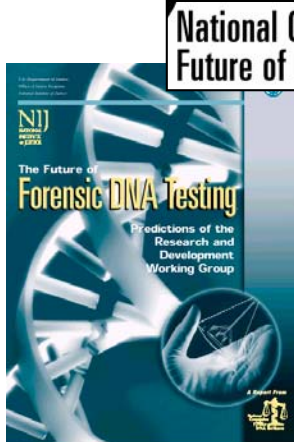


## Introduction



**National Commission on the Future of DNA Evidence**

NIJ  
The Future of Forensic DNA Testing  
Predictions of the Research and Development Working Group

• 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

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

## 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

## Basis of DNA Profiling

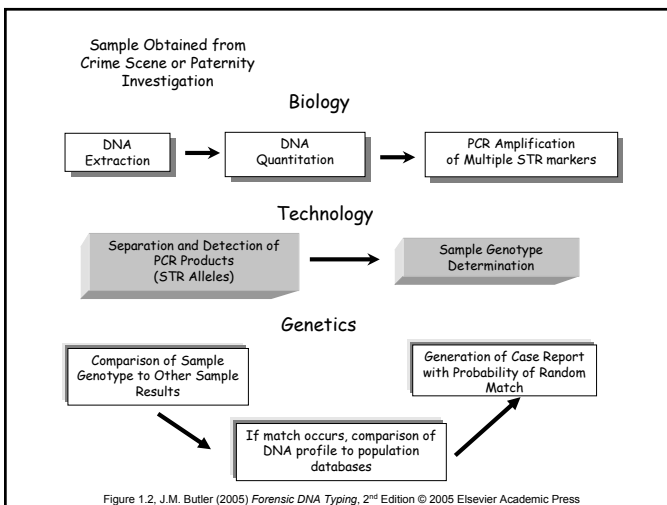
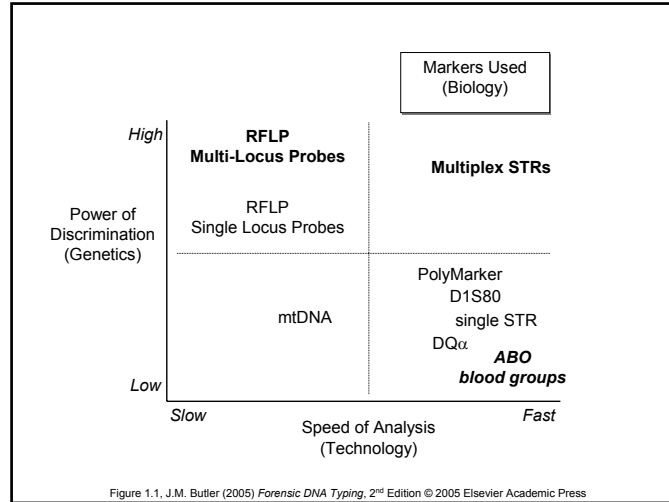
The genome of **each individual is unique** (with the exception of identical twins) and **is inherited from parents**

**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

# Overview and History of DNA Typing



# DNA Biology

Deoxyribonucleic acid consists of a sugar backbone attached to one of 4 different bases. The sugars are linked to each other by phosphates and the bases of each strand are hydrogen bonded to each other.

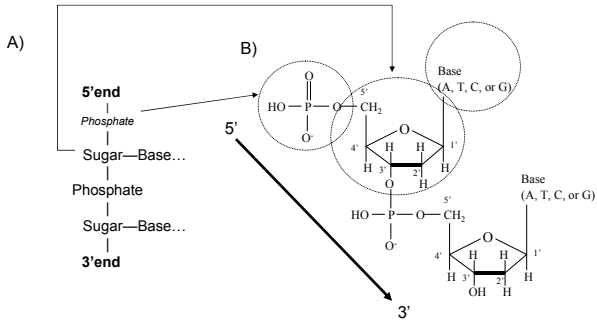


Figure 2.1, J.M. Butler (2005) *Forensic DNA Typing*, 2<sup>nd</sup> Edition © 2005 Elsevier Academic Press

