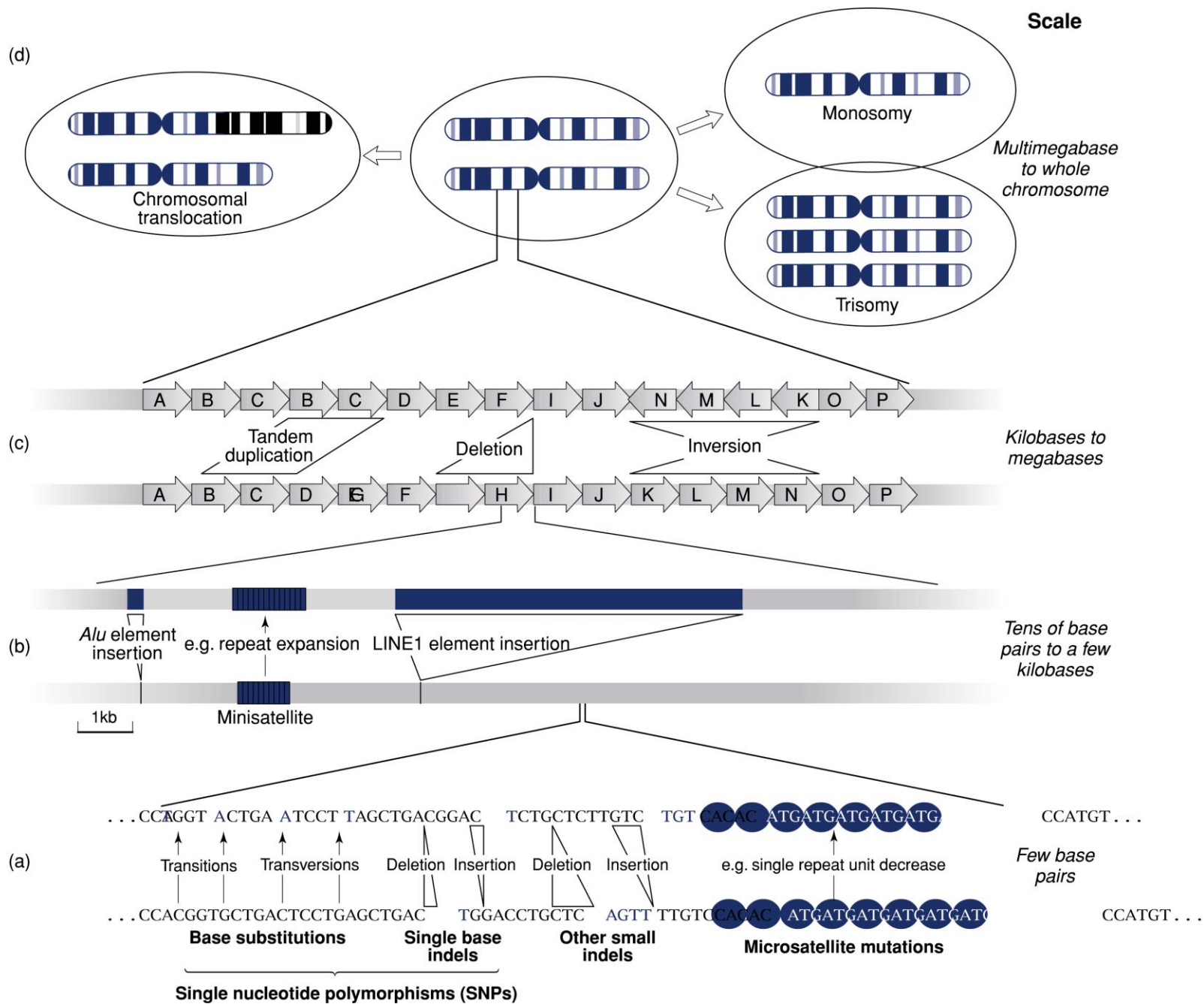




Genomika II.

Genomok variabilitása



DNS molekula szekvencia variabilitás: polimorfizmusok

(A) Pontmutációk: szubsztitúció, inzerció/delécio → szekvencia polimorfizmus → single nucleotide polymorphism

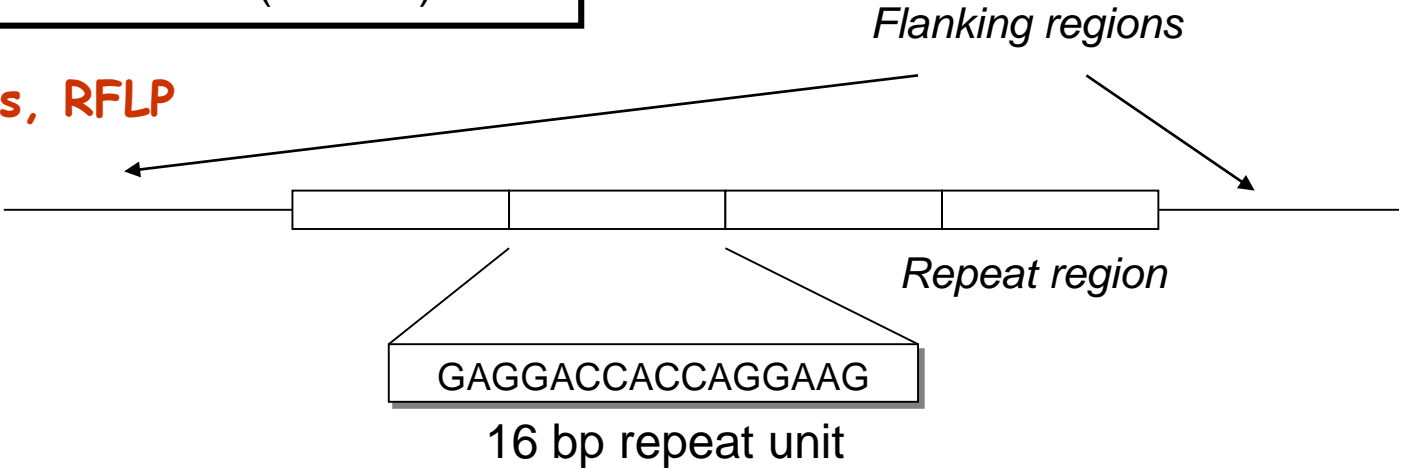
-----AGACTAGACATT-----
-----AGATTAG_CATT-----
SNPs

(B) Szekvencia ismétlődések: szatellit DNS, repetitív szekvenciák → hosszpolimorfizmus → short tandem repeats

-----**(AATG)(AATG)(AATG)**-----
-----**(AATG)(AATG)**-----
STRs
mikroszatelliták

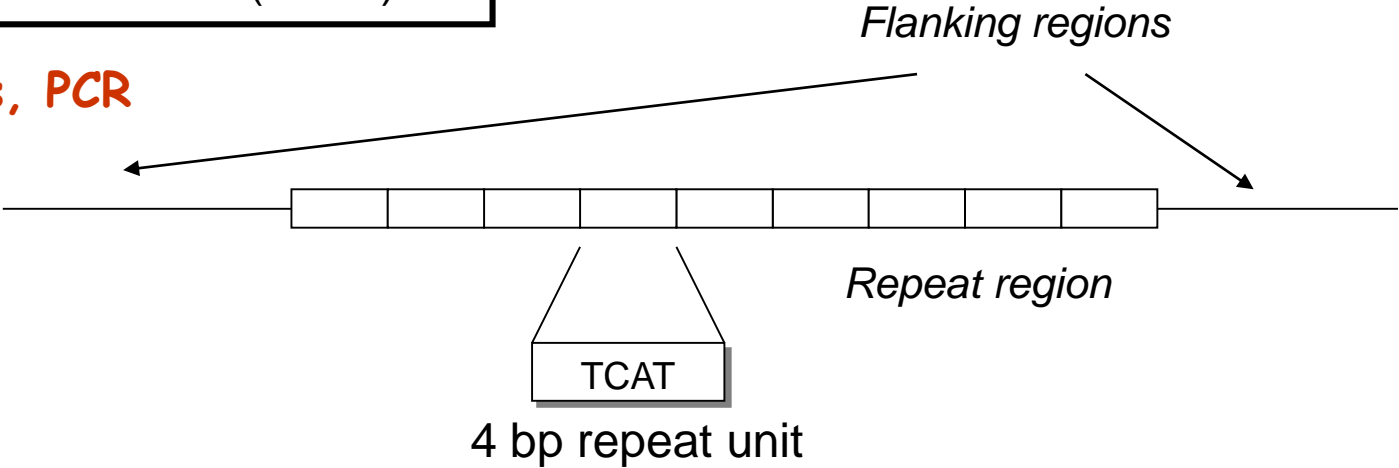
Minisatellite (D1S80)

VNTRs, RFLP



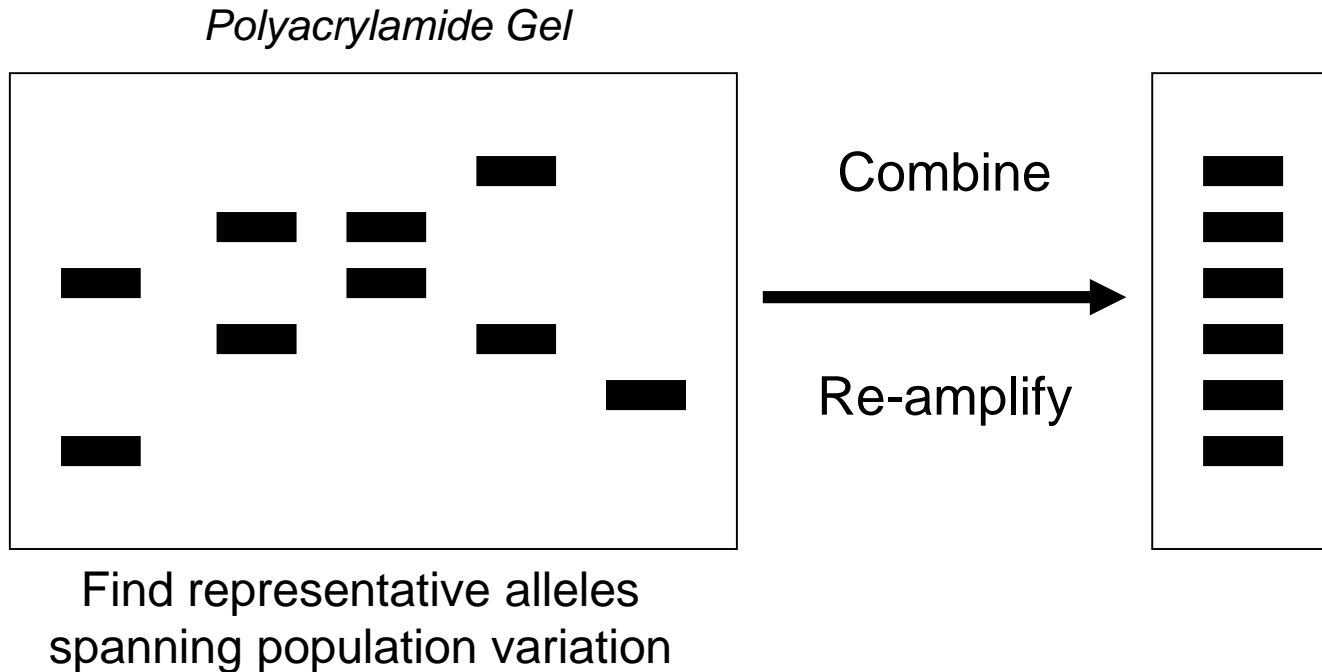
Microsatellite (TH01)

