GENETICS AND POPULATION GENETICS

Forensic Genetics

Population allele frequencies

DNA Profile
(with specific alleles)

Rarity estimate of DNA profile
(e.g., RMP or LR)

Genetic formulas

ELTE Faculty of Sciences Department of Genetics

Genotiping microsatellites by PCR
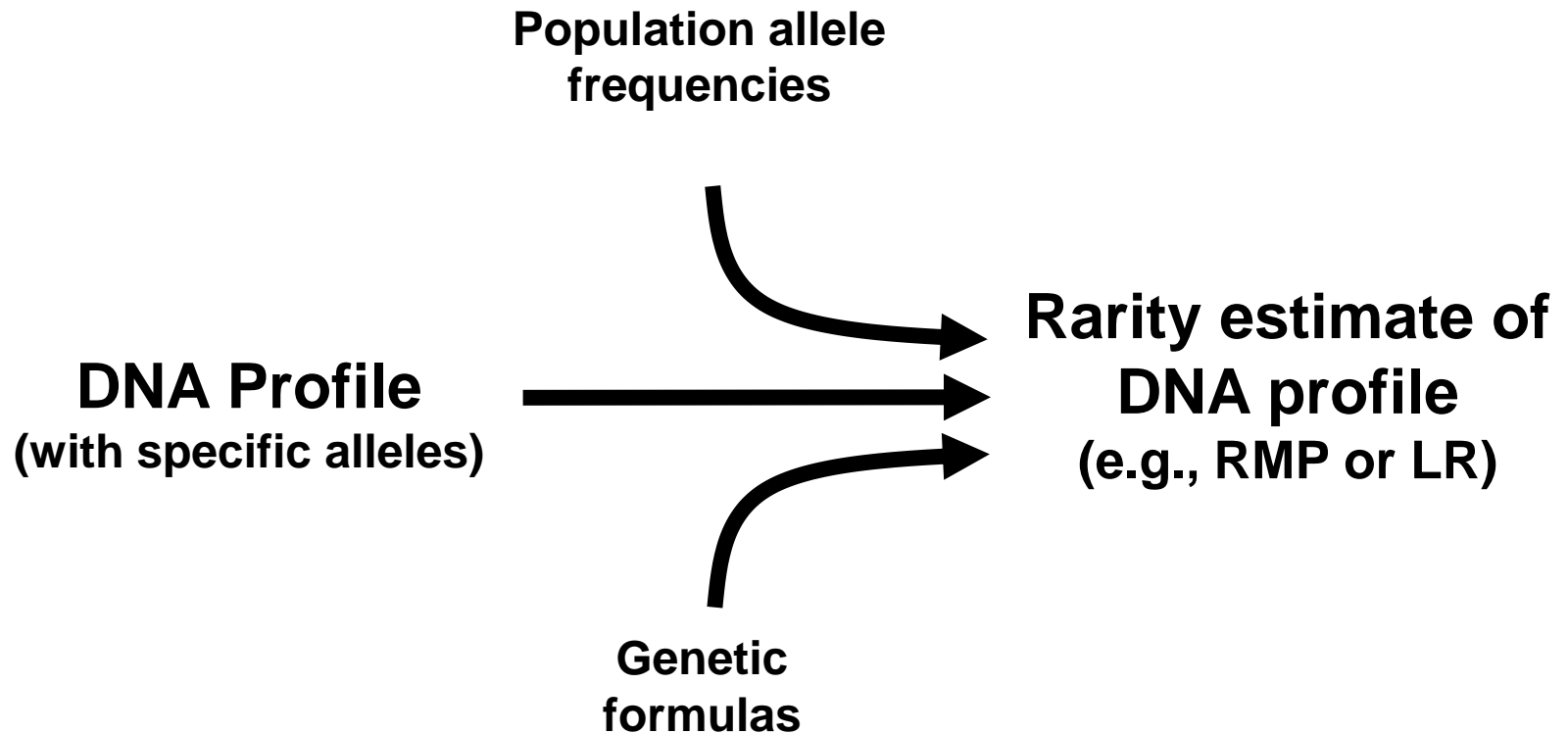
# Microsatellite locus genotyping: multiplex-PCR

TACAAAAAAATACAAAAATTAGTTGGGCATGGTGGCACGTGCCTGTAGTCTCAGCTACT
TGCAGGGCTGAGGCAGGAGGAG**TTCTTGAGCCCAGAA**[G→A]**GTTA**AGGCTGCAGTGAG
CCATGTTCATGCCACTGCACTTCACTCTGAGTGACAAATTGAGACCTTGTCTC/agaa/
agaa/agaa/agaa/agaa/agaa/agaa/agaa/agaa/agaa/agaa/agaa/agaa
/agaa/agaa/agaa/agaa/agaa/ *[AAAGAGAGA]*GGAAAGAAAGAGAAAAAGAAAA
GAAATAGTAGCAACTGTTATTGTAAGACATCTCCACACACCAGAGAAGTTAATTTTAAT
TTTAACATGTTAAGAACAGAGAGAAGCCAACATGTCCACCTTAGGCTGACGGTTT**GTTT**
ATTTGTGTTGTTGCTGGTAG*TC[G→A]GGTTTGTTATTTTTAAAGTAGCTTATCCAAT
ACTTCATTAACAATTTCAGTAAGTTATTTCATCTTTCAACATAAATACGCACAAGGATT
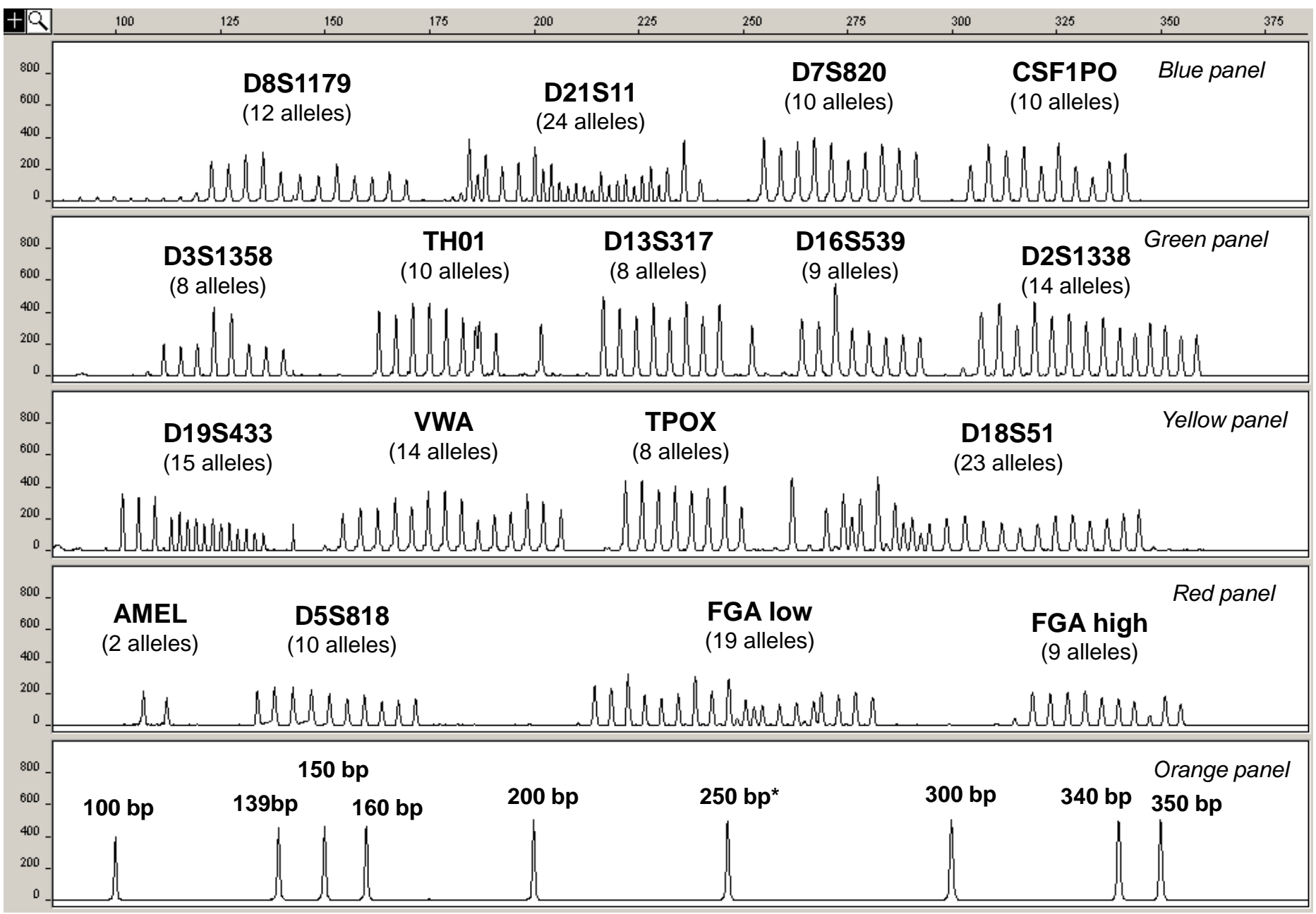TCTTCTGGTCAAGACCAAACTAATATTAGTCCATAGTAG

Polyacrylamide Gel
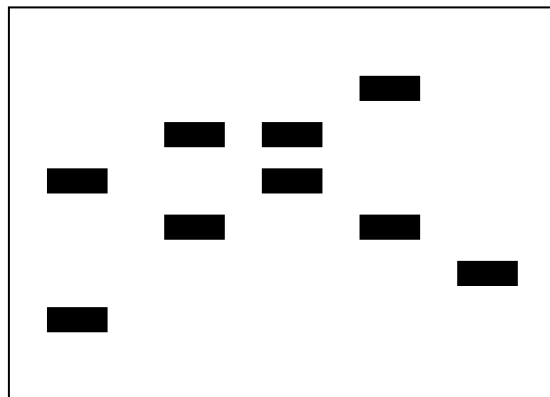
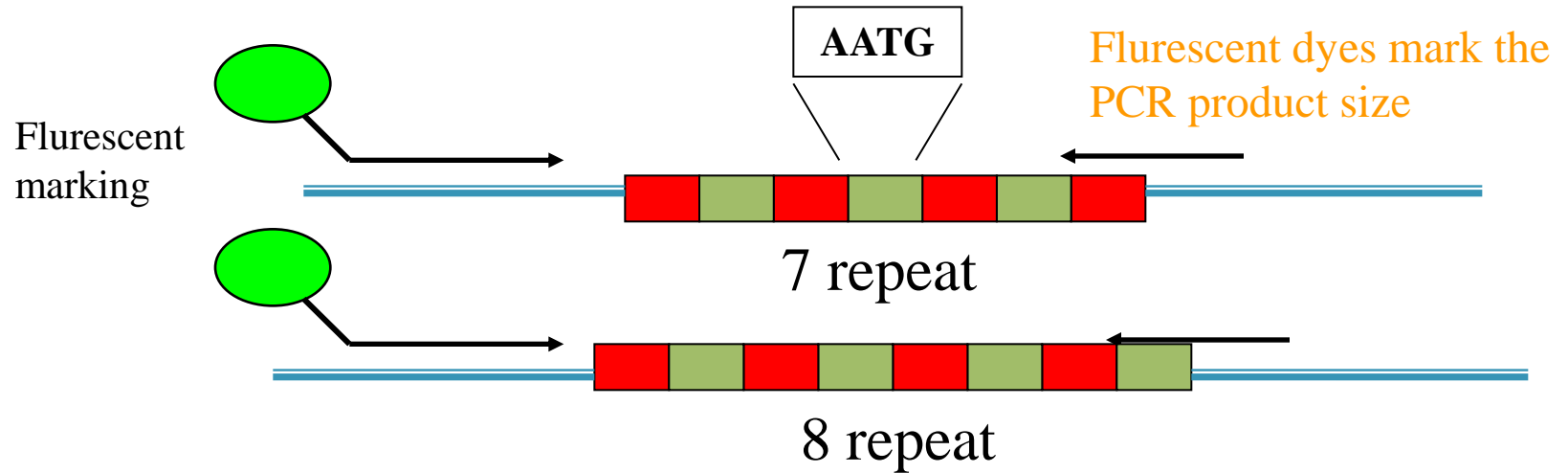Find representative alleles spanning population variation

