations
- Small size range allows multiplex development
- Discrete alleles allow digital record of data
- Allelic ladders simplify interpretation
- PCR allows use of small amounts of DNA material
- Small product size compatible with degraded DNA

Multiplex PCR

(Parallel Sample Processing)



- **Compatible primers are the key to successful multiplex PCR**
- **10 or more STR loci can be simultaneously amplified**
- **STR kits are commercially available**

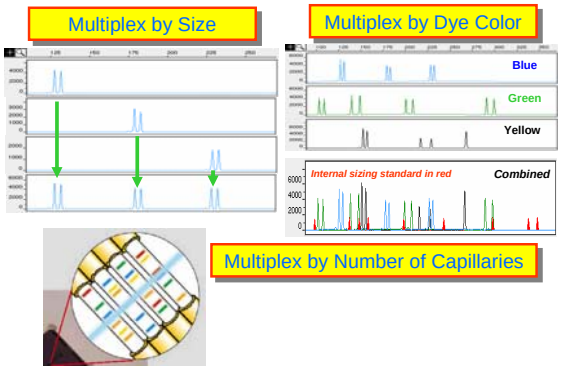
Challenges to Multiplexing

–primer design to find compatible primers (no program exists)
–reaction optimization is highly empirical often taking months

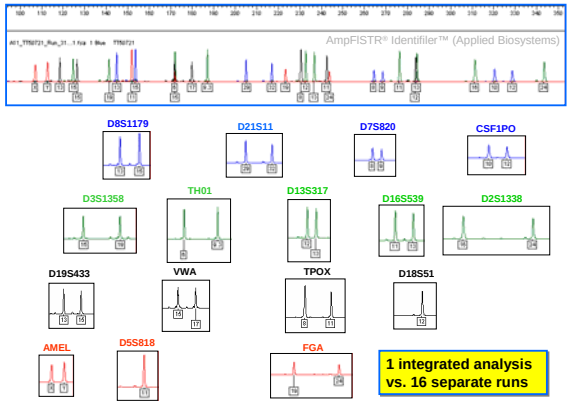
Advantages of Multiplex PCR

–Increases information obtained per unit time (increases power of discrimination)
–Reduces labor to obtain results
–Reduces template required (smaller sample consumed)

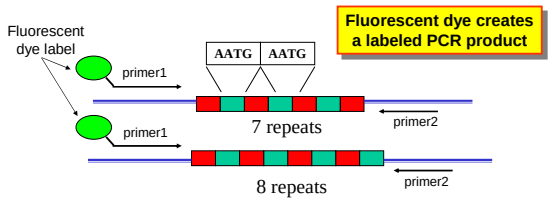
Methods for Parallel Sample Processing



Information is tied together with multiplex PCR and data analysis



Short Tandem Repeats (STRs)



the repeat region is variable between samples while the flanking regions where PCR primers bind are constant

Homozygote = both alleles are the same length
Heterozygote = alleles differ and can be resolved from one another

Primer positions define PCR product size

Types of STR Repeats

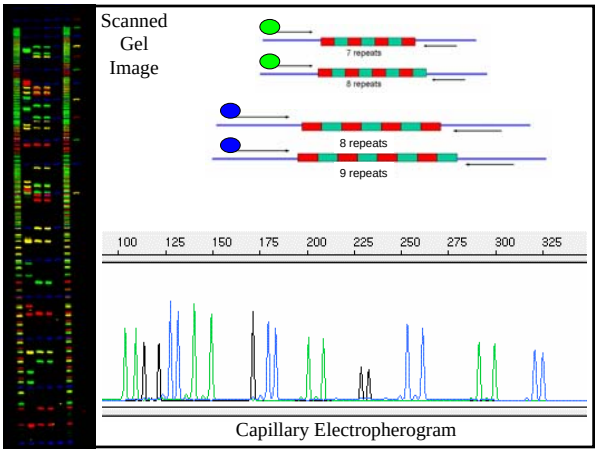
Microsatellite = simple sequence repeat (SSR) = short tandem repeat (STR)

High stutter

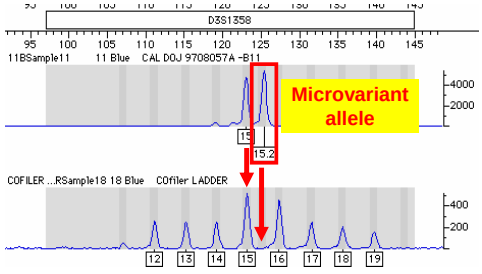
- **Di**nucleotide (CA)(CA)(CA)(CA)
- **Tri**nucleotide (GCC)(GCC)(GCC)
- **Tetra**nucleotide (AATG)(AATG)(AATG)
- **Penta**nucleotide (AGAAA)(AGAAA)
- **Hexa**nucleotide (AGTACA)(AGTACA)

Low stutter

Require size based DNA separation to resolve different alleles; dinucleotide repeats require higher resolution



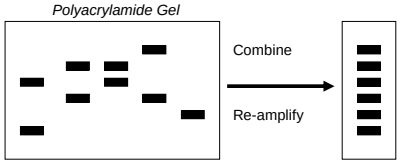
STR genotyping is performed by comparison of sample data to allelic ladders



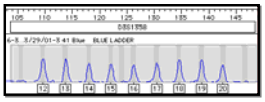
Allelic Ladder Formation

Separate PCR products from various samples amplified with primers targeted to a particular STR locus

Polyacrylamide Gel



Find representative alleles spanning population variation



Butler, J.M. (2001) *Forensic DNA Typing*, Figure 5.3, ©Academic Press

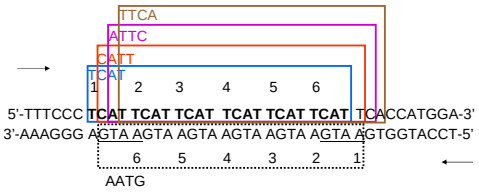
STR Repeat Nomenclature

International Society of Forensic Genetics (ISFG) – *Int. J. Legal Med.* (1997) 110:175-176