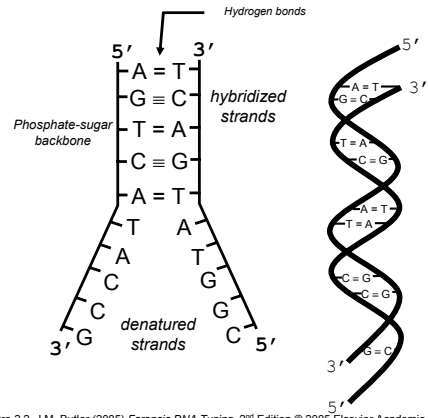


Figure 2.2, J.M. Butler (2005) *Forensic DNA Typing*, 2<sup>nd</sup> Edition © 2005 Elsevier Academic Press

### Human Genome

23 Pairs of Chromosomes + mtDNA

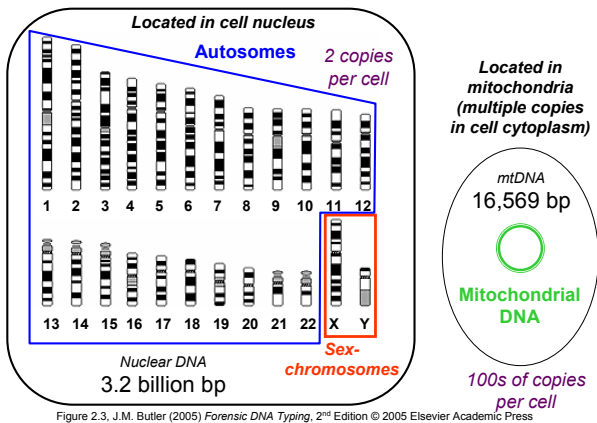


Figure 2.3, J.M. Butler (2005) *Forensic DNA Typing*, 2<sup>nd</sup> Edition © 2005 Elsevier Academic Press

### Chromosome 12

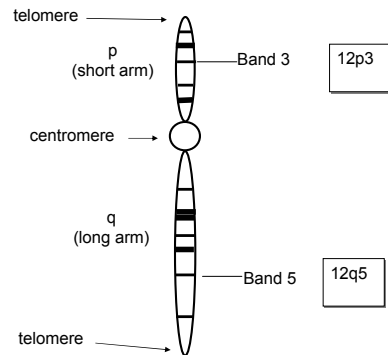


Figure 2.4, J.M. Butler (2005) *Forensic DNA Typing*, 2<sup>nd</sup> Edition © 2005 Elsevier Academic Press

## Genes

- DNA within a chromosome is organized into genes, sequences that code for proteins (**exons**) and non-coding regions called **introns**.
- There are genes for about 30,000 proteins.
- Genes make up just 5% of the human genome.
- Markers used for DNA testing are located in non-coding introns.
- These regions differ between individuals and are termed polymorphic regions or loci.

- A pair of chromosomes is called homologous because they are the same in size and structure.
- One chromosome was inherited from mother and the other from the father.
- The DNA sequences for each pair of chromosomes, however may not be the same.
- Alternative possibilities for a gene or genetic locus sequence is called an allele.
- Two alleles that are the same are homozygous. Two that are different are heterozygous.
- DNA identity testing is based on the ability to detect differences between alleles at various genetic loci.

- **HUMAN GENOME**
- - genome: total (haploid) genetic makeup of an organism
- - human genome contains 3X10<sup>9</sup> bp
- -- 4% is coding DNA
- - structural genes
- - regulatory genes
- -- 96% is noncoding DNA
- - regulatory regions
- - promoters, upstream regulatory sequences, response elements, enhancers
- - repetitive sequences
- - dispersed
- - tandem

- **FORENSIC DNA TYPING SYSTEMS**
- **Length polymorphisms**
- - Restriction Fragment Length Polymorphisms (**RFLP**)
- - analysis of variable number of tandem repeats (**VNTR**)
- - Polymerase Chain Reaction- Short Tandem Repeats (**PCR-STR**)
- **Sequence polymorphisms**
- - AMPLITYPE® PM (polymarker)
- - analysis of several genes
- - sequence specific oligonucleotide (**SSO**) probes
- - mitochondrial DNA (mtDNA)
- - sequencing of two hypervariable regions (**HV1** & **HV2**)
- within the D- loop

(A) Sequence polymorphism

-----AGACTAGACATT-----  
-----AGATTAGGCATT-----

(B) Length polymorphism

-----**(AATG)**(AATG)(AATG)-----  
          3 repeats  
-----**(AATG)**(AATG)-----  
          2 repeats

