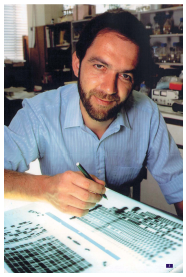
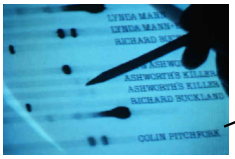
A Bit of History

- The double helix structure of DNA was discovered in 1953 by Watson & Crick, two British scientists who based their work on Rosalind Franklin's data
- Kary Mullis (American) invented a method for duplicating targeted strands of DNA in 1983
- This method is called polymerase chain reaction (PCR)
- In the course of his research on variability in human DNA, Alec Jeffreys (British) developed a method of forensic DNA typing in 1985, originally called "DNA fingerprinting"


Sir Alec Jeffreys

A Bit of History

- "DNA fingerprinting" was first used in 1986 to catch a rapist/ murderer in England named Colin Pitchfork
- The case is chronicled in The Blooding, by Joseph Wambaugh



DNA 101

What *is* it? *Where* is it?

- **DeoxyriboNucleic Acid**
  - The genetic material found in the nucleus of every every nucleated cell in the body
  - Has structure of a "double helix", like a twisted zipper or a twisted ladder
  - Made of long polymer strands of nucleotides (molecules made up of nitrogenous bases and sugar-phosphate groups)

DNA 101

- Blood (white blood cells)
- Roots of hair (epithelial cells)
- Saliva (epithelial cells)
- Semen (sperm cells)
- Skin, dandruff (epithelial cells)
- Sweat stains (epithelial cells)
- Vaginal fluids (epithelial cells)
- Nasal secretions (epithelial cells)
- Urine (epithelial cells)
- Feces (epithelial cells, rarely used)

DNA is the same in every nucleated somatic cell, and is robust and stable

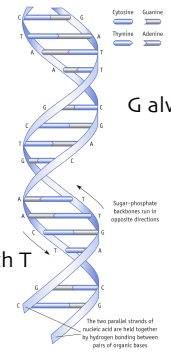
DNA 101

- Nucleotides contain either a purine or a pyrimidine base:

Pyrimidines
- Cytosine (C)
- Thymine (T)

Purines
- Guanine (G)
- Adenine (A)

OHT 6.4



G always pairs with C

A always pairs with T

Figure 6.2 (b) DNA structure

Jackson & Jackson: *Forensic Science*

DNA 101

- A "base pair" is like one wrung of the ladder (eg, one A with one T)
- There are approximately *3 billion* base pairs (C-G, A-T) of nucleotides in every nucleated cell.
- Some sequences of DNA code for certain proteins: genes.
- A gene is a hereditary unit that determines a particular characteristic in an organism.
- Genes are sequences of A,G,C,T *nucleotides.*
- The length and order of nucleotides determines the type of protein that is produced by that gene.
- Humans have ~20,000-25,000 genes

DNA 101

- DNA is organized into *chromosomes.*
- Every person gets 23 chromosomes from each parent, for a total of 46.
- Exactly half your DNA comes from each parent.
- The location of a gene on a chromosome is its *locus* (plural*: loci*).
- An *allele* is an alternate form of gene (for example, eye color)
- Each person inherits one allele from each parent at every locus.

DNA 101

- We share over 99% of our genome with other humans.
- However, that <1% of 3 billion nucleotides is still a significant and detectable level of variation.
- ***Most variation exists in the part of the genome that does not code for genes.***
- This DNA, called "junk DNA", had no known function until recently.
- ENCODE (Encyclopedia of DNA Elements) is a scientific working group at UCSC that recently released several reports on the role of "junk DNA"
- It is understood to be biochemically active, and to have a regulatory function on genes.

