A Primer on DNA Profiling Using STR Markers

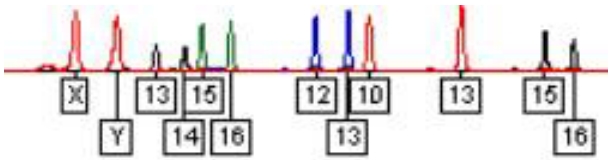
Dr. John M. Butler

National Institute of Standards and Technology

Biotechnology Division

Presentation Outline

- Biology and technology behind short tandem repeat (STR) DNA testing
- How statistical calculations are made with STRs
- Approaches for “challenging” samples: perspectives for the future
- Resources for additional information



Dr. John M. Butler, Project Leader, Biotechnology Division, Human Identity DNA Technologies Group National Institute of Standards and Technology

Human Identity Testing

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

Involves generation of DNA profiles usually with the same core STR (short tandem repeat) markers

Dr. John M. Butler, Project Leader, Biotechnology Division, Human Identity DNA Technologies Group National Institute of Standards and Technology

Basis of DNA Profiling

The genome of **each individual is unique** (with the exception of identical twins) and **is inherited from parents**

Probe subsets of genetic variation in order to differentiate between individuals (statistical probabilities of a random match are used)

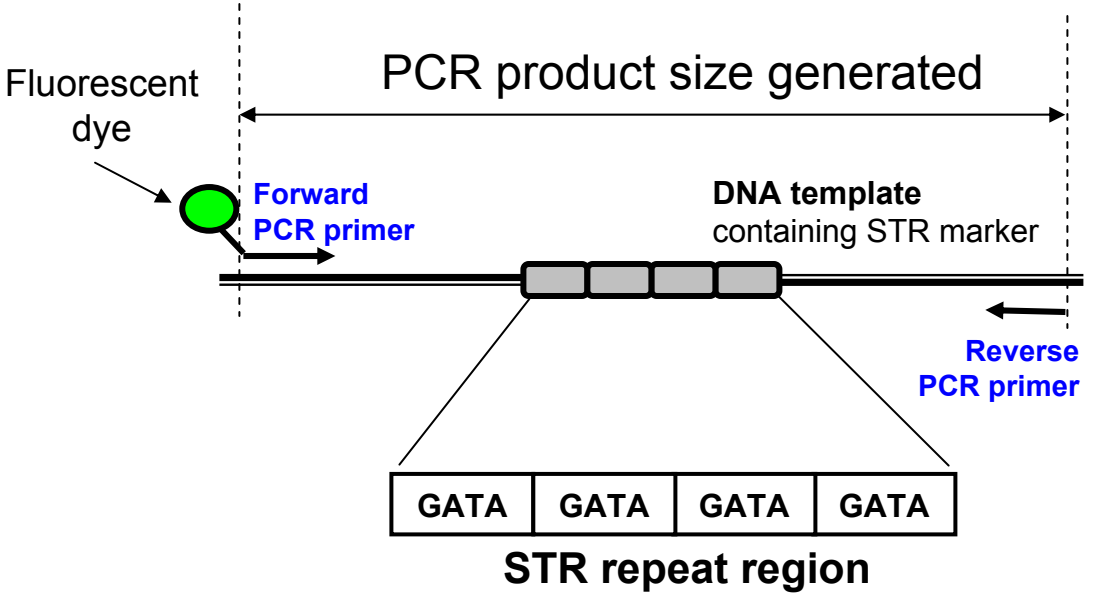
DNA typing must be **performed efficiently and reproducibly** (information must hold up in court)

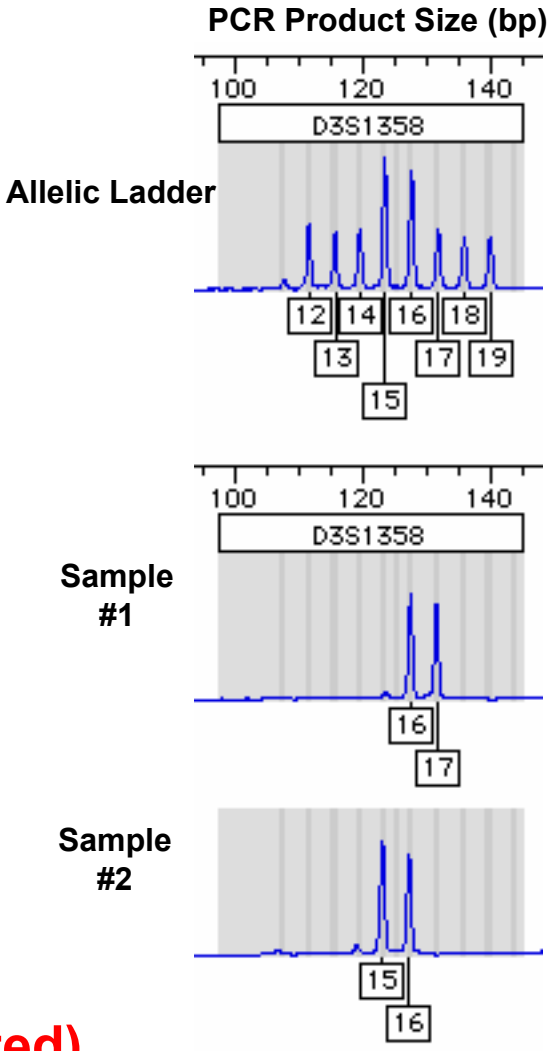
Current standard DNA tests **DO NOT look at genes** – little/no information about race, predisposal to disease, or phenotypical information (eye color, height, hair color) is obtained

Dr. John M. Butler, Project Leader, Biotechnology Division, Human Identity DNA Technologies Group National Institute of Standards and Technology

Short Tandem Repeat (STR) Markers

PCR primers anneal to unique sequences bracketing the variable STR repeat region





TCCCAAGCTCTTCCTCTTCCCTAGATCAATACAGACAGA
AGACAGGTG**GATAGATAGATAGATAGATAGATAGATA**
GATAGATAGATAGATATCATTGAAAGACAAAACAGAGA
TGGATGATAGATACATGCTTACAGATGCACAC

= 11 GATA repeats ("11" is all that is reported)

Advantages for STR Markers

- Use of the polymerase chain reaction (PCR) enables recovery of information from small amounts of material
- Small product sizes are generally compatible with degraded DNA
- 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