STRs, PCR

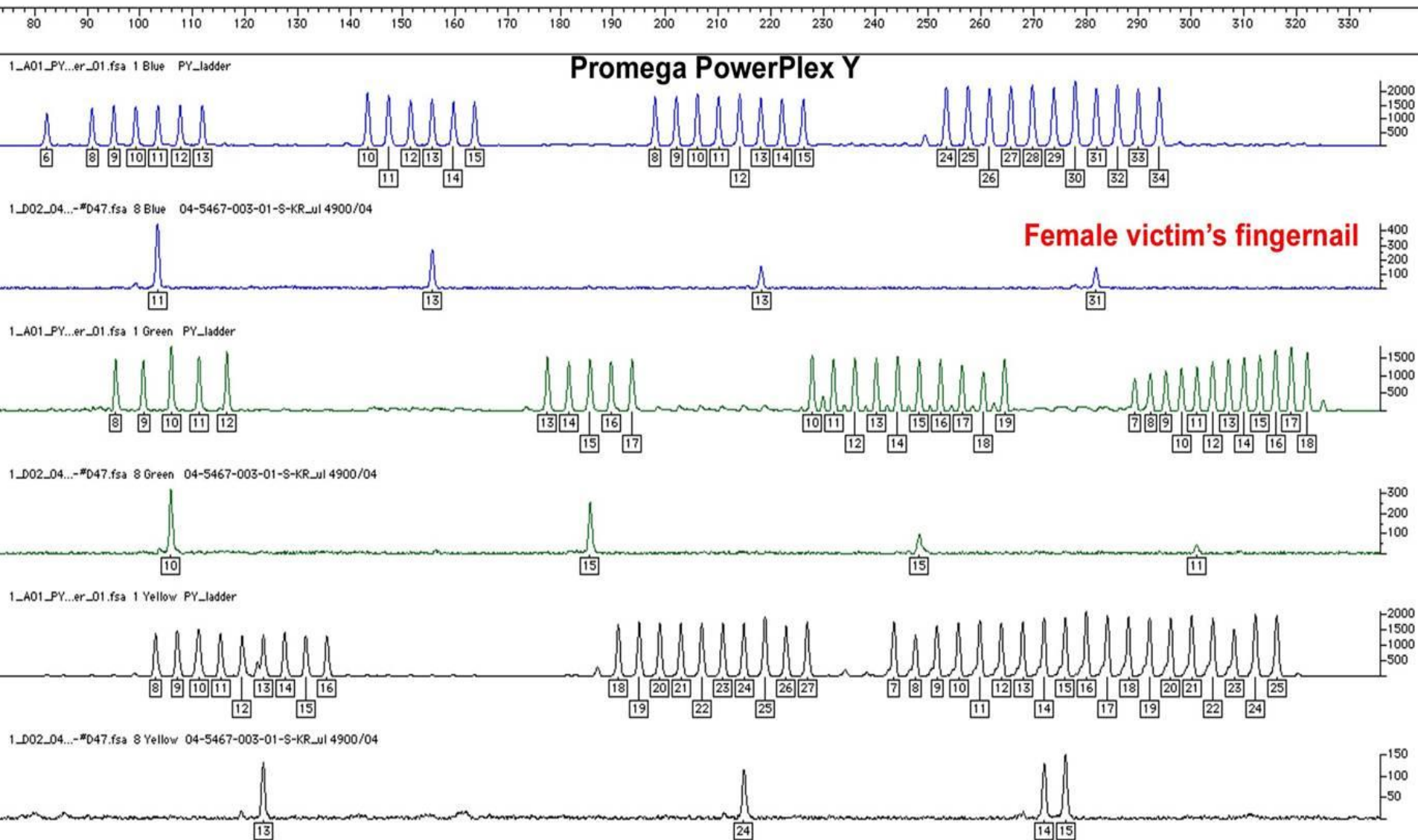


STR allélek genotipizálása: multiallélek

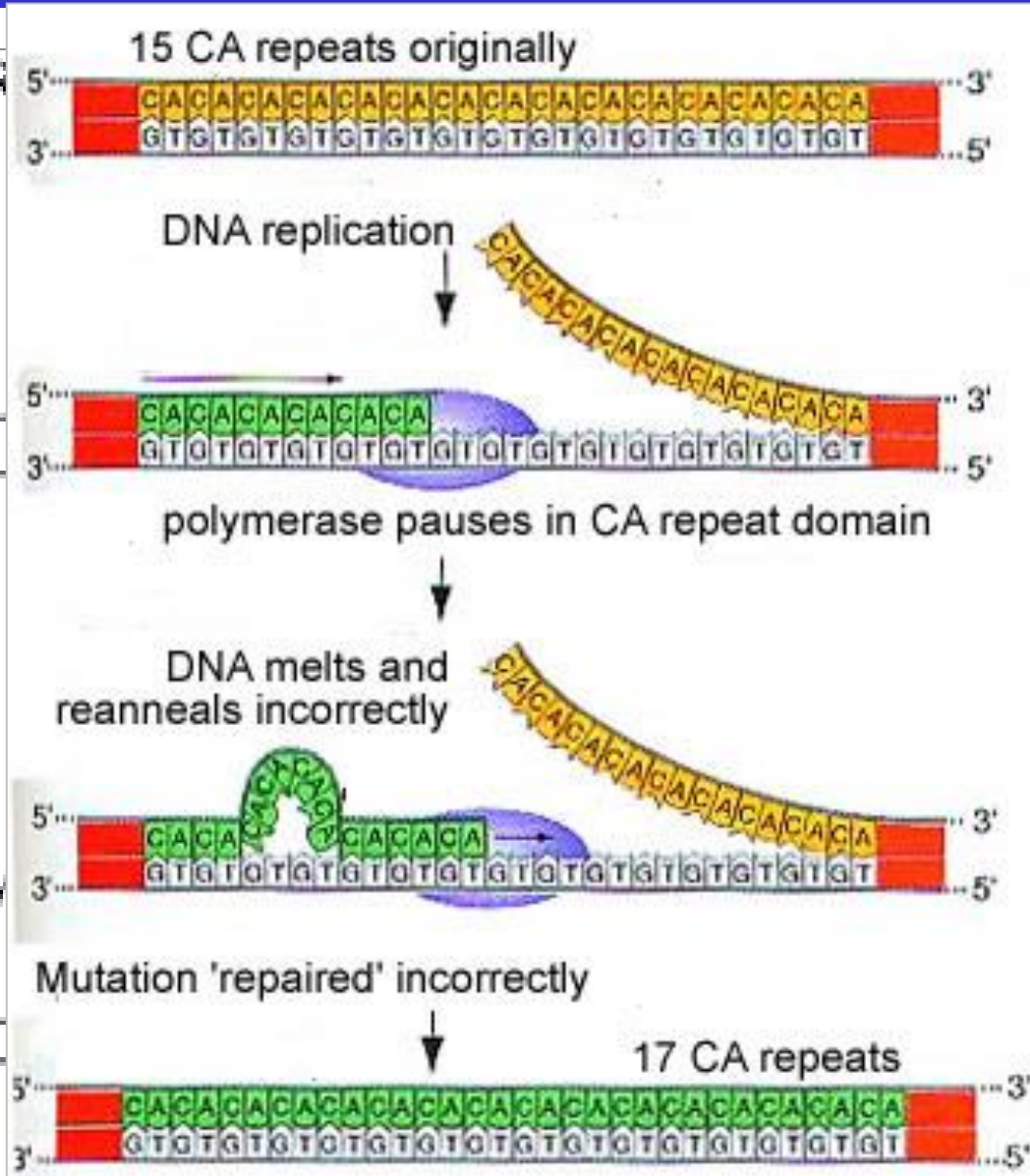
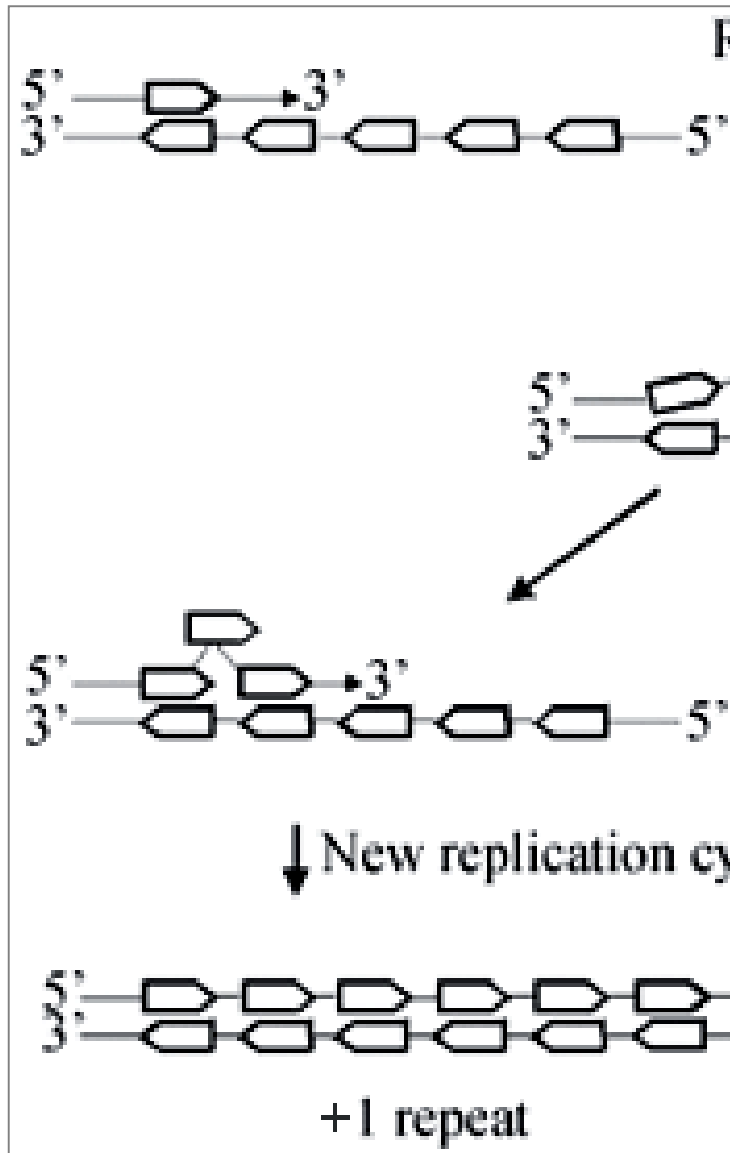
Separate PCR products from various samples amplified with primers targeted to a particular STR locus



Y kromoszómás mikroszatellita genetikai profil



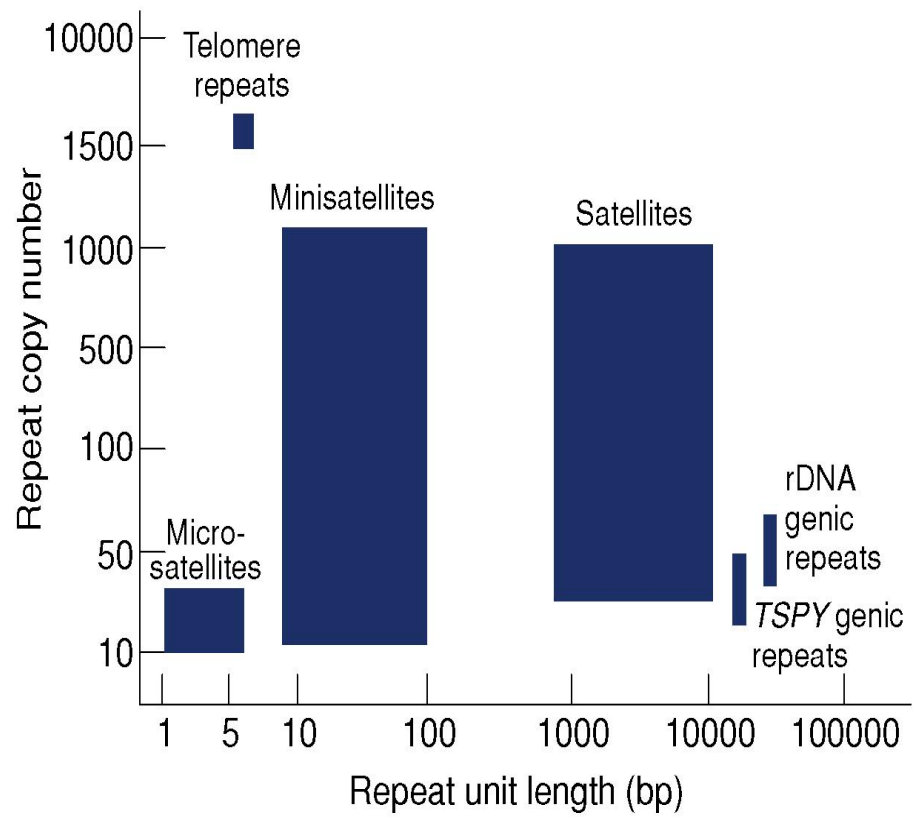
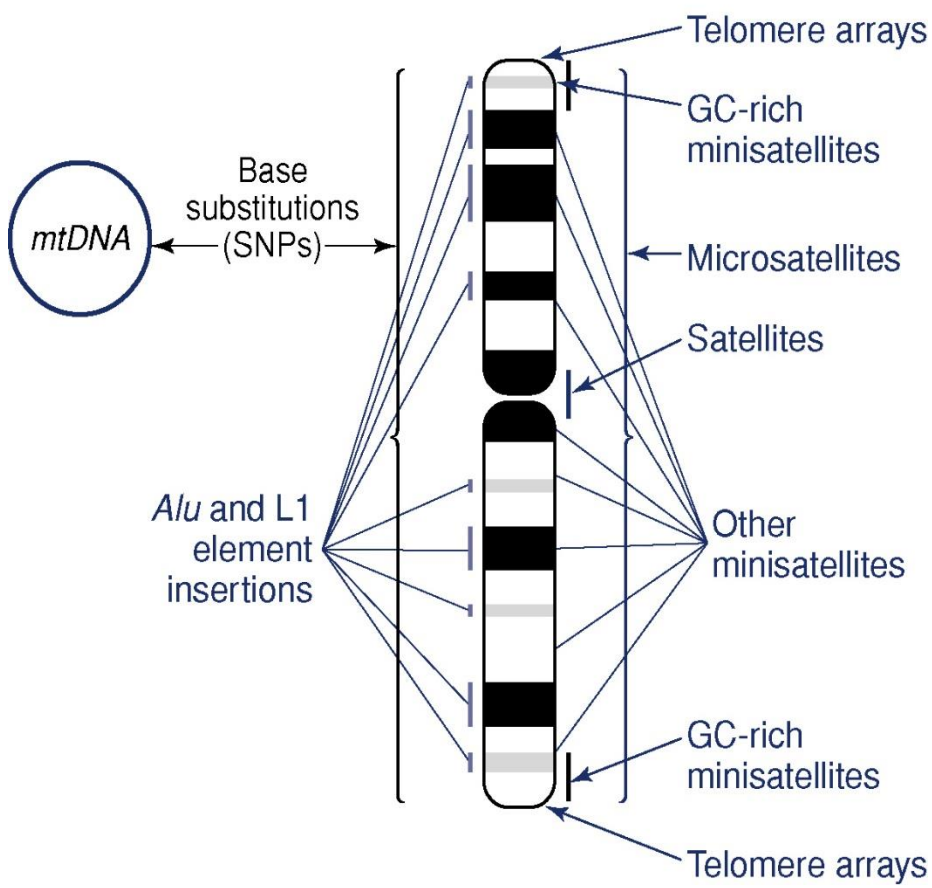
Mikroszatellita evolúció



