

# Basic Principles of Forensic Molecular Biology and Genetics

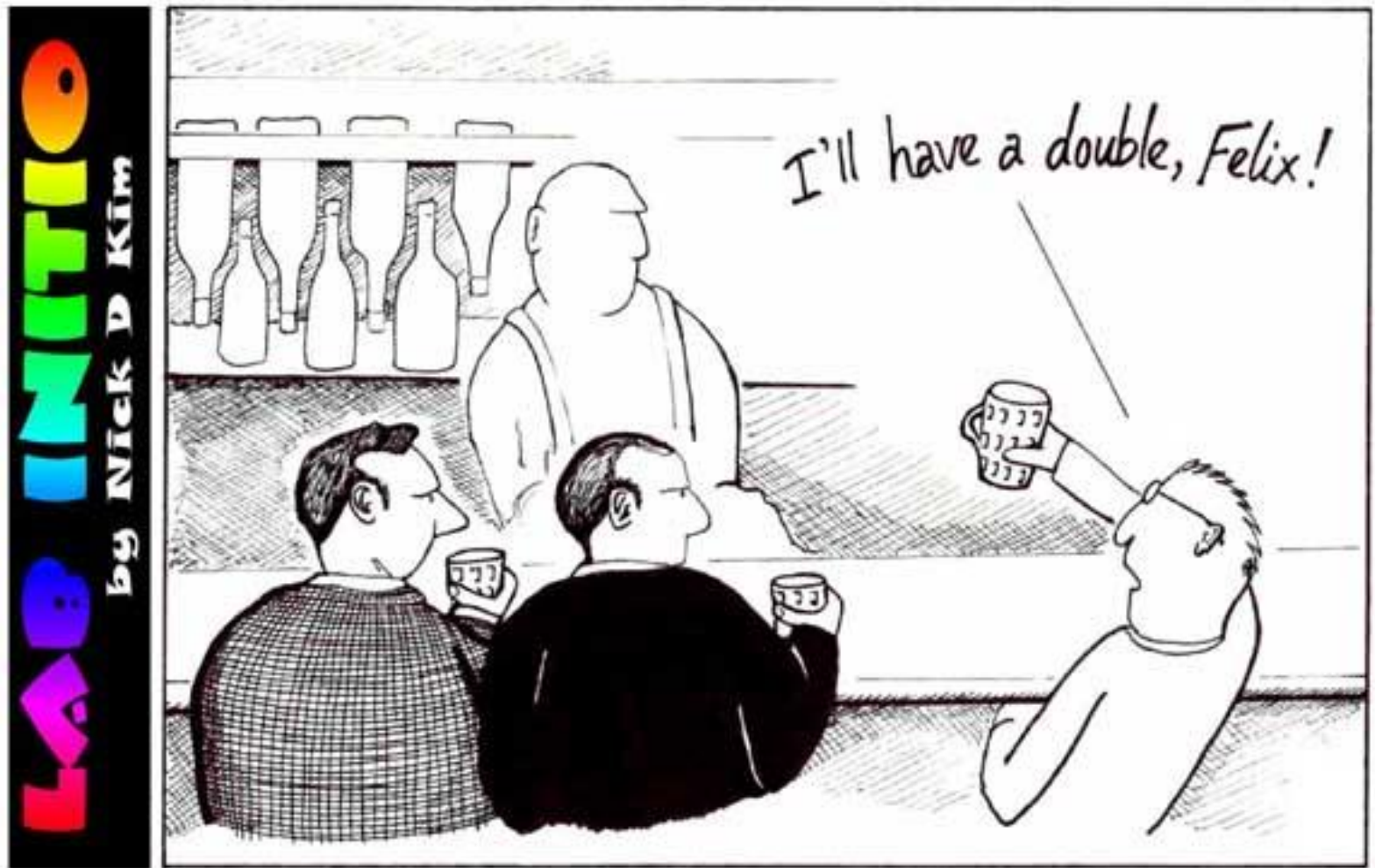


# Y-Chromosome



Biology

# Inspiration Comes Easier to Some



Cambridge, 1953. Shortly before discovering the structure of DNA, Watson and Crick, depressed by their lack of progress, visit the local pub.

# Human Genome

- Size: 3200 Mb
  - 5% coding, 50% repeat sequences
- Sequence coverage: > 94 % represented in combined high quality finished sequence in publicly available databases (>99 % of euchromatic portion)
  - Gaps heterochromatic ~200 Mb and euchromatic ~24.4 Mb
- Chromosome sizes:
  - autosomes: 279 Mb (#1) to 45 Mb (#21)
  - sex chromosomes: 163 Mb (X), 60 Mb (Y)

# human chromosome complement (haploid)

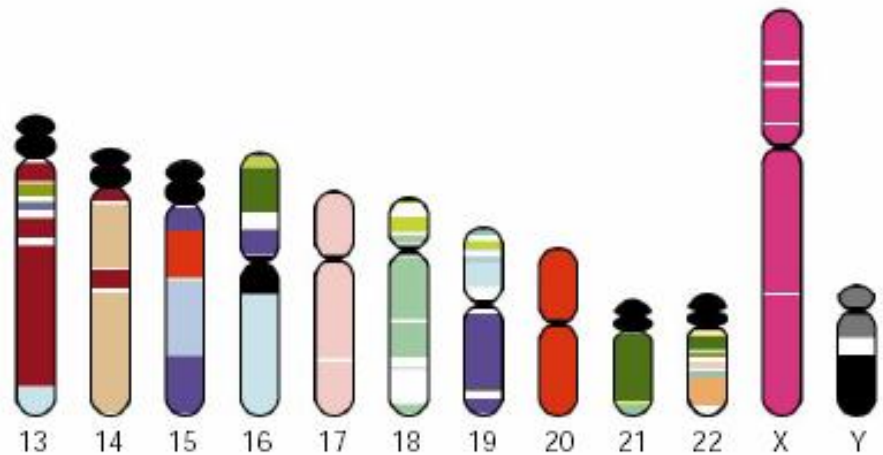
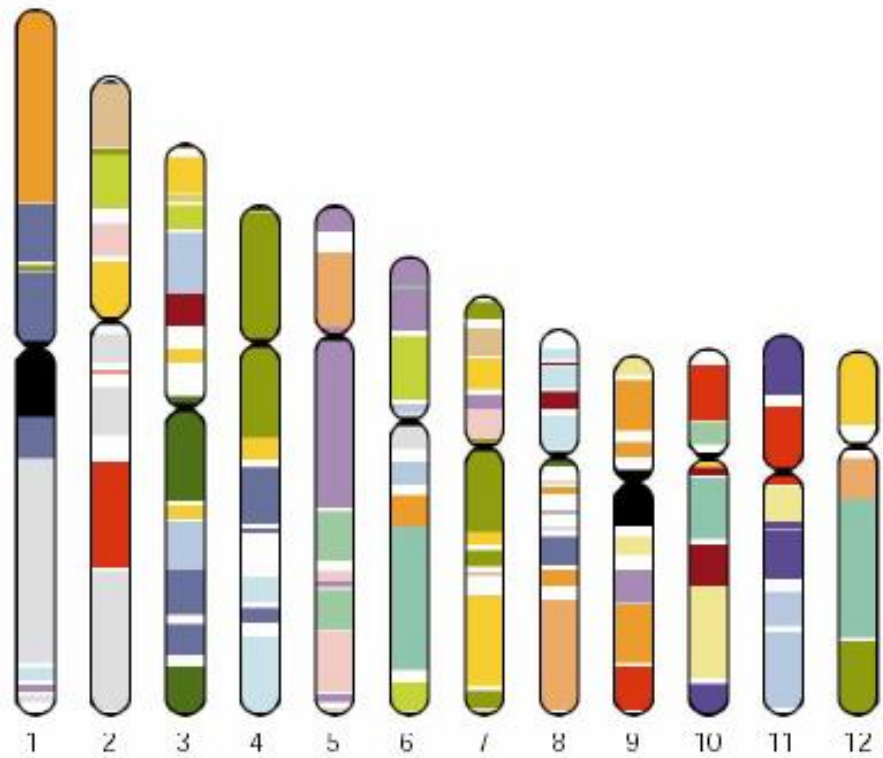
## chromosome size

(Mb)

Genome = 3289

Y = 60 (1.8 %)

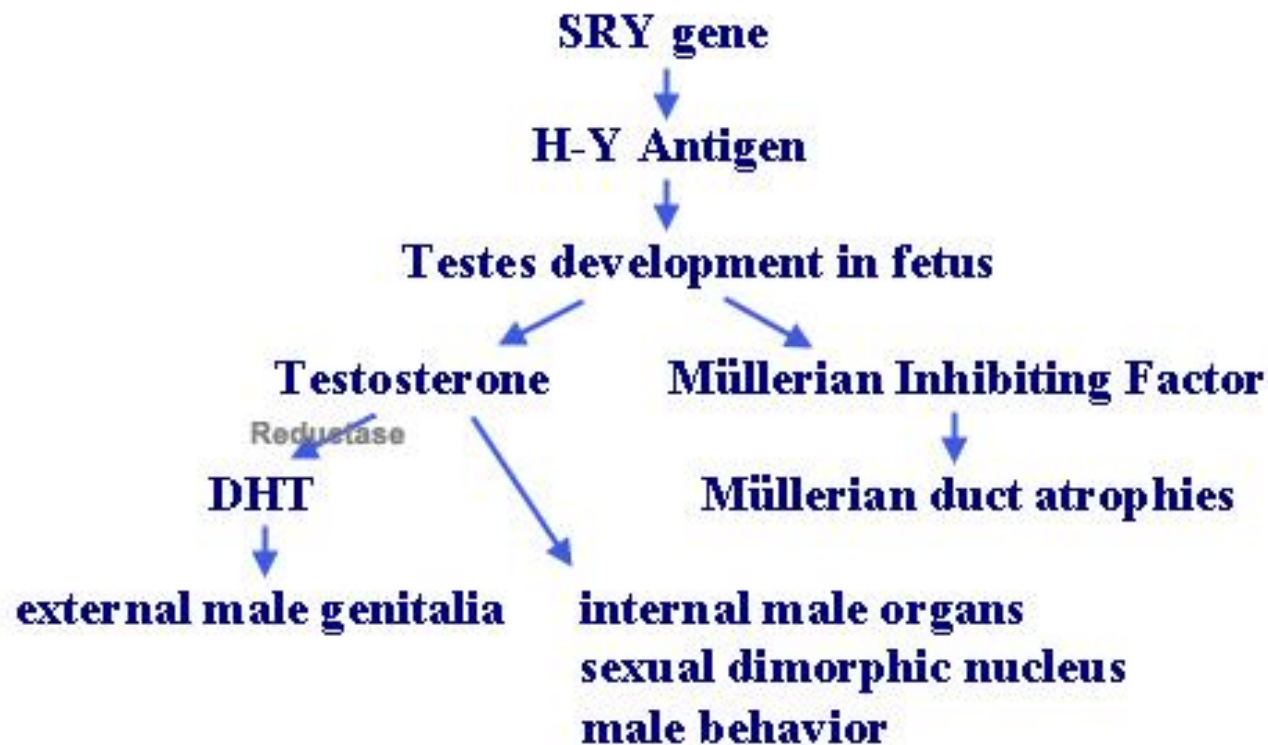
X = 163 (5%)



# Function of Y Chromosome?

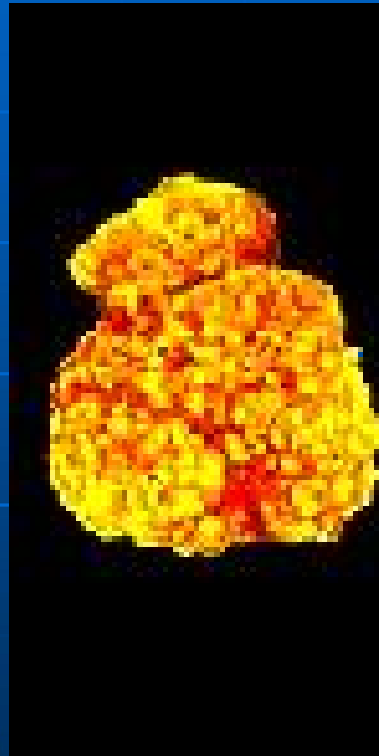
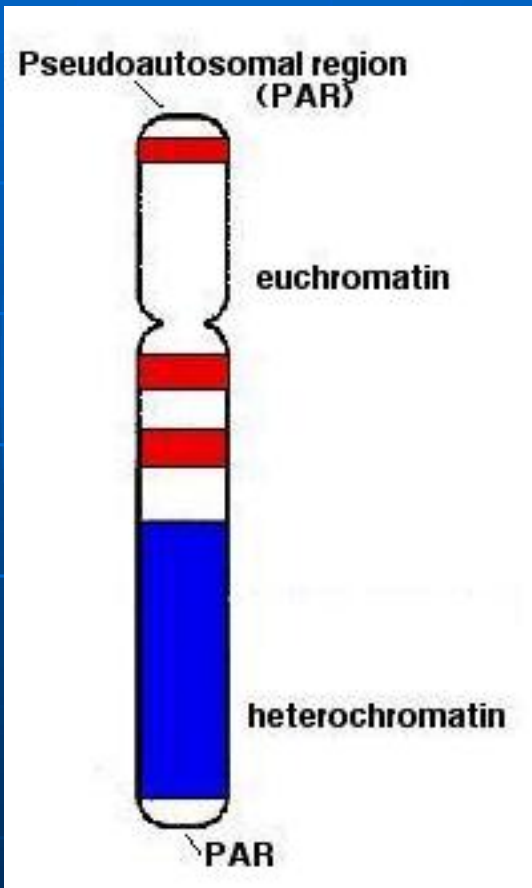
- Aneuploidies for the X and Y
  - 47, XXY (Klinefelter synd.) → males
    - 48, XXXY; 49, XXXXY; 50, XXXXXY → males
  - 45, XO (Turner synd.) → females
  - 47, XXX (triple-X karyotype) → 'normal' female
  - 47, XYY karyotype → 'normal' male
- Sex Reversed Humans
  - XY → female (Y minus TDF)
  - XX → male (X plus TDF)