- For sequences within genes, use the coding strand
- For other sequences, select the first GenBank database entry or original literature description
- Define the repeat sequence which will provide the largest number of consecutive repeats
- If two sequences are repeated, include both motifs in determining the repeat number
- Microvariants: should be designated by the number of complete repeats and the number of base pairs of the partial repeat separated by a decimal point (Int. J. Legal Med. 1994, 107:159-160) *e.g. TH01 allele 9.3*

Nomenclature Example

(TH01)

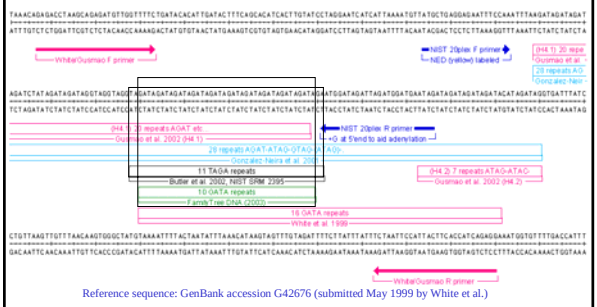


Edwards *et al.* (1991) used **AATG** (adopted early on by Promega)
Kimpton *et al.* (1993) used **TCAT** (Forensic Science Service) – ***most widely used now***

Differences in nomenclature can lead to confusion or even worse problems with database matches—standardization and consistency in use is essential…

Use of "TAGA" vs "GATA" results in a single repeat difference

(Y-GATA-H4)



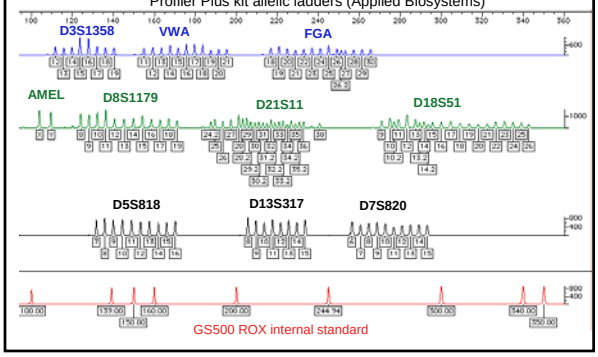
Reference sequence: GenBank accession G42676 (submitted May 1999 by White et al.)

This simple difference has impacted the genetic genealogy community

NIST SRM 2395 follows ISFG guidelines (for our primer pair): first adjacent repeat starting from 5'end is TAGA

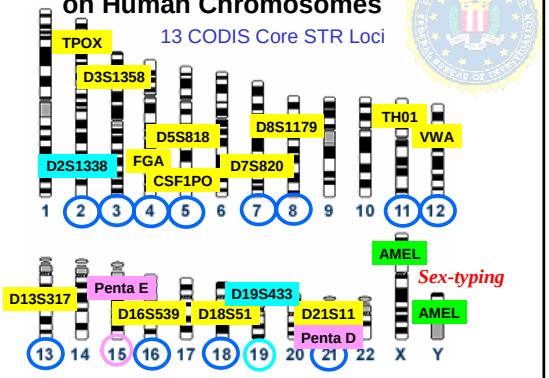
Companies Supply Allelic Ladders in STR Kits to Aid Interlaboratory Consistency

Profiler Plus kit allelic ladders (Applied Biosystems)

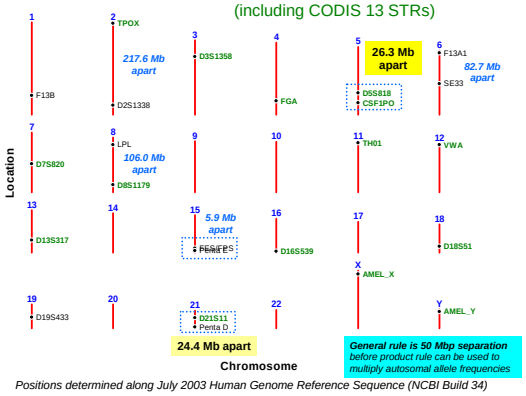


Position of Forensic STR Markers on Human Chromosomes

13 CODIS Core STR Loci



Commercial STR Kit Loci Positions
(including CODIS 13 STRs)



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

Information on 13 CODIS STRs

Locus Name	Chromosomal Location	Repeat Motif ISFG format	GenBank Accession	Allele in GenBank	Allele Range	Number of Alleles Seen*
CSF1PO	5q33.1	TAGA	X14720	12	5-16	20
FGA	4q31.3	CTTT	M64982	21	12.2-51.2	80
TH01	11p15.5	TCAT	D00269	9	3-14	20
TPOX	2p25.3	GAAT	M68651	11	4-16	15
VWA	12p13.31	[TCTG][TCTA]	M25858	18	10-25	28
D3S1358	3p21.31	[TCTG][TCTA]	NT_005997	18	8-21	24
D5S818	5q23.2	AGAT	G08446	11	7-18	15
D7S820	7q21.11	GATA	G08616	12	5-16	30
D8S1179	8q24.13	[TCTA][TCTG]	G08710	12	7-20	17
D13S317	13q31.1	TATC	G09017	13	5-16	17
D16S539	16q24.1	GATA	G07925	11	5-16	19
D18S51	18q21.33	AGAA	L18333	13	7-39.2	51
D21S11	21q21.1	Complex [TCTA][TCTG]	AP000433	29	12-41.2	82

*Butler, Forensic DNA Typing (2nd edition), Appendix I

Butler, J.M. (2005) *Forensic DNA Typing, 2nd Edition*, Table 5.2, ©Elsevier Science/Academic Press

Commercial STR Kits

Kit Contents:

Allelic Ladders for Genotyping

PCR Component Mix

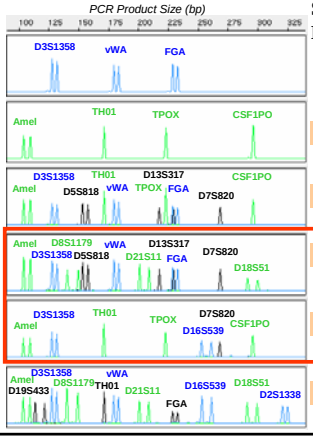
Primer Mix

Positive Control DNA Sample

Cost to User: $15-30 per DNA sample tested

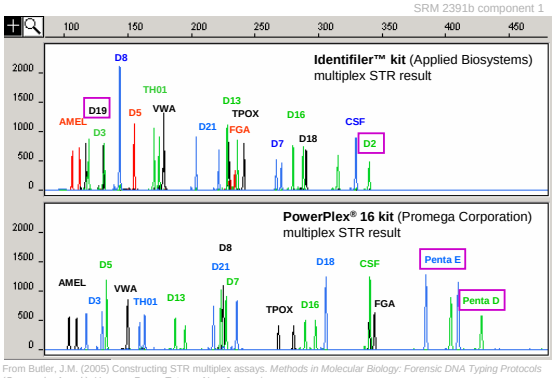
Currently 2 Suppliers: Applied Biosystems and Promega Corporation

Same DNA Sample Run with Each of the ABI STR Kits



	Power of Discrimination
Blue	1:5000
Green I	1:410
Profiler™	$1:3.6 \times 10^9$
Profiler Plus™	$1:9.6 \times 10^{10}$
COfiler™	$1:8.4 \times 10^5$
SGM Plus™	$1:3.3 \times 10^{12}$

Commercial STR 16plex Kits



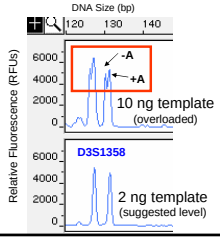
From Butler, J.M. (2005) Constructing STR multiplex assays. *Methods in Molecular Biology: Forensic DNA Typing Protocols* (Carracedo, A., ed.), Humana Press: Totowa, New Jersey, *in press*.

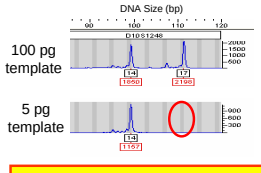
Impact of DNA Amount into PCR

Reason that DNA Quantitation is Important Prior to Multiplex Amplification

Generally 0.5 – 2.0 ng DNA template is best for STR kits

- Too much DNA
  - Off-scale peaks
  - Split peaks (+/-A)
  - Locus-to-locus imbalance
- Too little DNA
  - Heterozygote peak imbalance
  - Allele drop-out
  - Locus-to-locus imbalance





Stochastic effect when amplifying low levels of DNA produces allele dropout

Value of STR Kits

Advantages

- Quality control of materials is in the hands of the manufacturer
- Improves consistency in results across laboratories – same allelic ladders used
- Common loci and PCR conditions used – aids DNA databasing efforts
- Simpler for the user to obtain results

Disadvantages

- Contents may not be completely known to the user (e.g., primer sequences)
- Higher cost to obtain results

FSS: 5X higher cost with SGM Plus kit

PCR Primer Quality Control

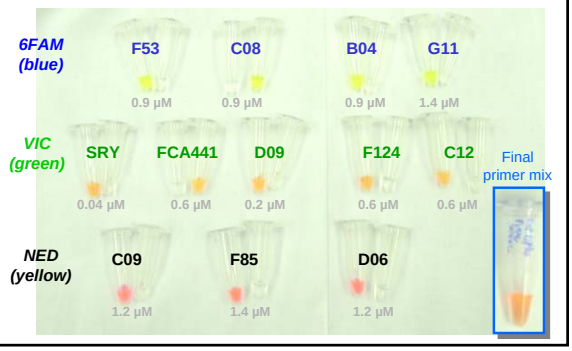
6FAM (yellow), VIC (orange), NED (red)

- UV Spec to determine concentration
- HPLC to evaluate purity
- TOF-MS to confirm correct sequence
- CE (ABI 310) to determine presence of residual dye molecules ("dye blobs")

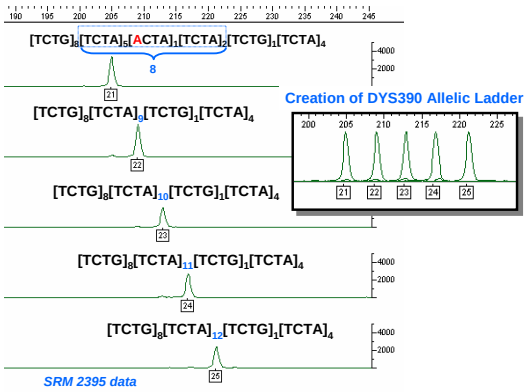
Butler *et al.* (2001) *Forensic Sci. Int.* 119: 87-96

Multiplex PCR Requires QC and Balancing of Many Primers

(24 primers used in cat STR 12plex assay)

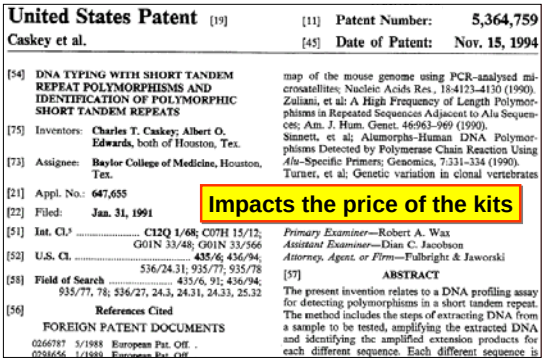


Allelic Ladder Characterization and Creation



Shipping Issues…

Core Patent Licensed to Promega Corporation and Applied Biosystems Inc



United States Patent [19]
Caskey et al.

[11] Patent Number: 5,364,759
[45] Date of Patent: Nov. 15, 1994

[54] DNA TYPING WITH SHORT TANDEM REPEAT POLYMORPHISMS AND IDENTIFICATION OF POLYMORPHIC SHORT TANDEM REPEATS

[75] Inventors: Charles T. Caskey; Albert O. Edwards, both of Houston, Tex.

[73] Assignee: Baylor College of Medicine, Houston, Tex.