Figure 2.5, J.M. Butler (2005) *Forensic DNA Typing*, 2<sup>nd</sup> Edition © 2005 Elsevier Academic Press

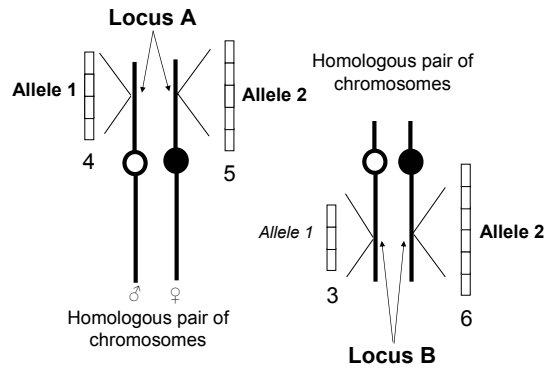
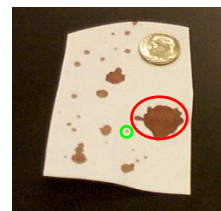


Figure 2.6, J.M. Butler (2005) *Forensic DNA Typing*, 2<sup>nd</sup> Edition © 2005 Elsevier Academic Press

Sample Collection, Extraction and  
Quantitation

Sources of Biological Evidence

- Blood
- Semen
- Saliva
- Urine
- Hair
- Teeth
- Bone
- Tissue



Blood stain

Only a very small amount of blood is needed to obtain a DNA profile

**FORENSIC DNA ANALYSIS**

**PROCEDURES**

**I. DNA isolation**

- **- Chelex extraction**
  - differential extraction for sperm/vaginal epithelial cell mixtures
  - yields ssDNA suitable for PCR amplification
- **- organic extraction**
  - yields large fragments of dsDNA suitable for RFLP analysis
  - purified DNA
- **silica column**
  - Qiamp
  - Purified DNA
  - ds DNA suitable for PCR and/or RFLP

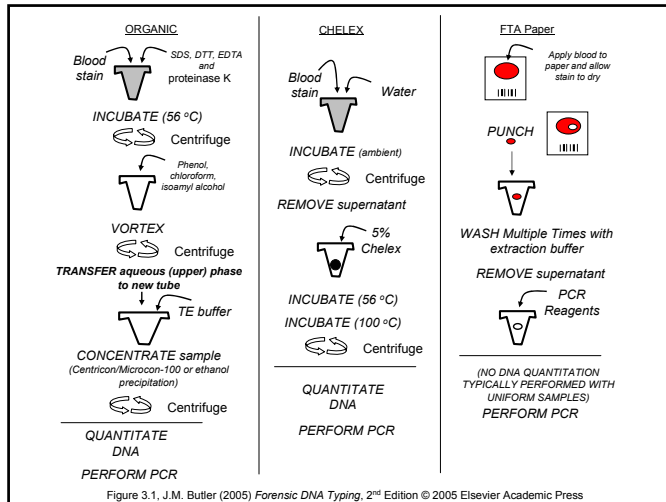


Figure 3.1, J.M. Butler (2005) *Forensic DNA Typing*, 2<sup>nd</sup> Edition © 2005 Elsevier Academic Press

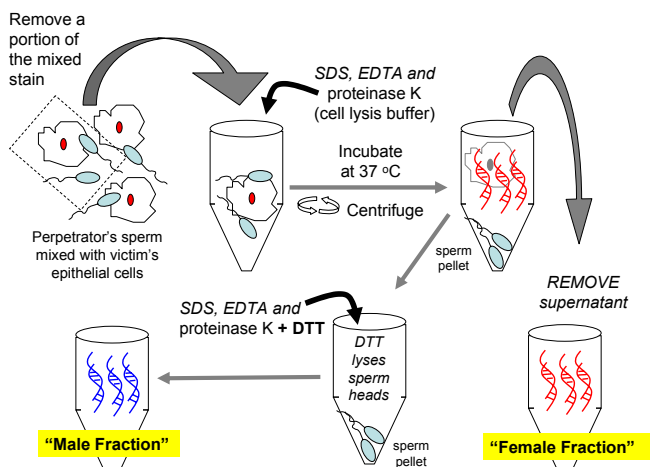
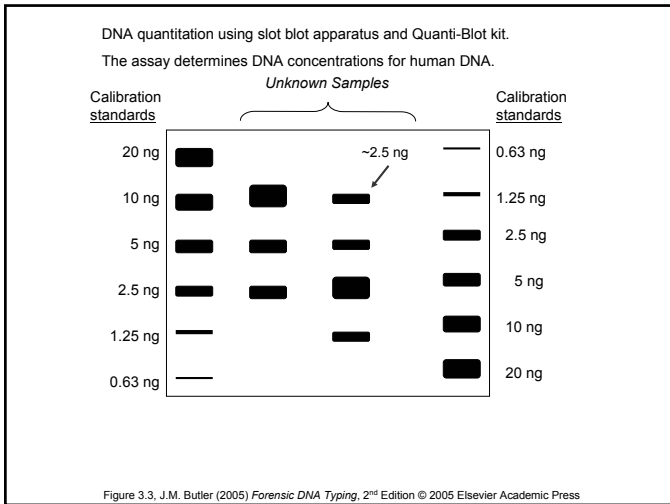


