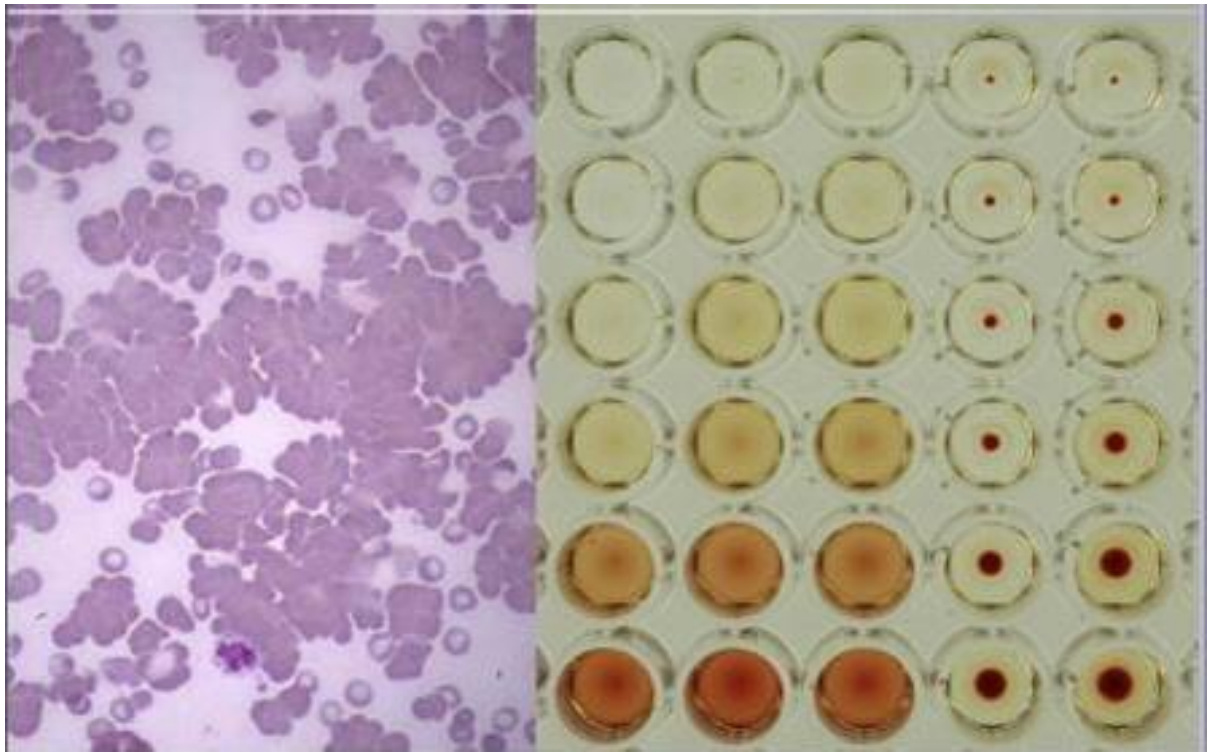
GENETICS AND POPULATION GENETICS

Genetic polymorphisms

ELTE Faculty of Sciences Department of Genetics

First genetic marker: ABO blood group system



Landsteiner, 1900

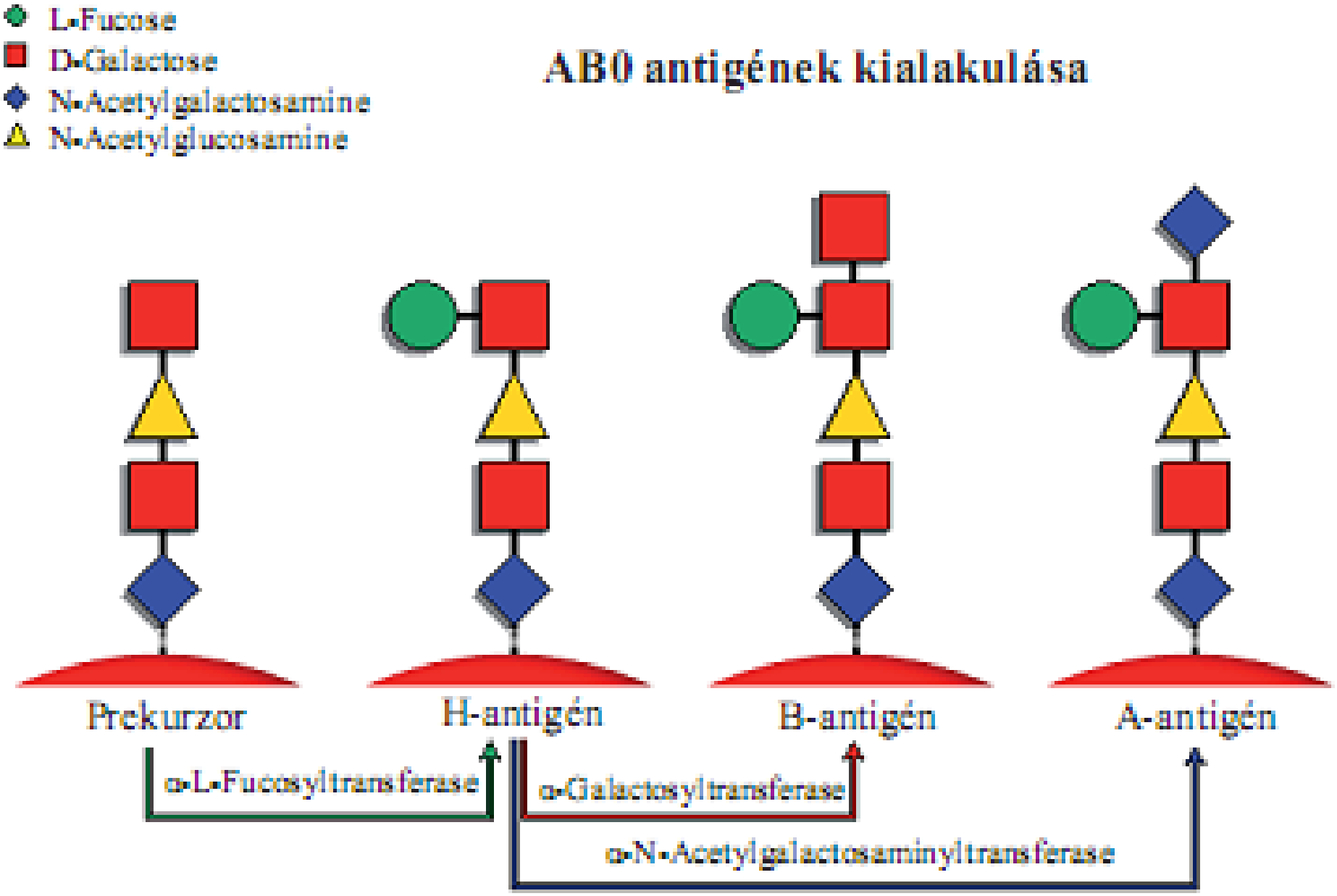
FELIX BERNSTEIN (1933)

Two hypotheses of blood group inheritance

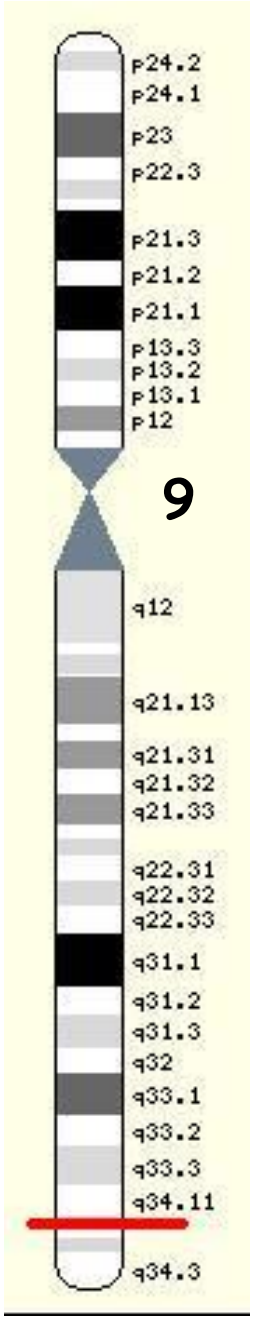
Group	VON DUNGERN and HIRZFELD		BERNSTEIN		Observed proportion
	Genotype	Expected proportion	Genotype	Expected proportion	
O	*aa bb*	$p_a^2\ p_b^2$	*OO*	p_O^2	0.294
A	*A– bb*	$(1 - p_a^2)p_b^2$	*AA, OA*	$p_A^2 + 2p_Op_A$	0.422
B	*aa B–*	$p_a^2\ (1 - p_b^2)$	*BB, OB*	$p_B^2 + 2p_Op_B$	0.206
AB	*A– B–*	$(1 - p_a^2)(1 - p_b^2)$	*AB*	$2p_Ap_B$	0.078
Total		1		1	1.000

The expected proportions assume Hardy-Weinberg ratios and linkage equilibrium. The observed proportions are from 502 Japanese (BERNSTEIN 1925).

AB0 antigének kialakulása



Various Alleles at the ABO Locus



Exon Number	6	7														
Nucleotide Position	261	297	467	526	664	675	681	703	771	796	802	803	829	871	930	1054	1060
A alleles																	
A101	G	A	C	C	T	C	G	G	C	C	G	G	G	G	G	C	C
A102	*	*	T	*	*	*	*	*	*	*	*	*	*	*	*	*	*
A201	*	*	T	*	*	*	*	*	*	*	*	*	*	*	*	*	Δ
A301	*	*	*	*	*	*	*	*	*	*	*	*	*	A	*	*	*
Ax01	*	*	*	*	A	*	*	*	*	*	*	*	*	*	*	*	*
cis-AB01	*	*	T	*	*	*	*	*	*	*	*	C	*	*	*	*	*
B alleles																	
B101	*	G	*	G	*	T	*	A	*	A	*	C	*	*	A	*	*
B301	*	G	*	G	*	T	*	A	*	A	*	C	*	*	A	T	*
B(A)01	*	G	*	G	*	*	*	*	*	A	*	C	*	*	A	*	*
O alleles																	
O01	Δ	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*
O02	Δ	G	*	*	A	*	A	*	T	*	*	*	A	*	*	*	*
O03	*	G	*	G	*	*	*	*	*	*	A	*	*	*	*	*	*
Possible Amino Acid Change	Frameshift	No change	P156L	R176G	F216I	No change	No change	G235S	No change	L266M	G268R	G268A	V277M	D291N	No change	R352W	Frameshift

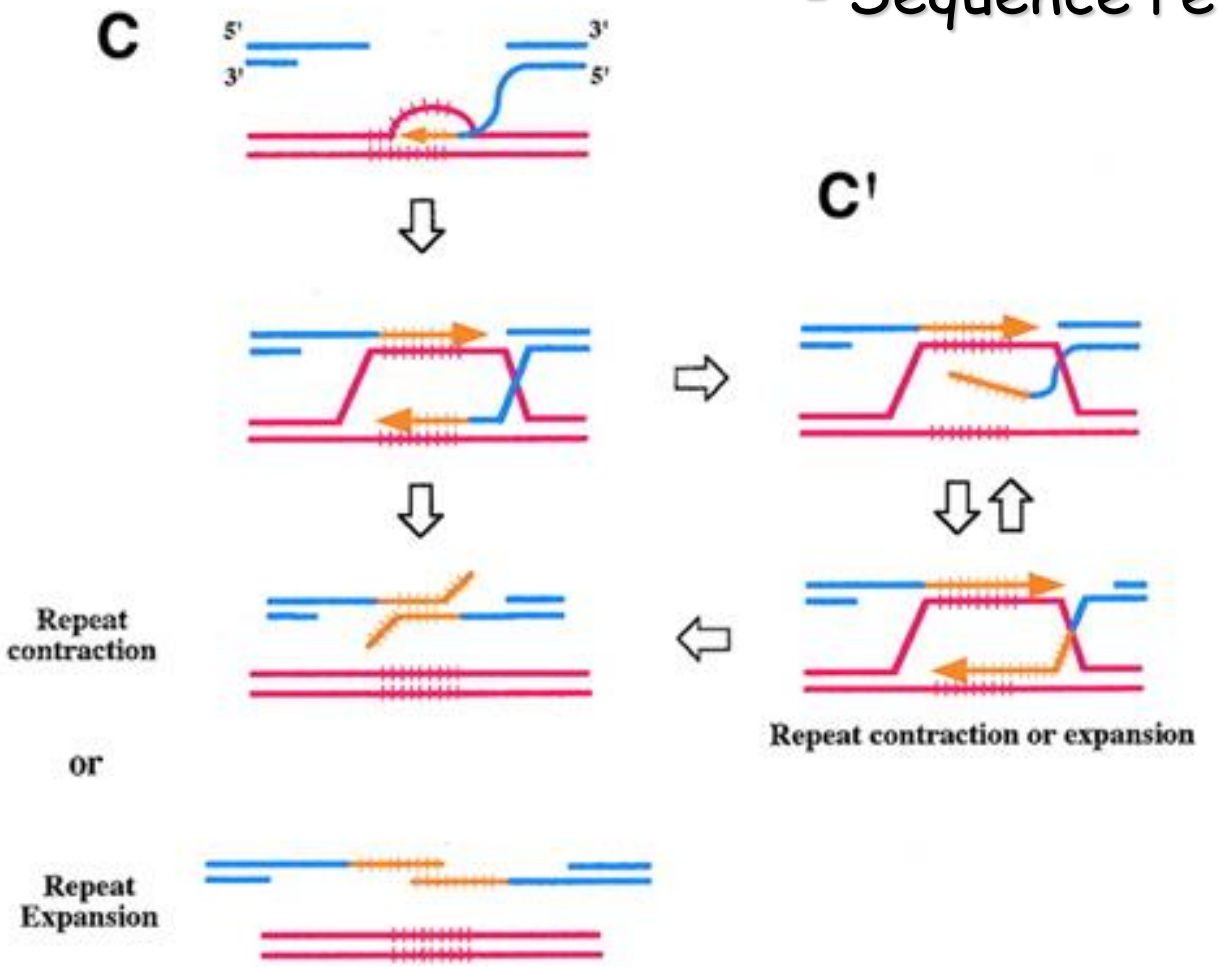
First results of Human Genome Project

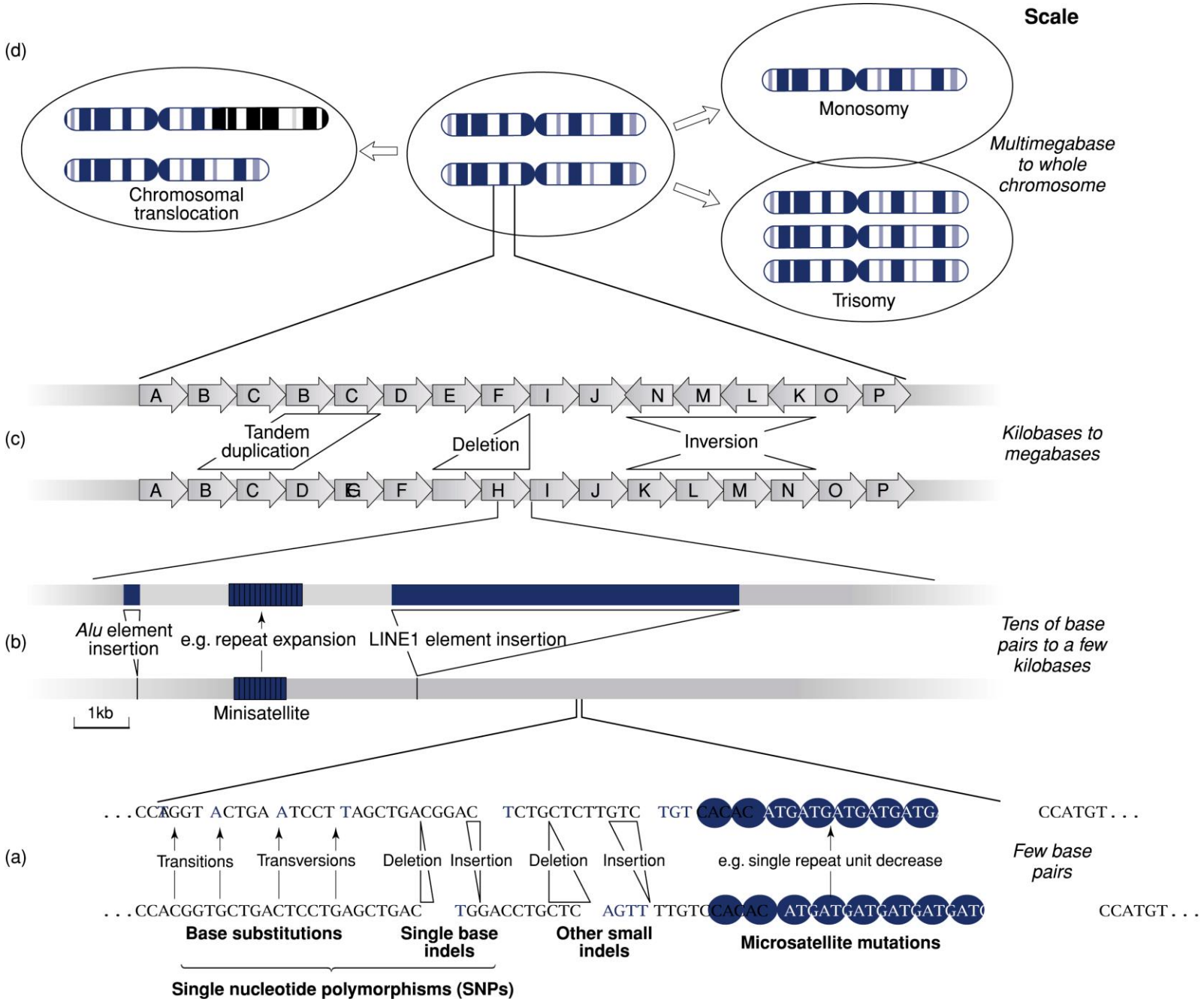
- First draft in 2001 (Science, Nature)
- The most large whole genome determined
- Structure and organisation similare to each eukaryotes (model organizms)
- Unexpectedly low amount of protein coding genes (~20000)
- Emerging number of RNA genes (snRNA, lcnRNA, miRNA)
- Low amount of protein coding sequences (exons): < 1 %
- **Excess amount of repetitive sequences: Mobile elements?**

