

GENETIKA ÉS POPULÁCIÓGENETIKA

Genetikai polimorfizmusok



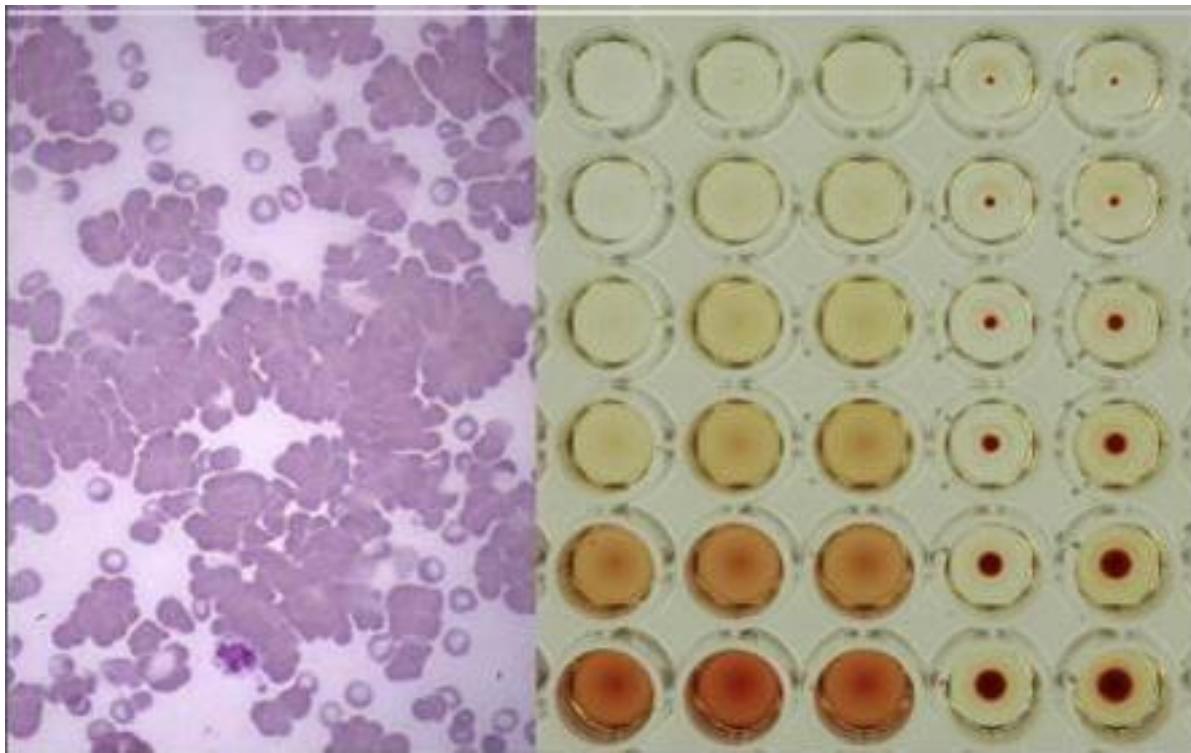
ELTE TTK Genetikai Tanszék

Recesszív gén hőmérséklet szenzitív expressziója

- TYR gén ► tirozináz enzim ► melanin szintézis (sötét szín)
 - Defektív tirozináz ► funkcióvesztés normál testhőmérsékleten
 - sötét színárnyalat csak az alacsonyabb testhőmérséklettű helyeken



Első genetikai marker: ABO vércsoport-rendszer



Landsteiner, 1900



FELIX BERNSTEIN (1933)

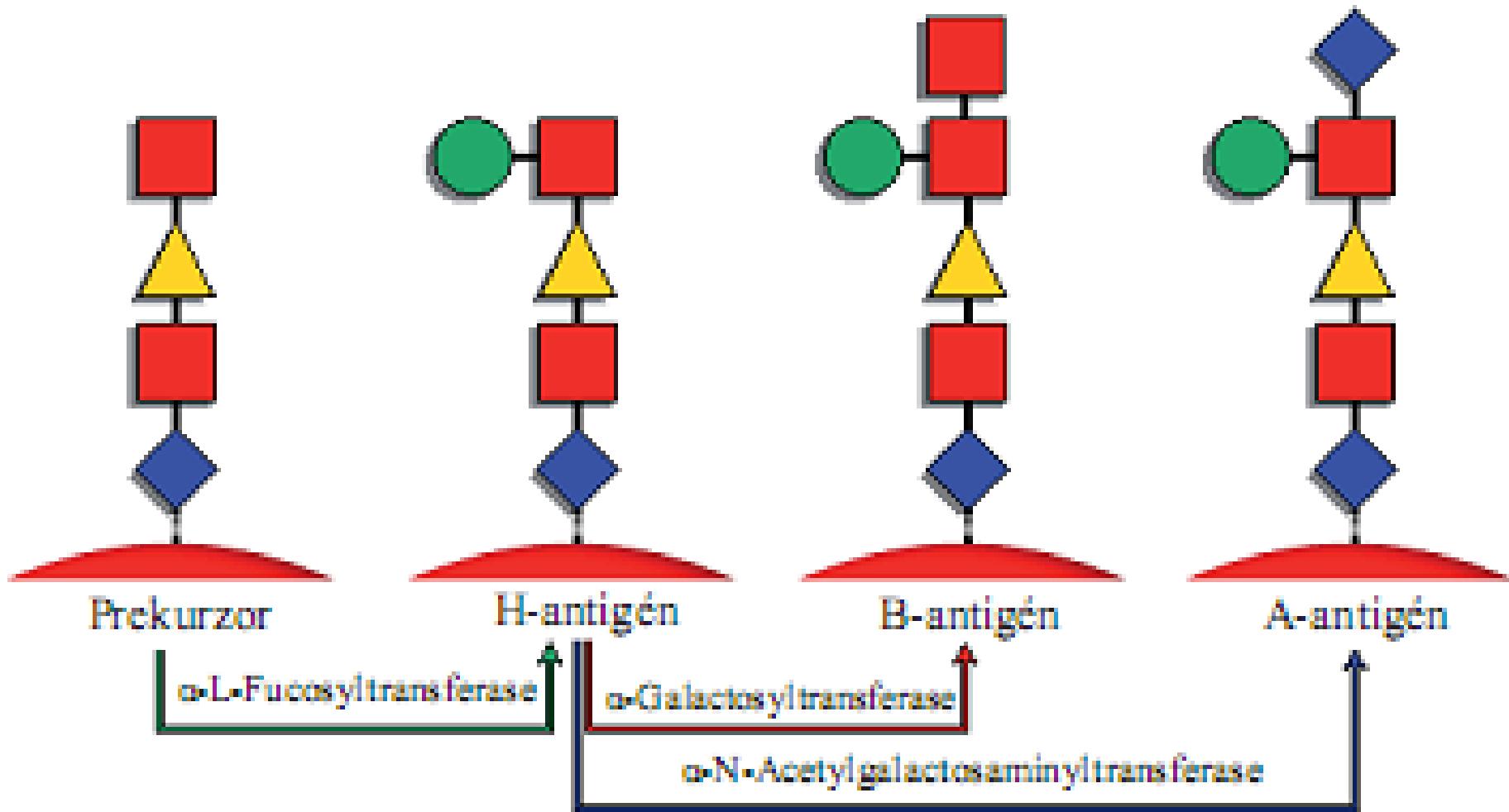
Two hypotheses of blood group inheritance

Group	VON DUNGERN and HIRZFIELD		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	$p_A^2 + 2p_O p_A$	0.422
B	$aa\ B-$	$p_a^2 (1 - p_b^2)$	BB, OB	$p_B^2 + 2p_O p_B$	0.206
AB	$A-\ B-$	$(1 - p_a^2)(1 - p_b^2)$	AB	$2p_A p_B$	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).

- L-Fucose
- D-Galactose
- ◆ N-Acetylgalactosamine
- ▲ N-Acetylglucosamine

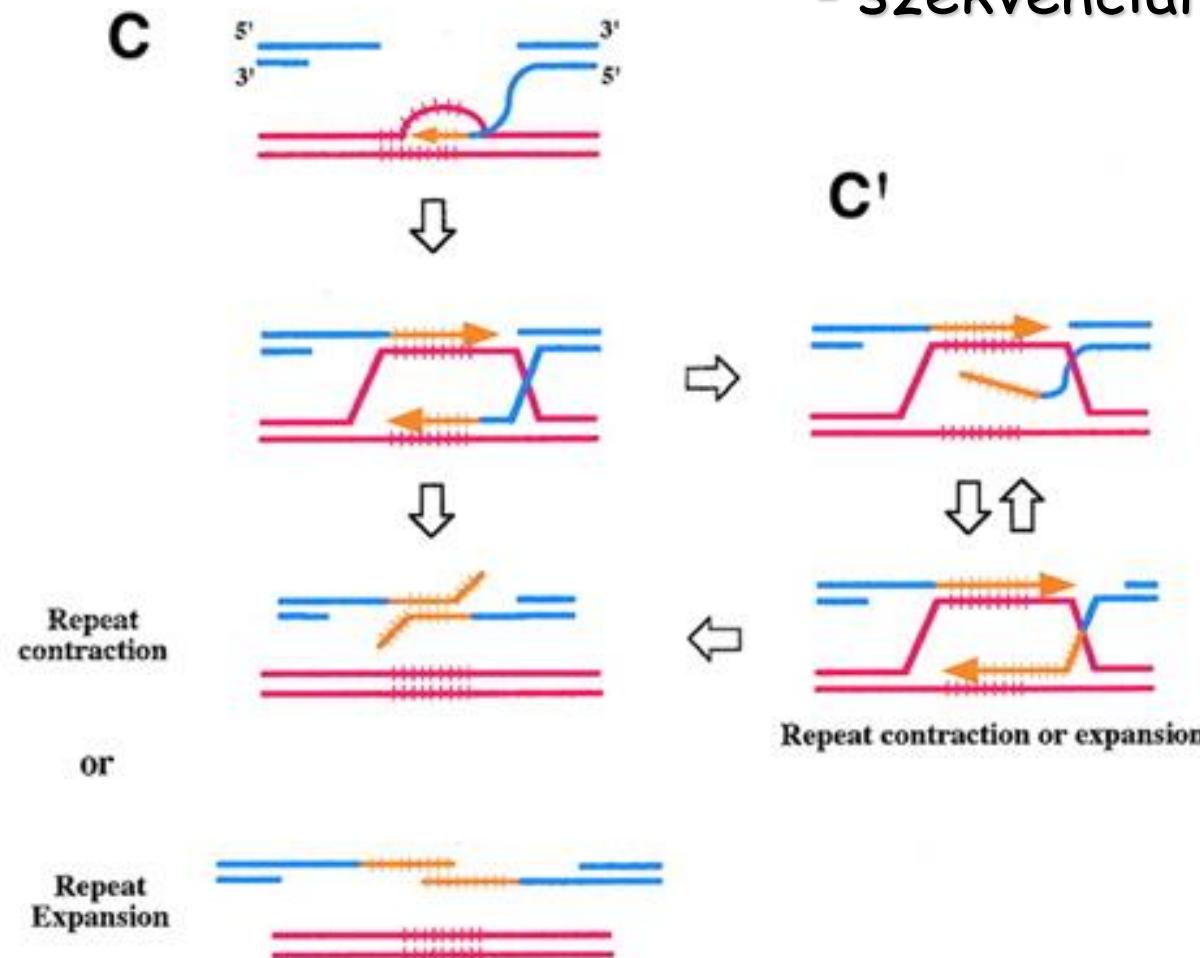
AB0 antigének kialakulása



A humán genom projekt eredményei

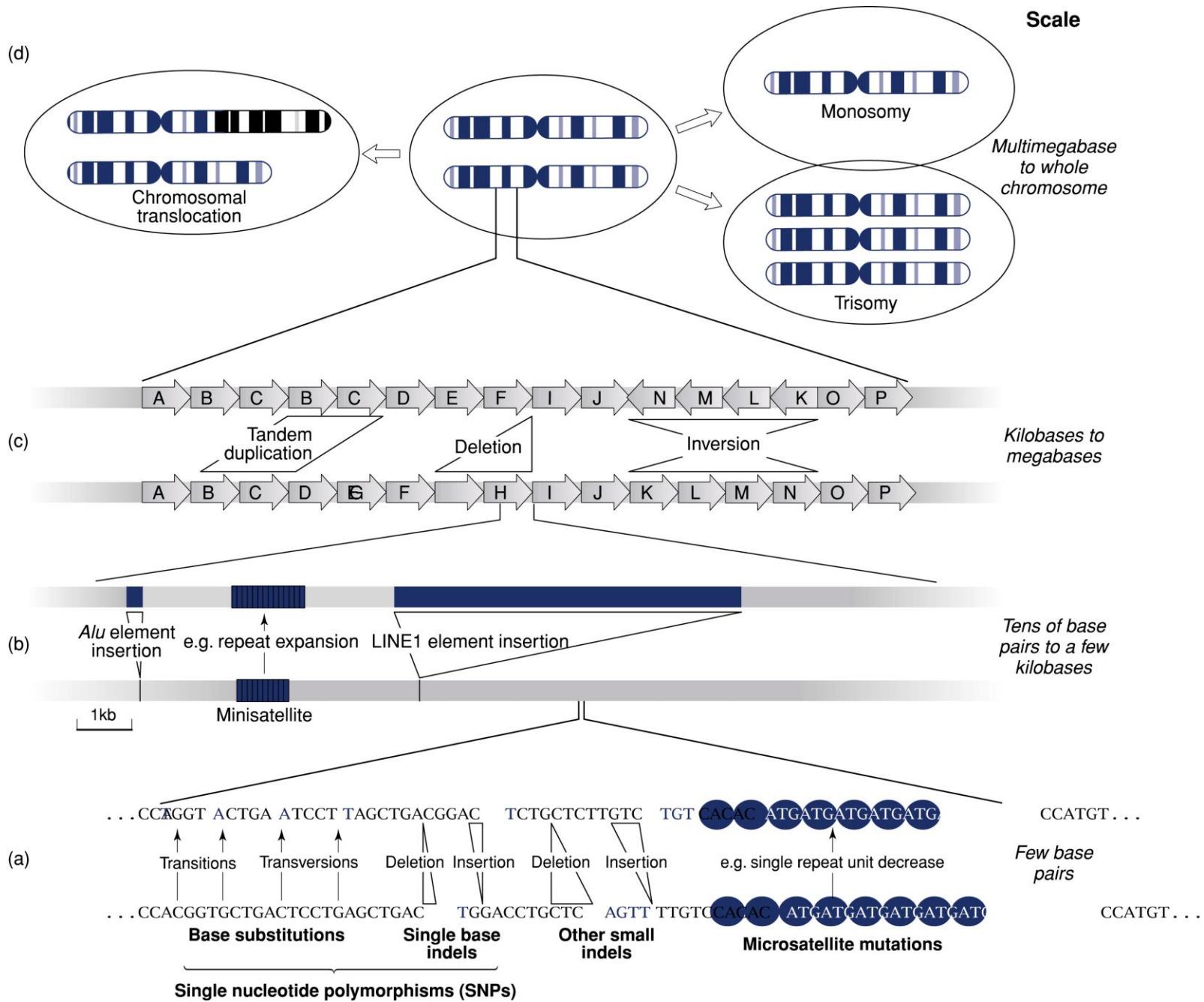
- Előzetes publikálás 2001-ben (Science, Nature)
- Az eddig leírt legnagyobb méretű teljes genom
- Az eukarióta genomra jellemző szerkezeti és szerveződési tulajdonságok (modellszervezetek)
- Meglepően alacsony a fehérje kódoló gének száma (~20000)
- RNS gének növekvő száma (snRNA, lcnRNA, miRNA, stb.)
- Orphan gének: nincs homológja, humán gének ~ 1/3-a
- Protein kódoló szekvenciák aránya: < 1 %
- **Repetitív szekvenciák túlnyomó aránya – Mobilis elemek!**

REKOMBINÁCIÓ



Polimorfizmus hajtóereje:

- pontmutáció
- szekvenciális átrendeződés



A global reference for human genetic variation

The 1000 Genomes Project Consortium*

IGSR: The International Genome Sample Resource

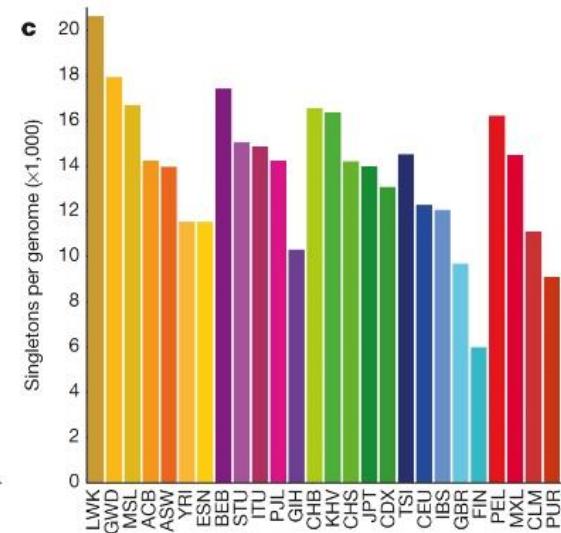
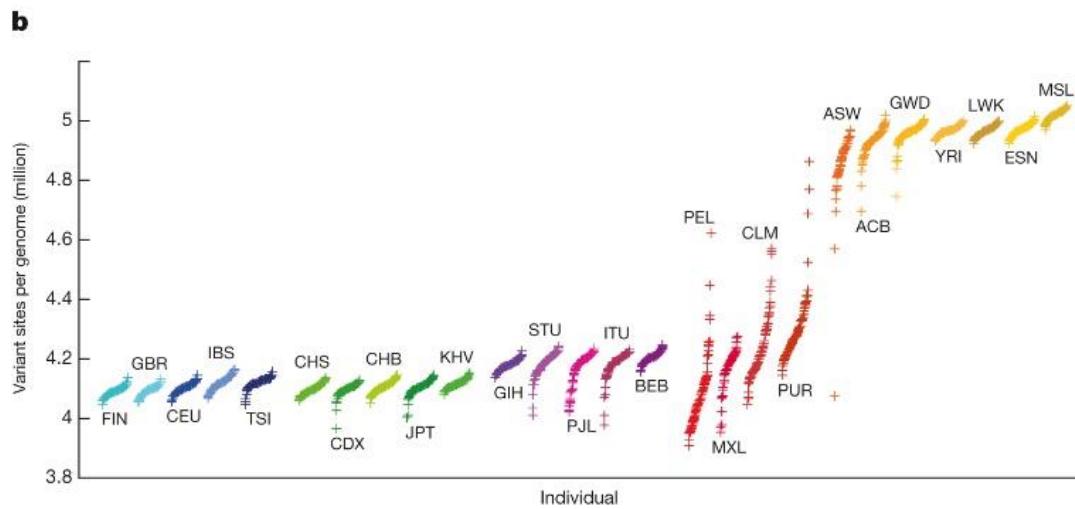
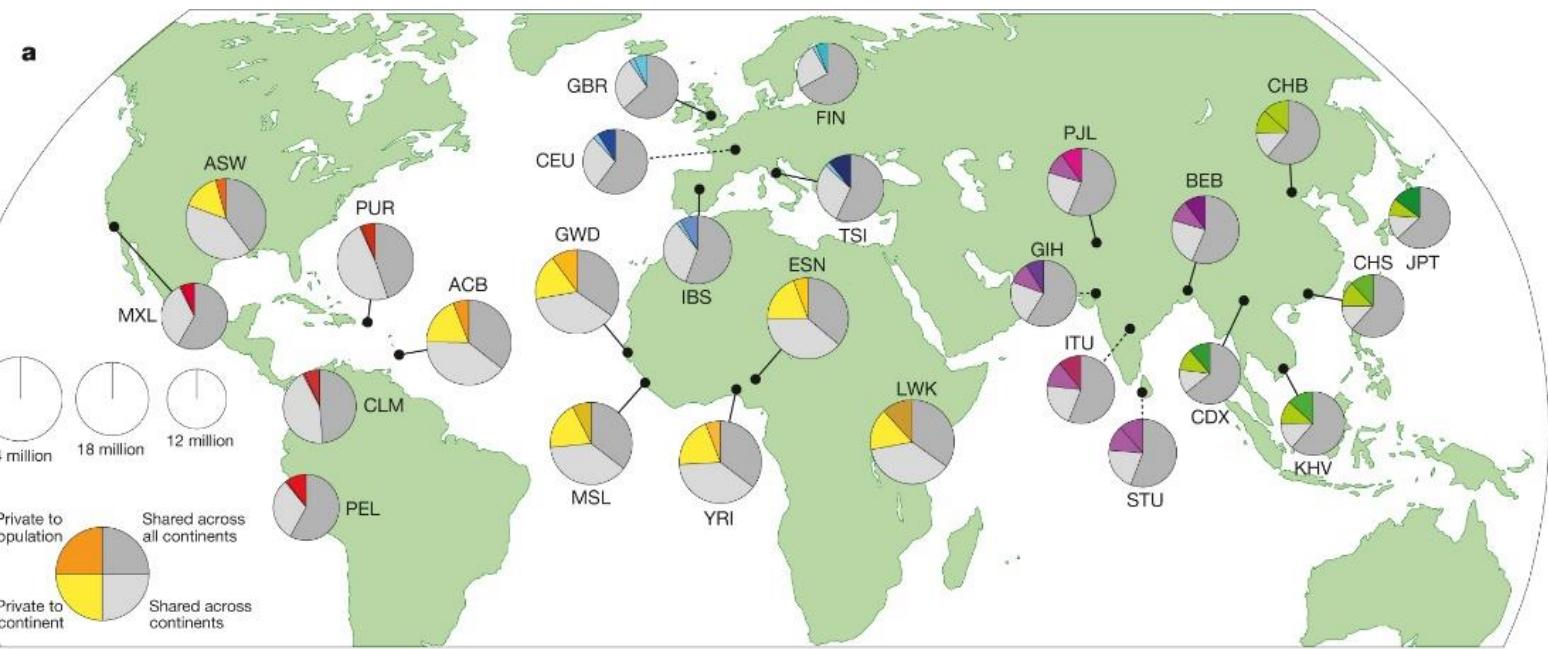
The 1000 Genomes Project set out to provide a comprehensive description of common human genetic variation by applying whole-genome sequencing to a diverse set of individuals from multiple populations. Here we report completion of the project, having reconstructed the genomes of 2,504 individuals from 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.