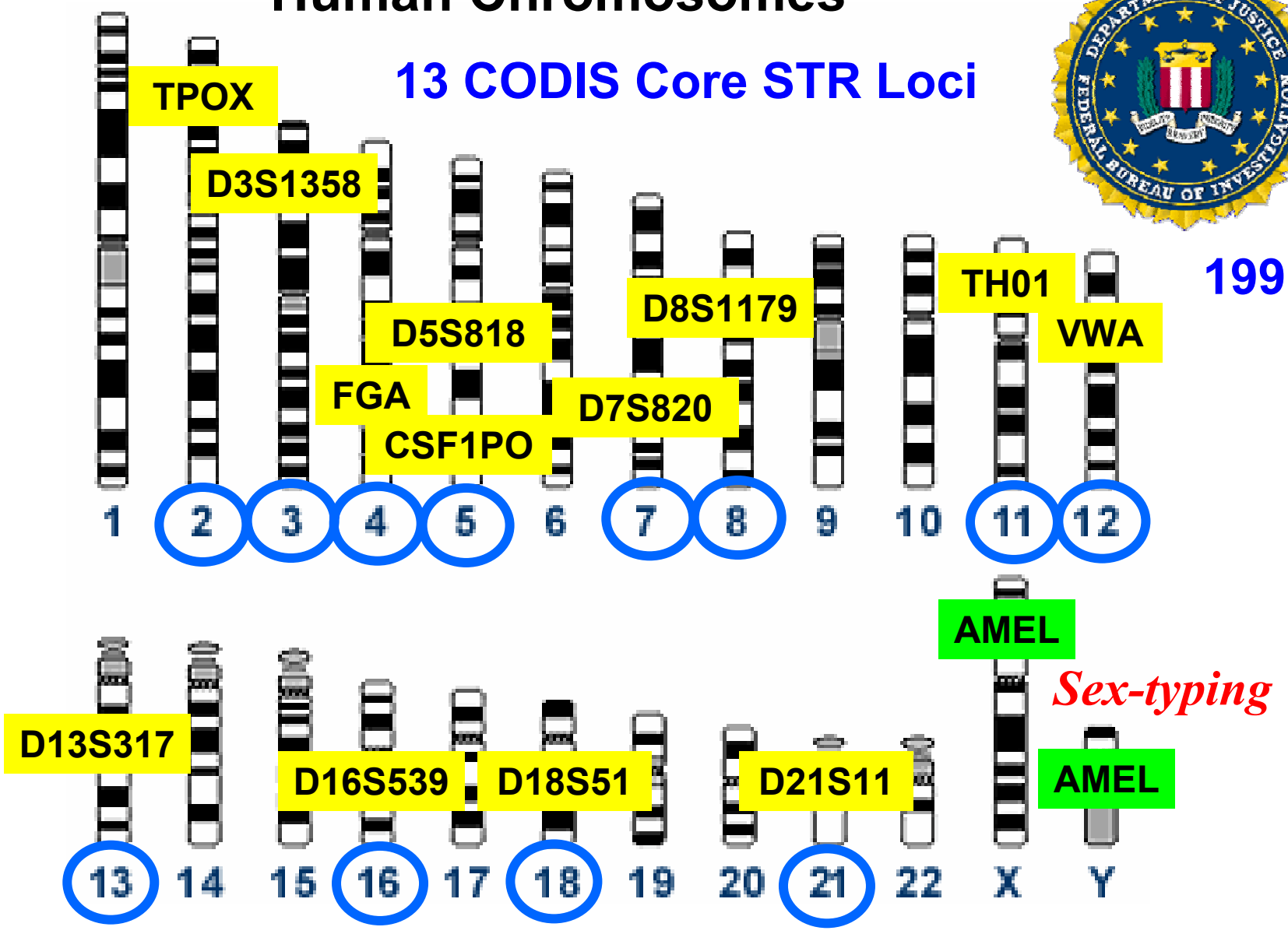
Dr. John M. Butler, Project Leader, Biotechnology Division, Human Identity DNA Technologies Group National Institute of Standards and Technology