RECOMBINATION

Drive of polymorphisms:

- Single nucleotide mutation
- Sequence re-arrangement





Human Evolutionary Genetics, Jobling, 2004

ARTICLE

OPEN

doi:10.1038/nature15393

A global reference for human genetic variation

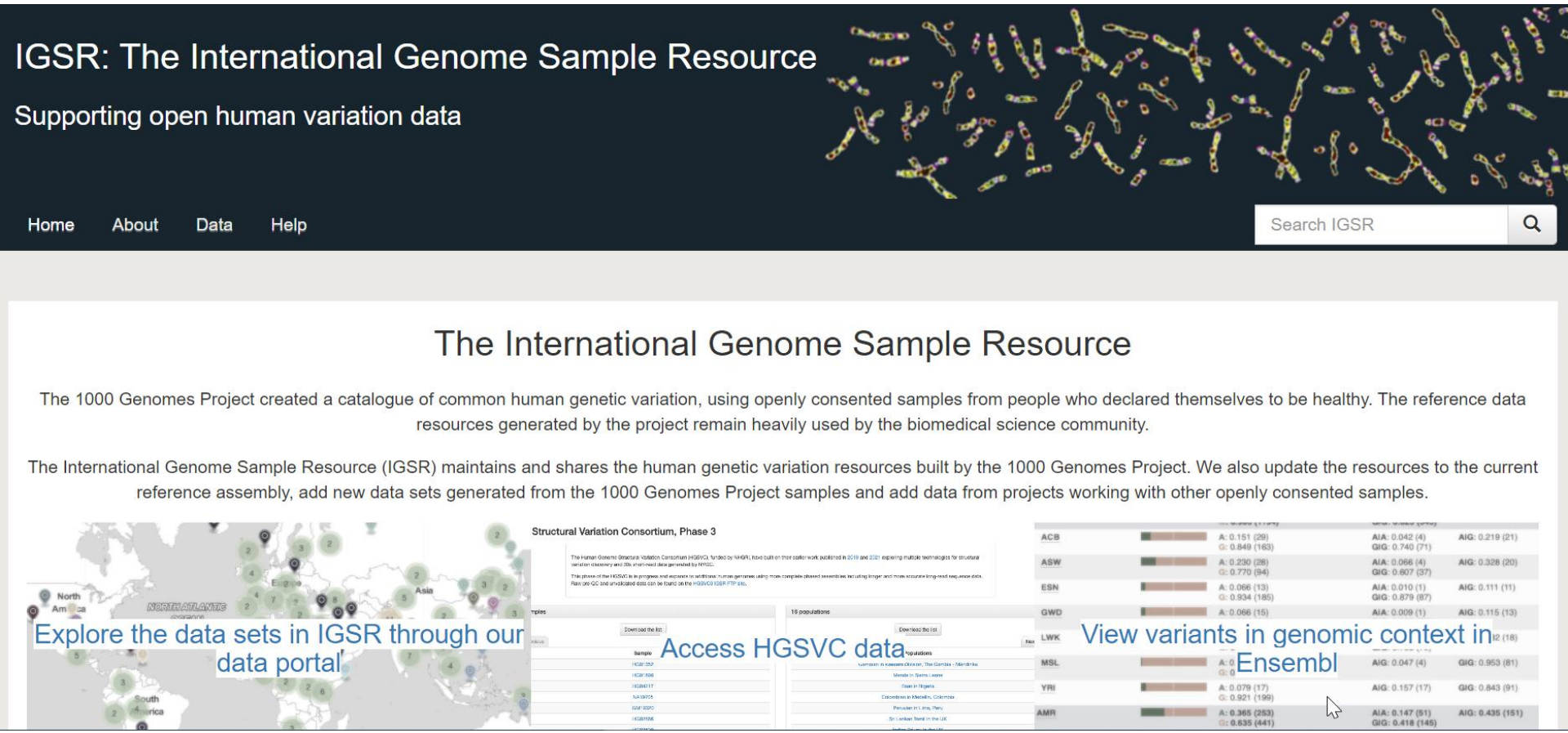
The 1000 Genomes Project Consortium*

The 1000 Genomes Project set out to provide a comprehensive description of common human genetic variation by applying whole-genome sequencing to a diverse set of individuals from multiple populations. Here we report completion of the project, having reconstructed the genomes of 2,504 individuals from 26 populations using a combination of low-coverage whole-genome sequencing, deep exome sequencing, and dense microarray genotyping. We characterized a broad spectrum of genetic variation, in total over 88 million variants (84.7 million single nucleotide polymorphisms (SNPs), 3.6 million short insertions/deletions (indels), and 60,000 structural variants), all phased onto high-quality haplotypes. This resource includes >99% of SNP variants with a frequency of >1% for a variety of ancestries. We describe the distribution of genetic variation across the global sample, and discuss the implications for common disease studies.

An integrated map of structural variation in 2,504 human genomes

A list of authors and their affiliations appears at the end of the paper.

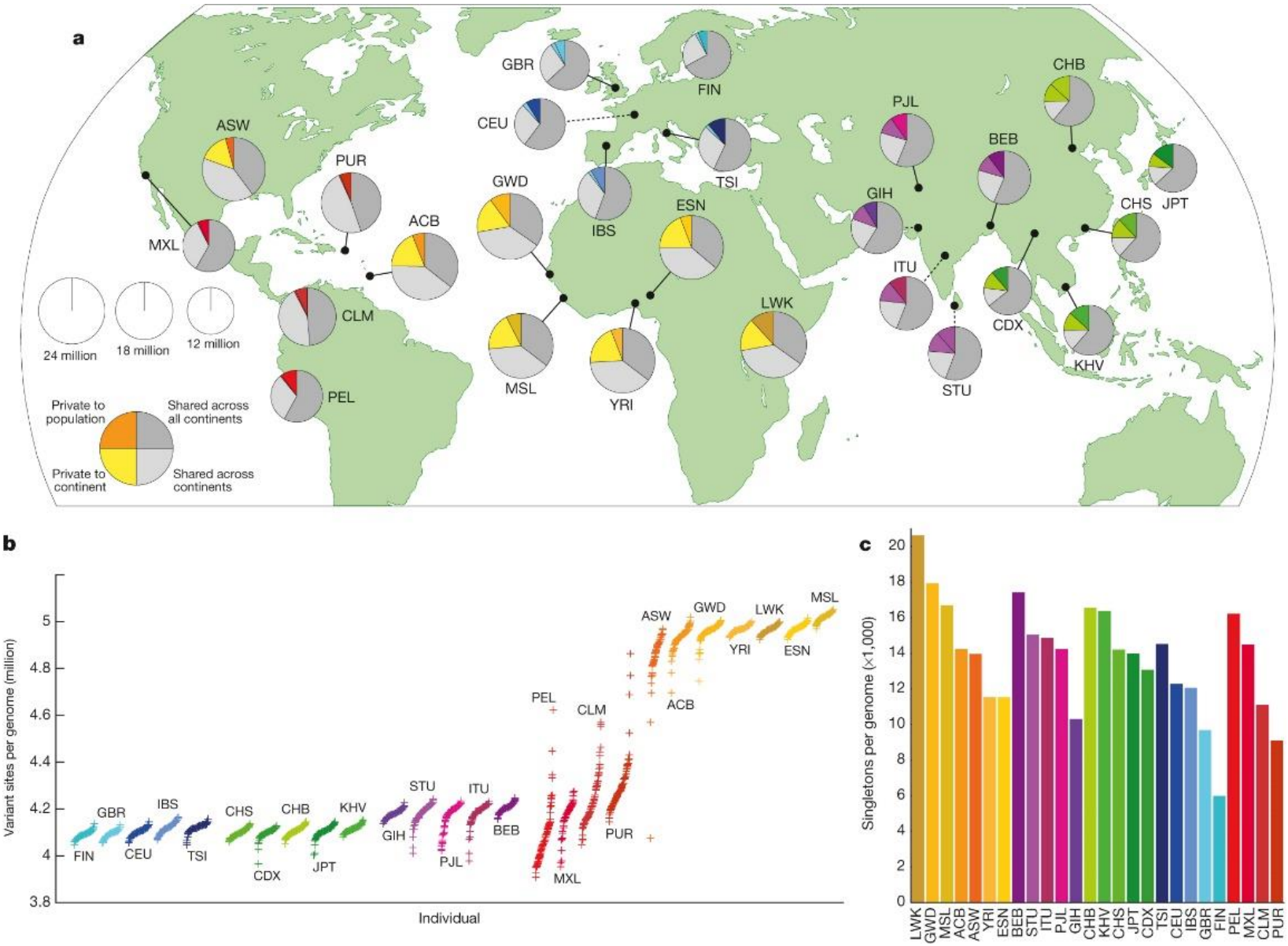
Structural variants are implicated in numerous diseases and make up the majority of varying nucleotides among human genomes. Here we describe an integrated set of eight structural variant classes comprising both balanced and unbalanced variants, which we constructed using short-read DNA sequencing data and statistically phased onto haplotype blocks in 26 human populations. Analysing this set, we identify numerous gene-intersecting structural variants exhibiting population stratification and describe naturally occurring homozygous gene knockouts that suggest the dispensability of a variety of human genes. We demonstrate that structural variants are enriched on haplotypes identified by genome-wide association studies and exhibit enrichment for expression quantitative trait loci. Additionally, we uncover appreciable levels of structural variant complexity at different scales, including genic loci subject to clusters of repeated rearrangement and complex structural variants with multiple breakpoints likely to have formed through individual mutational events. Our catalogue will enhance future studies into structural variant demography, functional impact and disease association.



https://www.internationalgenome.org/

nature

Population sampling



nature

ARTICLE

OPEN

doi:10.1038/nature15393

A global reference for human genetic variation

The 1000 Genomes Project Consortium*

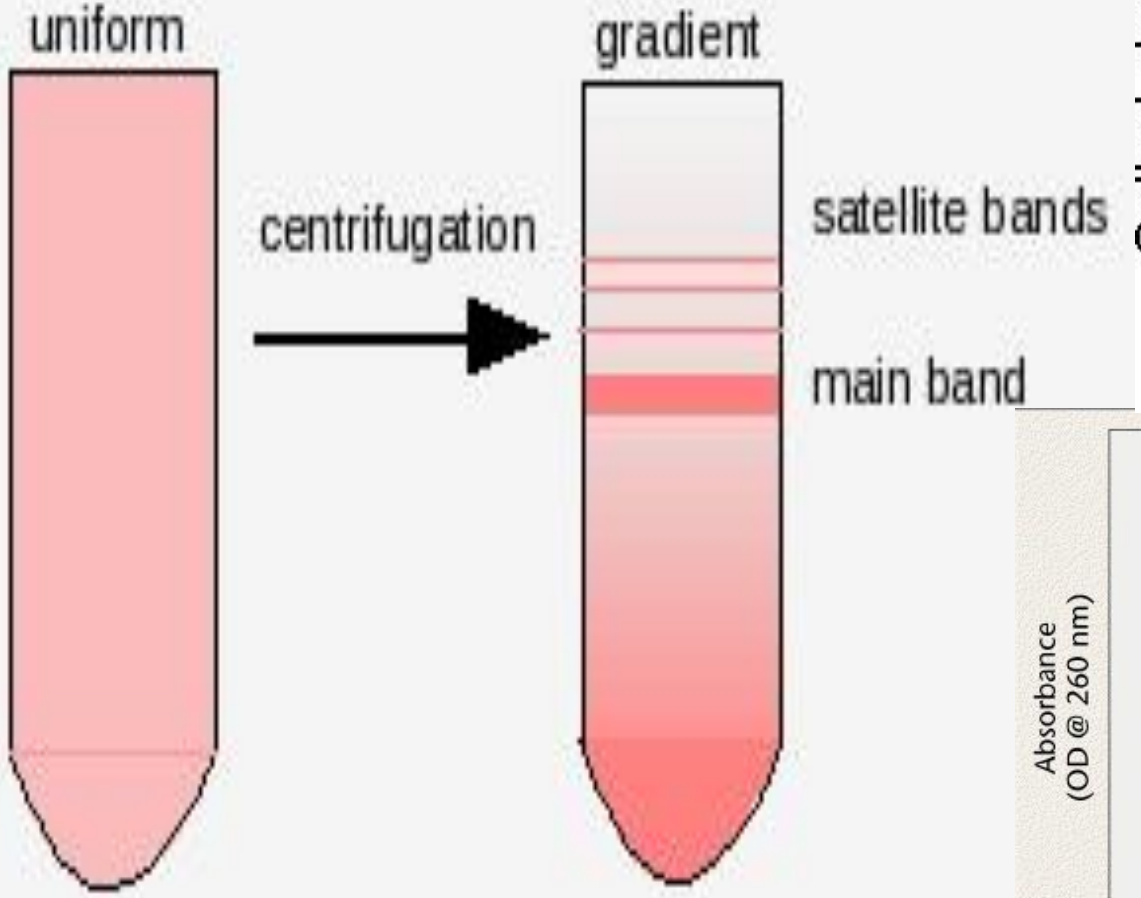
Table 1 | Median autosomal variant sites per genome

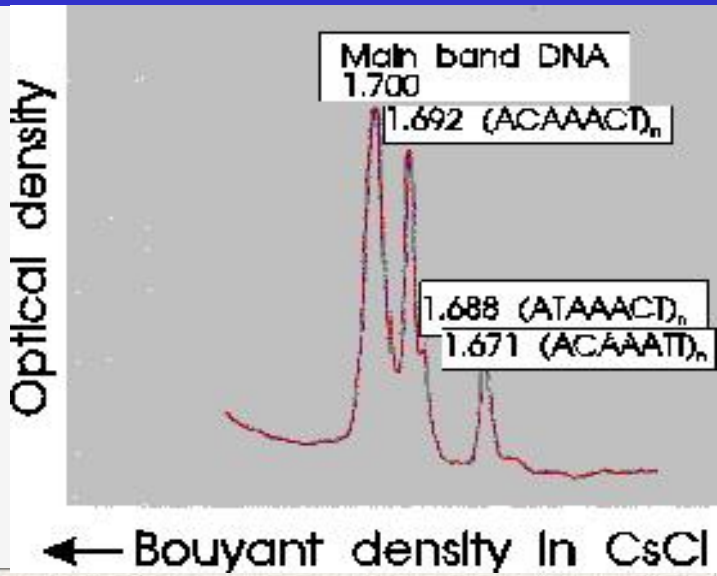
	AFR		AMR		EAS		EUR		SAS	
Samples	661		347		504		503		489	
Mean coverage	8.2		7.6		7.7		7.4		8.0	
	Var. sites	Singletons	Var. sites	Singletons	Var. sites	Singletons	Var. sites	Singletons	Var. sites	Singletons
SNPs	4.31M	14.5k	3.64M	12.0k	3.55M	14.8k	3.53M	11.4k	3.60M	14.4k
Indels	625k	-	557k	-	546k	-	546k	-	556k	-
Large deletions	1.1k	5	949	5	940	7	939	5	947	5
CNVs	170	1	153	1	158	1	157	1	165	1
MEI (Alu)	1.03k	0	845	0	899	1	919	0	889	0
MEI (L1)	138	0	118	0	130	0	123	0	123	0
MEI (SVA)	52	0	44	0	56	0	53	0	44	0
MEI (MT)	5	0	5	0	4	0	4	0	4	0
Inversions	12	0	9	0	10	0	9	0	11	0
Nonsynon	12.2k	139	10.4k	121	10.2k	144	10.2k	116	10.3k	144
Synon	13.8k	78	11.4k	67	11.2k	79	11.2k	59	11.4k	78
Intron	2.06M	7.33k	1.72M	6.12k	1.68M	7.39k	1.68M	5.68k	1.72M	7.20k
UTR	37.2k	168	30.8k	136	30.0k	169	30.0k	129	30.7k	168
Promoter	102k	430	84.3k	332	81.6k	425	82.2k	336	84.0k	430
Insulator	70.9k	248	59.0k	199	57.7k	252	57.7k	189	59.1k	243
Enhancer	354k	1.32k	295k	1.05k	289k	1.34k	288k	1.02k	295k	1.31k
TFBSs	927	4	759	3	748	4	749	3	765	3
Filtered LoF	182	4	152	3	153	4	149	3	151	3
HGMD-DM	20	0	18	0	16	1	18	2	16	0
GWAS	2.00k	0	2.07k	0	1.99k	0	2.08k	0	2.06k	0
ClinVar	28	0	30	1	24	0	29	1	27	1

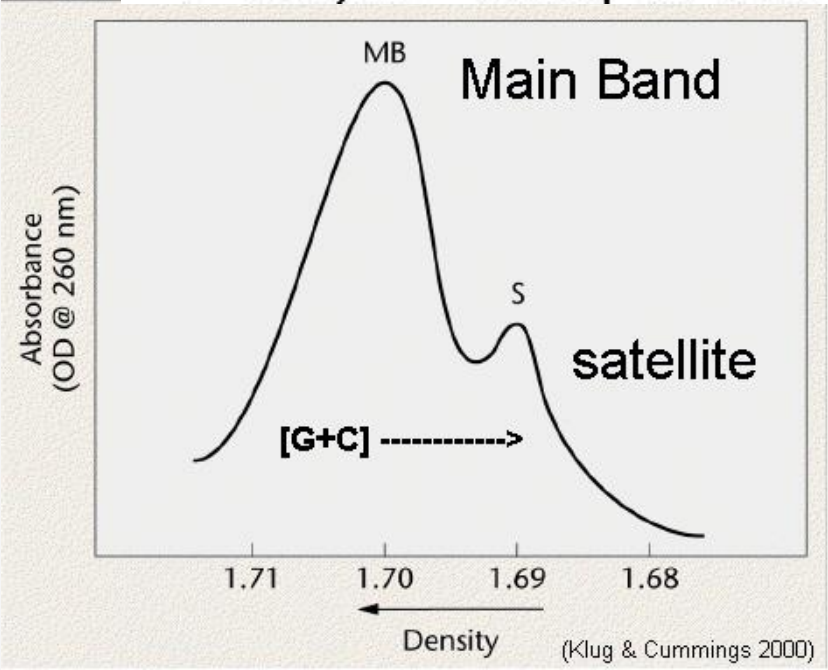
See Supplementary Table 1 for continental population groupings. CNVs, copy-number variants; HGMD-DM, Human Gene Mutation Database disease mutations; k, thousand; LoF, loss-of-function; M, million; MEI, mobile element insertions.