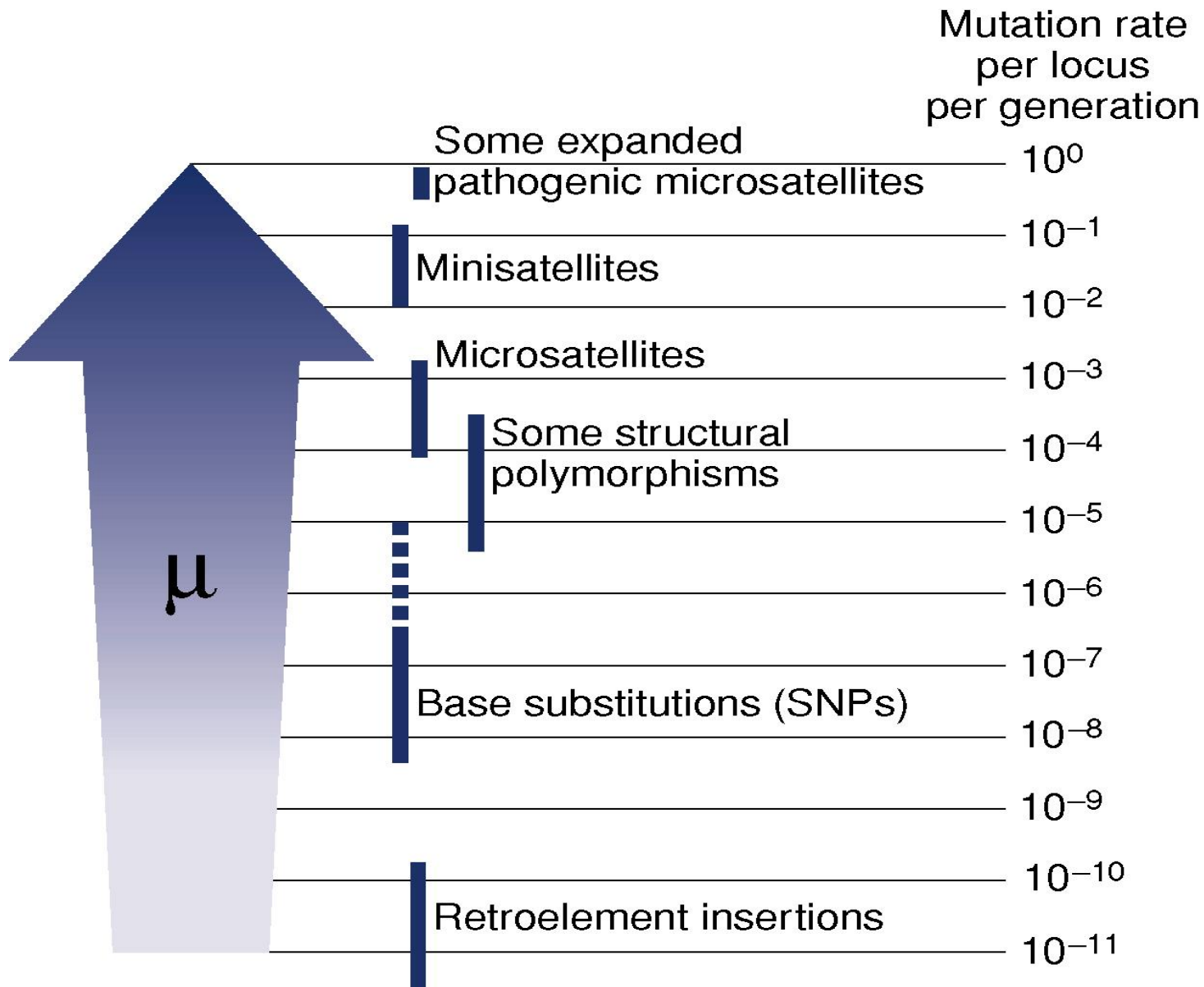
Ismétlődő elemek a genomban

BINARY MARKERS

MULTIALLELIC MARKERS

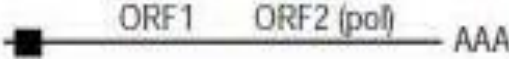

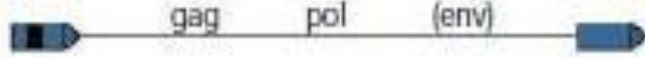
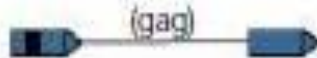
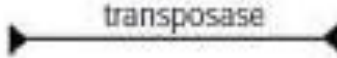



Polimorf genomik szekvenciák mutációs rátája (μ)



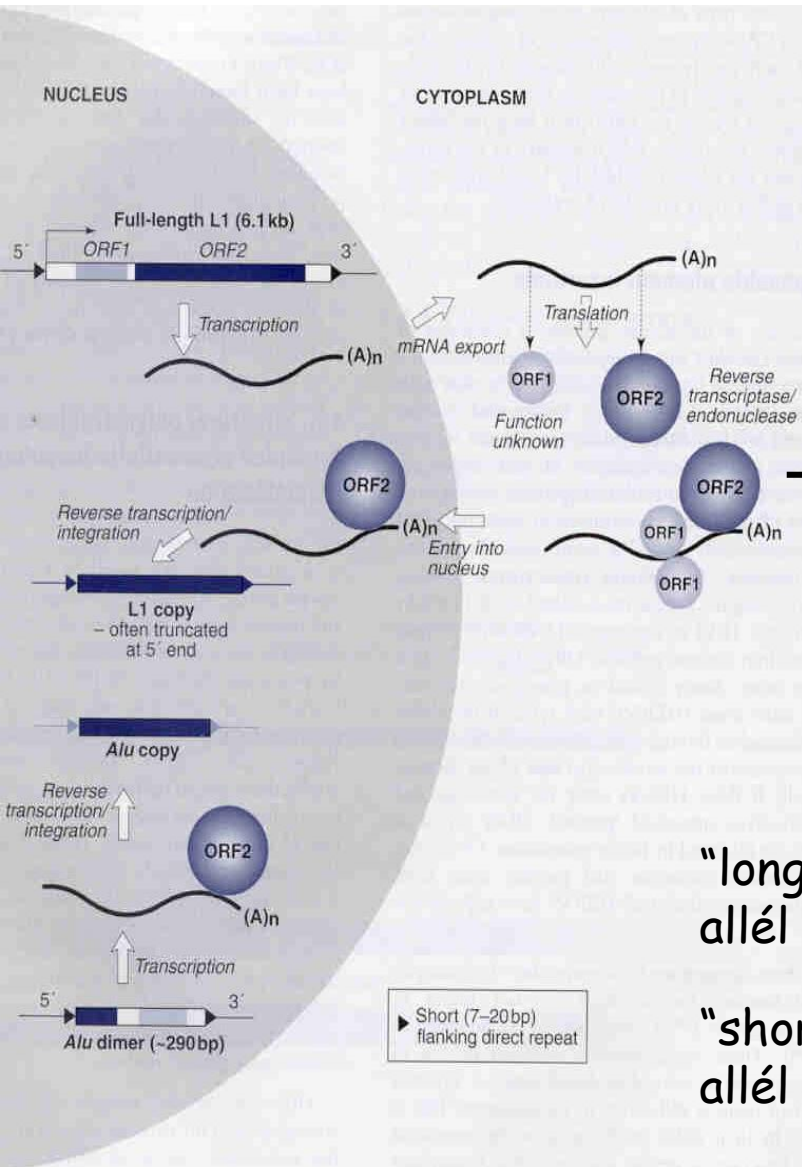
Mobilis genetikai elemek a humán genomban

Classes of interspersed repeat in the human genome

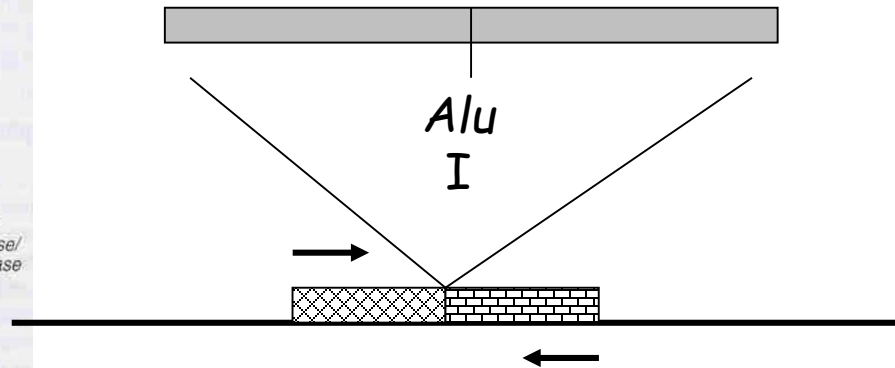
			Length	Copy number	Fraction of genome
LINEs	Autonomous		6–8 kb	850,000	21%
	Non-autonomous		100–300 bp		
Retrovirus-like elements	Autonomous		6–11 kb	450,000	8%
	Non-autonomous		1.5–3 kb		
DNA transposon fossils	Autonomous		2–3 kb	300,000	3%
	Non-autonomous		80–3,000 bp		

„copy-and-paste v. cut-and-paste“

Mobilis elemek: biallélikus hossz-polimorfizmus



Human *Alu* Repeat
(~300 bp)



Kétféle alléltípus

"long" (+)
allél

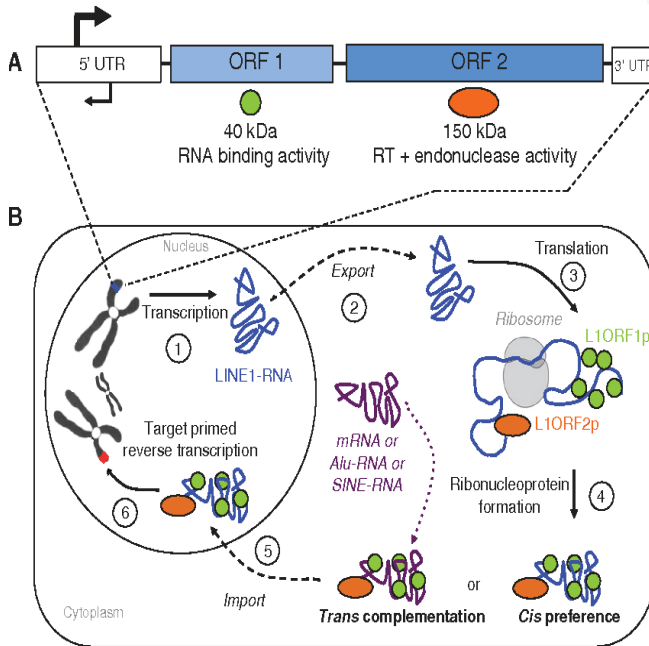


400
bp

"short" (-)
allél



100
bp



Regulation of LINE-1 in mammals
 Maxime Bodak / Jian Yu / Constance Ciaudo
 Published Online: 2014-09-25 | DOI: <https://doi.org/10.1515/bmc-2014-0018>