Combine

Re-amplify
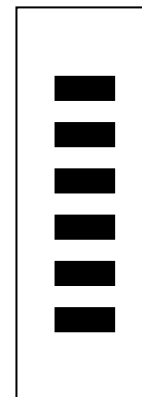
Separate PCR products from various samples amplified with primers targeted to a particular STR locus

Blue panel

D8S1179 (12 alleles)

D21S11 (24 alleles)

D7S820 (10 alleles)

CSF1PO (10 alleles)

Green panel

D3S1358 (8 alleles)

TH01 (10 alleles)

D13S317 (8 alleles)

D16S539 (9 alleles)

D2S1338 (14 alleles)

Yellow panel

D19S433 (15 alleles)

VWA (14 alleles)

TPOX (8 alleles)

D18S51 (23 alleles)

Red panel

AMEL (2 alleles)

D5S818 (10 alleles)

FGA low (19 alleles)

FGA high (9 alleles)

Orange panel

100 bp 139bp 150 bp 160 bp 200 bp 250 bp* 300 bp 340 bp 350 bp

LIZ-labeled DNA sizing standard

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

Microsatellite – STR – markers
(Short Tandem Repeat)

AATG

Flurescent dyes mark the PCR product size

Flurescent marking

7 repeat

8 repeat

Repeat region varies in length from alleles to alleles but flanking region where PCR primers bind is constant
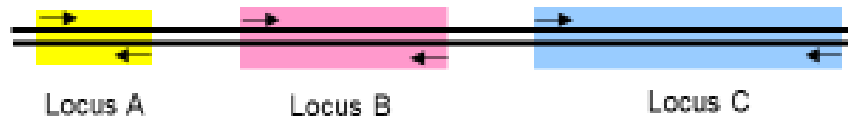
Homozygous = two homologous are the same

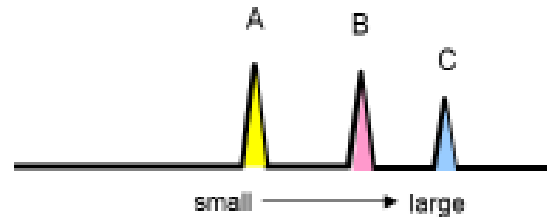Heterozygous = two homologous separate from each other

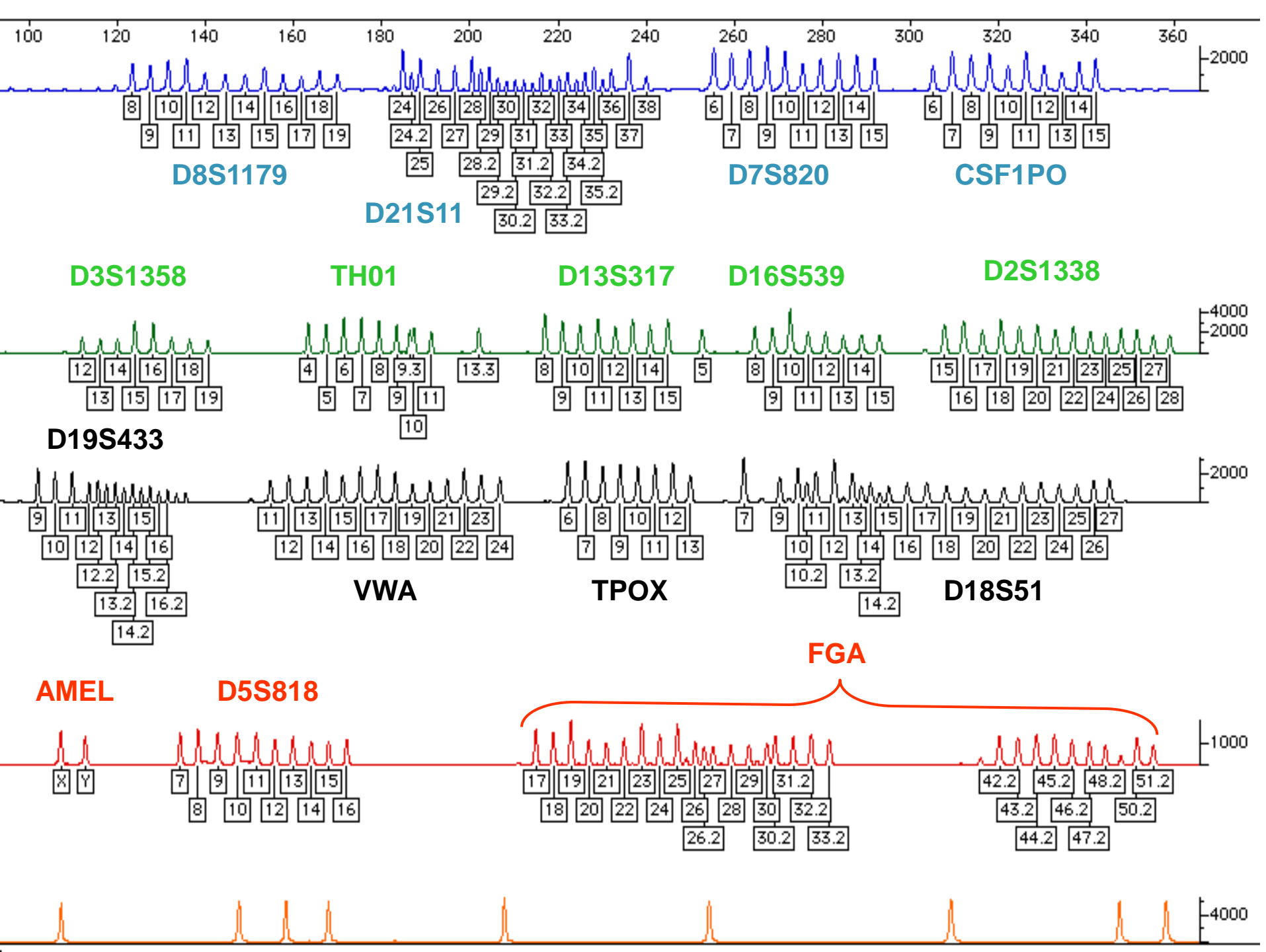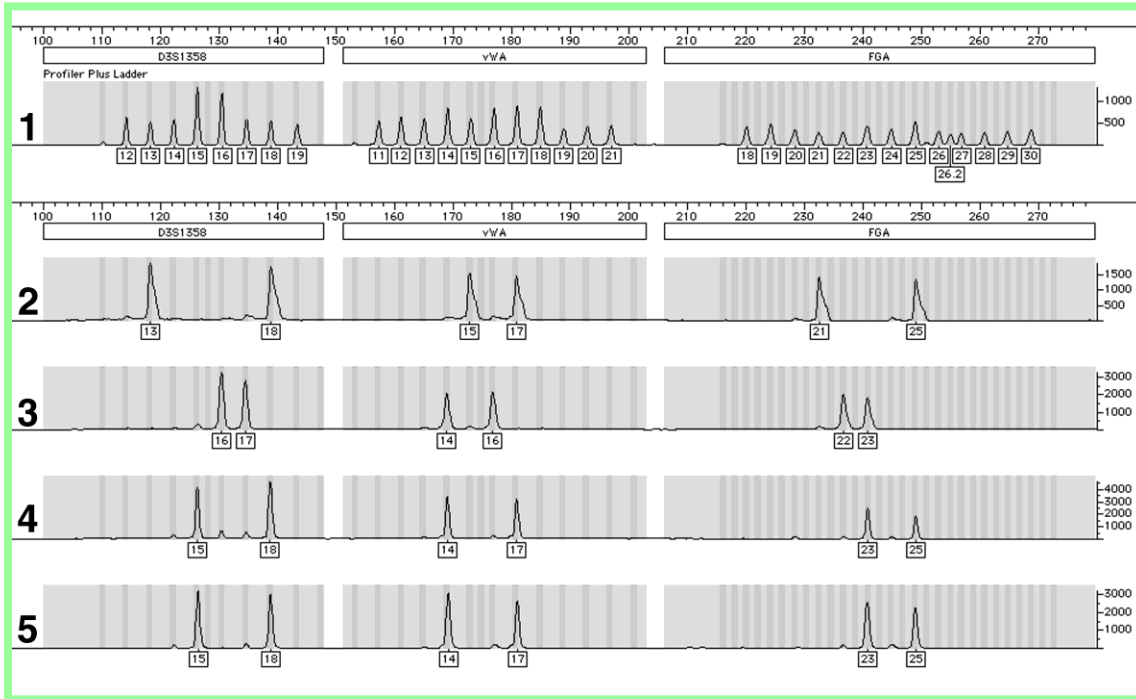**Primer binding sites determine the PCR product size!**

Multiplex - PCR

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

Locus A   Locus B   Locus C

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

A   B   C

small ⟶ large

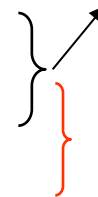Multiplex STR analyses on three microsatellite loci

1: allelic ladder

2, 3, 5: DNA profile of three different persons (suspects)

4: evidence DNA profile (cigarette butt)

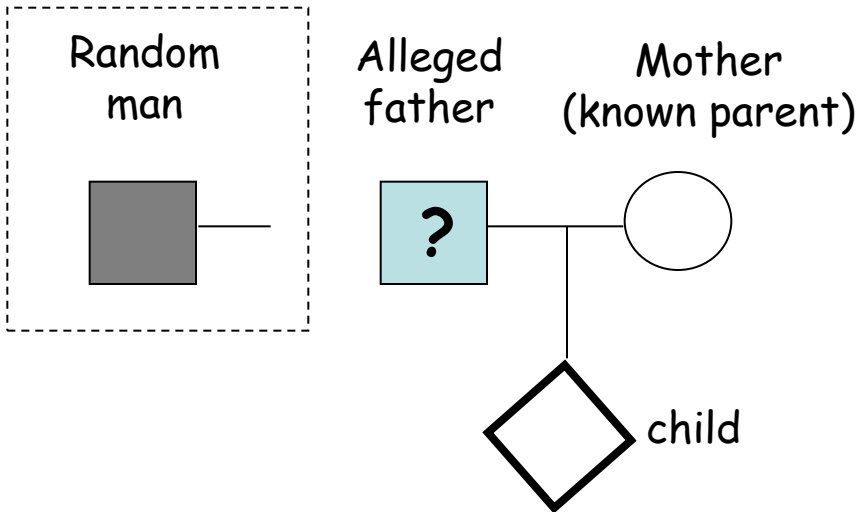|	D3S1358	vWA	FGA
3: A person	16/17	14/16	22/23
4: Evidence	15/18	14/17	23/25
5: B person	15/18	14/17	23/25

Different origin = Exclusion
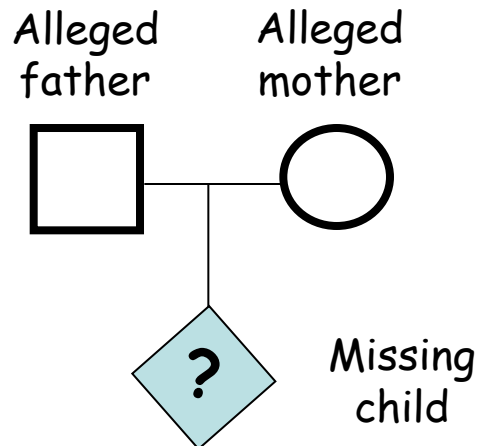
Identical? = Contribution?
Random match probability?