DNA 101

- Approximately 98% of our DNA is located in this "junk DNA" part of the genome (which, again, does NOT code for genes).
- *Mutations in this region are tolerated and can accumulate with no affect on the organism… leading to increasing variation.*

DNA 101

Phenotype: the physical characteristics (or exterior expression) of an organism's genetic makeup

Genotype: 1. the genetic makeup of an organism;
2. the combination of alleles present at a particular locus, or at all loci present

Mother's Genotype: Bb

Mother's phenotype: Brown eyes

Father's Genotype: Bb

Father's phenotype: Brown eyes

	B	b
B	BB	Bb
b	Bb	bb

If they each have a blue recessive gene, two brown-eyed parents may have a blue eyed child.

In reality, inheritance is much more complicated than this Punnett Square diagram, but it offers a useful illustration of genotype and phenotype.

DNA 101

PCR: Polymerase Chain Reaction

A method for duplicating and amplifying targeted regions of DNA outside the body

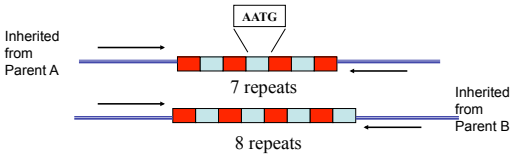
STR: Short Tandem Repeat

A repeat of a short sequence of nucleotides found at a specific locus

Short Tandem Repeats

- **S**hort **T**andem **R**epeats (STRs)
  - **Short** because the sequences are short – usually 1-4 nucleotides in length
  - **Tandem** because they occur one after the other
  - **Repeats** because they are repeats of the same sequence
  - ATCGACCTTG-GCCG-GCCG-GCCG-GCCG-ATCGATTGACCTAAC
    = 4 **s**hort **t**andem **r**epeats of GCCG
- These are found in sections of DNA that are non-coding (aka junk DNA).
- The number of repeats of a particular sequence may vary between individuals.

Short Tandem Repeats



This person has 7 repeats of AATG from parent A, and 8 repeats of AATG from parent B, at this locus.

This person's genotype at this particular locus is 7,8 (or 8,7).

An Example of a STR in locus D7S280

- D7S280 is a locus on human chromosome 7. Its DNA sequence, as obtained from GenBank (a public DNA database) is gata.

```
aatttttgta ttttttttag agacggggtt tcaccatgtt ggtcaggctg
actatggagt tattttaagg ttaatatata taaagggtat gatagaacac
ttgtcatagt ttagaacgaa ctaacgatag atagatagat agatagatag
atagatagat agatagatag atagacagat tgatagtttt tttttatctc
actaaatagt ctatagtaaa catttaatta ccaatatttg gtgcaattct
gtcaatgagg ataaatgtgg aatcgttata attcttaaga atatatattc
cctctgagtt tttgatacct cagattttaa ggcc
```

- Alleles at this locus have from 6 to 15 tandem repeats of the 'gata' sequence.
- In other words, every person has between 6 to 15 STRs of gata at this location (locus) in their DNA.

An Example of a STR in locus D7S280

- So what?
- If we look at only one locus (for example, D7S280), then the fact that every person has either 6, 7, 8, 9, 10, 11, 12, 13, 14, or 15 repeats of gata is not enough for us to tell them all apart, because too many people will have the SAME number of repeats at that locus. (For example, there may be *thousands* of people who have the genotype 7,8 at that locus.)
- HOWEVER, the more loci we look at, the fewer people will share numbers of repeats.
- AND, if we look at many loci, which we do in DNA profiling, then the number of people who share the numbers of repeats at all of the loci will be extremely close to ZERO.
- This is where probability becomes relevant.
- Recall that probability means *the frequency of occurrence of an event.*
- The probability that you will have the exact same number of repeats at the same loci as anyone else is infinitesimally small (unless you have an identical twin).
- This is why DNA evidence is so powerful.