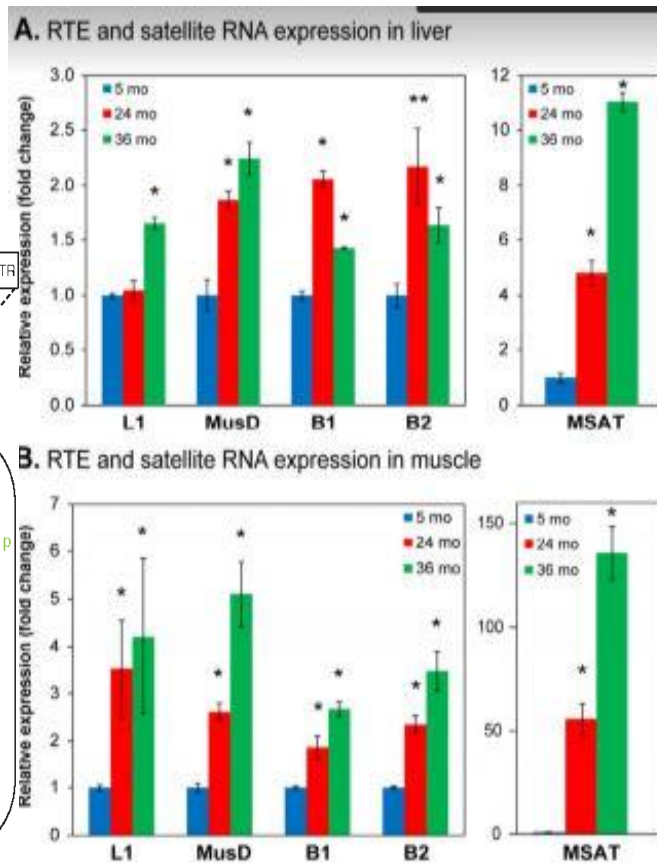


Figure 4. qPCR analysis of RNA expression of representative RTEs and SEs. Total RNA was extracted from (A) liver and (B) skeletal muscle, quantified by qPCR using indicated primers (Table S1) and normalized to GAPDH. Data were additionally normalized to the 5 month value for each element (shown as 1.0).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

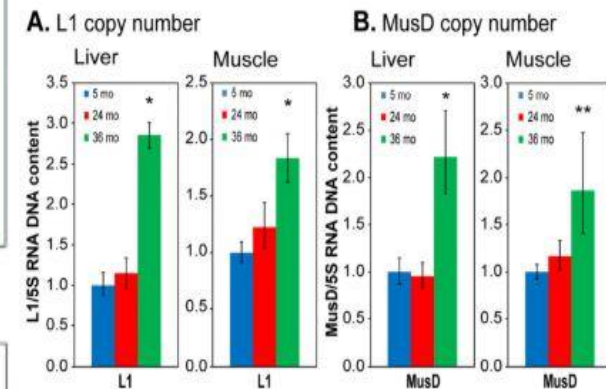


Figure 6. qPCR analysis of DNA to assess RTE genome copy number. (A) L1; (B) MusD. Total DNA was extracted from tissues of the same animals and tissues as used in Figure 4. Relative copy numbers were quantified using a multiplex TaqMan qPCR assay with the indicated primers (Table S1) and normalized to 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. (*) $p < 0.01$; (**) $p \leq 0.05$.

Aging (Albany NY). 2013 Dec;5(12):867-83.
 Transposable elements become active and mobile in the genomes of aging mammalian somatic tissues.
 De Cecco M¹, Criscione SW, Peterson AL, Neretti N, Sedivy JM, Kreiling JA

Human Y Chromosome Base-Substitution Mutation Rate Measured by Direct Sequencing in a Deep-Rooting Pedigree

Y kromoszóma
reszekvenálás:

ILLUMINA

Forensic Science International: Genetics 4 (2010) 59–61

Contents lists available at ScienceDirect

Forensic Science International: Genetics

journal homepage: www.elsevier.com/locate/fsig



Review

The hare and the tortoise: One small step for four SNPs, one giant leap for SNP-kind

Yali Xue, Chris Tyler-Smith*

The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambs CB10 1SA, UK

ARTICLE INFO

Article history:

Received 31 July 2009

Accepted 6 August 2009

Keywords:

Next-gen sequencing

Y-SNP

Y-STR

Haplotype resolution

Forensic applications

ABSTRACT

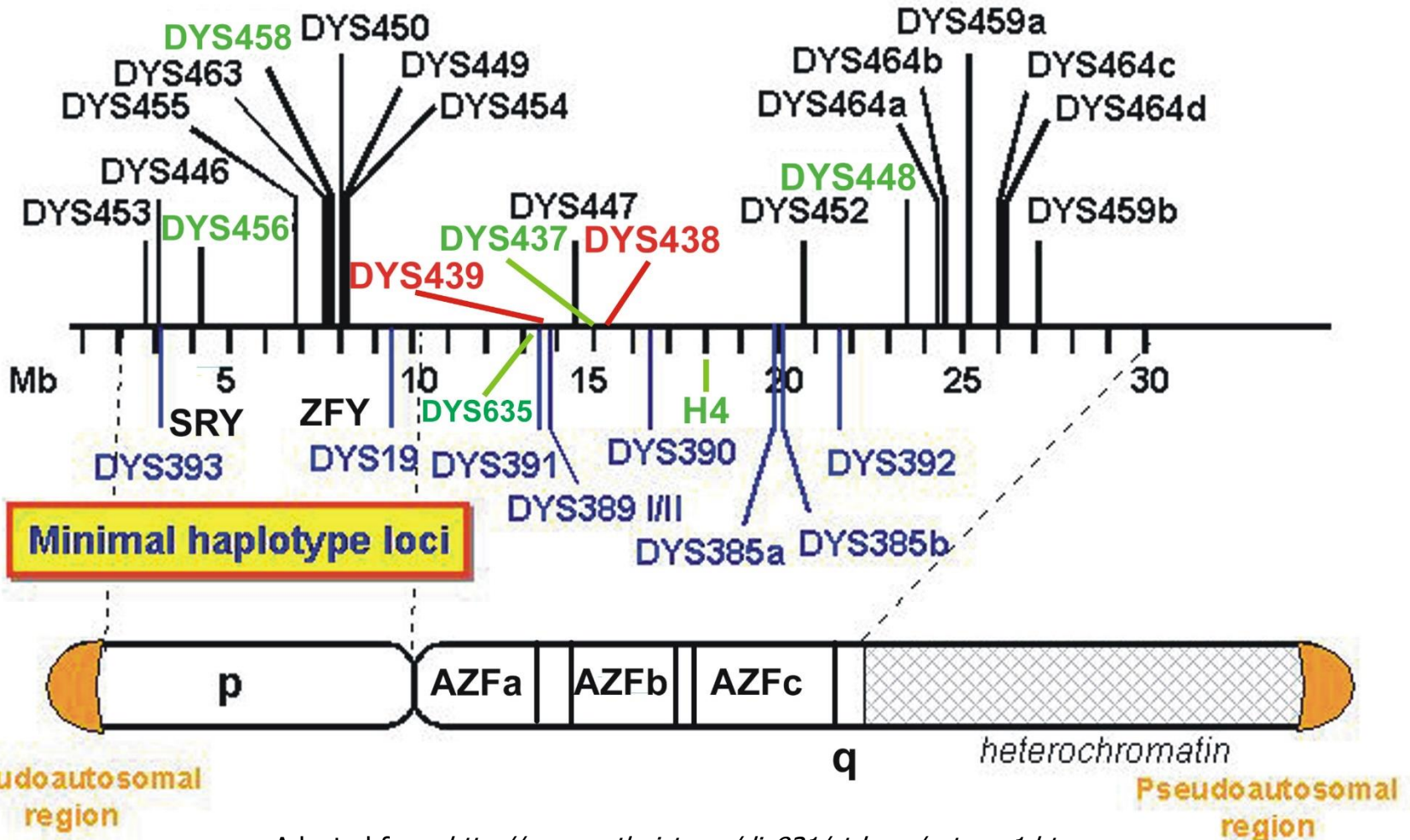
A recently published study has used next-gen sequencing technology to resequence two Y chromosomes separated by 13 generations and discovered four single-base differences in ~10 Mb DNA, suggesting that the Y chromosome euchromatin accumulates around one mutation per generation. Y-SNPs therefore now offer the best resolution of Y haplotypes and promise to distinguish almost every Y chromosome. This work illustrates the promise of current sequencing technology for forensically relevant applications.

© 2009 Elsevier Ireland Ltd. All rights reserved.

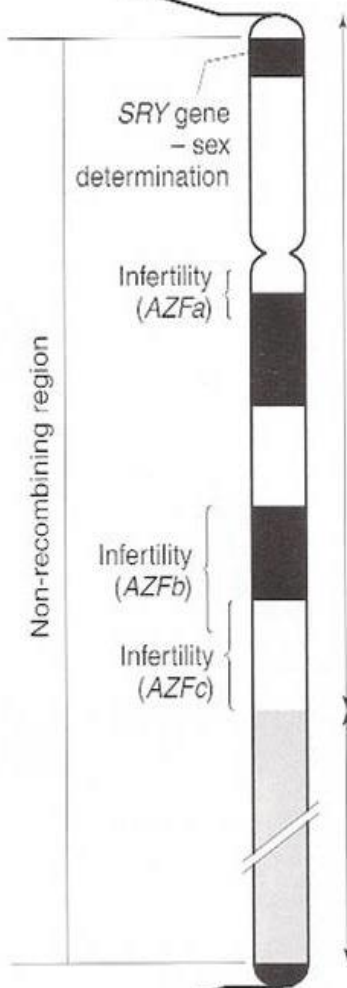
Y STR Positions along Y Chromosome

Extended **haplotype loci**

ABI AmpF/STR **Yfiler loci**



Pseudoautosomal region I: 2.6 Mb – obligatory recombination with the X



Euchromatin – ~30 Mb

Heterochromatin – variable in length; typically ~30 Mb

Pseudoautosomal region II: 0.32 Mb – recombination with the X not obligatory

Sample Number	M176	M5	M122	PN31	LLY22G	M134	M7	M113	M121	M159	M164	B_DYS388I	B_DYS389II	B_DYS390	B_DYS466	G_DYS19	G_DYS385a	G_DYS385b	G_DYS458	R_DYS437	R_DYS438	R_DYS448	R_Y_GATA_H4	Y_DYS391	Y_DYS392	Y_DYS393	Y_DYS439	Y_DYS635
66	A(1)	G(0)	G(1)	T(0)	C(0)	C(0)	G(0)	T(0)	A(0)	T(0)	A(0)	14	30	23	15	15	12	21	18	14	10	19	11	11	13	12	11	22
101	A(1)	G(0)	G(1)	T(0)	C(0)	C(0)	G(0)	T(0)	A(0)	T(0)	A(0)	14	30	23	15	15	12	21	18	14	10	19	11	11	13	12	11	22