Figure 3.2, J.M. Butler (2005) *Forensic DNA Typing*, 2<sup>nd</sup> Edition © 2005 Elsevier Academic Press

**FORENSIC DNA ANALYSIS**  
**PROCEDURES**

**II. Determination of the quantity/quality of the DNA**

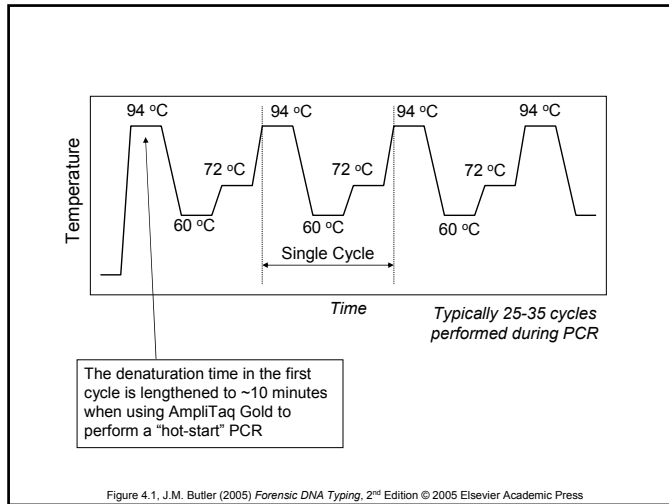
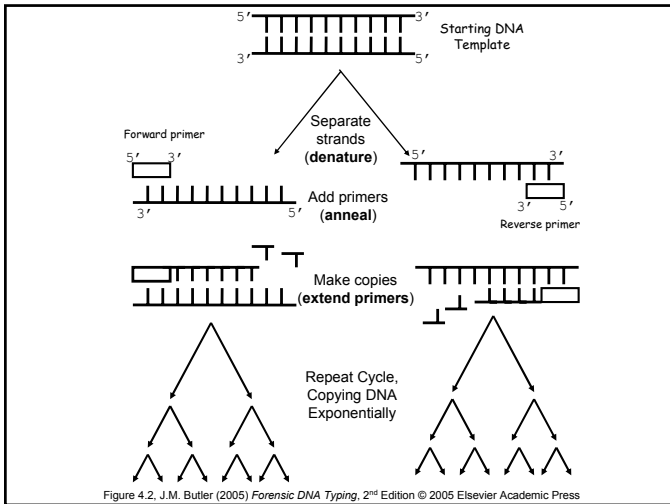
- **- Quantiblot (slot blot)** determines the
  - amount of DNA
  - species
  - primate specific probe to locus D17Z1
  - human origin
  - must be used to analyze DNA from extractions
- **- yield gel** determines the
  - quantity of DNA
  - quality
  - presence of ds high molecular weight (HMW) DNA fragments
- **- required for RFLP analysis**



- **PCR-STR**
- 
- **STRs (Short Tandem Repeats)**
- - length polymorphisms
- - moderately polymorphic
- - short sequences (blocks of 2-5 bp) repeated in series
- - repeats are well characterized
- - “alleles” are designated according to repeat number
- - not linked to any disease
- - low mutation rate
- - population data are available
- - large number
- - sufficient degree of discrimination

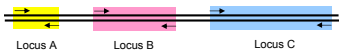
## The Polymerase Chain Reaction

- ### Why amplify DNA in a sample?
- Given that the DNA concentrations in most forensic samples are low and of poor quality (degraded) the ability to amplify what little good quality DNA is left makes it possible to analyze many “impossible” samples.
  - This is accomplished by use of the polymerase chain reaction or PCR.
  - PCR allows specific DNA sequences to be copied billions of times.