[21] Appl. No.: 647,655

[22] Filed: Jan. 31, 1991

[51] Int. Cl.[5] C12Q 1/68; C07H 15/12; G01N 33/48; G01N 33/566

[52] U.S. Cl. 435/6; 436/94; 536/24.31; 935/77; 935/78

[58] Field of Search 435/6, 91; 436/94; 935/77, 78; 536/27, 24.3, 24.31, 24.33, 25.32

[56] References Cited
FOREIGN PATENT DOCUMENTS
0266787 5/1988 European Pat. Off.
0284656 1/1989 European Pat. Off.

map of the mouse genome using PCR-analysed microsatellites; Nucleic Acids Res., 18:4123-4130 (1990). Zuliani, et al; A High Frequency of Length Polymorphisms in Repeated Sequences Adjacent to Alu Sequences; Am. J. Hum. Genet. 46:963-969 (1990). Sinnett, et al; Alumorphs-Human DNA Polymorphisms Detected by Polymerase Chain Reaction Using *Alu*-Specific Primers; Genomics, 7:331-334 (1990). Turner, et al; Genetic variation in clonal vertebrates

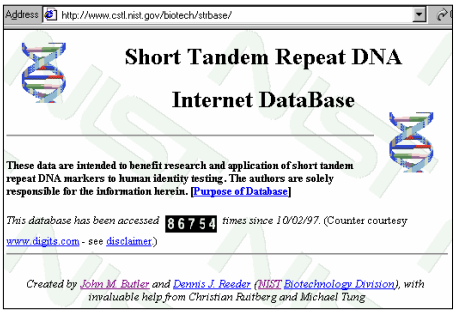
Impacts the price of the kits

Primary Examiner—Robert A. Wax
Assistant Examiner—Dian C. Jacobson
Attorney, Agent, or Firm—Fulbright & Jaworski

[57] ABSTRACT

The present invention relates to a DNA profiling assay for detecting polymorphisms in a short tandem repeat. The method includes the steps of extracting DNA from a sample to be tested, amplifying the extracted DNA and identifying the amplified extension products for each different sequence. Each different sequence is

Internet-Accessible Information Resources



Has been used in court cases to support application of forensic DNA technology

Up-to-date information is needed in a rapidly changing field such as DNA typing

STRBase

Short Tandem Repeat DNA Internet Database

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

General Information	Forensic Interest Data	Supplemental Info
•Intro to STRs (downloadable PowerPoint)	•FBI CODIS Core Loci	•Reference List 2172
•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	
	•Validation Studies	

STR Statistics

Very Brief Overview

Allele Frequency Table

D3S1358

Allele	Caucasian N= 302	African American N=258	Hispanic N=140
11	0.00166*		
13		0.00194*	0.00714*
14	0.10265	0.08915	0.07857
15	**0.26159**	0.30233	**0.29286**
15.2		0.00194*	
16	0.25331	**0.33527**	0.28571
17	0.21523	0.20543	0.20357
18	0.15232	0.06008	0.12500
19	0.01159	0.00388*	0.00714*
20	0.00166*		

Allele frequencies denoted with an asterisk (*) are **below the 5/2N minimum allele threshold** recommended by the National Research Council report (NRCII) *The Evaluation of Forensic DNA Evidence* published in 1996.

Most common allele

Butler, J.M., *et al.* (2003) Allele frequencies for 15 autosomal STR loci on U.S. Caucasian, African American, and Hispanic populations. *J. Forensic Sci.* 48(4):908-911.

DNA Statistics

For heterozygous loci

P = 2pq

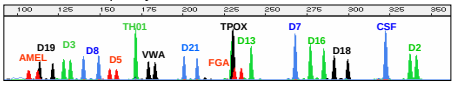
P = probability; p and q are frequencies of allele in a given population

Example: For the locus D3S1358 and individual is 16,17 with frequencies of 0.2315 and 0.2118 respectively

P = 2(0.2315)(0.2118) = 0.0981 or 1 in 10.2

For independent loci, the genotype frequencies can be combined through multiplication…
Profile Probability = (P1)(P2)…(Pn)
= *1 in a very large number…*

DNA Profile Frequency with all 13 CODIS STR loci



AmpFISTR® Identifiler™ (Applied Biosystems)

What would be entered into a DNA database for searching: 16,17-17,18-21,22-12,14-28,30-14,16-12,13-11,14-9,9-11,13-6,6-8,8-10,10

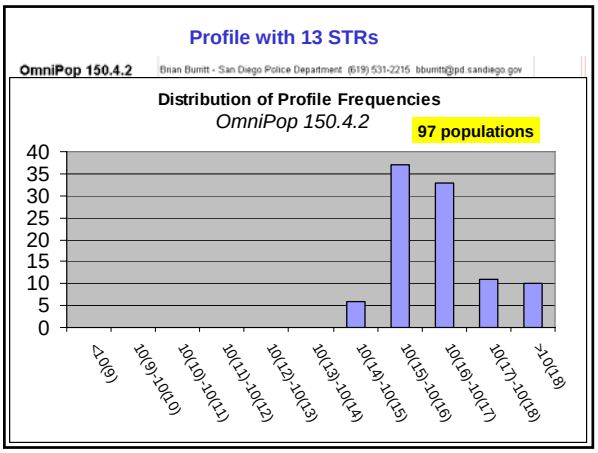
Locus	allele	value	allele	value	frequency, 1 in
D3S1358	**16.0**	*0.2315*	**17.0**	*0.2118*	**10.20**
VWA	**17.0**	*0.2628*	**18.0**	*0.2219*	**8.57**
FGA	**21.0**	*0.1735*	**22.0**	*0.1888*	**15.26**
D8S1179	**12.0**	*0.1454*	**14.0**	*0.2015*	**17.07**
D21S11	**28.0**	*0.1658*	**30.0**	*0.2321*	**12.99**
D18S51	**14.0**	*0.1735*	**16.0**	*0.1071*	**26.91**
D5S818	**12.0**	*0.3539*	**13.0**	*0.1462*	**9.66**
D13S317	**11.0**	*0.3189*	**14.0**	*0.0357*	**43.92**
D7S820	**9.0**	*0.1478*			**43.28**
D16S539	**11.0**	*0.2723*	**13.0**	*0.1634*	**11.24**
THO1	**6.0**	*0.2266*			**18.83**
TPOX	**8.0**	*0.5443*			**3.35**
CSF1PO	**10.0**	*0.2537*			**15.09**