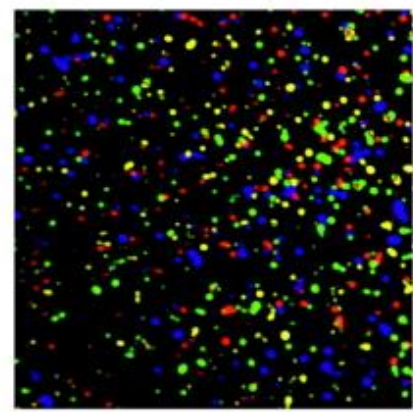
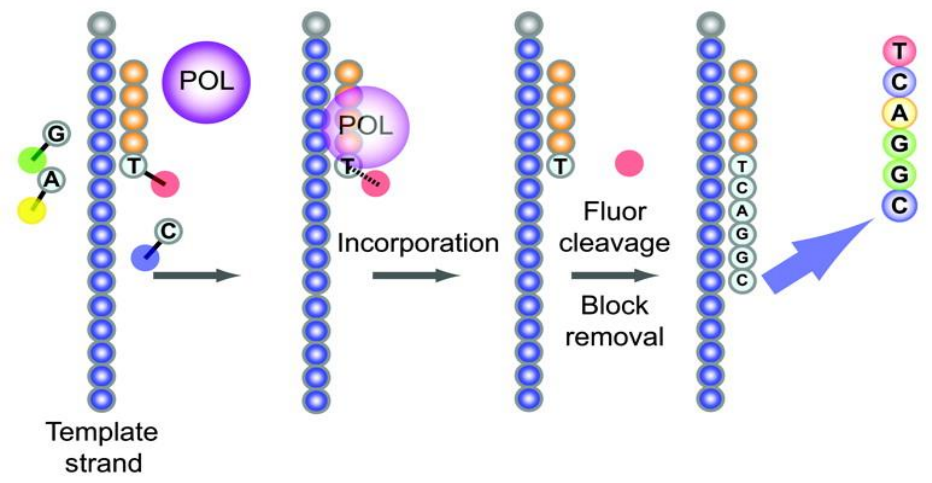
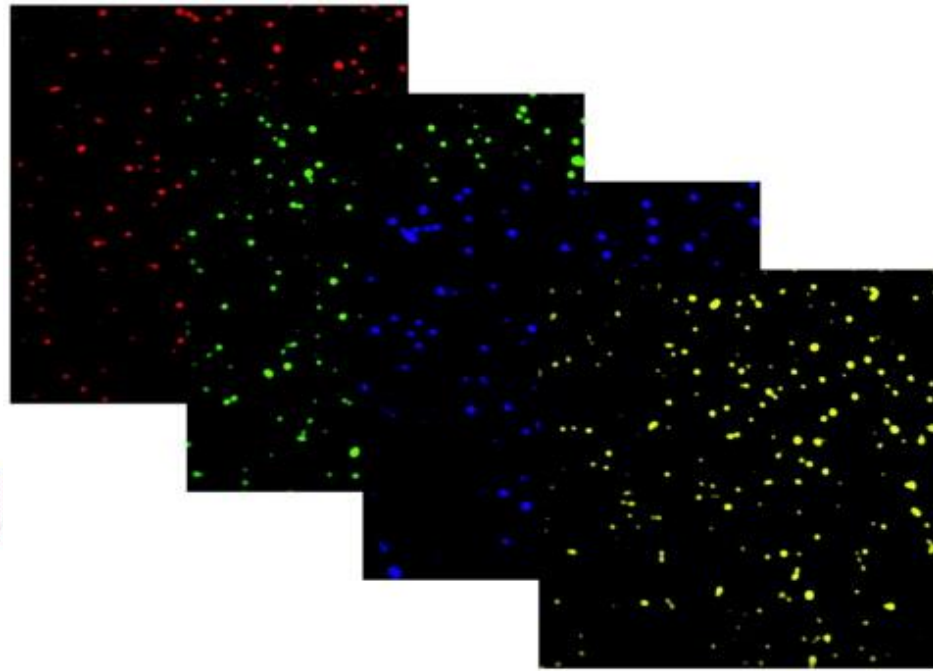
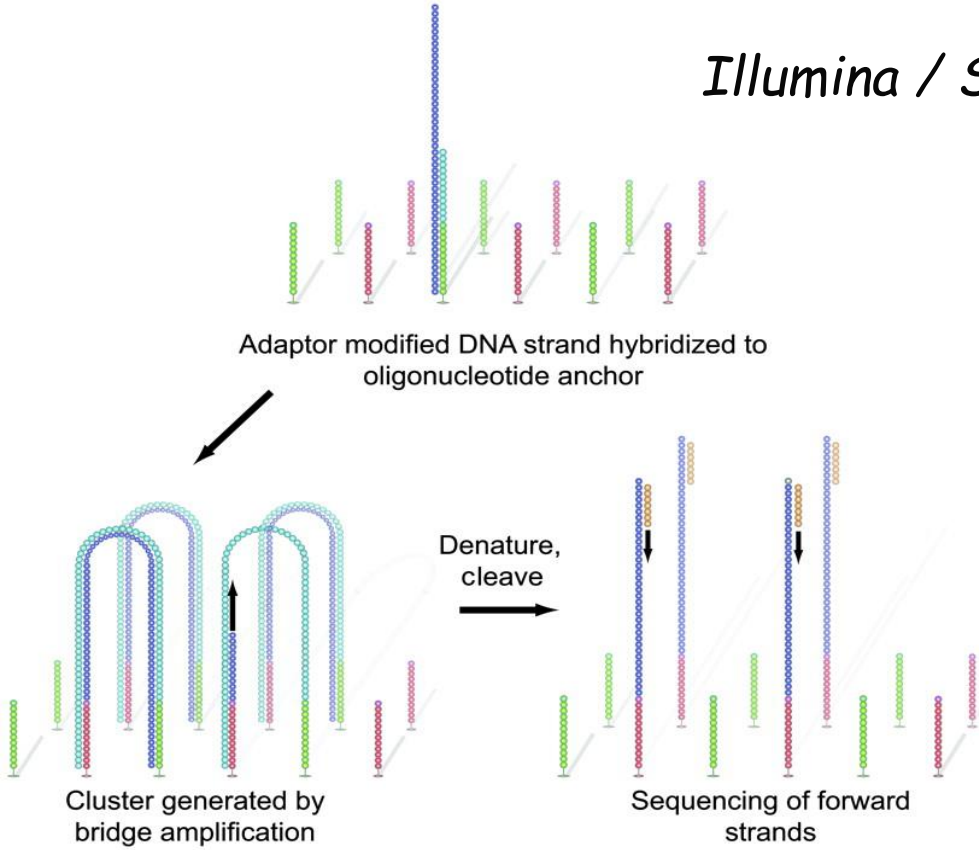
Sample number	DYS472 (B)	DYS508 (G)	DYS487 (Y)	DYS570 (G)	DYS583(B)	DYS579 (G)	DYS525 (Y)	DYS531 (B)	DYS488 (G)	DYS559 (Y)	DYS575(B)	DYS590(G1)	DYS636(Y)	DYS590(B)	DYS533(G1)	DYS617(Y)	DYS594(G2)	DYS505(B)	DYS641(G)	DYS638(Y)	DYS476(B)	DYS492(G)	DYS540(Y)	DYS537	DYS405 I(G)	DYS568(Y)	DYS480(B)	DYS572(G)
66	8	12	14	18	9	9	10	13	13	13	8	10	10	8	12	16	12	11	7	12	11	12	12	11	11	11	8	12
101	8	12	14	18	9	9	10	13	13	13	8	10	10	8	12	16	12	11	7	12	11	12	12	11	11	11	8	12

Sample number	DYS485(Y)	DYS490(B)	DYS495(G)	DYS667(Y)	DYS494(B)	DYS575(G)	DYS665(Y)	DYS48 I(G)	DYS576(B)	DYF390S I(G)	DYS669(Y)	DYS618(G)	DYS611(Y)	DYS643(B)	DYS666(G)	DYS673(Y)	DYS630(B)	DYS49 I(G)	DYS649(Y)	DYS640(G)	DYS654(B)	DYS497(G)
66	10	12	15	10	9	10	11	22	17	3	10	12	11	11	11	10	9	13	12	15	9	11
101	10	12	15	10	9	10	11	22	17	3	10	12	11	11	11	10	9	13	12	15	9	11

Table S1. Y-SNP and Y-STR haplotypes of the DFNY1-66 and DFNY1-101 chromosomes

Megegyező Y kromoszómás haplotípus
 67 mikroszatellita és 11 SNP markeren
 -generációs távolság: 13 generáció
 -markerek lokalizációja: eukromatin

Illumina / Solexa NGS genomszekvenálás



Sequencing by reversible dye terminators

Y kromoszómális eukromatin kandidáns pontmutációk

Table 2. Details of the Filtered Candidate Mutations

Chromosome Coordinate	Base	DFNY1_101 Pileup		DFNY1_66 Pileup		Confirmation	
		Coverage	Calls ¹	Coverage	Calls ¹	Cell-Line DNA	Blood DNA
First Class							
chrY:3,957,219	G	7	AAaaAAA	10	GGgGGGGgGG	Yes	No
chrY:4,633,474	C	4	tttT	6	cCCccc	Yes, het	No
chrY:4,939,256	T	13	cCccCcccCCCC	13	TTTTTTTTTtT	Yes	No
chrY:4,980,623	T	5	ggggg	7	TtTTTT	Yes, het	No
chrY:5,355,809*	C	12	TtTTTTTTtTt	9	cCccccCcC	Yes	Yes
chrY:6,555,594	G	13	TgTtTTtTTtTT	12	GGGGgGGgGGG	No	
chrY:7,381,330	G	7	cCcCCCc	12	GGGGgGGgGGG	No	
chrY:12,063,011	C	5	gggGG	8	ccccCCCC	Yes	No
chrY:14,745,277*	A	9	TtTtTtTt	6	aaAaAa	Yes	Yes
chrY:15,126,873	T	7	cccCccc	8	tttTtTT	Yes	No
chrY:15,146,905*	T	4	CCcC	9	tTtTTTTtT	Yes	Yes
chrY:20,627,064	C	9	gGGgGGGG.	5	Ccccc	Yes	No
chrY:27,095,961	T	7	CCcCCCc	8	TTtTTt	Yes	No
chrY:2,971,542*	A	4	aAAA	14	tTTtTtTtTtT	Yes	Yes
chrY:4,097,585	C	7	CCcaacc	2	aa	No	
chrY:4,876,956	T	11	aatTTTTTTTT	4	AAAA	No	
chrY:11,970,133	T	10	ttTTTTTTt	6	aaAAaa	No	
chrY:19,883,785	A	5	aAaaA	4	cccc	No	
Second Class							
chrY:13,445,456	G	4	GGGg	1	t	No	
chrY:13,568,272	G	13	aAAggggggggggg	11	aaaAaAaaAAa	No	
chrY:13,833,351	C	17	cCccCCggccCcCcccc	16	CCcCcCCcCttCtttc	No	
chrY:14,573,532	A	21	GAAAAaAaAAaAaaAAaAAg	5	AAggg	No	
chrY:15,375,202	G	4	GGGg	4	TTTT	No	

An asterisk denotes mutations that were confirmed in blood DNA.

¹ Upper case = forward strand; lower case = reverse strand.

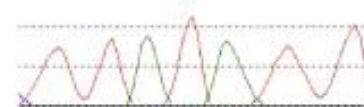
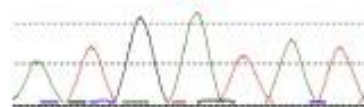
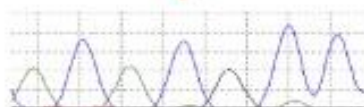
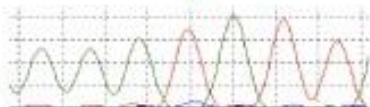
ChY: 2,971,542 (A→T)

ChY: 5,355,809 (C→T)

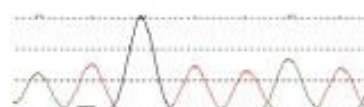
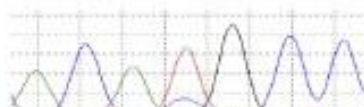
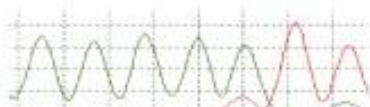
ChY: 14,745,277 (A→T)

ChrY: 15,146,905 (T→C)

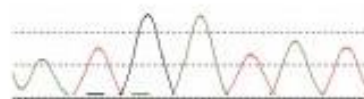
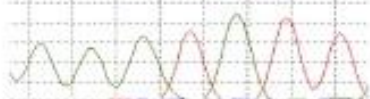
DFNY1-66
cell line DNA



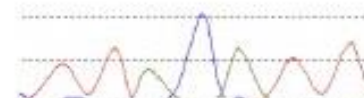
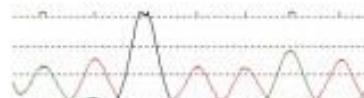
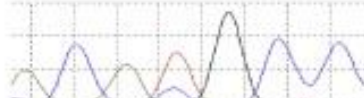
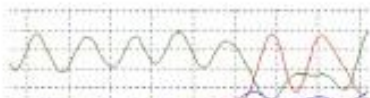
DFNY1-101
cell line DNA



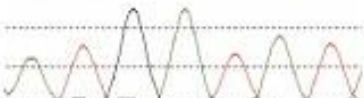
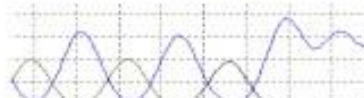
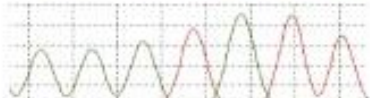
DFNY1-66
blood DNA