Population sampling



nature

A global reference for human genetic variation

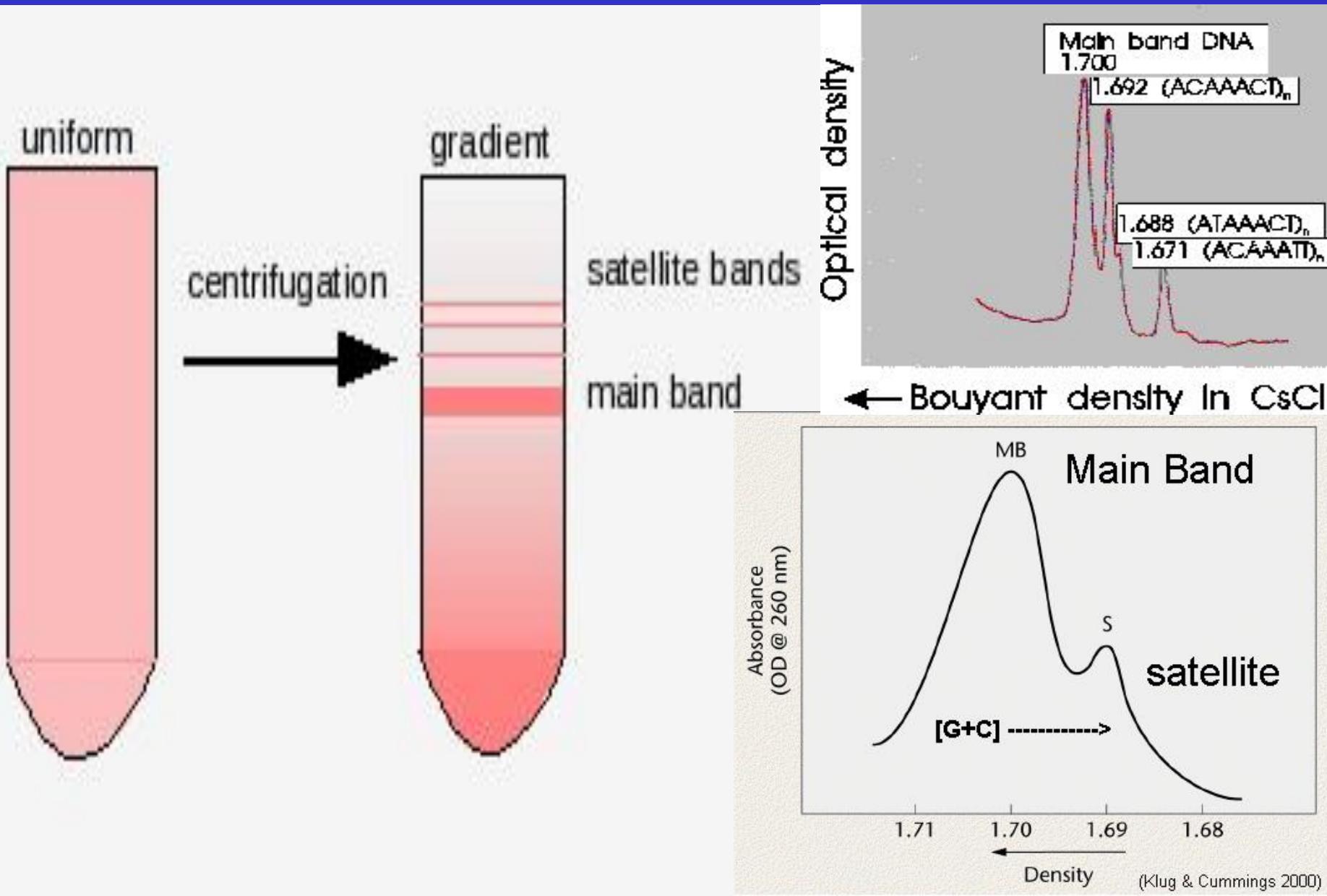
The 1000 Genomes Project Consortium*

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								
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.

Satellite DNA



„DNA fingerprinting“ – Restriction Fragment Length Polymorphism

Double-stranded DNA

Restriction enzymes

Gel electrophoresis

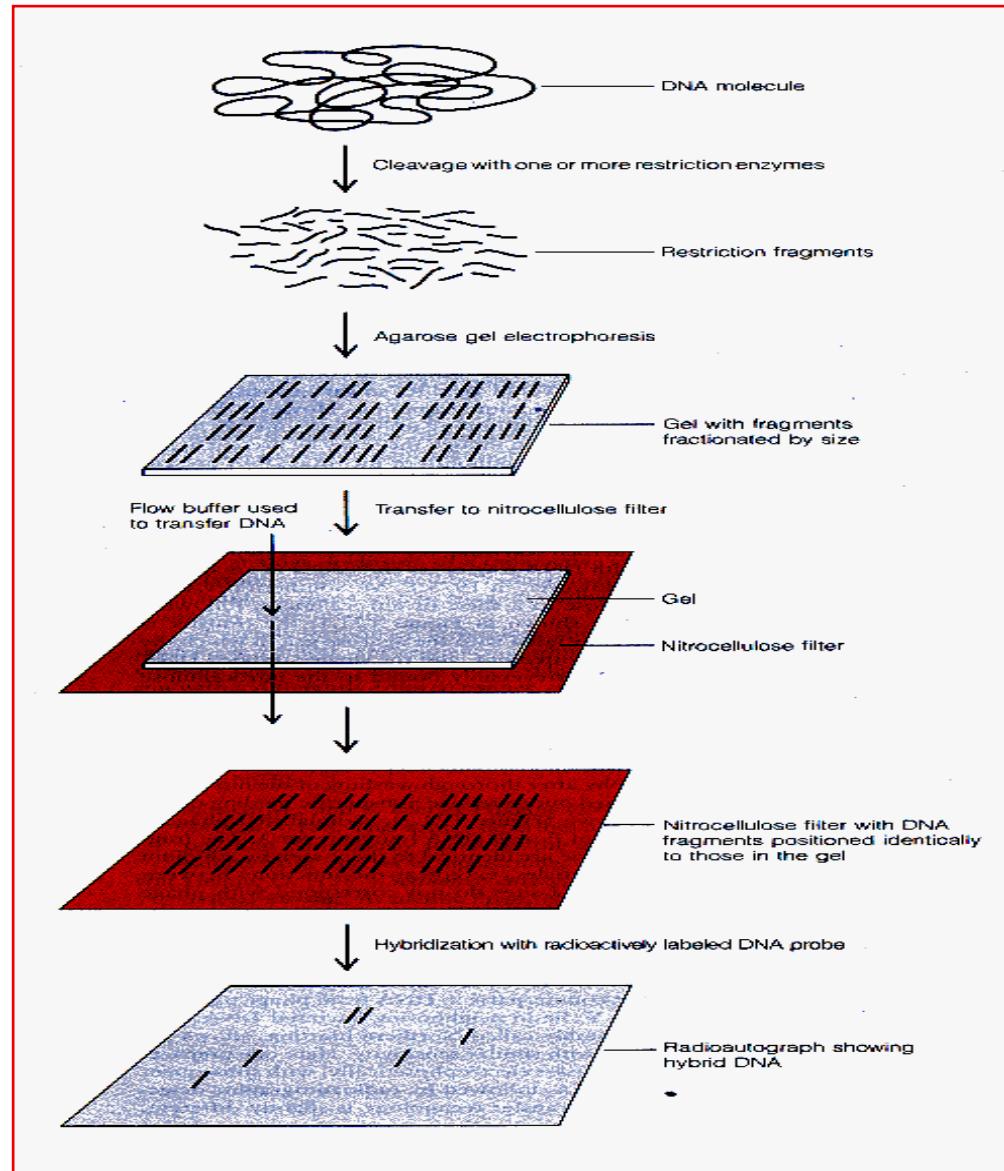
Southern-blot

Probe hybridization

Autoradiogram

- MLP-RFLP

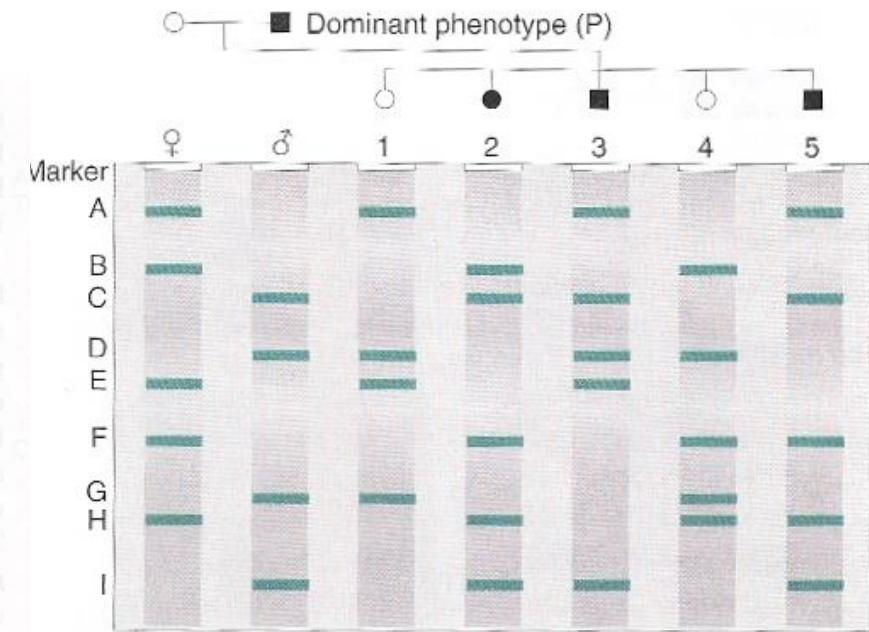
- SLP-RFLP





VNTR assay markerek: RFLP analitika

1985 - Sir Alec Jeffreys



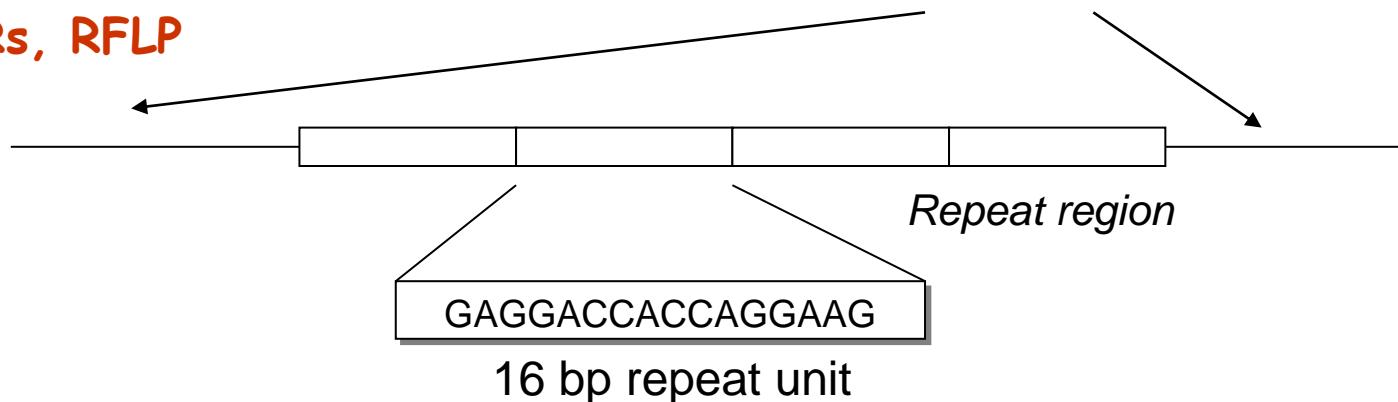
ANALYSIS EXAMPLES

- 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.

Genetikai térképezés

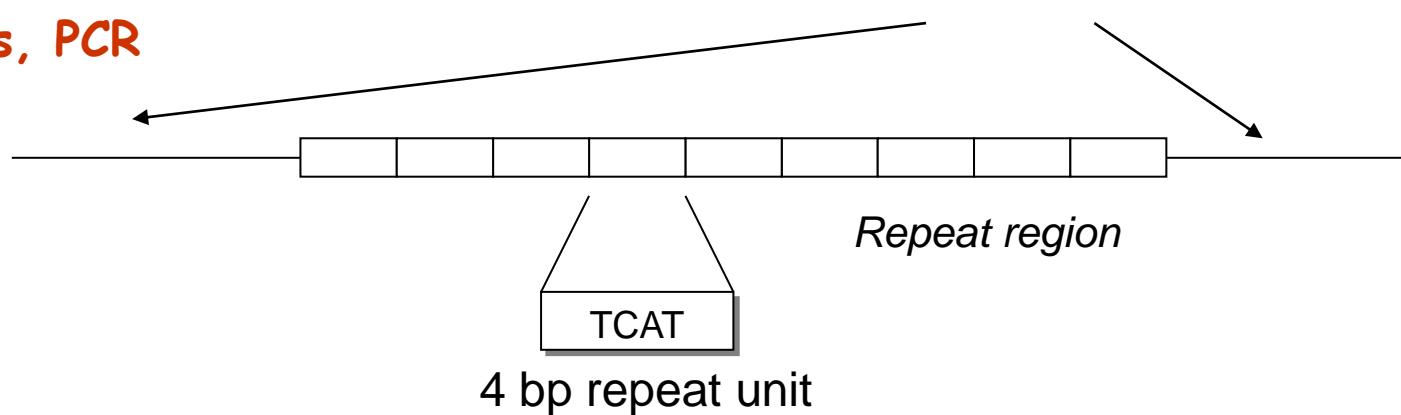
Minisatellite (D1S80)

VNTRs, RFLP



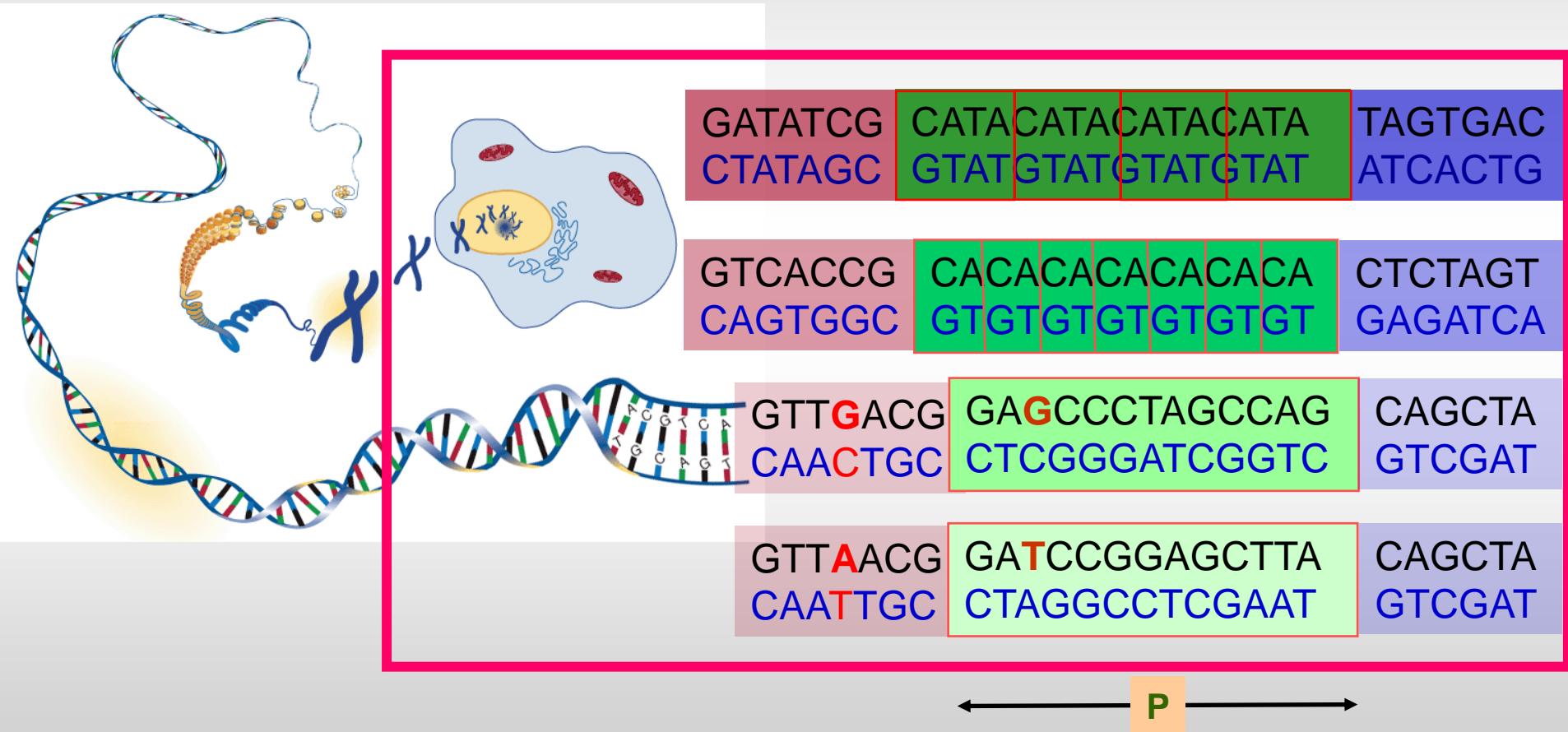
Microsatellite (TH01)

STRs, PCR



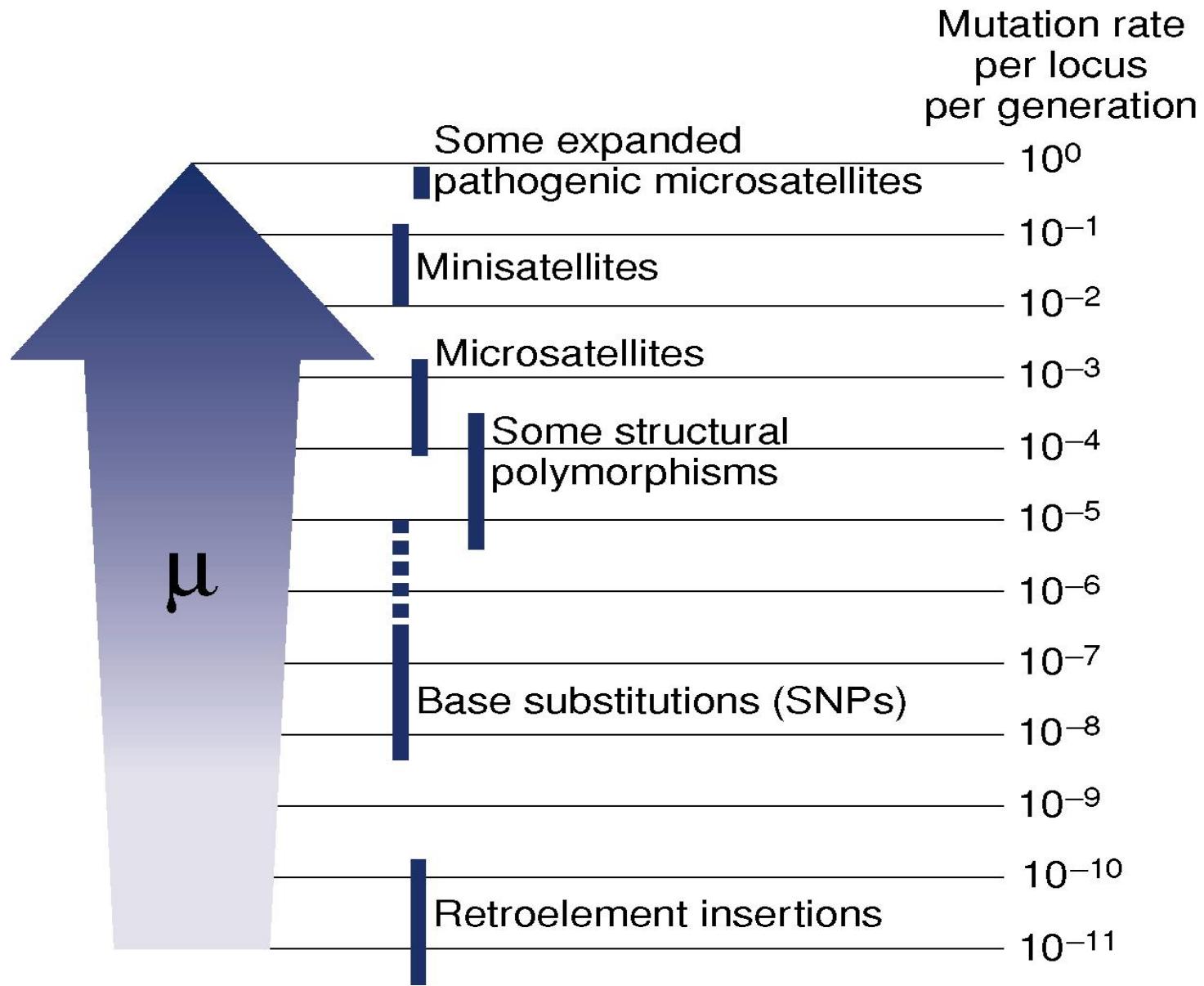
DNS polimorfizmusok

polimorfizmus - célrégió a PCR vizsgálatokban



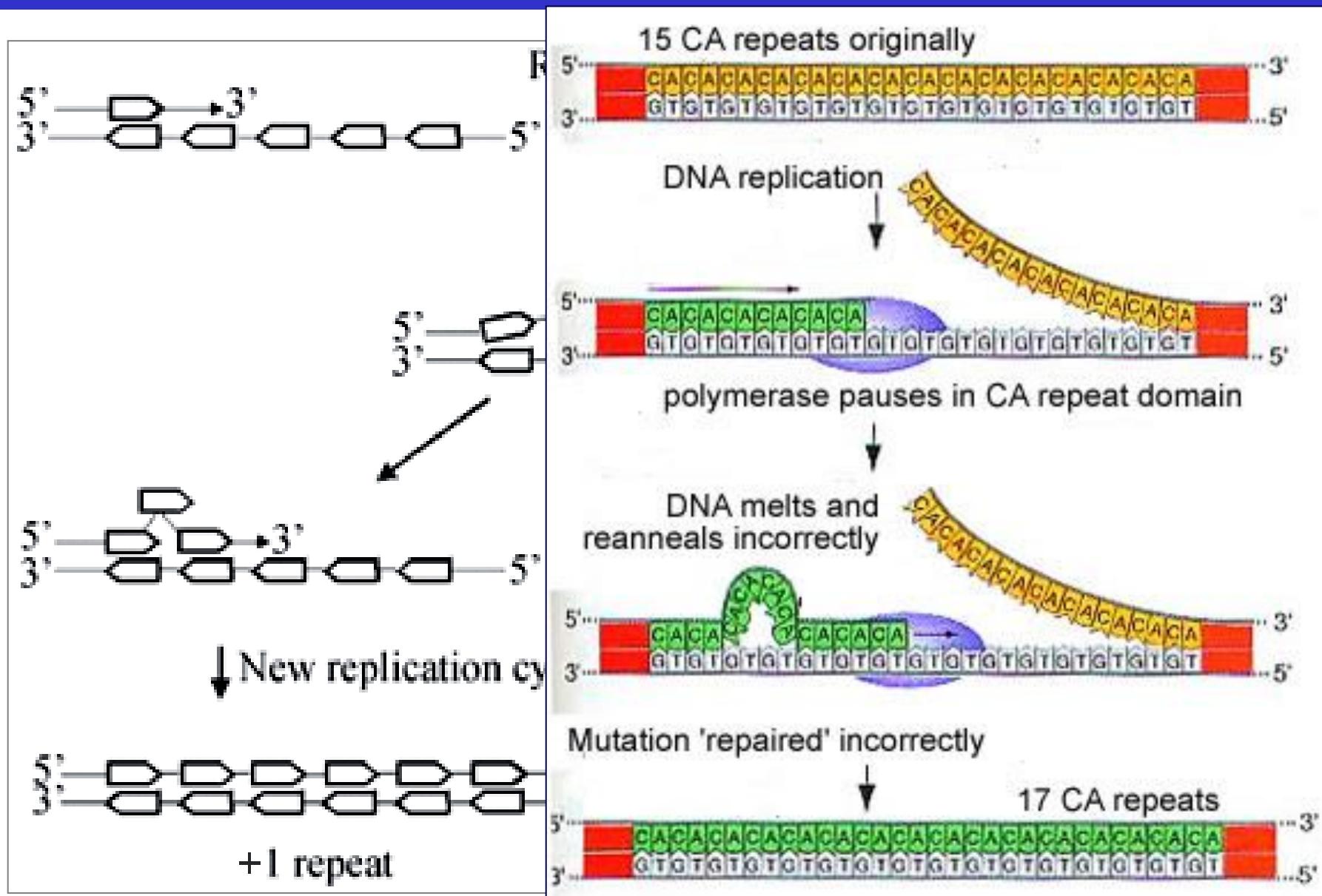
P : polimorf célszakasz (marker, lokusz, allél)

Polimorf szekvenciák mutációs rátája (μ)