CODIS

- Combined DNA Index System
- All forensic laboratories that use the CODIS system can contribute DNA profiles to the CODIS database.
  - The **Forensic Index** contains DNA profiles from crime scene evidence.
  - The **Offender Index** contains DNA profiles of individuals convicted of sex offenses (and other violent crimes) with many states now expanding legislation to include other felonies.
  - Looks at 13 loci

- A typical DNA case involves the comparison of two samples – an unknown or *evidence* sample, such as semen from a rape, and a known or *reference* sample, such as a blood sample from a suspect.
- If the DNA profiles obtained from the two samples are indistinguishable (they "match"), that is evidence for the court that the samples have a common source – in other words, that the suspect *is* the source of the semen.
- A logical and important legal question, then, is: well, how do we know that the suspect is the only person with this particular profile? How do we know that our suspect's DNA doesn't *just happen* to match the crime scene sample by chance?

A Sample Frequency Database

Locus	Alleles	#Times Observed	Dbase size	Frequency of Occurrence
CSF1PO	10 11	109 134	432	=.25 =.31
TPOX	8 8	229	432	=.53
THO1	6 7	102 64	428	=.24 =.15
vWA	16 16	91	428	=.21

It is slightly more complicated, but this gives you the general idea.

How rare is the profile?

- For a DNA profile to be considered useful in identification, it must be extremely rare -- otherwise it would describe too many people
- The uniqueness of a profile is determined using frequency databases (frequencies of alleles in a given population) and probability law
- The number of repeats found at each locus is an independent event (in other words, the number of repeats at a locus is not related to the number of repeats found at another locus)
- Probability Law: "The probability that two independent events may happen together is the *product* of their individual probabilities"
- This means we can multiply those frequencies together!

A Sample Profile

Locus	D3S1358	vWA	FGA	D8S1179	D21S11	D18S51	D5S818
Genotype	15, 18	16, 16	19, 24	12, 13	29, 31	12, 13	11, 13
Frequency	8.2%	4.4%	1.7%	9.9%	2.3%	4.3%	13%

Locus	D13S317	D7S820	D16S539	THO1	TPOX	CSF1PO	AMEL
Genotype	11, 11	10, 10	11, 11	9, 9.3	8, 8	11, 11	X Y
Frequency	1.2%	6.3%	9.5%	9.6%	3.52%	7.2%	(Male)

- The *genotype* row tells you the number of repeats this person has at a particular locus.
- The *frequency* row tells you how often this particular combination of alleles appears in the general population. (Frequencies are determined simply by counting the number of alleles in a given population.)
- Because we know that the number of repeats occurring at each locus is an independent event, we can multiply their frequencies together to determine how probable that combination is.

A Sample Profile

Locus	D3S1358	vWA	FGA	D8S1179	D21S11	D18S51	D5S818
Genotype	15, 18	16, 16	19, 24	12, 13	29, 31	12, 13	11, 13
Frequency	8.2%	4.4%	1.7%	9.9%	2.3%	4.3%	13%

Locus	D13S317	D7S820	D16S539	THO1	TPOX	CSF1PO	AMEL
Genotype	11, 11	10, 10	11, 11	9, 9.3	8, 8	11, 11	X Y
Frequency	1.2%	6.3%	9.5%	9.6%	3.52%	7.2%	(Male)

- FOR EXAMPLE: If we want to know how often "19, 24" is found at FGA, we only need to look at its frequency in the chart to see that 1.7% of the people in this population have that genotype at that locus.
- If we want to know how often "19, 24" is found at FGA **TOGETHER WITH** "11, 13" at D5S818, we simply multiply 1.7% x 13% (or .017 x .13). We can do this because they are independent events.
- The answer is .00221 -- or .221%
- So, the combination of those alleles at those 2 loci is quite rare! Only 2 in 1000 people have that combination of alleles.