- a typical genome differs from the reference human genome at 4.1 million to 5.0 million sites.
- >99.9% of variants consist of SNPs and short indels.
- structural variants affect more bases:
- typical genome contains an estimated 2,100 to 2,500 structural variants (1,000 large deletions, 160 copy-number variants, 915 Alu insertions, 128 L1 insertions, 51 SVA insertions, 4 NUMTs and 10 inversions) affecting 20 million bases of sequence.

Satellite DNA







Restriction Fragment Length Polymorphism (RFLP) - „DNA fingerprinting"

Double-stranded DNA

Restriction enzymes

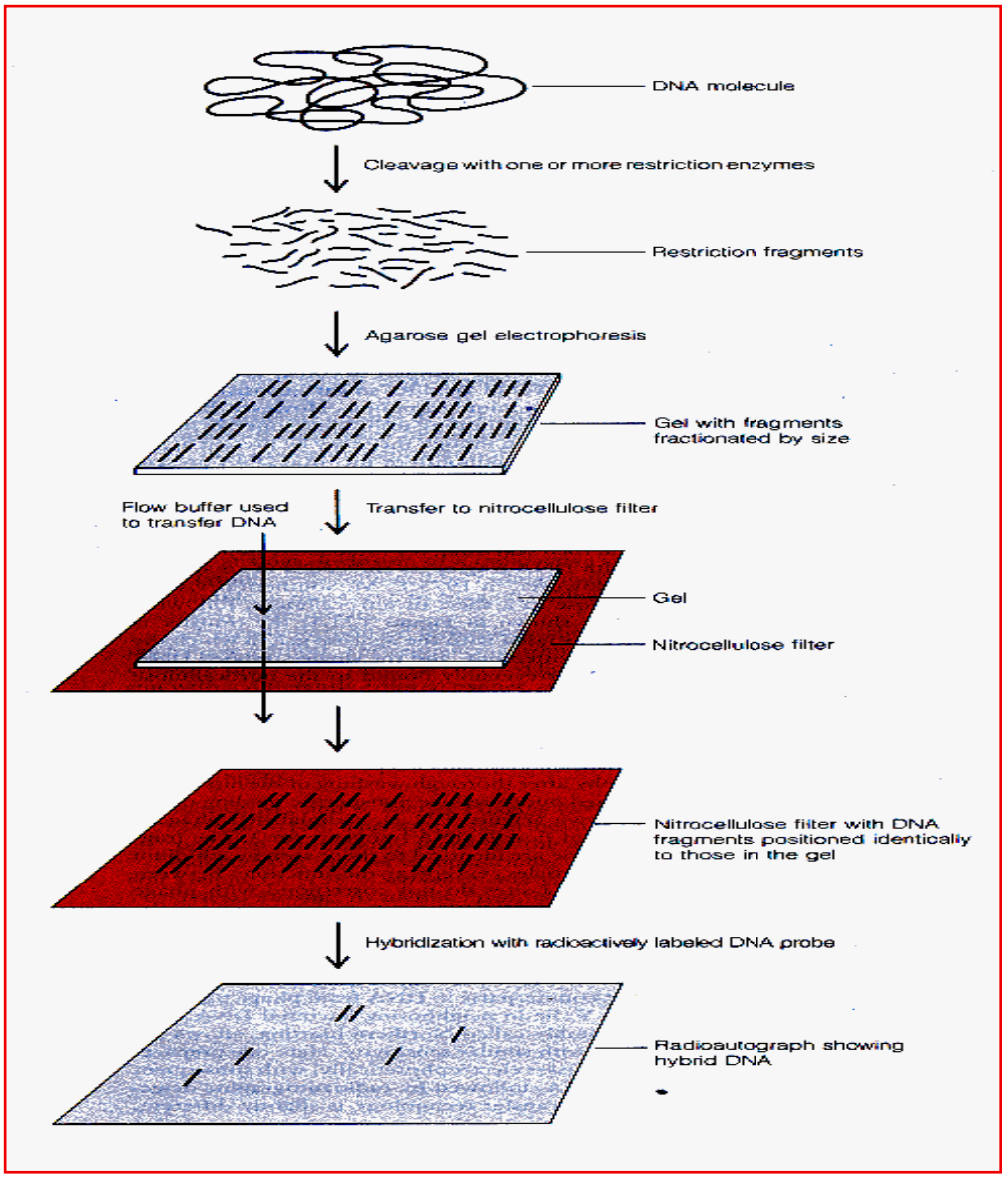
Gel electrophoresis

Southern-blot

Probe hibridization

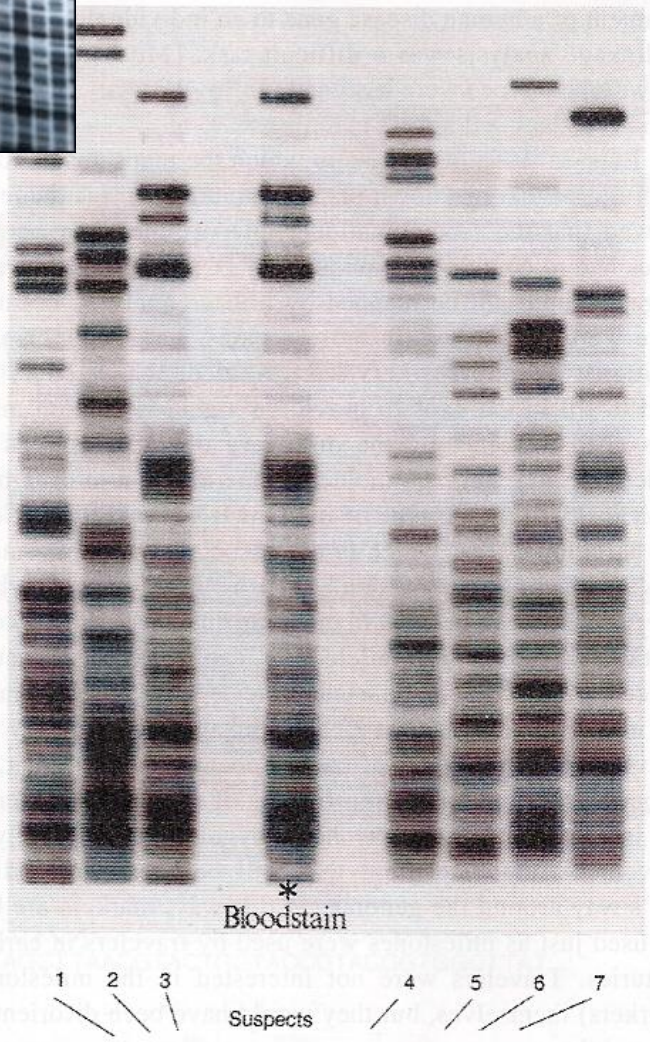
Autoradiogram

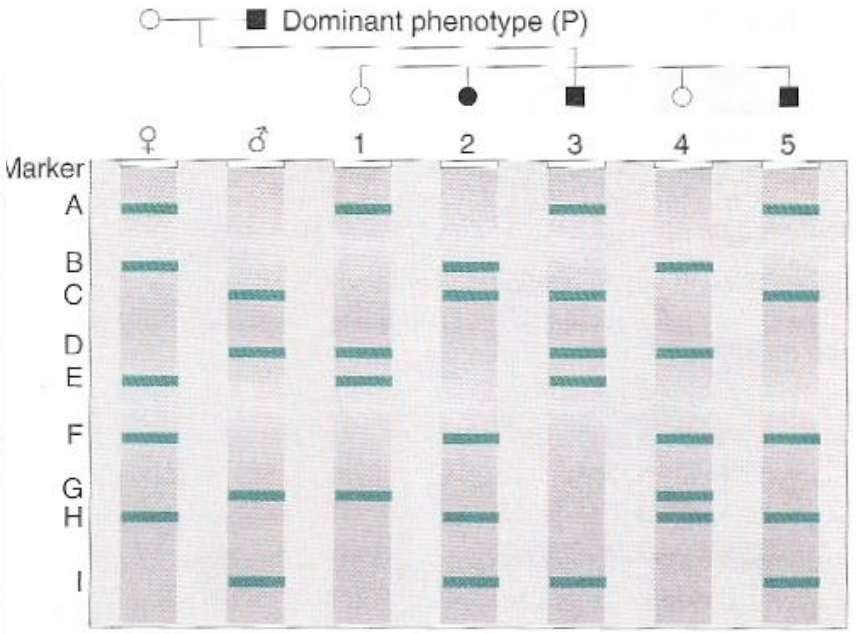
- ***MLP-RFLP***
- ***SLP-RFLP***



VNTR assay markers: RFLP analytics

1985 – Sir Alec Jeffreys





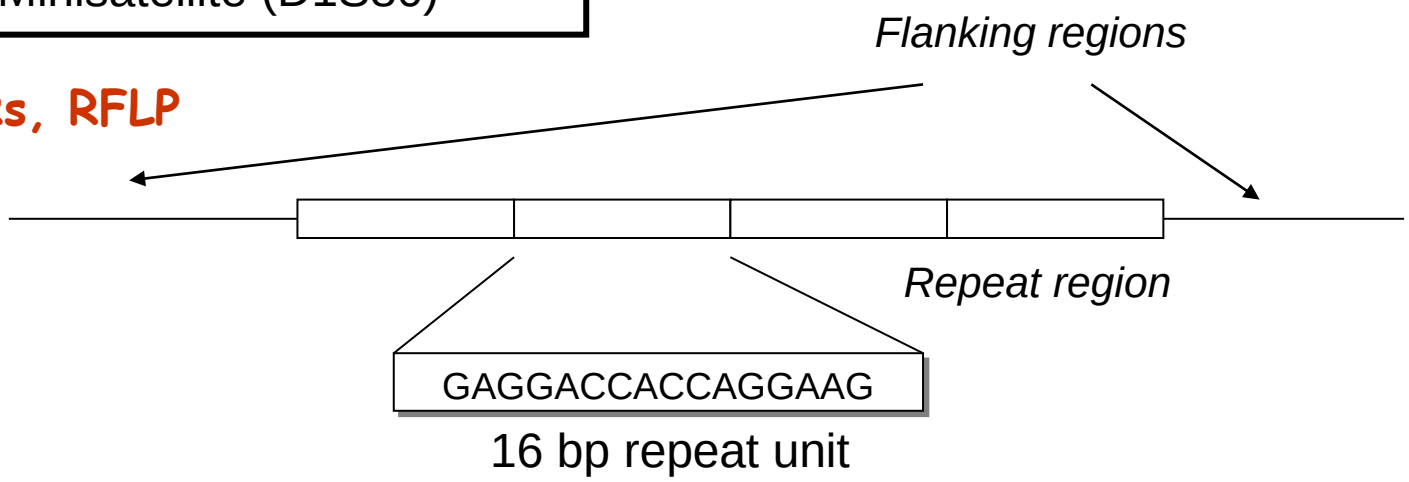
ANALYSIS EXAMPLES

F and H Always inherited together — linked?
A and B In progeny, always *either* A *or* B — "allelic"?
A and D Four combinations; A and D, A, D, or neither — unlinked?
F, H, and E Always *either* F and H *or* E — closely linked in trans?
Allele *P* Possibly linked to I and C.

Genetic mapping

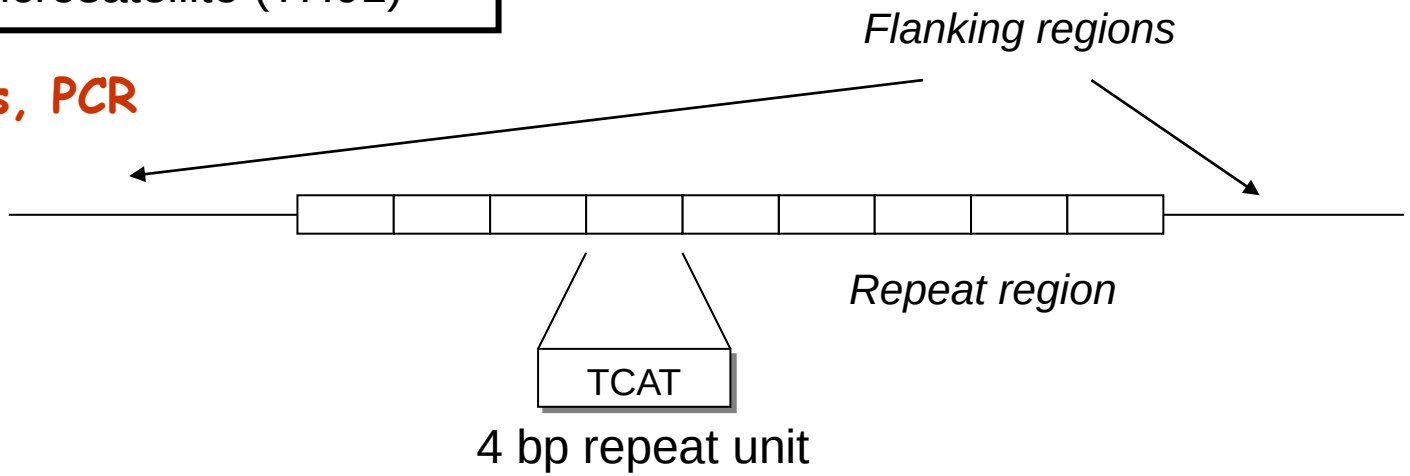
Minisatellite (D1S80)

VNTRs, RFLP

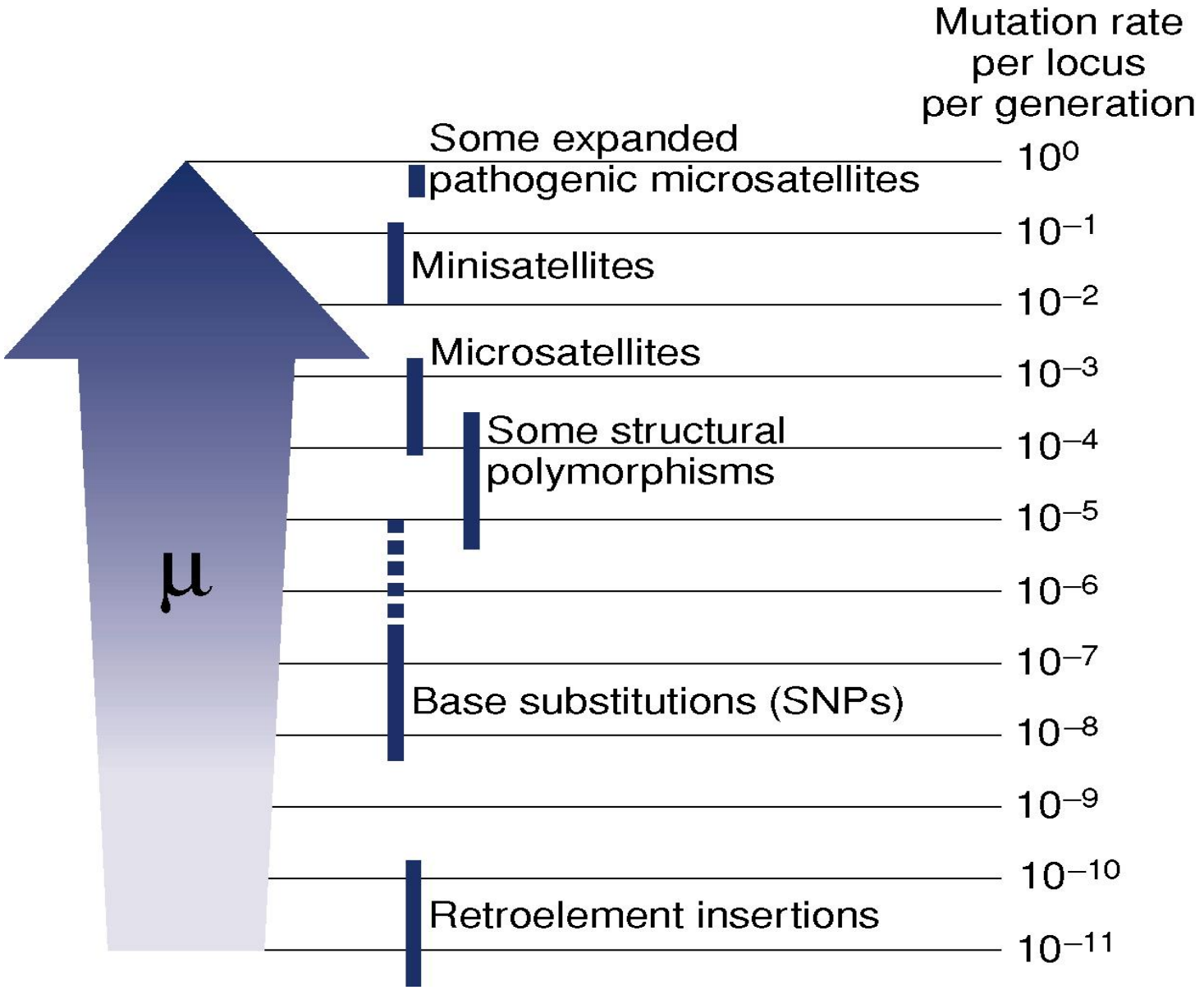


Microsatellite (TH01)