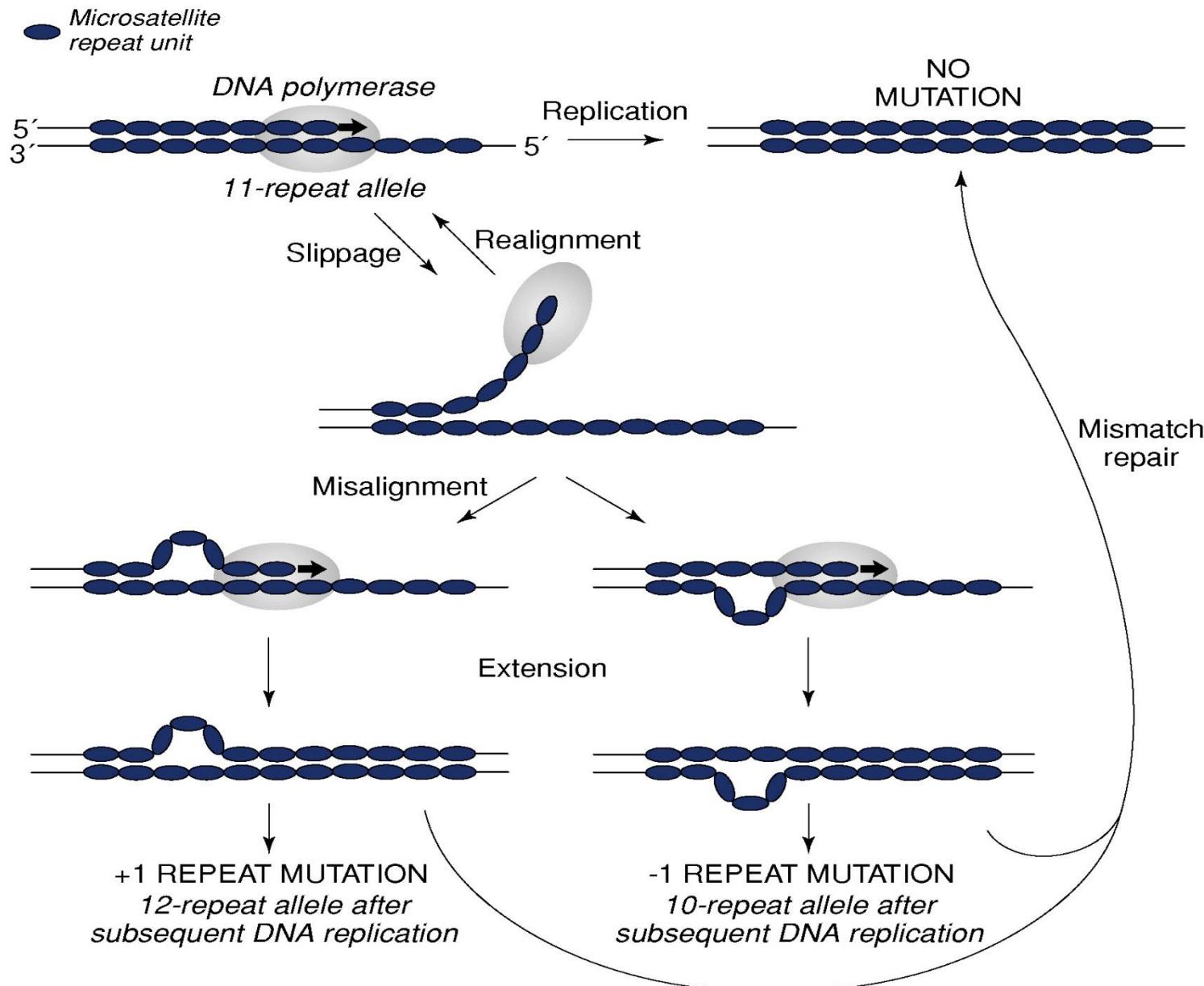


Microsatellite structure

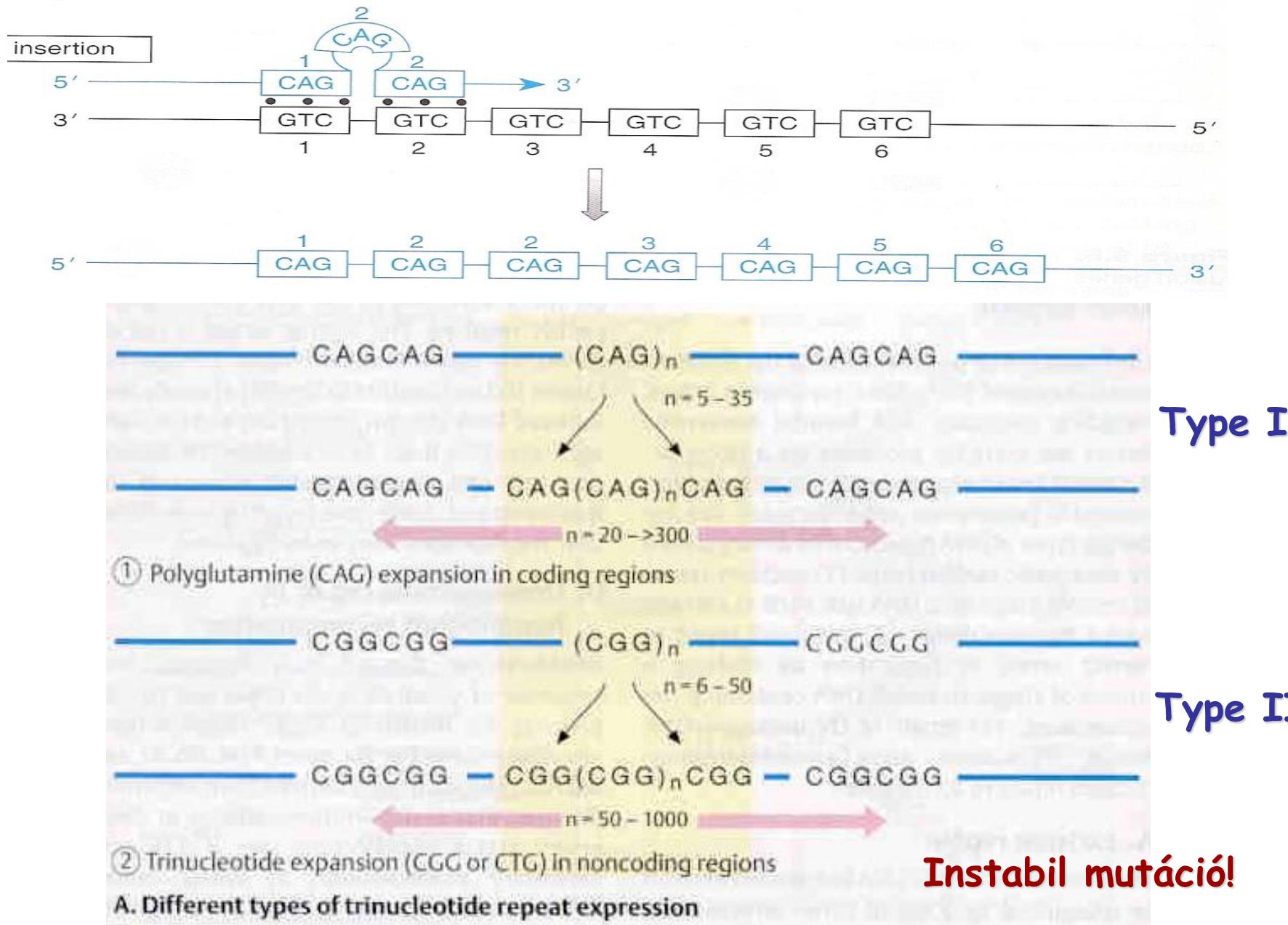
Mikroszatellita evolúció



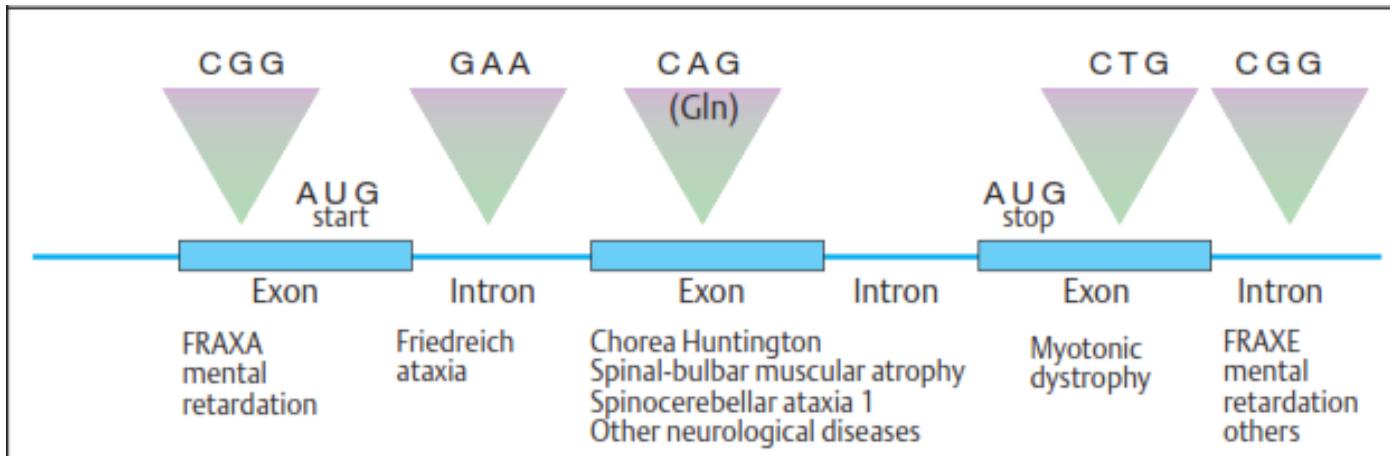
„Replikációs slippage” - Mikroszatellita mutáció



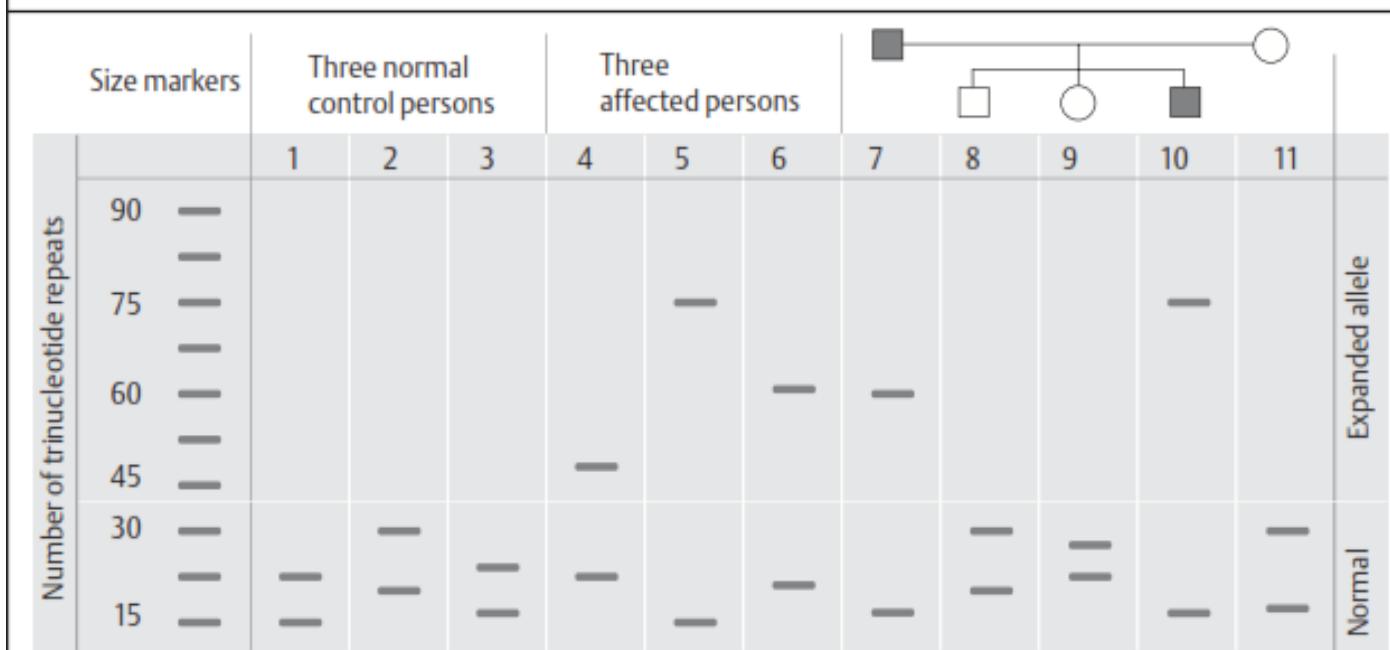
Trinukleotid repeat expansió



Trinucleotide repeat expansion



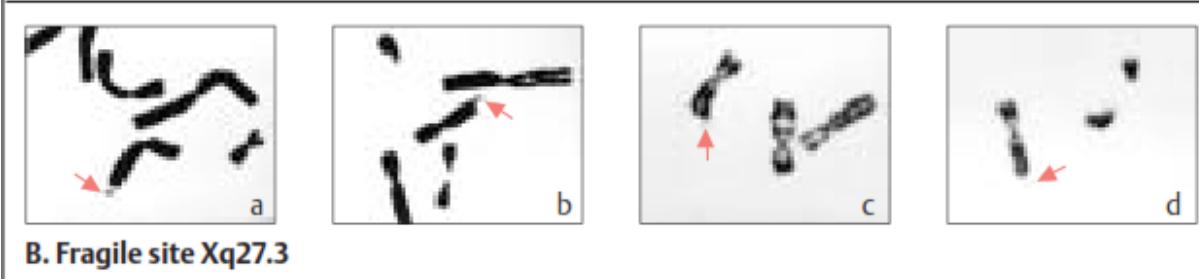
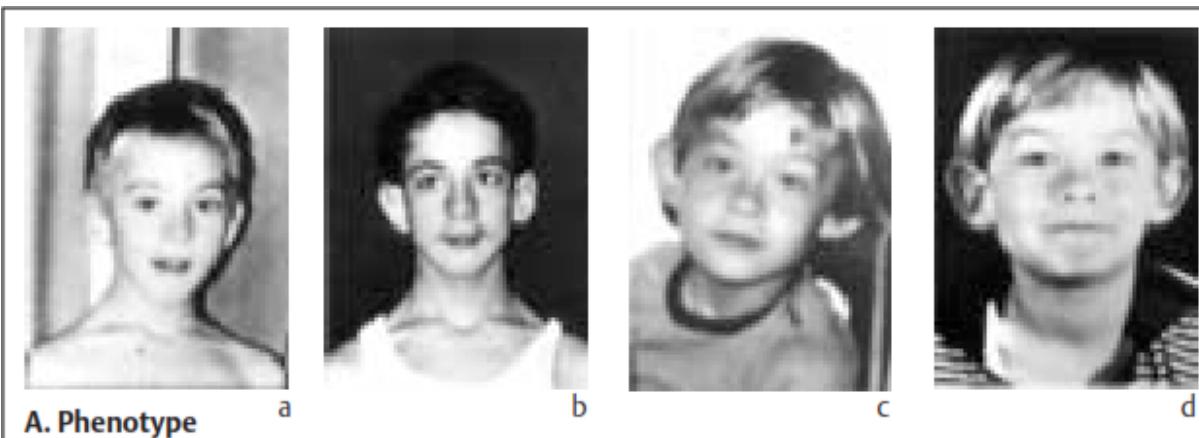
B. Unstable trinucleotide repeats in different diseases



C. Principle of laboratory diagnosis of unstable trinucleotide repeats leading to expansion

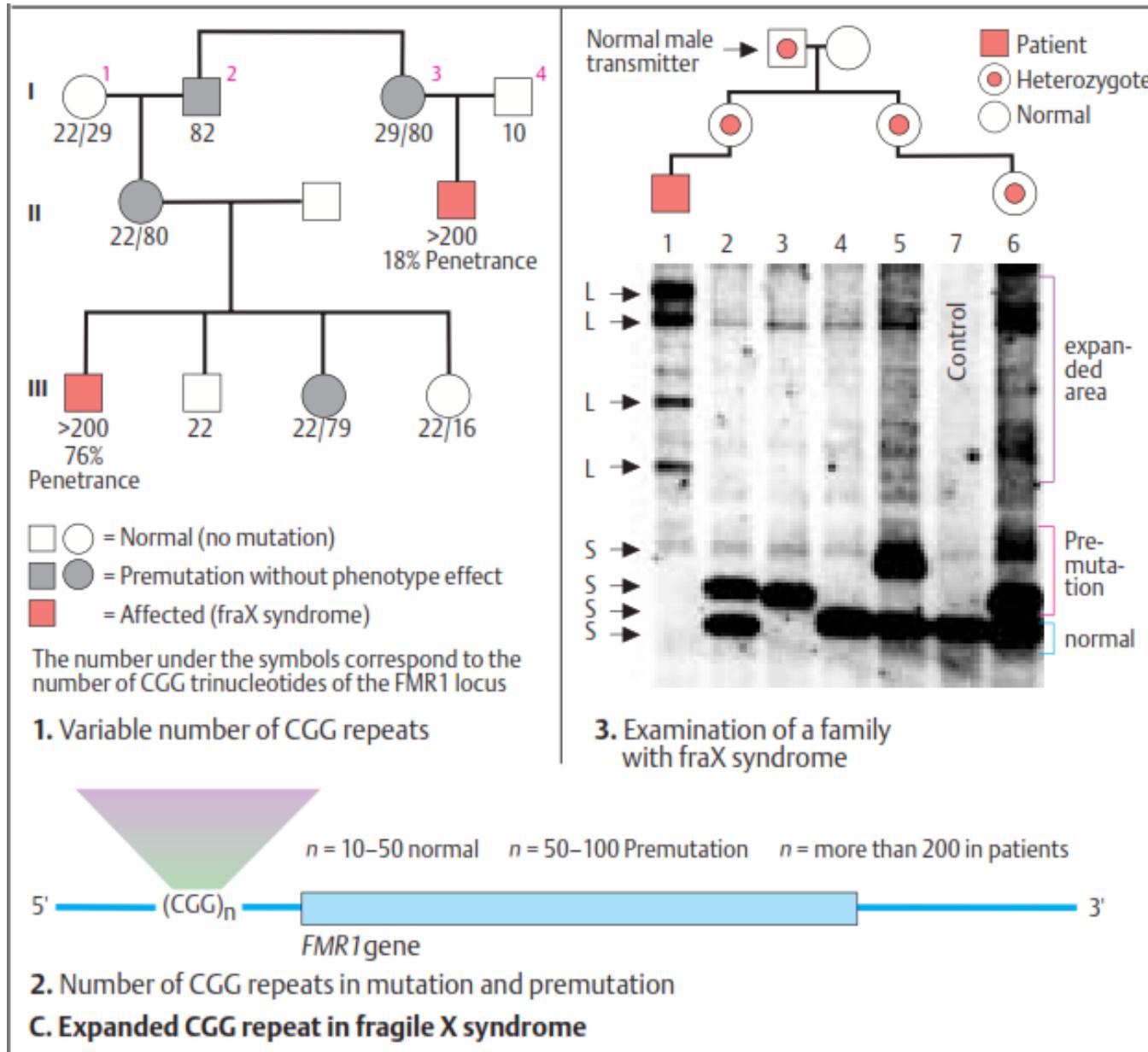
Genetic diseases due to repeat expansion

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



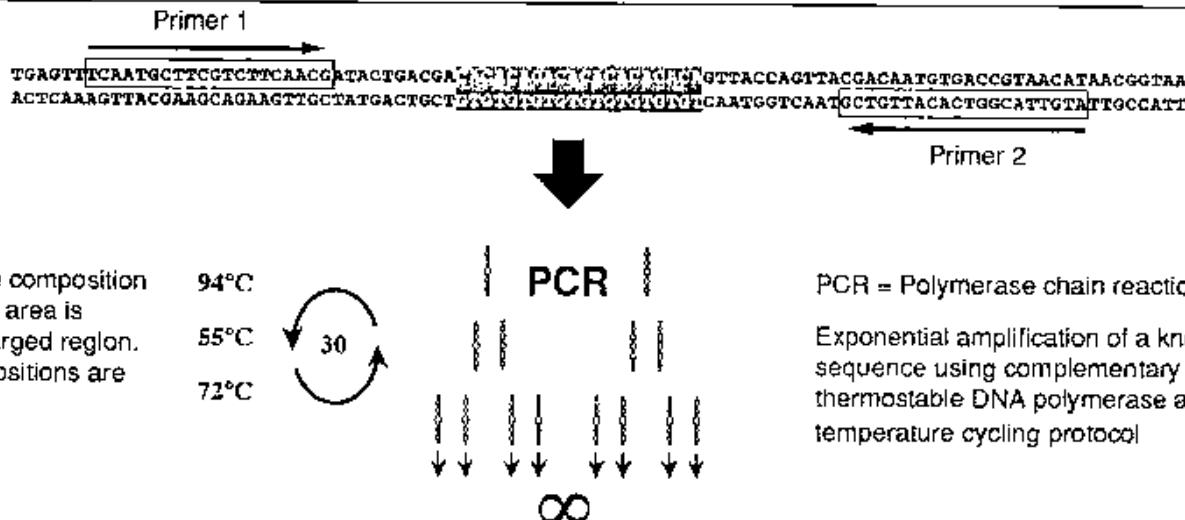
Fragile X
Huntington disease
Myotonic dystrophy
Friedrich ataxia
SMA
etc.

Expanded CGG repeats in the Fragile X syndrome



Mikroszatellita genotipizálás PCR-rel

Note: the nucleotide composition of the most variable area is indicated in the enlarged region. Bases at variable positions are highlighted.