A Sample Profile

Locus	D3S1358	vWA	FGA	D8S1179	D21S11	D18S51	D5S818
Genotype	15, 18	16, 16	19, 24	12, 13	29, 31	12, 13	11, 13
Frequency	8.2%	4.4%	1.7%	9.9%	2.3%	4.3%	13%

Locus	D13S317	D7S820	D16S539	THO1	TPOX	CSF1PO	AMEL
Genotype	11, 11	10, 10	11, 11	9, 9.3	8, 8	11, 11	X Y
Frequency	1.2%	6.3%	9.5%	9.6%	3.52%	7.2%	(Male)

- CONTINUING OUR EXAMPLE…
- You can probably see that the more frequencies we multiply together, the more rare this combination of alleles becomes.
- .082x.044x.017x.099x.023x.043x.13x.012x.063x.095x.096x.0352x.072 = (drumroll) .000000000000000013642127!!!
- So the answer to our question: how rare is this profile?
  LESS THAN 1 IN 100 QUADRILLION!
- (Unless I pressed the wrong button on my calculator somewhere)

A Sample Profile

Locus	D3S1358	vWA	FGA	D8S1179	D21S11	D18S51	D5S818
Genotype	15, 18	16, 16	19, 24	12, 13	29, 31	12, 13	11, 13
Frequency	8.2%	4.4%	1.7%	9.9%	2.3%	4.3%	13%

Locus	D13S317	D7S820	D16S539	THO1	TPOX	CSF1PO	AMEL
Genotype	11, 11	10, 10	11, 11	9, 9.3	8, 8	11, 11	X Y
Frequency	1.2%	6.3%	9.5%	9.6%	3.52%	7.2%	(Male)

- So, let's say you collect a DNA sample at a sexual assault scene, or off of some evidence in the crime lab, and get this profile
- And it just so happens that there is a "matching" profile in the CODIS database. This matching person (let's call him XA) becomes a suspect.
- XA says, of course, that he was nowhere near the crime scene and had nothing to do with it, doesn't know the victim, etc.
- Presuming he is innocent, which our justice system does, we have to answer this question: What is the probability that XA matches the crime scene sample *by chance*, and not because he is the perpetrator? In other words, what are the chances that the DNA found at the scene *is not his DNA*?

A Sample Profile

Locus	D3S1358	vWA	FGA	D8S1179	D21S11	D18S51	D5S818
Genotype	15, 18	16, 16	19, 24	12, 13	29, 31	12, 13	11, 13
Frequency	8.2%	4.4%	1.7%	9.9%	2.3%	4.3%	13%

Locus	D13S317	D7S820	D16S539	THO1	TPOX	CSF1PO	AMEL
Genotype	11, 11	10, 10	11, 11	9, 9.3	8, 8	11, 11	X Y
Frequency	1.2%	6.3%	9.5%	9.6%	3.52%	7.2%	(Male)

UNFORTUNATELY FOR XA, IT'S LESS THAN 1 IN 100 QUADRILLION!

How do we get the profile? How do we find out what a person's alleles are?

- I'm glad you asked that.

The Process (in Summary)

1. Extract DNA from sample
2. Quantify the DNA
3. Amplify the DNA (PCR)
4. Separate & detect PCR products via Capillary Electrophoresis
9. Determine genotype
10. Compare to reference profiles
11. Consult population dbase & determine frequency of profile

- Remember, in DNA profiling, we are not concerned with genes, because that's not where the variation is.
- Rather, the DNA we look at is in the junk DNA regions.
- Every cell has only one copy of nuclear DNA, which is not enough for us to "see".
- So, we must first target the areas of DNA that we want to see more closely (the 13 loci) and amplify or copy it, so that there is enough for us to analyze.
- We do this using PCR…

PCR: Polymerase Chain Reaction

- Because most tissue samples from a crime scene contain very little DNA, whatever amount is present must be amplified (many copies made of)
- In STR analysis, we want to amplify the DNA containing the tandem repeats *and* ***only this DNA***
- PCR machines, or thermocyclers, use repeated cycles of heating and cooling to denature (unzip) and replicate (rezip) the DNA using many of the same enzymes found in cells which facilitate DNA replication naturally.