Position of Forensic STR Markers on Human Chromosomes

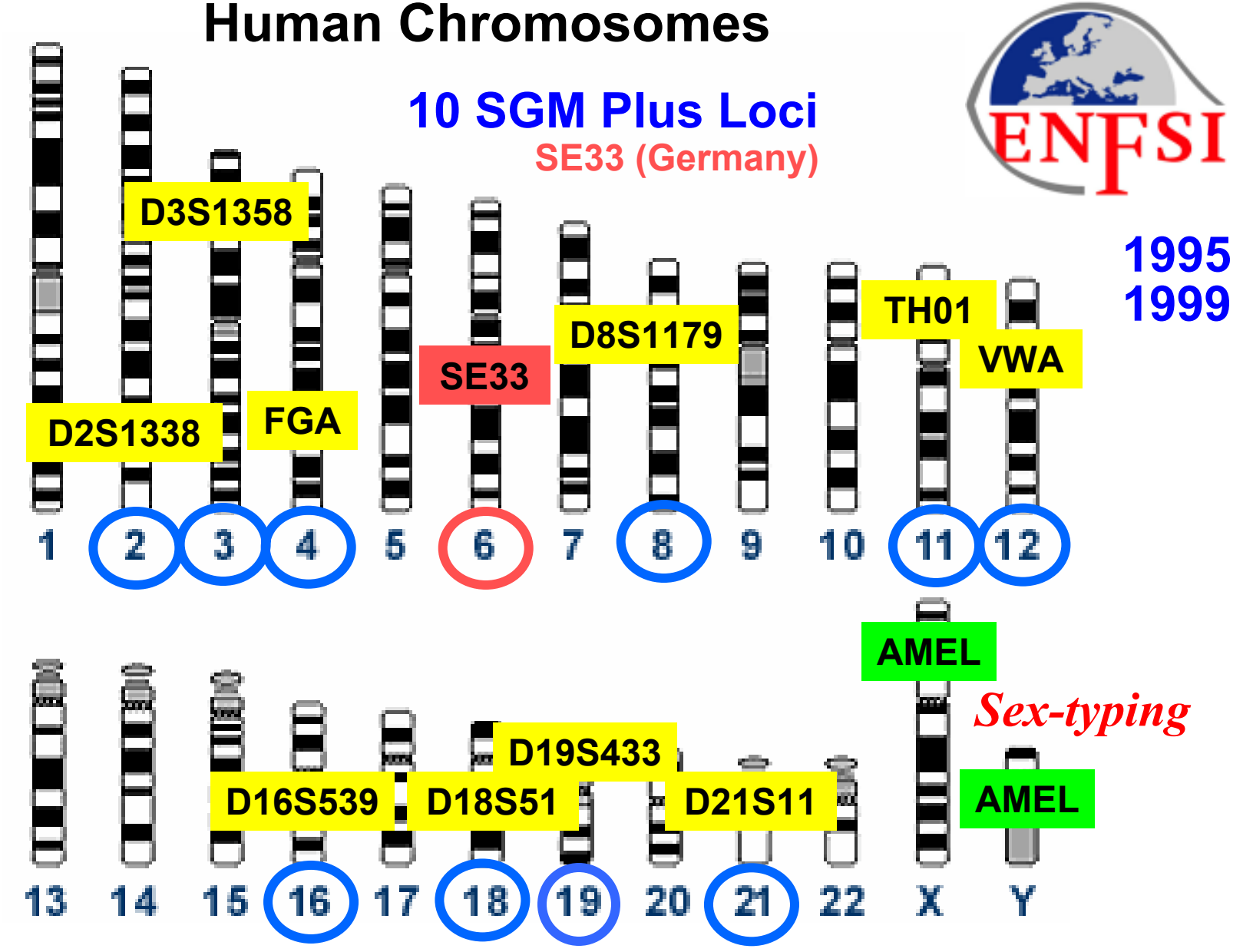
Core STR Loci for the United States

13 CODIS Core STR Loci

1997



Position of Forensic STR Markers on Human Chromosomes



Typical Instruments Used for STR Typing

GeneAmp 9700

Thermal Cycler for PCR Amplification

Capillary electrophoresis instruments for separating and sizing PCR products

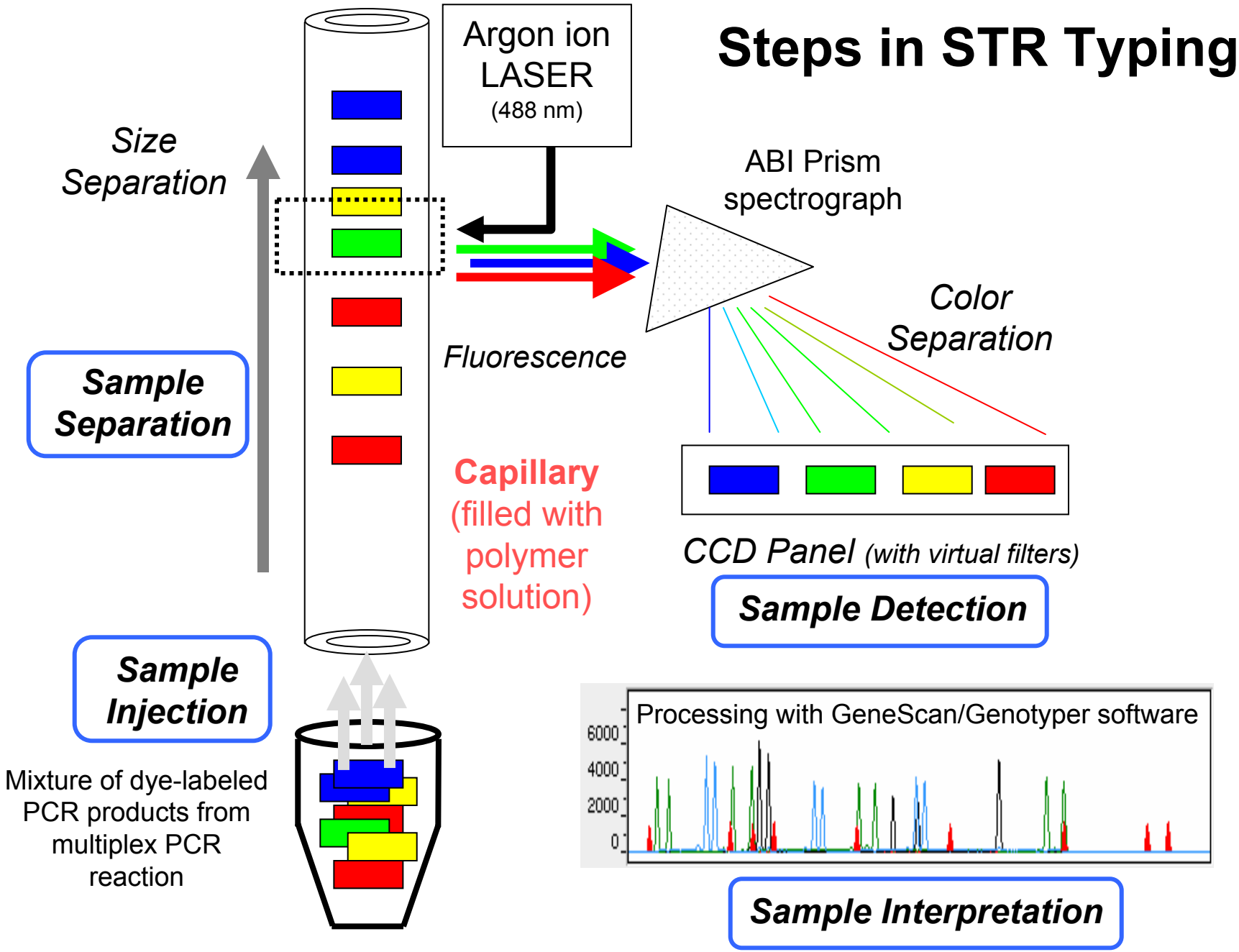
single capillary

ABI 310

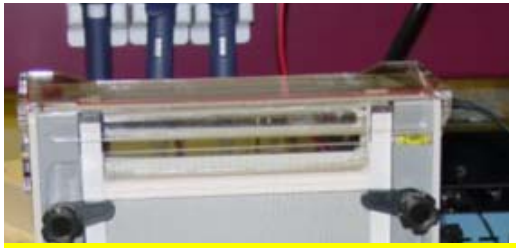
16-capillary array

ABI 3100

Steps in STR Typing



FMBIO III Gel Imager System

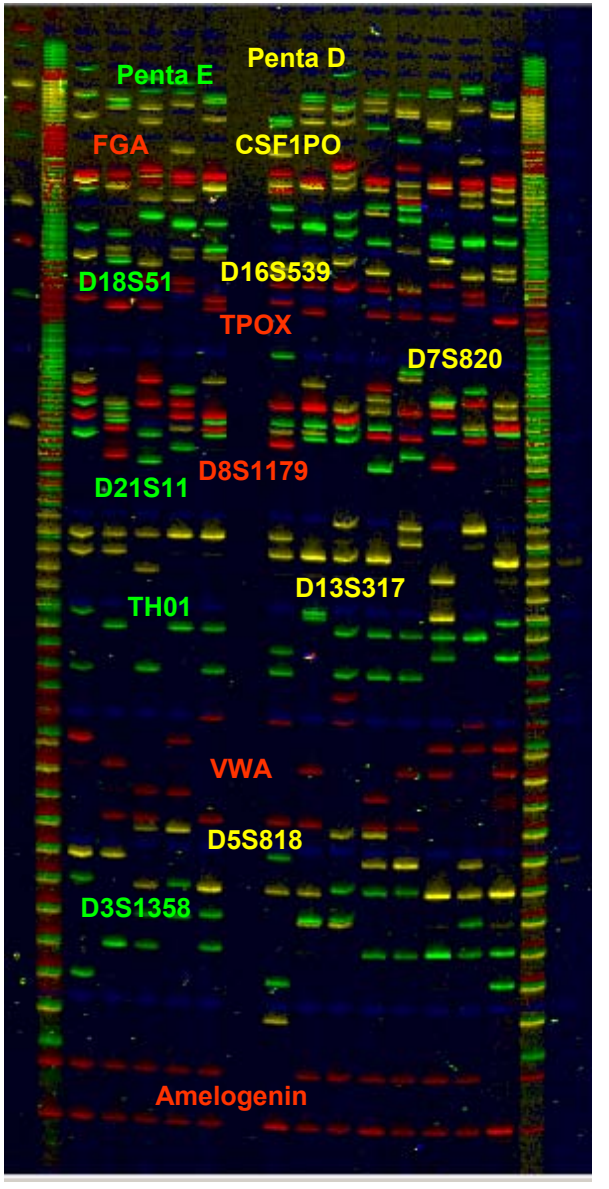


Gel Electrophoresis

HITACHI FMBIO III Multi-View

Gel Scanner

Gel Image

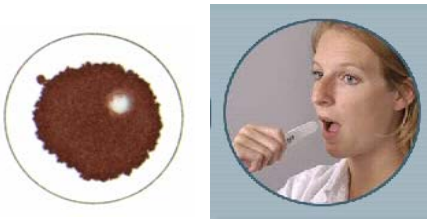


PowerPlex 16 BIO

Steps in DNA Analysis

Usually 1-2 day process (a minimum of ~5 hours)

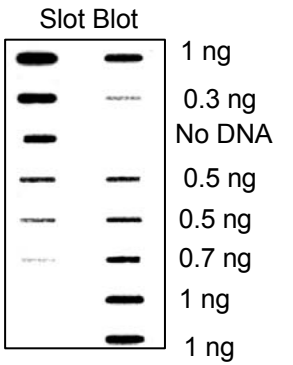