## Sex Determination in Mammals



Female system is  
always the default

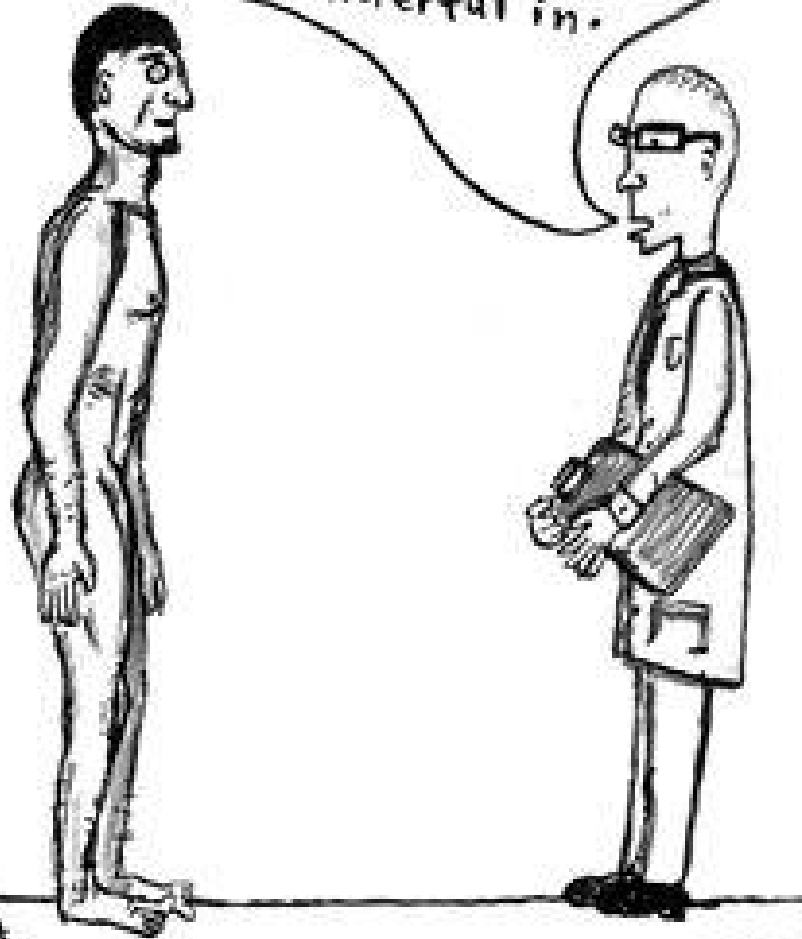
# Classic View of Y-Chromosome



- TDF master gene
- patrilineal inheritance
- no recombination in NRY
- recombination in PAR
- junk-rich, gene poor

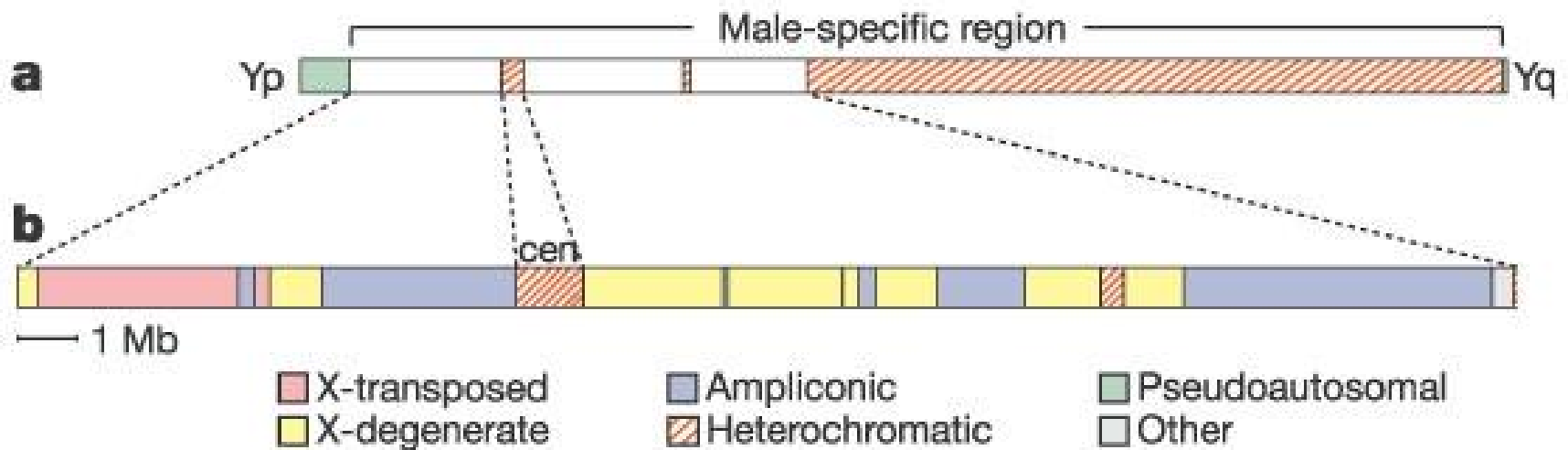


the bad news Mr. Sapien  
is your Y chromosome is  
shrinking at an alarming rate,  
the good news is I have a  
dress I'm sure you'll look  
wonderful in.



parnick

# Y Chromosome NRY is a Mosaic of Discrete Sequence Classes



# Sequence Classes of the Y Chromosome NRY

Sequence class	Defining characteristics	Evolutionary origins	Distribution	Aggregate length (Mb)	No. of coding genes	No. of non-coding transcription units
X-transposed	99% identity to X	Single transposition from X	2 blocks on Yp	3.4	2	0
X-degenerate	Single-copy gene or pseudogene homologues of X-linked genes	Relics of ancient autosomes from which X and Y evolved	8 blocks on Yp and Yq	8.6	16, most expressed widely	4
Ampliconic	Lengthy similarity to other MSY sequences	Acquired from diverse sources, then amplified	7 blocks on Yp and Yq	10.2	60 (in 9 families), expressed mainly or only in testes	74 (9 single-copy; 65 in 15 families), expressed mainly or only in testes

Table 2 **MSY genes and gene families demonstrated or hypothesized to encode proteins**

MSY sequence class	Gene symbol	Gene name	Number of copies†	Tissue expression	X-linked homologue	Autosomal homologue	
X-transposed	<i>TGIF2LY*</i>	TGF (beta)-induced transcription factor 2-like Y	1	Testis	<i>TGIF2LX</i>	–	
	<i>PCDH11Y</i>	Protocadherin 11 Y	1	Fetal brain, brain	<i>PCDH11X</i>	–	
Total			2				
X-degenerate	<i>SRY</i>	Sex determining region Y	1	Predominantly testis	<i>SOX3</i>	–	
	<i>RPS4Y1</i>	Ribosomal protein S4 Y isoform 1	1	Ubiquitous	<i>RPS4X</i>	–	
	<i>ZFY</i>	Zinc finger Y	1	Ubiquitous	<i>ZFX</i>	–	
	<i>AMELY</i>	Amelogenin Y	1	Teeth	<i>AMELX</i>	–	
	<i>TBL1Y*</i>	Transducin (beta)-like 1 protein Y	1	Fetal brain, prostate	<i>TBL1X</i>	–	
	<i>PRKY</i>	Protein kinase Y	1	Ubiquitous	<i>PRKX</i>	–	
	<i>USP9Y</i>	Ubiquitin-specific protease 9 Y	1	Ubiquitous	<i>USP9X</i>	–	
	<i>DBY</i>	Dead box Y	1	Ubiquitous	<i>DBX</i>	–	
	<i>UTY</i>	Ubiquitous TPR motif Y	1	Ubiquitous	<i>UTX</i>	–	
	<i>TMSB4Y</i>	Thymosin (beta)-4 Y	1	Ubiquitous	<i>TMSB4X</i>	–	
	<i>NLGN4Y</i>	Neurologin 4 isoform Y	1	Fetal brain, brain, prostate, testis	<i>NLGN4X</i>	–	
	<i>CYorf15A*</i>	Chromosome Y open reading frame 15A	1	Ubiquitous	<i>CXorf15</i>	–	
	<i>CYorf15B*</i>	Chromosome Y open reading frame 15B	1	Ubiquitous	<i>CXorf15</i>	–	
	<i>SMCY</i>	SMC (mouse) homologue, Y	1	Ubiquitous	<i>SMCX</i>	–	
	<i>EIF1AY</i>	Translation initiation factor 1A Y	1	Ubiquitous	<i>EIF1AX</i>	–	
	<i>RPS4Y2*</i>	Ribosomal protein S4 Y isoform 2	1	Ubiquitous	<i>RPS4X</i>	–	
	Total			16			
	Amplificonic	<i>TSPY</i>	Testis-specific protein Y	~35	Testis	–	–
		<i>VCY</i>	Variable charge Y	2	Testis	<i>VCX</i>	–
		<i>XKRY</i>	XK related Y	2	Testis	–	–
<i>CDY</i>		Chromodomain Y	4	Testis	–	<i>CDYL</i>	
<i>HSFY*</i>		Heat shock transcription factor Y	2	Testis	–	–	
<i>RBMY</i>		RNA-binding motif Y	6	Testis	<i>RBMX</i>	–	
<i>PRY</i>		PTP-BL related Y	2	Testis	–	–	
<i>BPY2</i>		Basic protein Y 2	3	Testis	–	–	
<i>DAZ</i>		Deleted in azoospermia	4	Testis	–	<i>DAZL</i>	
Total				~60			
Grand total			~78				

Table 3 **MSY palindromes**

Palindrome	Arm length (kb)	Arm-to-arm identity (%)	Spacer length (kb)	Palindrome span (kb)
P1	1,450	99.97	2.1	2,902
P1.1*	9.9	99.95	3.9	24
P1.2*	9.9	99.95	3.9	24
P2	122	99.97	2.1	246
P3	283	99.94	170	736
P4	190	99.98	40	419
P5	496	99.98	3.5	996
P6	110	99.97	46	266
P7	8.7	99.97	12.6	30
P8	36	99.997	3.4	75
Total span (kb)				5,670

\* Palindromes P1.1 and P1.2 are located within, respectively, the distal and the proximal arms of palindrome P1 (Fig. 2).

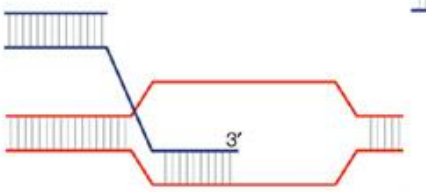
**a** DNA double-strand break



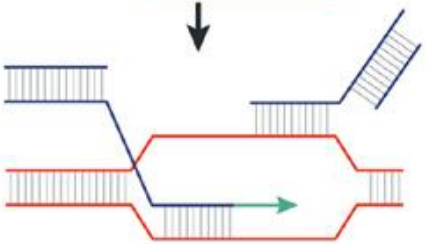
**b** Resection



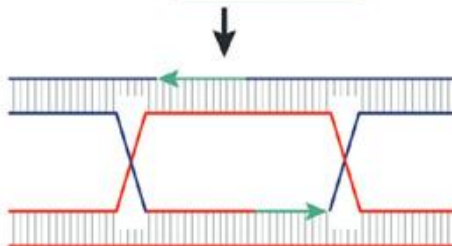
**c** Invasion of the first end into a homologous duplex



