

DNA profiling and Genome Structure

About.com US Politics


US Politics | The Issues | Elections & Leaders | Politics 101

Ex-Illinois Governor Ryan Convicted on 18 Charges

From Kathy Gill, About.com Guide April 17, 2006

Update 1

A federal jury in Chicago has convicted former Illinois Governor George Ryan (R, age 72) of 18 corruption-related charges stemming from an FBI investigation into allegations that unqualified truck drivers could secure driver licenses in exchange for a bribe. Ryan faces a maximum penalty of 95 years in prison and \$4.5 million in fines for racketeering conspiracy, mail fraud, tax fraud and making false statements.



Ryan was governor from 1999-2003 and secretary of state from 1991-1999. The FBI investigation has resulted in charges against 79 state workers and lobbyists; 75 people have been convicted. No one has been acquitted.

The prosecutor, Patrick Fitzgerald, is also investigating the [Valerie Plame](#) case in Washington, DC. The sole indictment in that case to date is [Lewis "Scooter" Libby](#), former chief of staff to Vice President Cheney.

Media report that Ryan was nominated for the Nobel Peace Prize in 2003. However, there is no inherent prestige in being nominated -- anyone can be nominated. The prestige comes from actually being considered -- and that list of persons is [not made available by the Nobel committee](#) for 50 years after the nomination.

From The Prosecutor

"Ryan is charged with betraying the citizens of Illinois for over a decade on state business, both large and small. By giving friends free rein over state employees and state business to make profits- and by steering those profits to his friends and, at times, his family-defendant Ryan sold his office." US Attorney Patrick Fitzgerald of Chicago on 17 December 2003, the date of Ryan's indictment.

The state of Illinois has a grand tradition of corrupt politicians (Re: Former governor Rod Blagojevich.)



Citing Issue of Fairness, Governor Clears Out Death Row in Illinois

New York Times Jan 12, 2003

Condemning the capital punishment system as fundamentally flawed and unfair, Gov. George Ryan commuted all Illinois death sentences today to prison terms of life or less, the largest such emptying of death row in history.

In one sweep, Governor Ryan, a Republican, spared the lives of 163 men and 4 women who have served a collective 2,000 years for the murders of more than 250 people. His bold move was seen as the most significant statement questioning capital punishment since the Supreme Court struck down states' old death penalty laws in 1972. It seemed sure to secure Mr. Ryan's legacy as a leading critic of state-sponsored executions even as he faces possible indictment in a corruption scandal that stopped him from seeking re-election.

"The facts that I have seen in reviewing each and every one of these cases raised questions not only about the innocence of people on death row, but about the fairness of the death penalty system as a whole," Governor Ryan said this afternoon. "Our capital system is haunted by the demon of error: error in determining guilt and error in determining who among the guilty deserves to die."

Governor Ryan, a pharmacist who was among the Illinois legislators who voted in 1977 to revive the death penalty, acknowledged in his speech the unlikelihood of his crusade. ***But when he found himself at the helm of a state that had conducted 12 executions and exonerated 13 death row inmates, one of whom came within 48 hours of the electric chair, Mr. Ryan called a moratorium on capital punishment.***

What was the basis of most if not all of these 13 exonerations?

Innocence Project

<http://www.innocenceproject.org/know/>

Death penalty by state:

<http://fire.biol.wvu.edu/trent/trent/DeathPenaltyThe.pdf>

- *To date, there have been 289 post-conviction DNA exonerations in the United States.*
- *In almost 40 percent of the cases profiled on the Innocence Project, the actual perpetrator has been identified by DNA testing*



After serving 14 years in prison, a DNA fingerprinting test showed that Thomas Webb had been wrongly convicted of rape.

The common themes that run through these cases — from global problems like poverty and racial issues to criminal justice issues *like eyewitness misidentification, invalid or improper forensic science, overzealous police and prosecutors and inept defense counsel* — cannot be ignored and continue to plague our criminal justice system.


*This lecture is about genetics and not about our system of justice
Take note though of the interesting case before the US Supreme Court (the decision is
due out in June)*

The New York Times
The Opinion Pages

WORLD U.S. N.Y. / REGION BUSINESS TECHNOLOGY SCIENCE HEALTH SPORTS OPINI

EDITORIAL
DNA and the Constitution
Published: February 24, 2013


http://www.nytimes.com/2013/02/25/opinion/dna-and-the-constitution.html?ref=opinion&_r=0


 news arts & life music listen

news > u.s. > law

Supreme Court Considers If Warrantless DNA Swab Violates Constitution

by NINA TOTENBERG
February 26, 2013 3:23 AM

 **Listen to the Story**
Morning Edition

 5 min 7 sec

- + Playlist
- + Download
- = Transcript

<http://www.npr.org/2013/02/26/172886713/supreme-court-considers-if-warrantless-dna-swab-violates-constitution>

What is so compelling about the science underlying DNA profiling that it can be used to overturn a conviction of an individual?

What is so compelling about the science underlying DNA profiling that it can be used to convict an individual of a serious crime?

To address this question, we need to consider

- *the structure of our genome*
- *how to use PCR to genotype highly polymorphic sites*
- *basic issues in population genetics*

Good web sites on DNA forensics

http://www.ornl.gov/sci/techresources/Human_Genome/elsi/forensics.shtml

<http://www.dnai.org/d/>

If you were working up a DNA fingerprinting strategy, what criteria would you consider in determining whether a particular allele or locus or site on a chromosome would be useful for distinguishing different individuals based on their DNA pattern?