The PCR Song

- Believe it or not, you can learn a lot about PCR from this song….
- http://www.youtube.com/watch?v=x5yPkxCLads

The PCR Song

"There was a time when to amplify DNA…You had to grow tons and tons of tiny cells.Then along came a guy named Dr. Kary Mullis, Said you can amplify in vitro just as well. Just mix your template with a buffer and some primers, nucleotides and polymerases, too. Denaturing, annealing, and extending. Well it's amazing what heating and cooling and heating will do.

PCR, when you need to detect mutations. PCR, when you need to recombine. PCR, when you need to find out who the daddy is (who's your daddy?). PCR, when you need to solve a crime."

OHT 6.14

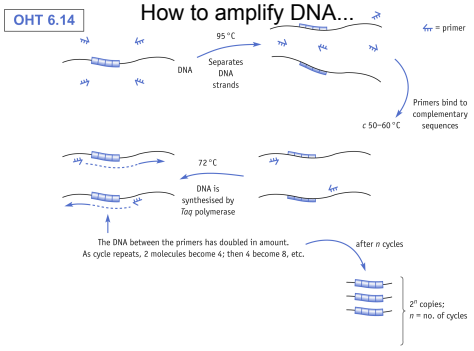


Figure 6.7 The polymerase chain reaction (PCR) (a) the basis of the process

Jackson & Jackson: *Forensic Science*

OHT 6.15

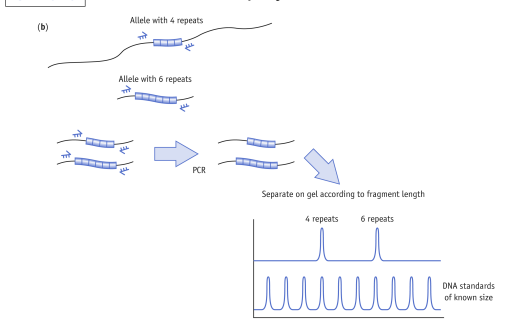


Figure 6.7 The polymerase chain reaction (PCR) (b) STR analysis using PCR

Jackson & Jackson: *Forensic Science*

PCR

- Using a thermocycler, the sample is denatured (double helix is unzipped) @ ~95°C
- Primers (oligonucleotides, typically 20 bases long) adhere to complimentary open strands (A-T, G-C) @ ~55°C
- The new DNA strand is synthesized with Taq polymerase @ 72 °C
- Process is repeated until there are enough copies of the DNA for detection using electrophoresis

Electrophoresis

Electrophoresis is the process of moving charged particles through a gel plate by applying an electric field (eg, negatively charged DNA will move toward a positive source).

- Following PCR, amplified DNA samples are separated by size (smaller particles move faster than larger ones) through the gel during electrophoresis.
- The process is automated using an instrument (a genetic analyzer) that reads the DNA by size -- a laser scans and detects the DNA samples as they electrophorese.

OHT 6.11

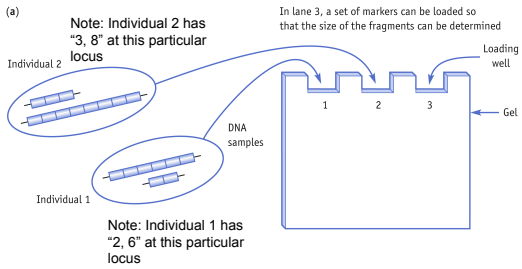


Figure 6.6 Separating DNA molecules according to their length: gel electrophoresis (a) loading the gel

OHT 6.12

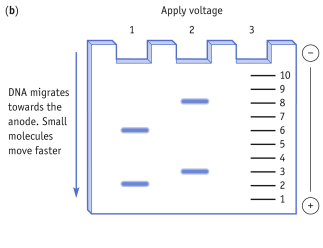
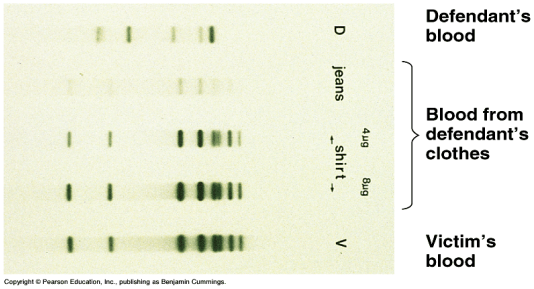