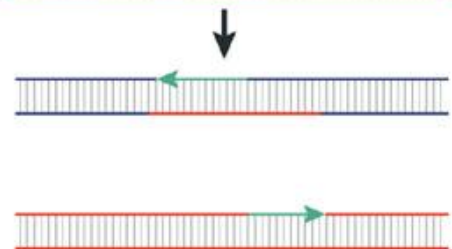
**d** Second-end capture



**e** Continued strand exchange and DNA synthesis

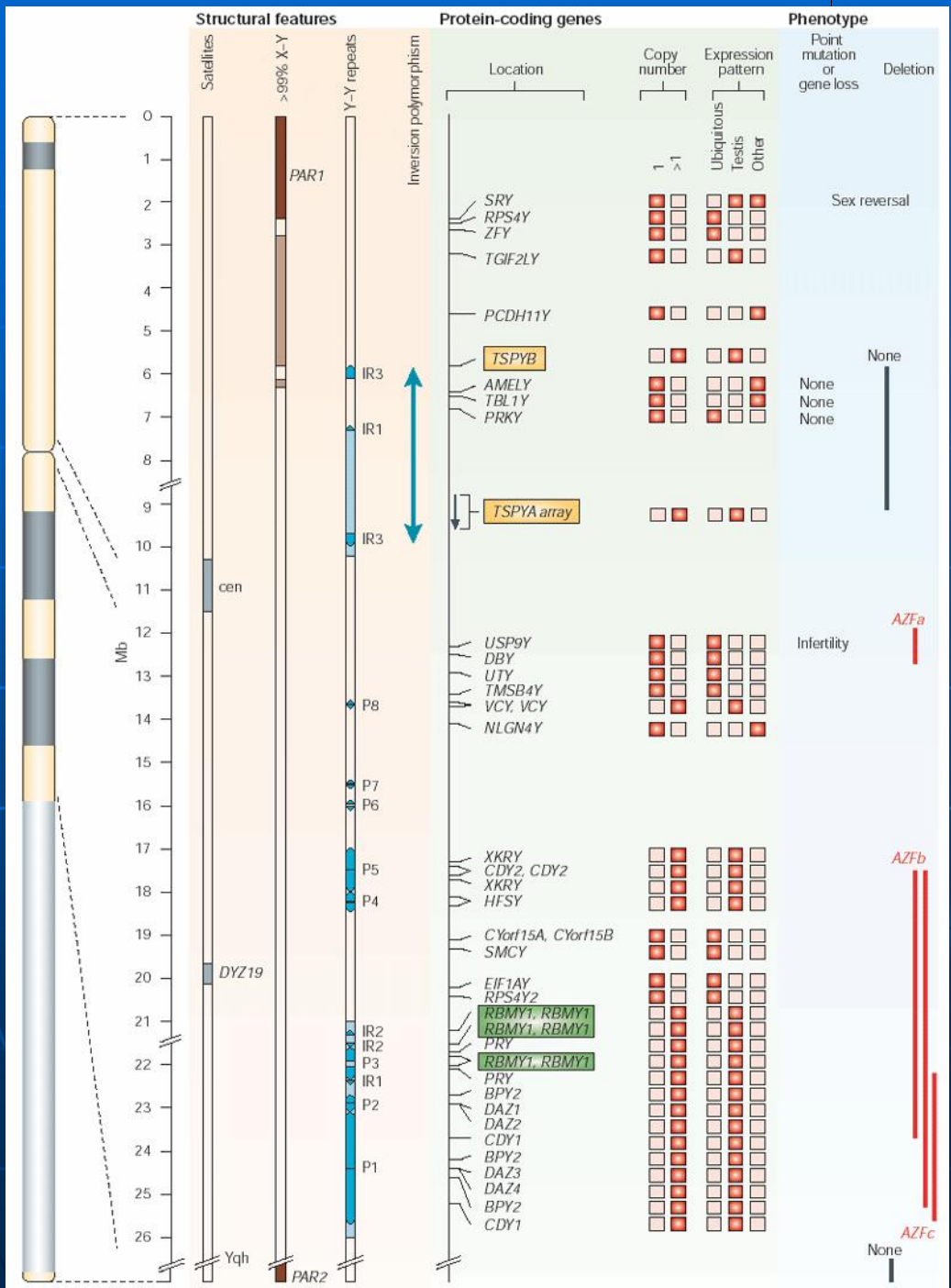


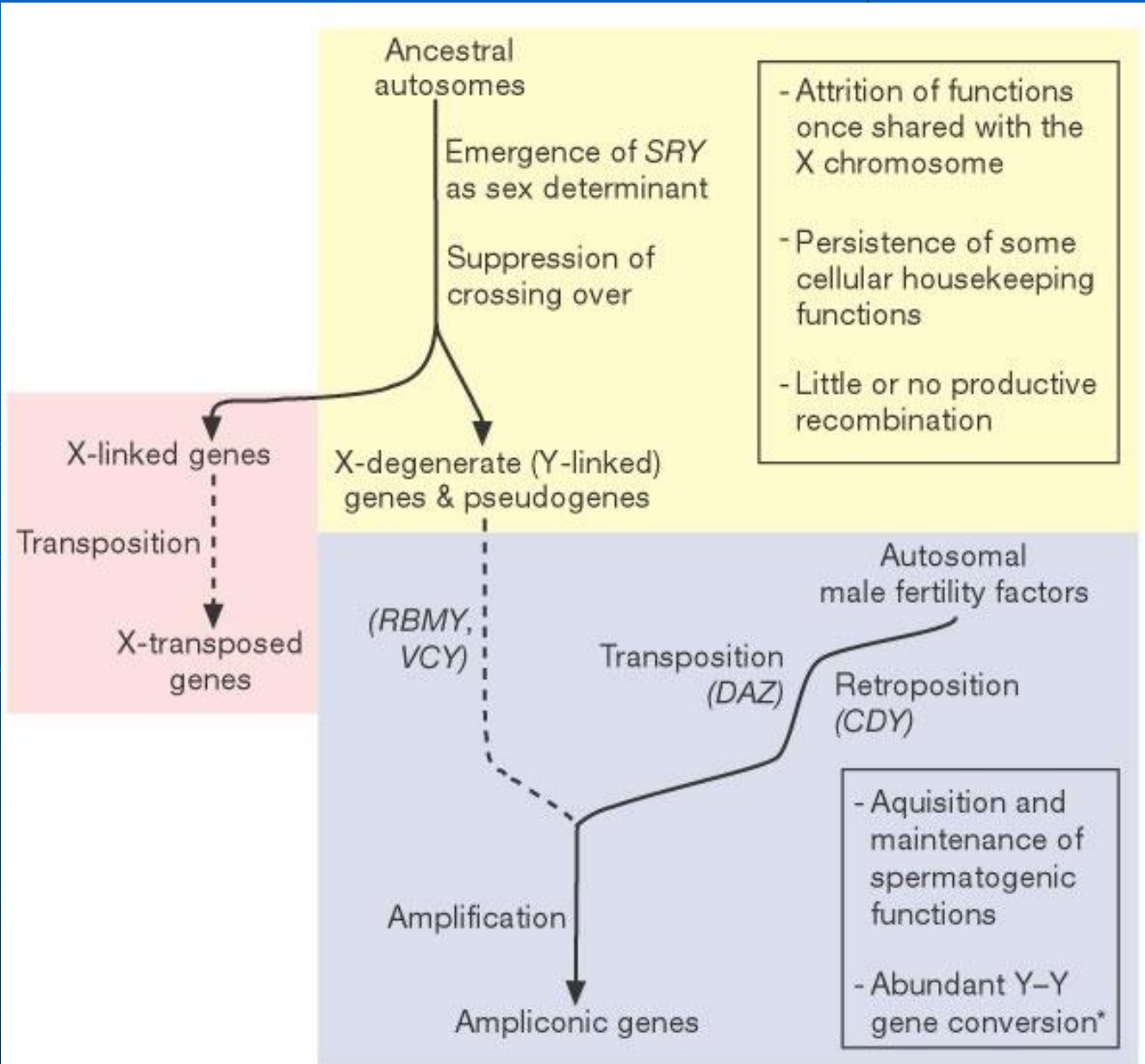
**f** Holiday junction resolution



gene conversion:

nonreciprocal transfer of genetic information from one homologous chromosome to another (i.e. no crossing over)





## evolution of the NRY sequence classes



# BRAINS



JAVA MAN

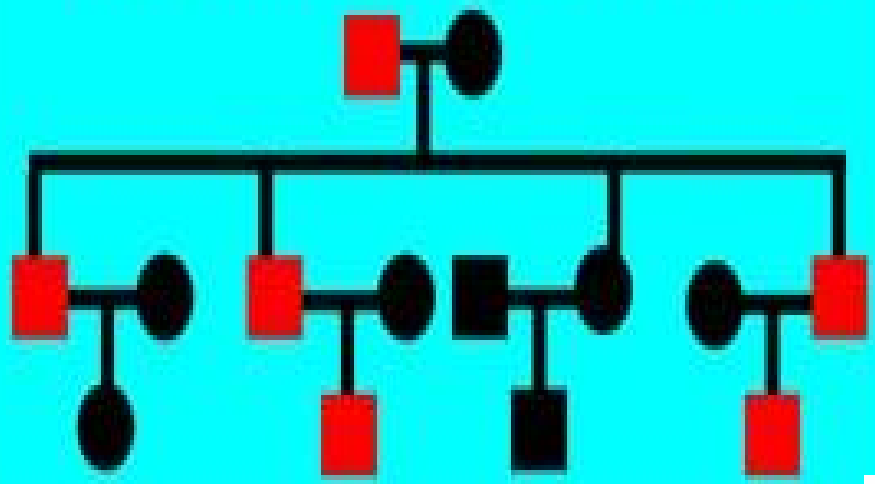


NEANDERTHAL MAN



MODERN MAN

# Y-chromosome Inheritance Pattern



# Y-Chromosome Inheritance

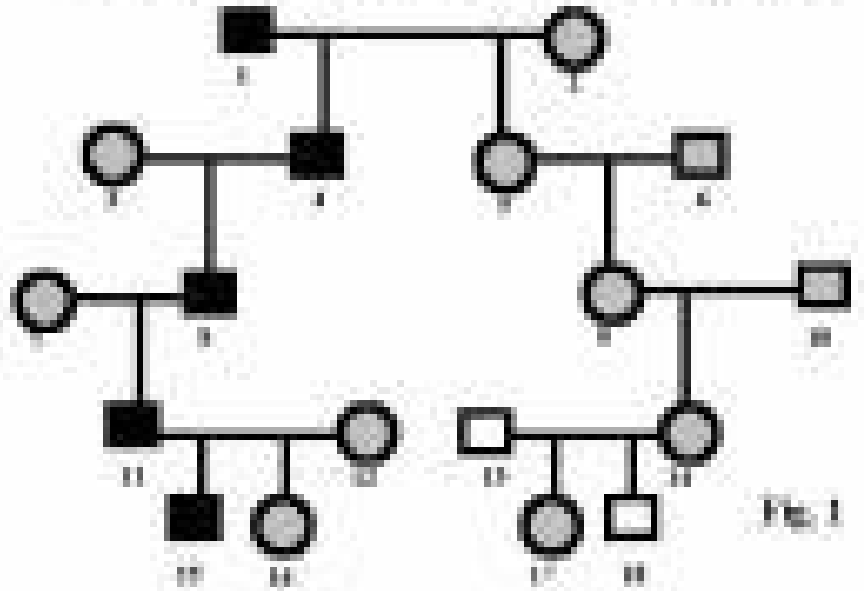


Fig. 1

# Y Chromosome Polymorphisms

- ~ 200 binary polymorphisms (Y-SNPs) characterized
- > 200 microsatellites (Y-STRs) characterized
- 1 minisatellite (MSY1)
  - A-T rich

# Mutation Rates: STRs

## ■ Autosomes

- $1.2 \times 10^{-3}$  (Weber and Wong, 1993)
- $2.1 \times 10^{-3}$  (Brinkmann et al, 1998)
- $2.7 \times 10^{-3}$  (Henke and Henke, 1999)
- $0.6 \times 10^{-3}$  (Sajantila et al, 1999)

## ■ Y-chromosome

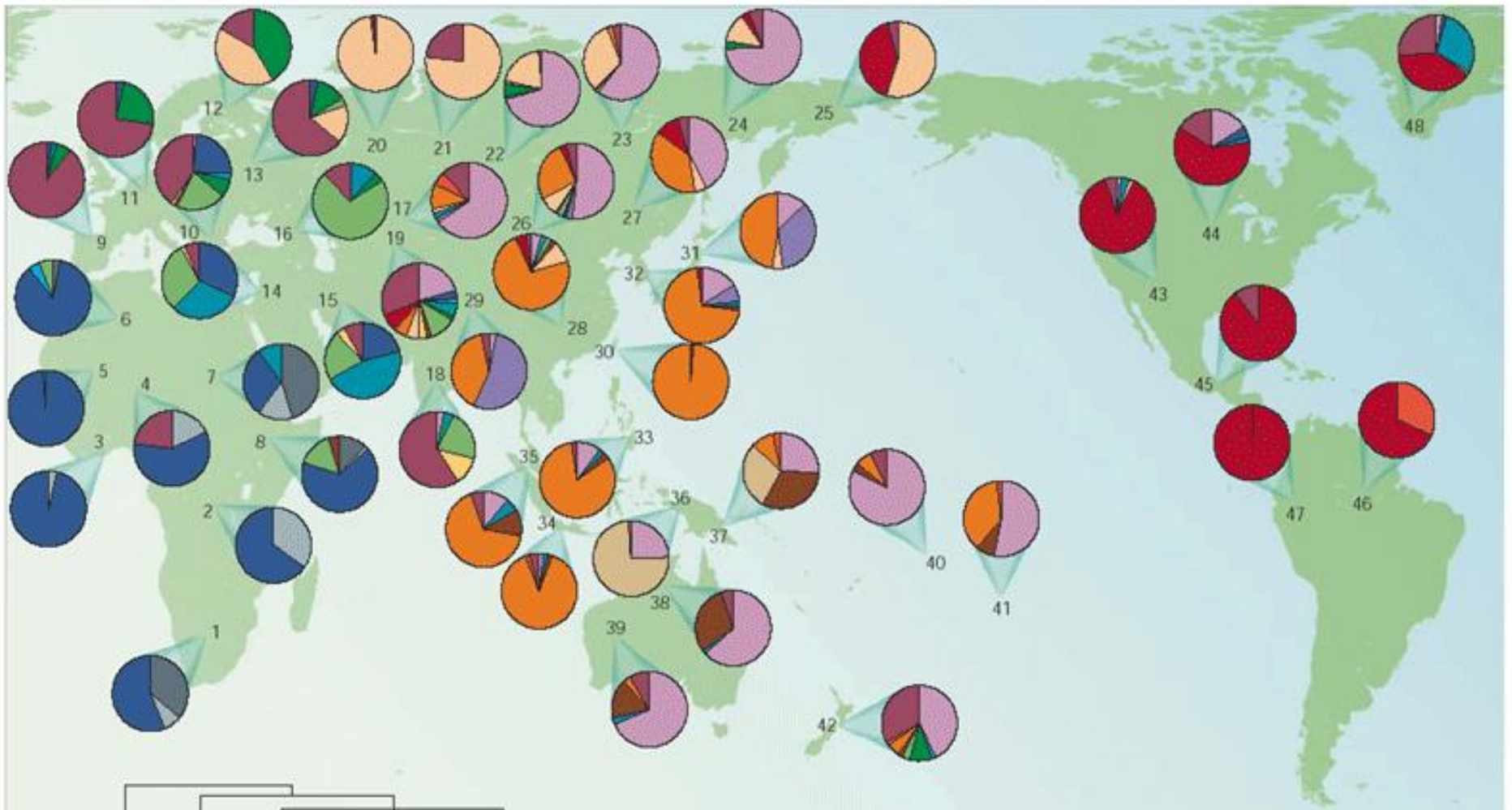
- $2.8 \times 10^{-3}$  (Kayser et al, 2000)
- $2.0 \times 10^{-3}$  (Heyer et al, 1997)
- $3.09 \times 10^{-3}$  (Bianchi et al, 1998)