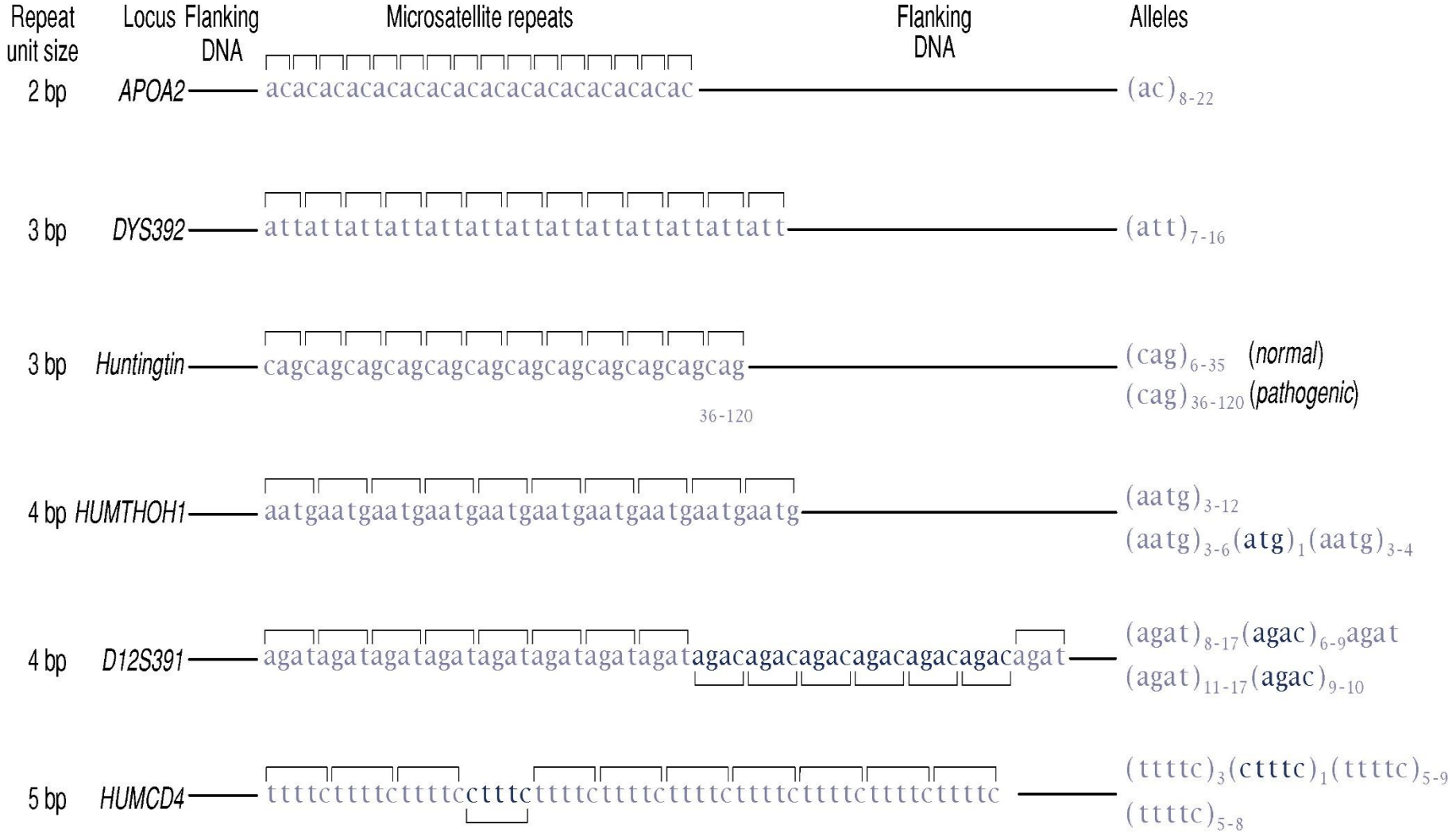
STRs, PCR



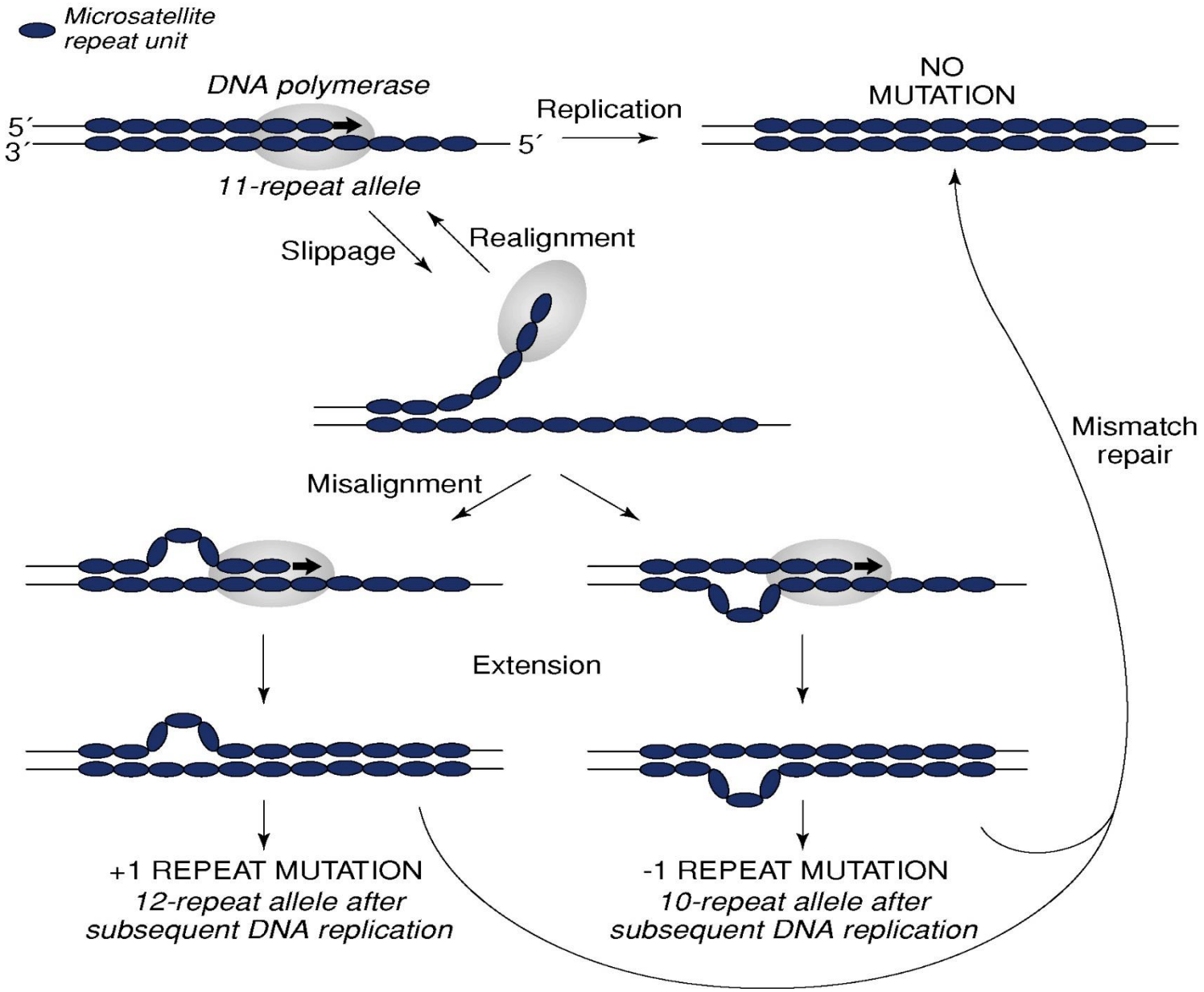
Mutation rate of polymorphic sequences (μ)



Microsatellite structure

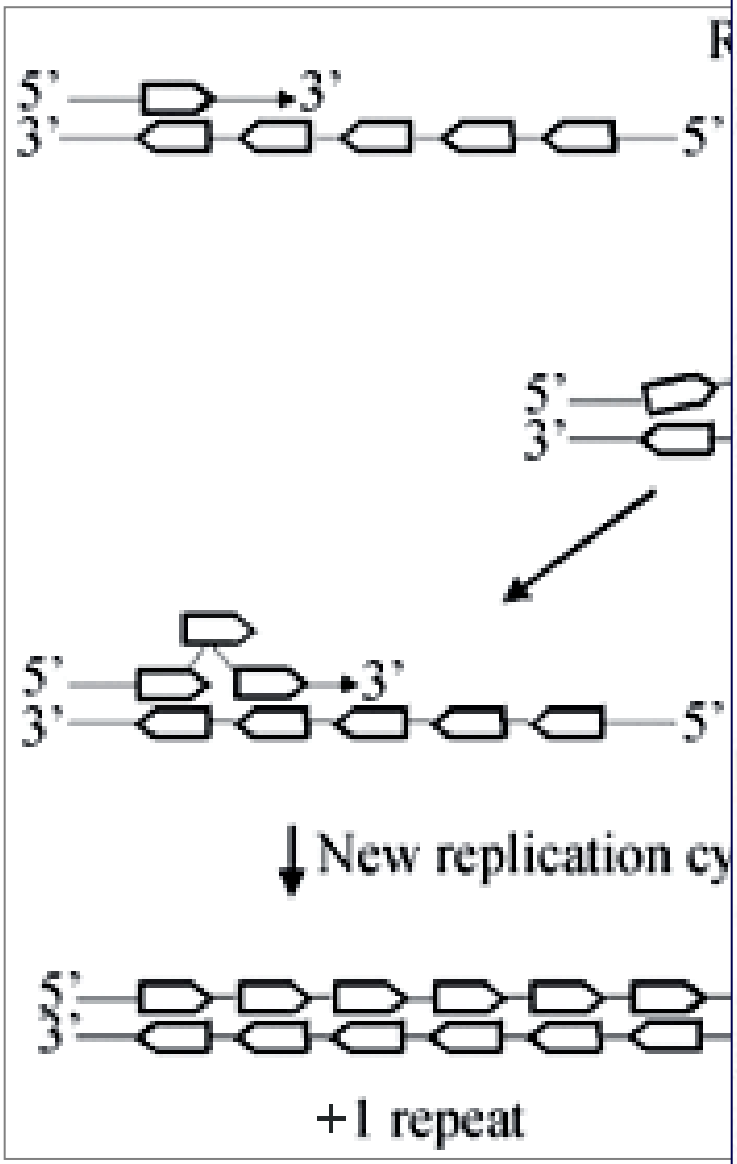


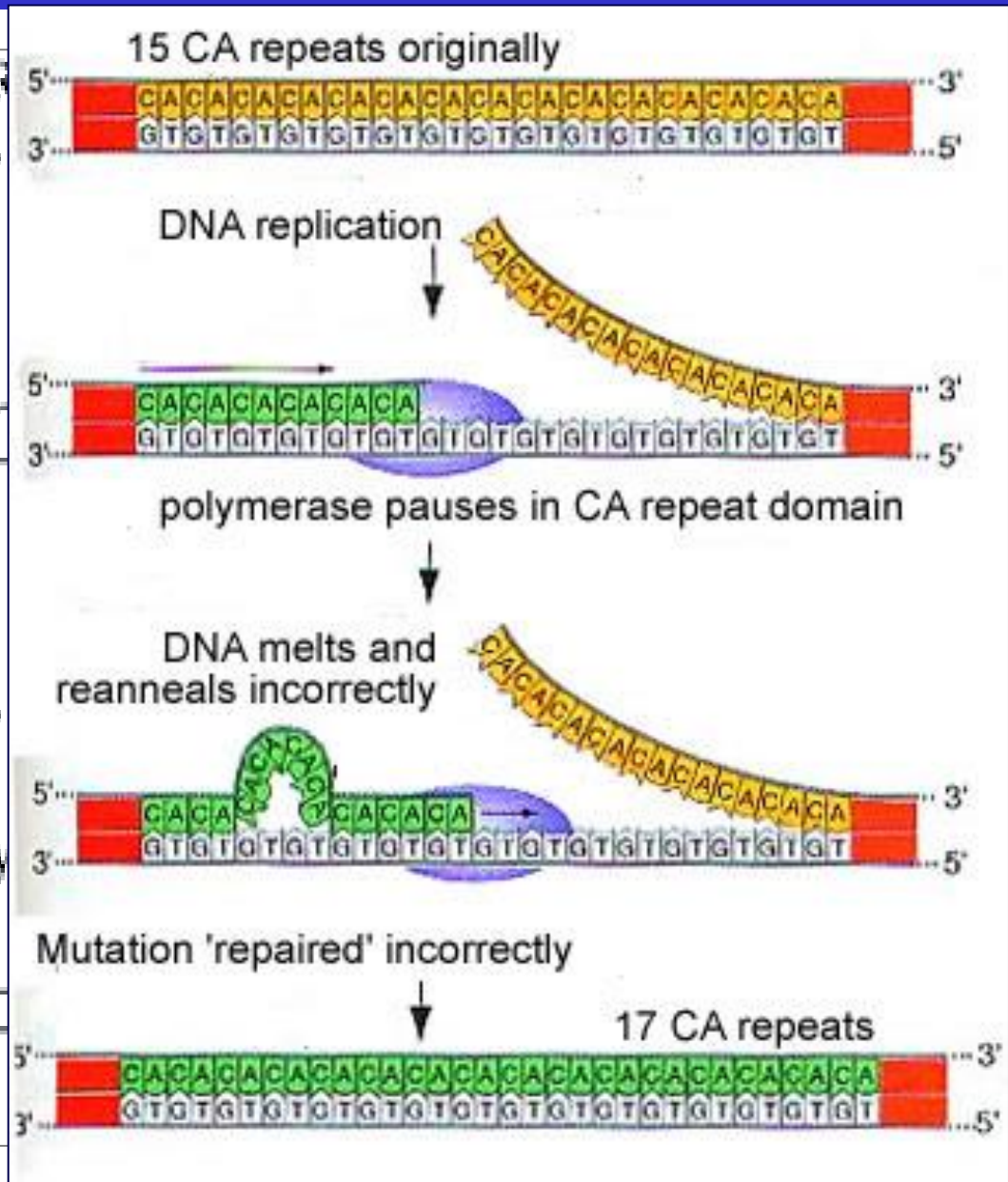
„Replication slippage" – Microsatellite mutation



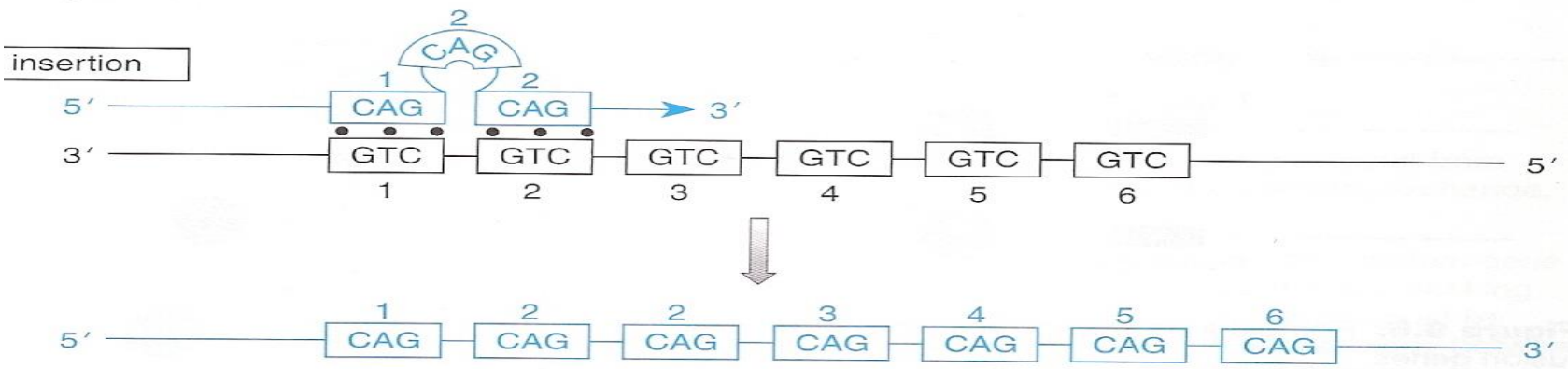
Human Evolutionary Genetics, Jobling, 2004