Figure 6.6 Separating DNA molecules according to their length: gel electrophoresis (b) DNA migration

Electrophoresis Autoradiograph

We used to get "bands" of DNA that looked like this (below). Now we have automated DNA instruments, which show DNA not as bands but as peaks (next slide).



OHT 6.13

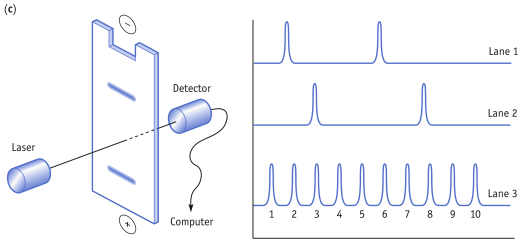


Figure 6.6 Separating DNA molecules according to their length: gel electrophoresis (c) detecting fluorescently tagged DNA by laser

OHT 6.2

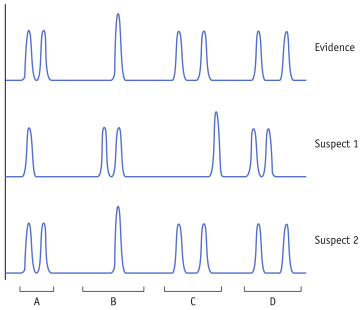


Figure 6.1 (continued) Simplified examples of DNA profiles

OHT 6.16

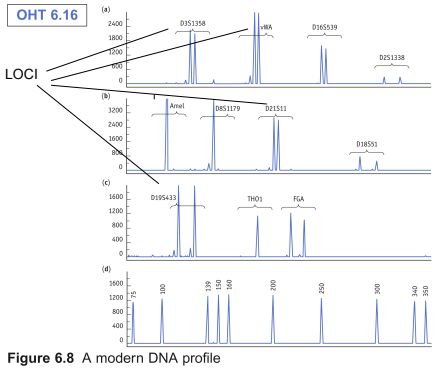


Figure 6.8 A modern DNA profile

Jackson & Jackson: *Forensic Science*

Other Types of DNA

- We've been talking about nuclear DNA (nDNA)
- There are other types of DNA that are useful for establishing genetic groups, but are otherwise not as discriminating
  - MtDNA (mitochondrial DNA) -- used to determine maternal relationships and maternal groups
    - EVERYONE has mtDNA. You inherited your mtDNA from your mother, without recombination with your father. Your mtDNA is an exact match with your mother's mtDNA, your mother's siblings, your mother's mother and all her siblings, and so on. If you are female, your daughters will continue this mtDNA.
  - Y-STR (short tandem repeats found on the Y chromosome -- used to determine paternal relationships and paternal groups in males
    - ONLY MALES HAVE YSTR. It is an exact match with your father, your brothers, your sons, your father's father and his male siblings….

mtDNA & Y-STR

Mitochondrial DNA

- 37 genes, 16569 base pairs
- 100-10,000 copies per cell (significantly more than nuclear DNA) make it useful for analyzing degraded samples
- Circular structure
- Non-coding D loop used for testing
- *Inherited maternally by all offspring*

Y-STR

- Short tandem repeats located on Y chromosome
- Useful for resolving mixtures between males & females
- *Inherited paternally by males only*

Some Considerations

- Although DNA is relatively stable, it does denature or get destroyed through enzyme action, from bacteria or through oxidation
- Therefore, samples should be collected quickly and preserved (usually by freezing if possible)
- Care should also be taken not to cross contaminate during collection -- including from collector!
- Blood is also a potential pathogen, so care must be taken to avoid exposing oneself to blood borne viruses like Hep B, tuberculosis or HIV