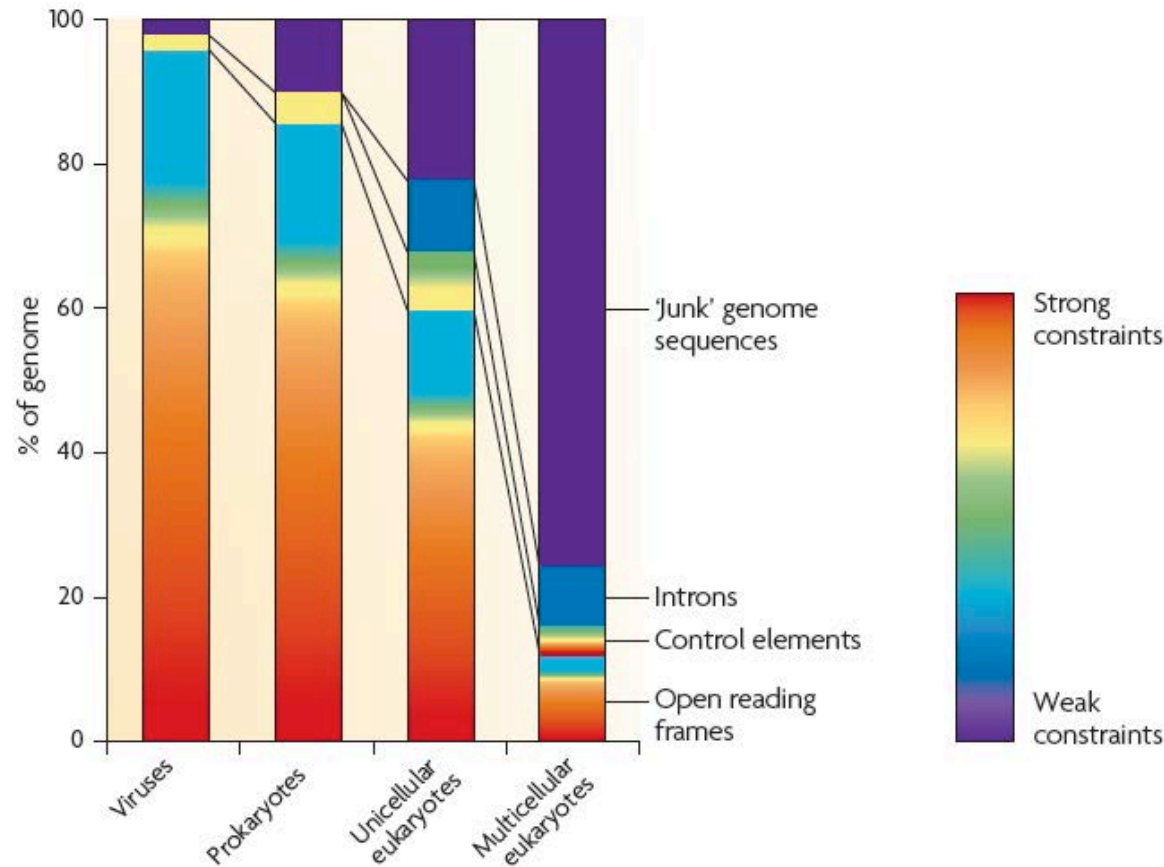


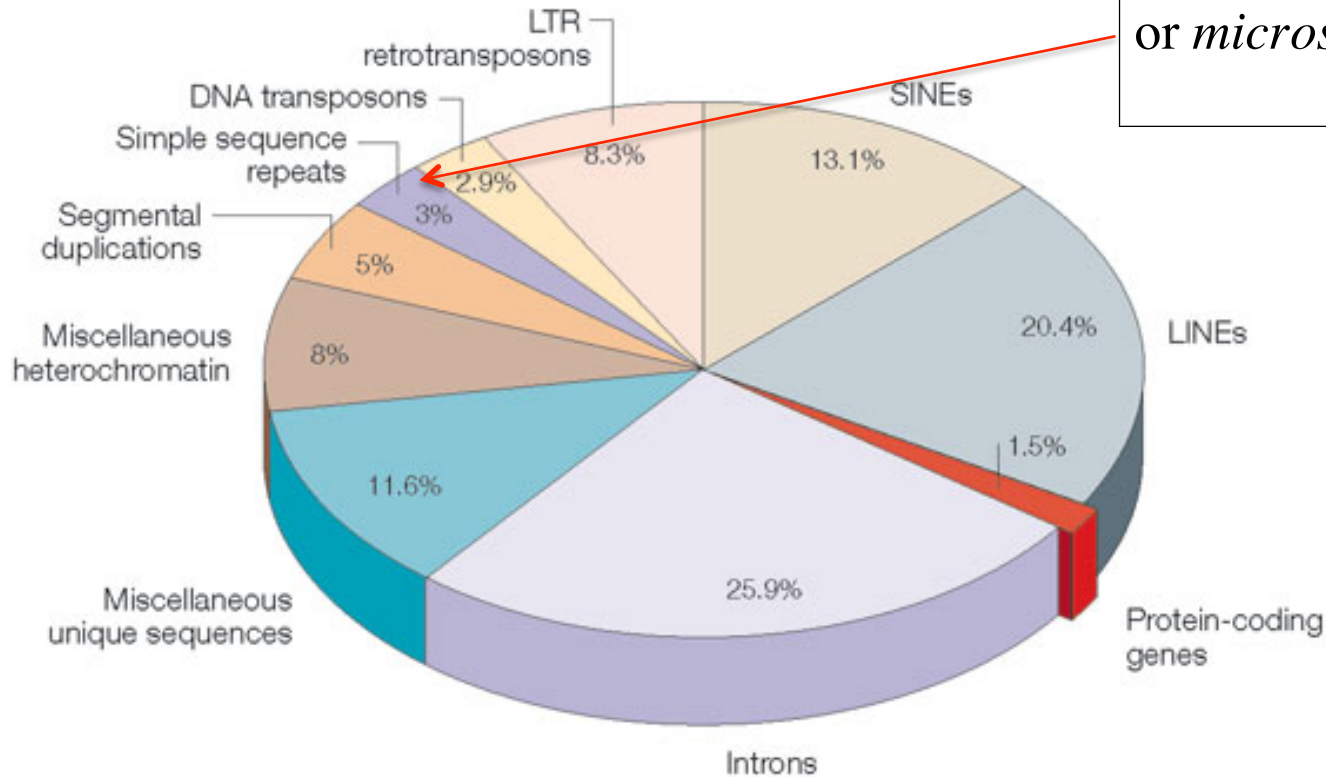
Figure 1 | **Approximate distribution of evolutionary constraints across genomes with different architectures.** The fractions of different classes of sequences that are subject to constraints of varying strength are shown as rough approximations of the values that are typical of the respective class of genomes. The data are from REFS 2,23,29,58, as discussed in the main text.