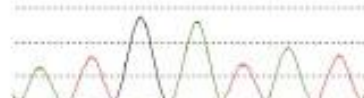
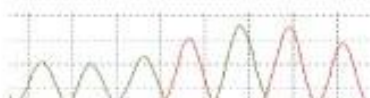
DFNY1-101
blood DNA



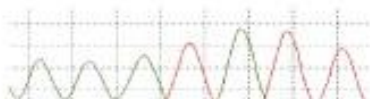
DFNY1-63
blood DNA



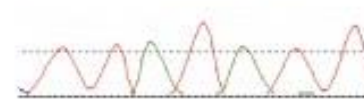
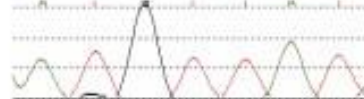
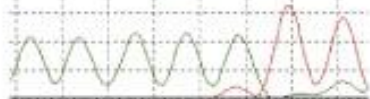
DFNY1-67
blood DNA



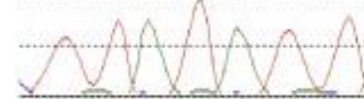
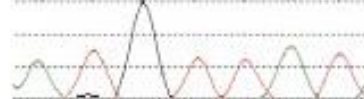
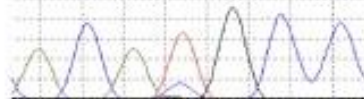
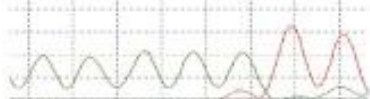
DFNY1-77
blood DNA

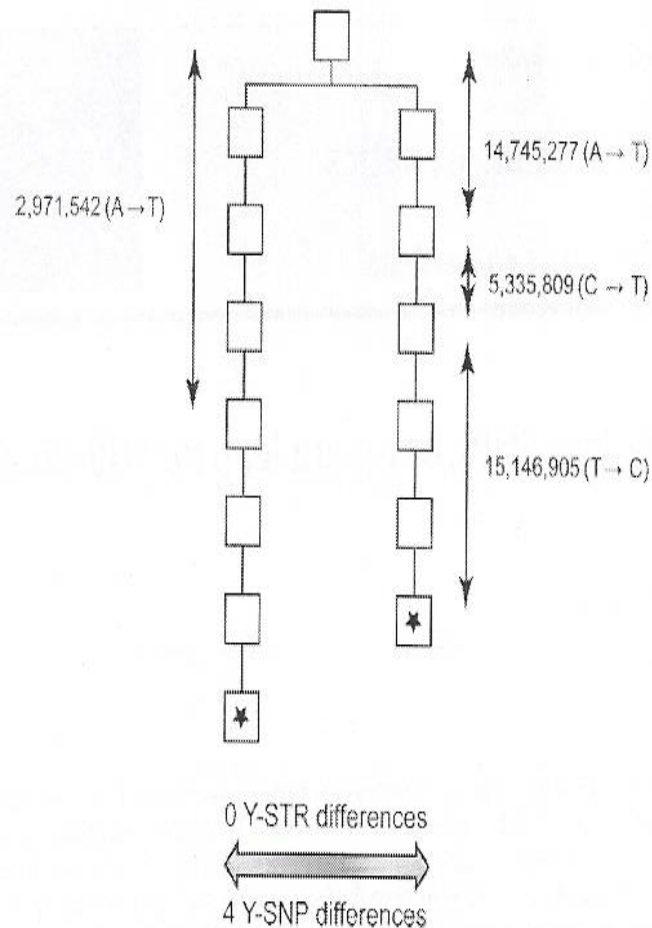


DFNY1-102
blood DNA

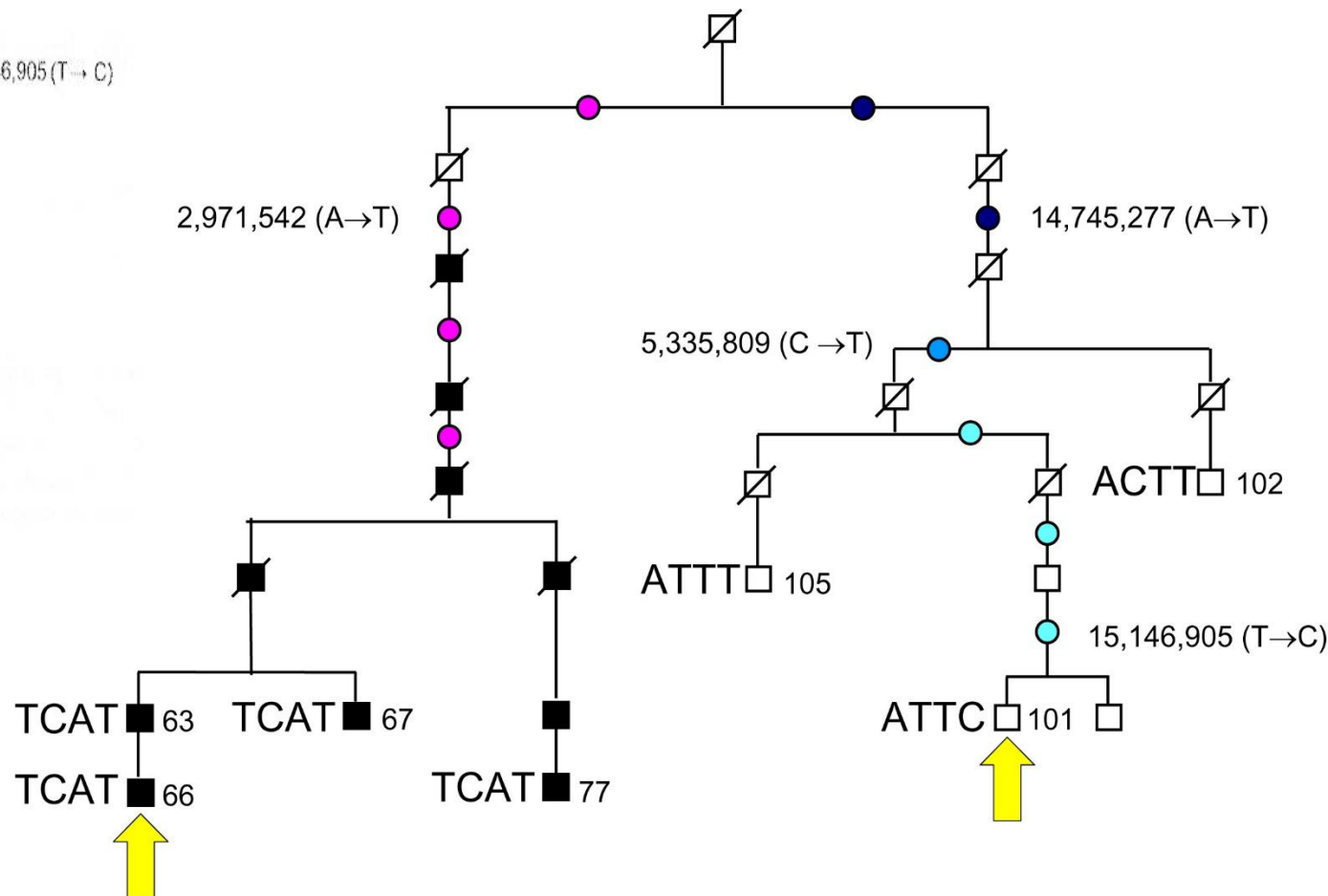


DFNY1-105
blood DNA





de novo szubsztitúciós mutációk az Y kromoszómás eukromatinban családfa analízis alapján



**Y chromosome
mutation rate:**

3.0×10^{-8}

**Consistent with human-chimp
Y chr. comparison**

Genetic History

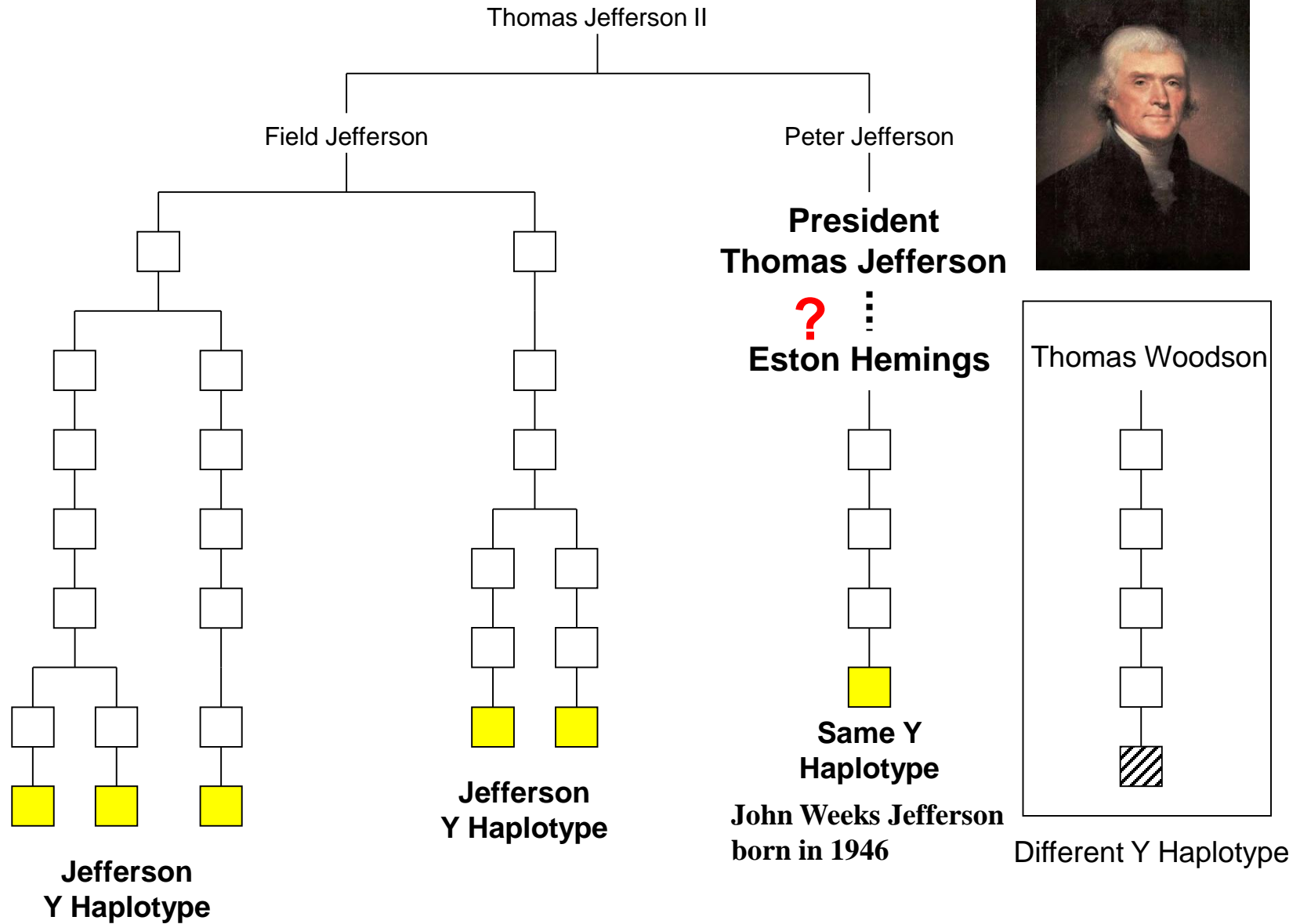


Figure 9.10, J.M. Butler (2005) *Forensic DNA Typing*, 2nd Edition © 2005 Elsevier Academic Press

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 (5Y81)	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	(3)-5	(1)-17 ←	(1)-16 ←
	(1)-14	(1)-14	(3)-36 ←	(3)-27 ←
	(3)-32	(3)-32	(4)-21 ←	(4)-21 ←
	(4)-16	(4)-16		

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.