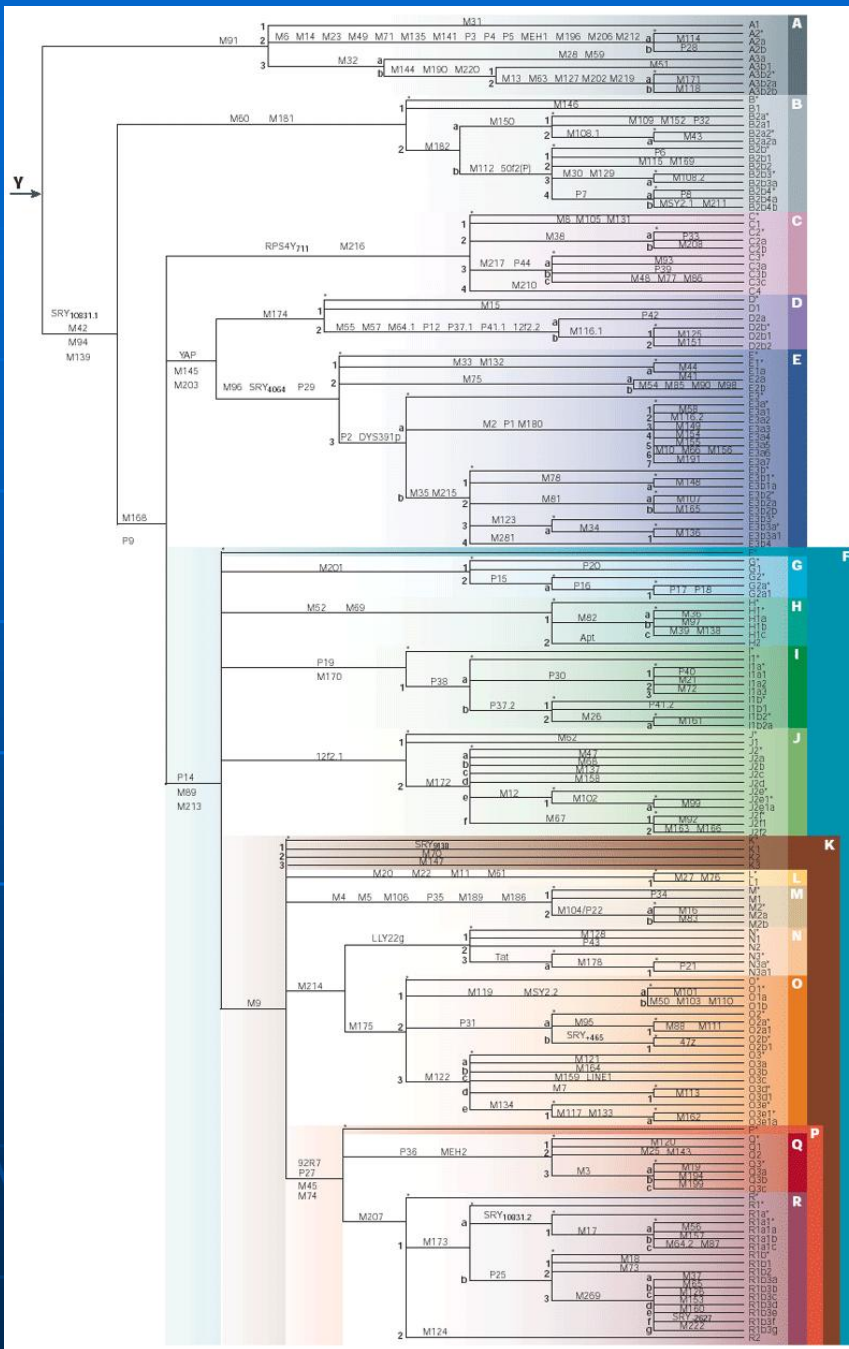
# Mutation Rates: SNPs

- $\sim 2 \times 10^{-8}$  per base per generation

# Population Differentiation

- Effective population size of Y chromosome is  $1/4$  of autosome or  $1/3$  of X
  - lower sequence diversity on Y
  - more susceptible to genetic drift
    - random changes in frequency of haplotypes due to sampling from one generation to next
    - accelerates differences between populations
- Geographical clustering due to patrilocal behavior of men
  - women move closer to man's birthplace
  - local geographical differentiation enhanced

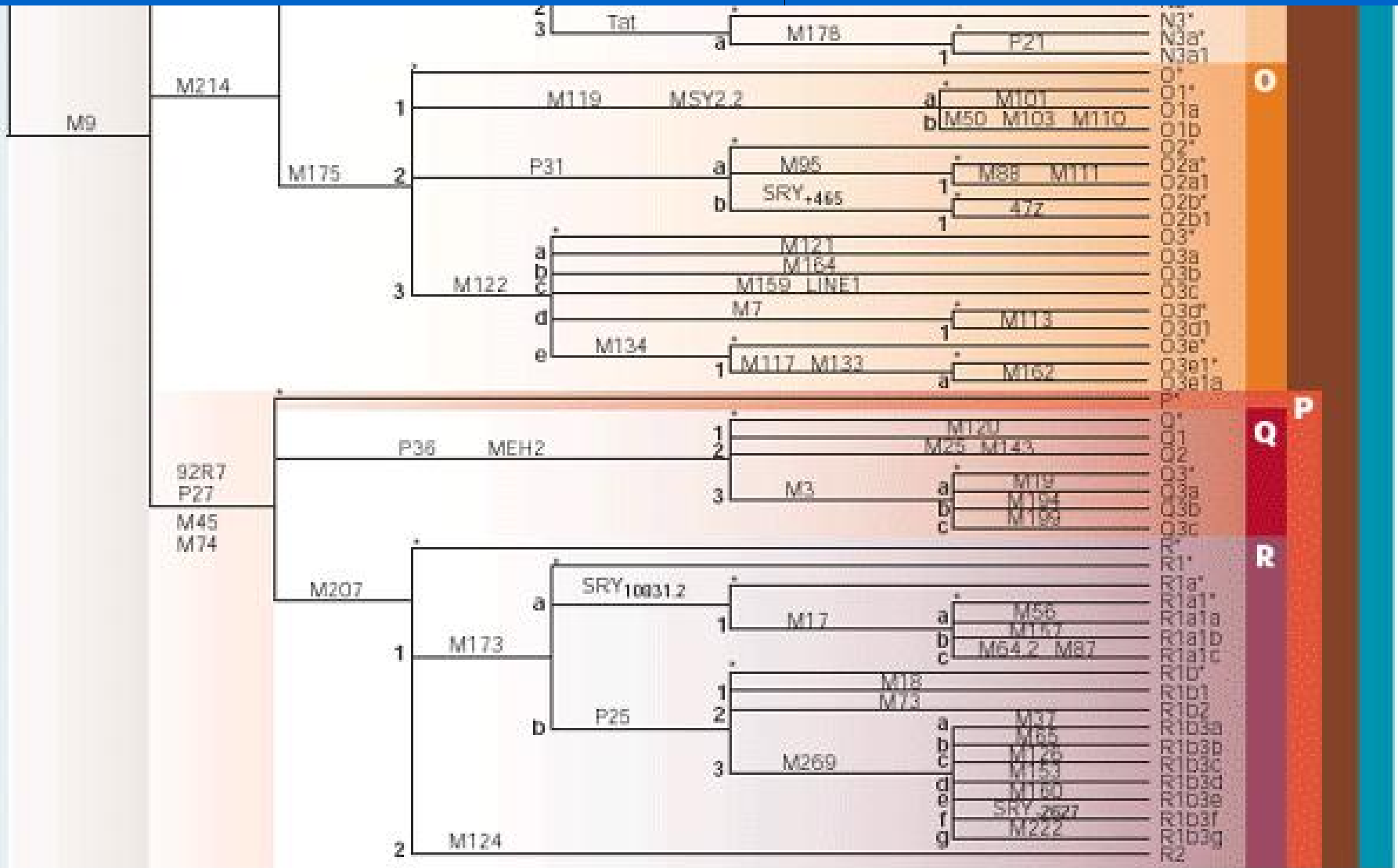




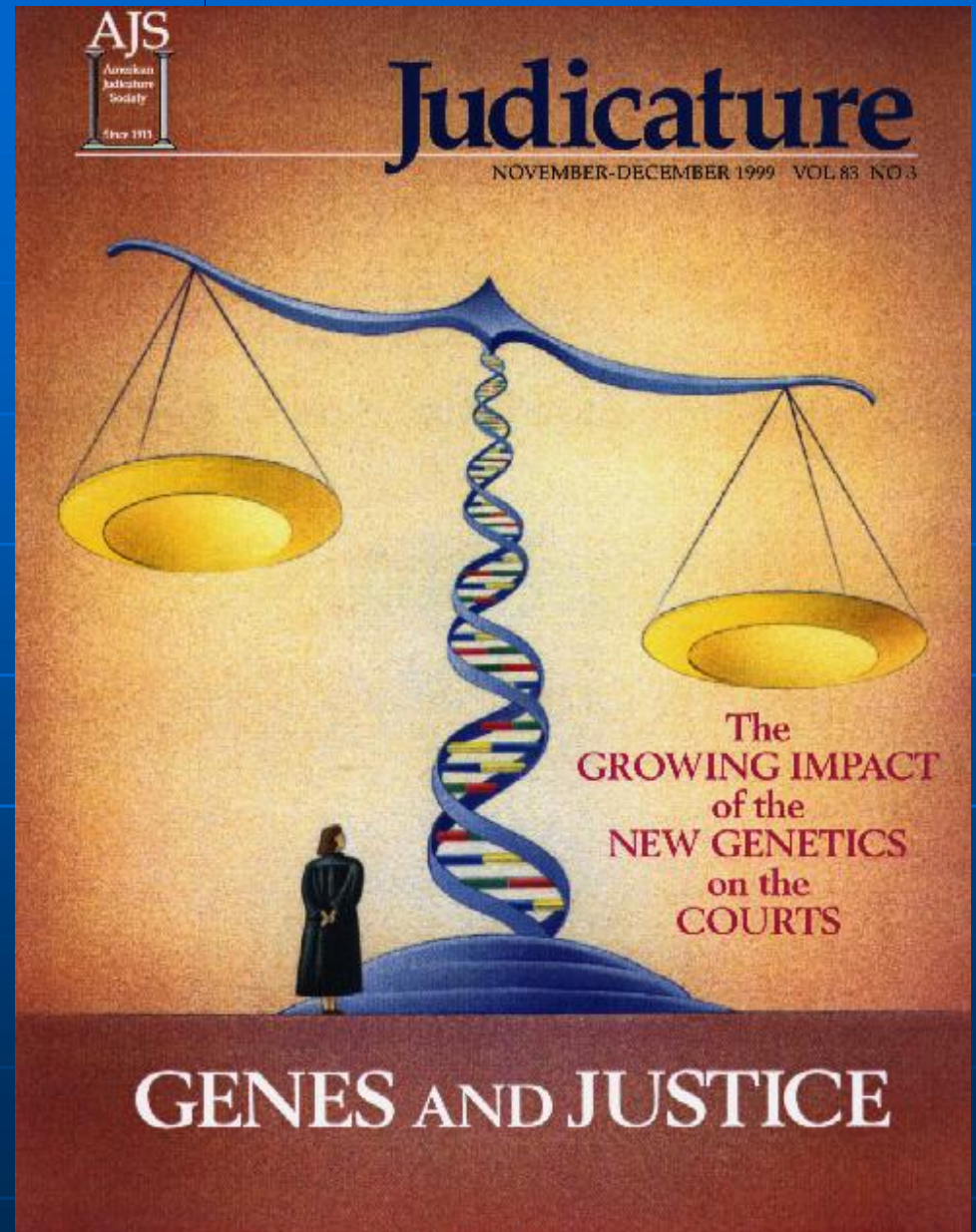
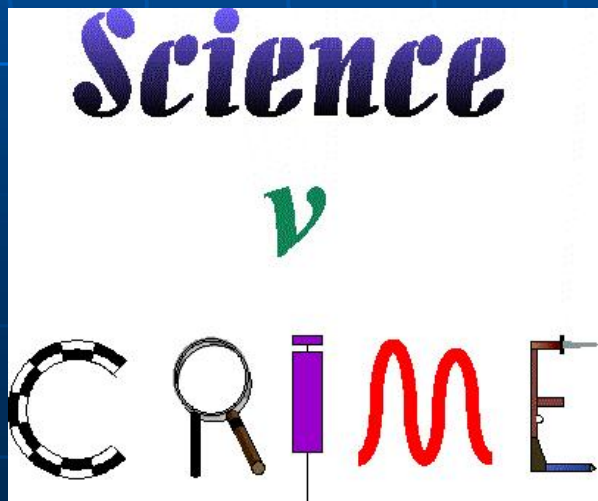
# phylogenetic tree of binary Y haplogroups



# detail of phylogenetic tree



# Forensic Applications of Y Chromosome Markers



# Reasons Y?

- **Males** (Criminal Victimization in United States, BJS 2001)
  - 80% of all violent crime
  - 95% of all sex offenses
- When trying to determine the genetic profile of the male donor in a male/female DNA admixture (when F/M > 20, often >1000) and autosomal STR analysis fails (is not informative) or not possible
  - sexual assault cases (saliva/saliva; saliva/vaginal secretions; extended interval post coital samples)
    - pre-mature lysis into non-sperm fraction
  - aspermia/oligospermia
  - normal degradation/loss over time

# Reasons Y? (cont'd)

- No need for differential extraction
- Determination of number of semen donors
- Missing persons (MP)
  - criminal paternity/disaster victim ID
  - haplotype of MP determined by typing male relative
    - son, brother, father, uncle, nephew
- Additional statistical discrimination
  - mixture/relative cases
- Y-SNPs: useful for ethnogeographic ancestry prediction