What do we mean by Open Reading Frames (ORFs)?

What do we mean by Junk sequences?

What do we mean by constraints?

Which category of sequences would be most useful for exploring genomic “individuality”?



STR aka SSRs
or *microsatellites*

Copyright © 2005 Nature Publishing Group
Nature Reviews | **Genetics**

Components of the Human Genome

http://en.wikipedia.org/wiki/Human_genome

Chromosome 11 “Flyover”

<http://www.dnalc.org/ddnalc/resources/chr11.html>

The human genome provides a rich source of genetic variability especially in non-coding regions that have more limited constraints on their evolution
(SEE also last page of these notes)

Single nucleotide polymorphisms (SNP's)

Frequency in genome: ~1 per every 1000 bp Number per genome: ?

Mutation rate per site per generation per haploid gamete: 3×10^{-8}

[Compare to Gene mutation rate: 10^{-4} - 10^{-5} per gene per gamete]

98% of genetic diversity is in this category [but SNP's have fewer possible alleles than micro and mini satellites]

Microsatellites =STRs (simple tandem repeats)=SSRs (simple sequence repeat)

repeat unit: typically 2-6 bp (up to 10 bps)

up to ~25 tandem repeats

Mutation rate per site per gamete: $\sim 1 \times 10^{-3}$

Also called: *Simple Sequence Repeats* or *Simple Tandem Repeats* Nature 409: 888 2/15/01

Minisatellites aka VNTR = variable number of tandem repeats

repeat unit: 15-100 bp in length

up to about 50 tandem repeats

Mutation rate per site per gamete: $\sim 1 \times 10^{-3}$

An example of a microsatellite or STR polymorphism

* STRs are short sequences of DNA, normally of length 2-5 base pairs, that are repeated numerous times in a head-tail manner.

* The 16 bp sequence of "gatagatagatagata" would represent 4 head-tail copies of the tetramer "gata".

* Example: **D7S280**

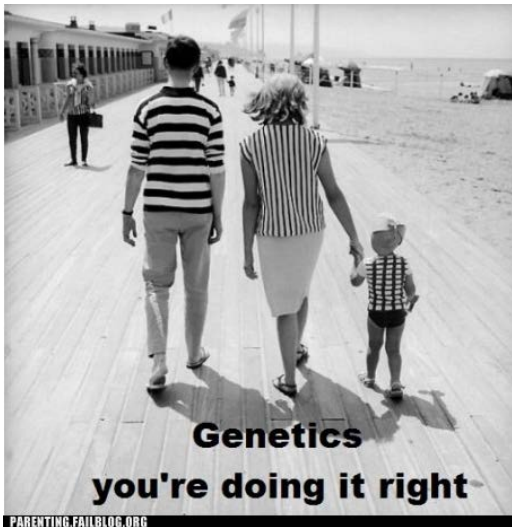
```
1 aatTTTTgta ttttttttag agacgggggtt tcaccatggt ggtcaggctg actatggagt
61 tattttaagg ttaatata taaagggtat gatagaacac ttgtcatagt ttagaacgaa
121 ctaacgatag atagatagat agatagatag atagatagat agatagatag atagacagat
181 tgatagtttt tttttatctc actaaatagt ctatagtaaa catttaatta ccaatatttg
241 gtgcaattct gtcaatgagg ataatgtgg aatcgttata attcttaaga atatatatc
301 cctctgagtt ttgatacct cagattttaa ggcc
```

* The polymorphisms in STRs are due to the different number of copies of the repeat element that can occur in a population of individuals.

* *THIRTY ALLELES of the D7S280 site have been described reflecting different numbers of repeats*

Polymorphisms in mini and microsatellites are used for DNA profiles

- highly polymorphic *mutational hotspots* that are under no obvious selection pressure (anonymous” sites with respect to function)
- easy to assay using PCR
- codominant Mendelian alleles (no dominance complications)

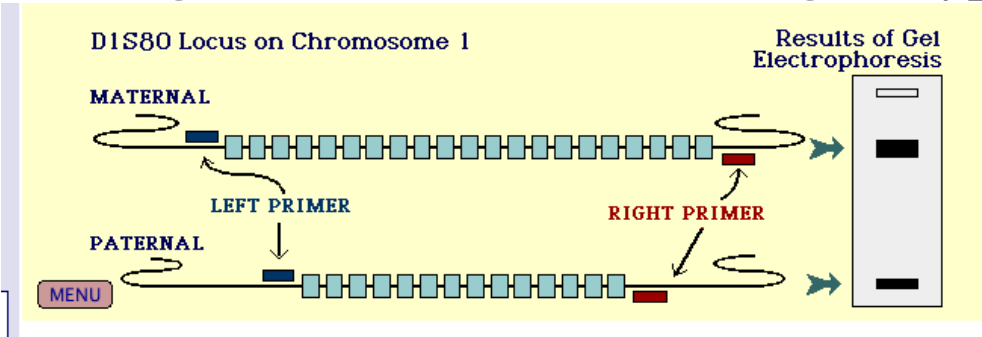


Estimated mutation rate at a given mini/microsatellite site is 1×10^{-3} per gamete