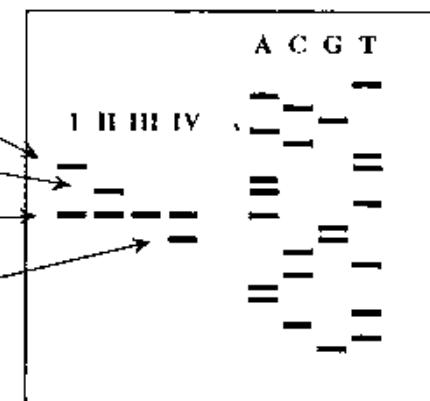


PCR = Polymerase chain reaction

Exponential amplification of a known DNA sequence using complementary primers, a thermostable DNA polymerase and a temperature cycling protocol

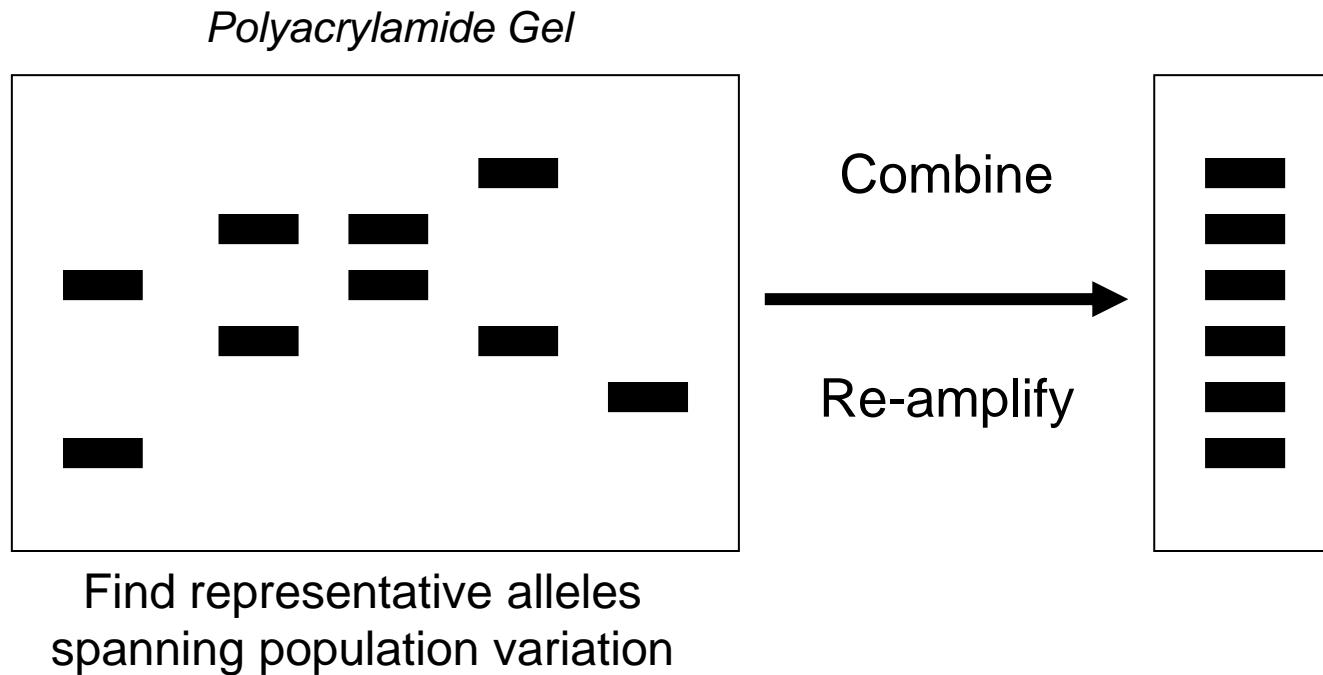
Allele length classes

Fractionated on sequencing gel

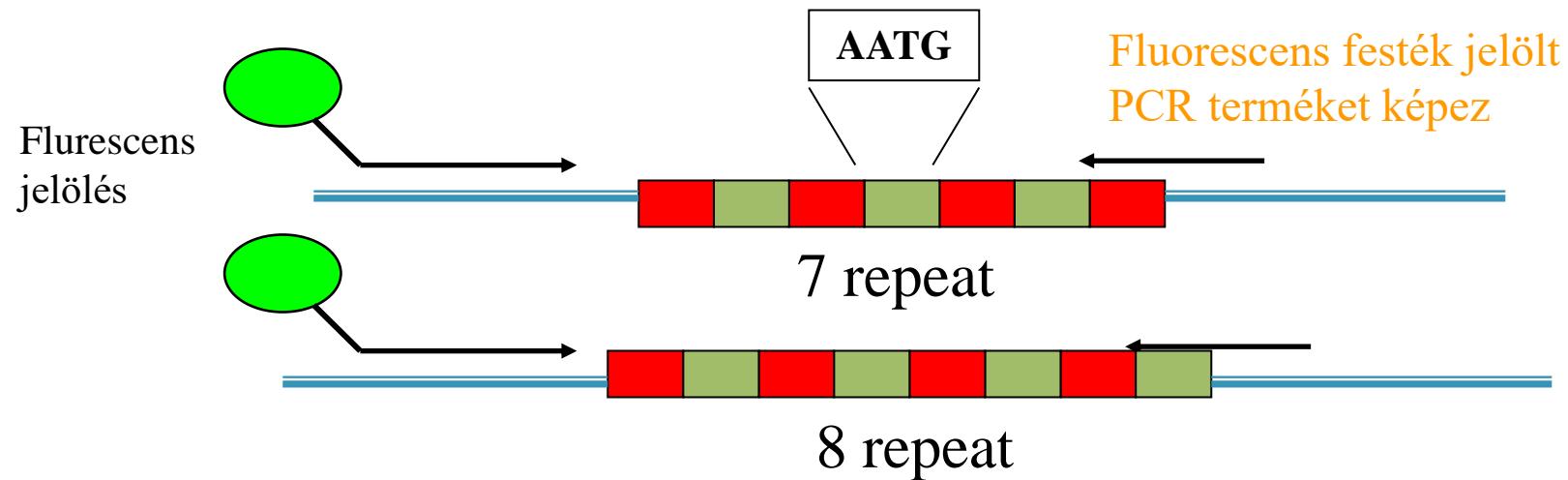


STR allélek genotipizálása: multiallélek

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



Mikroszatellita - STR - markerek (Short Tandem Repeat)



a repeat régió variabilis az egyes minták között míg a flanking régió ahol a PCR primerek tapadnak állandó

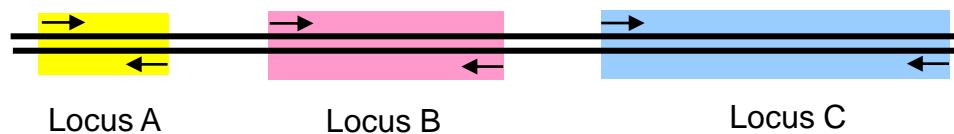
Homozigóta = a két allél megegyezik

Heterozigóta = allélek különbözők és egymástól elválaszthatók

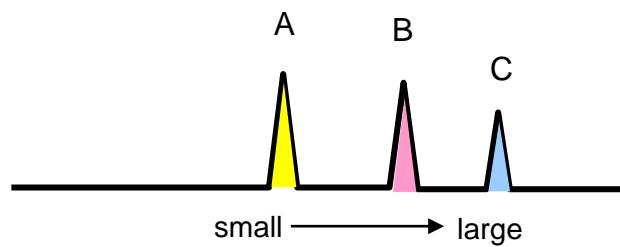
A primerek kötési helye meghatározza a PCR termék méretét!

Multiplex - PCR

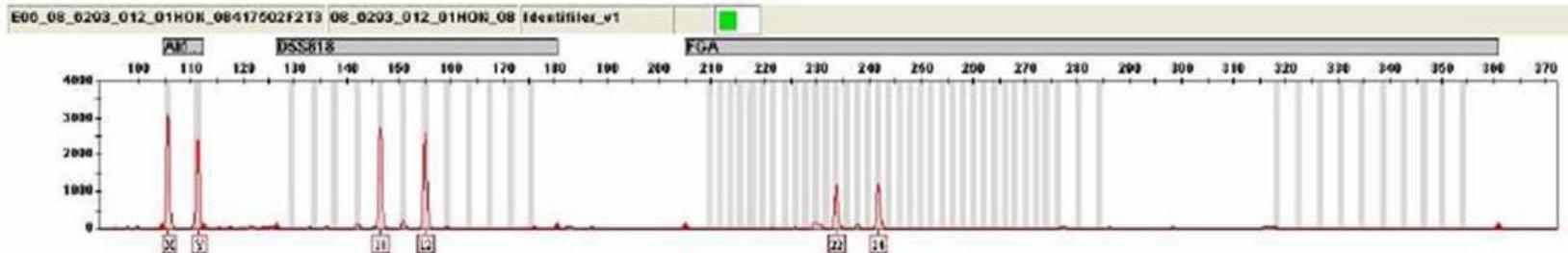
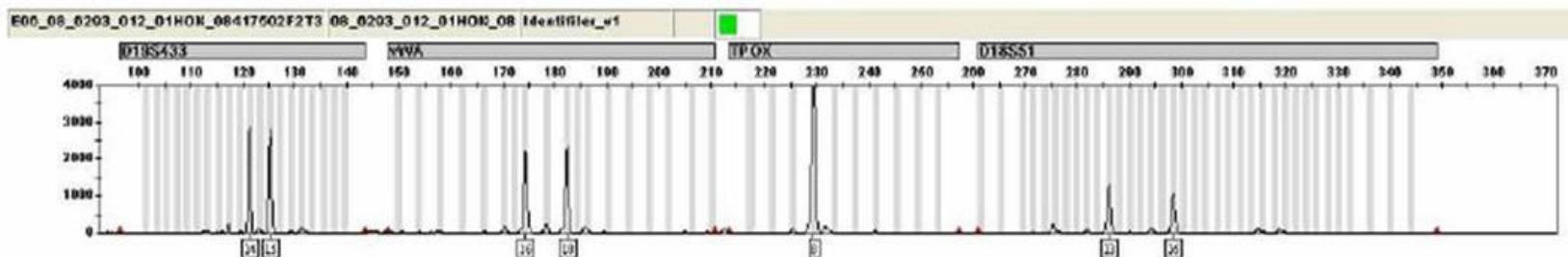
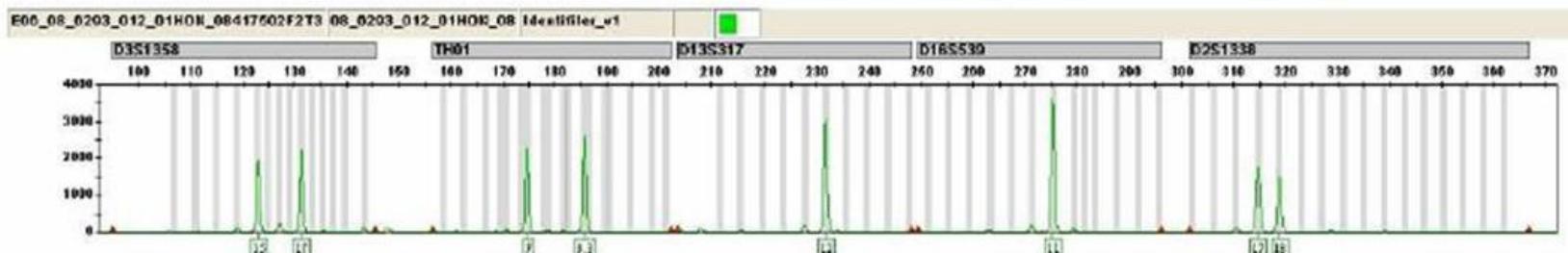
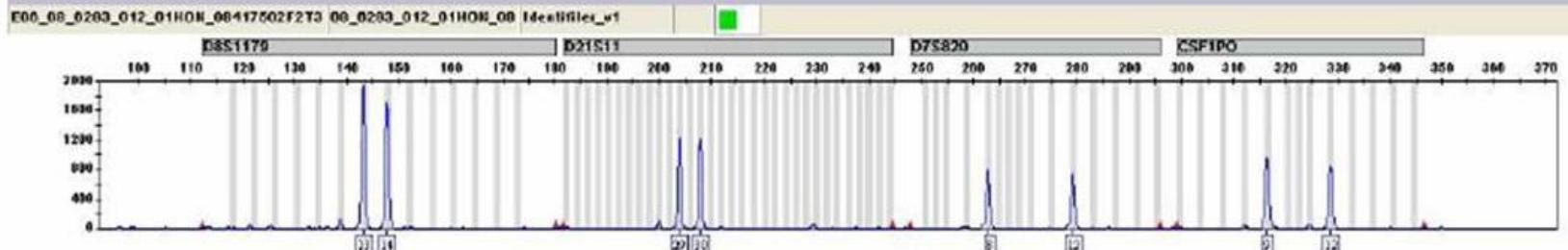
(A) Három lokusz szimultán amplifikálása egy reakcióban

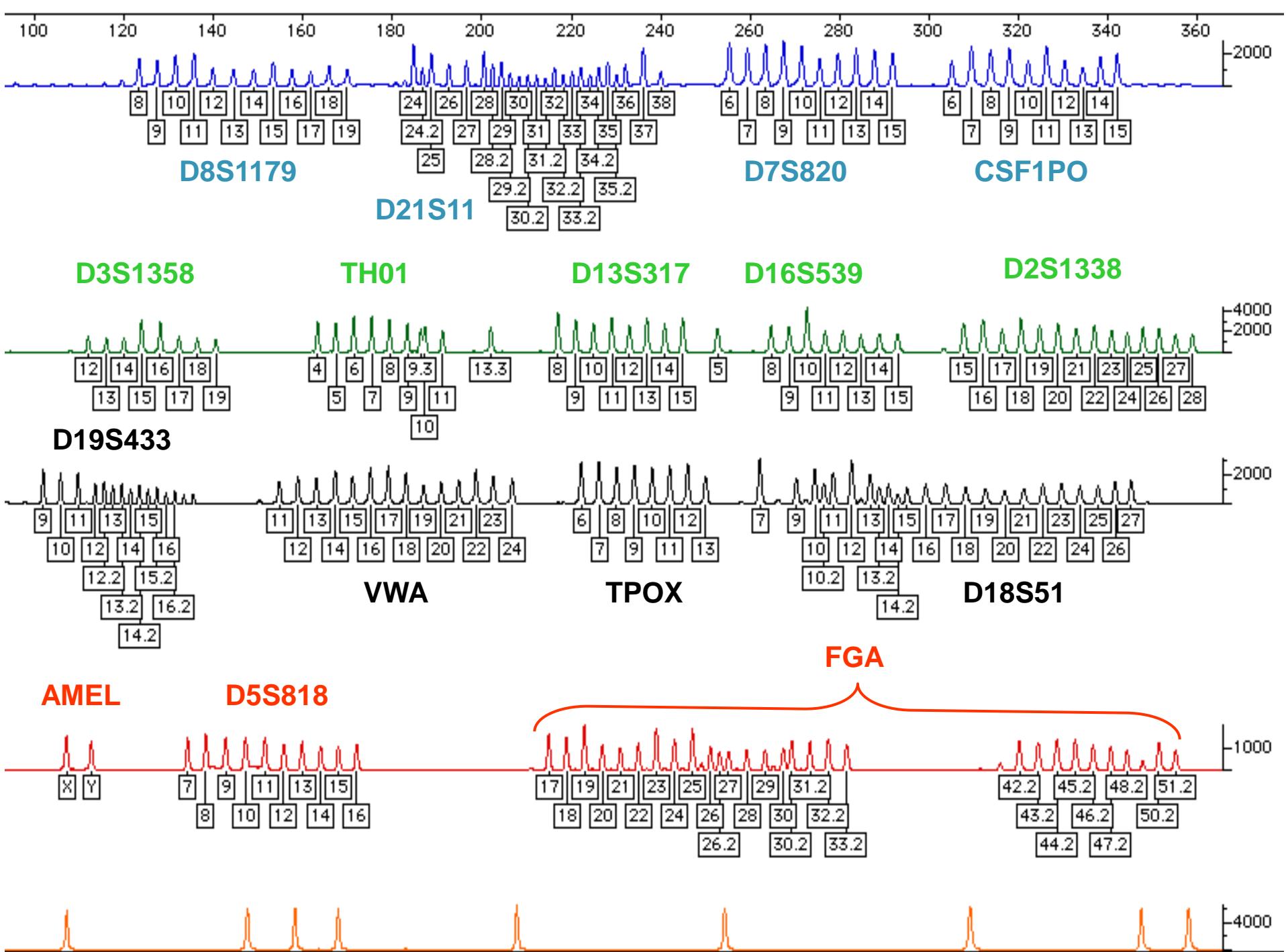


(B) PCR termékek elválasztása fragmens méret alapján



15 autószmás lókusz ...

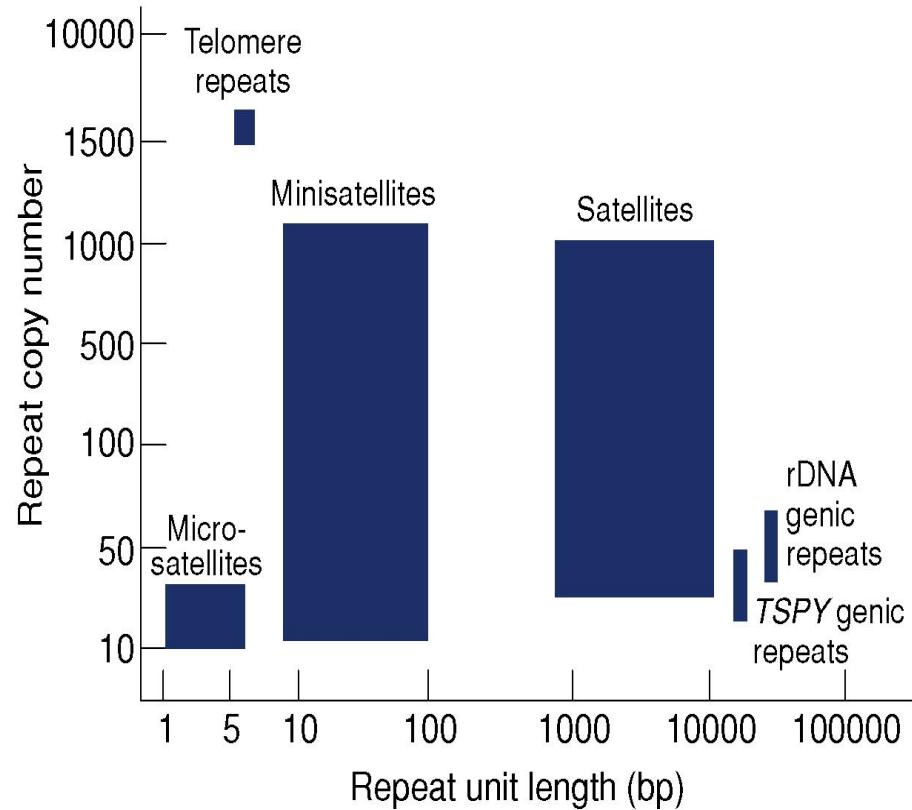
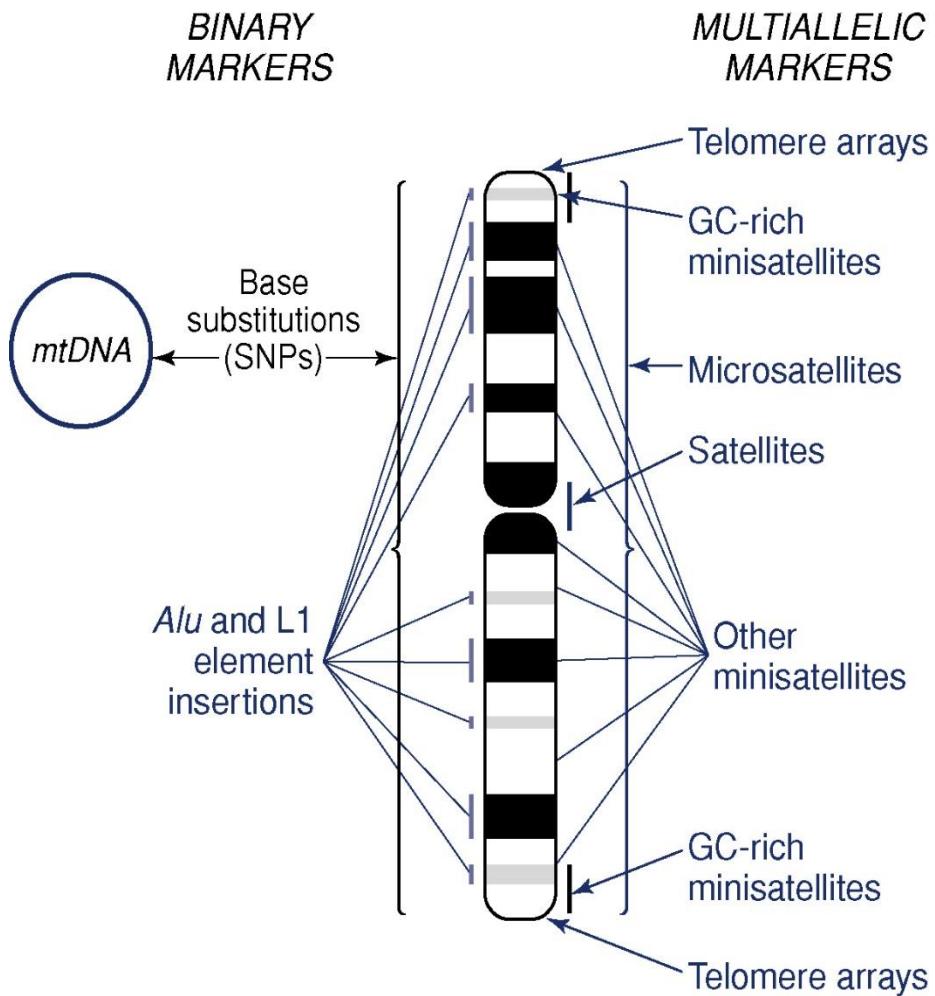




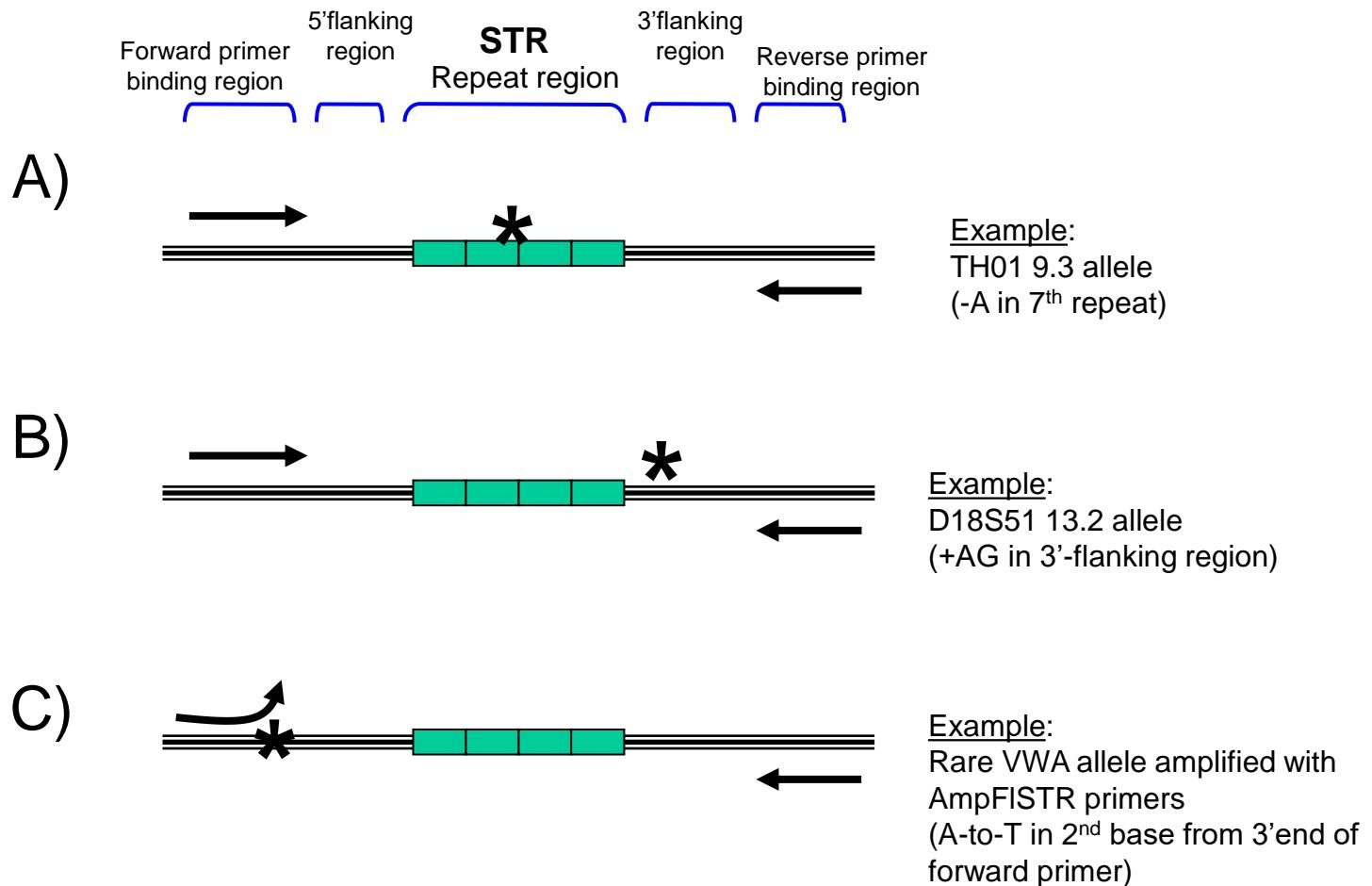
STR System	Maternal Meioses (%)	Paternal Meioses (%)	Number from either	Total Number of Mutations	Mutation Rate
CSF1PO	95/304,307 (0.03)	982/643,118 (0.15)	410	1,487/947,425	0.16%
FGA	205/408,230 (0.05)	2,210/692,776 (0.32)	710	3,125/1,101,006	0.28%
TH01	31/327,172 (0.009)	41/452,382 (0.009)	28	100/779,554	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)	881/621,118 (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 lókuszok mutációs rátája: 10^{-3} - 10^{-4} / meiózis

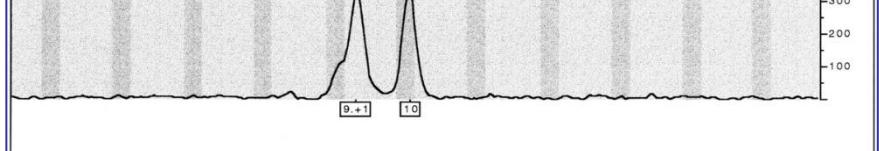
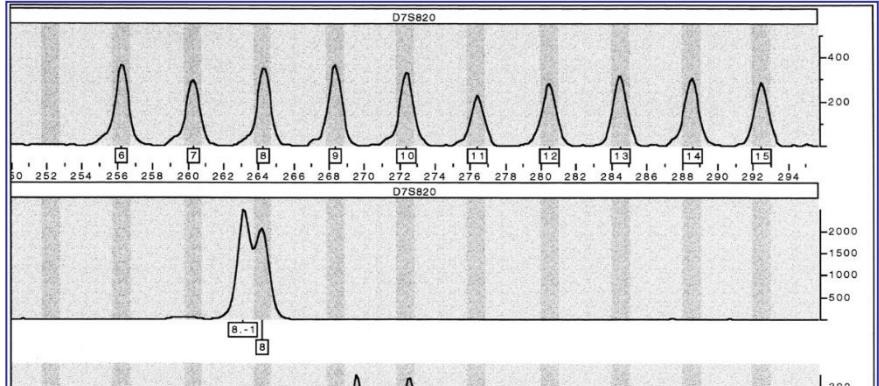
Polimorf markerek eloszlása a genomban



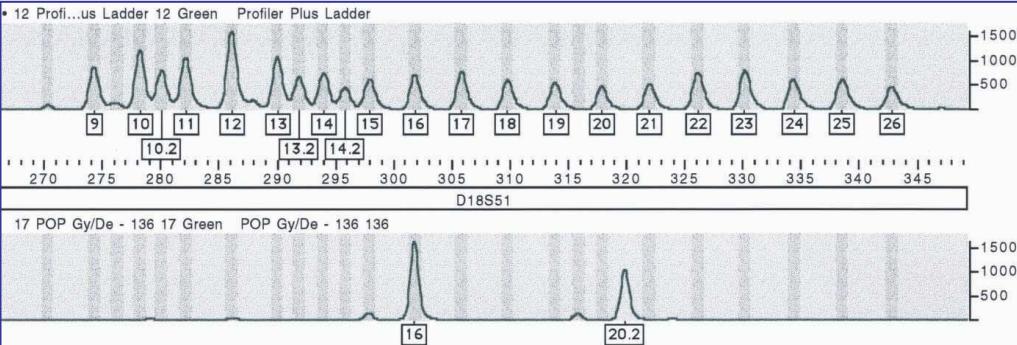
Mikroszatellita pontmutációk



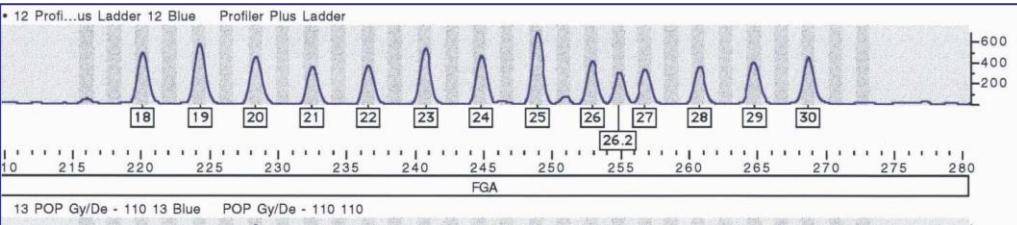
Variáns mikroszatellita allélek: Null-allélek



Allél	Fragmens méret	5' Flanking régió	Repeat regió	3' Flanking régió
8.-1	205 bp	[redacted]	- (GATA) ₈ -	- (T) ₈ ATCT -
9.+1	211 bp	[redacted]	- (GATA) ₉ -	- (T) ₁₀ ATCT -
10	214 bp	[redacted]	- (GATA) ₁₀ -	- (T) ₈ AATCT -
12 (ref.)	222 bp	[redacted]	- (GATA) ₁₂ -	- (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]



Homozigóta fenotípus!