(a) Parentage (Paternity) Testing

Random man

Alleged father

Mother (known parent)

?

child

(b) Reverse Parentage Testing (Missing Persons Investigation)

Alleged father

Alleged mother

?   Missing child
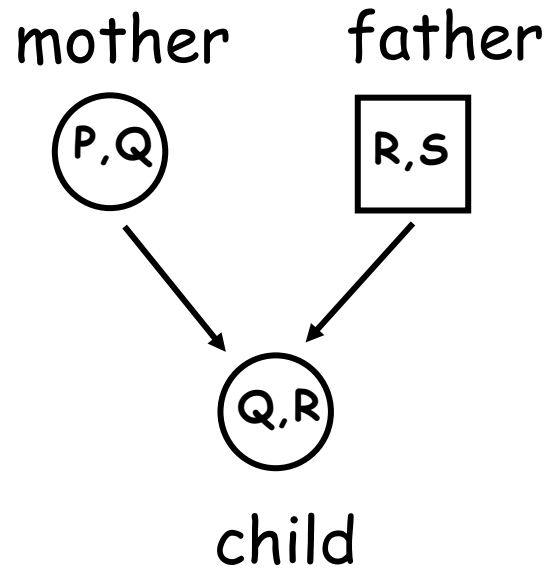
Rules of Inheritance

1) Child has two alleles for each autosomal marker *(one from mother and one from biological father)*

2) Child will have mother's mitochondrial DNA haplotype (barring mutation)

3) Child, if a son, will have father's Y-chromosome haplotype (barring mutation)

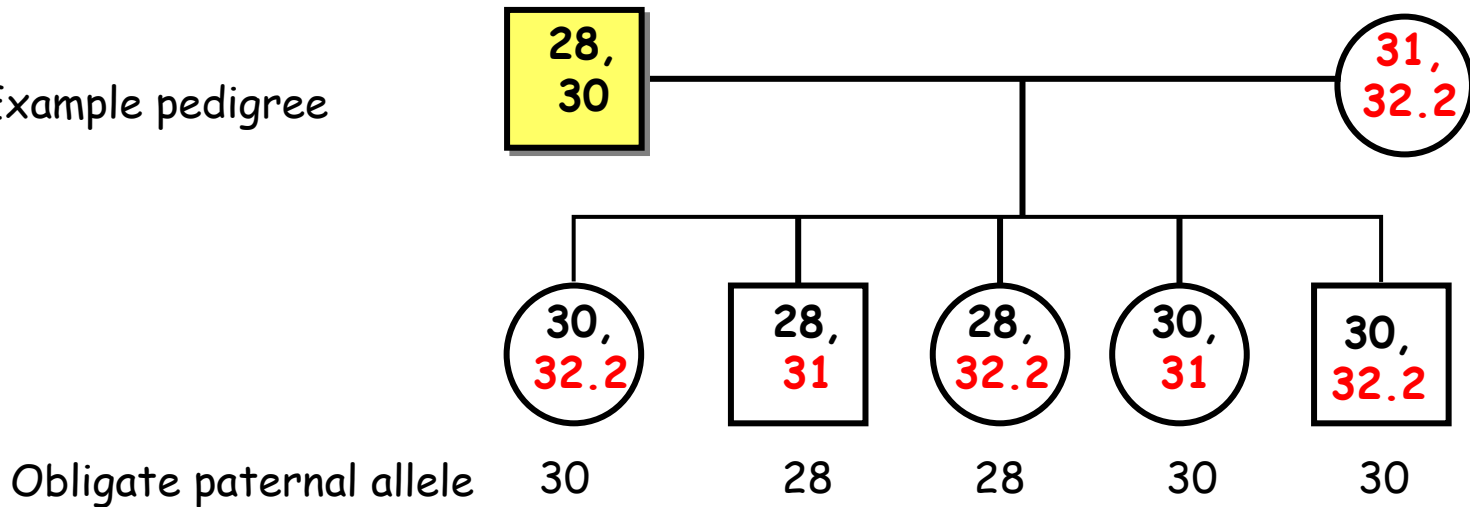**(a)** Mendelian Inheritance

mother   father

P,Q   R,S

Q,R

child

(b) Example pedigree

28, 30

31, 32.2

30, 32.2   28, 31   28, 32.2   30, 31   30, 32.2

Obligate paternal allele   30   28   28   30   30

(c) Example data

Punnett Square

Mother gametes (egg)

A p a q
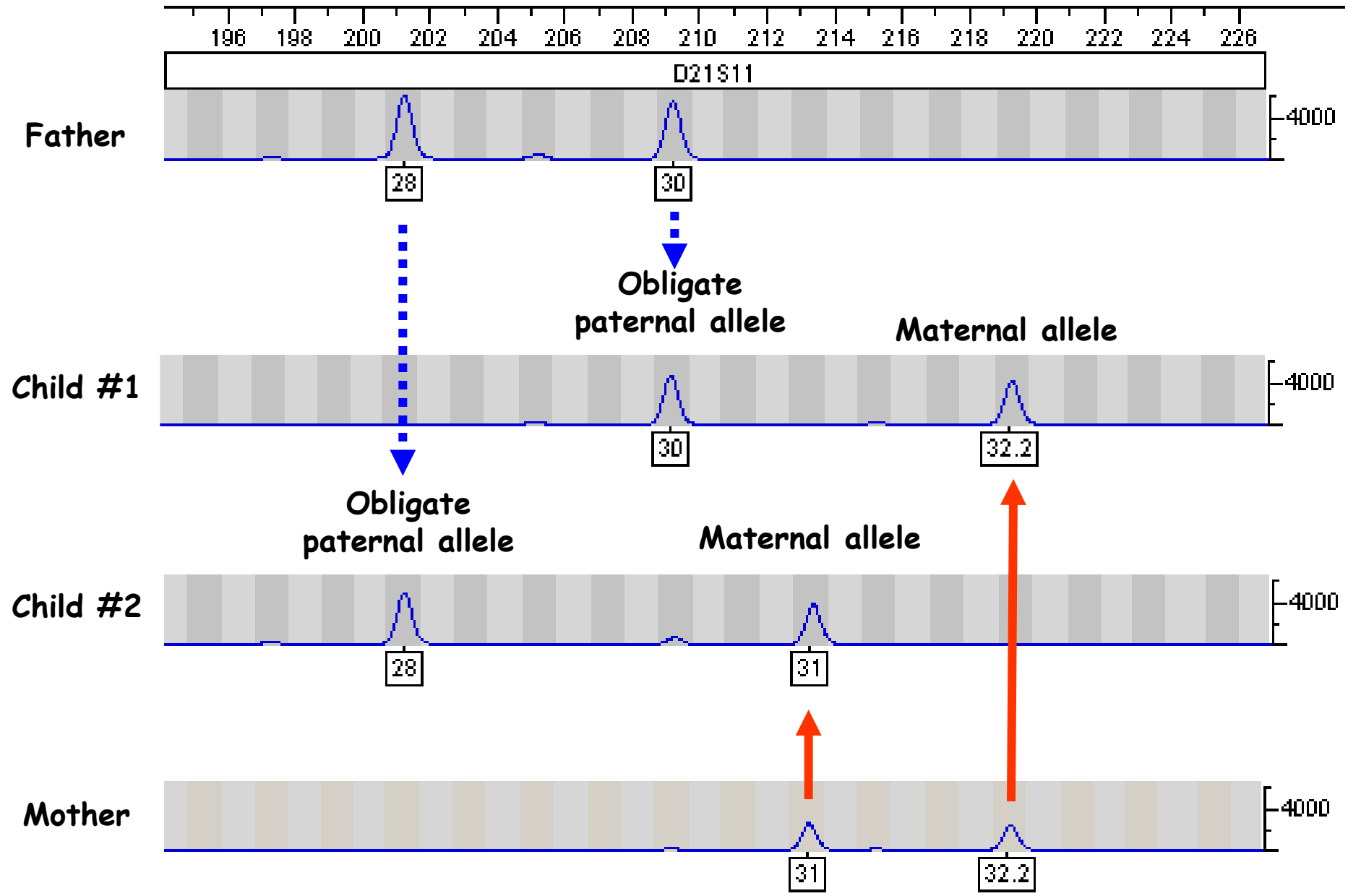
Resulting genotype combinations and frequencies

Father gametes (sperm)