The Random Match Probability for this profile in the FBI Caucasian population is **1 in 1.56 quadrillion (10^{15})**

http://www.csfs.ca/pplus/profiler.htm

OmniPop 150.4.2

- Published allele frequencies
  - from 97 populations containing all 13 CODIS loci
  - From 166 populations with 9 loci (Profiler Plus)
- From 64 publications
- Available from Brian Burritt (San Diego Police Dept)
  - (619) 531-2215
  - bburritt@pd.sandiego.gov



Forensic DNA Typing, 2nd Edition:

Biology, Technology, and Genetics of STR Markers

(John M. Butler, Elsevier Science/Academic Press, Jan 2005)

5 chapters on statistical issues

- **Basic Genetic Principles and Statistics**
- **STR Database Analyses**
- **Profile Frequency Estimates**
- **Approaches to Statistical Analysis of Mixtures**
- **Kinship and Paternity Testing**

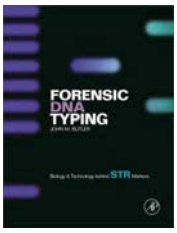
Examples are carefully worked through using the same U.S. population database to illustrate concepts

STR Biology

Biological "Artifacts" of STR Markers

- Stutter Products
- Non-template nucleotide addition
- Microvariants
- Null alleles
- Mutations

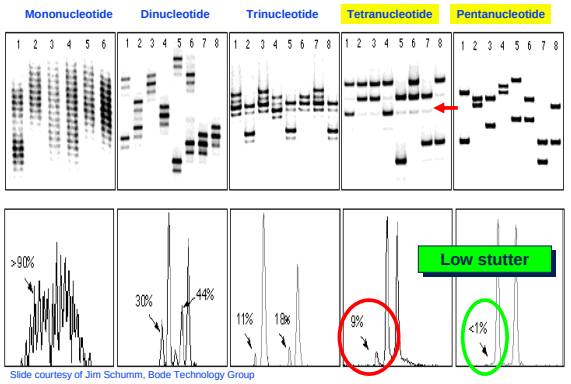
Chapter 6 covers these topics in detail



Stutter Products

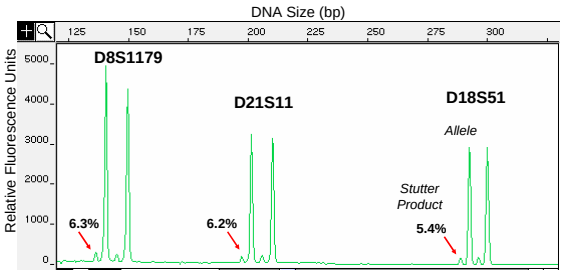
- Peaks that show up primarily one repeat less than the true allele as a result of strand slippage during DNA synthesis
- Stutter is less pronounced with larger repeat unit sizes
  (dinucleotides > tri- > tetra- > penta-)
- Longer repeat regions generate more stutter
- Each successive stutter product is less intense
  (allele > repeat-1 > repeat-2)
- Stutter peaks make mixture analysis more difficult

Stutter Product Intensity versus Repeat Length



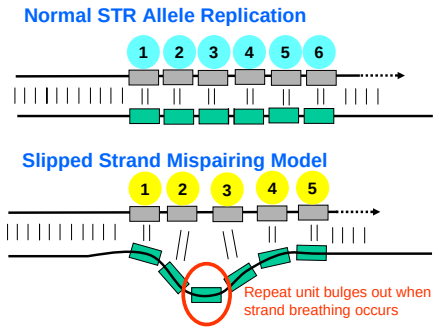
Slide courtesy of Jim Schumm, Bode Technology Group

STR Alleles with Stutter Products

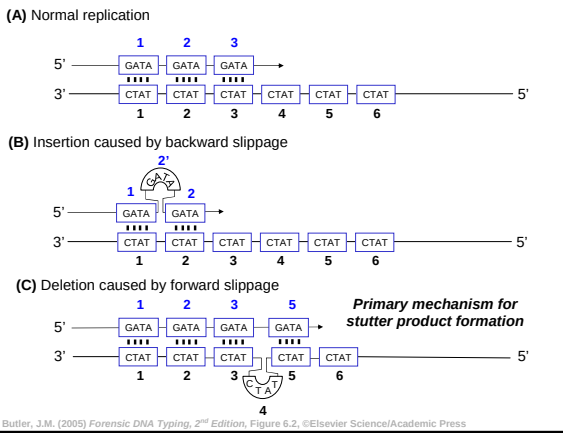


Butler, J.M. (2001) *Forensic DNA Typing*, Figure 6.1, ©Academic Press

Schematic of Stutter Product Formation Process



Walsh *et al* (1996) *Nucleic Acids Res.* 24: 2807-2812

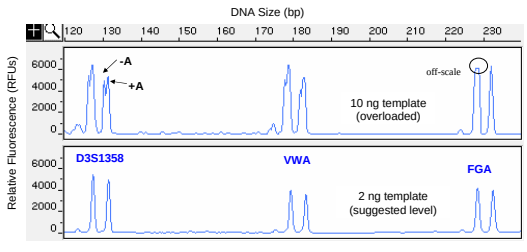


Butler, J.M. (2005) *Forensic DNA Typing, 2nd Edition*, Figure 6.2, ©Elsevier Science/Academic Press