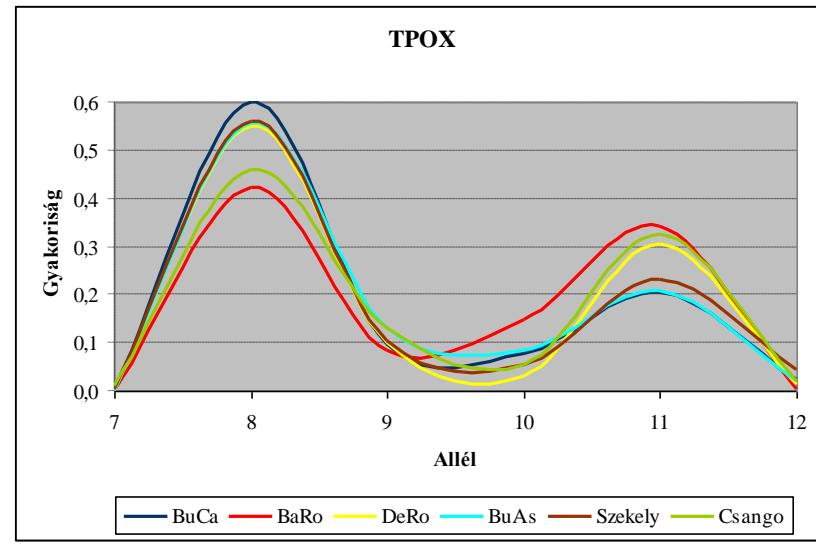
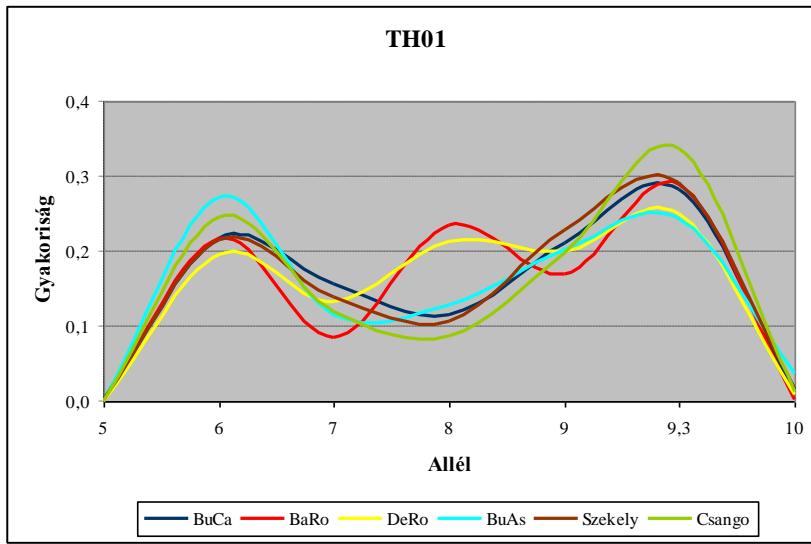
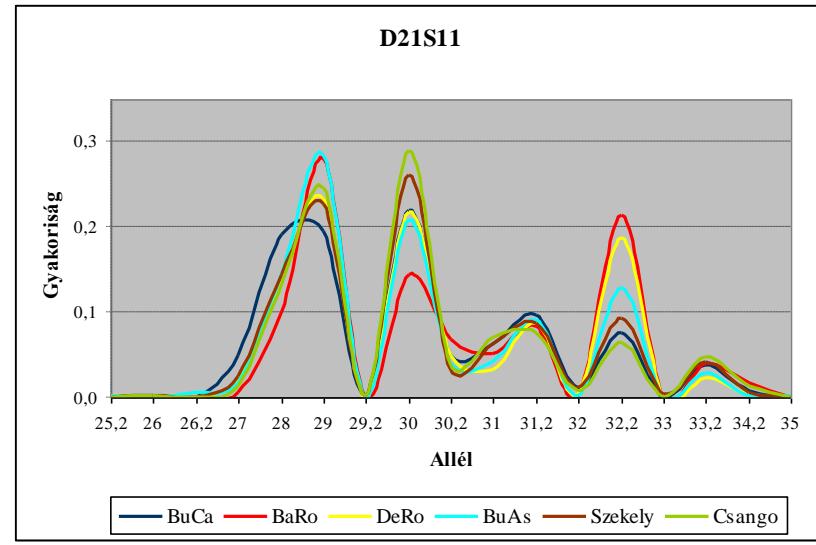
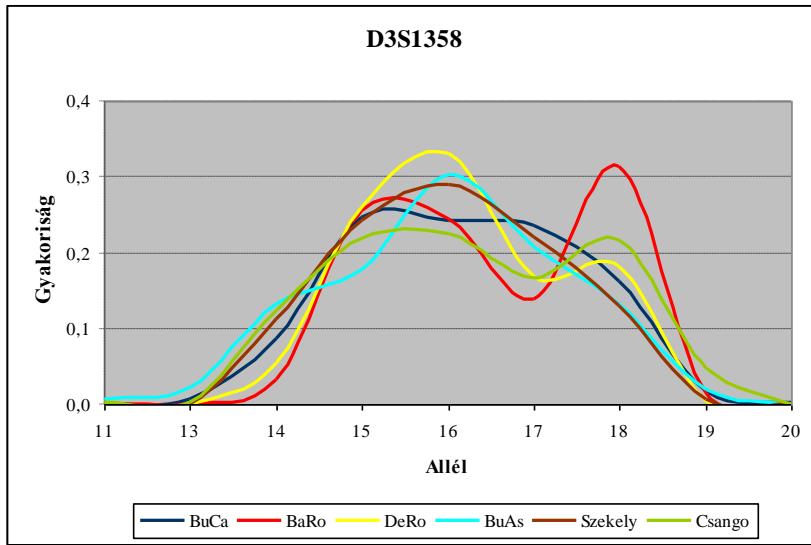
D13S317 allele 12

5'-gggttgctggacatggtatcACAGAAGTCTGGGATGTGGA--N82---(ATCT)₁₂ATCAATC(ATCT)₃TTCTGTCTGTCTTTTGGGC--N36---
gaccaacaattcaagctctc-3'

D13S317 allele 7 (variant)

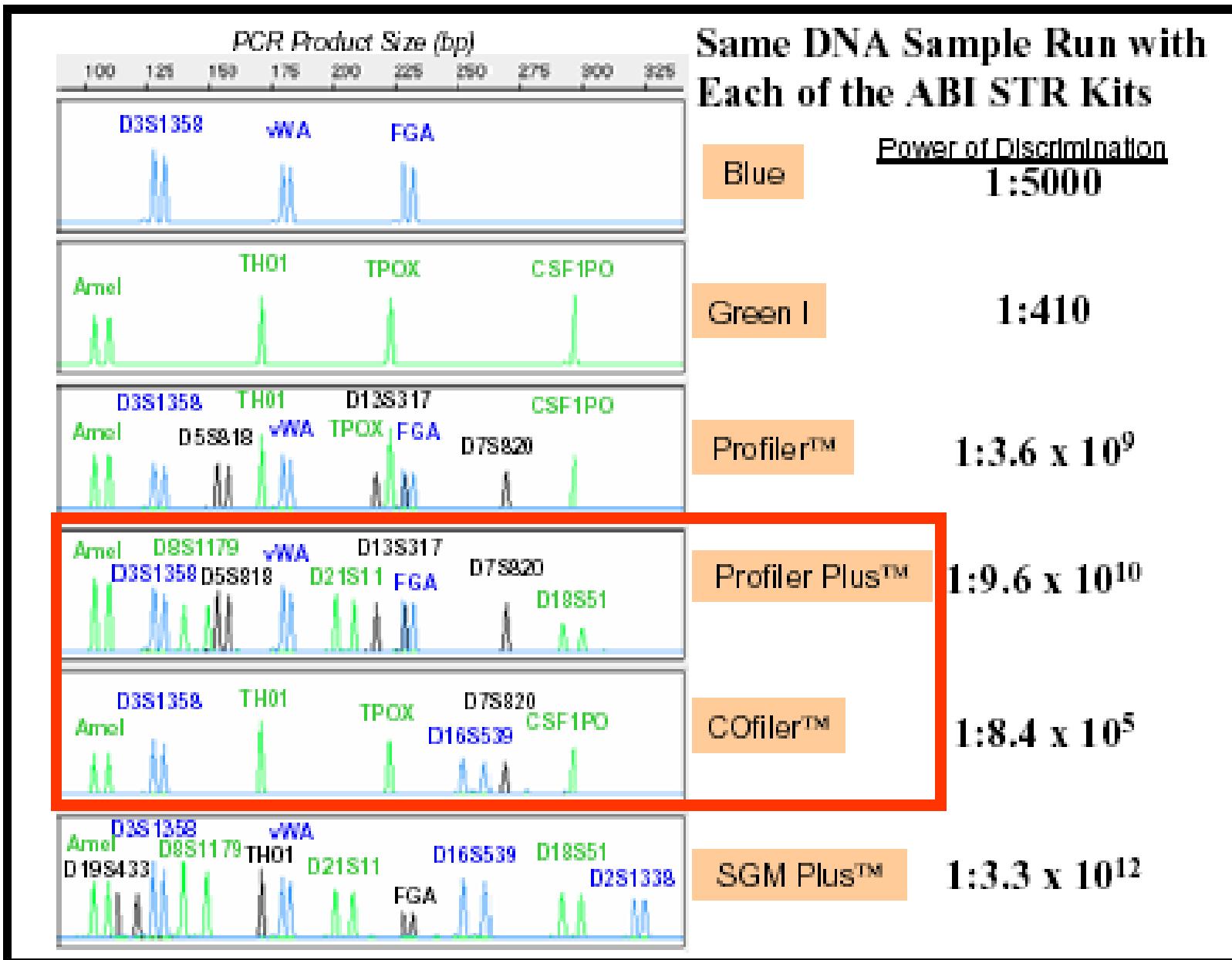
5'-gggttgctggacatggtatcACAGAAGTCTGGGATGTGGA--N82---(ATCT)₇**ATCAATCAATC(ATCT)₃TTCTGTCTTTTGGGC**--N36---
gaccaacaattcaagctctc-3'

Mikroszatellita allélgyakoriság diagramok



Genotípus gyakoriságok statisztikai interpretálása

- Allél-gyakorisági adatok generálása a vizsgált populációban gyűjtött minták sorozatából (általában 100-200 nem rokon egyed / populáció autoszóma STR lókuszok esetén)
- Allél-gyakorisági értékek meghatározása minden lókuszon és az összes megfigyelt allélra
- Allél-gyakorisági értékek felhasználásával a kérdéses DNS-profil előfordulási gyakoriságának kiszámolása
 - Homozigóta (p^2), Heterozigóta ($2pq$)
 - Product rule: független lókuszok genotípus gyakoriságának multiplikálhatósága $PM = (P_1)(P_2) \dots (P_n)$



Fajtaazonosítás?

STRUCTURE analízis

1. Ősi



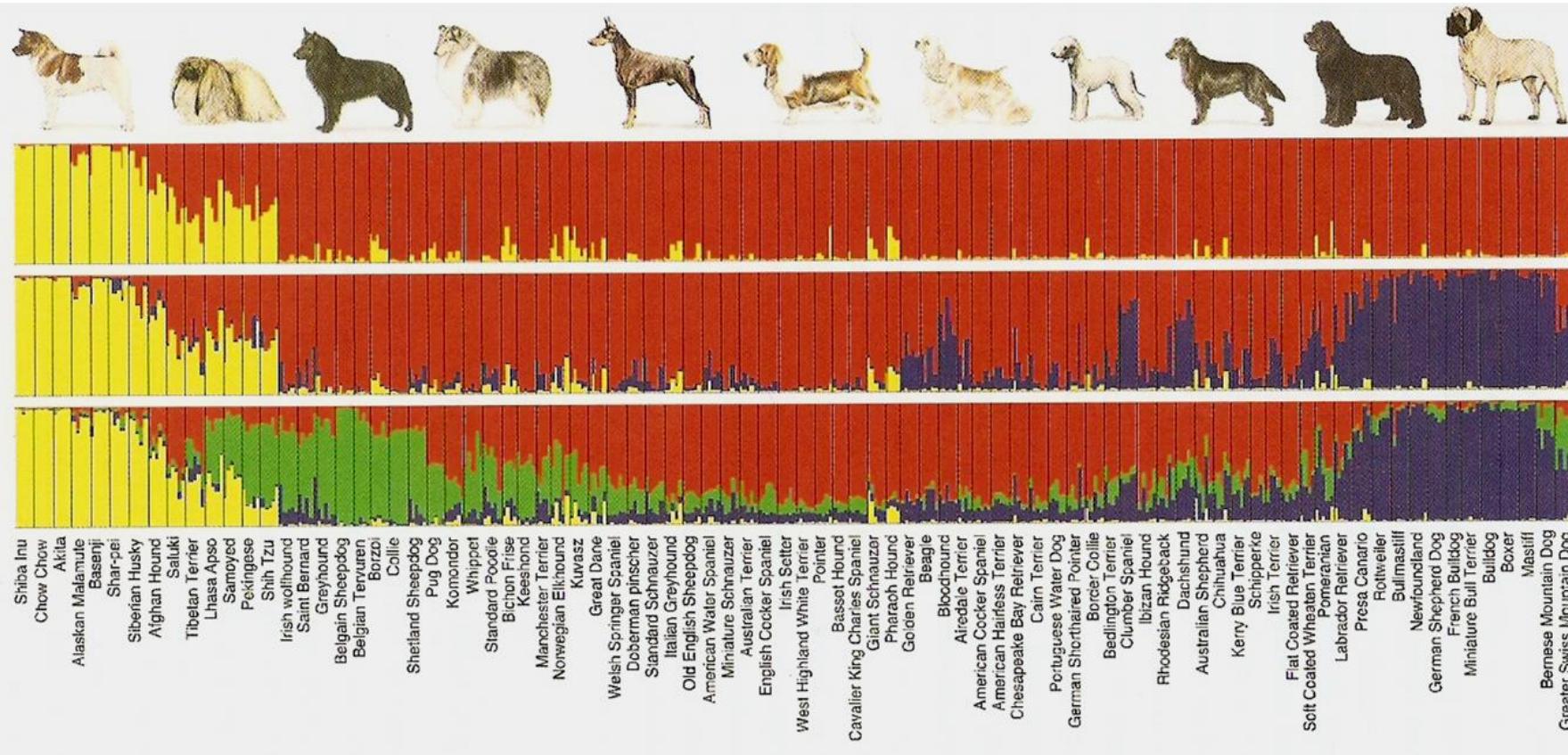
2. Terelő



3. Vadász



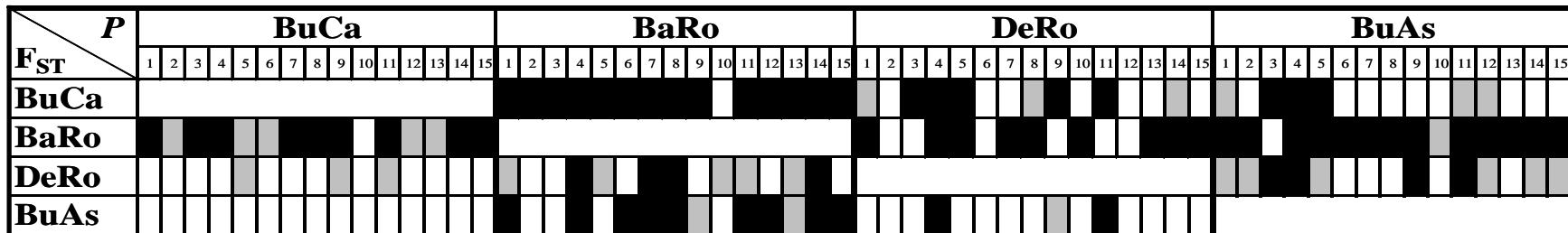
4. Védő



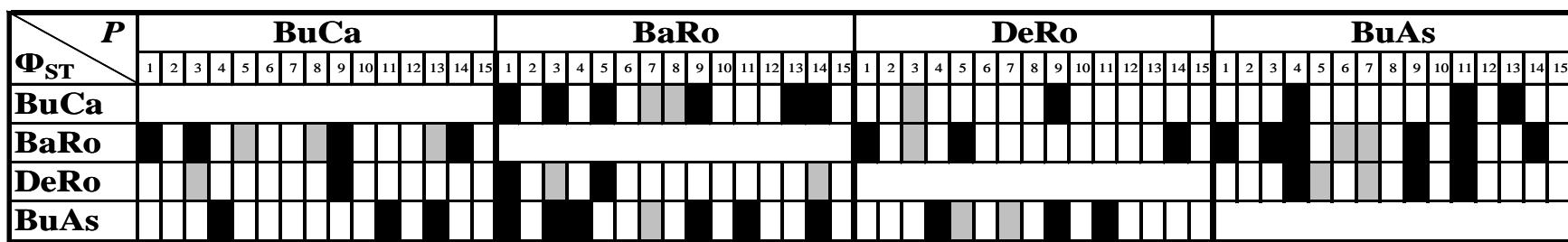
- 85 fajta ($n=414$)
- 95 mikroszatellita marker (dinukleotid repeat)
- 99%-ban helyes besorolás

Genetic Structure - Analysing of MOlaculare VAriance (AMOVA)

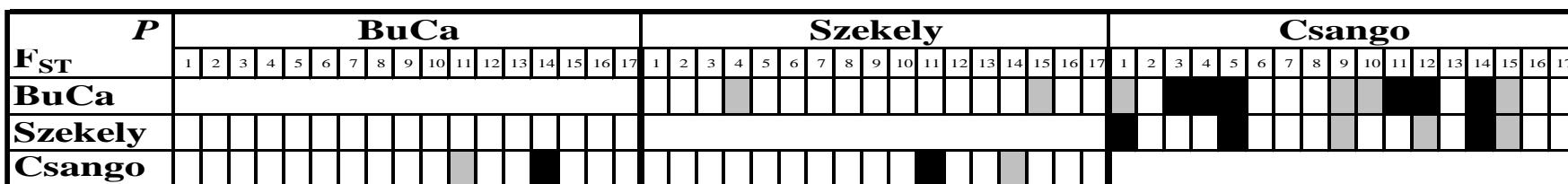
a,



b,



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



b,



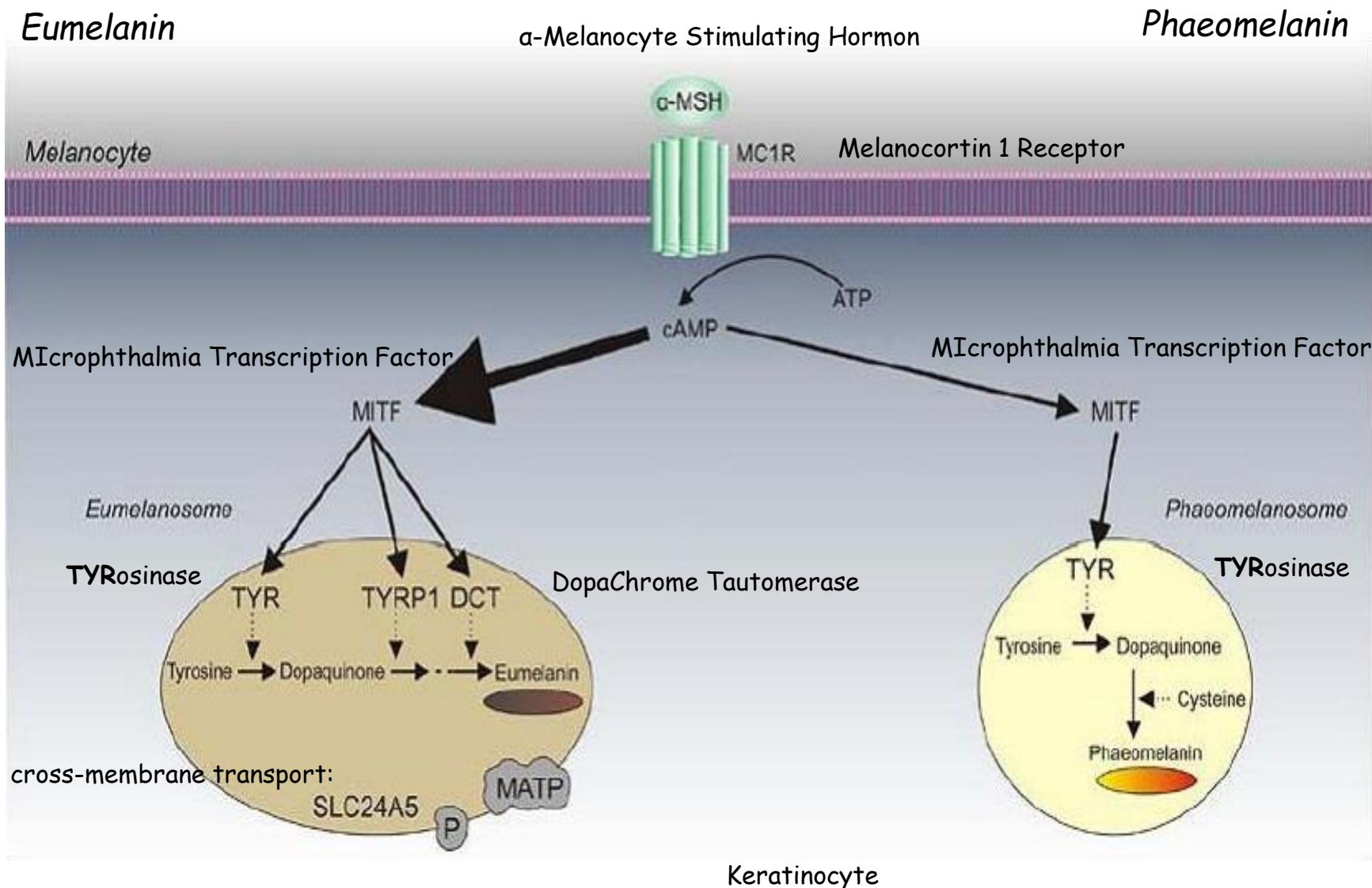
Jelölések: ■ ($F_{ST}, \Phi_{ST} > 0,02; P < 0,1$) ■ 0,02 > ($F_{ST}, \Phi_{ST} > 0,01; 0,01 < P < 0,05$) □ ($F_{ST}, \Phi_{ST} < 0,01; P > 0,05$)

Autosome SNPs in the Human Genome

TABLE 12.2 Categories of SNP Markers (See Budowle & van Daal 2008, Butler et al. 2008).