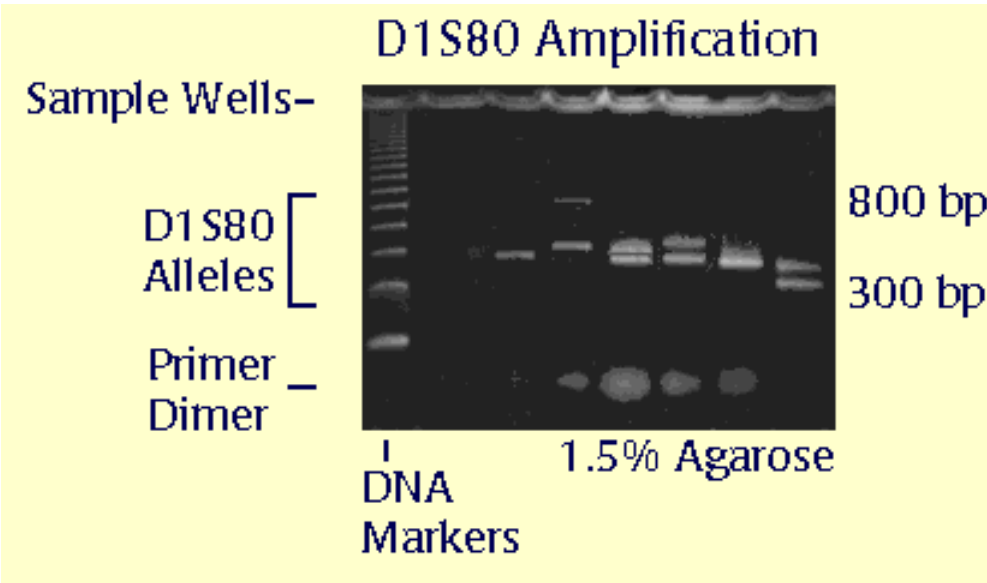
- This means ~1 change in a given site in every 1000 gametes
- Results in lots of variation between unrelated individuals in a population
- But mutation rate is low enough that within a family allele changes do not occur readily
- *Why are micro-satellite regions so mutable? What mechanisms of mutation could explain this?*

<http://www.cstl.nist.gov/div831/strbase/index.htm>

Using PCR to determine the genotype at the D1S80 STR locus



The D1S80 repeat unit is 16 base pairs (bp) in length and there are dozens of known alleles ranging from approximately 350 to 1,000 bp.



Alleles are distinguished by the size of PCR products generated with primers that match unique (single-copy) sequences that flank the repeat units

STR typically loci used for DNA profiling

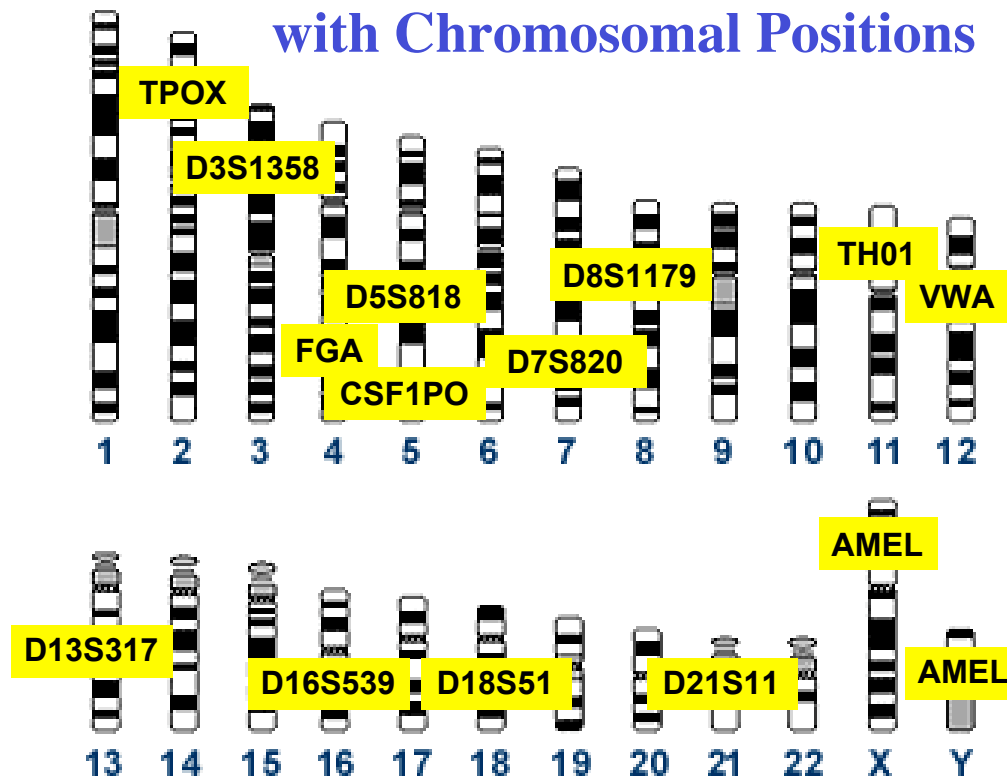
FBI CODIS = Combined DNA Index System

<http://www.fbi.gov/hq/lab/html/codis1.htm>

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

Typically *13 core CODIS* loci plus Amelogenin to determine sex of individual

13 CODIS Core STR Loci with Chromosomal Positions



From: Molecular Biology Principles and Practice

Table 1 Properties of the Loci Used for the CODIS Database

Locus Name	Chromosome	Repeat Motif*	Repeat Length (range) [†]	Number of Alleles Seen [‡]
CSF1PO	5	TAGA	5-16	20
FGA	4	CTTT	12.2-51.2	80
TH01	11	TCAT	3-14	20
TPOX	2	GAAT	4-16	15
VWA	12	[TCTG][TCTA]	10-25	28
D3S1358	3	[TCTG][TCTA]	8-21	24
D5S818	5	AGAT	7-18	15
D7S820	7	GATA	5-16	30
D8S1179	8	[TCTA][TCTG]	7-20	17
D13S317	13	TATC	5-16	17
D16S539	16	GATA	5-16	19
D18S51	18	AGAA	7-39.2	51
D21S11	21	[TCTA][TCTG]	12-41.2	82
Amelogenin [§]	X, Y	Not applicable		



Source: Adapted from J. M. Butler, *Forensic DNA Typing*, 2nd ed., Academic Press, 2006, p. 96.

*Brackets indicate alternating repeats.

[†]Repeat lengths observed in the human population. Partial or imperfect repeats are seen in some alleles.

[‡]Number of different alleles observed in the human population. Careful analysis of the same locus in many individuals is a prerequisite to its use in forensic DNA typing.