Non-template Addition

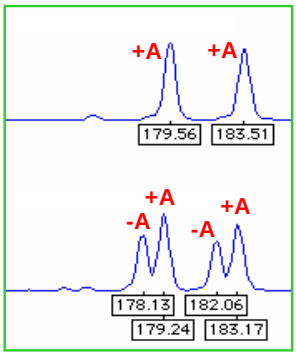
- Taq polymerase will often add an extra nucleotide to the end of a PCR product; most often an "A"
- Dependent on 5'-end of the reverse primer
- Can be enhanced with extension soak at the end of the PCR cycle (e.g., 15-45 min @ 60 or 72 °C)
- Can be reduced with new polymerase
- Best if there is NOT a mixture of "+/- A" peaks

Higher Levels of DNA Lead to Incomplete Adenylation



Butler, J.M. (2001) *Forensic DNA Typing*, Figure 6.4, ©Academic Press

Impact of the 5' nucleotide on Non-Template Addition

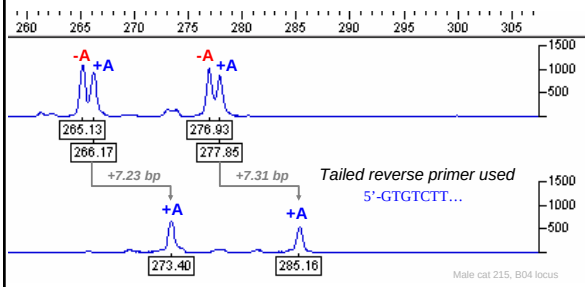


5'-ACAAG…

Last Base for Primer Opposite Dye Label

5'-CCAAG…

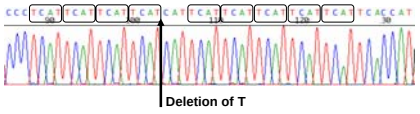
Improving Adenylation with Tailing of Reverse PCR Primer



Makes product 7 bases longer and improves adenylation

Microvariant Alleles

- Defined as alleles that are not exact multiples of the basic repeat motif or sequence variants of the repeat motif or both
- Alleles with partial repeat units are designated by the number of full repeats and then a decimal point followed by the number of bases in the partial repeat
- Example: **TH01 9.3 allele**: $[TCAT]_4$ -CAT $[TCAT]_5$



Variation in the Flanking Region Can Cause Variant Alleles

D7S820 Example

GATAGAACACTTGTCATAGTTTAGAACGAACTAAC**GATAGATAGATAGATAGATAGATAG**
CTATCTTGTGAACAGTATCAAATCTTGCTTGATTG**CTATCTATCTATCTATCTATCTATC**

13 repeat units = $(GATA)_{13}$

TTTTTTTT x.3
TTTTTTTTT x.1

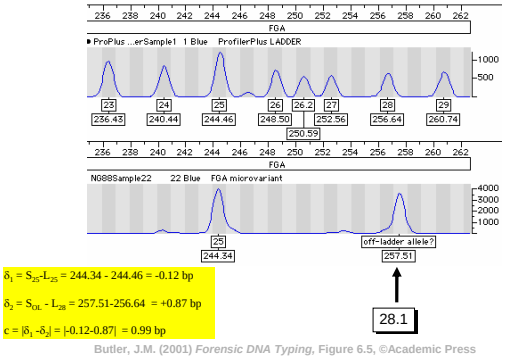
ATAGATAGATAGATAGATAGATAGATAGACAGATTGATAG**TTTTTTTTT**AATCTCACTAAA
TATCTATCTATCTATCTATCTATCTATCTGTCTAACTATC**AAAAAAAAA**TTAGAGTGATTT

The commonly observed x.3 and x.1 alleles are likely the result of a variation in the number of T's found in a poly(T) stretch 13 bases downstream of the core GATA repeat. "On-ladder" alleles contain 9Ts while the x.3 alleles contain 8Ts and the x.1 alleles contain 10Ts (Egyed, B. *et al. Forensic Sci. Int.* 2000, 113, 25-27).

http://www.cstl.nist.gov/biotech/strbase/var_d782.htm

Allele Designation	Allele Size	Instrument	Amp Kit*	Contributor	Verification/Conformation Method(s)	Notes	Frequency
12.1 [3]	281.65	ABI 310	PR	Kelly Duffy/R.Rubocki			
12.1 [4]	281.5	ABI 310	PR	Gintautas Svilpa	Observed both from suspect and crime scene stain		1
12.1 [5]	283.85	ABI 377	PS	Catherine Allor	Reamplified and Reanalyzed	Paternity samples only	1 in 11100
12.3	285.43	ABI 377	CO	Kelly Solis, Texas DPS	Re-extraction	Convicted offender	1 in 68000
13.1	286.8	ABI 310	PR, MP	Margaret Kline	Reamplified with two kits.		1 in 600 samples
13.1 [2]	287.58	ABI 377	CO, PS	Nicole Swinton	Re-extracted and Reamplified		

Detection of a Microvariant Allele at the STR locus FGA



Butler, J.M. (2001) *Forensic DNA Typing*, Figure 6.5, ©Academic Press

Variant Alleles Cataloged in STRBase

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

Off-Ladder Alleles

224 total variants reported as of 04/21/04

Currently **224** at 13/13 CODIS loci

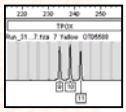
- CSF1PO (10)
- D2S1338 (3)
- D3S1358 (16)
- D5S818 (5)
- D7S820 (20)
- D8S1179 (4)
- D13S317 (8)
- D16S539 (10)
- D18S51 (28)
- D19S433 (8)
- D21S11 (21)
- FES/FPS (1)
- FGA (62)
- HUMTH01 (4)
- Penta D (8)
- Penta E (6)
- TPOX (7)
- VWA (5)

Tri-Allelic Patterns

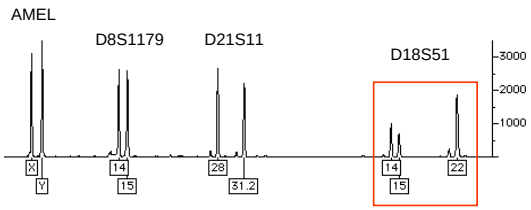
56 total patterns reported as of 04/09/04