- Collection
- Specimen Storage
- Extraction
- Quantitation
- Genotyping
- Interpretation of Results
- Database
  Storage & Searching



Blood Stain Buccal swab

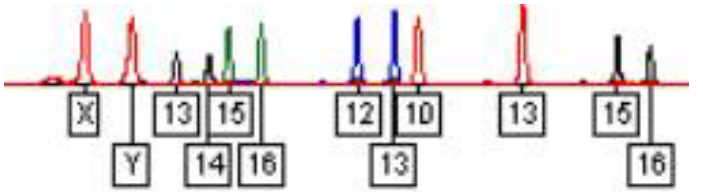
Sample Collection & Storage

DNA Extraction



DNA Quantitation

Multiplex PCR
(**Amplification of STR Loci**)



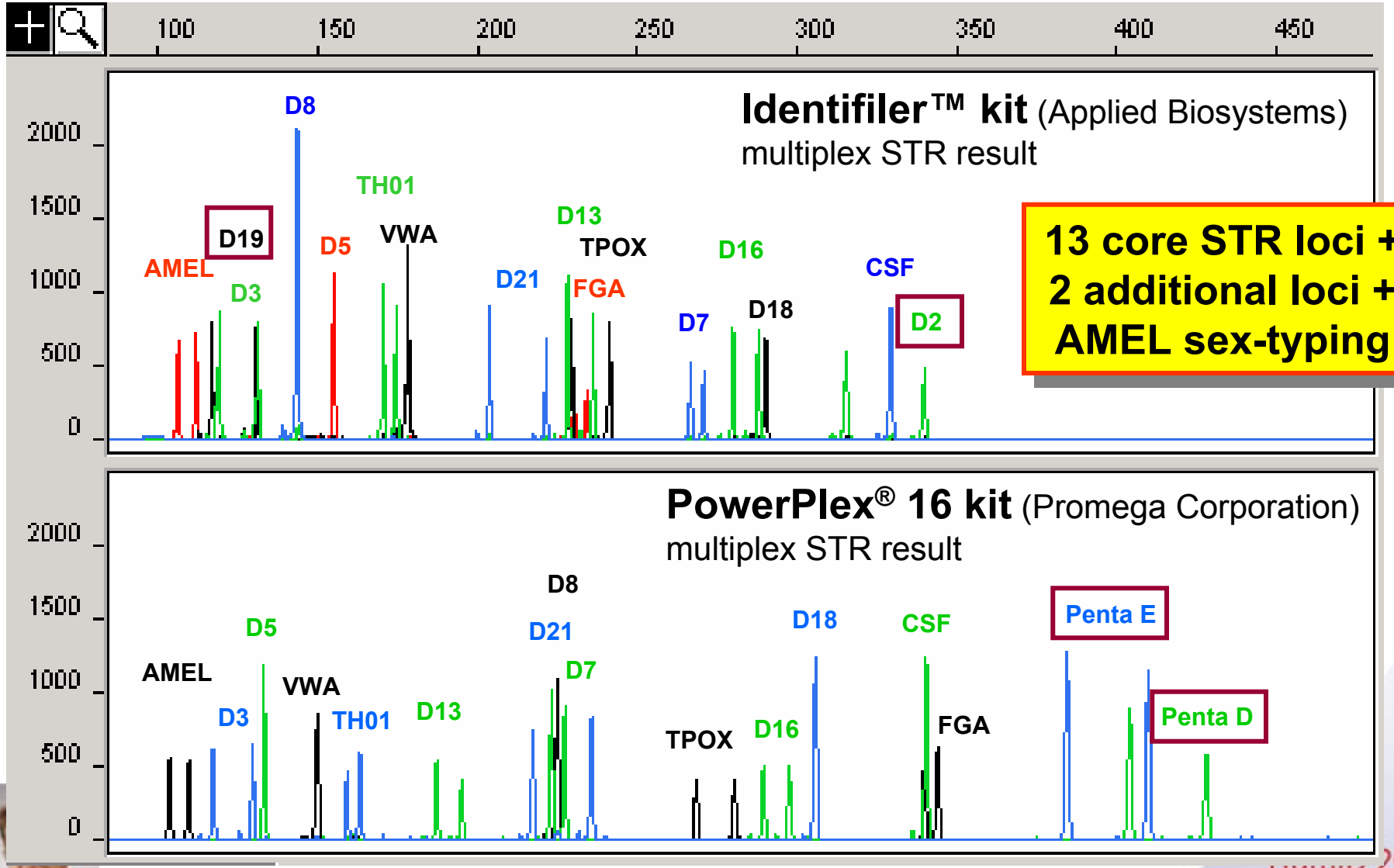
STR Typing (**DNA separation**)

Male: 13,14-15,16-12,13-10,13-15,16

Interpretation of Results

DNA Database

Commercial STR 16plex Kits



Dr. John M. Butler, Project Leader, Biotechnology Division, Human Identity DNA Technologies Group National Institute of Standards and Technology

SRM 2391b component 1

How Statistical Calculations are Made

- Generate data with set(s) of samples from desired population group(s)
  - – Generally only 100-150 samples are needed to obtain reliable allele frequency estimates