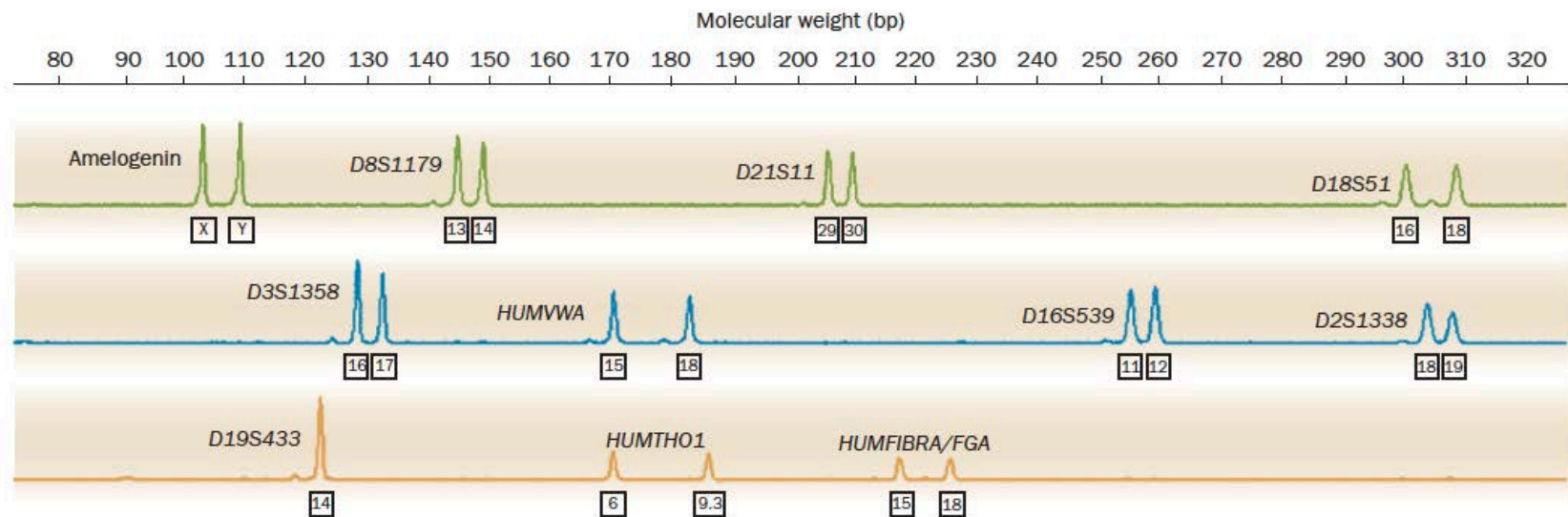
[§]Amelogenin is a gene, of slightly different size on the X and Y chromosomes, that is used to establish gender.

Complete listing of alleles for each site: <http://www.cstl.nist.gov/biotech/strbase/index.htm>

Complete listing of alleles for FGA: http://www.cstl.nist.gov/biotech/strbase/str_FGA.htm

Multiplex STR profile including X and Y specific products as analyzed by capillary gel electrophoresis

- Multiplex means that a single PCR reaction is performed with more than one set of primers -- 11 primer pairs in this case
- By carefully adjusting the positions of the primers relative to the repeat sequence, amplification products from different sites will not overlap during gel electrophoresis
- Each peak represents an allele and numbers below the peak indicate allele sizes in repeat units.
- The primer pairs are tagged with one of three fluorescent dyes --yellow, blue or green
- The STR profile is displayed in the green, blue and yellow channels of a four-color fluorescent system with the red channel being used for size markers (not shown).
- Amelogenin is found on both the X & the Y chromosomes. But, the Y linked copy results in a larger PCR product
- Standard number of PCR cycles used is 28 but 34 cycles are used when little DNA is available: *typically <100 pg or < 17diploid genomes!*



DNA fingerprint (or profile): the multi-locus pattern produced by the detection of genotype at a group of unlinked, highly polymorphic loci

Comparing two DNA fingerprints to determine if they represent the same person:

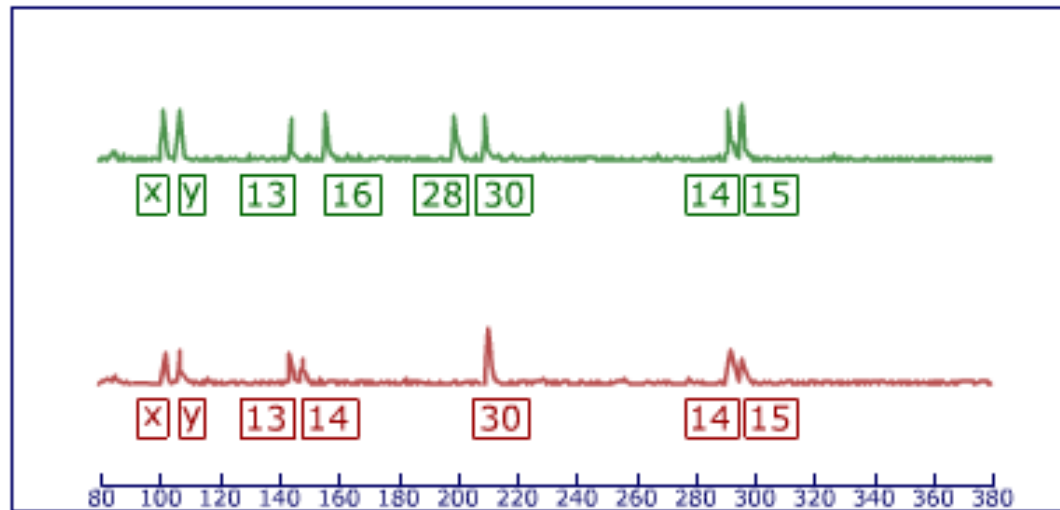
Exclusion: if the patterns do not match at every microsatellite locus tested, then the DNA must have come from different individuals

Inclusion: if the pattern of bands match at every locus, then the DNA **may have** come from the same source

Exclusion: if the patterns do not match at every micro/minisatellite locus tested, then the DNA must have come from different individuals

<http://www.dnai.org/d/>

click on Innocence project



Sample	Sex	D8S1179	D21S11	D18S51
Anderson	XY	13, 16	28, 30	14, 15
Crime Scene	XY	13, 14	30	14, 15

Inclusion: if the pattern of bands match at every locus, then the DNA *may have come from the same source*

Are the data sufficient to conclude identity between the suspect and the forensic sample?