A p

a q

	A p	a q
A p	AA p^2	aA qp
a q	Aa pq	aa q^2

Punnett square

AA p^2

Aa $2pq$

aa q^2
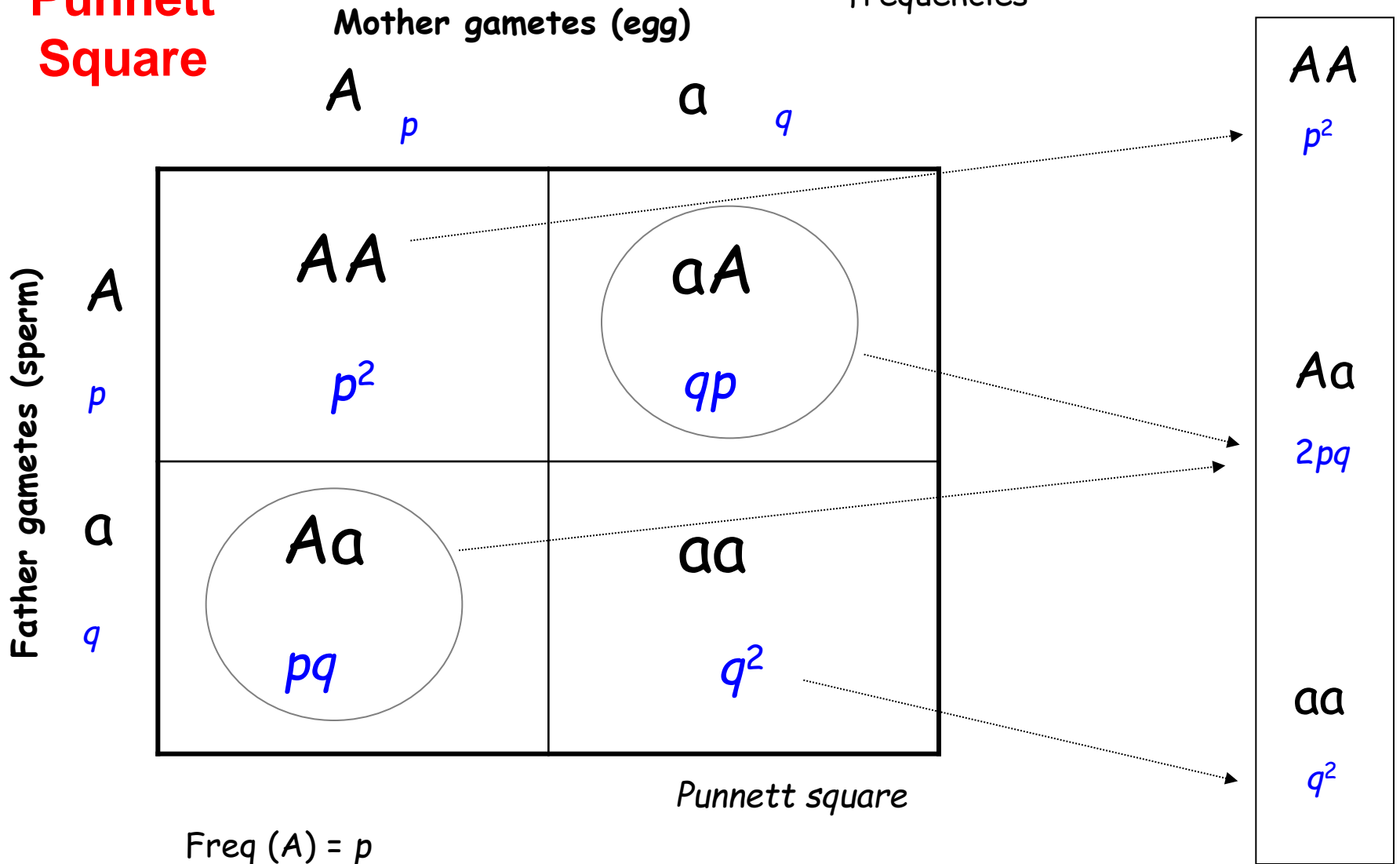
Freq (A) = p

Freq (a) = q

$$p + q = 1 \qquad (p + q)^2 = p^2 + 2pq + q^2$$

Relationship between Allele Frequency and Genotype Frequency

Frequency of *A* allele (*p*)

1.0 0.8 0.6 0.4 0.2 0.0

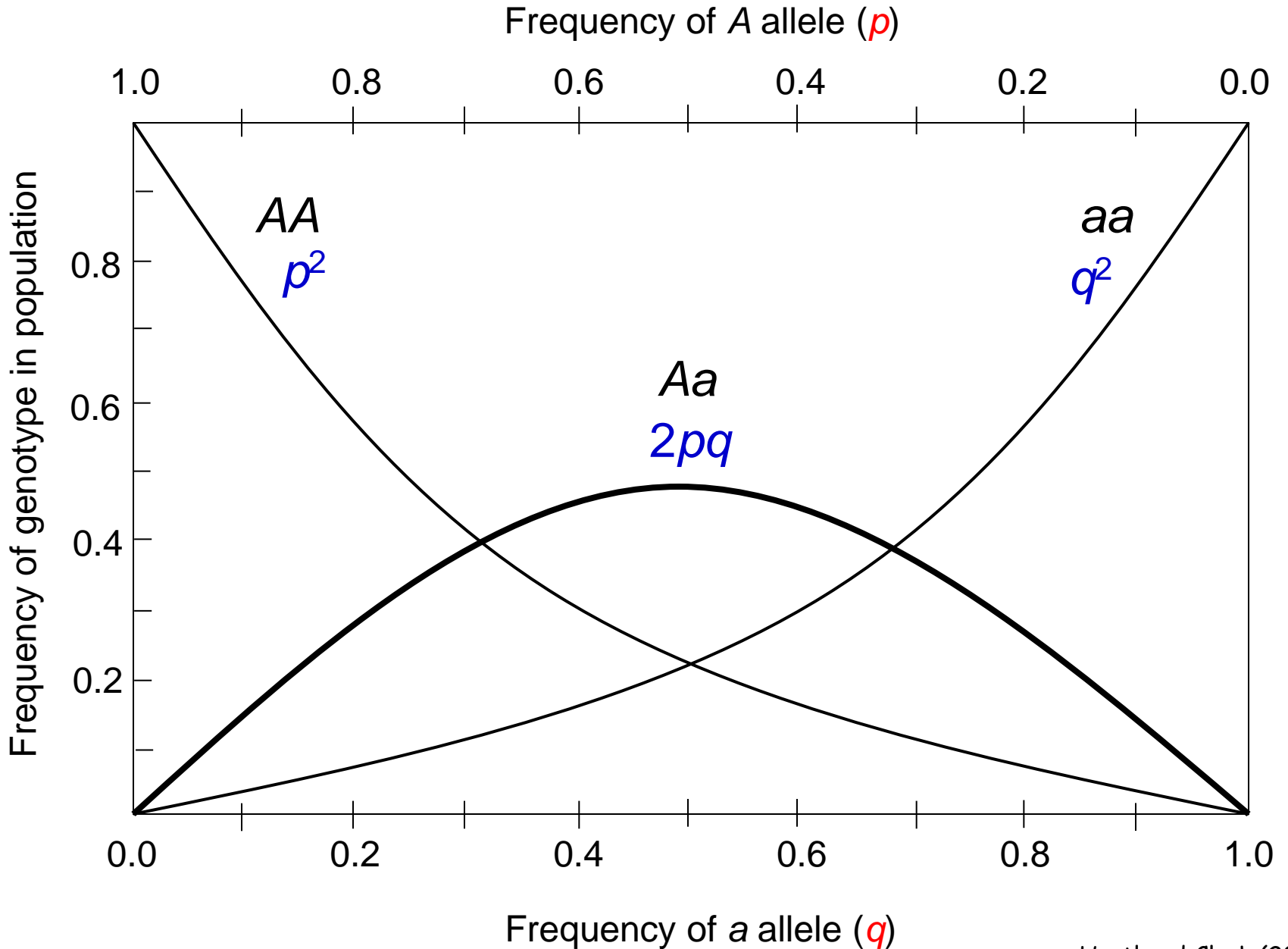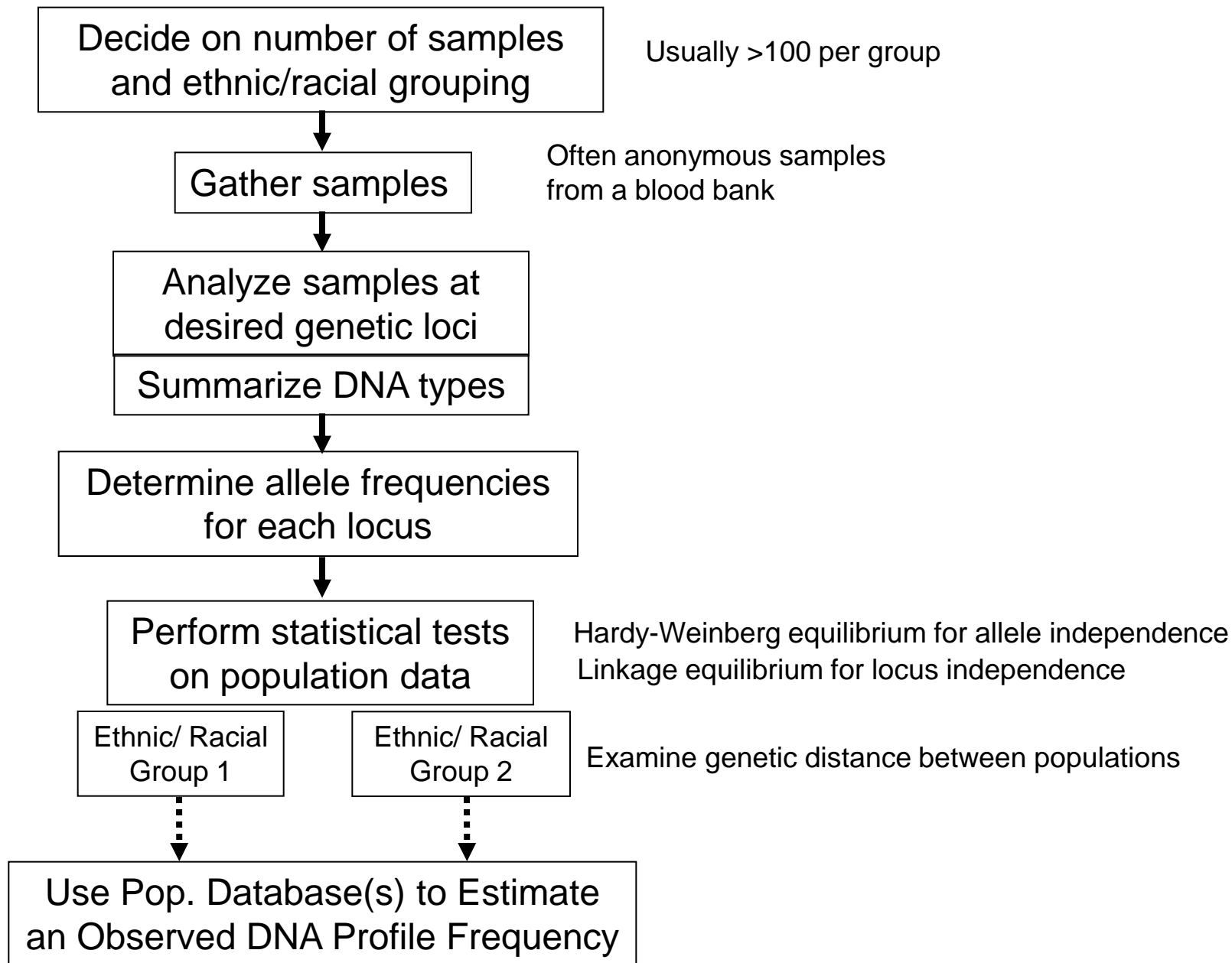Frequency of genotype in population

AA
p^2

Aa
2*pq*

aa
q^2

Frequency of *a* allele (*q*)

0.0 0.2 0.4 0.6 0.8 1.0

Hartl and Clark (2007)

Decide on number of samples
and ethnic/racial grouping

Usually >100 per group

Gather samples

Often anonymous samples
from a blood bank

Analyze samples at
desired genetic loci

Summarize DNA types

Determine allele frequencies
for each locus

Perform statistical tests
on population data

Hardy-Weinberg equilibrium for allele independence
Linkage equilibrium for locus independence