Currently **56** at 13/13 CODIS loci

- CSF1PO (2)
- D3S1358 (4)
- D5S818 (1)
- D7S820 (3)
- D8S1179 (5)
- D13S317 (3)
- D16S539 (1)
- D18S51 (4)
- D21S11 (4)
- FGA (9)
- HUMTH01 (1)
- TPOX (12)
- VWA (7)



Three-Peak Pattern at D18S51



Butler, J.M. (2001) *Forensic DNA Typing*, Figure 6.6, ©Academic Press

Three-Peak Patterns

"Type 1"

Balanced peak heights

Most common in TPOX and D21S11

"Type 2"

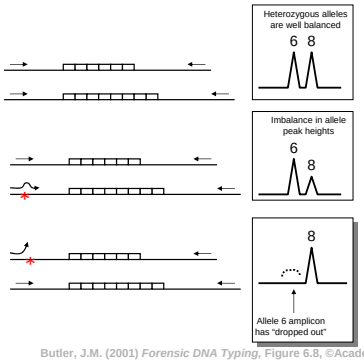
Sum of heights of two of the peaks is equal to the third

Most common in D18S51 and …..

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

Impact of DNA Sequence Variation in the PCR Primer Binding Site



Butler, J.M. (2001) *Forensic DNA Typing*, Figure 6.8, ©Academic Press

Apparent Null Alleles Observed During Concordance Studies

10/13 CODIS loci affected so far

Locus	STR Kits/Assays Compared	Results	Reference
VWA	PP1.1 vs ProPlus	Loss of allele 19 with **ProPlus**; fine with PP1.1	Kline et al. (1998)
D5S818	PP16 vs ProPlus	Loss of alleles 10 and 11 with **PP16**; fine with ProPlus	Alves et al. (2003)
D13S317	Identifiler vs miniplexes	Shift of alleles 10 and 11 due to deletion outside of **miniplex assay**	Butler et al. (2003), Drabek et al. (2004)
D16S539	PP1.1 vs PP16 vs COfiler	Loss of alleles with **PP1.1**; fine with PP16 and COfiler	Nelson et al. (2002)
D8S1179	PP16 vs ProPlus	Loss of alleles 15, 16, 17, and 18 with **ProPlus**; fine with PP16	Budowle et al. (2001)
FGA	PP16 vs ProPlus	Loss of allele 22 with **ProPlus**; fine with PP16	Budowle and Sprecher (2001)
D18S51	SGM vs SGM Plus	Loss of alleles 17, 18, 19, and 20 with **SGM Plus**; fine with SGM	Clayton et al. (2004)
CSF1PO	PP16 vs COfiler	Loss of allele 14 with **COfiler**; fine with PP16	Budowle et al. (2001)
TH01	PP16 vs COfiler	Loss of allele 9 with **COfiler**; fine with PP16	Budowle et al. (2001)
D21S11	PP16 vs ProPlus	Loss of allele 32.2 with **PP16**; fine with ProPlus	Budowle et al. (2001)

Recovery of Null Alleles Using Degenerate Primers

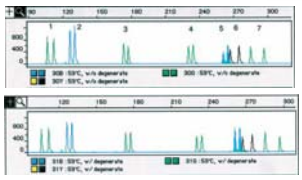
(Applied Biosystems STR Kits)

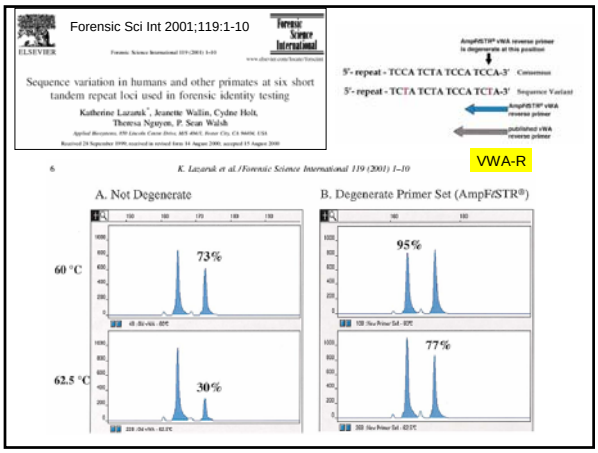
Jeanette M. Wallin,[1] M.P.H.; Cydne L. Holt,[1] Ph.D.; Katherine D. Lazaruk,[1] Ph.D.; Theresa H. Nguyen,[1] B.S.; and P. Sean Walsh,[1] M.P.H.

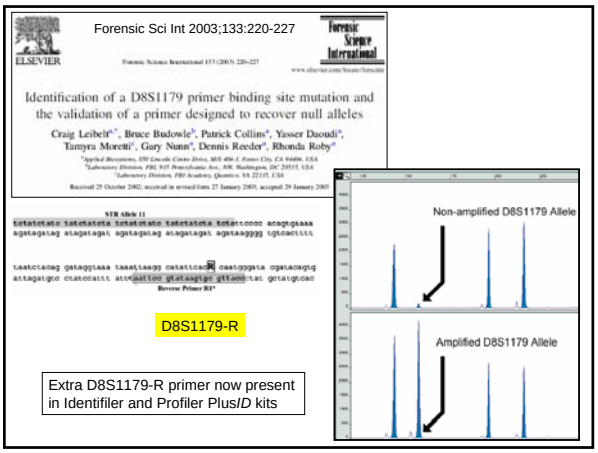
Constructing Universal Multiplex PCR Systems for Comparative Genotyping

J Forensic Sci 2002;47(1):52-65.