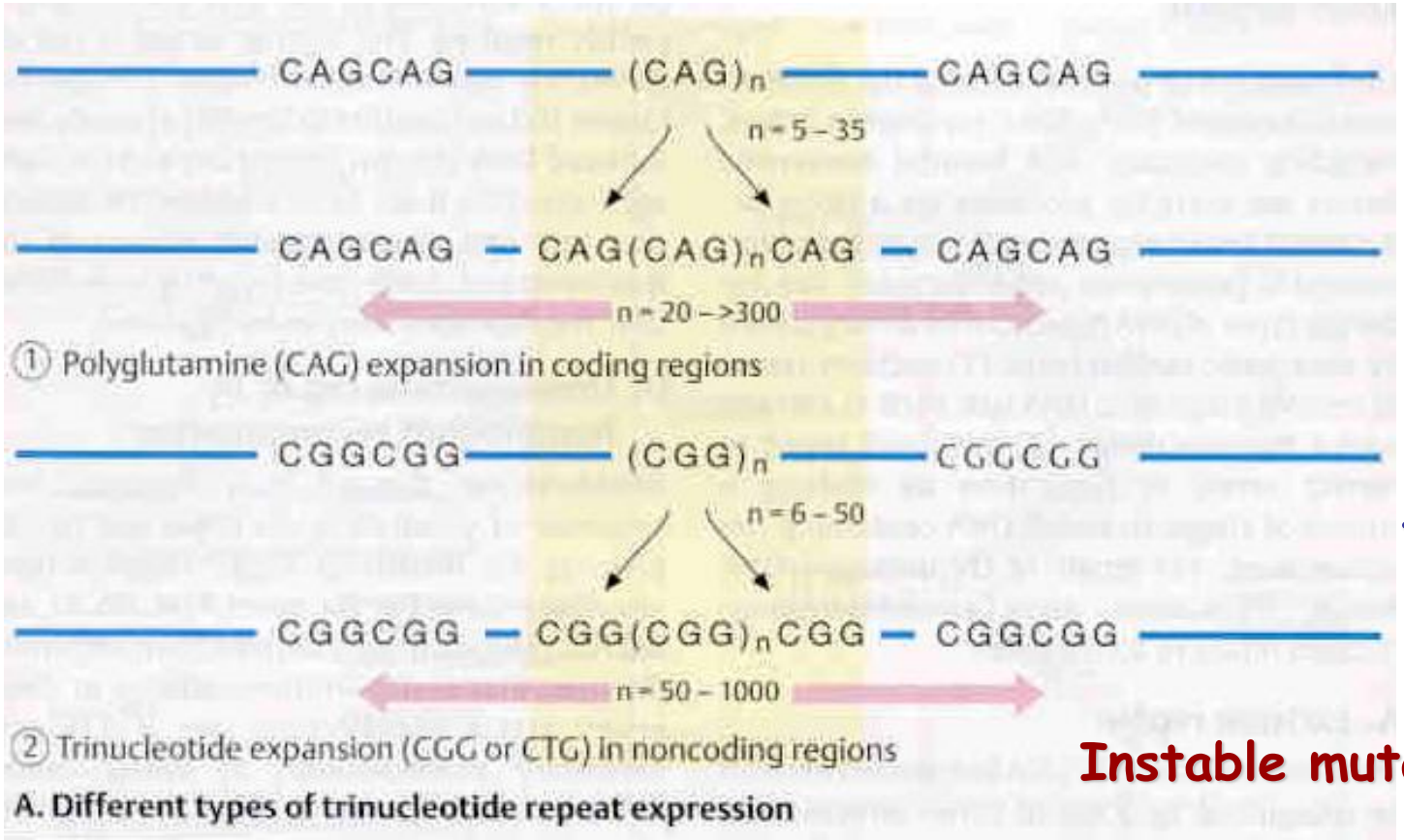
Microsatellite evolution





Trinucleotide repeat expansion



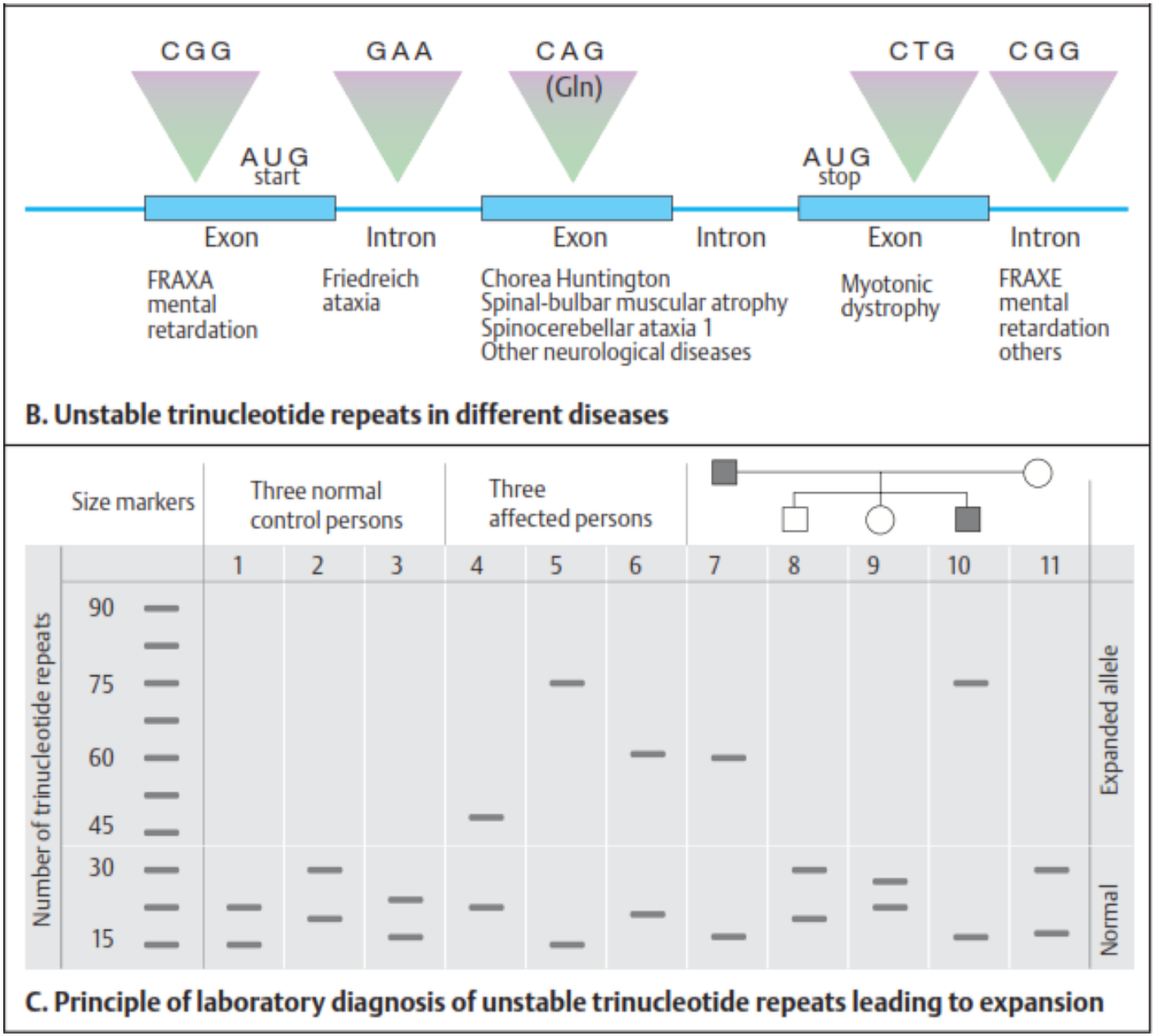


Type I

Type II

Instable mutation

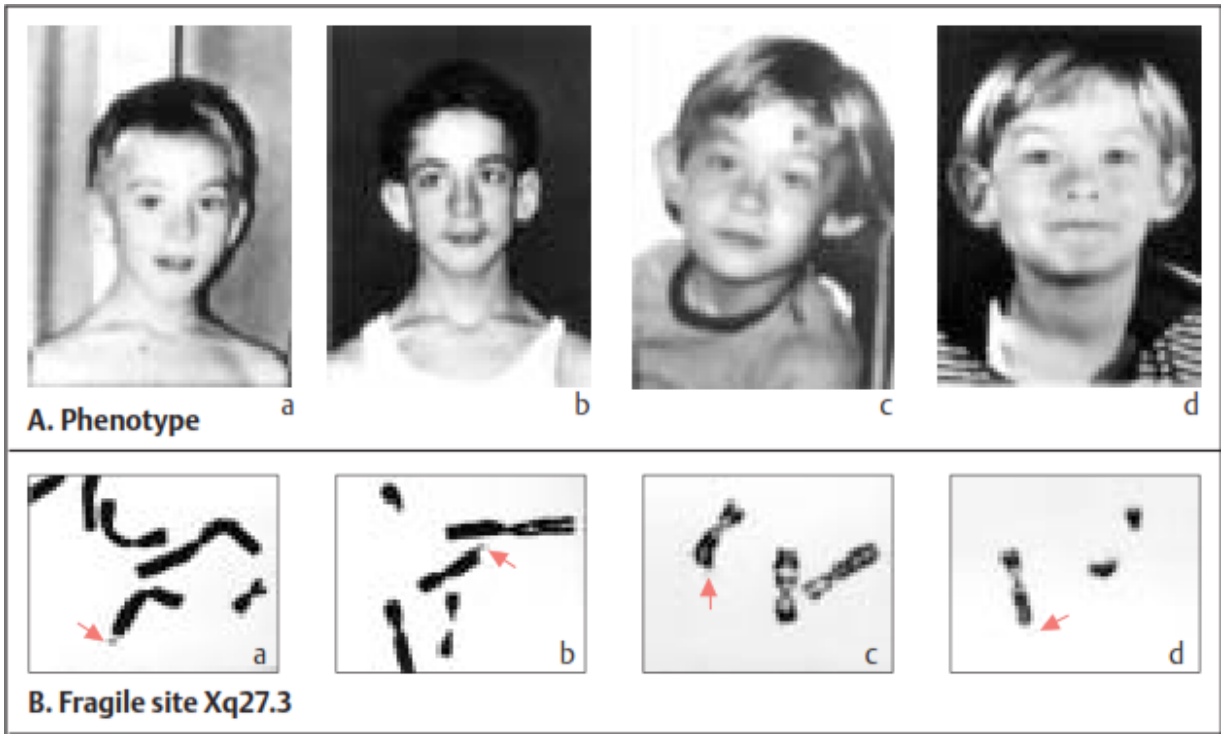
Trinucleotide repeat expansion



Passarge, 2001

Genetic diseases due to repeat expansion

Disease (Examples)	Gene	Frequency	Tri-nucleotide	Normal Number	Mutant Allele	Chromosome
Huntington disease	*HD*	1:10 000	$(CAG)_n$	0–26	36–121	4p16.3
Fragile X syndrome	*FMR1*	1:5 000	$(CGG)_n$	6–50	52–500	Xq27.3
Myotonic dystrophy	*DMPK*	1:8 000	$(CTG)_n$	5–37	50–500	19q13.2
Spinal-bulbar muscular atrophy (Kennedy)	*SBMA*	<1:50 000	$(CAG)_n$	11–31	36–65	Xq11-12



Fragile X

Huntington disease

Myotonic dystrophy

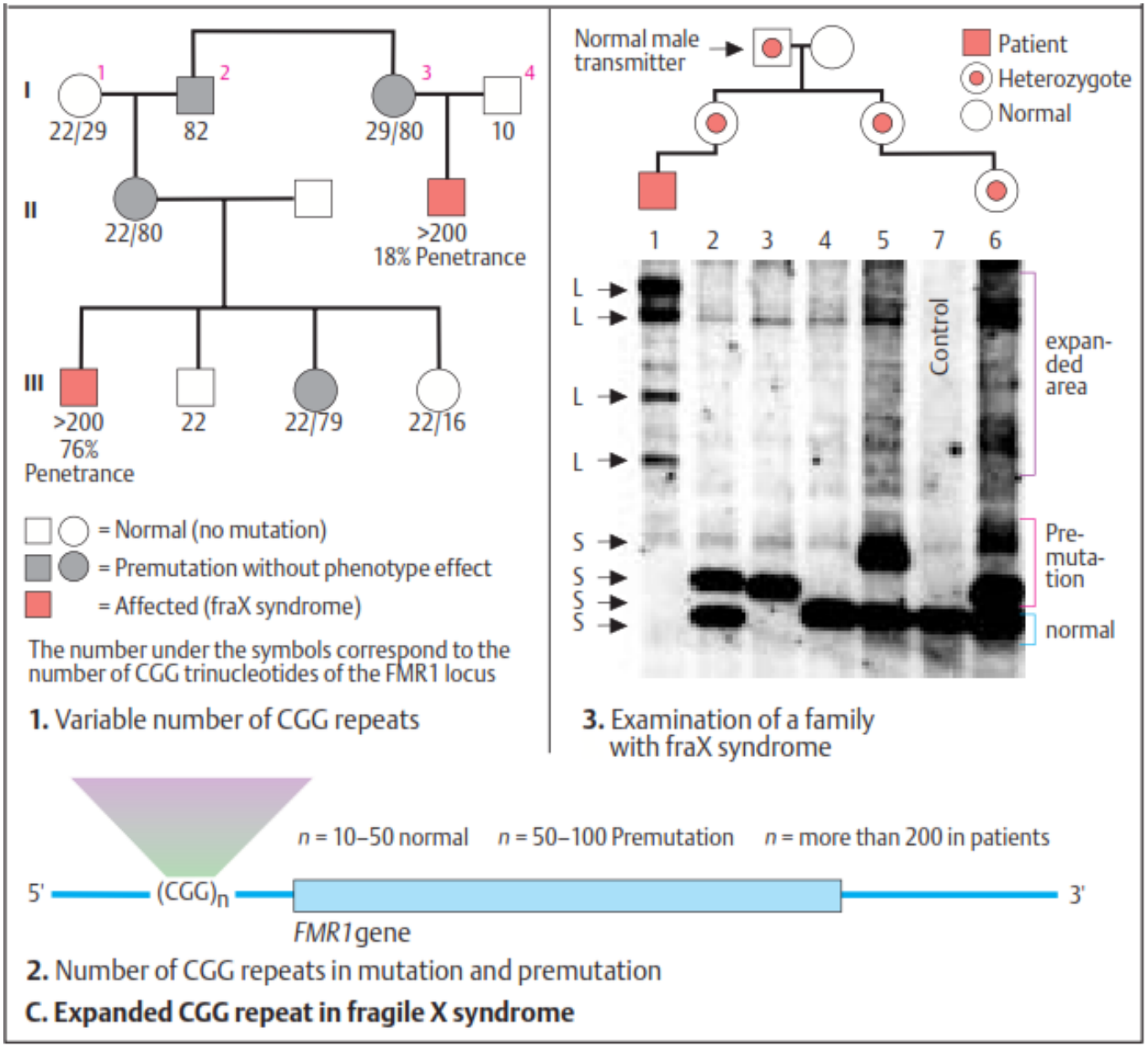
Friedrich ataxia

SMA

etc.