# Minimal Haplotype

- DYS393
- DYS392
- DYS391
- DYS390
- DYS19
- DYS389I
- DYS389II
- DYS385 a/b

# Y-STR Markers in Commercial kits

MHL

DYS19  
DYS385

DYS388

DYS389I/II

DYS390

DYS391

DYS392

DYS393

DYS425

DYS426

DYS434

DYS435

DYS436

DYS437

DYS438

DYS439

Reliagene  
Y-Plex 12

Promega  
Powerplex Y

SWGDM

Core Loci

DYS441

DYS442

DYS443

DYS444

DYS445

DYS446

DYS447

DYS448

DYS449

DYS452

DYS453

DYS454

DYS455

DYS456

DYS458

DYS462

DYS463

DYS464

DYS468

DYS484

DYS522

DYS527

DYS531

DYS557

DYS558

Y-GATA A7.1

Y-GATA A7.2

Y-GATA A10

Y-GATA C4

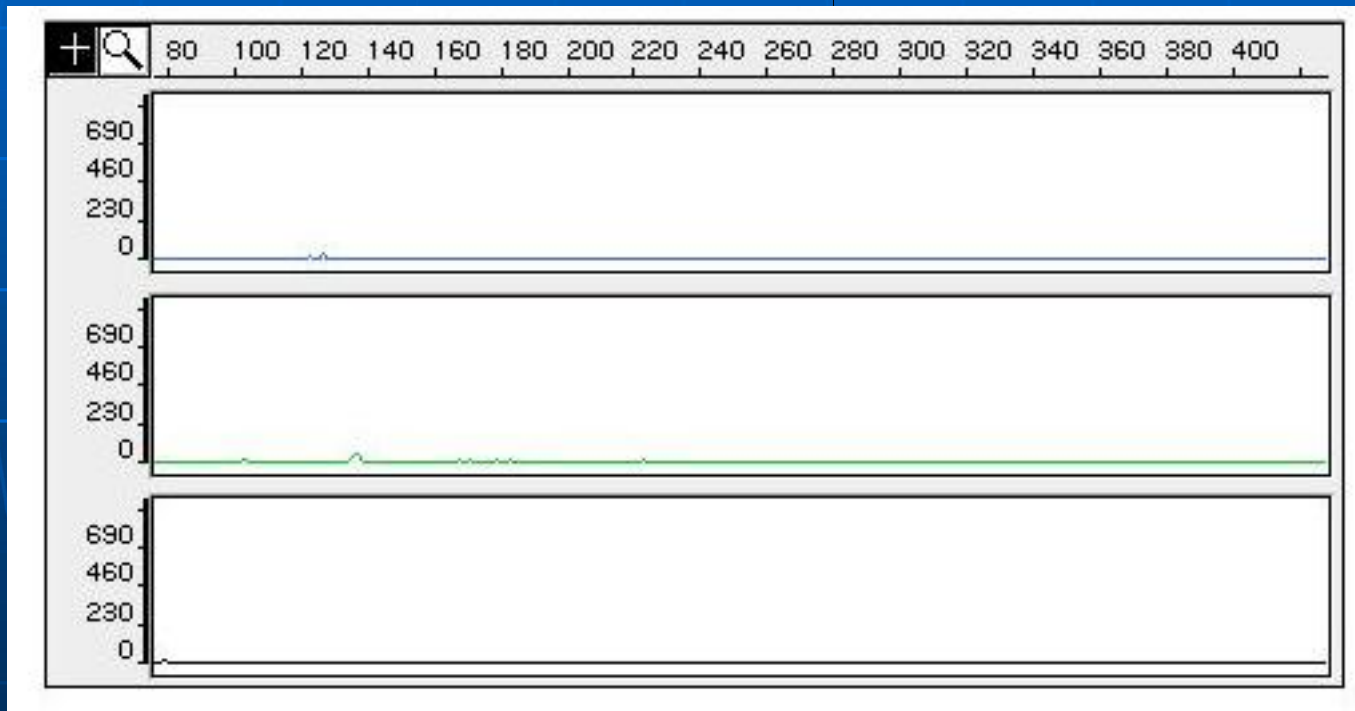
Y-GATA H4

YAP

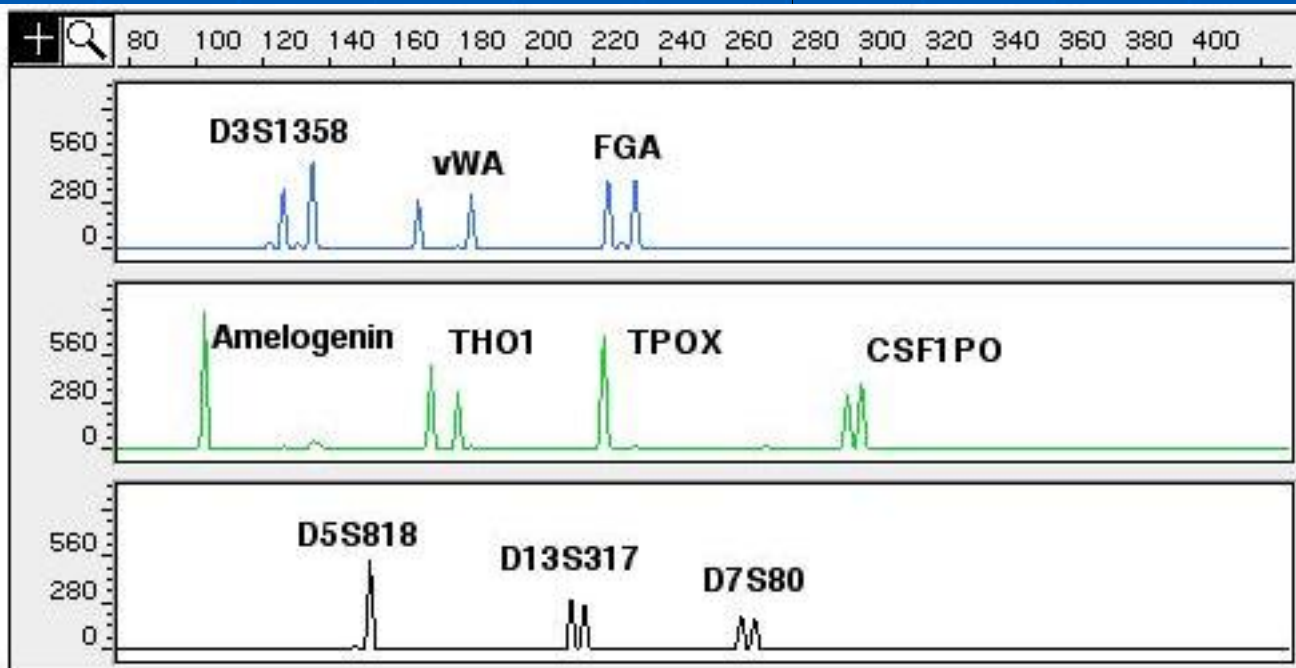
ABI

Y-Filer

# *Profiler:* 48h Post-Coital Sample, Sperm Fraction

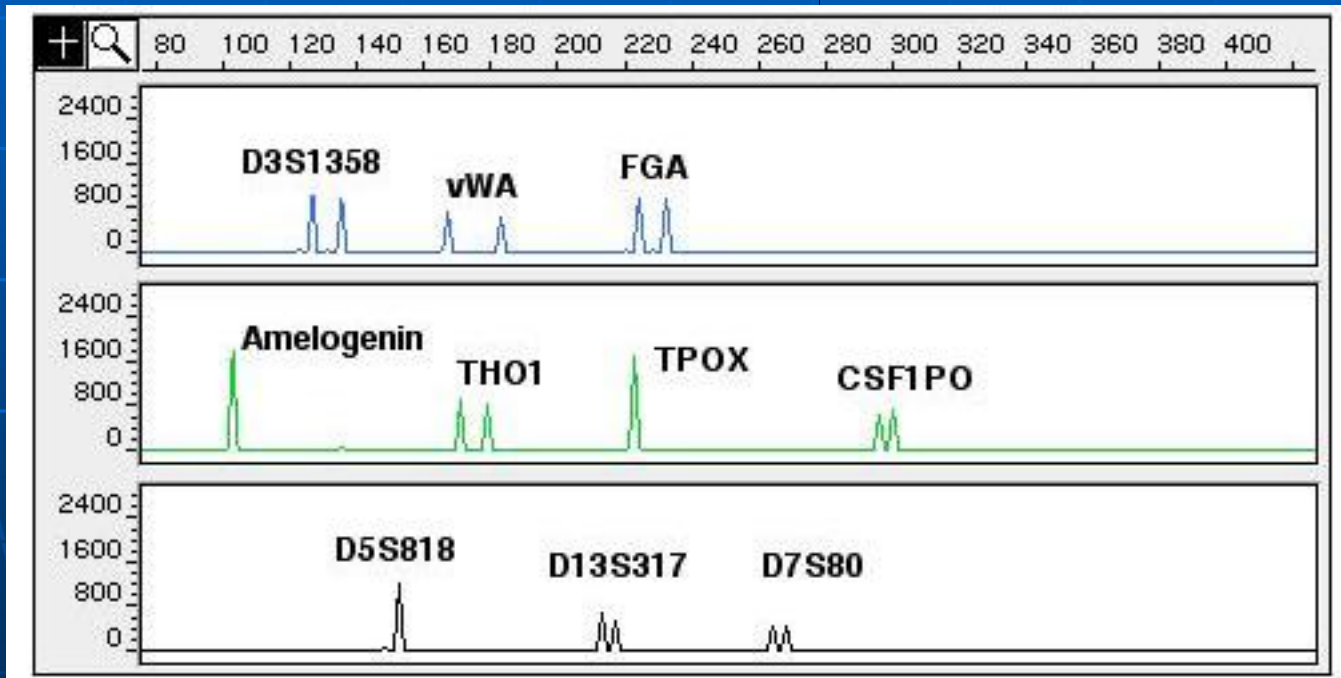


*Profiler:*  
48h Post-Coital Sample,  
Non-Sperm Fraction

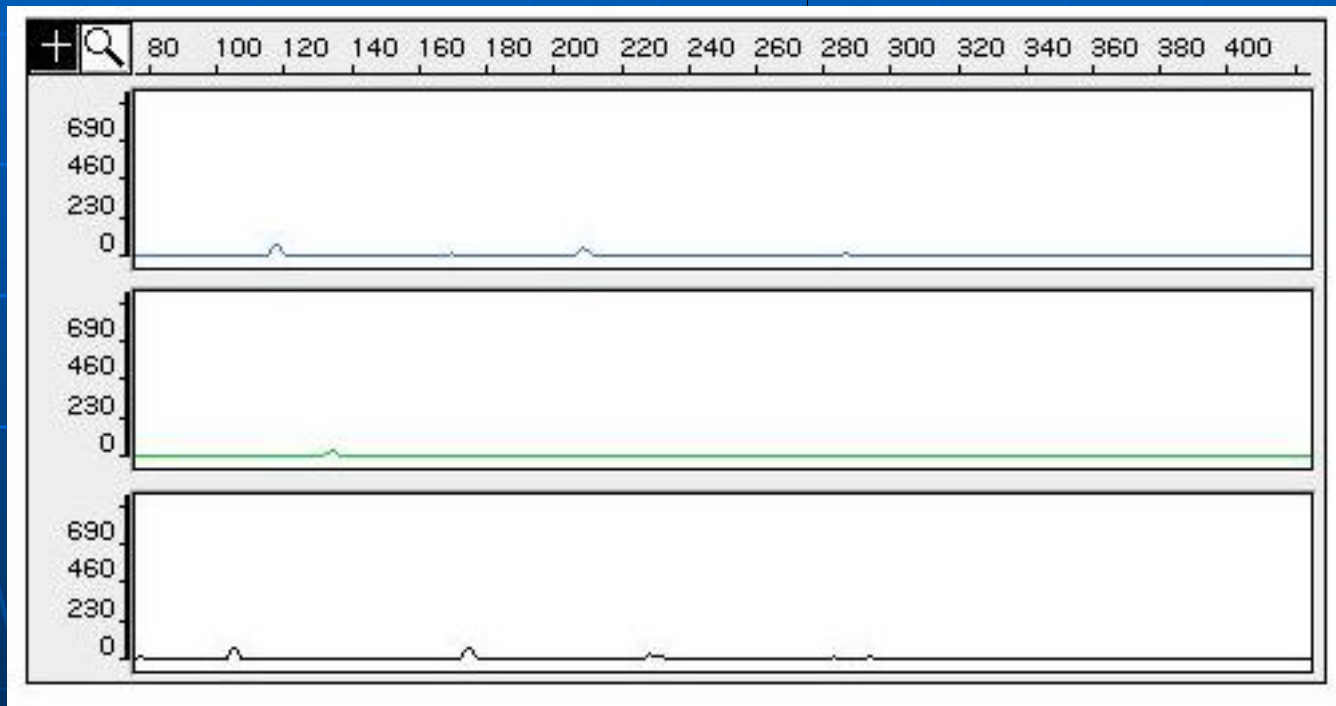




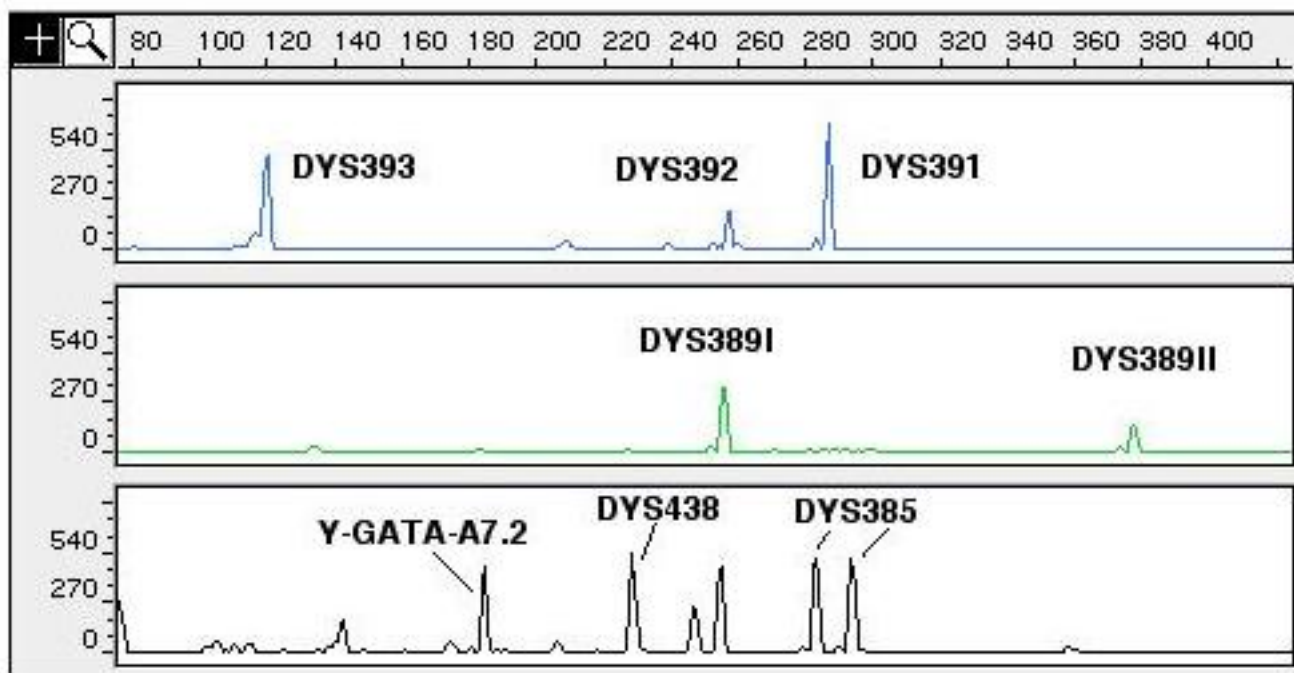
# *Profiler:* 48h Post-Coital Sample, No Differential Extraction