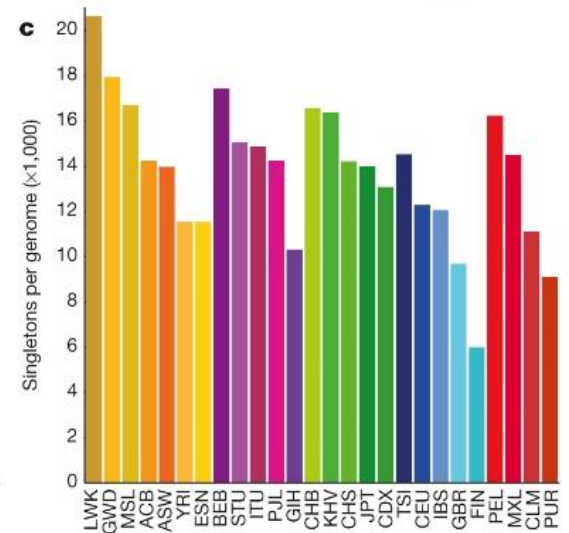
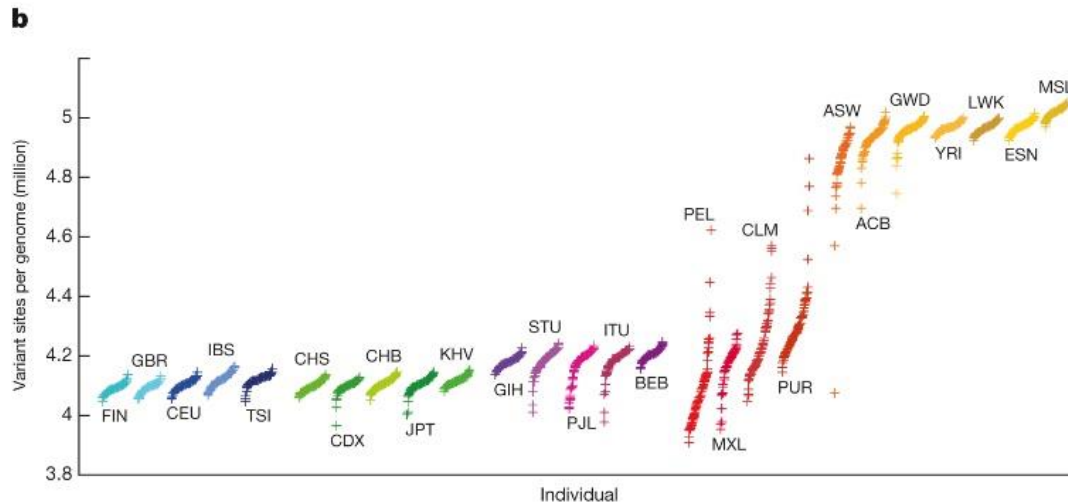
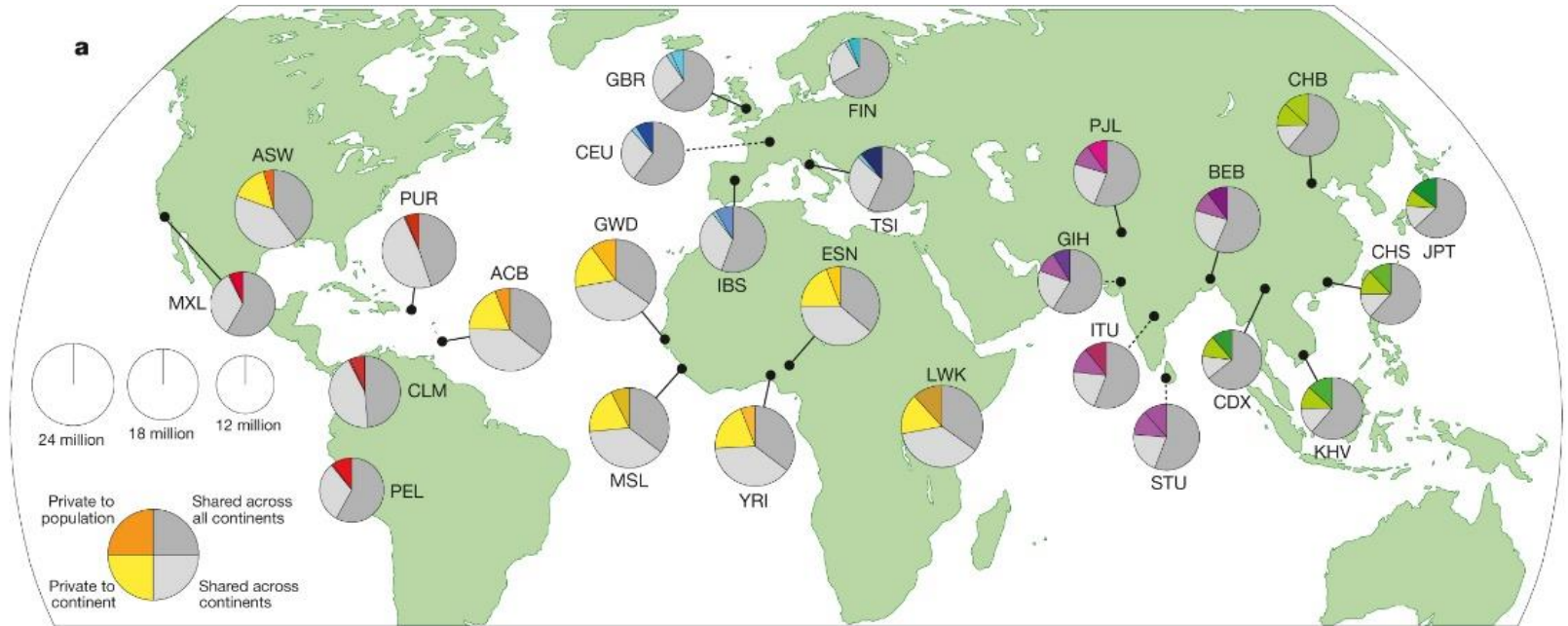
Structural variants are implicated in numerous diseases and make up the majority of varying nucleotides among human genomes. Here we describe an integrated set of eight structural variant classes comprising both balanced and unbalanced variants, which we constructed using short-read DNA sequencing data 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.

	Autosomes	Exome target regions**	chrX***	chrY***	Totals
Samples	2,504	2,504	2,504	1,233	-
Total Raw Bases (Gb)	85,426	18,273	3,213	291	-
Mean Mapped Depth (X)*	8.45	75.25	6.20	2.60	-
Total Variant Sites	84,801,880	1,416,049	3,468,093	62,042	88,332,015
Biallelic SNPs	81,102,777	1,383,927	3,223,927	60,505	84,387,209
Indels	3,196,364	19,832	212,196	1,427	3,409,987
Mean Indel Length (bp)	2.94	3.46	2.64	2.00	-
Multiallelic sites	444,026	6,153	30,996	-	475,022
Multiallelic SNPs	274,425	4,706	15,055	-	289,480
Multiallelic Indels	169,601	1,447	15,941	-	185,542
Structural Variants	58,713	6,137	974	110	59,797
ALU Insertion	12,491	52	-	-	12,491
LINE1 Insertion	2,910	10	-	-	2,910
Large Deletion	33,336	2,684	974	-	34,310
Duplication	5,896	2,513	-	-	5,896
SVA Insertion	822	5	-	-	822
Other Insertion	165	1	-	-	165
Inversion	100	8	-	-	100
CNV	2,993	864	-	110	3,103

Supplementary Information Table 3: Integrated callset summary. *Assuming 2.84Gb as the genome size. The mapping of exome sequence to targeted pull down regions was calculated by Picard function *calculateHsMetrics*. **The exome targeted regions were exome pulldown targets derived from CCDS (NimbleGen EZ Exome v1 and Agilent SureSelect v2). These variant totals are included in the other columns. ***chrX and chrY statistics are for the entire chromosomes.

- 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.

Population sampling



Egynukleotid polimorfizmusok (SNPs)

- Biallélikus markerek (6 lehetőség)
 - (A / G, C / T, A / T, C / G, T / G, A / C)
- Több millió SNP a genomban
 - kb. 500-1000 bázispáronként
 - pontmutációk génekben és/vagy regulátor régiókban
- Fenotípus kapcsolatok
 - pigmentáció, testalkat, ...
- Leszármazási vonalak (Y-SNP's)
- Diagnosztika
 - multifaktoriális poligénes betegségek

SNP markerek a humán genomban

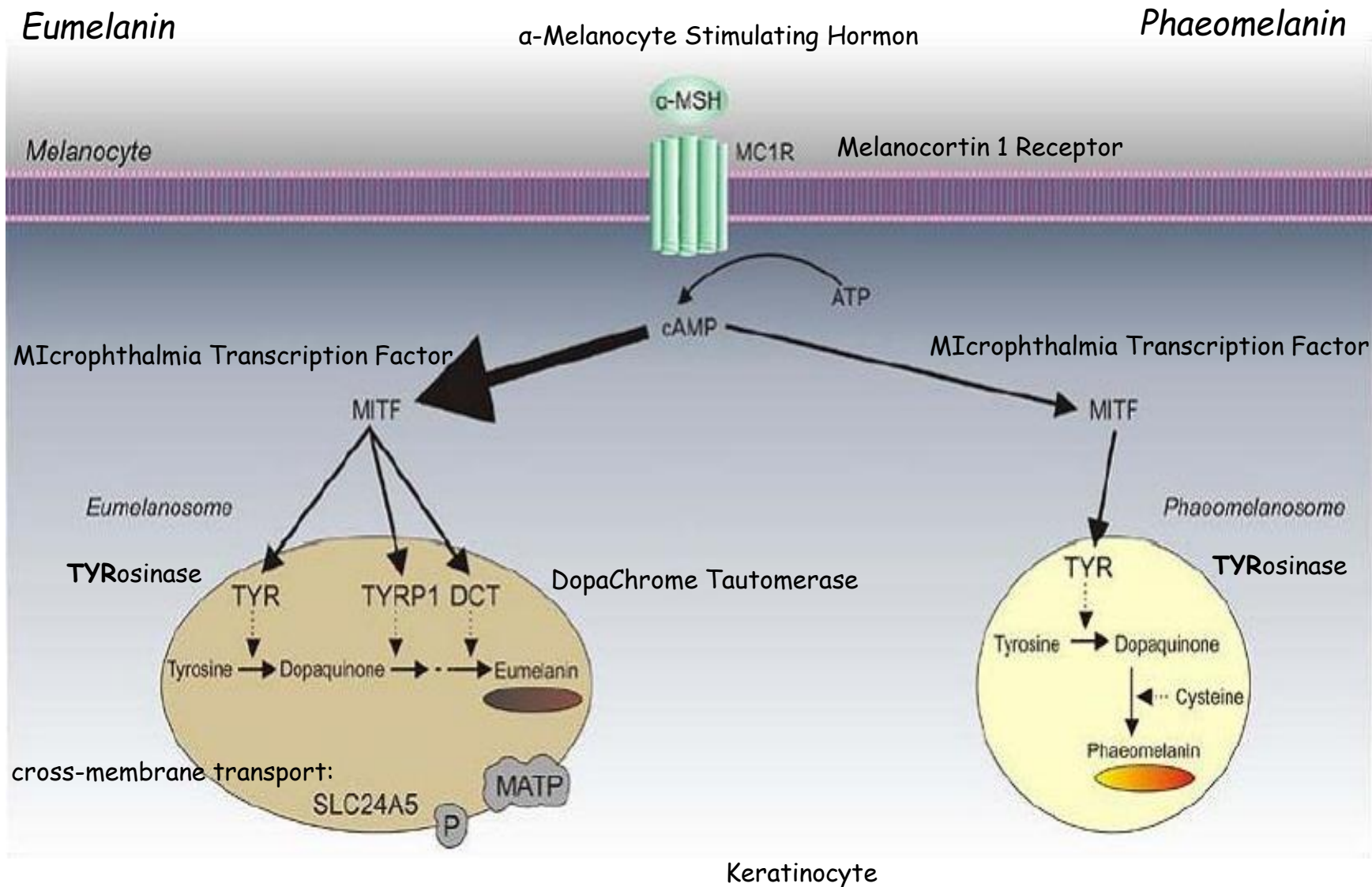
TABLE 12.2 Categories of SNP Markers (See Budowle & van Daal 2008, Butler et al. 2008).