Ethnic/ Racial
Group 1

Ethnic/ Racial
Group 2

Examine genetic distance between populations

Use Pop. Database(s) to Estimate
an Observed DNA Profile Frequency

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) PM = (P1)(P2)…(Pn)

Microsatellite allele frequency diagrams
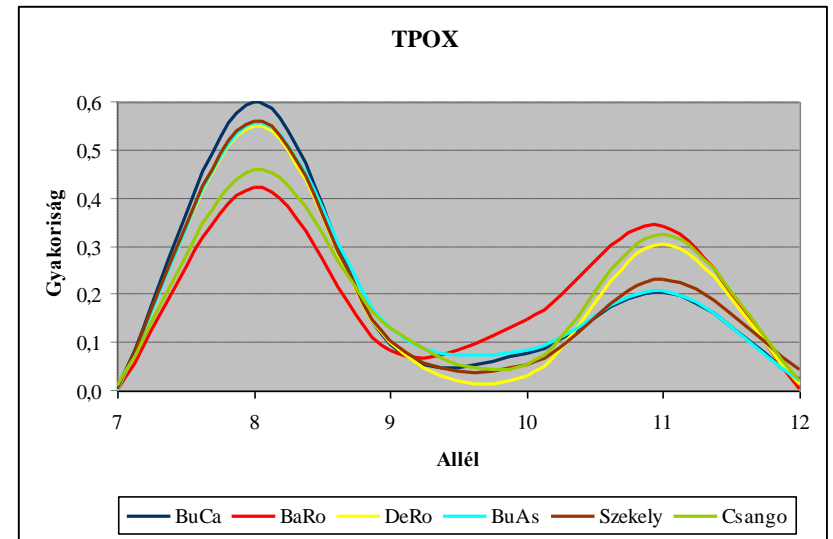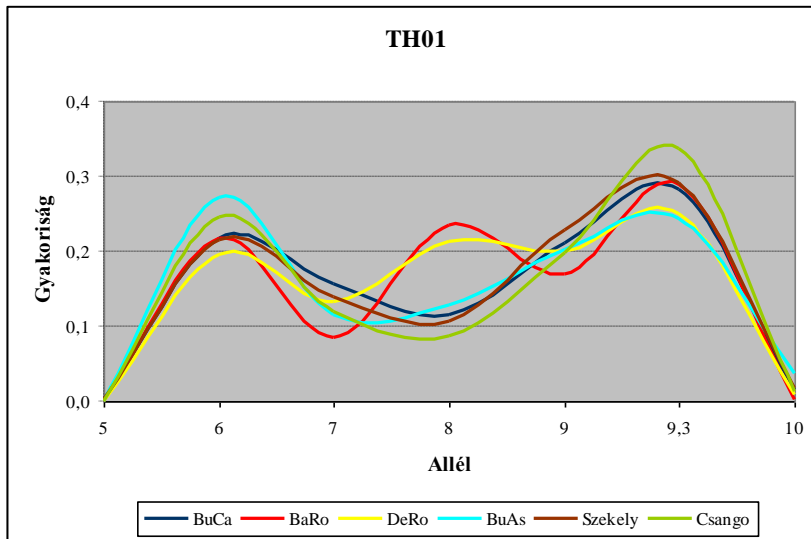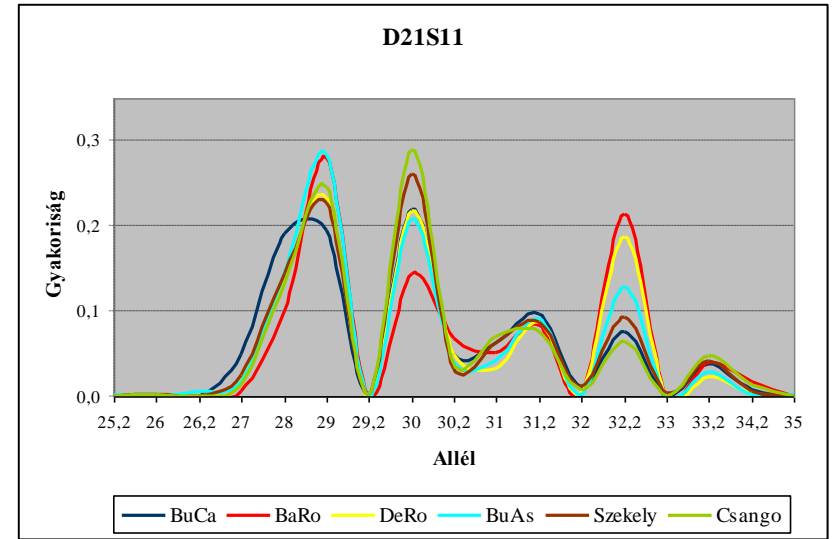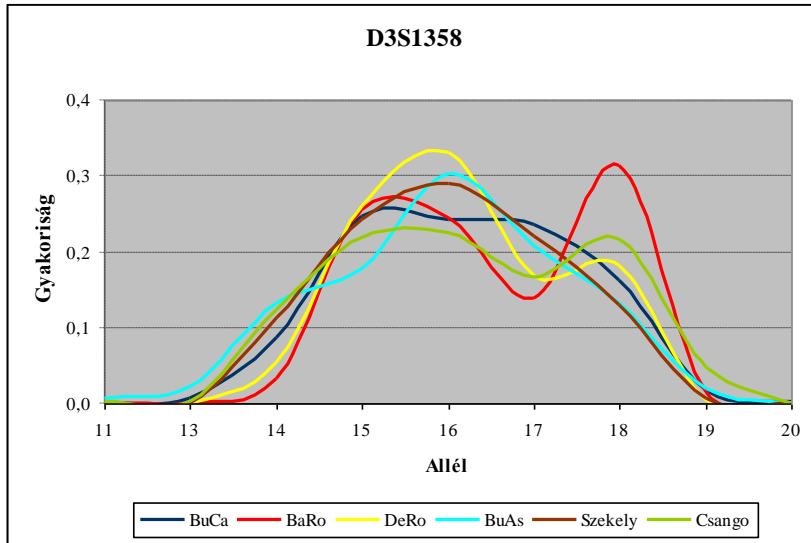
Allele- and genotype frequencies

Genotype Array	8	9	10	11	12	13	14	15		Allele Count	Observed Frequency
D13S317	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	

> Genotype **11,12** was observed **54 times** in 302 persons (604 analysed chromosomes)

ORIGINAL ARTICLE

Same DNA Sample Run with Each of the ABI STR Kits

PCR Product Size (bp)

Kit	Power of Discrimination
Blue	1:5000
Green I	1:410
Profiler™	$1:3.6 \times 10^9$
Profiler Plus™	$1:9.6 \times 10^{10}$
COfiler™	$1:8.4 \times 10^5$
SGM Plus™	$1:3.3 \times 10^{12}$
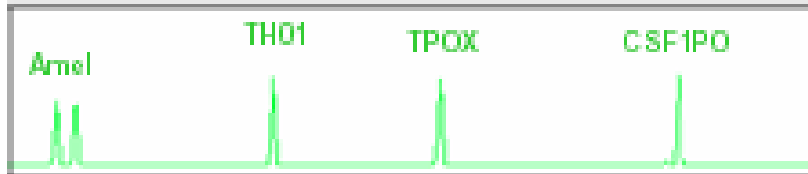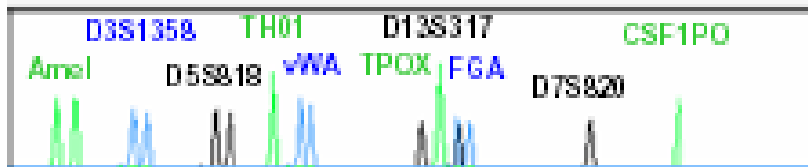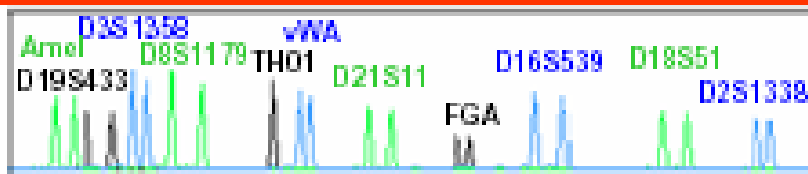
Likelihood ratio calculation /Bayes theorem/
Hypothesis testing

in case of matching DNA profiles:

$$LR = \frac{Pr(E|H_1)}{Pr(E|H_2)} = \frac{Pr(STR\ profile|the\ DNA\ comes\ from\ the\ suspect)}{Pr(STR\ profile|the\ DNA\ comes\ from\ a\ random\ person)}$$

The LR equals the probability of the evidence given the hypothesis of the prosecution (H1 or Hp) divided by the probability of the evidence given the hypothesis of the defense (H2 or Hd).
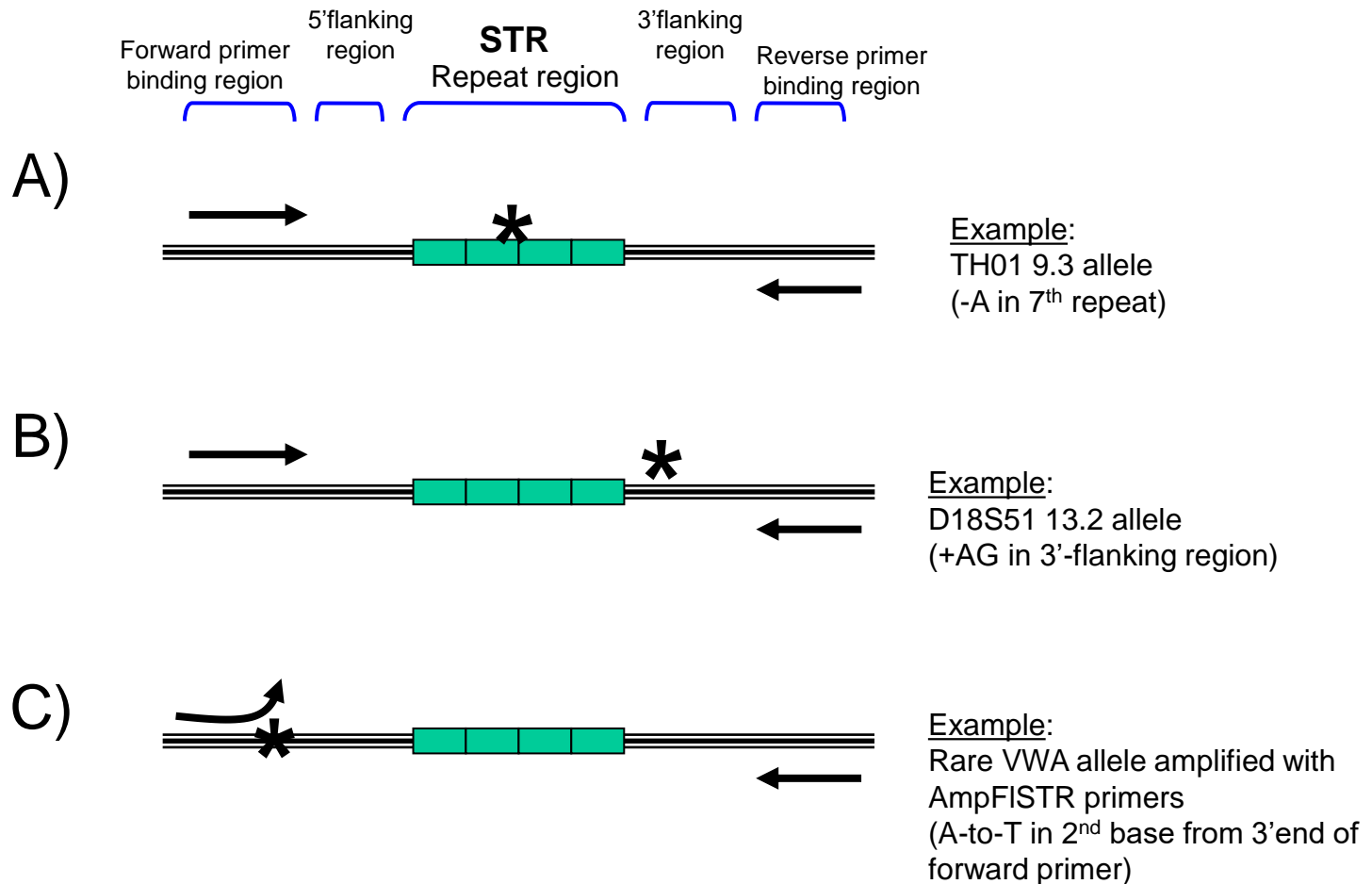
Example:

$$LR = \frac{Pr(E|H_p)}{Pr(E|H_d)} = \frac{1}{p^2} = \frac{1}{(0.139)^2} = \frac{1}{0.0193} = 51.8.$$ Homozygous DNA profile

$$LR = \frac{Pr(E|H_p)}{Pr(E|H_d)} = \frac{1}{2pq} = \frac{1}{2(0.134)(0.147)} = \frac{1}{0.0394} = 25.4.$$ Heterozygous DNA profile
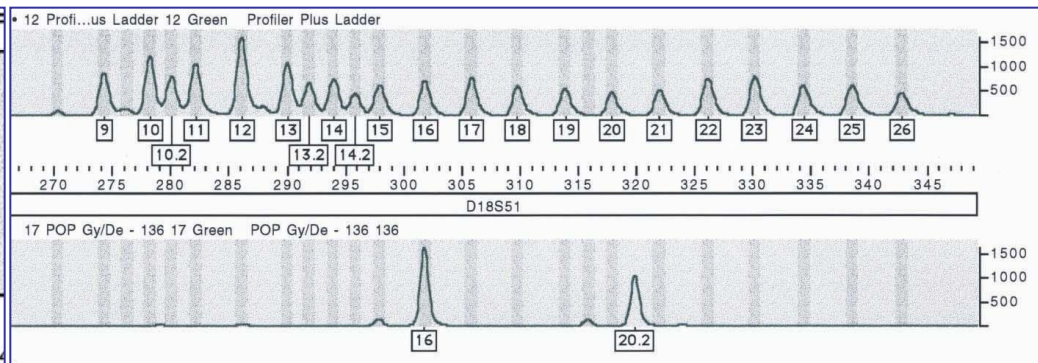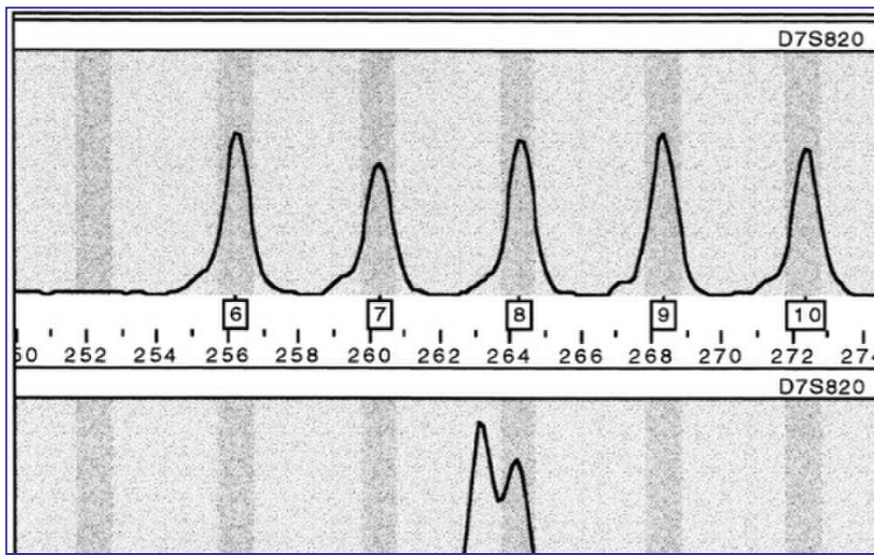
STR System	Maternal Meioses (%)	Paternal Meioses (%)	Number from either	Total Number of Mutations	Mutation Rate
CSF1PO	95/304,307 (0.03)	982/643,118 (0.15)	410	1,487/947,425	**0.16%**
FGA	205/408,230 (0.05)	2,210/692,776 (0.32)	710	3,125/1,101,006	**0.28%**
TH01	31/327,172 (0.009)	41/452,382 (0.009)	28	100/779,554	**0.01%**
TPOX	18/400,061 (0.004)	54/457,420 (0.012)	28	100/857,481	**0.01%**
VWA	184/564,398 (0.03)	1,482/873,547 (0.17)	814	2,480/1,437,945	**0.17%**
D3S1358	60/405,452 (0.015)	713/558,836 (0.13)	379	1,152/964,288	**0.12%**
D5S818	111/451,736 (0.025)	763/655,603 (0.12)	385	1,259/1,107,339	**0.11%**
D7S820	59/440,562 (0.013)	745/644,743 (0.12)	285	1,089/1,085,305	**0.10%**
D8S1179	96/409,869 (0.02)	779/489,968 (0.16)	364	1,239/899,837	**0.14%**
D13S317	192/482,136 (0.04)	881/621,146 (0.14)	485	1,558/1,103,282	**0.14%**
D16S539	129/467,774 (0.03)	540/494,465 (0.11)	372	1,041/962,239	**0.11%**
D18S51	186/296,244 (0.06)	1,094/494,098 (0.22)	466	1,746/790,342	**0.22%**
D21S11	464/435,388 (0.11)	772/526,708 (0.15)	580	1,816/962,096	**0.19%**
Penta D	12/18,701 (0.06)	21/22,501 (0.09)	24	57/41,202	**0.14%**
Penta E	29/44,311 (0.065)	75/55,719 (0.135)	59	163/100,030	**0.16%**
D2S1338	15/72,830 (0.021)	157/152,310 (0.10)	90	262/225,140	**0.12%**
D19S433	38/70,001 (0.05)	78/103,489 (0.075)	71	187/173,490	**0.11%**
SE33 (ACTBP2)	0/330 (<0.30)	330/51,610 (0.64)	None reported	330/51,940	**0.64%**

STR loci mutation rate: 10^{-3} - 10^{-4} / meiosis

'Stepwise' mutation model

Microsatellite point mutations

5'flanking region

3'flanking region

Forward primer binding region

STR Repeat region

Reverse primer binding region

A)

Example:
TH01 9.3 allele
(-A in 7th repeat)

B)

Example:
D18S51 13.2 allele
(+AG in 3'-flanking region)

C)

Example:
Rare VWA allele amplified with AmpFlSTR primers
(A-to-T in 2nd base from 3'end of forward primer)

Figure 6.8, J.M. Butler (2005) *Forensic DNA Typing*, 2nd Edition © 2005 Elsevier Academic Press

Variant microsatellite alleles: Null-alleles

D7S820

Allél	Fragmens méret	5' Flanking régió
8.-1	205 bp	
9.+1	211 bp	

D13S317 allele 12

5'-gggttgctggacatggtatcACAGAAGTCTGGGGATGTGGA---N82---(A
gaccaacaattcaagctctc-3'

D18S51

Allél	Fragmenthossz	5' flanking régió	Repeat régió	3' flanking régió
16	299 bp		- (AGAA)16	- AAAG AGAGAG -
20.2	317 bp		- (AGAA)21	- AG AGAGAG -
15*	295 bp		- (ATAG)15	- AAAG AGAGAG -

FGA

Homozygous phenotype
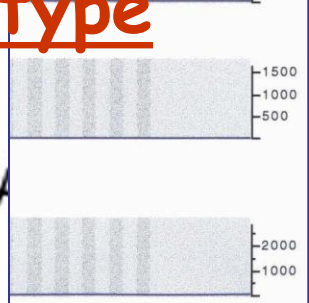
Genetic Structure – Analysing of MOlaculare VAriance (AMOVA)

Breed identification?

STRUCTURE statistics

1. Ancient   2. Sheperd   3. Hunting   4. Job
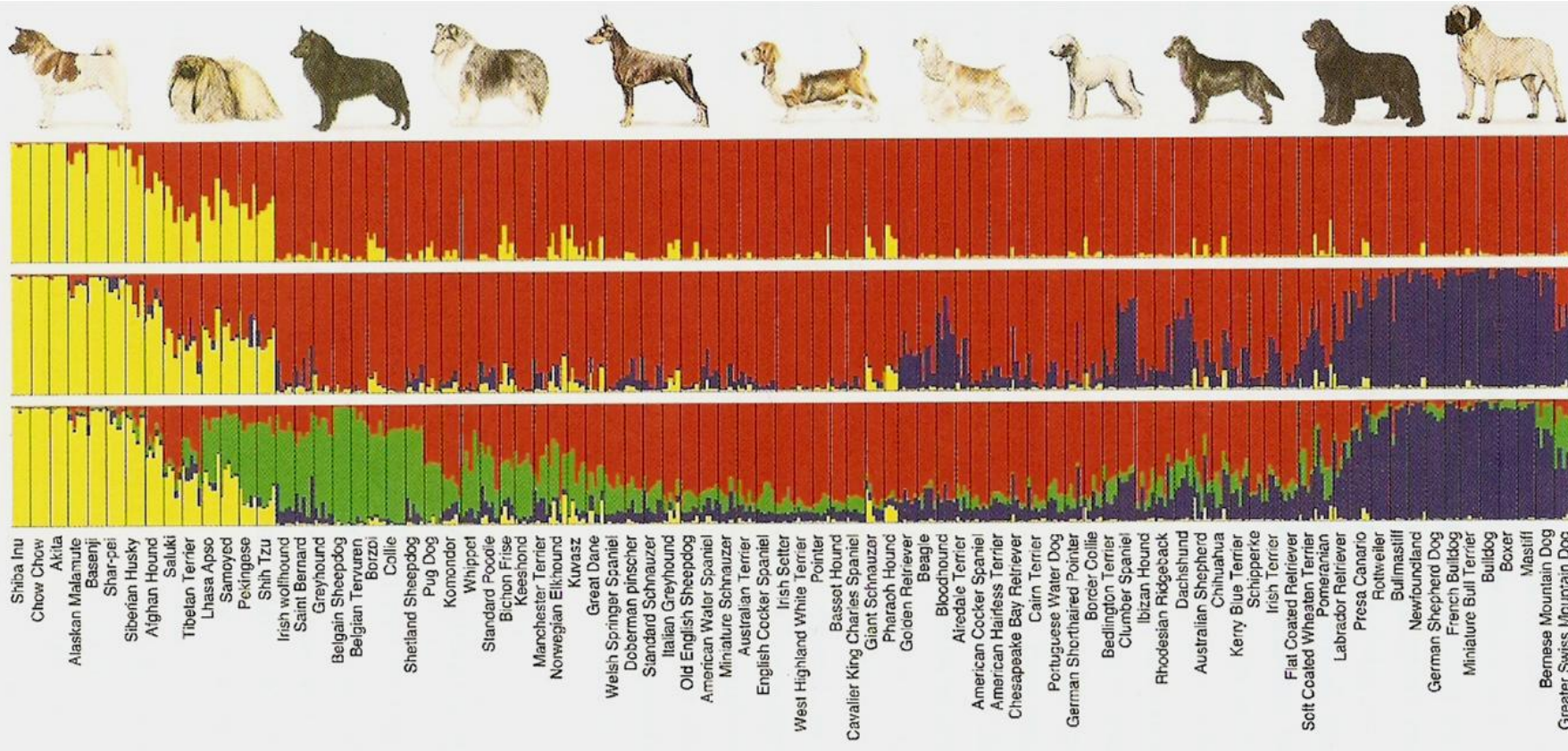
- 85 breeds (n=414)
- 95 microsatellites (dinucleotide repeat)
- 99% correct prediction!