If matches appear in multiple tests*, a statistical conclusion can be reached by calculating the *probability of a chance match*.

What sort of information do we need to calculate the probability of such a random match?

Inclusion: if the pattern of bands match at every locus, then the DNA *may have come from the same source:*

How many individuals in a given population carry these specific alleles: in other words, what is the overall frequency of such a genotype?

See web site for info on sample calculations

http://www.biology.arizona.edu/human_bio/activities/blackett2/act_probability1.html

Calculating genotype frequency from allele frequency?

The DIS80 repeat unit is 16 base pairs (bp) in length and there are dozens of known alleles

DIS80 Allele Frequencies in Unrelated Finns and U.S. Caucasians

ALLELE ^a (no. of core units)	FREQUENCY IN	
	Unrelated Finns (<i>n</i> = 140)	U.S. Caucasian. (<i>n</i> = 94)
1 (18)307	.293
2 (19)011	.011
3 (20)032	.021
4 (21)018	.032
5 (22)014	.043
6 (23)014	.016
7 (24)311	.335
8 (25)075	.037
9 (26)011	.016
10 (27)007	.000
11 (28)068	.059
12 (29)032	.059
13 (30)043	.016
14 (31)079	.043
20 (37)007	.000

^a Nomenclature is that of Budowle et al. (1991).

Calculating genotype frequency from allele frequency

p = frequency of A allele

q = frequency of a allele

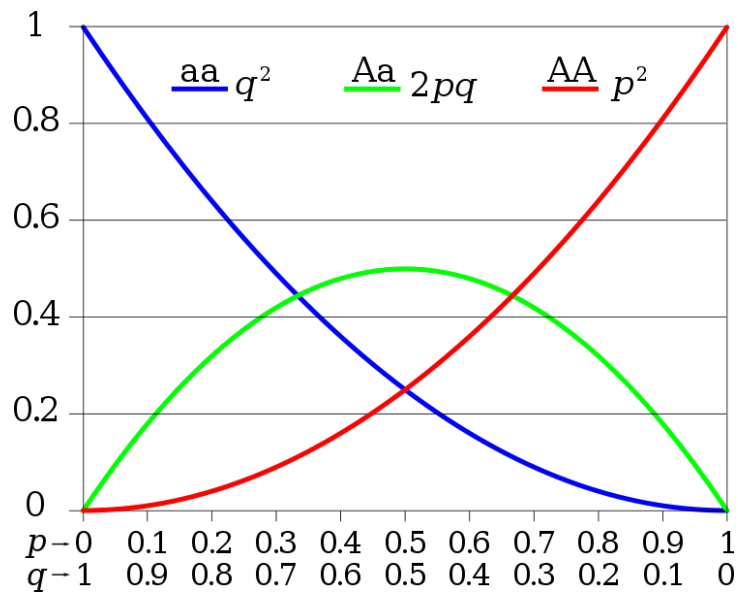
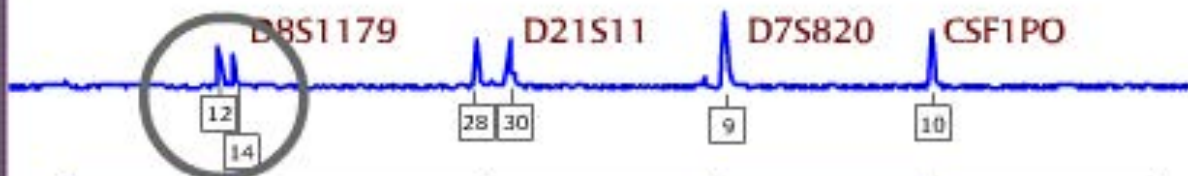


Table 1: Punnett square for Hardy-Weinberg equilibrium

		Females	
		A (p)	a (q)
Males	A (p)	AA (p^2)	Aa (pq)
	a (q)	Aa (pq)	aa (q^2)

Logic can be extended to loci with many alleles

Using the Hardy-Weinberg equation, let's calculate the frequency of a Caucasian being heterozygous for D8S1179 for alleles 12 and 14.



D8S1179 (Alleles)	Asians (N = 196)	African American (N = 210)	Caucasians (N = 203)	Hispanic (N = 209)
9	0.0024	0.0056	0.0102	0.0025
10	0.0119	0.0250	0.1020	0.0936
11	0.1214	0.0361	0.0587	0.0616
12	0.2905	0.1083	0.1454	0.1207
13	0.3071	0.2222	0.3393	0.3251
14	0.2000	0.3333	0.2015	0.2463
15	0.0548	0.2139	0.1097	0.1158
16	0.0048	0.0444	0.0128	0.0246
17	0.0048	0.0000	0.0026	0.0074

$$2pq = 2(0.1454)(0.2015) = 0.0586$$

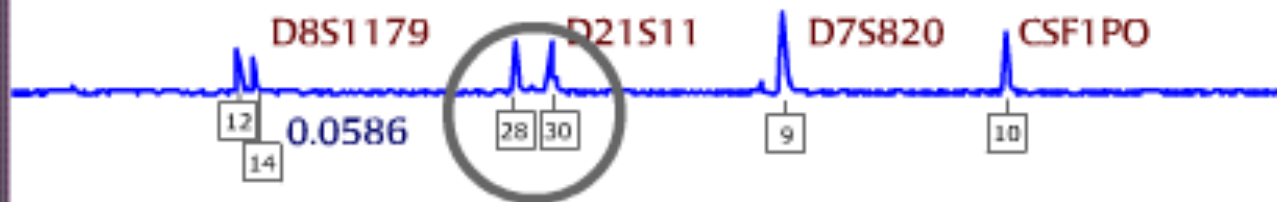
Table adapted from:

Budowle et. al., 2001. Journal of Forensic Science 46(3): 453-489.