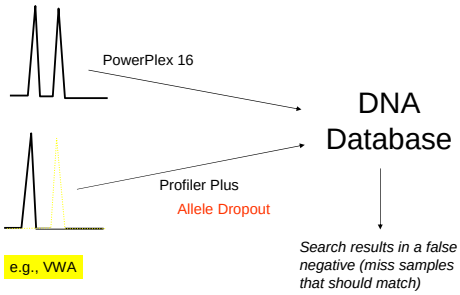
D16S539-R

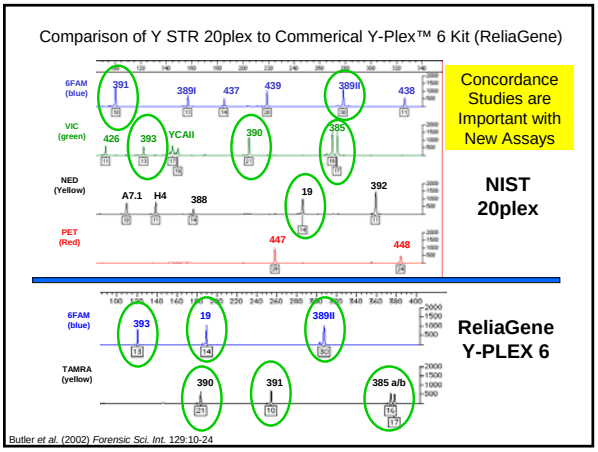




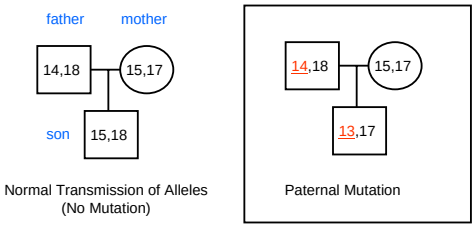


Concordance between STR primer sets is important for DNA databases





Mutation Observed in Family Trio



Butler, J.M. (2001) *Forensic DNA Typing*, Figure 6.9, ©Academic Press

STR Measured Mutation Rates http://www.cstl.nist.gov/biotech/strbase/mutation.htm

STR Locus	Maternal Meioses (%)	Paternal Meioses (%)	Either Parent	Total Mutations	Rate
13 CODIS core loci					
CSF1PO	70/179,353 (0.04)	727/504,342 (0.14)	303	1,100/683,695	0.16%
FGA	134/238,378 (0.06)	1,481/473,924 (0.31)	495	2,110/712,302	0.30%
TH01	23/189,478 (0.01)	29/346,518 (0.008)	23	75/535,996	0.01%
TPOX	16/299,186 (0.005)	43/328,067 (0.01)	24	83/627,253	0.01%
VWA	133/400,560 (0.03)	907/646,851 (0.14)	628	1,668/1,047,411	0.16%
D3S1358	37/244,484 (0.02)	429/336,208 (0.13)	266	732/580,692	0.13%
D5S818	84/316,102 (0.03)	537/468,366 (0.11)	303	924/784,468	0.12%
D7S820	43/334,886 (0.01)	550/461,457 (0.12)	218	811/796,343	0.10%
D8S1179	54/237,235 (0.02)	396/264,350 (0.15)	225	675/501,585	0.13%
D13S317	142/348,395 (0.04)	608/435,530 (0.14)	402	1,152/783,925	0.15%
D16S539	77/300,742 (0.03)	350/317,146 (0.11)	256	683/617,888	0.11%
D18S51	83/130,206 (0.06)	623/278,098 (0.22)	330	1,036/408,304	0.25%
D21S11	284/258,795 (0.11)	454/306,198 (0.15)	423	1,161/564,993	0.21%
Penta D	12/18,701 (0.06)	10/15,088 (0.07)	21	43/33,789	0.13%
Penta E	22/39,121 (0.06)	58/44,152 (0.13)	55	135/83,273	0.16%
D2S1338	2/25,271 (0.008)	61/81,960 (0.07)	31	94/107,231	0.09%
D19S433	22/28,027 (0.08)	16/38,983 (0.04)	37	75/67,010	0.11%
F13A01	1/10,474 (0.01)	37/65,347 (0.06)	3	41/75,821	0.05%
FES/FPS	3/18,918 (0.02)	79/149,028 (0.05)	None reported	82/167,946	0.05%
F13B	2/13,157 (0.02)	8/27,183 (0.03)	1	11/40,340	0.03%
LPL	0/8,821 (<0.01)	9/16,943 (0.05)	4	13/25,764	0.05%
SE33 (ACTBP2)	0/330 (<0.30)	330/51,610 (0.64)	None reported	330/51,940	0.64%

*Data used with permission from American Association of Blood Banks (AABB) 2002 Annual Report.

Summary of STR Mutations

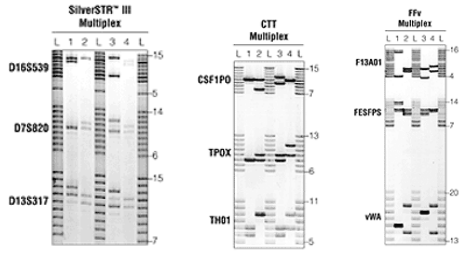
- Mutations happen and need to be considered
- Usually 1 in ~1000 meioses
- Paternal normally higher than maternal
- VWA, FGA, and D18S51 have highest levels
- TH01, TPOX, and D16S539 have lowest levels

Mutations impact paternity testing and missing persons investigations but not forensic direct evidence-suspect matches…

Advantages for STR Markers

- Small product sizes are generally compatible with degraded DNA and PCR enables recovery of information from small amounts of material
- Multiplex amplification with fluorescence detection enables high power of discrimination in a single test
- Commercially available in an easy to use kit format
- Uniform set of core STR loci provide capability for national and international sharing of criminal DNA profiles

STR Typing with Silver Stain Gels



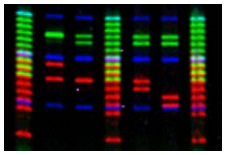
http://www.promega.com/applications/hmnid/productprofiles/geneprint/gp_sstr/silverstr.htm

Advantages: Inexpensive, requires no expensive equipment purchase

Disadvantages: Labor intensive, not automated, cannot analyze large multiplex (only a single "color" channel is available)

Multi-Color Fluorescence Detection

Gel Image of STRs
(Multiple Samples and Ladders)



Advantages: Detection can be automated and large multiplexes can be analyzed due to multi-channel detection that enables overlapping PCR product sizes with different dye labels

Disadvantages: Expensive equipment required, fluorescent artifacts can interfere with interpretation

CE Electropherogram (Single Sample with Multiple STRs)