Human melanogenesis

Eumelanin

Phaeomelanin

α-Melanocyte Stimulating Hormon

Melanocortin 1 Receptor

MIcrophthalmia Transcription Factor

MIcrophthalmia Transcription Factor

TYRosinase

DopaChrome Tautomerase

TYRosinase

cross-membrane transport:

Keratinocyte

Bouakaze et al., Int J Legal Med (2009) 123:315
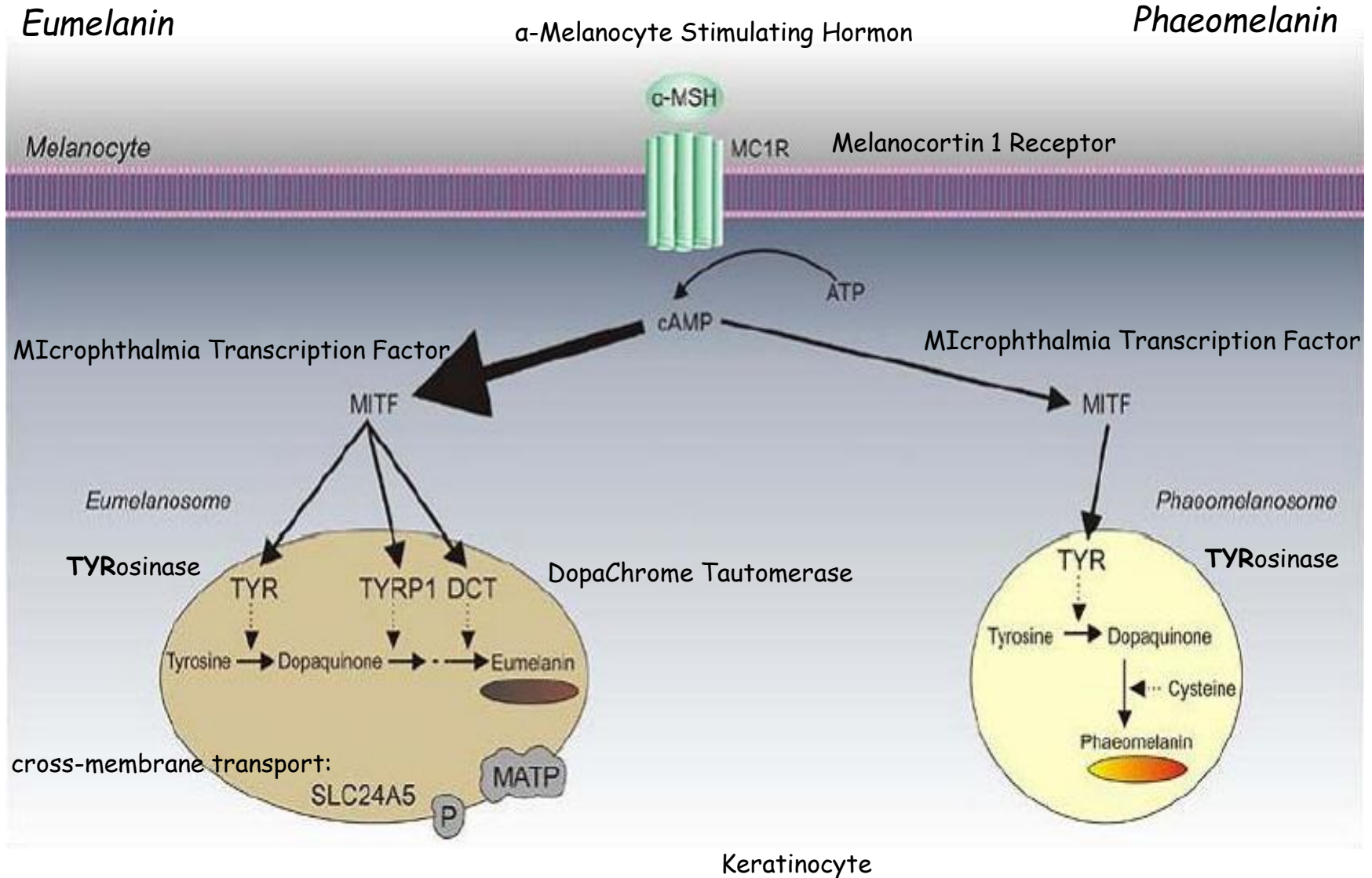
Genes responsible for skin pigmentation

Principal skin pigmentation candidate genes

Locus	Chromosome	Protein	Mut phenotype	Function
Melanosome proteins				
TYR	11q14-11q21	Tyrosinase	OCA1	Oxidation of tyrosine
TYRP1	9p23	Gp75, TRYP1	OCA3	DHICA-oxidase, TYR stabilisation
DCT	13q32	DCT, TRYP2		Dopachrome tautomerase
OCA2	15q11.2-15q12	P-protein	OCA2 (eye)	pH of melanosome
SLC45A2	5p14.3-5q12.3	MATP, AIM-1	OCA4 (skin)	Melansome maturation
SLC24A5	15q21.1	Cation exchanger		Melanosome precursor
Signal proteins				
ASIP	20q11.2-20q12	Agouti signal protein		MC1R antagonist
MC1R	16q24.3	MSH receptor	Red hair (skin)	G-protein coupled receptor
POMC	16q24.3	MSH receptor	Red hair	MC1R antagonist
OA1	Xp22.3	OA1 protein	OA1	G-protein coupled receptor
MITF	3p12.3-3p14.1	MITF	Waardenburg	Transcription factor
Proteins involved in melanosome transport or uptake by keratinocytes				
MYO5A	15q21	MyosinVa	Griscelli	Motor protein
RAB27A	15q15-15q21.1	Rab27a	Griscelli	RAS family protein
HPS1	10q23.1-10q23.3	HPS1	Hermansky-Pudlak	Organelle biogenesis and size
HPS6	10q24.32	HPS6	Hermansky-Pudlak	Organelle biogenesis

ACTH: adrenocorticotrophin hormone; DCT: dopachrome tautomerase; DHICA: 5,6-dihydroxyindole-2-carboxylic acid; MATP: membrane-associated transporter protein; MC1R: melanocortin-1 receptor; MITF: microphthalmia-associated transcription factor; MSH: melanocyte stimulating hormone; OCA: oculocutaneous albinism; POMC: pro-opiomelanocortin; TYRP1: tyrosinase-related protein 1.

MC1R gene mutations

Mutations in the MC1R gene, their penetrance and functional significance (where known)

Mutation	Type	Designation	Penetrance (odds ratio)	Functional significance	References (for functional significance and penetrance)
R151C	Mis-sense	R	63.3	Altered cellular location	[16,26]
R160W	Mis-sense	R	63.3	Altered cellular location	[16,26]
D294H	Mis-sense	R	63.3	Impaired G coupling ability	[26,27]
D84E	Mis-sense	R	63.3	Altered cellular location	[16,26]
I155T	Mis-sense	Lack of statistical data—strong familial association		Altered cellular location	[16,26]
V92M	Mis-sense	r	5.1	Reduced α-MSH binding	[26,28,29]
V60L	Mis-sense	r	5.1		[26]
R163Q	Mis-sense	r	5.1	Slightly reduced α-MSH binding	[26,29]
R142H	Mis-sense	Lack of statistical data—strong familial association			[26]

- MC1R alleles possess different activity levels
- 317 AA and 7 transmembran domen
- SNPs: RHC phenotype - neanderthal pigmentation
- Phenotype prediction? Genetic tests?

SNaPshot: A Primer Extension Assay Capable of Multiplex Analysis

Minisequencing (SNaPshot assay)

Allele-specific primer extension across the SNP site with fluorescently labeled ddNTPs; mobility modifying tails can be added to the 5'-end of each primer in order to spatially separate them during electrophoresis.
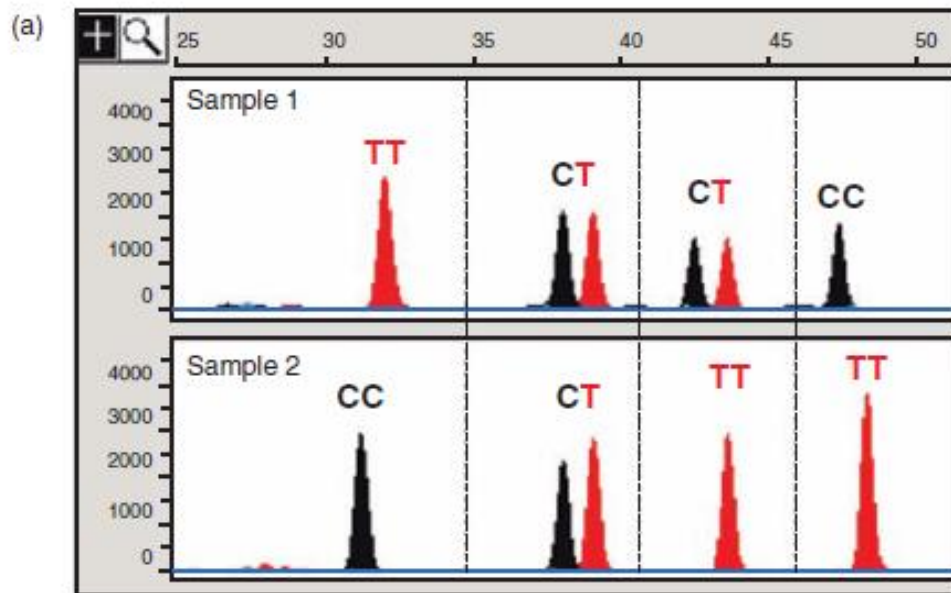
(b) (TTTTT)–**primer1** (chromosome 20)-**ddT/ddT**

(TTTTT)–(TTTTT)–**primer2** (chromosome 6)-**ddC/ddT**

(TTTTT)–(TTTTT)–(TTTTT)–**primer3** (chromosome 14)-**ddC/ddT**

(TTTTT)–(TTTTT)–(TTTTT)–(TTTTT)–**primer4** (chromosome 1)-**ddC/ddC**

FIGURE 12.2   Allele-specific primer extension results using four autosomal SNP markers on two different samples (a). SNP loci are from separate chromosomes (1, 6, 14, and 20) and therefore unlinked. Electrophoretic resolution of the SNP primer extension products occurs due to poly(T) tails that are 5 nucleotides different from one another (b).