Category	Characteristics	Examples
Identity SNPs Individual Identification SNPs (IISNPs)	SNPs that collectively give very low probabilities of two individuals having the same multi-locus genotype	FSS 21plex (Dixon et al. 2005) SNPforID 52plex (Sanchez et al. 2006) Kidd group SNPs (Pakstis et al. 2010)
Lineage SNPs Lineage Informative SNPs (LISNPs)	Sets of tightly linked SNPs that function as multi-allelic markers that can serve to identify relatives with higher probabilities than simple bi-allelic SNPs	mtDNA coding region SNPs (Coble et al. 2004) Japanese Y-SNPs (Mizuno et al. 2010) Haplotype blocks (Ge et al. 2010)
Ancestry SNPs Ancestry Informative SNPs (AISNPs)	SNPs that collectively give a high probability of an individual's ancestry being from one part of the world or being derived from two or more areas of the world	SNPforID 34plex (Phillips et al. 2007b) 24 SNPs (Lao et al. 2010) FSS YSNPs (Wetton et al. 2005)
Phenotype SNPs Phenotype Informative SNPs (PISNPs)	SNPs that provide a high probability that the individual has particular phenotypes, such as a particular skin color, hair color, eye color, etc.	Red hair (Grimes et al. 2001) "Golden" gene pigmentation (Lamason et al. 2005) IrisPlex eye color (Walsh et al. 2010)

A humán melanogenezis



A bőrpigmentáció kialakulásáért felelős gének

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 gén mutációk

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 allélváltozatok különböző aktivitással rendelkeznek.
- 317 AS, 7 transzmembrán domén,
- SNPs: RHC fenotípus - neandervölgyi pigmentáció
- genetikai tesztek, fenotípus predikció

SNPs - pigmentációs gének

- ASIP (aguti): 3'UTR 8818A - MSH antagonista - phaeomelanin termelés
- MATP: melanoszóma pH reguláció, 374Leu allél - sötét szín, albinizmus
- SLC24A5: „arany” gén, zebrafish, Ala111Thr allél, világos árnyalat, europid rasszban fixált, szelekciós nyomás?
- OCA2: albinizmus gén, 305 Arg/Trp, Afrika / Európa

Gene	Location	Protein	Reference SNP ID (rs#) ^a	Alleles	Variation type
<i>MC1R</i>	16q24.3	MC1R: melanocortin 1 receptor	rs1805007	C/T	ns coding, c.451C>T, p.R151C
			rs1805008	C/T	ns coding, c.478C>T, p.R160W
<i>HERC2</i>	15q13	Unknown	rs12913832	A/G	Non-coding, intron 86
<i>OCA2</i>	15q11.2-15q12	P-protein: NA+/H+ antiporter or glutamate transporter	rs7495174 rs6497268 or rs4778241 rs11855019 or rs4778138 rs1545397	T/C G/T T/C G/A	Non-coding, intron 1 Non-coding intronic
<i>SLC45A2</i>	5p13.3	MATP: membrane-associated transporter protein	rs16891982	C/G	ns coding, c.1122C>G, p.F374L
<i>SLC24A5</i>	15q21.1	SLC24A5 (or NCKX5): solute carrier family 24, member 5; potassium-dependent sodium- calcium ion exchanger	rs1426654	G/A	ns coding, p.A111T
<i>DCT</i>	13q32	DCT or TYRP2/TRP-2: dopachrome tautomerase or tyrosinase-related protein-2	rs2031526	G/A	Non-coding, intronic

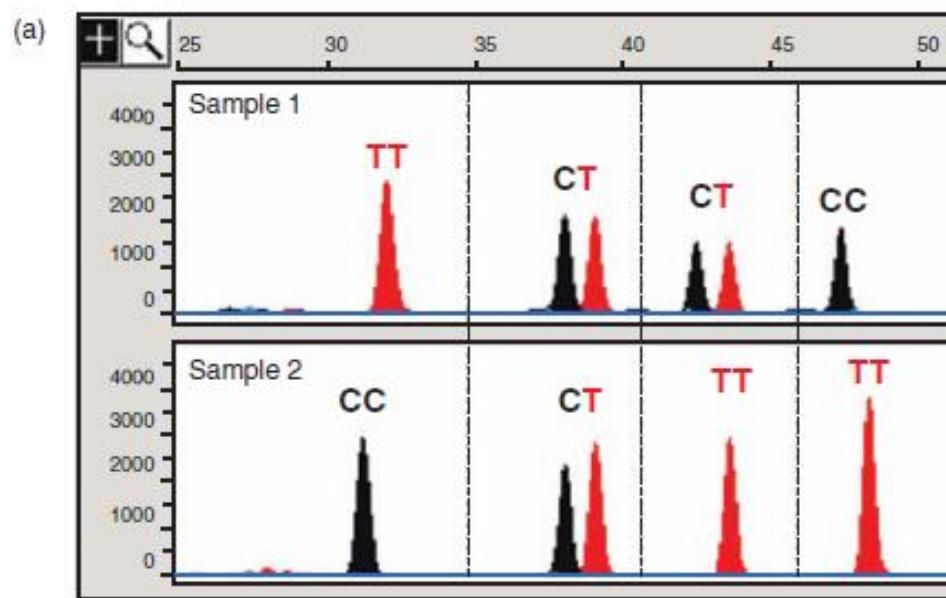
ns non-synonymous

^b Reference SNP ID refer to the reference sequence identifier given to the SNP in the dbSNP database

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.

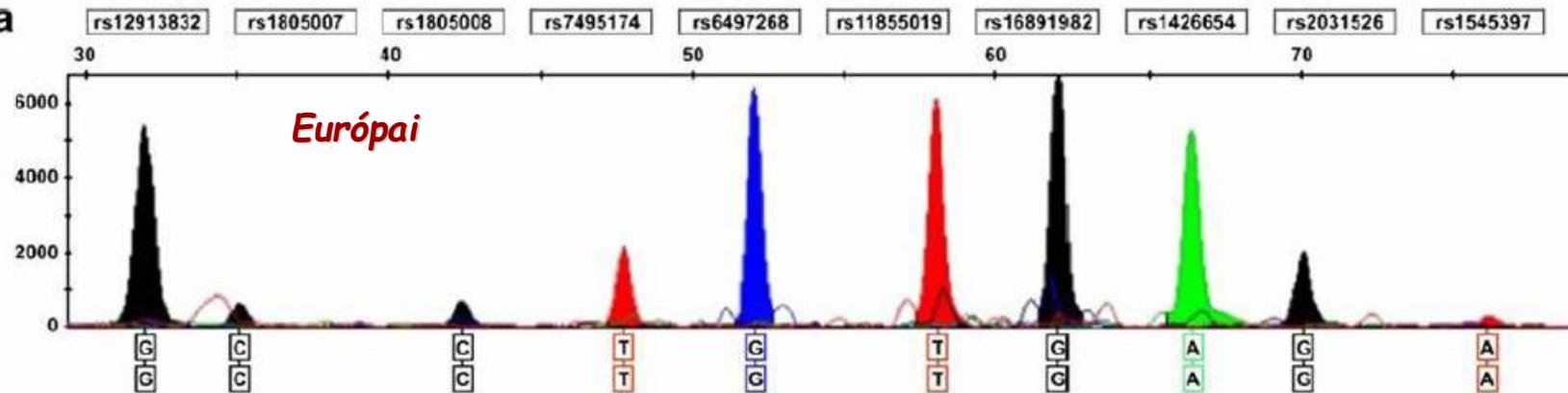


- (b) (TTTTT)-primer1 (chromosome 20)-ddT/ddT
(TTTTT)-(TTTTT)-primer2 (chromosome 6)-ddC/ddT
(TTTTT)-(TTTTT)-(TTTTT)-primer3 (chromosome 14)-ddC/ddT
(TTTTT)-(TTTTT)-(TTTTT)-(TTTTT)-primer4 (chromosome 1)-ddC/ddC

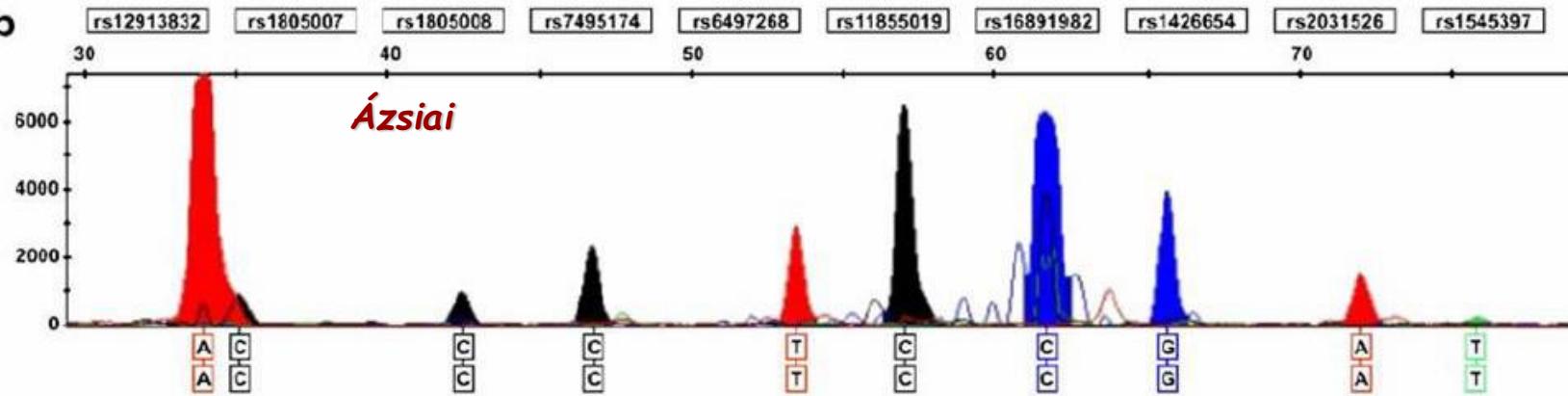
FIGURE 12.2 Allele-specific primer extension results using four autosomal SNP markers on two different samples (a). SNP loci are from separate chromosomes (1, 6, 14, and 20) and therefore unlinked. Electrophoretic resolution of the SNP primer extension products occurs due to poly(T) tails that are 5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

10 pigmentációs gén SNP genotipizálás (SNaPshot)

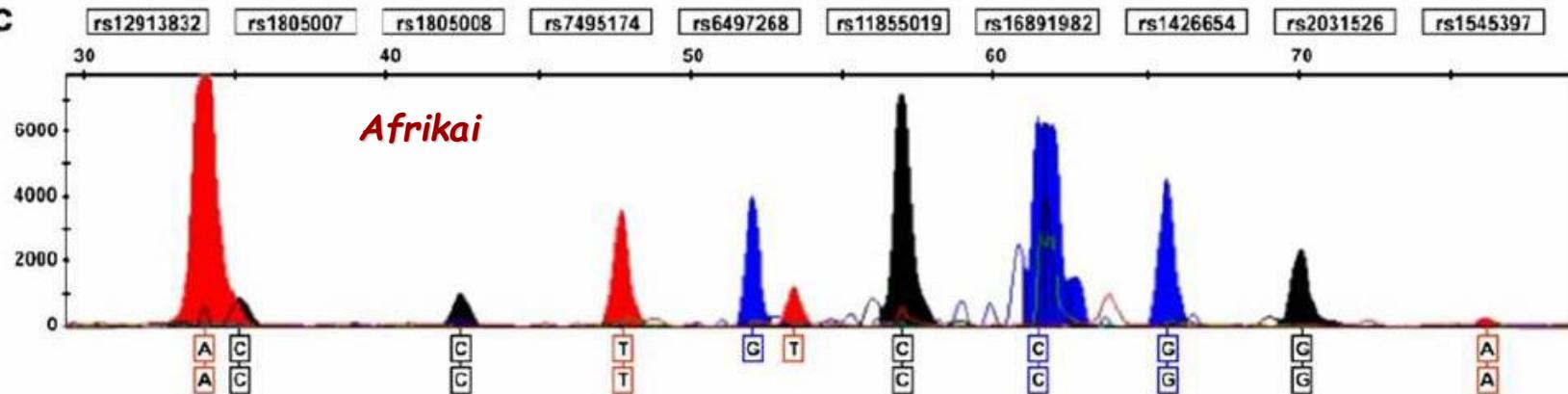
a



b



c



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
E1	Blue	Red	Fair	<u>G/G</u>	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	<u>G/G</u>	C/C	C/C	TGT/TGT	G/G	A/A	A/G	A/A	0.954	0.021	0.025
E3	Blue	Blond	Fair	<u>G/G</u>	C/C	C/C	TGT/TGT	G/G	A/A	A/G	A/A	0.954	0.024	0.022
E4	Blue	Blond	Fair	<u>G/G</u>	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	<u>G/G</u>	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	<u>G/G</u>	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	<u>G/G</u>	C/C	C/C	TGT/TGT	C/G	A/A	G/G	A/A	0.789	0.049	0.163
E11	Green	Auburn	Fair	<u>G/G</u>	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	<u>G/G</u>	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	<u>G/G</u>	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	<u>G/G</u>	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	<u>G/G</u>	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	<u>G/G</u>	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	<u>G/G</u>	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	<u>G/G</u>	C/C	C/C	TGT/TGT	G/G	A/A	A/G	A/A	0.954	0.021	0.025
E27	Blue	Red	Fair	<u>G/G</u>	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

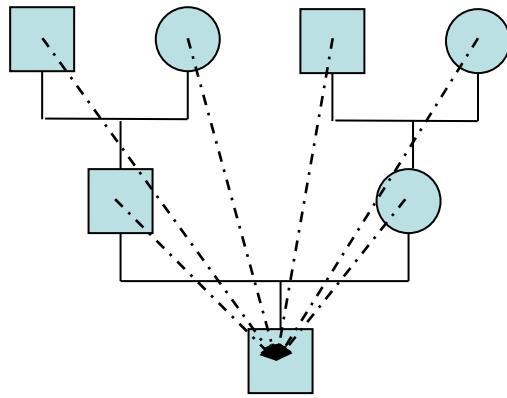
E European modern sample, Af African modern sample, As Asian modern sample

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

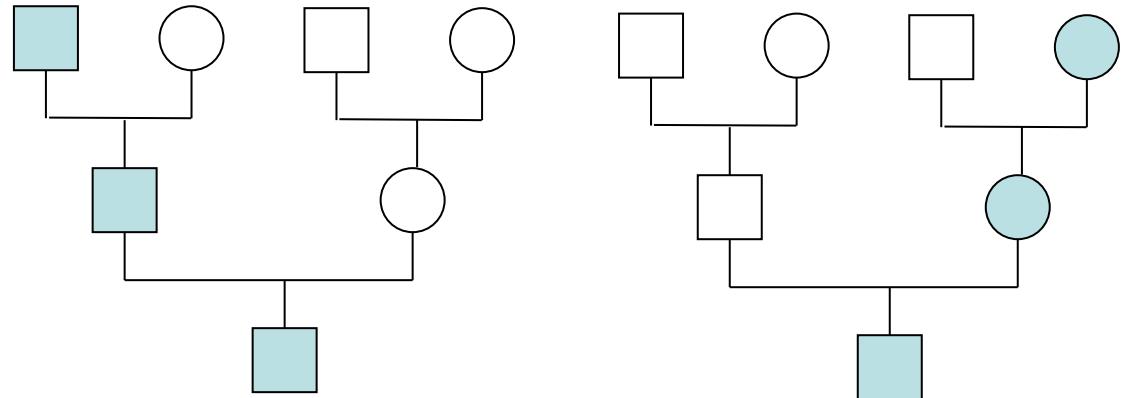
^b Probability of being from European/Asian/African population determined using the STRUCTURE program. The greatest probability, most likely estimate of ancestry, is indicated in bold

Eltérő öröklődési mintázatok

Leszármazási markerek



Autoszómás marker
(az összes felmenőnktől
örököljük részletekben)



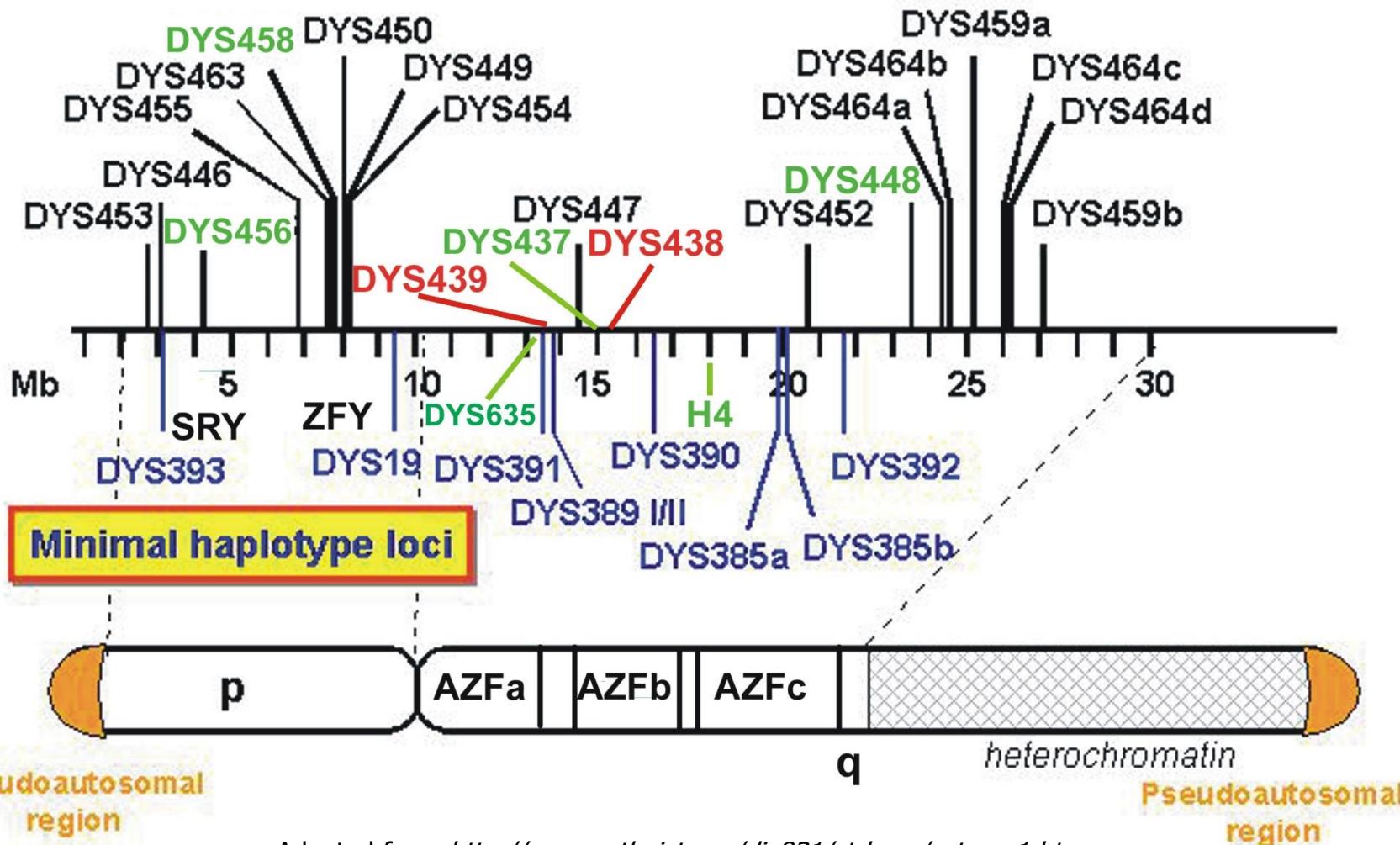
Y-kromoszóma
(csak a fiú gyermekek
öröklik apjuktól)

Mitokondriális DNS
(minden gyermek az
édesanyjától örökli)