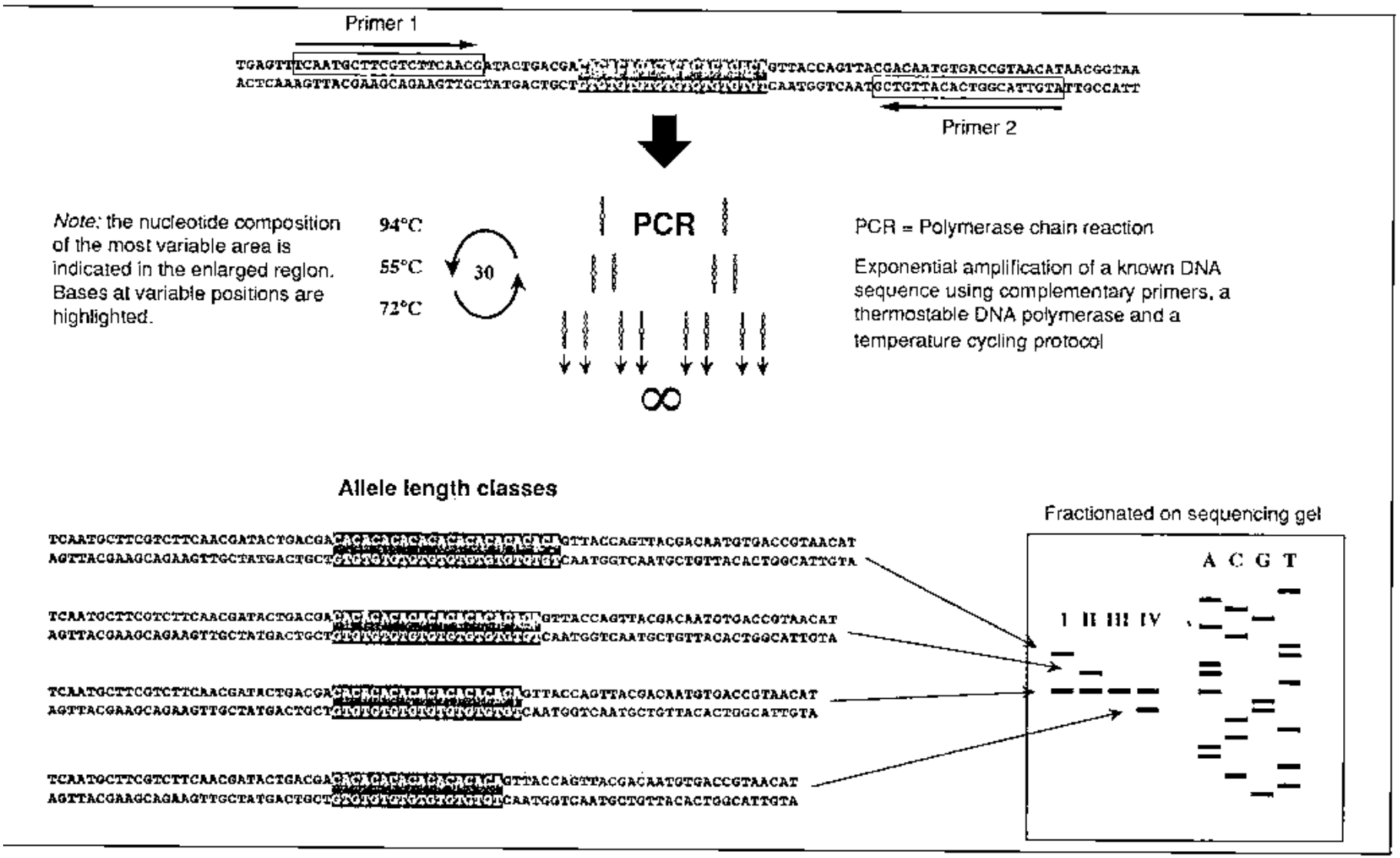
Passarge, 2001

Diagnostics of expanded *CGG* repeats in Fragile X



Passarge, 2001

Genotiping microsatellites by PCR



Microsatellite allele genotyping: multiallelic

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

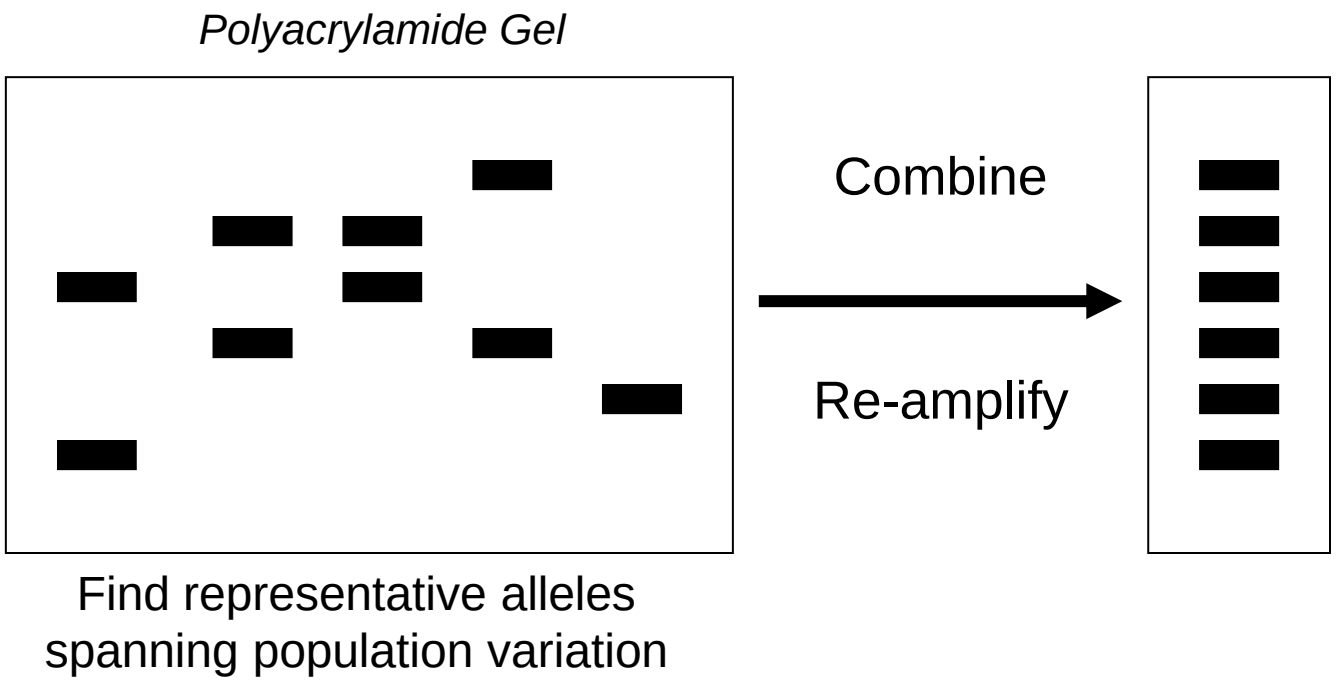
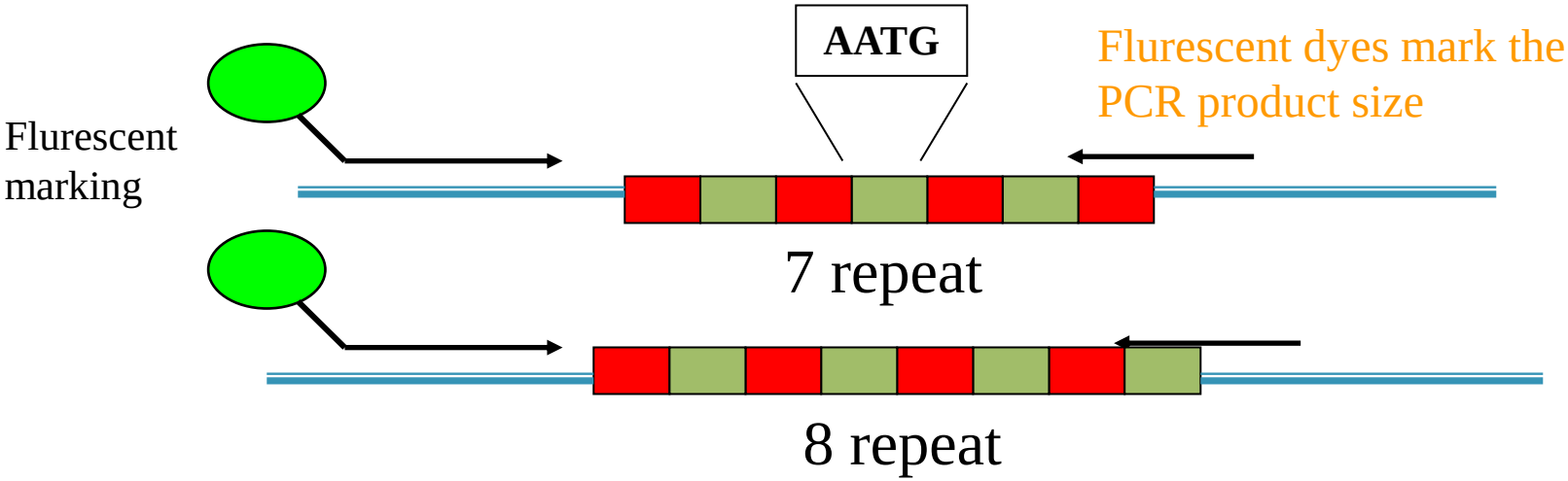


Figure 5.3, J.M. Butler (2005) *Forensic DNA Typing*, 2nd Edition © 2005 Elsevier Academic Press

Microsatellite – STR – markers (Short Tandem Repeat)



Repeat region varies in length from alleles to alleles but flanking region where PCR primers bind is constant

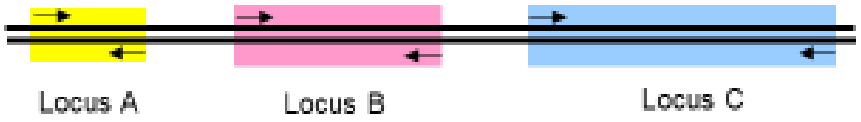
Homozygous = two homologous are the same

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



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

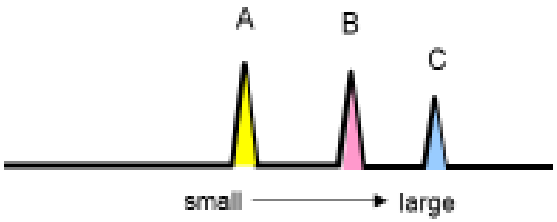
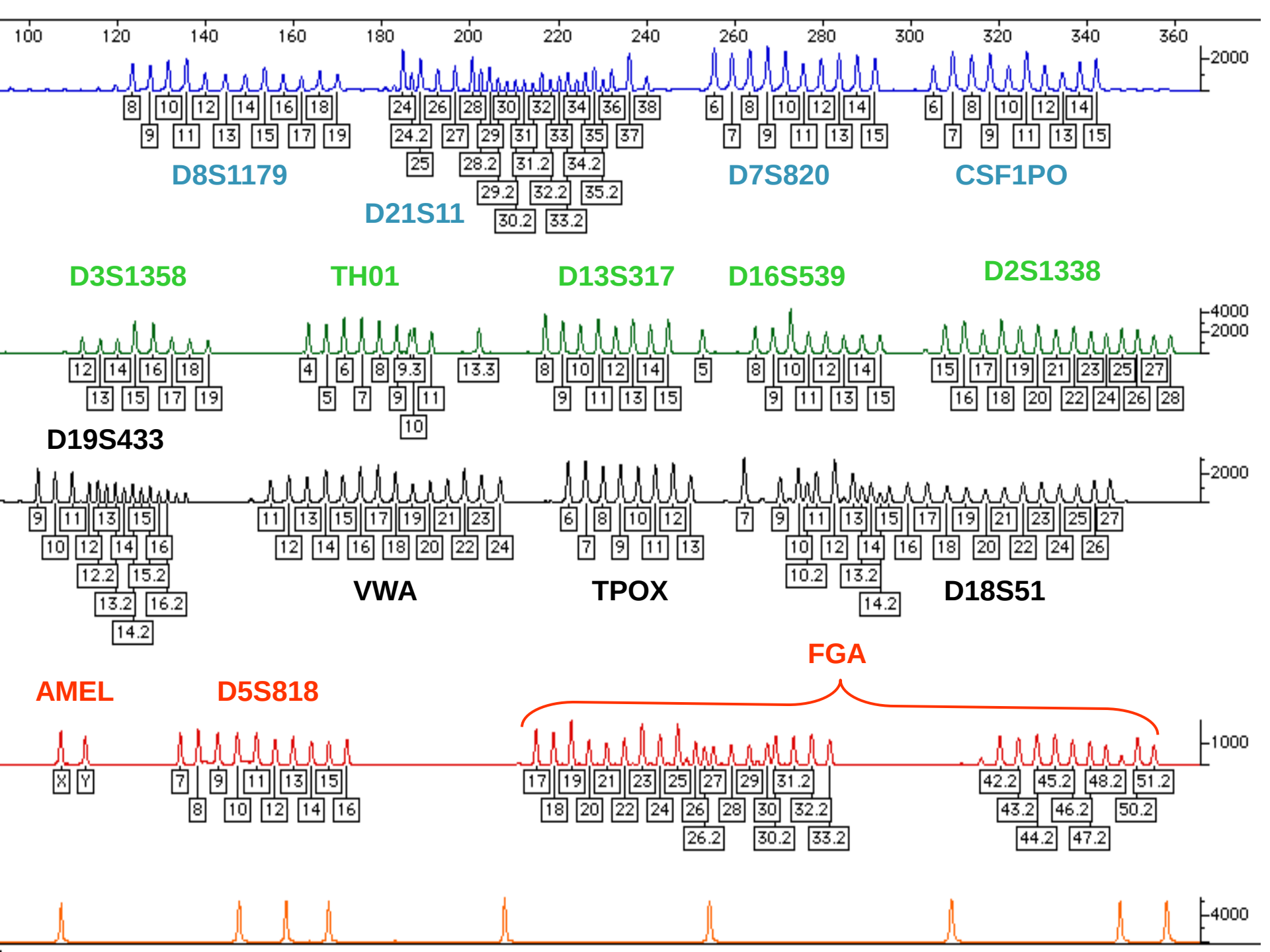


Figure 4.3, J.M. Butler (2005) *Forensic DNA Typing*, 2nd Edition © 2005 Elsevier Academic Press

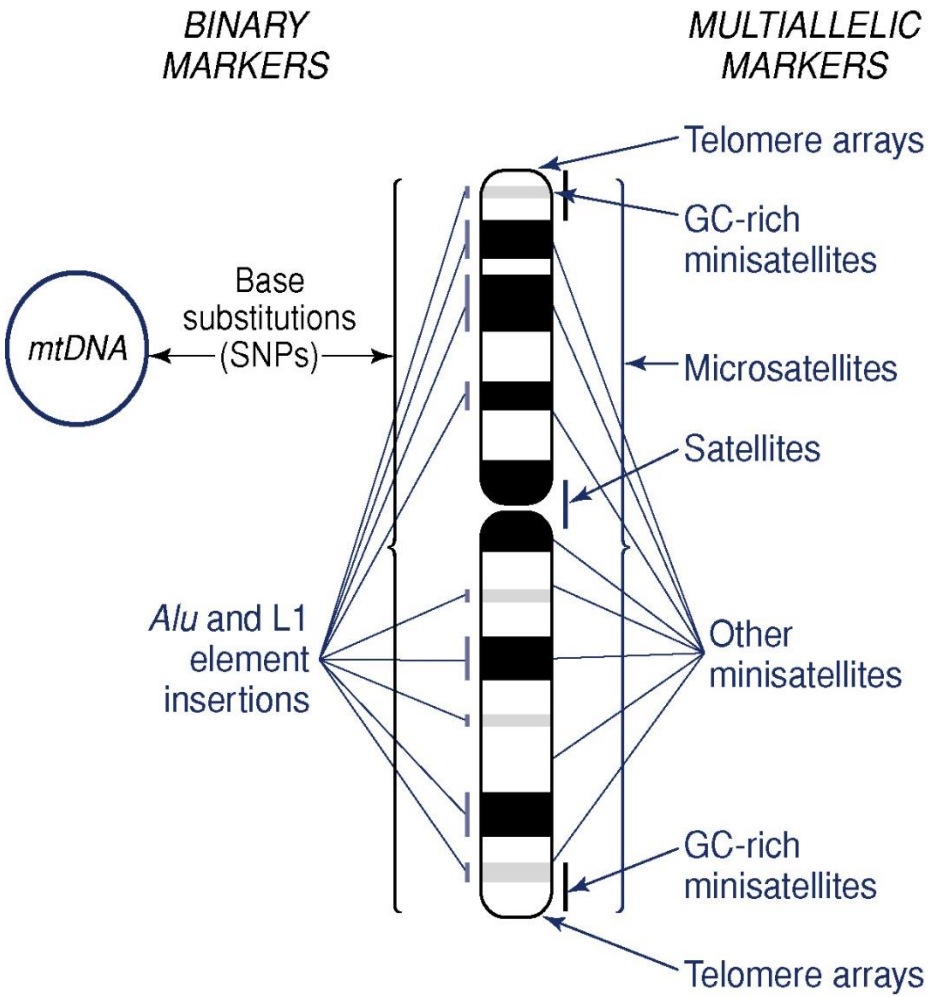

100 120 140 160 180 200 220 240 260 280 300 320 340 360
2000
8 10 12 14 16 18
9 11 13 15 17 19
D8S1179
24 26 28 30 32 34 36 38
24.2 27 29 31 33 35 37
25 28.2 31.2 34.2
29.2 32.2 35.2
30.2 33.2
D21S11
6 8 10 12 14
7 9 11 13 15
D7S820
6 8 10 12 14
7 9 11 13 15
CSF1PO
D3S1358
TH01
D13S317
D16S539
D2S1338
4000
2000
12 14 16 18
13 15 17 19
4 6 8 9.3 13.3
5 7 9 11
10
8 10 12 14
9 11 13 15
5
8 10 12 14
9 11 13 15
15 17 19 21 23 25 27
16 18 20 22 24 26 28
D19S433
2000
9 11 13 15
10 12 14 16
12.2 15.2
13.2 16.2
14.2
11 13 15 17 19 21 23
12 14 16 18 20 22 24
VWA
6 8 10 12
7 9 11 13
TPOX
7 9 11 13 15 17 19 21 23 25 27
10 12 14 16 18 20 22 24 26
10.2 13.2
14.2
D18S51
FGA
AMEL
D5S818
1000
X Y
7 9 11 13 15
8 10 12 14 16
17 19 21 23 25 27 29 31.2
18 20 22 24 26 28 30 32.2
26.2 30.2 33.2
42.2 45.2 48.2 51.2
43.2 46.2 50.2
44.2 47.2
4000

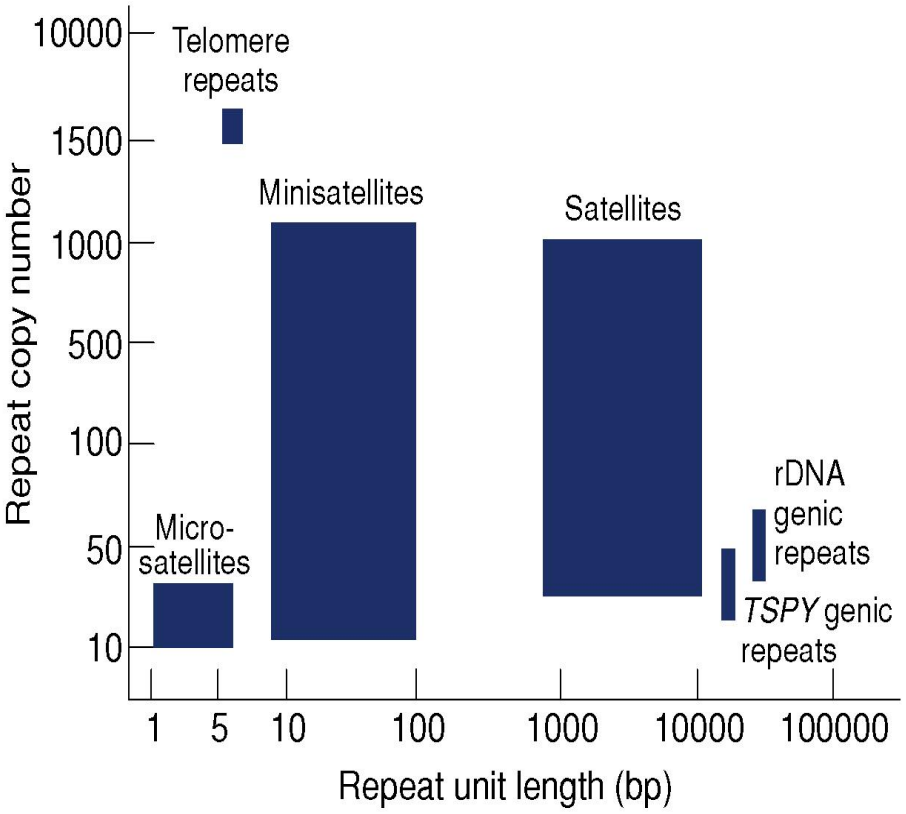
STR System	Maternal Meioses (%)	Paternal Meioses (%)	Number from either	Total Number of Mutations	Mutation Rate
CSF1PO	95/304,307 (0.03)	982/643,118 (0.15)	410	1,487/947,425	**0.16%**
FGA	205/408,230 (0.05)	2,210/692,776 (0.32)	710	3,125/1,101,006	**0.28%**
TH01	31/327,172 (0.009)	41/452,382 (0.009)	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%**