- Determine allele frequencies at each locus
  - Count number of each allele seen

- Allele frequency information is used to estimate the rarity of a particular DNA profile
  - Homozygotes (p^2), Heterozygotes (2pq)
  - Product rule used (multiply locus frequency estimates)

For more information, see Chapters 20 and 21 in *Forensic DNA Typing*, 2nd Edition

Dr. John M. Butler, Project Leader, Biotechnology Division, Human Identity DNA Technologies Group National Institute of Standards and Technology

Individual Genotypes Are Summarized and Converted into Allele Frequencies

Genotype Array	8	9	10	11	12	13	14	15		Allele Count	Observed Frequency
	8,8	8,9	8,10	8,11	8,12	8,13	8,14	8,15			
8	9	9	1	17	13	10	0	0	**8**	68	0.11258
		9,9	9,10	9,11	9,12	9,13	9,14	9,15			
9		1	2	15	10	4	3	0	**9**	45	0.07450
			10,10	10,11	10,12	10,13	10,14	10,15			
10			2	12	6	3	2	1	**10**	31	0.05132
				11,11	11,12	11,13	11,14	11,15			
11				37	54	21	12	0	**11**	205	0.33940
					12,12	12,13	12,14	12,15			
12					21	18	7	0	**12**	150	0.24834
						13,13	13,14	13,15			
13						7	5	0	**13**	75	0.12417
							14,14	14,15			
14							0	0	**14**	29	0.04801
								15,15			
15								0	**15**	**1**	**0.00166**
										604	

The **11,12** genotype was seen **54 times** in 302 samples (604 examined chromosomes)

Allele Frequency Tables

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.

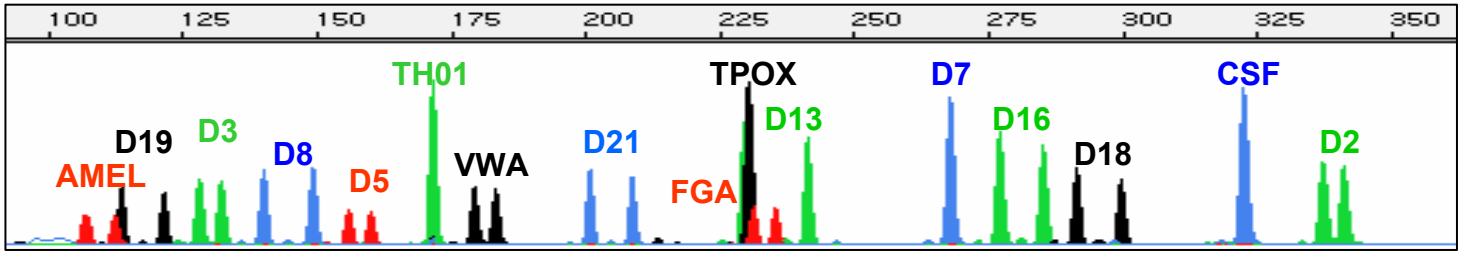
D3S1358

Most common allele

Allele	Butler *et al*. (2003) *JFS* 48(4):908-911 **Caucasian** *N= 302*	Einum *et al*. (2004) *JFS* 49(6) **Caucasian** ***N= 7,636***	Allele	**African American** *N=258*	**African American** ***N= 7,602***
11	0.0017*	0.0009	**11**	--	0.0003*
12	0.0017*	0.0007	**12**	--	0.0045
13	--	0.0031	**13**	0.0019*	0.0077
14	0.1027	0.1240	**14**	0.0892	0.0905
15	**0.2616**	**0.2690**	**15**	0.3023	0.2920
15.2	--	--	**15.2**	0.0019*	0.0010
16	0.2533	0.2430	**16**	**0.3353**	**0.3300**
17	0.2152	0.2000	**17**	0.2054	0.2070
18	0.15232	0.1460	**18**	0.0601	0.0630
19	0.01160	0.0125	**19**	0.0039*	0.0048
20	0.0017*	0.0001*	**20**		

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-
9,11-
6,6-
8,8-
10,10

Locus	allele	value	allele	value	1 in	Combined
D3S1358	**16**	*0.2533*	**17**	*0.2152*	**9.17**	**9.17**
VWA	**17**	*0.2815*	**18**	*0.2003*	**8.87**	**81**
FGA	**21**	*0.1854*	**22**	*0.2185*	**12.35**	**1005**
D8S1179	**12**	*0.1854*	**14**	*0.1656*	**16.29**	**16,364**
D21S11	**28**	*0.1589*	**30**	*0.2782*	**11.31**	**185,073**
D18S51	**14**	*0.1374*	**16**	*0.1391*	**26.18**	**4,845,217**
D5S818	**12**	*0.3841*	**13**	*0.1407*	**9.25**	**44,818,259**
D13S317	**11**	*0.3394*	**14**	*0.0480*	**30.69**	**1.38×10^{9}**
D7S820	**9**	*0.1772*			**31.85**	**4.38×10^{10}**
D16S539	**9**	*0.1126*	**11**	*0.3212*	**13.8**	**6.05×10^{11}**
THO1	**6**	*0.2318*			**18.62**	**1.13×10^{13}**
TPOX	**8**	*0.5348*			**3.50**	**3.94×10^{13}**
CSF1PO	**10**	*0.2169*			**21.28**	**8.37×10^{14}**

PRODUCT RULE

The Random Match Probability for this profile in the U.S. Caucasian population is **1 in 837 trillion (10^{12})**

The Same 13 Locus STR Profile in Different Populations

1 in 837 trillion

1 in 0.84 quadrillion (10^{15}) in U.S. Caucasian population (NIST)
1 in 2.46 quadrillion (10^{15}) in U.S. Caucasian population (FBI)*
1 in 1.86 quadrillion (10^{15}) in Canadian Caucasian population*

1 in 16.6 quadrillion (10^{15}) in African American population (NIST)
1 in 17.6 quadrillion (10^{15}) in African American population (FBI)*

1 in 18.0 quadrillion (10^{15}) in U.S. Hispanic population (NIST)

These values are **for unrelated individuals**
assuming no population substructure (using only p^2 and 2 pq)

NIST study: 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. (http://www.cstl.nist.gov/biotech/strbase/NISTpop.htm)

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

Approaches for "challenging" samples: perspectives for the future

- Limited sample material (highly degraded DNA)
  - **mtDNA** (in use for this purpose since mid-1990s due to high copy number per cell)