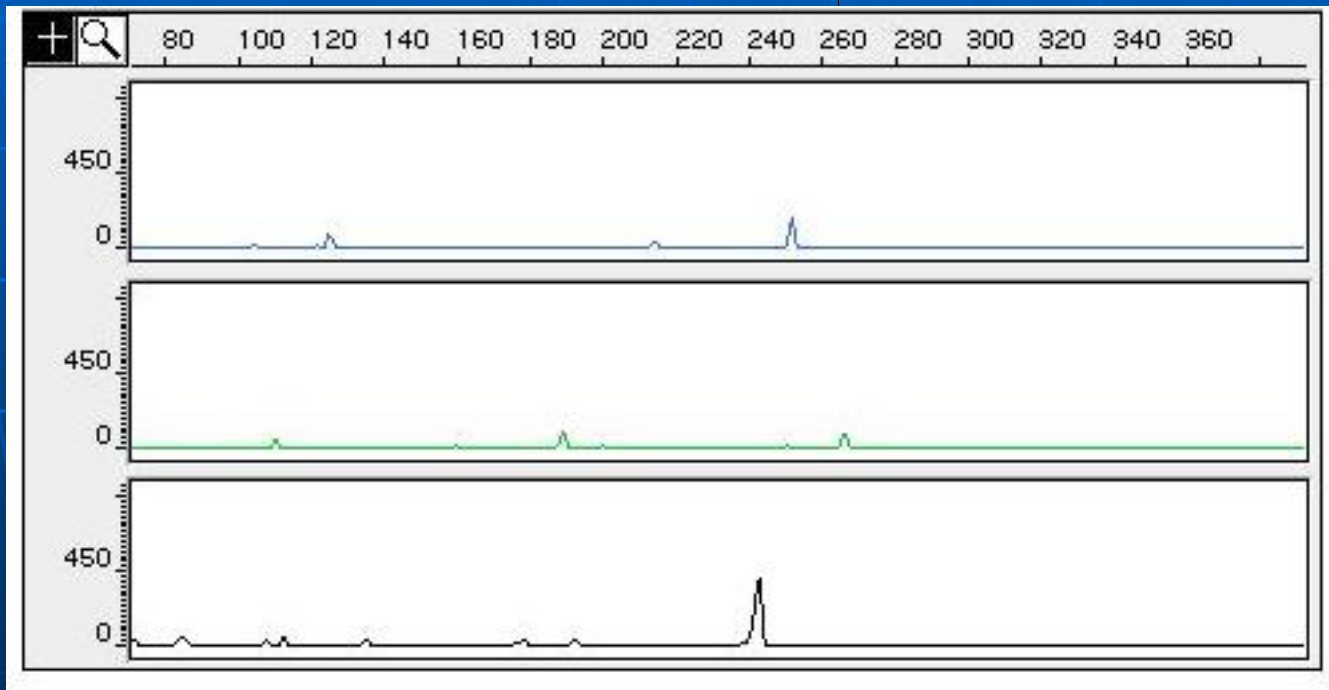
# Multiplex I: 48h Post-Coital Sample, Sperm Fraction



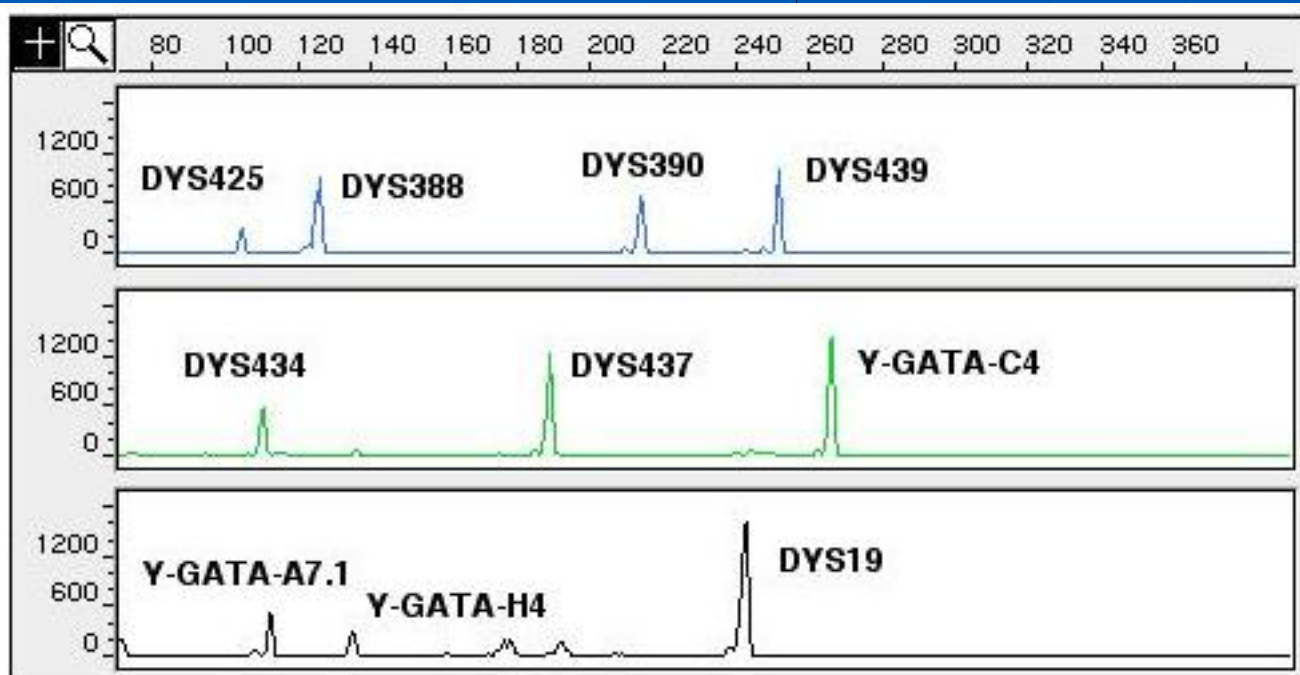
# Multiplex I: 48h Post-Coital Sample, No Differential Extraction



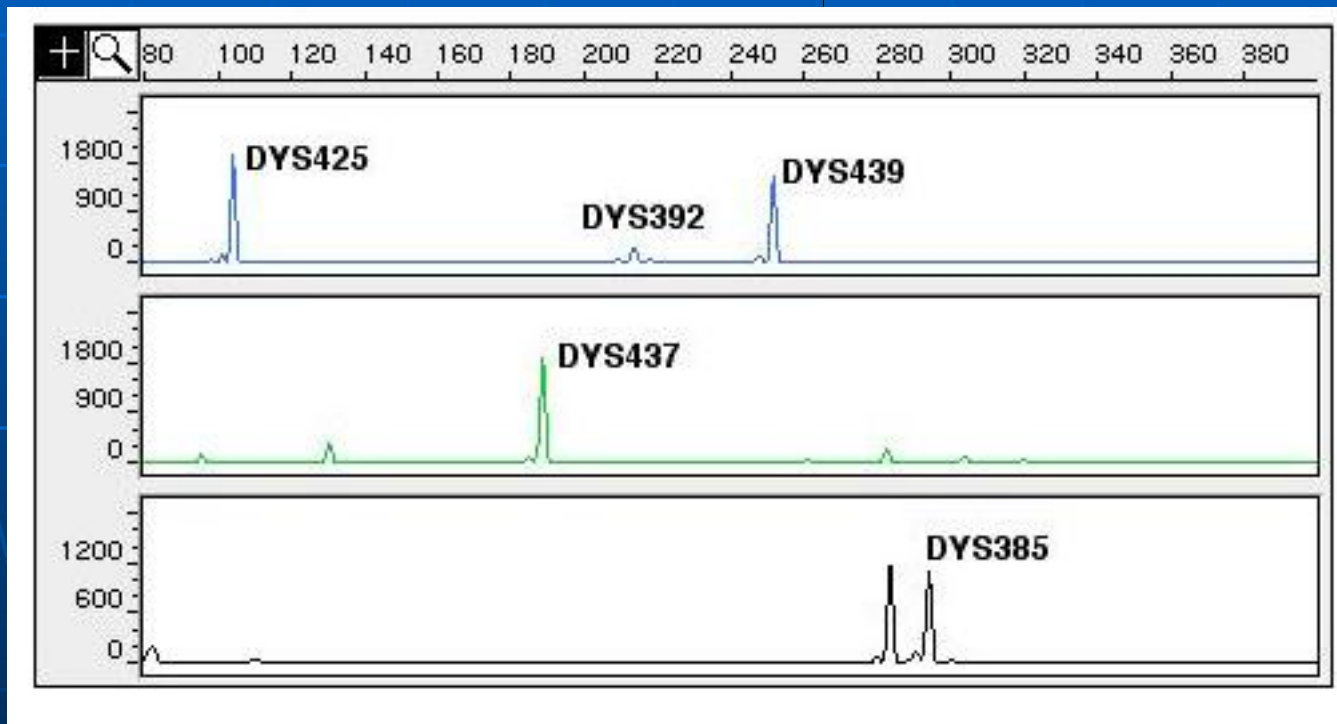
# Multiplex II: 48h Post-Coital Sample, Sperm Fraction



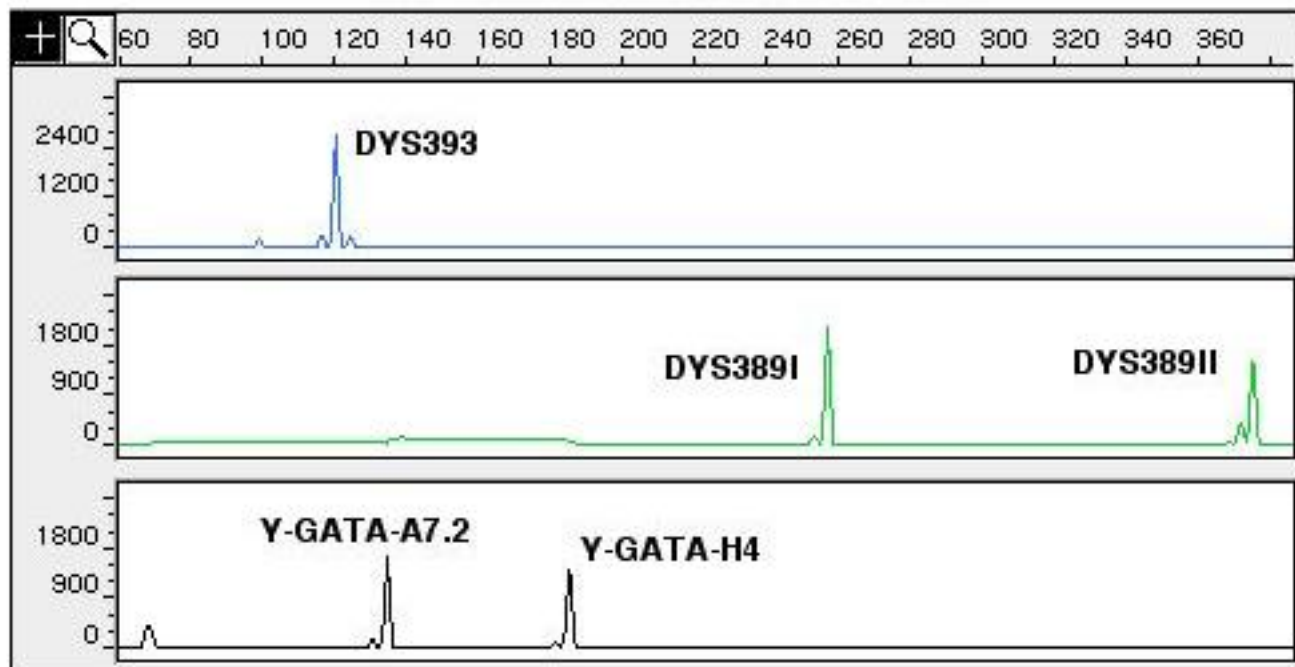
# Multiplex II: 48h Post-Coital Sample, No Differential Extraction



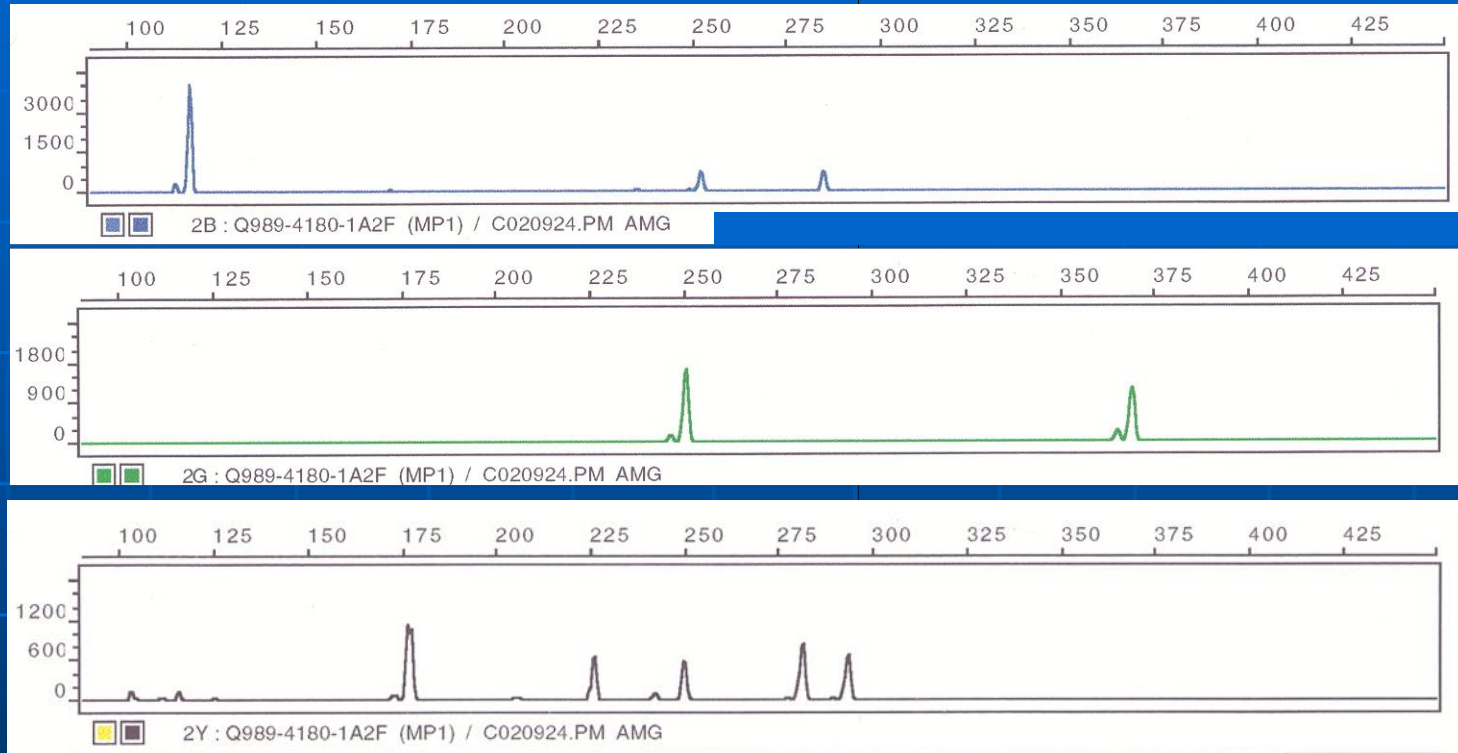
# Multiplex A: 85 h Post-Coital Sample, No Differential Extraction