Y STR Positions along Y Chromosome

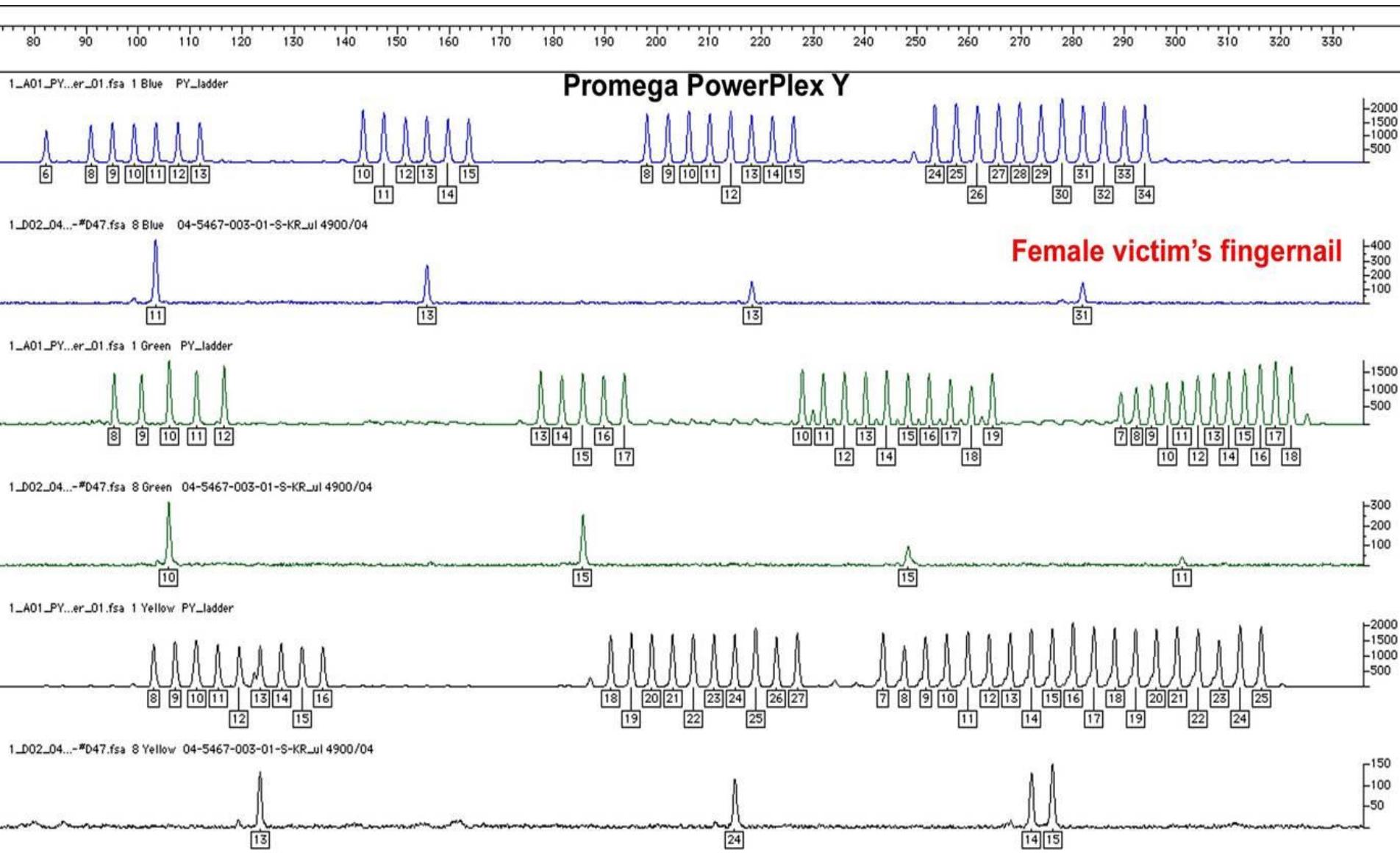
Extended haplotype loci

ABI AmpF/STR Yfiler loci

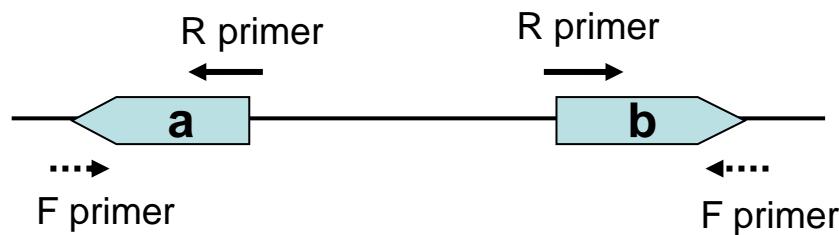


Adapted from <http://www.cstl.nist.gov/div831/strbase/ystrpos1.htm>

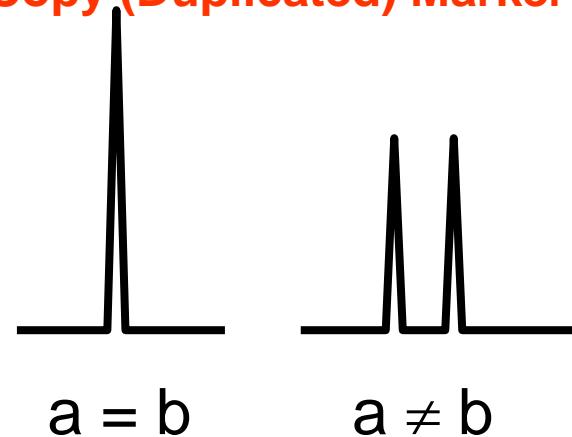
Y kromoszómás vizsgálat bűnügyi mintákon



(A) DYS385 a/b

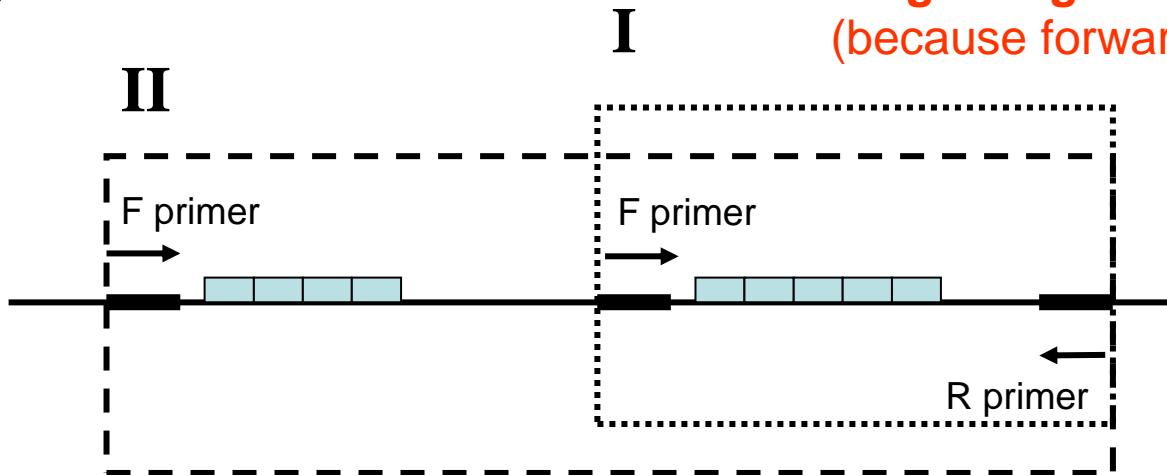


Multi-Copy (Duplicated) Marker

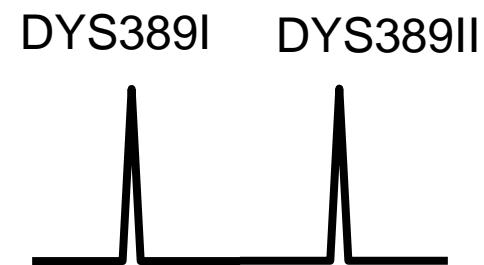


Duplicated regions are
40,775 bp apart and facing
away from each other

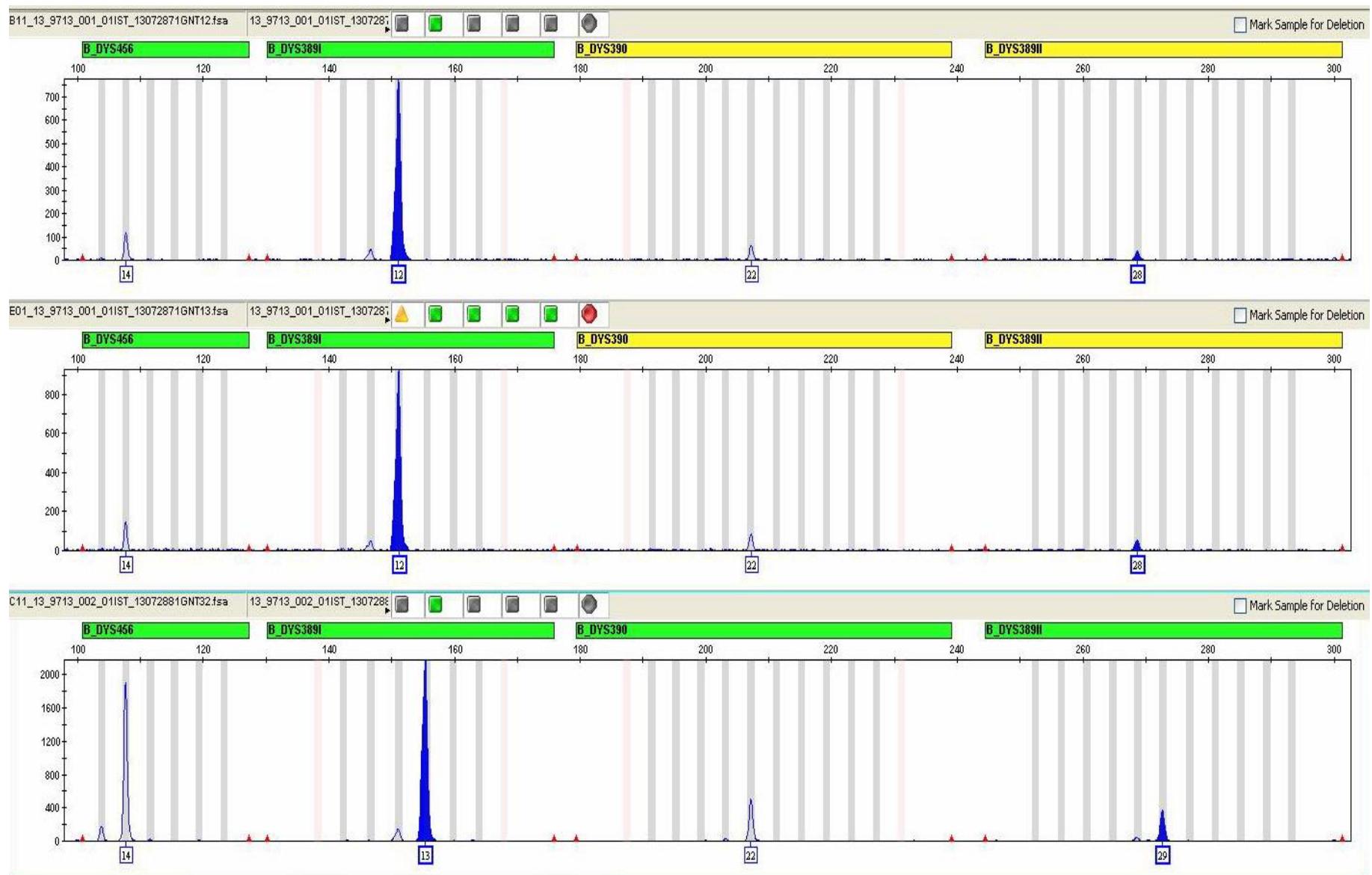
(B) DYS389 I/III



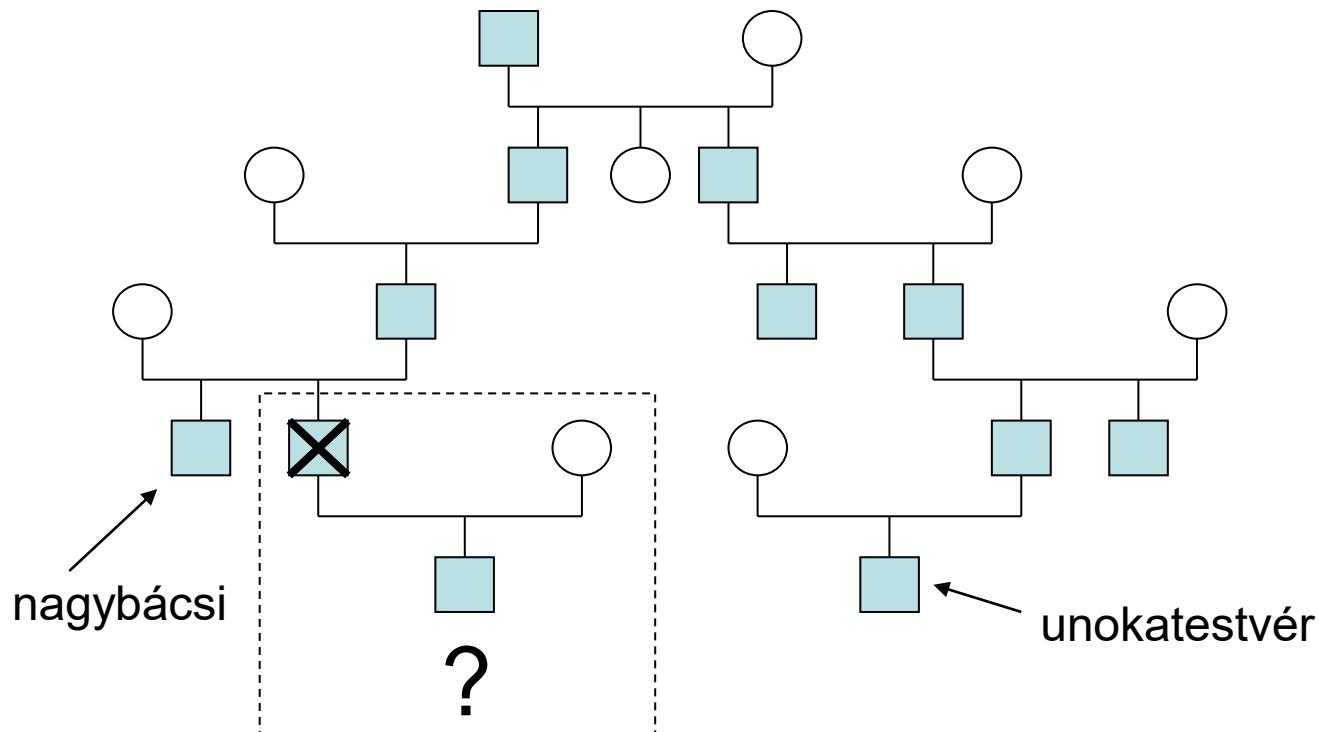
Single Region but Two PCR Products
(because forward primers bind twice)



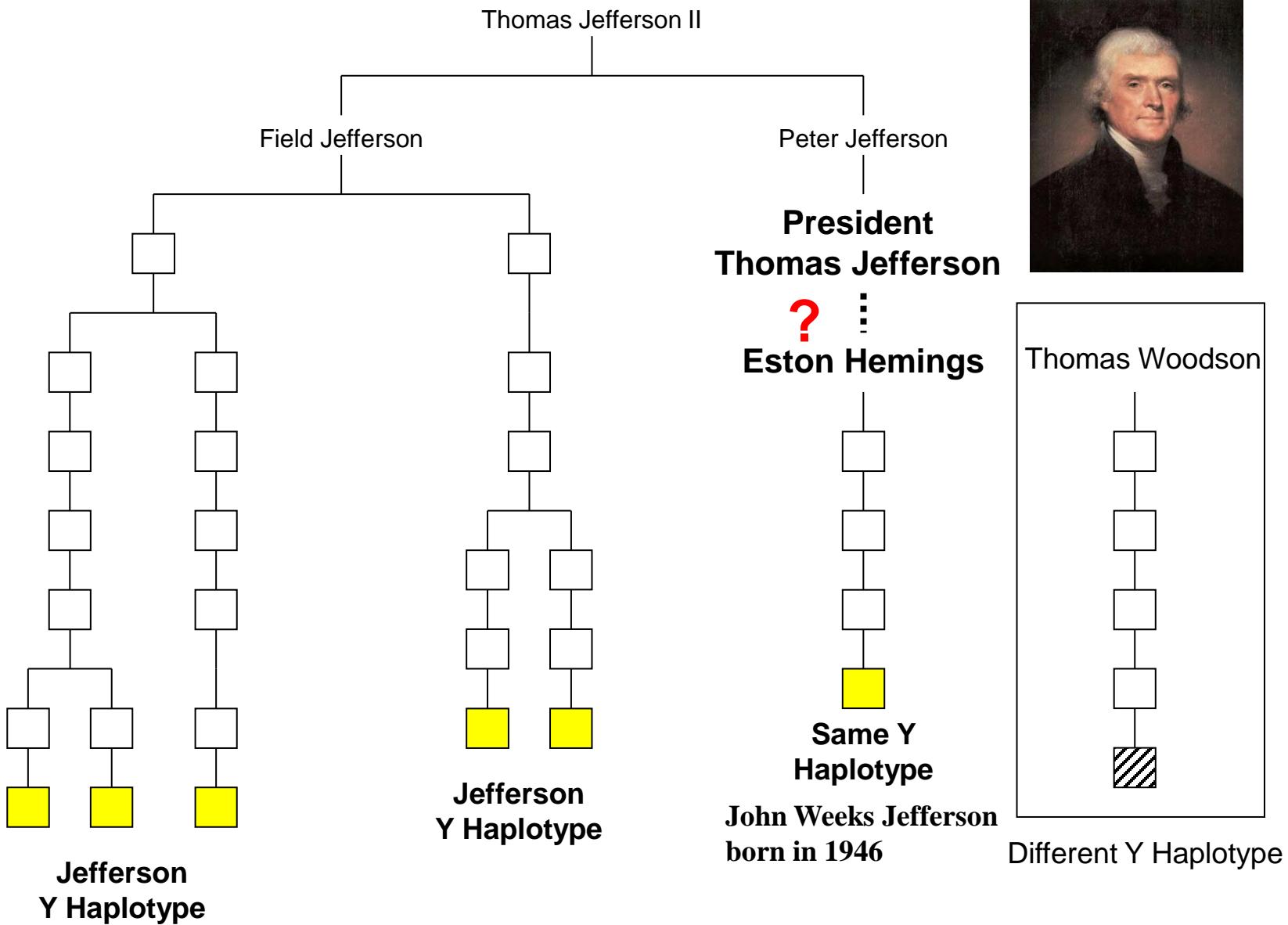
DYS389 I-II mutation



Y kromoszóma vizsgálat során a férfi ági fiú rokonok mindegyike potenciális mintaadó



Genetic History



DNA Marker Tested	Field Jefferson Male-Line	Eston Hemings Male-Line	John Carr Male-Line	Thomas Woodson Male-Line
Number of individuals typed	5	1	3	5
Y STR Loci				
DYS19	15	15	14	14
DYS388	12	12	12	12
DYS389A	4	4	5	5
DYS389B	11	11	12	11
DYS389C	3	3	3	3
DYS389D	9	9	10	10
DYS390	11	11	11	11
DYS391	10	10	10	13
DYS392	15	15	13	13
DYS393	13	13	13	13
DXYS156Y	7	7	7	7
Y SNP Loci	(0 = ancestral state; 1 = derived state)			
DYS287 (YAP)	0	0	0	0
SRYm8299	0	0	0	0
DYS271 (SY81)	0	0	0	0
LLY22g	0	0	0	0
Tat	0	0	0	0
92R7	0	0	1	1
SRYm1532	1	1	1	1
Minisatellite Locus				
MSY1	(3)-5 (1)-14 (3)-32 (4)-16	(3)-5 (1)-14 (3)-32 (4)-16	(1)-17 (3)-36 (4)-21	(1)-16 (3)-27 (4)-21

Table 9.8, J.M. Butler (2005) *Forensic DNA Typing*, 2nd Edition © 2005 Elsevier Academic Press



Search

Haplotypes

SNPs

Populations

Contributors

Contributions

Analyse

Research

Contribute

Meet

DYS19	DYS389I	DYS389II	DYS390	DYS391	DYS392	DYS393	DYS385
16	13	31	25	12	11	13	14.16
<input checked="" type="checkbox"/>							
DYS438	DYS439	DYS437	DYS448	DYS456	DYS458	DYS635	YGATAH4
10	13	15	20	15	18	23	11
<input checked="" type="checkbox"/>							

[National database](#) | [Metapopulations](#) | [SNP](#)[Whole database](#)

Search

Please note: The database size will vary based on the loci you have entered.

- 7 loci haplotype (DYS19, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393): **101055 haplotypes**
- 9 loci haplotype (+ DYS385a/b): **99258 haplotypes**
- 11 loci haplotype (+ DYS438, DYS439): **72171 haplotypes**
- 12 loci haplotype (+ DYS437): **52628 haplotypes**
- 17 loci haplotype (+ DYS448, DYS456, DYS458, DYS635, YGATAH4): **40987 haplotypes**

Y-SNPs:

- 124 Y-SNP branches (defined by **134 Y-SNP markers**)
- **9039 haplotypes** with Y-SNP information



YHRD by Sascha Willuweit & Lutz Roewer is licensed under a Creative Commons Attribution-Noncommercial-Share Alike 3.0 Unported License.

Supported by



Endorsed by



**Haplotypes**

SNPs

Populations

Contributors

Contributions

Analyse

Research

Contribute

Meet

DYS19	DYS389I	DYS389II	DYS390	DYS391	DYS392	DYS393	DYS385
16	13	31	25	12	11	13	14.16
DYS438	DYS439	DYS437	DYS448	DYS456	DYS458	DYS635	YGATAH4
10	13	15	20	15	18	23	11

National database

Metapopulations

SNP

Whole database

Search

Matches grouped by Metapopulations

Matches grouped by Continents

Matches grouped by Haplogroups

Frequency surveying estimates

- » **All Metapopulation:** Found 0 of 40987 matching haplotypes [$f=0$ (95% CI: $0 - 9 \times 10^{-5}$)] in 0 of 282 populations.
- » **Eurasian Metapopulation:** Found 0 of 16733 matching haplotypes [$f=0$ (95% CI: $0 - 2.204 \times 10^{-4}$)] in 0 of 130 populations.
- » **East Asian Metapopulation:** Found 0 of 12674 matching haplotypes [$f=0$ (95% CI: $0 - 2.91 \times 10^{-4}$)] in 0 of 64 populations.
- » **Australian Aboriginal Metapopulation:** Found 0 of 766 matching haplotypes [$f=0$ (95% CI: $0 - 4.804 \times 10^{-3}$)] in 0 of 1 populations.
- » **African Metapopulation:** Found 0 of 1533 matching haplotypes [$f=0$ (95% CI: $0 - 2.403 \times 10^{-3}$)] in 0 of 10 populations.
- » **Native American Metapopulation:** Found 0 of 384 matching haplotypes [$f=0$ (95% CI: $0 - 9.56 \times 10^{-3}$)] in 0 of 9 populations.
- » **Eskimo Aleut Metapopulation:** Found 0 of 301 matching haplotypes [$f=0$ (95% CI: $0 - 1.218 \times 10^{-2}$)] in 0 of 2 populations.
- » **Afro-Asiatic Metapopulation:** Found 0 of 1854 matching haplotypes [$f=0$ (95% CI: $0 - 1.988 \times 10^{-3}$)] in 0 of 21 populations.
- » **Admixed Metapopulation:** Found 0 of 6742 matching haplotypes [$f=0$ (95% CI: $0 - 5.47 \times 10^{-4}$)] in 0 of 45 populations.

www.yhrd.org/Search/Haplotypes;j;js

R39: 101055 haplotypes

Contributors Contributions Analyse Research Contribute Meet

DYS19	DYS389I	DYS389II	DYS390	DYS391	DYS392	DYS393	DYS385	National database	Metapopulations	SNP
16	13	31	25	12	11	13	14.16	Whole database		
DYS438	DYS459	DYS437	DYS448	DYS456	DYS458	DYS635	YGATAH4			
10	13	15	20	15	18	23	11			

Search Reset

Matches grouped by Metapopulations Matches grouped by Continents Matches grouped by Haplogroups Frequency surveying estimates

African - Afro-American

Frequency estimates with given haplotype not included in the database: Mean: 3.366×10^{-4} , Mode: 2.843×10^{-4}

Frequency estimates with given haplotype included in the database: Mean: 3.889×10^{-4} , Mode: 3.366×10^{-4}

Afro-Asiatic - Semitic

Frequency estimates with given haplotype not included in the database: Mean: 4.267×10^{-4} , Mode: 4.064×10^{-4}

Frequency estimates with given haplotype included in the database: Mean: 4.47×10^{-4} , Mode: 4.267×10^{-4}

East Asian - Japanese

Frequency estimates with given haplotype not included in the database: Mean: 5.41×10^{-4} , Mode: 4.877×10^{-4}

Frequency estimates with given haplotype included in the database: Mean: 6.143×10^{-4} , Mode: 5.41×10^{-4}

East Asian - Korean

Frequency estimates with given haplotype not included in the database: Mean: 1.786×10^{-4} , Mode: 1.395×10^{-4}

Frequency estimates with given haplotype included in the database: Mean: 2.177×10^{-4} , Mode: 1.786×10^{-4}

East Asian - Sino-Tibetan - Chinese

Frequency estimates with given haplotype not included in the database: Mean: 6.028×10^{-5} , Mode: 3.951×10^{-5}

Frequency estimates with given haplotype included in the database: Mean: 8.104×10^{-5} , Mode: 6.028×10^{-5}

Eurasian - Altaic

Frequency estimates with given haplotype not included in the database: Mean: 5.634×10^{-4} , Mode: 4.953×10^{-4}

Frequency estimates with given haplotype included in the database: Mean: 6.315×10^{-4} , Mode: 5.634×10^{-4}

Eurasian - European - Eastern European

Frequency estimates with given haplotype not included in the database: Mean: 7.667×10^{-5} , Mode: 3.381×10^{-5}

Frequency estimates with given haplotype included in the database: Mean: 1.193×10^{-4} , Mode: 7.658×10^{-5}

Eurasian - European - South-Eastern European

Frequency estimates with given haplotype not included in the database: Mean: 3.772×10^{-4} , Mode: 2.878×10^{-4}

Frequency estimates with given haplotype included in the database: Mean: 4.867×10^{-4} , Mode: 3.772×10^{-4}

Eurasian - European - Western European

Frequency estimates with given haplotype not included in the database: Mean: 2.893×10^{-5} , Mode: 1.444×10^{-5}

Frequency estimates with given haplotype included in the database: Mean: 3.941×10^{-5} , Mode: 2.893×10^{-5}

Counting Method

95% confidence interval

$$1 - (0.05)^{1/N} \approx 3/N$$

Mobilis Genetikai Elemek

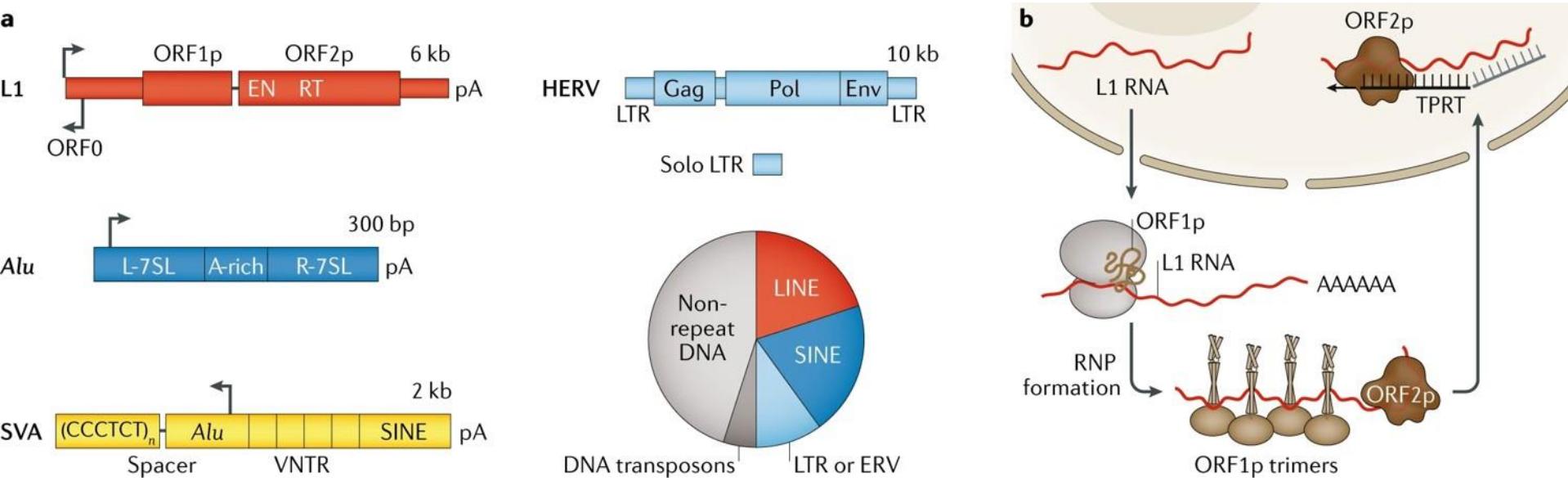
- Extra mennyisége a humán genomban: 100/gén (~ 45 %)
- Humán genom ~ 75 %-a lehet mobilis elem maradványa
- Mutagén hatások: pl. inzerció, nem homológ rekombináció,
→ negatív konzekvencia a gazdára nézve
- Retrotranszpozon: „copy-and-paste”, LINEs, SINEs, LTRs
- Transzpozon: „cut-and-paste”

TABLE 2.2: CLASSES OF DISPERSED REPEATS IN THE HUMAN GENOME.