Chapter 10 in *Forensic DNA Typing*, 2nd Edition

- Mixed male-female DNA
  - **Y-chromosome STRs**

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

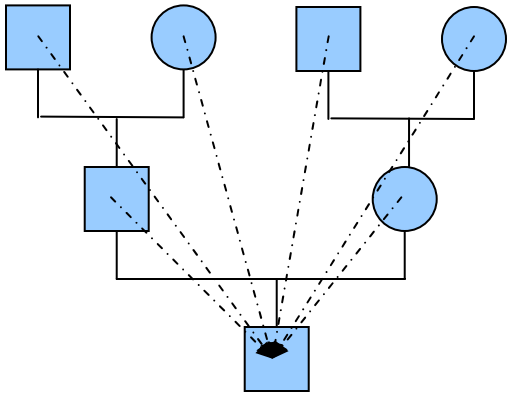
- Degraded DNA
  - **miniSTRs** **http://www.cstl.nist.gov/biotech/strbase/miniSTR.htm**
  - SNPs (?) **http://www.cstl.nist.gov/biotech/strbase/SNP.htm**

Dr. John M. Butler, Project Leader, Biotechnology Division, Human Identity DNA Technologies Group National Institute of Standards and Technology

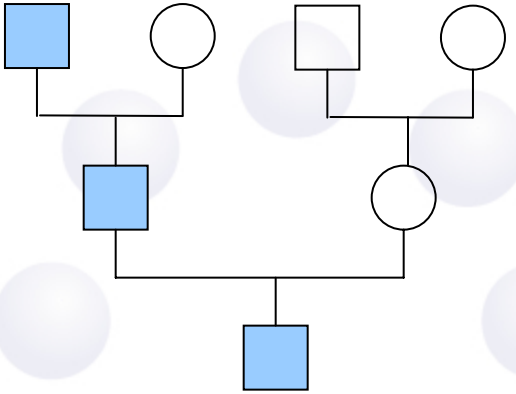
Different Inheritance Patterns

Lineage Markers

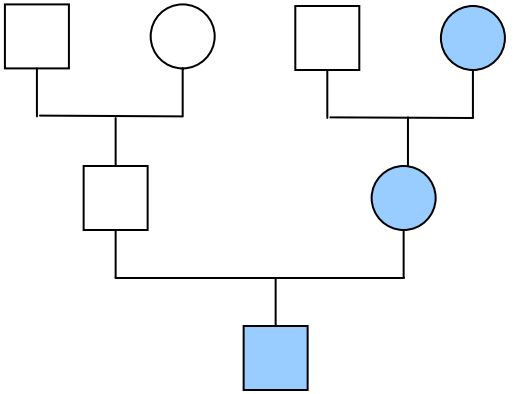
CODIS STR Loci



Autosomal
(**passed on in part**,
from all ancestors)



Y-Chromosome
(**passed on complete**,
but only by sons)

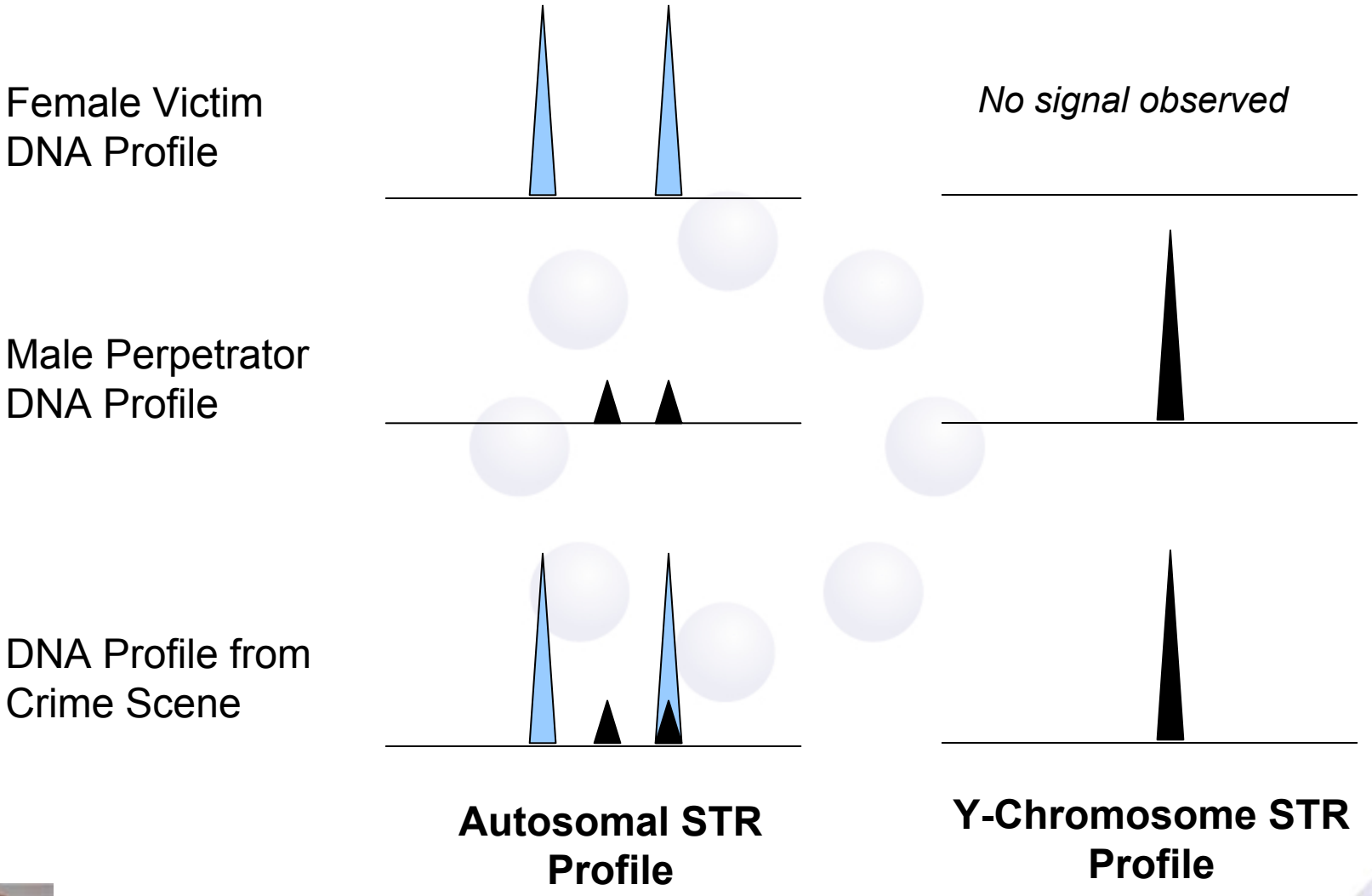


Mitochondrial
(**passed on complete**,
but only by daughters)

Butler, J.M. (2005) *Forensic DNA Typing, 2nd Edition,* Figure 9.1, ©Elsevier Science/Academic Press

Dr. John M. Butler, Project Leader, Biotechnology Division, Human Identity DNA Technologies Group National Institute of Standards and Technology

Y-STRs can permit simplification of male DNA identification in sexual assault cases