## Multiplexing

(A) Simultaneous amplification of three locations on a DNA template



(B) Resolution of PCR products with size-based separation method

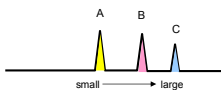


Figure 4.3, J.M. Butler (2005) *Forensic DNA Typing*, 2<sup>nd</sup> Edition © 2005 Elsevier Academic Press

## Advantages of PCR for Forensic DNA Analysis

- Very small amounts of DNA may be used from as little as a single cell
- Degraded DNA can serve as a template.
- Large number of copies of specific DNA sequences amplified simultaneously.
- Contaminate DNA (animal, bacterial) sources will not be amplified.
- Commercial kits available for PCR reactions.

### Disadvantages of PCR for Forensic DNA Analysis

- Target DNA may not be amplified due to PCR inhibitors in the extract.
- Amplification may fail due to changes in primer-binding regions of the genomic DNA template (null-alleles).
- Contamination from other DNA sources (human) that are not part of the evidence can skew the results.

### STR Primers

### STR Primers

- Thousands of microsatellites but they only account for 3% of the total human genome. STRs are scattered throughout the genome and occur on an average of 1/10,000 bases.
- STRs are named for the length of the repeat unit (di, tri, tetra, etc.)
- Tetranucleotide repeats are the most popular STR in use today.
- Repeats can be classified as **simple**, **compound** or **complex** depending on the repeat pattern.

PCR primers are designed to target invariant flanking regions of the STR

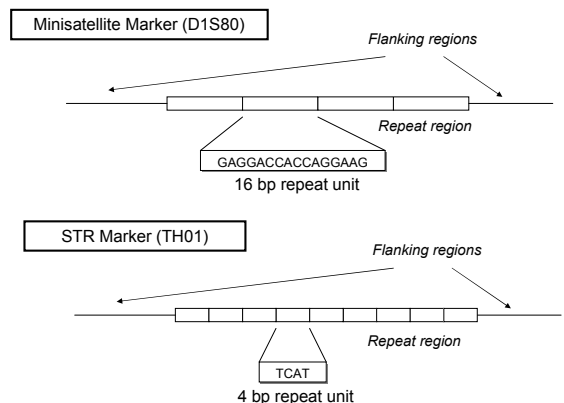


Figure 5.1. J.M. Butler (2005) *Forensic DNA Typing*, 2<sup>nd</sup> Edition © 2005 Elsevier Academic Press

- STRs found in coding regions have the name of the protein that it codes for (TPOX, CF1PO).
- STRs in introns (noncoding regions) are assigned numbers (D1S80).
- The small size of STRs (~100-400bp) compared to minisatellite VNTR (~400-1000 bp) make STRs better candidates for forensic analysis where DNA may be degraded.