SNP genotyping of 10 pigmentation genes (SNaPshot)

Bouakaze et al., Int J Legal Med (2009) 123:315
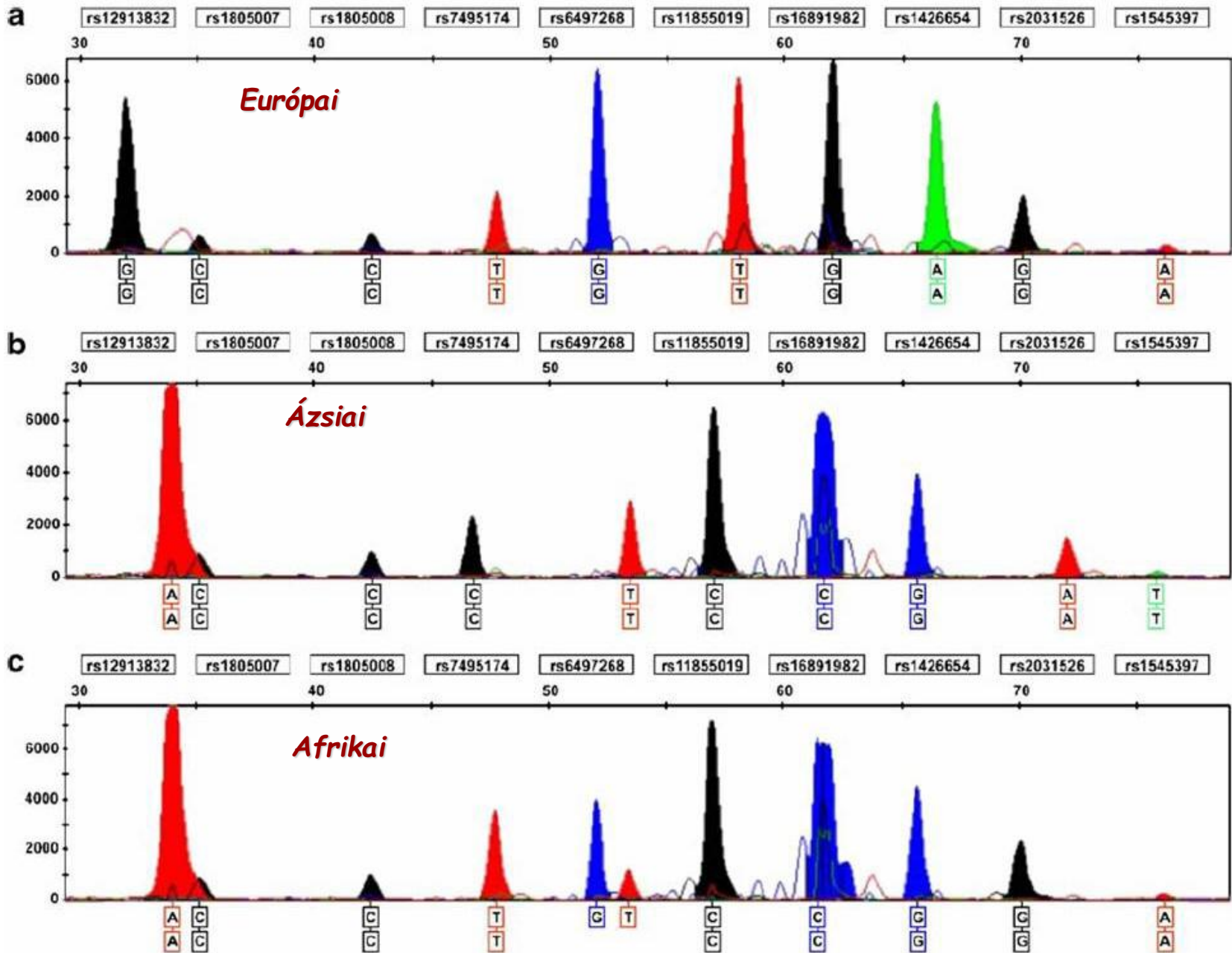
Sample	Self-reported pigmentary traits			rs12913832 HERC2	rs1805007 MC1R	rs1805008 MC1R	OCA2 diplotype[a]	rs16891982 SLC24A2	rs1426654 SLC24A5	rs2031526 DCT	rs1545397 OCA2	Inferred ancestry of individuals[b]		
	Eye color	Hair color	Skin color									European	Asian	African
E1	Blue	Red	Fair	G/G	C/C	C/T	TGT/TGT	G/G	A/A	G/G	A/A	0.963	0.012	0.024
E2	Green	Light brown	Fair	G/G	C/C	C/C	TGT/TGT	G/G	A/A	A/G	A/A	0.954	0.021	0.025
E3	Blue	Blond	Fair	G/G	C/C	C/C	TGT/TGT	G/G	A/A	A/A	A/A	0.954	0.024	0.022
E4	Blue	Blond	Fair	G/G	C/C	C/C	TGT/TGT	G/G	A/A	A/G	A/A	0.960	0.020	0.020
E5	Blue/gray	Auburn	Fair	G/G	C/T	C/C	TGT/TGT	G/G	A/A	G/G	A/A	0.961	0.013	0.026
E6	Green/gray	Light brown	Fair	G/G	C/C	C/C	TGT/TGT	C/G	A/A	G/G	A/A	0.787	0.038	0.175
E7	Green/hazel	Light brown	Fair	A/G	C/C	C/C	TGT/TGT	G/G	A/A	A/G	A/A	0.955	0.022	0.024
E8	Green/hazel	Dark brown	Fair	A/A	C/C	C/C	TGT/CTC	G/G	A/A	G/G	A/A	0.961	0.013	0.027
E9	Green/hazel	Dark brown	Fair	A/A	C/C	C/C	TTT/CTC	G/G	A/A	G/G	A/A	0.963	0.013	0.024
E10	Blue	Light brown	Fair	G/G	C/C	C/C	TGT/TGT	C/G	A/A	G/G	A/A	0.789	0.049	0.163
E11	Green	Auburn	Fair	G/G	C/T	C/C	TGT/TGC	G/G	A/A	G/G	A/A	0.958	0.014	0.028
E12	Blue/hazel	Light brown	Fair	A/G	C/C	C/C	TGT/TTT	G/G	A/A	G/G	A/A	0.962	0.012	0.026
E13	Blue/hazel	Light brown	Fair	A/G	C/C	C/C	TGT/TTT	G/G	A/A	G/G	A/A	0.965	0.013	0.022
E14	Green	Light brown	Fair	G/G	C/C	C/C	TGT/TGT	C/G	A/A	G/G	A/T	0.763	0.165	0.073
E15	Brown	Dark brown	Fair	A/G	C/C	C/C	TGT/TGT	G/G	A/A	A/G	A/A	0.957	0.022	0.021
E16	Brown	Dark brown	Fair	A/A	C/C	C/C	TGT/CTC	C/G	A/A	A/G	A/T	0.669	0.283	0.048
E17	Green/hazel	Dark brown	Medium	A/G	C/C	C/C	TGT/TTT	C/G	A/A	G/G	A/T	0.755	0.170	0.076
E18	Blue	Light brown	Fair	G/G	C/C	C/C	TGT/TGT	G/G	A/A	G/G	A/T	0.935	0.045	0.021
E19	Brown	Red	Fair	A/G	C/T	C/C	TGT/TGT	G/G	A/A	G/G	A/A	0.964	0.013	0.022
E20	Green	Light brown	Fair	G/G	C/C	C/C	TGT/TGT	C/G	A/A	G/G	A/A	0.792	0.047	0.161
E21	Green/gray	Blond	Fair	G/G	C/C	C/C	TGT/TGT	G/G	A/A	A/G	A/A	0.957	0.022	0.021
E22	Blue	Light brown	Fair	G/G	C/C	C/C	TGT/TGT	G/G	A/A	G/G	A/A	0.959	0.014	0.026
E23	Green/hazel	Light brown	Fair	A/G	C/C	C/C	TGT/TTT	G/G	A/A	A/G	A/A	0.957	0.020	0.022
E24	Green	Light brown	Fair	G/G	C/C	C/C	TGT/TGT	C/G	A/A	G/G	A/A	0.786	0.049	0.166
E25	Brown	Red	Fair	A/G	C/C	T/T	TGT/TGC	G/G	A/A	G/G	A/A	0.963	0.014	0.023
E26	Blue	Light brown	Fair	G/G	C/C	C/C	TGT/TGT	G/G	A/A	A/G	A/A	0.954	0.021	0.025
E27	Blue	Red	Fair	G/G	C/C	C/T	TGT/TGT	G/G	A/A	G/G	A/A	0.958	0.014	0.028
Af1	Brown	Black	Dark	A/A	C/C	C/C	TGC/TTC	C/C	G/G	A/G	A/A	0.028	0.094	**0.878**
Af2	Brown	Black	Dark	A/A	C/C	C/C	TGC/TTC	C/C	G/G	G/G	A/A	0.023	0.031	**0.946**
Af3	Brown	Black	Dark	A/A	C/C	C/C	TGC/TTC	C/C	A/G	G/G	A/A	0.164	0.041	**0.795**
As1	–	–	–	A/A	C/C	C/C	TTT/CTC	C/C	G/G	A/G	A/T	0.042	**0.649**	0.308
As2	–	–	–	A/A	C/C	C/C	CTC/CTC	C/C	G/G	A/G	T/T	0.020	**0.921**	0.060
As3	–	–	–	A/A	C/C	C/C	CTC/CTC	C/C	G/G	A/A	T/T	0.013	**0.964**	0.023
As4	–	–	–	A/G	C/C	C/C	TTT/CGC	C/C	A/G	A/A	A/T	0.212	**0.708**	0.080
As5	–	–	–	A/A	C/C	C/C	TTC/CGC	C/C	A/G	A/G	T/T	0.019	**0.922**	0.059
As6	–	–	–	A/A	C/C	C/C	CTC/CTC	C/G	G/G	A/A	T/T	0.119	**0.858**	0.023

E European modern sample, *Af* African modern sample, *As* Asian modern sample

[a] OCA2 diplotype correspond to markers rs7495174/rs6497268/rs11855019. OCA2 diplotype and rs12913832 genotype predictive of blue eye color phenotype are underlined

[b] Probability of being from European/Asian/African population determined using the STRUCTURE program. The greatest probability, most likely estimate of ancestry, is indicated in bold