# Multiplex B: 85h Post-Coital Sample, No Differential Extraction



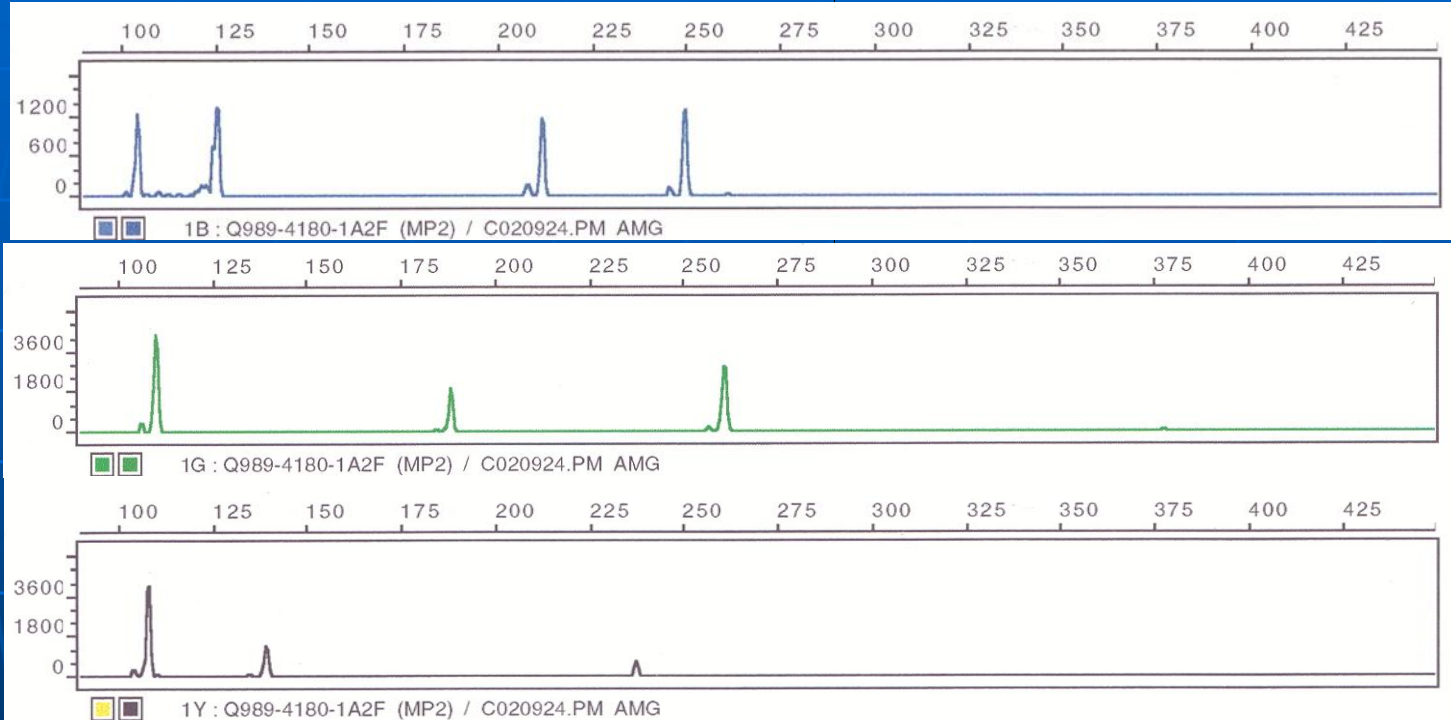
# Non –sperm fraction



UCF MPI: Female fraction of  
differential organic extract

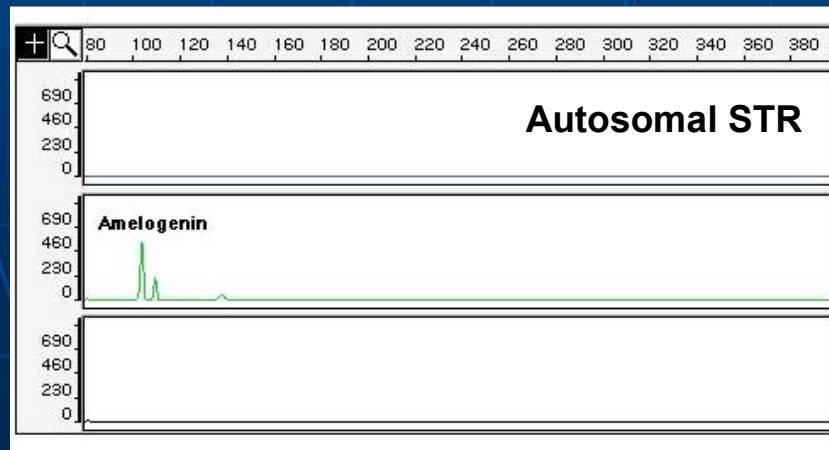
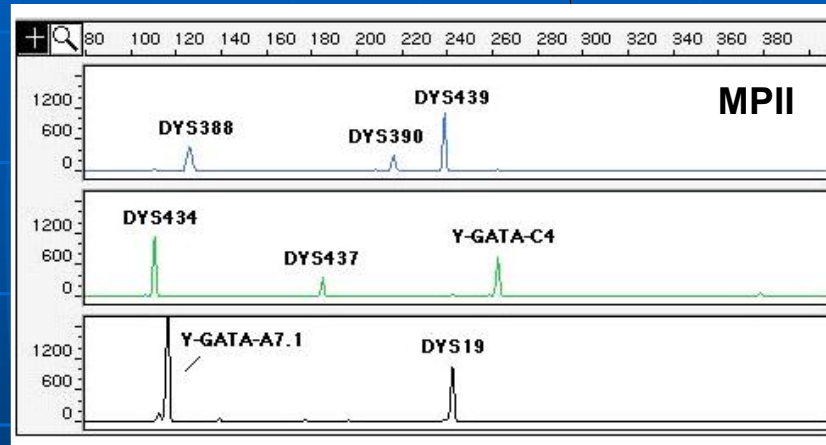
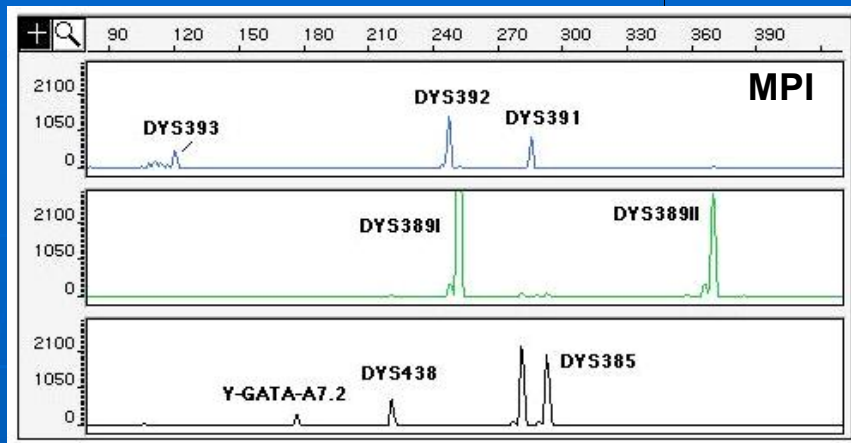


# Non –sperm fraction



UCF MPII: Female fraction of a  
differential extract

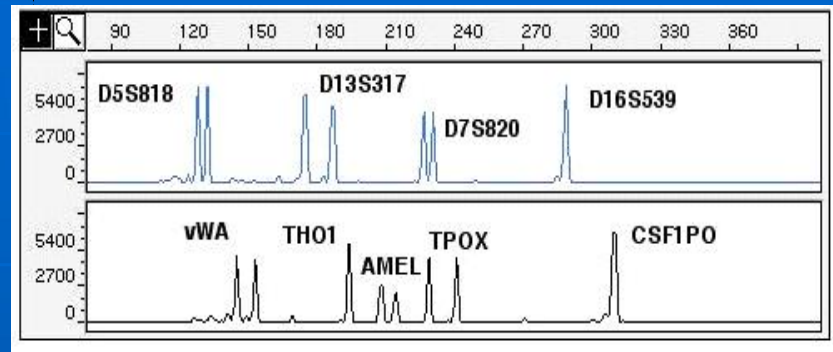
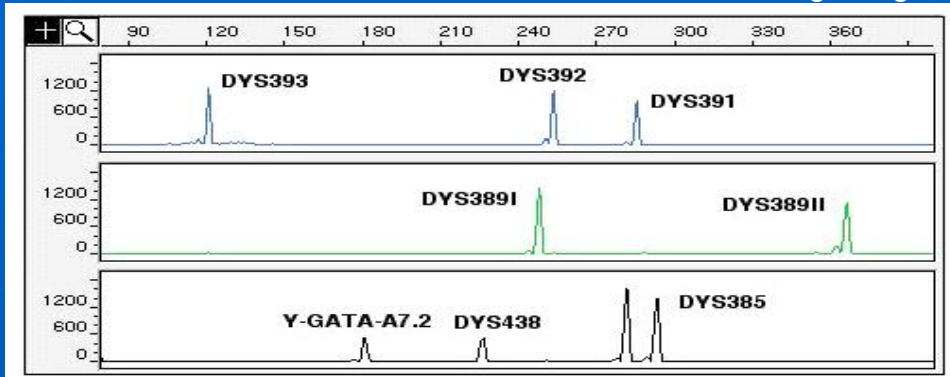
# Vasectomized Male



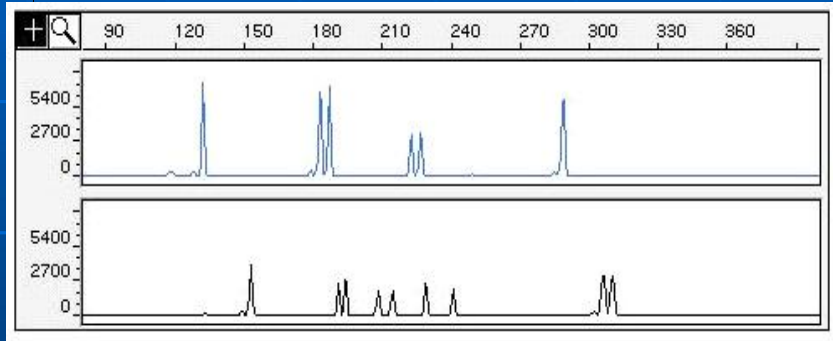
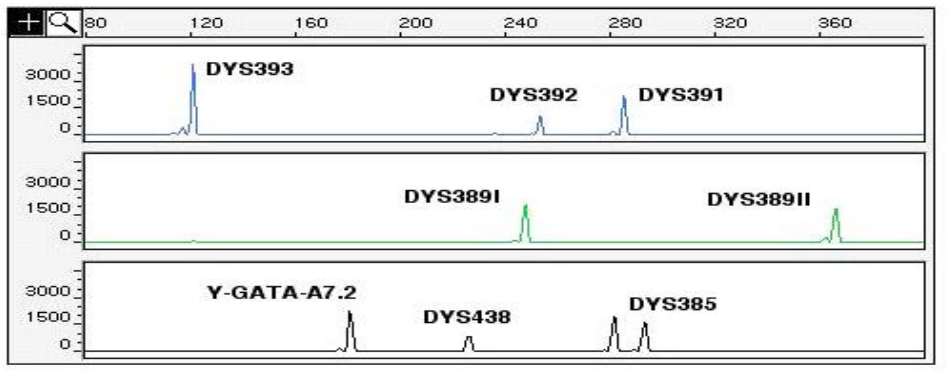
# Multigenerational Paternity Case

- Patriarch reported to have mistress and second family
- Family 1 (known relatives):
  - great-grandson 1
- Family 2 (purported relatives):
  - great-grandson 2
  - great-grandson 3

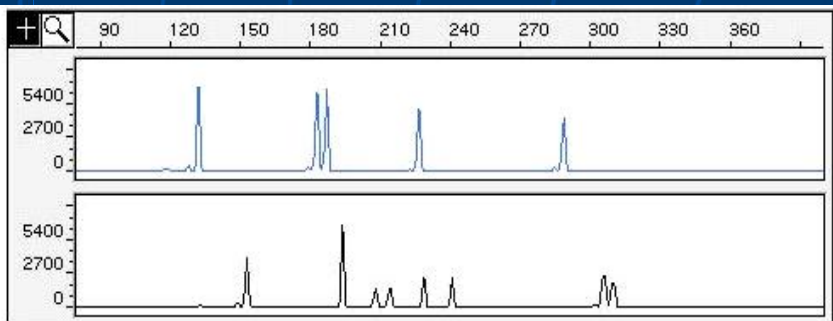
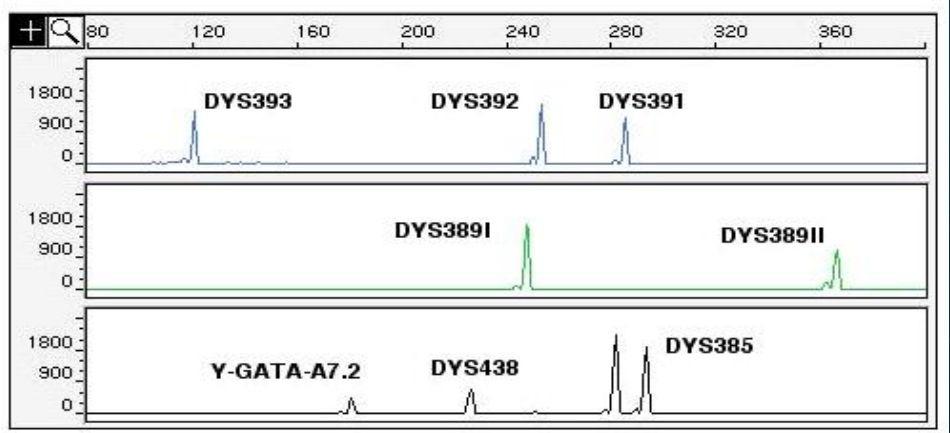
great-grandson 1