~ 1 in every 17 Caucasians has this genotype at the D8S1179 STR locus

<http://www.dnai.org/d/>

Similarly, the frequency can be calculated for D21S11.



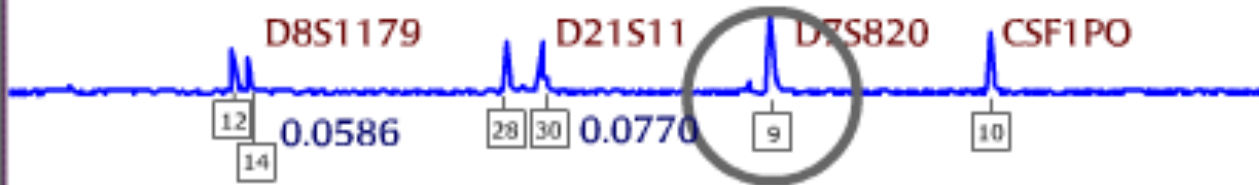
D21S11 (Alleles)	Asians (N = 196)	African American (N = 180)	Caucasians (N = 196)	Hispanic (N = 203)
24	0.0000	0.0000	0.0000	0.0000
25	0.0000	0.0000	0.0000	0.0000
26	0.0026	0.0028	0.0000	0.0000
27	0.0102	0.0615	0.0459	0.0099
28	0.0969	0.2151	0.1658	0.0690
29	0.2474	0.1899	0.1811	0.2044
30	0.2015	0.1188	0.2321	0.3300
31	0.1071	0.0922	0.0714	0.0690
32	0.0408	0.0084	drag	0.0123
33	0.0051	0.0084	0.0000	0.0000
34	0.0000	0.0000	0.0000	0.0000

$$2pq = 2(0.1658)(0.2321) = 0.0770$$

Table adapted from:
Budowle et. al., 2001. Journal of Forensic Science 46(3): 453-489.

~ 1 in every 13
Caucasians has this
genotype at the
D21S11 STR locus

Again, using the Hardy-Weinberg equation, let's calculate the frequency of a Caucasian being homozygous for D7S820 for allele 9.



D7S820 (Alleles)	Asians (N = 196)	African American (N = 180)	Caucasians (N = 196)	Hispanic (N = 203)
6	0.0000	0.0000	0.0025	0.0024
7	0.0102	0.0071	0.0172	0.0215
8	0.1684	0.1738	0.1626	0.0981
9	0.0816	0.1571	0.1478	0.0478
10	0.2168	0.3364	0.2906	0.3062
11	0.3138	0.2238	0.2020	0.2895
12	0.1760	0.0905	0.1404	0.1914
13	0.0255	0.0190	0.0296	0.0383
14	0.0077	0.0048	0.0074	0.0048

$$p^2 = (0.1478)(0.1478) = 0.0218$$

Table adapted from:

Budowle et. al., 2001. Journal of Forensic Science 46(3): 453-489.

~ 1 in every 46
Caucasians has this
genotype at the
D7S820 STR locus

These sites are on different chromosomes and therefore assorting independently. Use the product rule to determine the overall probability of this DNA profile.

Similarly the calculation can be done for CSF1PO. For this DNA profile, the random match probability is the product of the individual frequencies of each of the four STRs.



$$\begin{aligned} \text{RANDOM MATCH PROBABILITY} &= \text{D8S1179} \times \text{D21S11} \times \text{D7S820} \times \text{CSF1PO} \\ &= (0.0586) \times (0.0770) \times (0.0218) \times (0.0644) \\ &= 0.000006334 \\ &= 6.334 \times 10^{-6} \end{aligned}$$

$$6.3 \times 10^{-6} = 1 \text{ in } 160,000$$

CONSIDER ALSO – this calculation would be extended to include all 13 CODIS loci

These calculations appear deceptively trivial, but they are not.

- Initially there was tremendous contention and disagreement among population geneticists about how these calculations should be done
- Such calculations are only valid and useful to determine the probability of a chance match *if there is reliable information on allele frequency and if the population database used is appropriate*

Validity of conclusions drawn from DNA fingerprinting test depends on

- the number of loci tested
- the number of possible allele variations at each site examined
- the integrity of the population database used to determine allele frequencies -- is it large enough and does it reflect differences in ethnic and racial groups?
- The quality of the laboratory work

DNA evidence must always be considered within the framework of other evidence of many types

Consider the case of Josiah Sutton

- In 1999 Josiah Sutton, then 16 years old, was sentenced to 25 years in prison for rape.
- The evidence seemed airtight: the victim had spotted Sutton walking down a Houston street five days after her attack, and crime lab analysis from Houston's police department showed his DNA was an exact match with semen from the crime.
- The DNA evidence was convincing: the analyst testified that the probability of a chance match was 1 in 694,000.
- But a later analysis showed that one in 8 black men would have had a similar match.
- Keep in mind that if a DNA profile exhibits a series of rare alleles, a chance match is much less likely than if the alleles are common in a given population

Eye-witness testimony is notoriously unreliable. According to the Innocence Project, 75 % of wrongful convictions are due to eyewitness mistakes

HVS, hypervariable site; Mb =megabase; mtDNA, mitochondrial DNA; SGM, second generation multiplex; STR, short tandem repeat.