Butler, J.M. (2005) *Forensic DNA Typing, 2nd Edition,* Figure 9.2, ©Elsevier Science/Academic Press

Dr. John M. Butler, Project Leader, Biotechnology Division, Human Identity DNA Technologies Group National Institute of Standards and Technology

Y-Chromosome Haplotype Reference Database (YHRD)

http://www.yhrd.org

As of 12/17/04: 28,650 haplotypes

6,281 haplotypes
with all US required loci

Commercial Y-STR kits exist to amplify all of the core loci in a single reaction (plus a few additional markers)

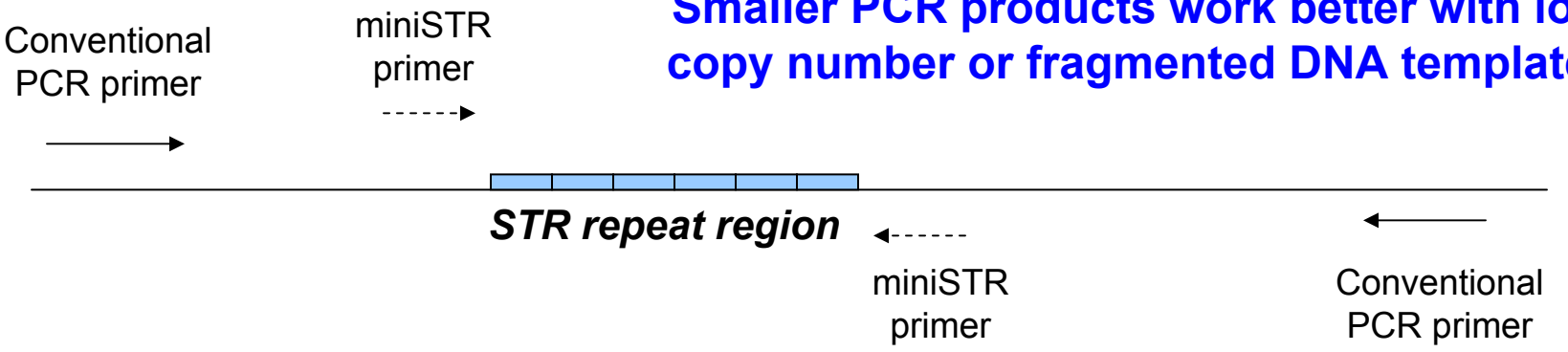
Run only with minimal haplotype

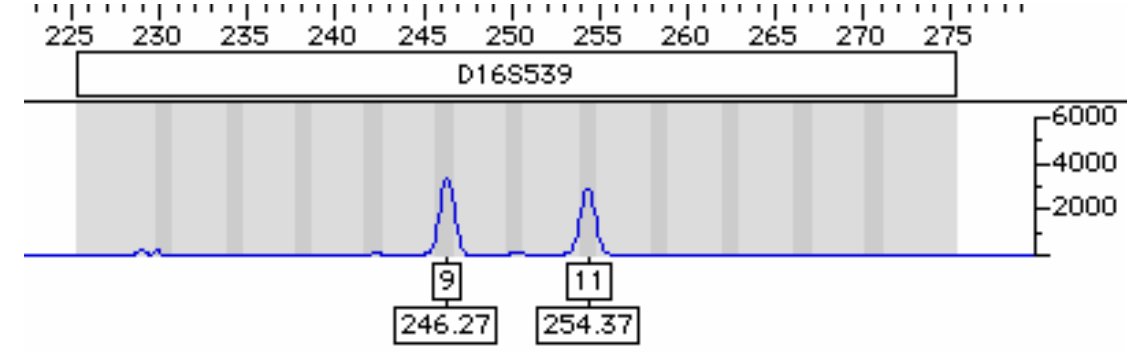
DYS19
DYS389I/II
DYS390
DYS391
DYS392
DYS393
DYS385 a/b

US haplotype requires 2 additional loci:
DYS438
DYS439

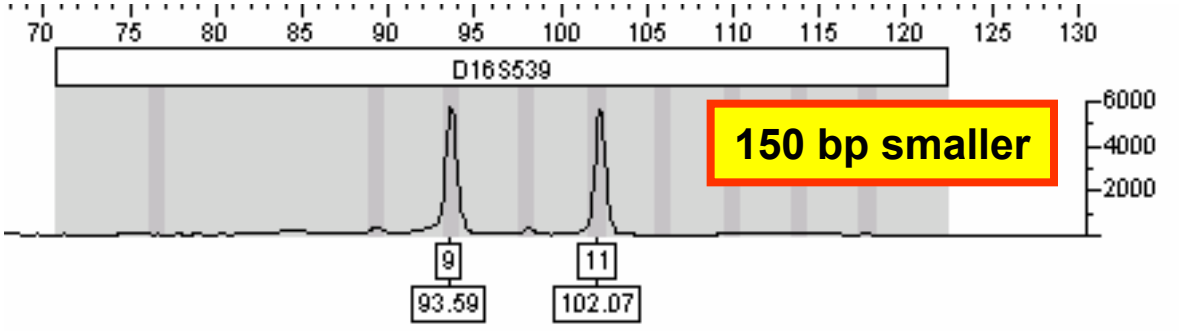
miniSTRs: new tool for degraded DNA

Smaller PCR products work better with low copy number or fragmented DNA templates