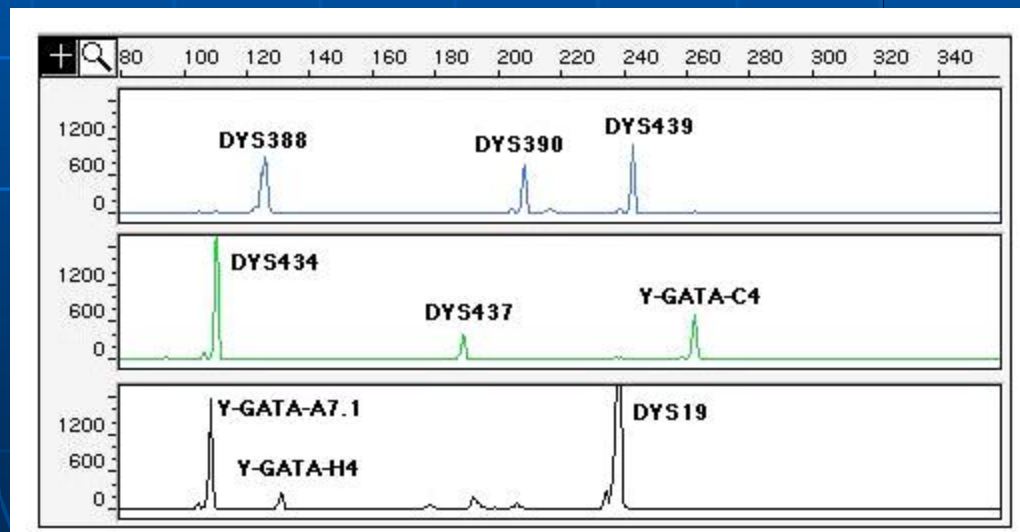
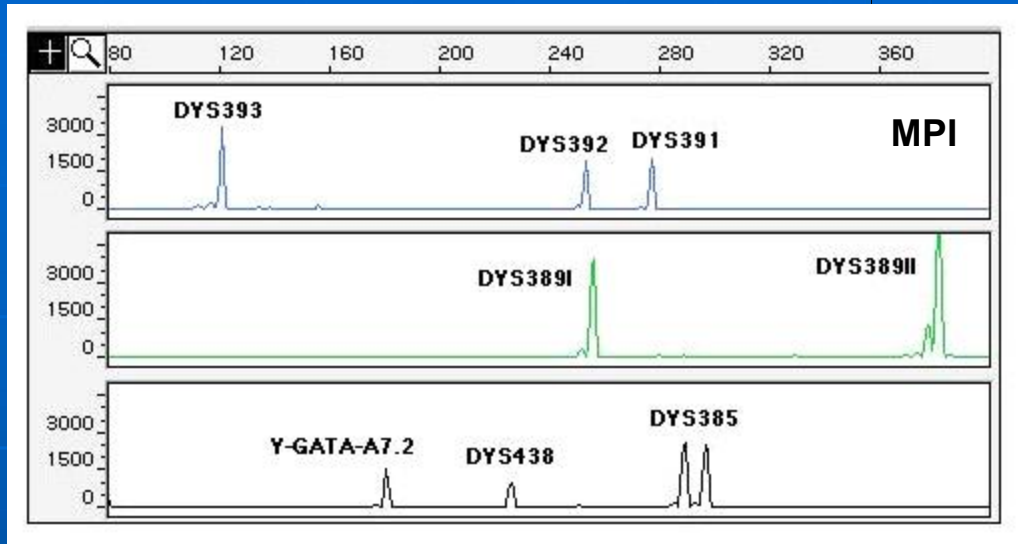


great-grandson 2



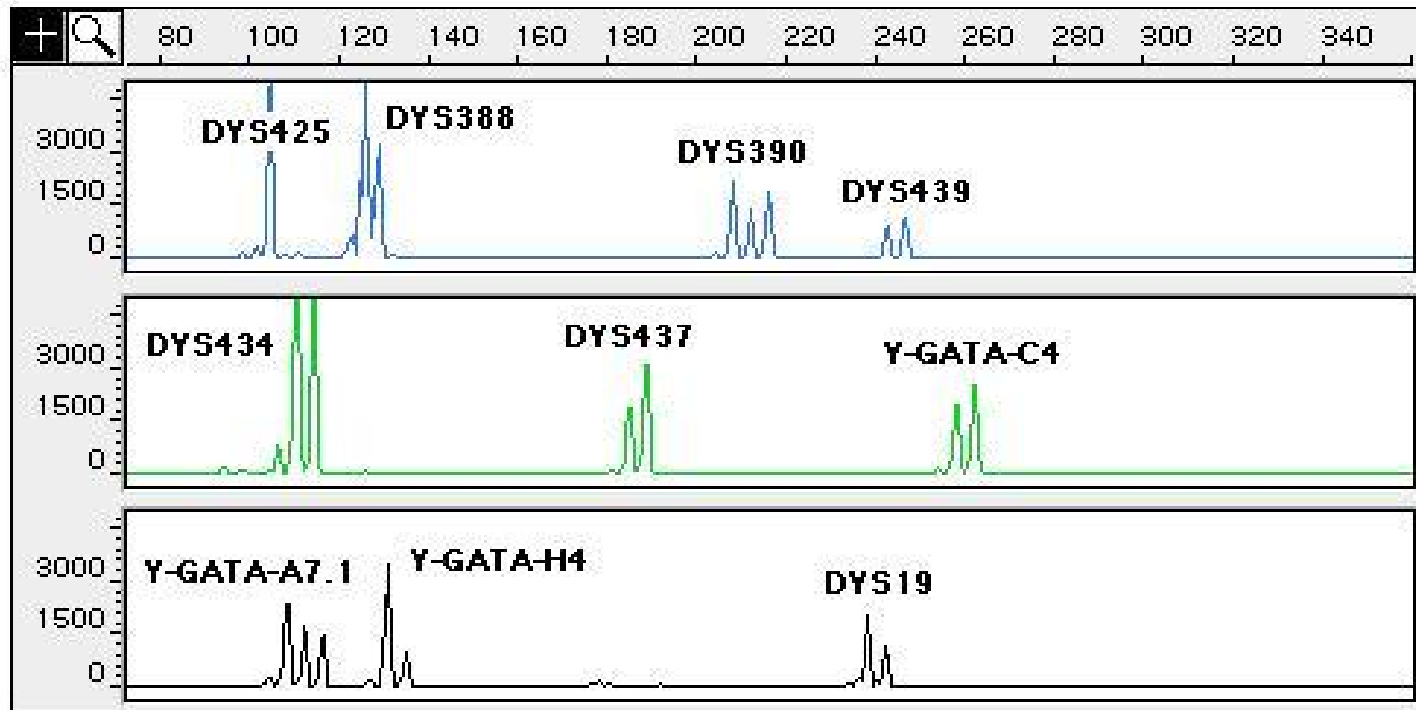
great-grandson 3





non-  
differential  
extraction  
of sample  
found on  
pants  
from a  
rape victim

# No. of Semen Donors?



# Significance of a Y-STR Match?

- **Unique biology of the Y-chromosome**
  - no recombination along most of its length
  - genetic markers are inherited as blocks of physically linked marker ---> haplotypes
- **Need to be able to estimate the frequency of occurrence of a multi-locus Y-STR haplotype**
  - since different markers (e.g. Y-STR loci) are not independent need to use counting method
    - how frequent a particular haplotype has been observed in a population
- **Need Database of Multi-Locus Haplotypes!**

# Current Y-STR Databases

- **European/Asian Y-STR Database**
  - ([www.ystr.org/europe](http://www.ystr.org/europe)) ([www.ystr.charite.de/index\\_asia](http://www.ystr.charite.de/index_asia))
  - created by Willuweit and Roewer
  - submissions from the International Forensic Y-User's Group
- **American Y-STR Database**
  - generated by Kayser and Stoneking at Max Plank Institute Leipzig, Germany
- **Databases are based upon the 9 Minimal Core Loci:**
  - **DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS385a/b**
- **SWGAM Endorsed Loci:**
  - **DYS438, DYS439**



# Current Y-STR Databases

**1. Databases comprising minimal haplotype (MHL) loci: (DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS385a/b)**

- <http://www.ystr.org> (Roewer et al (2001), *Forensic Sci. Int.* **118** 106-113)  
94 European populations n=13,892
- <http://www.ystr.org/usa> (Kayser et al (2002), *J. Forensic Sci.* **47** 513-519)  
30 regional US populations: 599 AA, 628 EA, 478 HA
- <http://www.ystr.org/asia> (Lessig et al (2003), *Legal Medicine* **5** S160-S163)  
22 Asian populations n = 2576

Worldwide Search Option (<http://www.ystr.org>)  
189 populations n = 20,865

# Current Y-STR Databases (cont')

## 2. Databases maintained by Commercial Companies

comprising SWGDAM Core: MHL + 438, 439

- Reliagene (<http://www.reliagene.com>)

617 AA, 628 EA, 347 HA, 84 NA

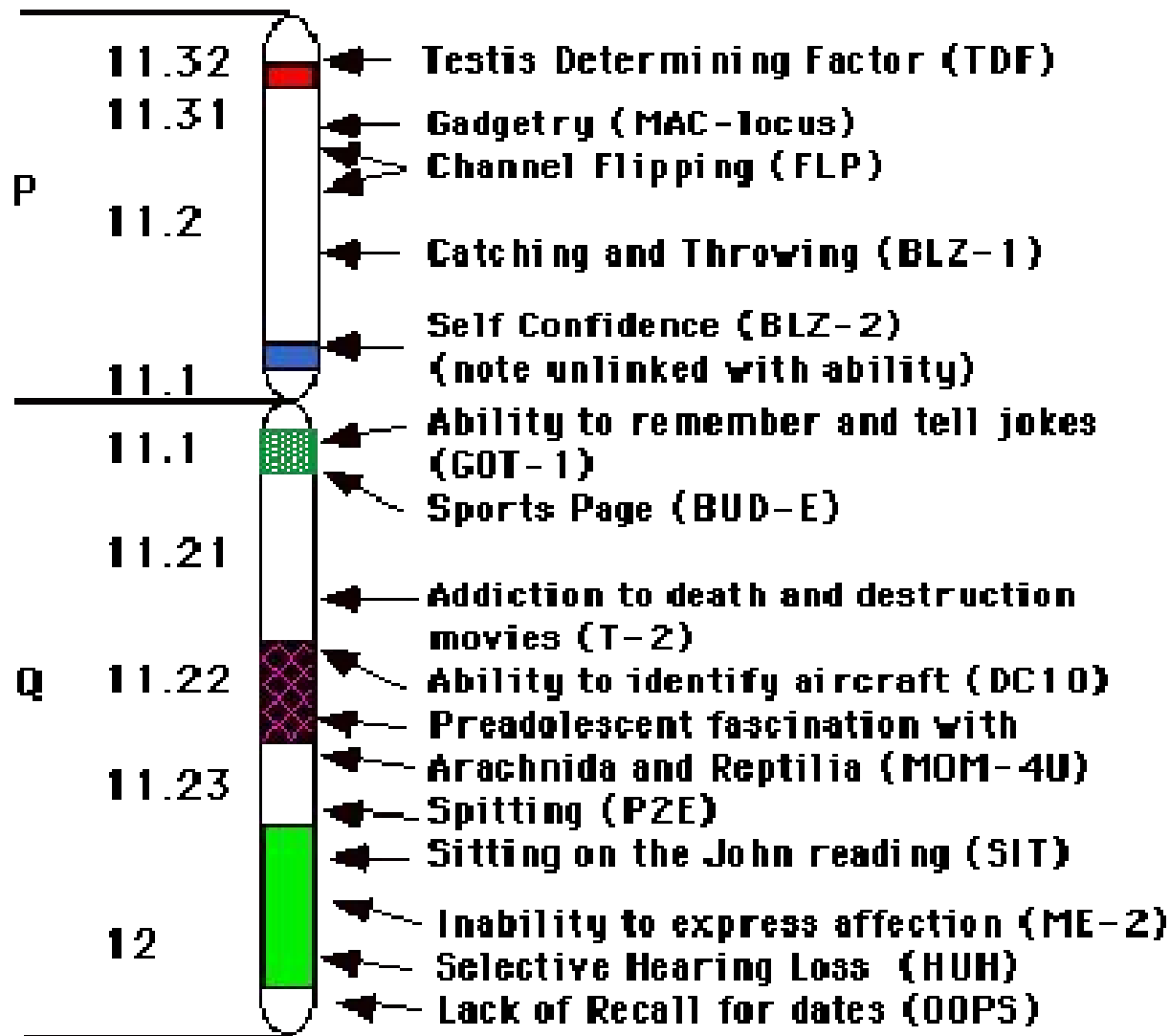
## 3. Databases maintained by Commercial Companies

comprising SWGDAM Core: MHL + 437, 438, 439

- Promega (<http://www.promega.com>)

577 AA, 595 EA, 630 HA, 357 NA, 284 A (n = 2443)

# Y Chromosome map



- Jane Gitschier, UCSF Science, 261, 679 (Aug. 93)

**Thank You!**



The end crowns all,  
And that old common  
arbitrator, Time,  
Will one day end it.

*William Shakespeare*  
*Troilus and Cressida*