"Stepwise" mutation model

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

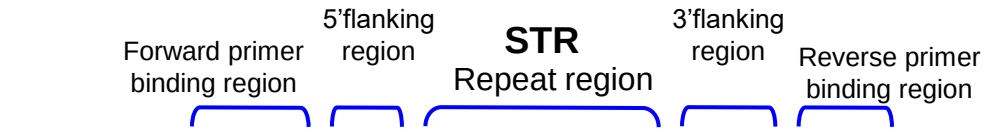
Distribution of polimorphic markers in the genome





Human Evolutionary Genetics, Jobling, 2004

Microsatellite point mutations



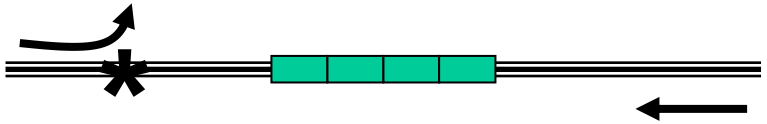
A)

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

B)

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

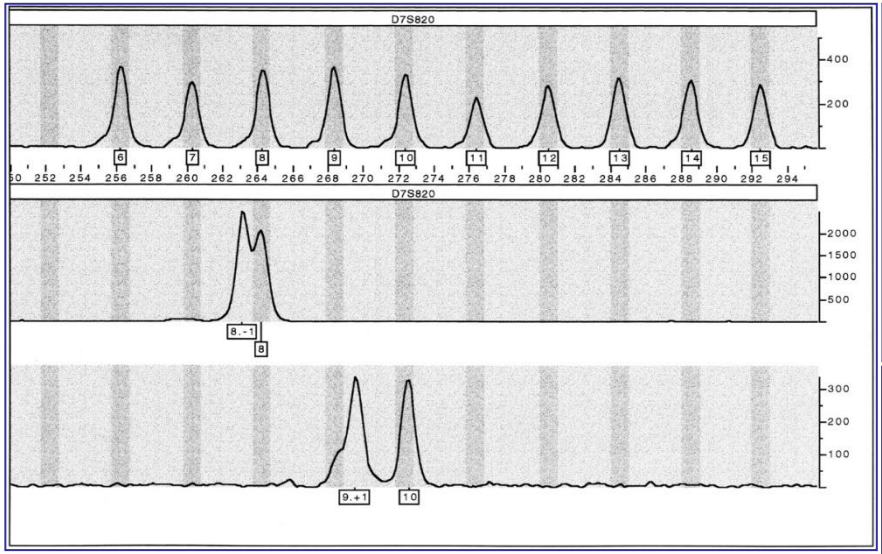
C)

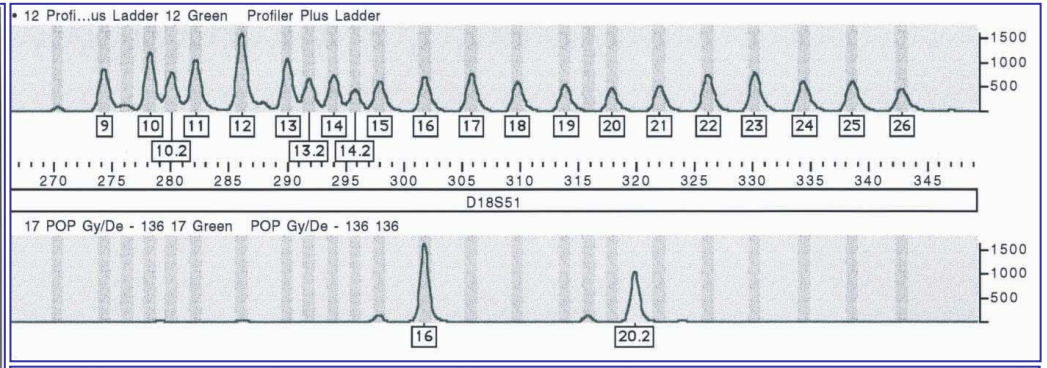


Example:
Rare VWA allele amplified with AmpFlSTR primers
(A-to-T in 2nd base from 3'end of forward primer)

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

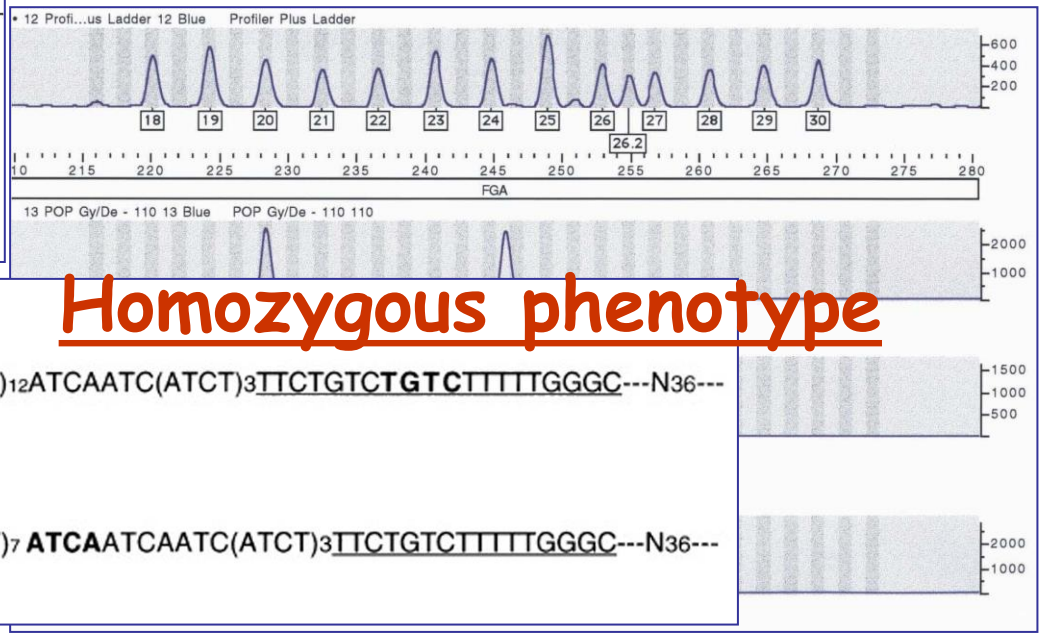
Variant microsatellite alleles: Null-alleles





Allél	Fragmens méret	5' Flanking régió	Repeat régió	3' Flanking régió
8.-1	205 bp	[redacted]	- $(GATA)_8$ -	[redacted] - $(T)_8$ ATCT -
9.+1	211 bp	[redacted]	- $(GATA)_9$ -	[redacted] - $(T)_{10}$ ATCT -
10	214 bp	[redacted]	- $(GATA)_{10}$ -	[redacted] - $(T)_8$ AATCT -
12 (ref.)	222 bp	[redacted]	- $(GATA)_{12}$ -	[redacted] - $(T)_9$ ATCT -
		24 bp		13 bp 124 bp

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



Homozygous phenotype

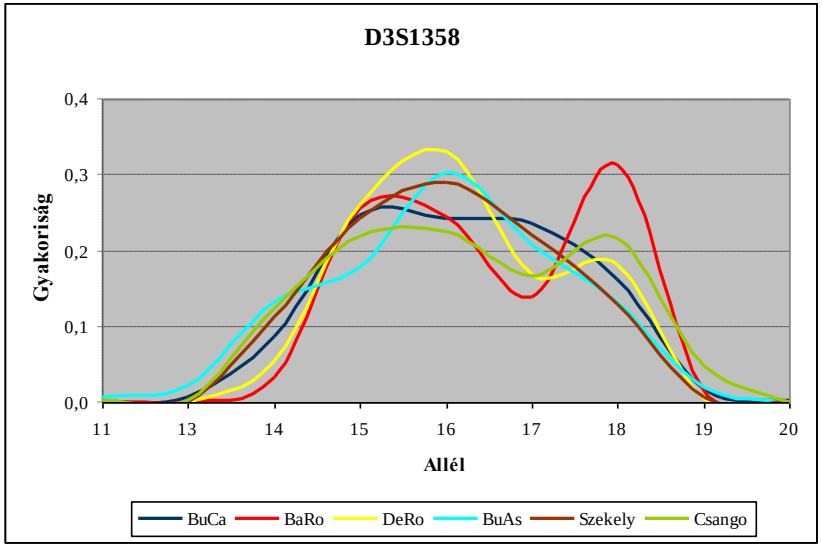
D13S317 allele 12

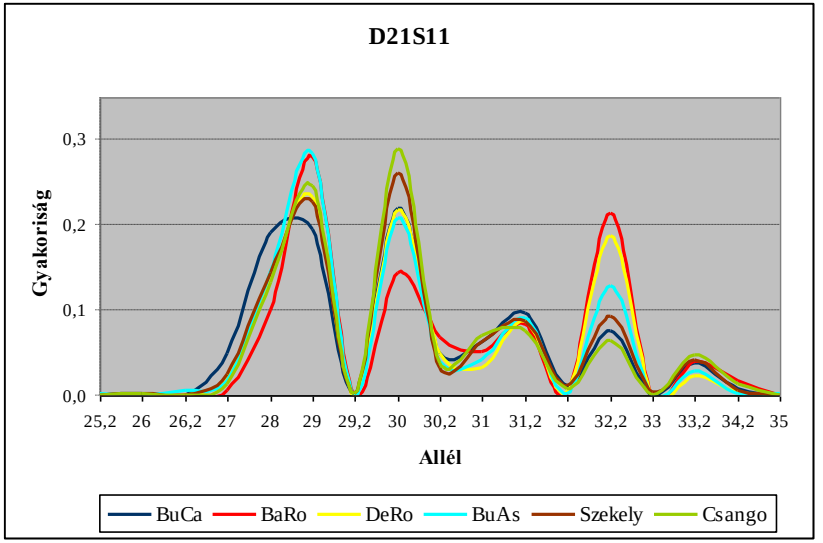
5'-gggttgctggacatggtatcACAGAAGTCTGGGATGTGGA---N82---$(ATCT)_{12}$ATCAATC$(ATCT)_3$TTCTGTC**TGTC**TTTTTGGGC---N36---gaccaacaattcaagctctc-3'

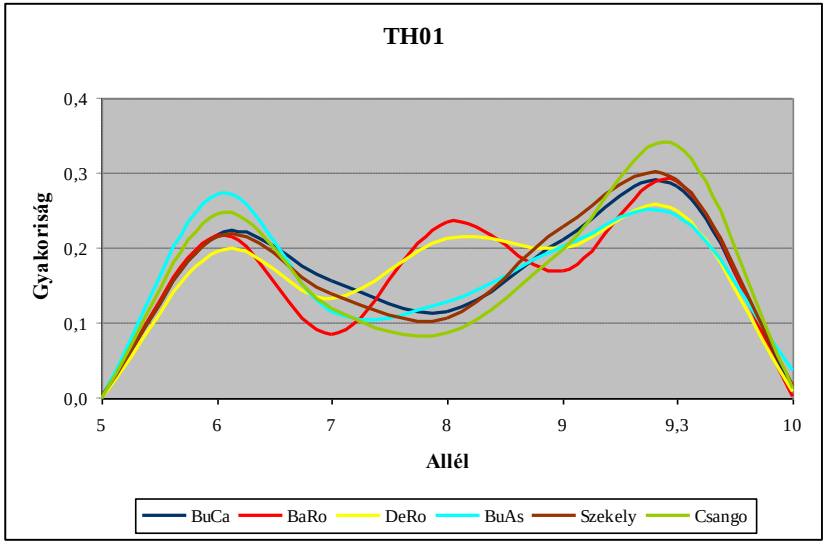
D13S317 allele 7 (variant)

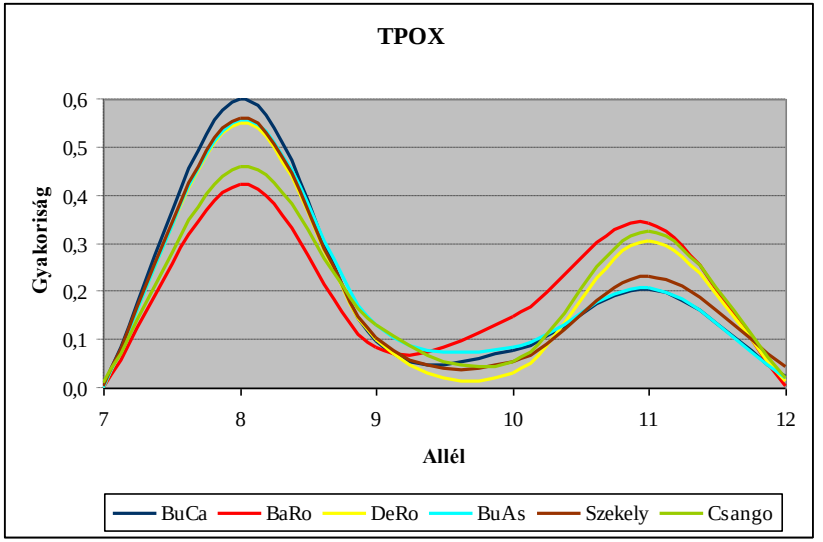
5'-gggttgctggacatggtatcACAGAAGTCTGGGATGTGGA---N82---$(ATCT)_7$ **ATCA**ATCAATC$(ATCT)_3$TTCTGTCTTTTTGGGC---N36---gaccaacaattcaagctctc-3'

Microsatellite allele frequency diagrams







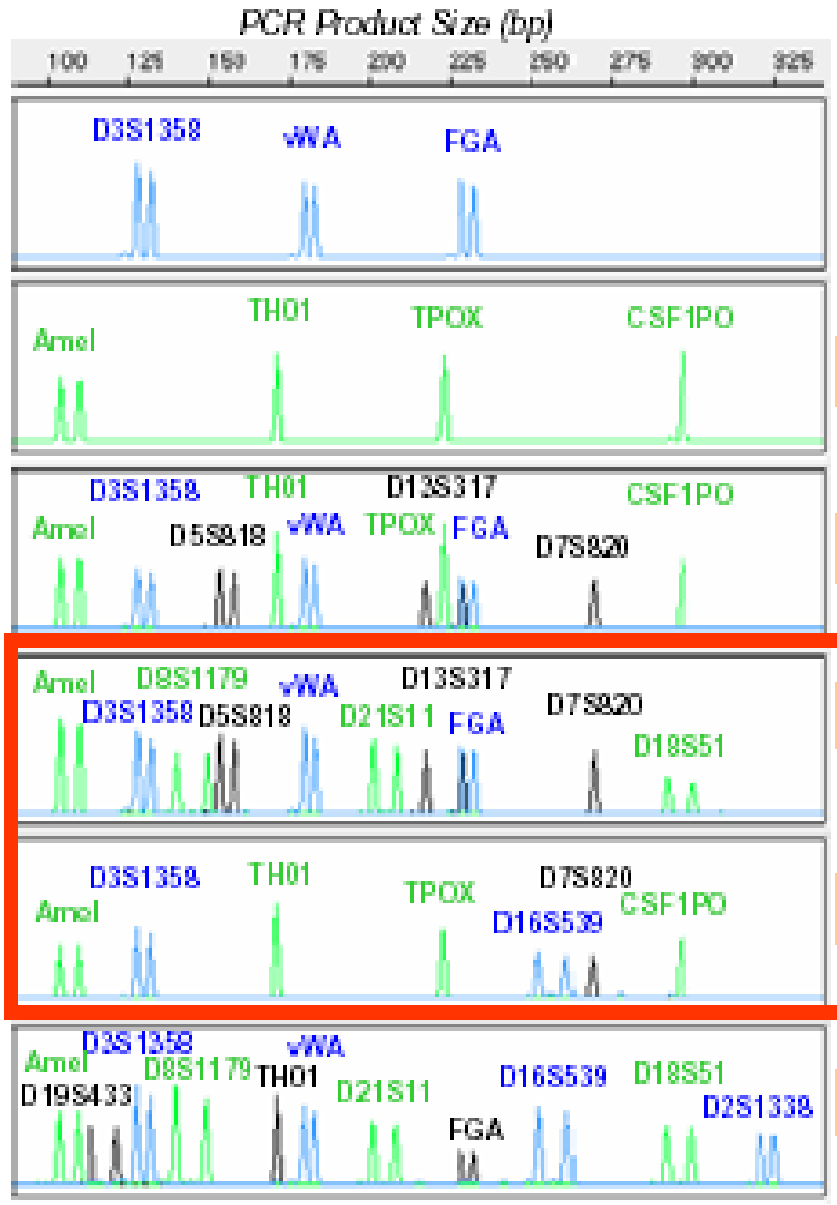


How Statistical Calculations are Made

- 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)



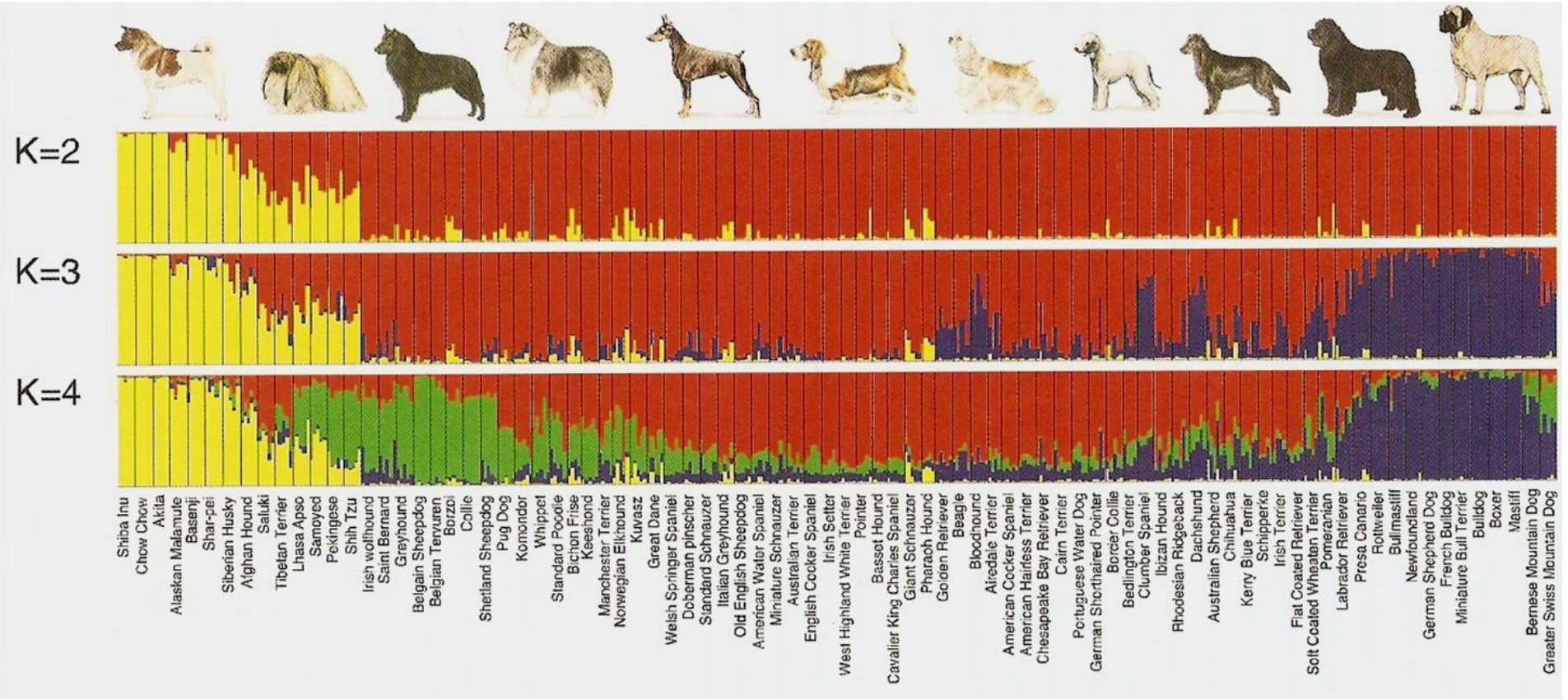
Same DNA Sample Run with Each of the ABI STR Kits

	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™	$1:3.3 \times 10^{12}$

Breed identification?