Some uses of DNA Profiling

- Forensic work on crime scenes
- Parentage testing
- Victim identification in mass disasters
- Animal identification- e.g. racehorse paternity, endangered species poaching
- Conservation biology and evolutionary studies

DNA Profiling can solve crimes

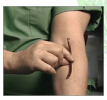
- The DNA profile is compared with those of the victim and the suspect.
- If the profile matches the suspect, it provides strong evidence that the suspect was present at the crime scene (**Note: this does not necessarily prove he or she committed the crime).**
- If the profile doesn't match the suspect, then that suspect may be eliminated from the inquiry.

As the technology gets smarter, so too do the criminals

- A physician in Canada eludes authorities for years
- Accused of drugging and sexually assaulting patients, DNA profiles from semen samples from the assaulted women do not match Dr. Schneeberger
- Blood was drawn on 3 occasions in 1992, 1993 and 1996, but never came back as a match
- Finally police obtain blood from a finger prick, swabbed the inside of his cheek and took hair samples
- The results matched the DNA from the semen of the victims
- How did he get away with it?

As the technology gets smarter, so too do the criminals

- On the previous 3 occasions, blood was drawn from the same arm
- The last time the blood was drawn, the technician stated that the blood looked brown and "old"
- Schneeberger had surgically implanted a piece of rubber tubing in his arm and filled it with stored blood from a patient



- DNA is also used in the identification of remains recovered in mass disasters


The Angel of Death: Josef Menegle

- Josef Mengele was a Nazi war criminal notorious for grotesque human experiments that he carried out at the Auschwitz concentration camp.
- After the Second World War he fled from the Allies and escaped to South America. The fugitive succeeded in living out the rest of his days without being caught.
- In 1985 investigators went to the cemetery of Nossa Senhora do Rosario in the small Brazilian town of Embu to dig up the skeleton of a man who had been drowned in a swimming accident six years previously.
- Using DNA extracted from blood provided by Mengele's wife and son, it was concluded that it was more than 99.94% certain that the skeleton was Mengele's.

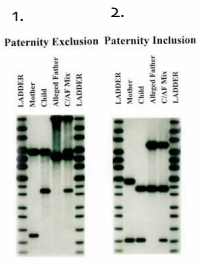

Paternity Cases

- Who's your daddy?

1.


2.


1. 2.



Paternity Cases

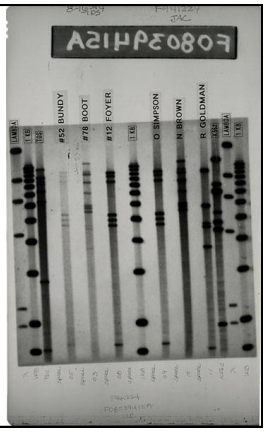
DNA Marker	Mother	Child	Alleged Father
D21S11	28, 30	28, 31	29, 31
D7S820	9, 10	10, 11	11, 12
TH01	14, 15	15, 15	15, 16
D13S317	7, 8	7, 9	8, 9
D19S433	14, 16.2	14, 15	15, 17

Exoneration

- Kirk Bloodsworth
  - Convicted in 1985 for the rape and strangulation of a 9-year old girl and sent to death row
  - In 1992, defense attorneys were successful in having a dime-sized semen stain on the girl's underpants tested against Bloodsworth's DNA
  - He was exonerated

Exoneration

OJ Simpson



DNA Profiling

- **"I didn't understand the DNA stuff at all. To me, it was just a waste of time. It was way out there and carried absolutely no weight with me at all."**
- Post-trial commentary from a juror in the O.J. Simpson trial: V. Bugliosi, Outrage (New York: Dell Publishing, 1996).
- **"In a forensic setting, … an innocent suspect has little to fear from DNA evidence, unless he or she has an evil twin."**
- N. Risch & B. Devlin, "On the Probability of Matching DNA Fingerprints" (1992) 255 Science.