Class	Copy no. per haploid genome	Fraction of genome	Autonomous transposition or retrotransposition?	Length
LINEs	850 000	21%	Yes	Up to 6–8 kb
SINEs	1 500 000	13%	No	Up to 100–300 bp
Retrovirus-like elements	450 000	8%	Complete copies, yes	6–11 kb (1.5–3 kb)
DNA transposon copies	300 000	3%	Complete copies, yes	2–3 kb (80–3000 bp)

Values given in parentheses are lengths of incomplete elements, incapable of autonomous transposition (see Section 3.4). Adapted from Lander *et al.* (2001).

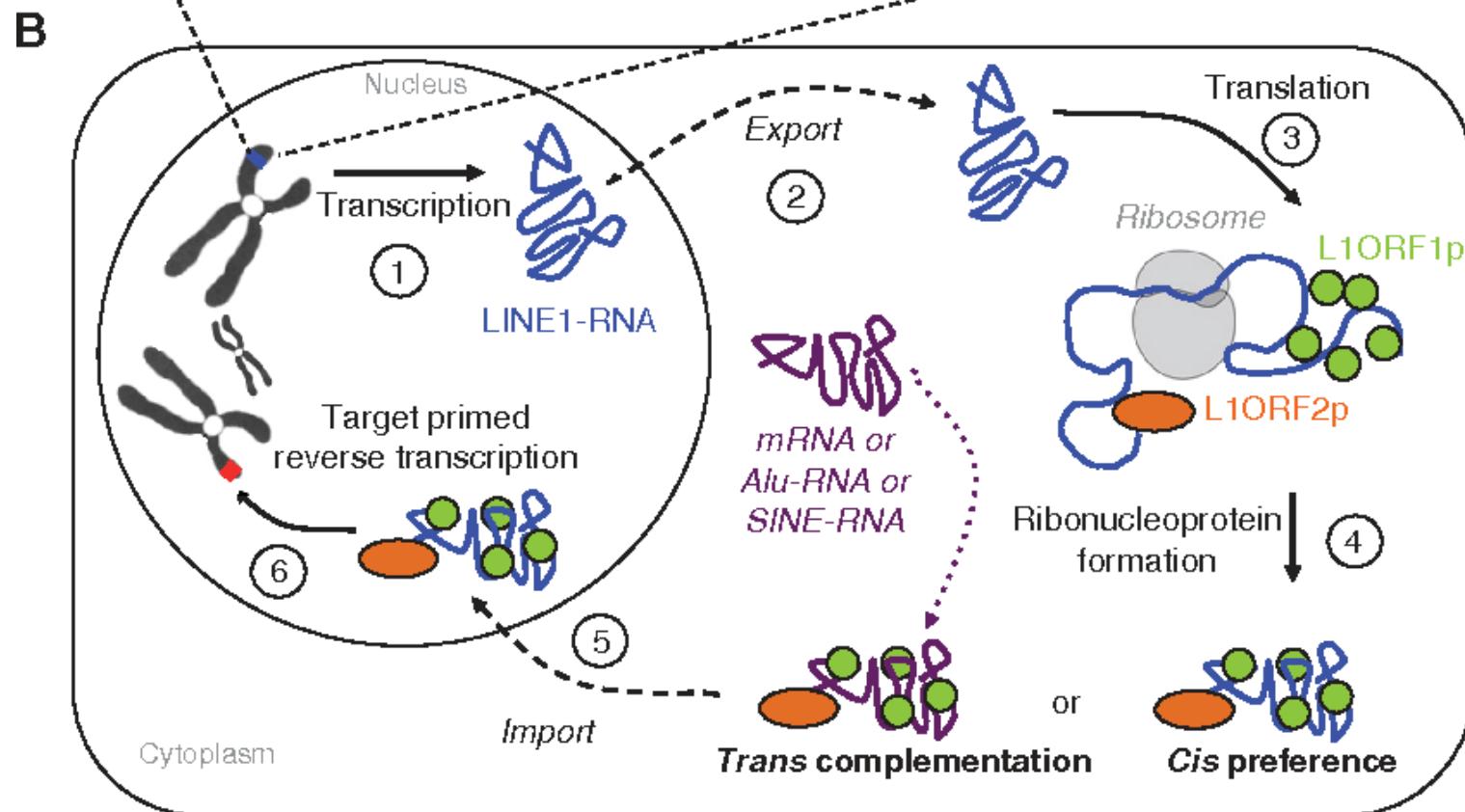
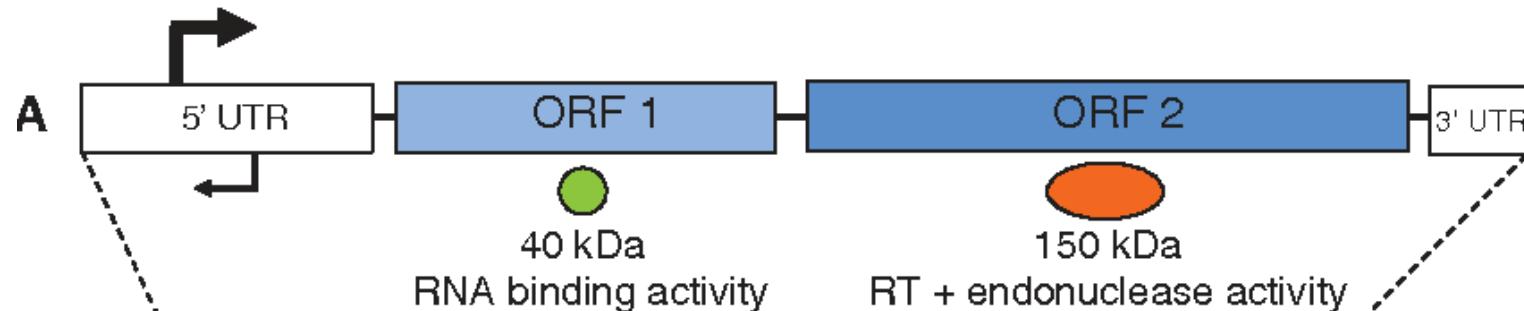
Structure of Transposable Elements



A schematic of common human transposable elements with their full-length size denoted. Long interspersed element 1 (LINE-1 or L1) encodes two open reading frames (ORFs). ORF2p protein has endonuclease (EN) and reverse transcriptase (RT) domains.

Alu elements are bipartite, with the two arms derived from 7SL RNA separated by an A-rich region. SVA is a composite element containing variable number tandem repeats (VNTRs). Human endogenous retroviruses (HERVs) are flanked by long terminal repeats (LTRs) and encode three essential viral proteins, including envelope (Env). ERVs also exist in the genome as solo LTRs.

The pie chart shows the proportion of the human genome made up of these repetitive sequences.

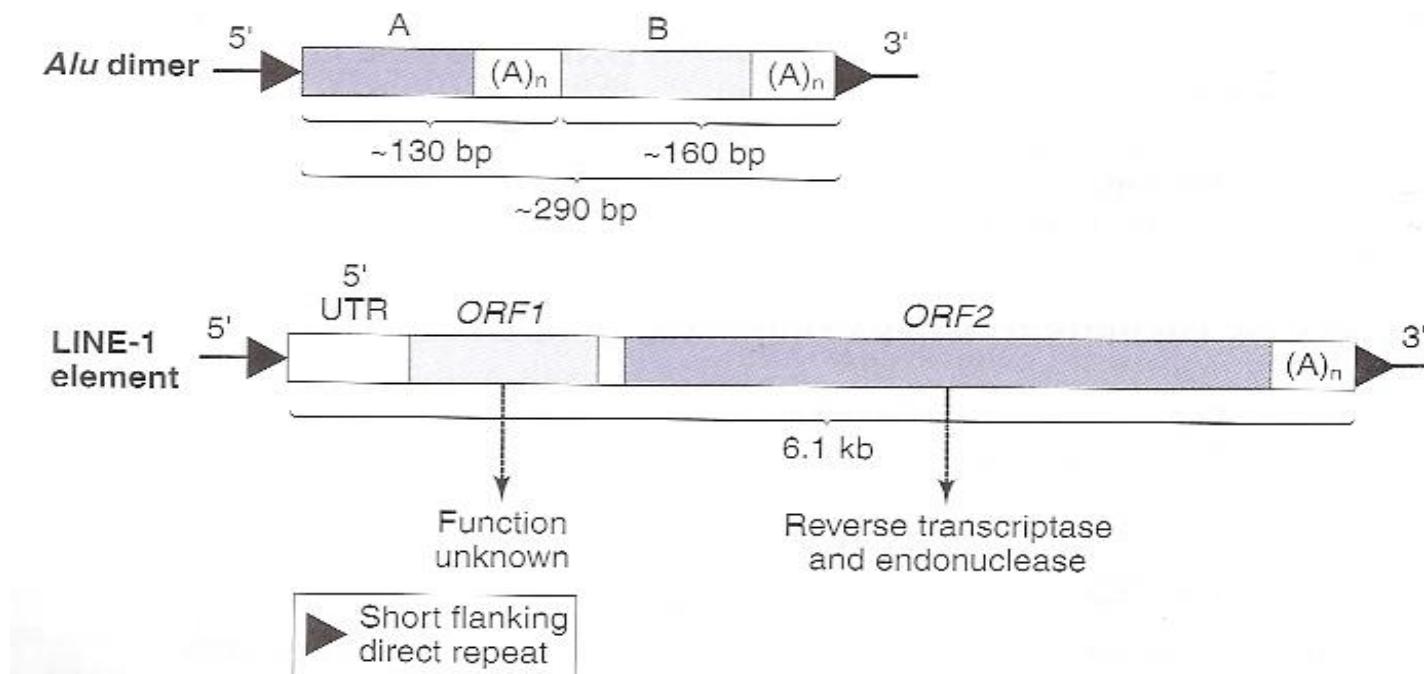


Long Interspersed Nuclear Elements - LINEs

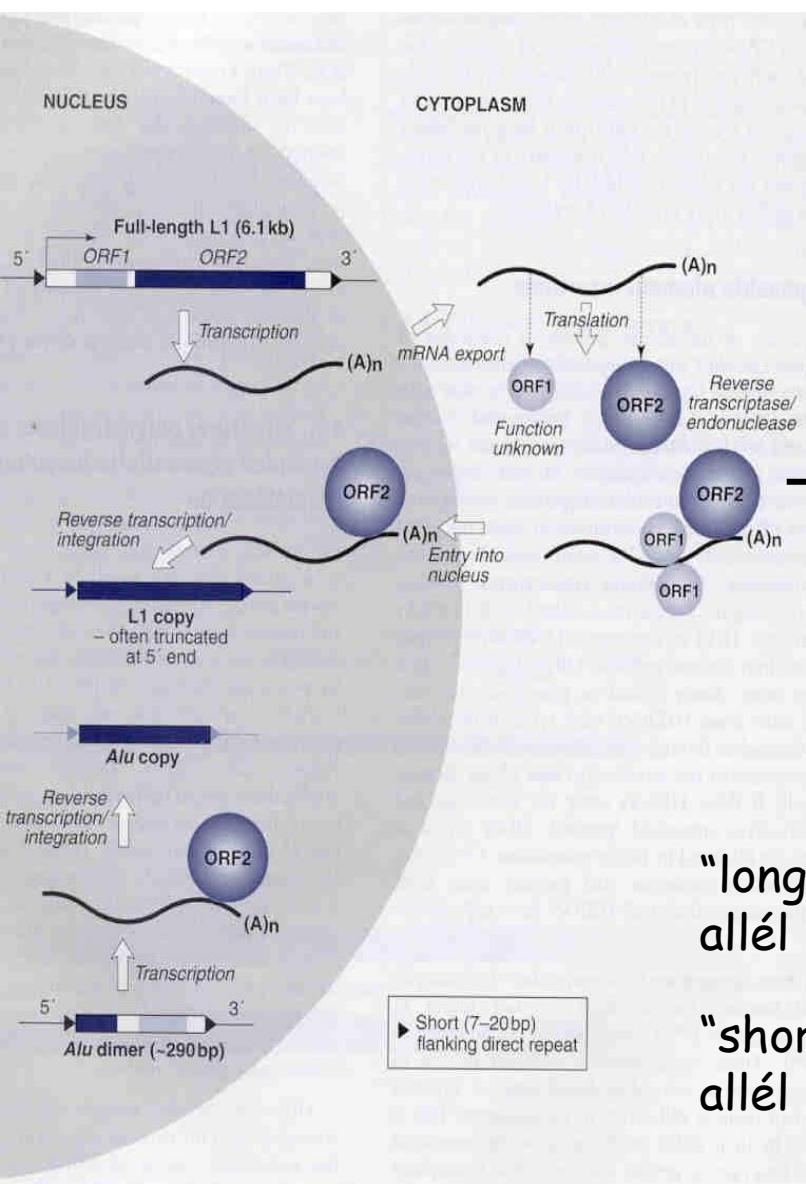
- LINEs v. Kpn: emberi genom 20 %-a, 870.000 kópia
- kb. 100 LINEs működőképes retrotranszpozonként
 - ~ 6 kb, belső 5' promóter, 2 ORF (RNS-kötő fehérje, endonukleáz + reverz transzkriptáz), poly(A)-farok,
- Target-primed reverz transzkripció: TT | AAAA - target
- Hanyag másolás
 - (transzkripció „read-through”, „dead-on-arrival”, nagyobb szekvenciarészek átrendeződése, egyéb nem autonóm szekvenciákhoz való kötődés)
- Önmaguktól nem tudnak a genomból kivágódni (deléció)
- ősi és relatíve új szekvenciák (pl. LINE-1: 5 Mys)

Short Interspersed Nuclear Elements - SINEs

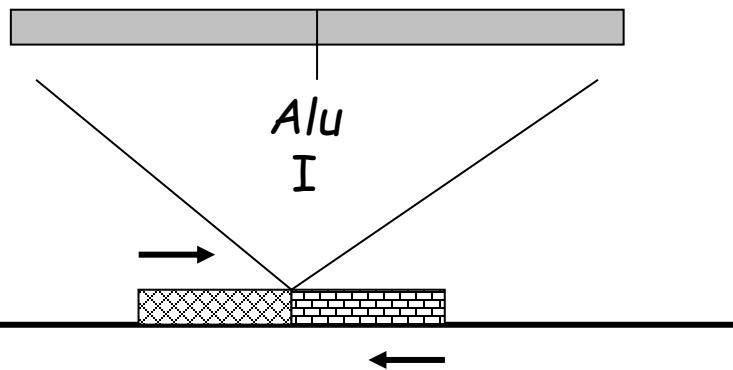
- SINEs v. *Alu*: 1.500.000 kópia, 70 % *AluI*, 300 bp,
- Főemlős specifikus, *Alu I* : AGCT, polimorfizmusok,
- Nem kódoló szekvencia, önállóan nem mobilizálódik
- *Alu* - LINE-1 retrotranszpozíció, 0.05 /genom / generáció



Mobilis elemek: biallélikus hossz-polimorfizmus



Human *Alu* Repeat
(~300 bp)



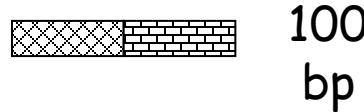
Kétféle alléltípus

"long" (+)
allél



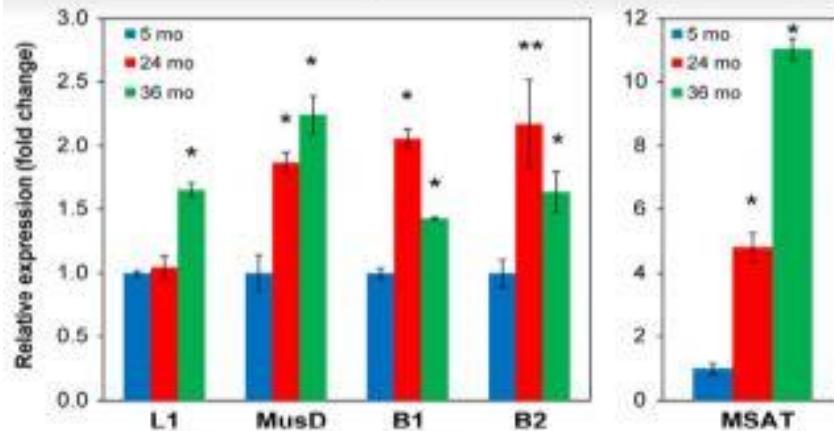
400
bp

"short" (-)
allél



100
bp

A. RTE and satellite RNA expression in liver



B. RTE and satellite RNA expression in muscle

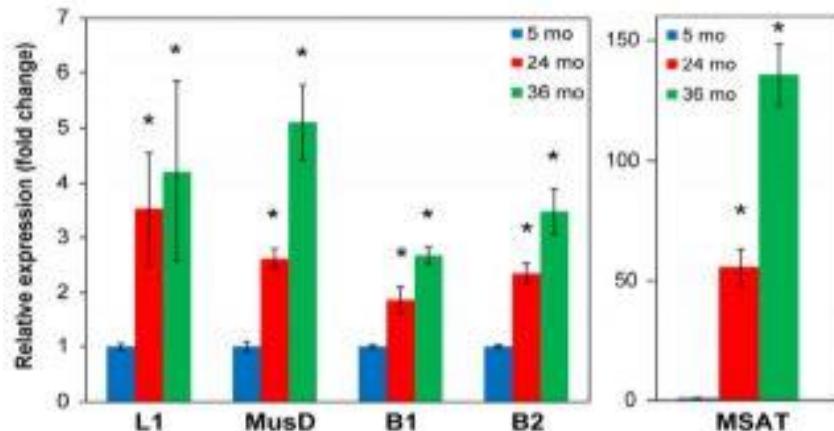
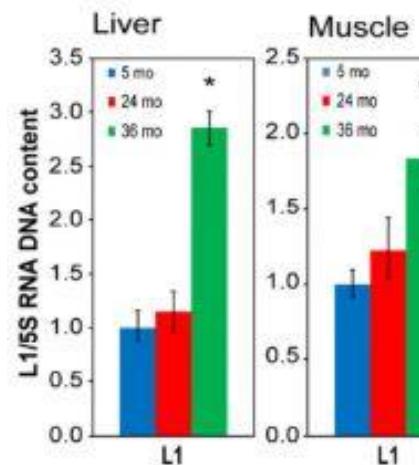


Figure 4. qPCR analysis of RNA expression of representative RTEs and SEs. Total RNA was extracted from (A) liver and (B) skeletal muscle, quantified by qPCR using indicated primers (Table S1) and normalized to GAPDH. Data were additionally normalized to the 5 month value for each element (shown as 1.0). L1, LINE L1; MusD, LTR RTE MusD/ETn; B1, SINE B1; B2, SINE B2; MSAT, major (also known as γ) SE. (*) p<0.01; (**) p≤0.05.

A. L1 copy number



B. MusD copy number

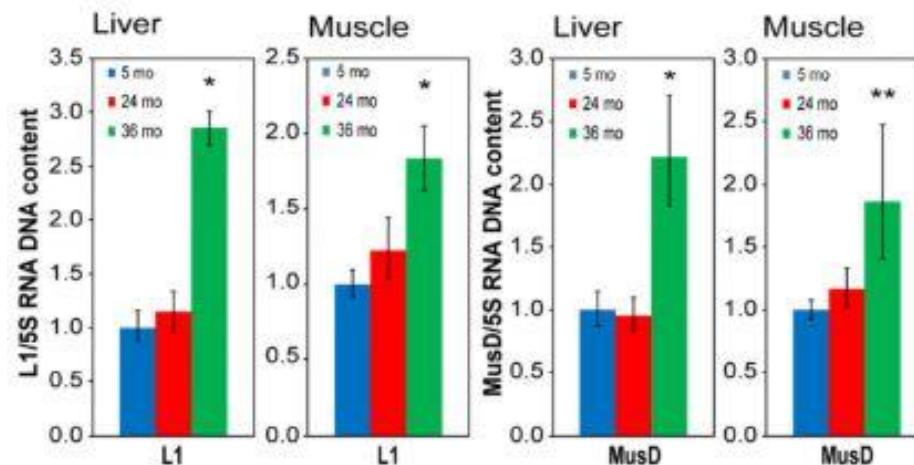


Figure 6. qPCR analysis of DNA to assess RTE genome copy number. (A) L1; (B) MusD. Total DNA was extracted from tissues of the same animals and tissues as used in Figure 4. Relative copy numbers were quantified using a multiplex TaqMan qPCR assay with the indicated primers (Table S1) and normalized to 5S ribosomal DNA. Data were additionally normalized to the 5 month value for each element (shown as 1.0). 5S DNA copy number was independently verified not to vary with age or between animals or tissues using qPCR against known single copy sequences. Means and standard deviations are shown. (*) p<0.01; (**) p≤0.05.

[Aging \(Albany NY\).](#)

[De Cecco M](#) [Criscione SW](#) [Peterson AL](#) [Neretti N](#) [Sedivy JM](#) [Kreiling JA](#)