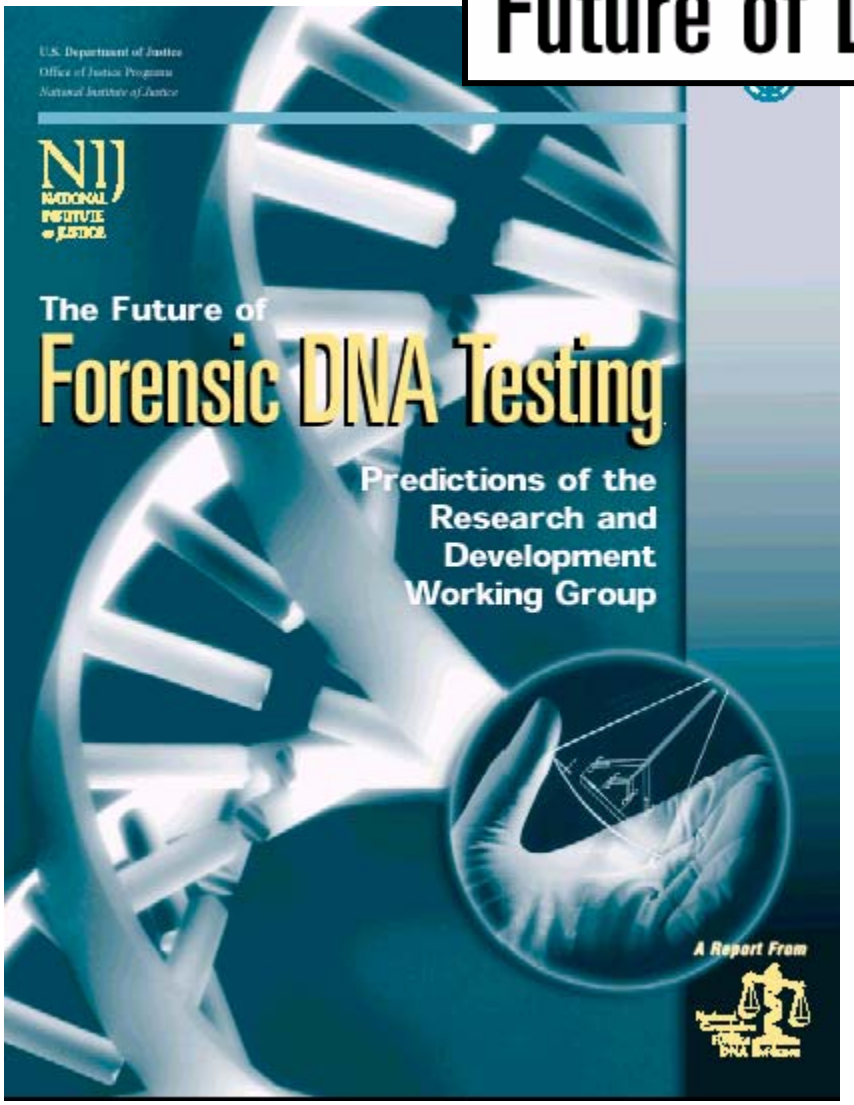


Conventional STR test (COfiler™ kit)



MiniSTR assay (using Butler *et al.* 2003 primers)

National Commission on the Future of DNA Evidence



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

- Report published in Nov 2000

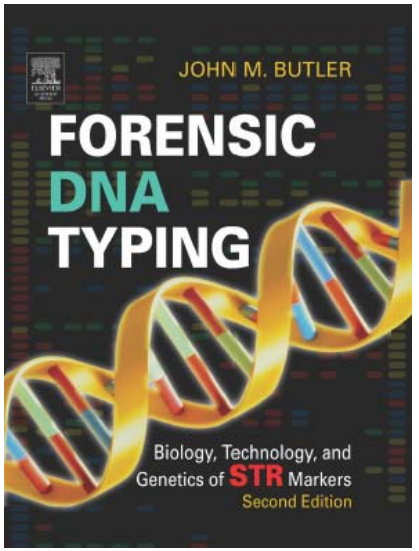
- Asked to estimate where DNA testing would be 2, 5, and 10 years into the future

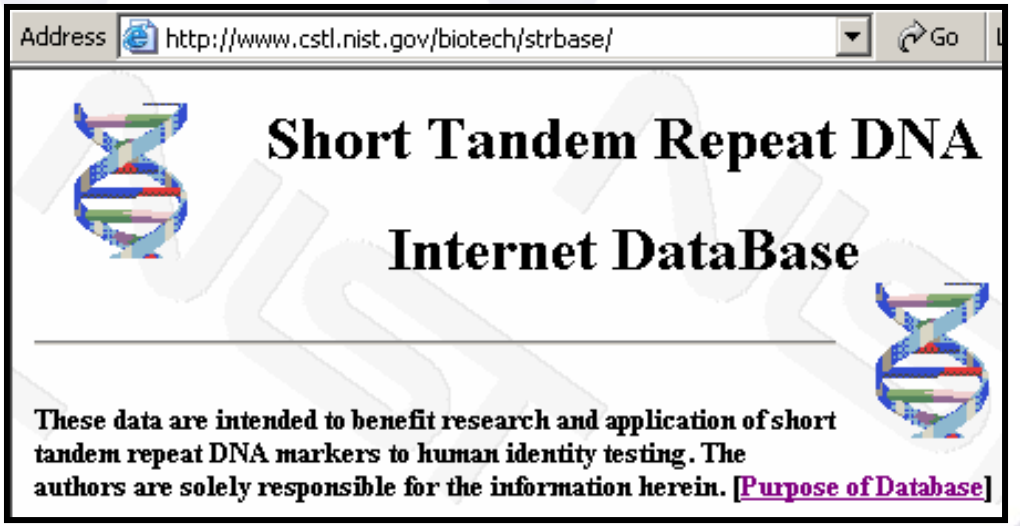
Conclusions

STR typing is here to stay for a few years because of DNA databases that have grown to contain millions of profiles

Additional Resources

- ***Forensic DNA Typing: Biology,Technology, and Genetics of STR Markers (2nd Edition)*** **by John M. Butler, Elsevier Academic Press, 2005 – please go to the Human Identification E-Symposium Delegate Zone, for more information**
- **Butler, J.M., *et al.* (2004) Forensic DNA typing by capillary electrophoresis using the ABI Prism 310 and 3100 genetic analyzers for STR analysis. *Electrophoresis* 25: 1397-1412.**
- **NIST website: http://www.cstl.nist.gov/biotech/strbase**





Dr. John M. Butler, Project Leader, Biotechnology Division, Human Identity DNA Technologies Group National Institute of Standards and Technology

Content of STRBase Website

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

- …/str_fact.htm — STR Fact Sheets on Core Loci
- …/multiplx.htm — Multiplex STR Kit Information
- …/y_strs.htm — Y-Chromosome Information
- …/var_tab.htm — Variant Alleles Reported
- …/mutation.htm — Mutation Rates for Common STRs
- …/str_ref.htm — Reference List with ~2,300 Papers
- …/training.htm — Downloadable PowerPoints for Training
- …/validation.htm — Validation Information
- …/miniSTR.htm — miniSTR Information
- …/address.htm — Addresses for Scientists
- …/NISTpub.htm — Publications & Presentations from NIST

Dr. John M. Butler, Project Leader, Biotechnology Division, Human Identity DNA Technologies Group National Institute of Standards and Technology

Summary of Key Points

- **STRs are highly variable genetic markers**
- **Core STR loci have been chosen to enable a common currency for use in national DNA databases**
- **STR kits permit co-amplification of up to 15 STRs plus amelogenin for sex-typing**
- **Capillary electrophoresis with fluorescence detection has become the method of choice for STR typing**
- **STR allele frequencies are used to estimate the rarity of a particular STR profile**
- **The core STR markers of today should remain in widespread use due to the millions of profiles already in DNA databases**
- **Y-STRs and miniSTRs will likely play a growing role in the future**

Dr. John M. Butler, Project Leader, Biotechnology Division, Human Identity DNA Technologies Group National Institute of Standards and Technology

Questions

Your turn, any questions?

>> Click on the Q&A tab, type your name & question, hit send and I will answer it live now!

John Butler

john.butler@nist.gov

301-975-4049

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

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

Dr. John M. Butler, Project Leader, Biotechnology Division, Human Identity DNA Technologies Group National Institute of Standards and Technology