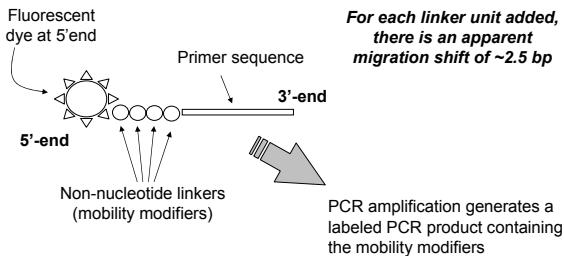
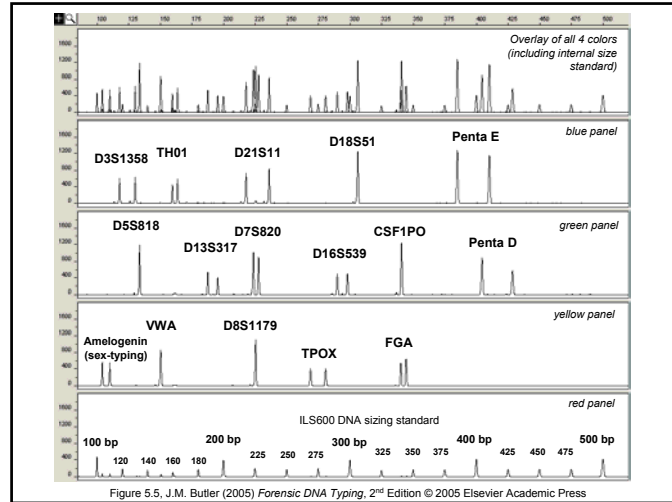
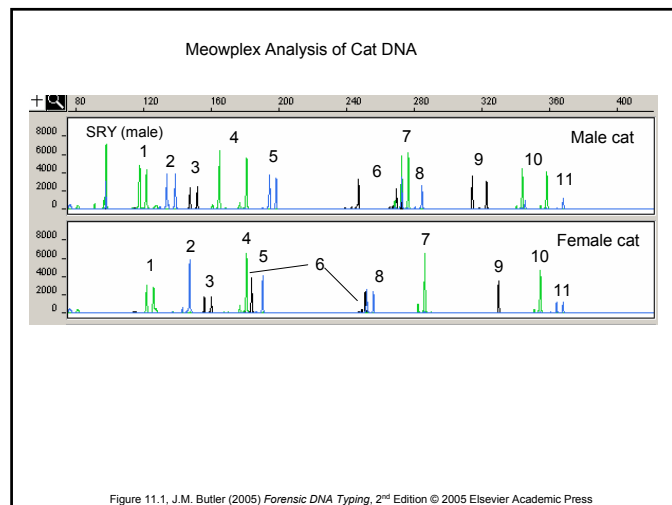
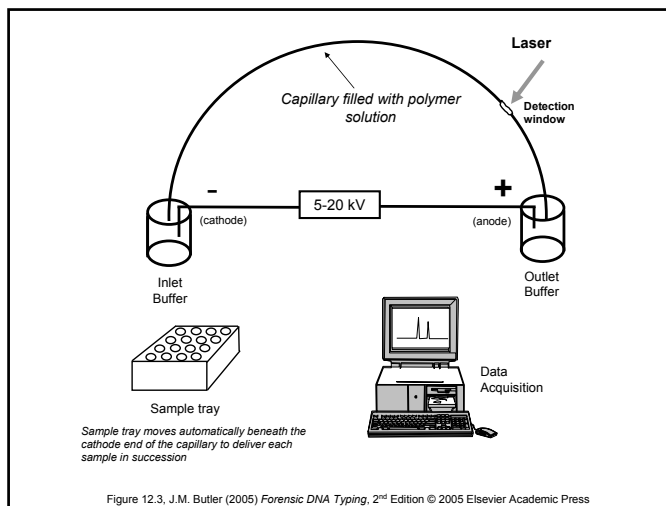


Figure 5.7, J.M. Butler (2005) *Forensic DNA Typing*, 2<sup>nd</sup> Edition © 2005 Elsevier Academic Press

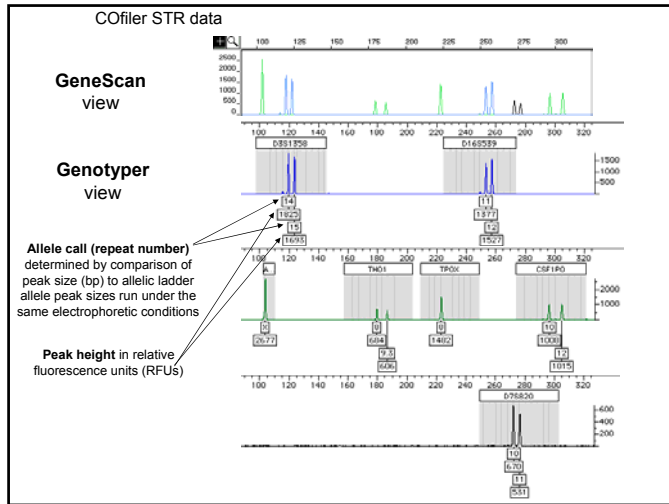
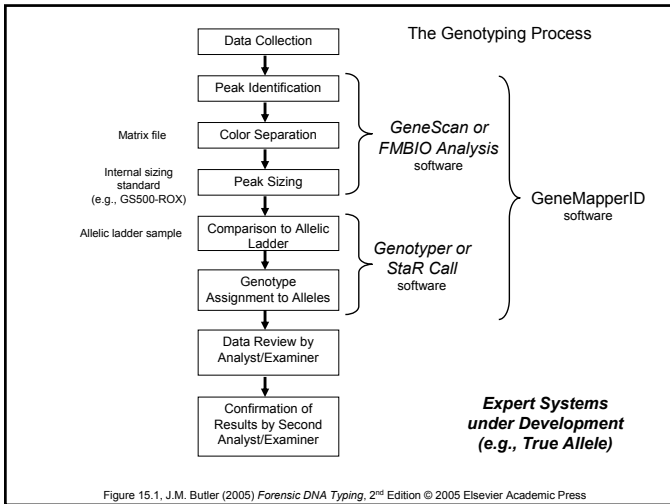