STRUCTURE statistics

1. Ancient 2. Sheperd 3. Hunting 4. Job

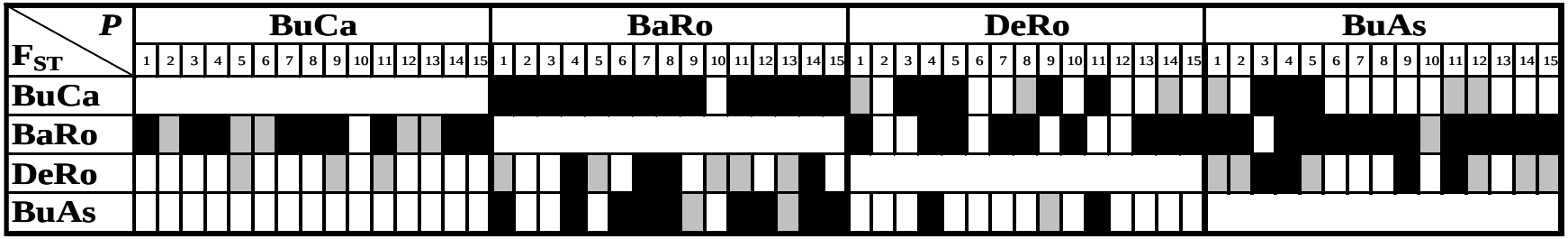


- 85 breeds (n=414)
- 95 microsatellites (dinucleotide repeat)
- 99% correct prediction!

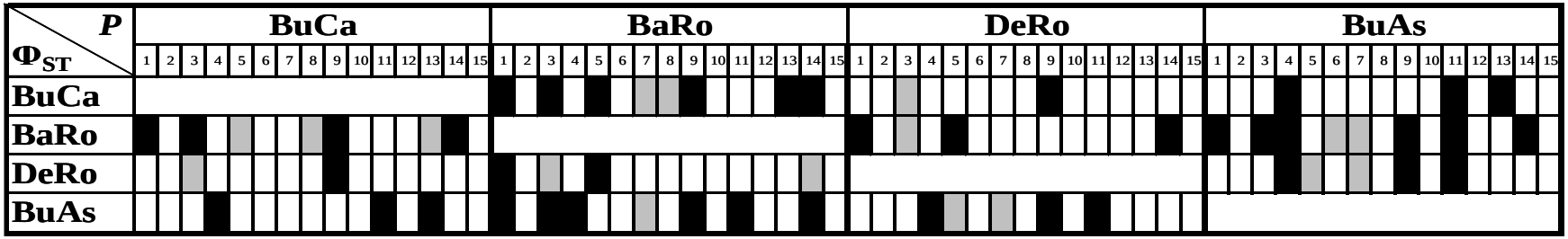
https://web.stanford.edu/group/pritchardlab

Genetic Structure – Analysing of MOlaculare VAriance (AMOVA)

a,

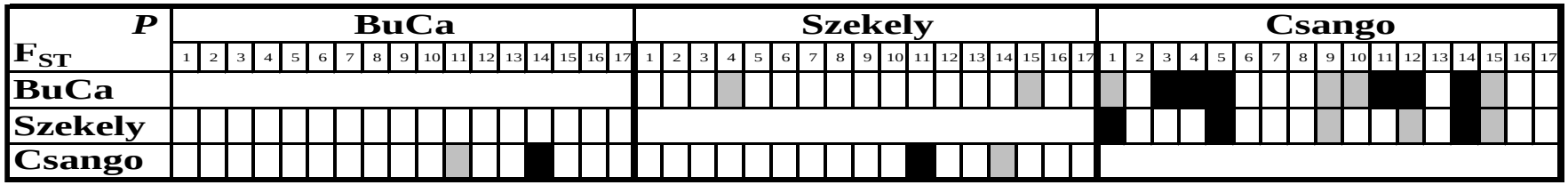


b,

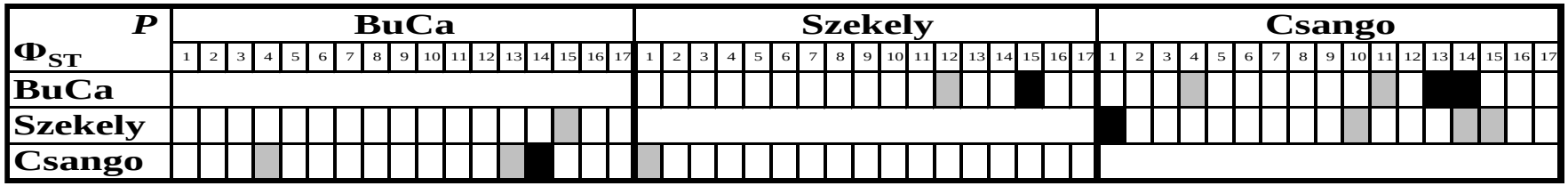


Jelölések: ■ $(F_{ST}, \Phi_{ST}) > 0,02$; $P < 0,1$ ▒ $0,02 > (F_{ST}, \Phi_{ST}) > 0,01$; $0,01 < P < 0,05$ □ $(F_{ST}, \Phi_{ST}) < 0,01$; $P > 0,05$

a,

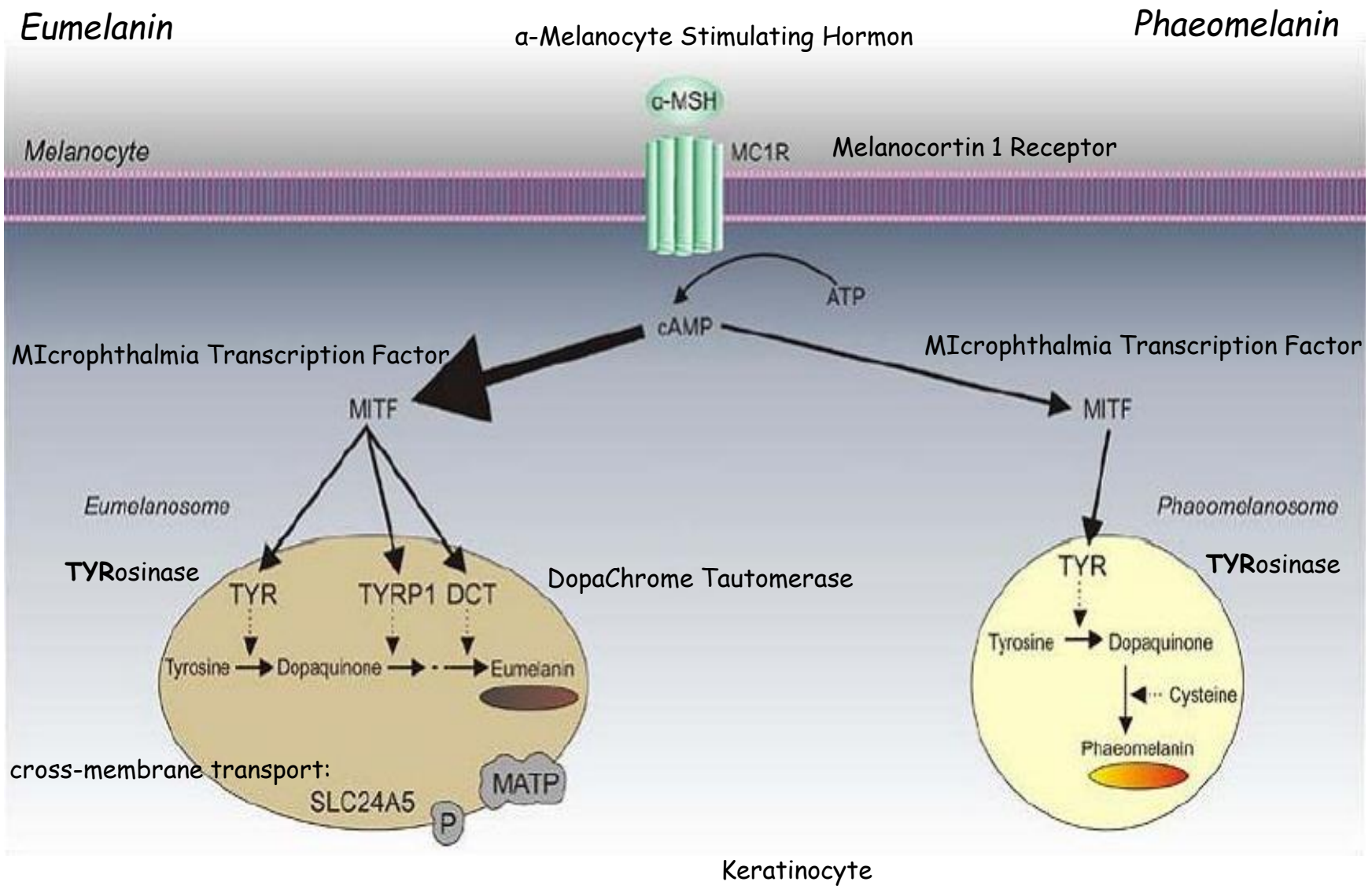


b,



Jelölések: ■ $(F_{ST}, \Phi_{ST}) > 0,02$; $P < 0,1$ ▒ $0,02 > (F_{ST}, \Phi_{ST}) > 0,01$; $0,01 < P < 0,05$ □ $(F_{ST}, \Phi_{ST}) < 0,01$; $P > 0,05$

Human melanogenesis



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.

Pulker et al., FSI Genet (2007) 1:100

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]

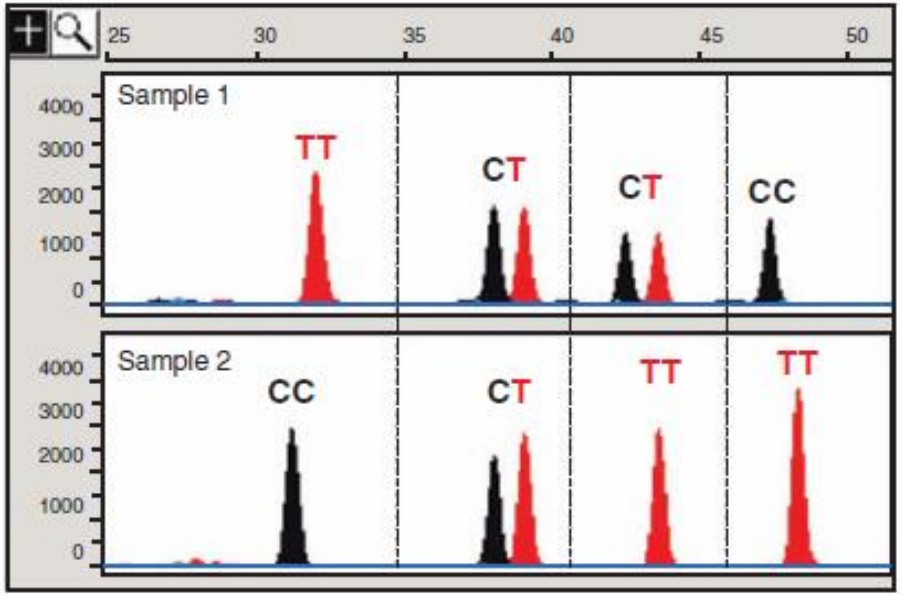
- MC1R alleles possess different activity levels
- 317 AA and 7 transmembran domen
- SNPs: RHC phenotype - neanderthal pigmentation
- Phenotype prediction? Genetic tests?

Tully, FSI Genetics (2007) 1:105

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.

(a)



(b) (TTTTT)–**primer1** (chromosome 20)-**ddT/ddT**

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

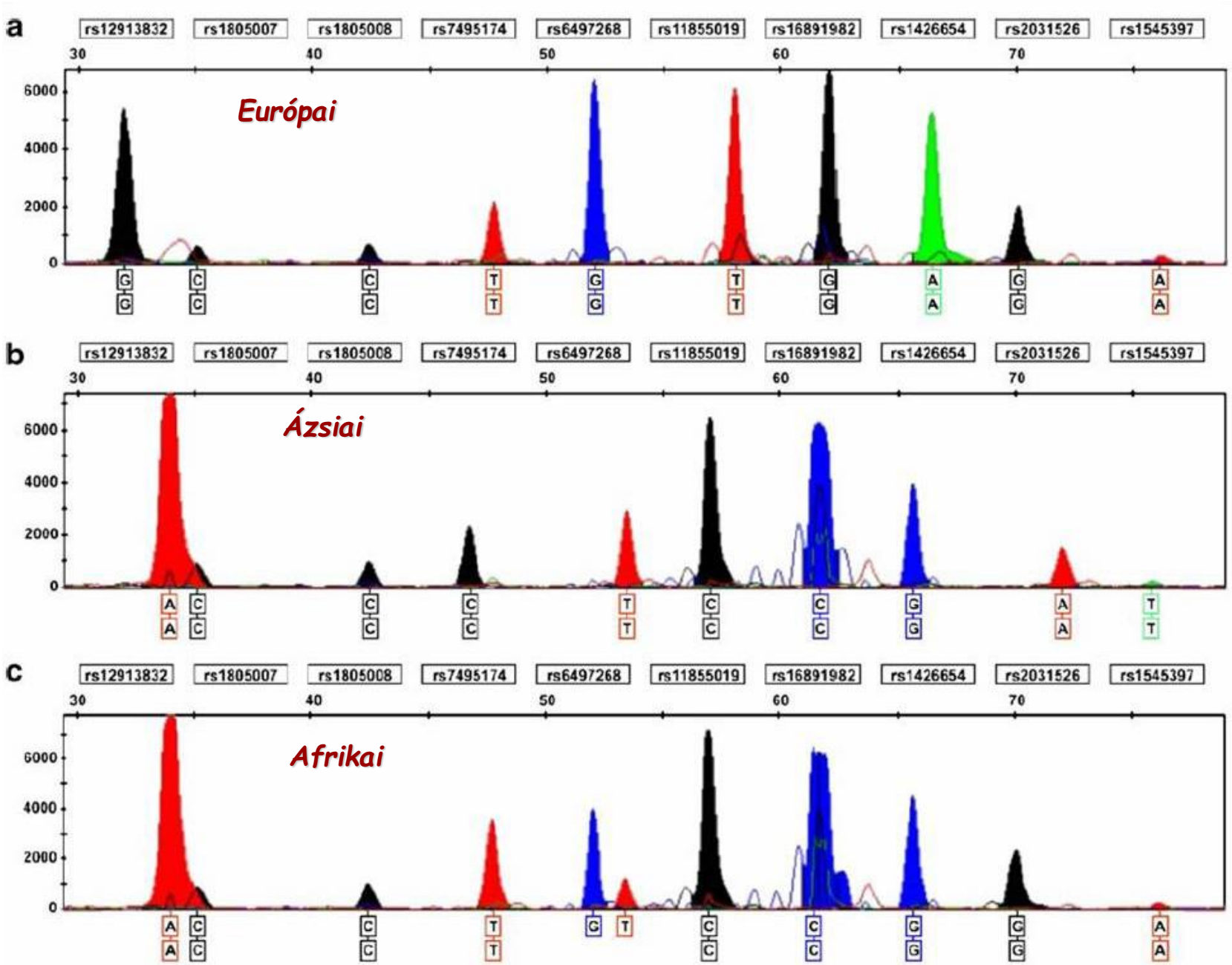
(TTTTT)–(TTTTT)–(TTTTT)–**primer3** (chromosome 14)-**ddC/ddT**

(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).

Butler, J.M. (2011) Advanced Topics in Forensic DNA Typing, p. 354

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/G	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	G/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