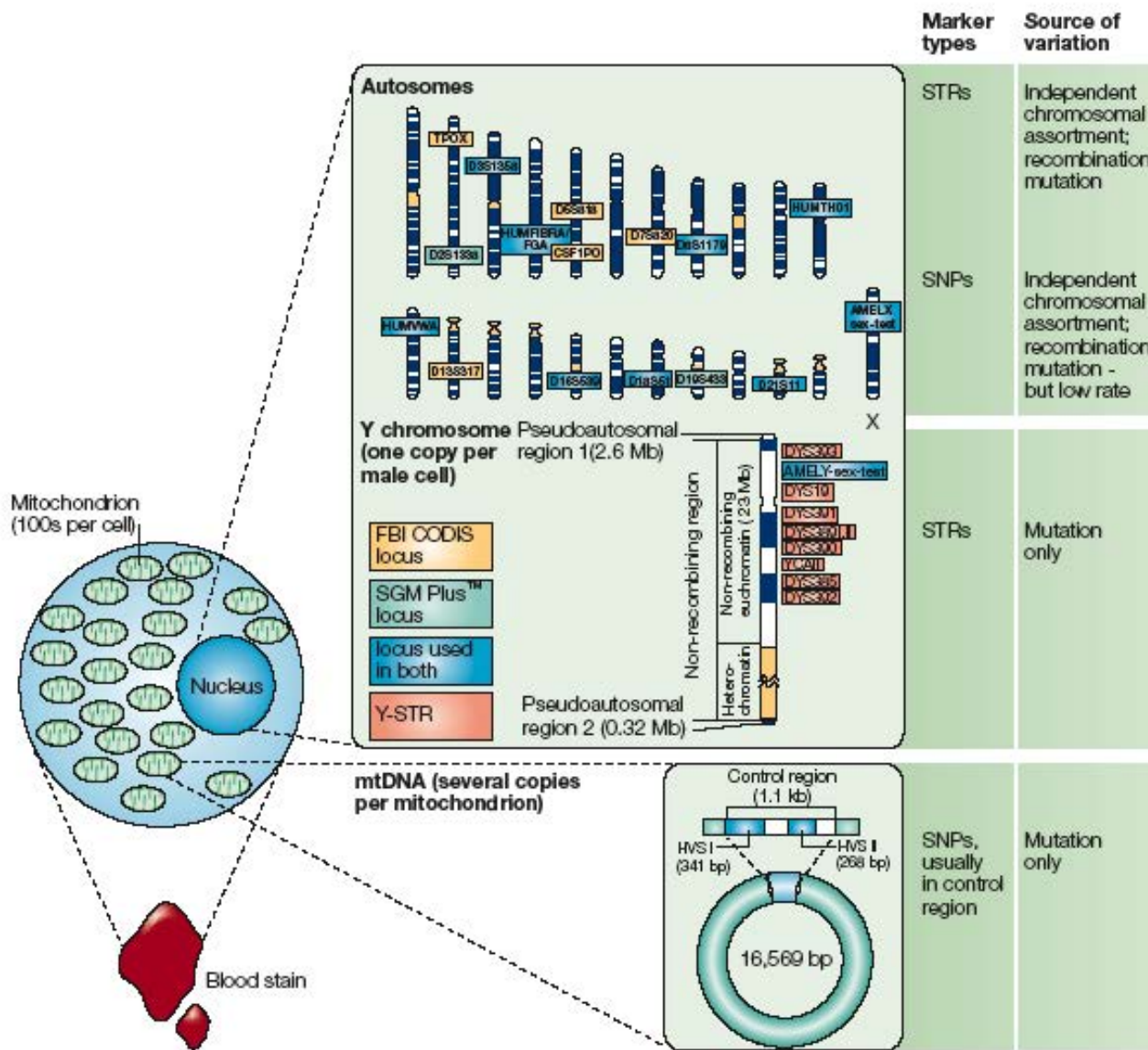
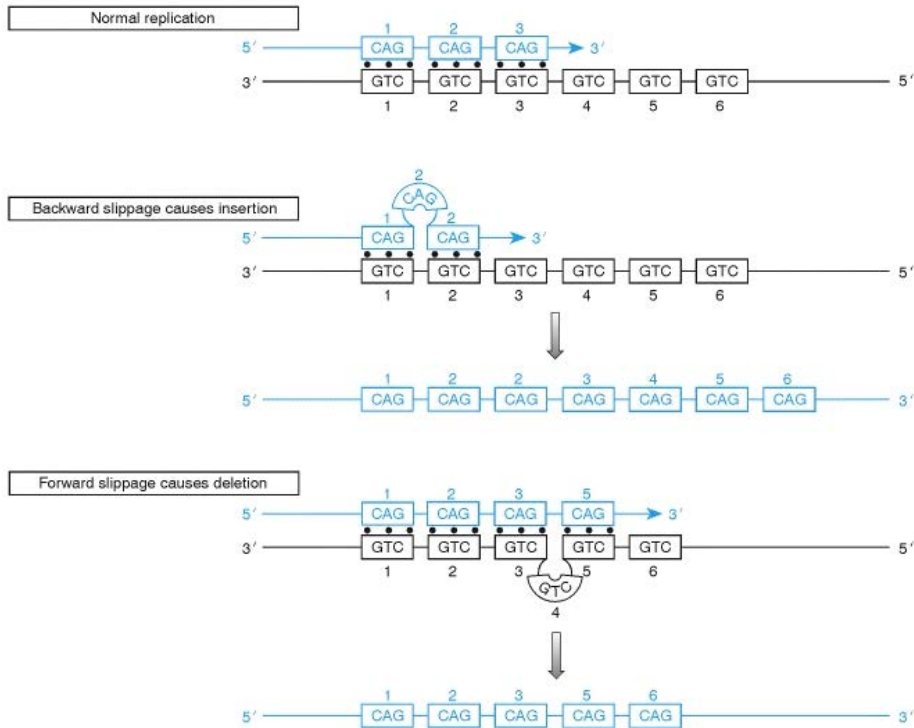


Figure 1 | Sources of human genetic variation used in forensic analysis. Further details of the properties of different loci can be found in the text.

Heteroplasmy describes the presence of two or more different mitochondrial DNA sequences in the same cell, or individual. FBI CODIS, US Federal Bureau of Investigation Combined DNA Index System;



Short tandem repeats are thought to be particularly prone to slipped strand mispairing, i.e. mispairing of the complementary DNA strands of a single DNA double helix. The examples show how slipped strand mispairing can occur during replication, with the lower strand representing a parental DNA strand and the upper blue strand representing the newly synthesized complementary strand. In such cases, slippage involves a region of nonpairing (shown as a bubble) containing one or more repeats of the newly synthesized strand (backward slippage) or of the parental strand (forward slippage), causing, respectively, an insertion or a deletion on the newly synthesized strand.