## DNA Separation Methods



## STR Genotyping

- The end result of STR analysis is the assignment of a genotype.
- The fluorescent peak results from the capillary electrophoresis must be converted to STR genotypes.
- A locus genotype is the allele, in the case of a homozygote, or alleles for heterozygotes.
- These are normally reported as the number of repeats present in the allele.
- A sample genotype or STR profile is produced by the combination of all of the locus genotypes into a single series of numbers.



- A forensic laboratory will typically have two independent reads for each sample.
- Comparison of the sample peaks to internal size standards and the allelic ladder produces allele results for the unknown sample.

- ### Three Possible Outcomes
- Butler, J.M. (2005) *Forensic DNA Typing*, 2<sup>nd</sup> Edition, p. 385
- **Match** – Peaks between the compared STR profiles have the same genotypes and no unexplainable differences exist between the samples. Statistical evaluation of the significance of the match is usually reported with the match report (see Chapter 21).
  - **Exclusion (Non-match)** – The genotype comparison shows profile differences that can only be explained by the two samples originating from different sources.
  - **Inconclusive** – The data does not support a conclusion as to whether the profiles match. This finding might be reported if two analysts remain in disagreement after review and discussion of the data and it is felt that insufficient information exists to support any conclusion.

## Combined DNA Index System (CODIS)

- On Oct. 13, 1998, The FBI officially launched its nation-wide DNA database. By the end of 2003, this database, named Combined DNA Index System (CODIS), contained of 1.5 million STR profiles and is linked to all 50 states in the US with capability to search criminal DNA profiles similar to those for fingerprints.

- These databases are effective because a majority of crimes are committed by repeat offenders.
- More than 60% of those put in prison for violent offenses and subsequently release were re-arrested for a similar offense in less than 3 yrs.
- CODIS databases and similar databases in other countries serve several functions.

1. To locate suspects in violent crime cases that would otherwise never have been solved.
2. To make associations between groups of unsolved cases.

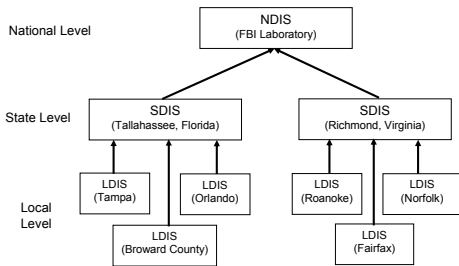


Figure 18.1, J.M. Butler (2005) *Forensic DNA Typing*, 2<sup>nd</sup> Edition © 2005 Elsevier Academic Press

## Statistical Analysis

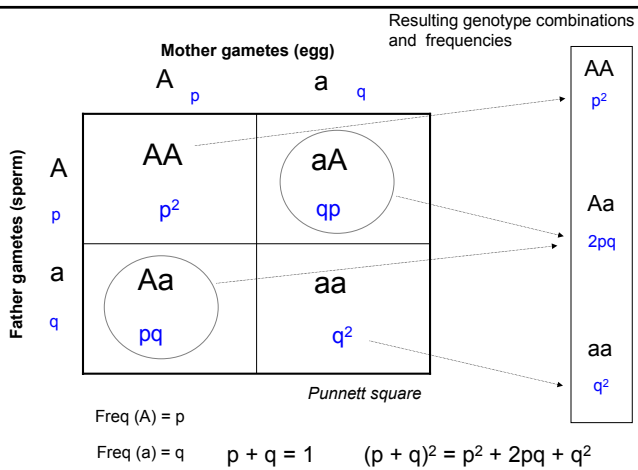


Figure 19.3, J.M. Butler (2005) *Forensic DNA Typing*, 2<sup>nd</sup> Edition © 2005 Elsevier Academic Press

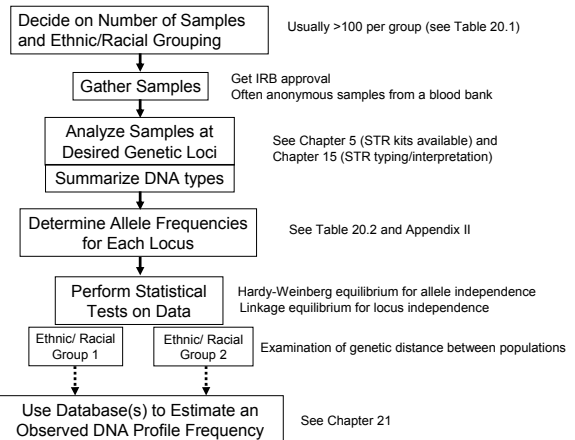


Figure 20.1, J.M. Butler (2005) *Forensic DNA Typing*, 2<sup>nd</sup> Edition © 2005 Elsevier Academic Press

## 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*, 2<sup>nd</sup> Edition

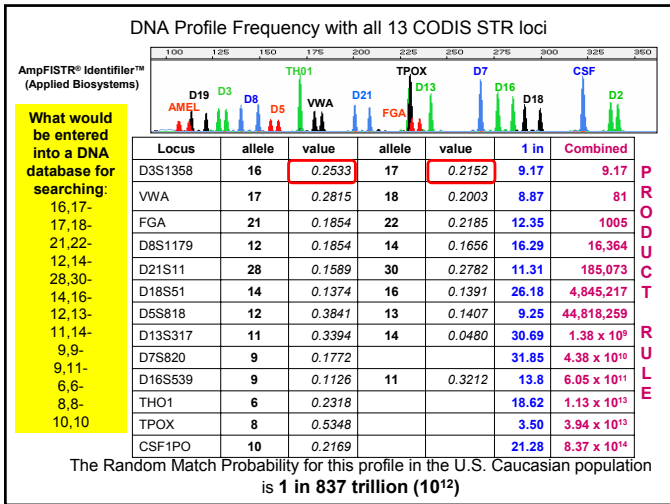
## Allele Frequency Tables

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.

		Butler et al. (2003) JFS 48(4):908-911		Einum et al. (2004) JFS 49(6)			
		Caucasian N= 302		Caucasian N= 7,636		African American N=258	
						African American N= 7,602	
	Allele					Allele	
	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	
<b>Most common allele</b>	15	<b>0.2616</b>	<b>0.2690</b>	15	0.3023	0.2920	
	15.2	--	--	15.2	0.0019*	0.0010	
	16	0.2533	0.2430	16	<b>0.3353</b>	<b>0.3300</b>	
	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			

## Calculating STR Frequency

- We can calculate the frequency of a given profile in the general population as follows:
- If an allele is homozygous the probability is squared =  $p^2$ .
- If an allele is heterozygous we use the following formula –  $2pq$ .
- Individual frequencies for each locus are multiplied by each other to give a total frequency for all loci tested.



- D3S1358 16 0.2533 17 0.2152 9.17 9.17
- So, for locus D3S1358 allele 16 has a frequency of .2533 and allele 17 frequency is 0.2152.
- So using the product rule for a heterozygote (2pq) we would do the calculations as follows:  $2(.2533 \times .2152) = .1076$ .
- Taking the reciprocal of this gives you 1 in 9.29 frequency for these two alleles.
- With additional loci (the next one listed in our table was 8.87 so  $9.17 \times 8.87 = 1$  in 81 frequency etc.) and alleles we would simply multiply them times our 9.17 to get a final frequency.

**The Same 13 Locus STR Profile  
in Different Populations**

**1 in 837 trillion**

**1 in 0.84 quadrillion (10<sup>15</sup>)** in U.S. Caucasian population (NIST)

**1 in 2.46 quadrillion (10<sup>15</sup>)** in U.S. Caucasian population (FBI)\*

**1 in 1.86 quadrillion (10<sup>15</sup>)** in Canadian Caucasian population\*

**1 in 16.6 quadrillion (10<sup>15</sup>)** in African American population (NIST)

**1 in 17.6 quadrillion (10<sup>15</sup>)** in African American population (FBI)\*

**1 in 18.0 quadrillion (10<sup>15</sup>)** in U.S. Hispanic population (NIST)

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

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>