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 pigmentáció genetikai szabályozása

- Melanoszóma: eumelanin v. phaeomelanin útvonal
- 127 pigmentációs gén az egér genomban
- Humán vonalon 12 gén azonosítása (2009)
- SNPs: fenotípus asszociált, ancestry-informative (AIMs)
- MC1R allélek aktivitása: RHC fenotípus, szeplősödés

A melanogenesis biokémiai útvonal



SNPs - pigmentációs gének

- ASIP (aguti): 3'UTR 8818A - MSH antagonist - phaetomelanin 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	T/C	Non-coding, intron 1
			rs6497268 or rs4778241	G/T	
			rs11855019 or rs4778138	T/C	
			rs1545397	G/A	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

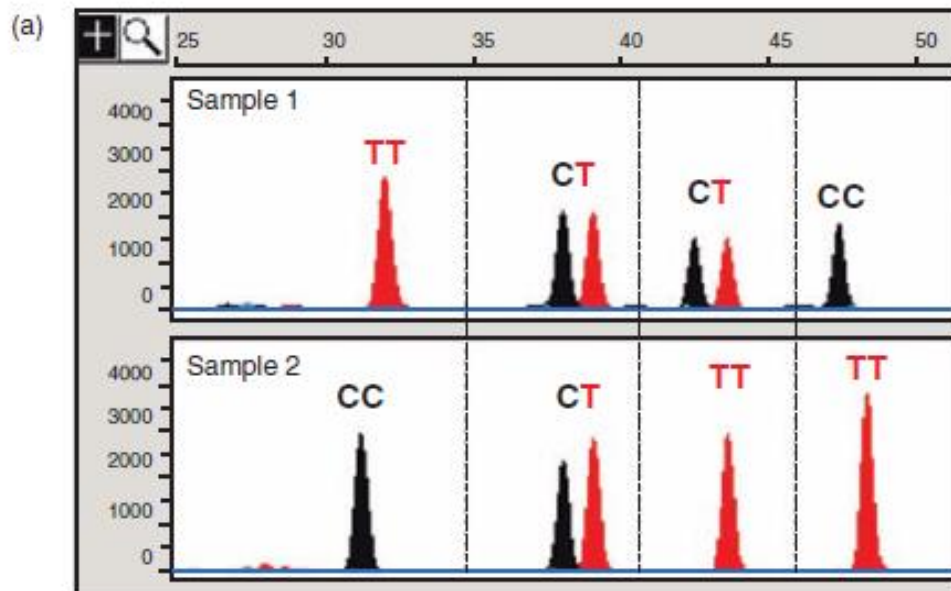
^a ns non-synonymous

^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)

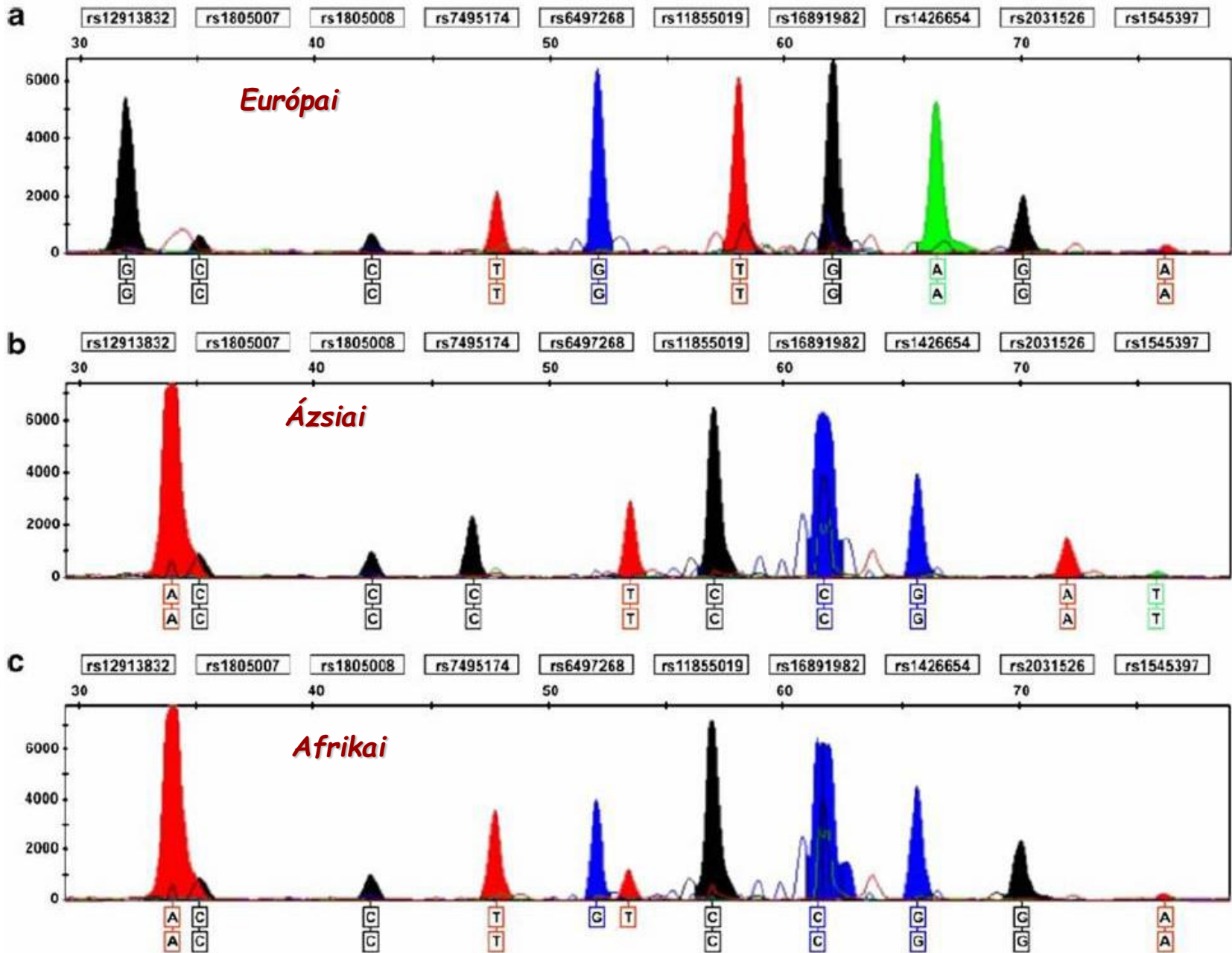
Allele-specific primer extension across the SNP site with fluorescently labeled ddNTPs; mobility modifying tails can be added to the 5'-end of each primer in order to spatially separate them during electrophoresis.



- (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 SNPs genotipizálás (SNaPshot)



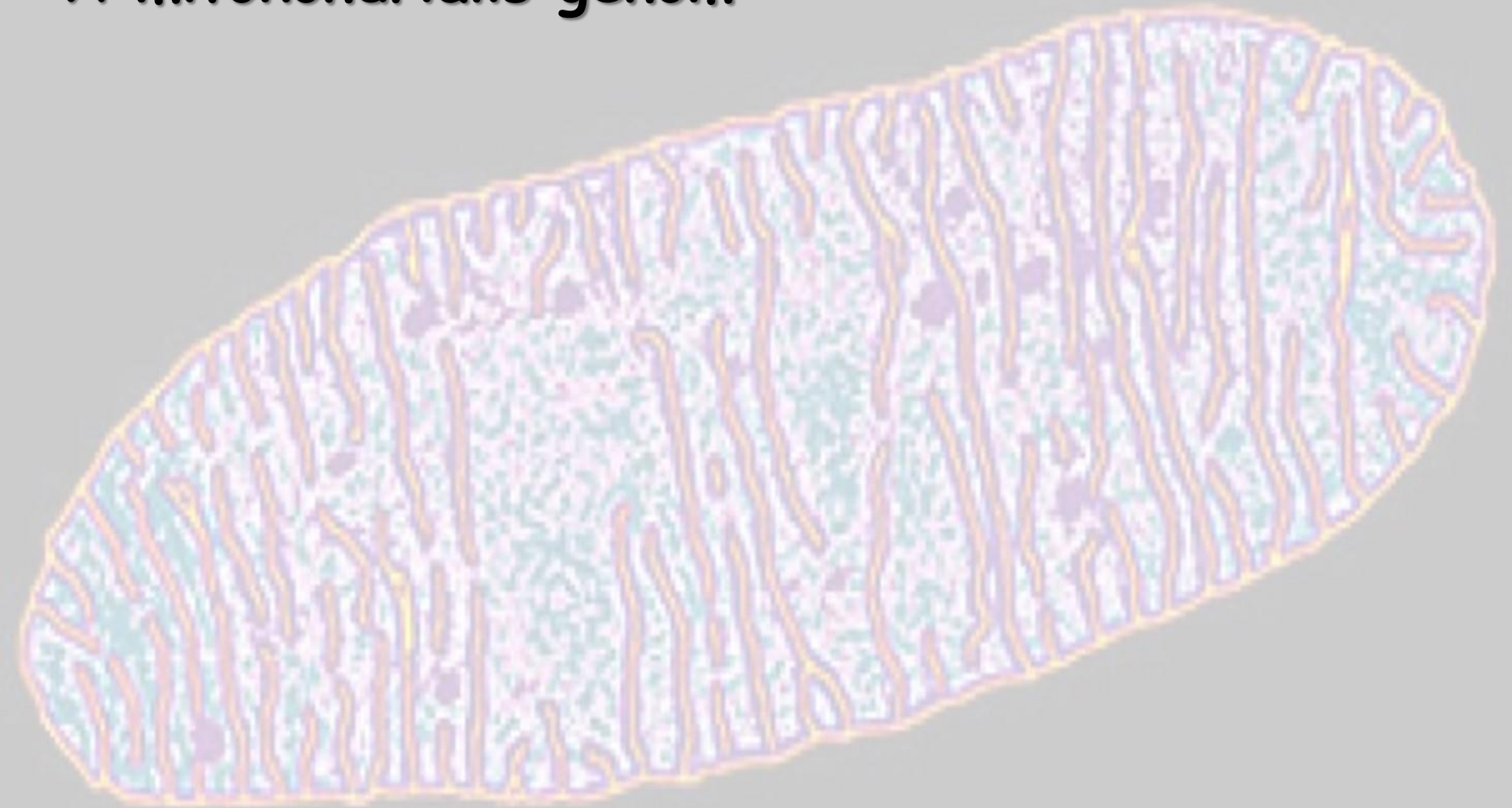
Sample	Self-reported pigmentary traits			rs12913832 HERC2	rs1805007 MC1R	rs1805008 MC1R	OCA2 diplotype ^a	rs16891982 SLC24A2	rs1426654 SLC24A5	rs2031526 DCT	rs1545397 OCA2	Inferred ancestry of individuals ^b		
	Eye color	Hair color	Skin color									European	Asian	African
E1	Blue	Red	Fair	<u>G/G</u>	C/C	C/T	<u>TGT/TGT</u>	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	<u>TGT/TGT</u>	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	<u>TGT/TGT</u>	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	<u>TGT/TGT</u>	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	<u>TGT/TGT</u>	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	<u>TGT/TGT</u>	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	<u>TGT/TGT</u>	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	<u>TGT/CTC</u>	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	<u>TTT/CTC</u>	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	<u>TGT/TGT</u>	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	<u>TGT/TGC</u>	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	<u>TGT/TTT</u>	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	<u>TGT/TTT</u>	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	<u>TGT/TGT</u>	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	<u>TGT/TGT</u>	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	<u>TGT/CTC</u>	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	<u>TGT/TTT</u>	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	<u>TGT/TGT</u>	G/G	A/A	G/G	A/T	0.935	0.045	0.021
E19	Brown	Red	Fair	A/G	C/T	C/C	<u>TGT/TGT</u>	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	<u>TGT/TGT</u>	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	<u>TGT/TGT</u>	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	<u>TGT/TGT</u>	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	<u>TGT/TTT</u>	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	<u>TGT/TGT</u>	C/G	A/A	G/G	A/A	0.786	0.049	0.166
E25	Brown	Red	Fair	A/G	C/C	T/T	<u>TGT/TGC</u>	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	<u>TGT/TGT</u>	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	<u>TGT/TGT</u>	G/G	A/A	G/G	A/A	0.958	0.014	0.028
Af1	Brown	Black	Dark	A/A	C/C	C/C	<u>TGC/TTC</u>	C/C	G/G	A/G	A/A	0.028	0.094	0.878
Af2	Brown	Black	Dark	A/A	C/C	C/C	<u>TGC/TTC</u>	C/C	G/G	G/G	A/A	0.023	0.031	0.946
Af3	Brown	Black	Dark	A/A	C/C	C/C	<u>TGC/TTC</u>	C/C	A/G	G/G	A/A	0.164	0.041	0.795
As1	-	-	-	A/A	C/C	C/C	<u>TTT/CTC</u>	C/C	G/G	A/G	A/T	0.042	0.649	0.308
As2	-	-	-	A/A	C/C	C/C	<u>CTC/CTC</u>	C/C	G/G	A/G	T/T	0.020	0.921	0.060
As3	-	-	-	A/A	C/C	C/C	<u>CTC/CTC</u>	C/C	G/G	A/A	T/T	0.013	0.964	0.023
As4	-	-	-	A/G	C/C	C/C	<u>TTT/CGC</u>	C/C	A/G	A/A	A/T	0.212	0.708	0.080
As5	-	-	-	A/A	C/C	C/C	<u>TTC/CGC</u>	C/C	G/G	A/G	T/T	0.019	0.922	0.059
As6	-	-	-	A/A	C/C	C/C	<u>CTC/CTC</u>	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

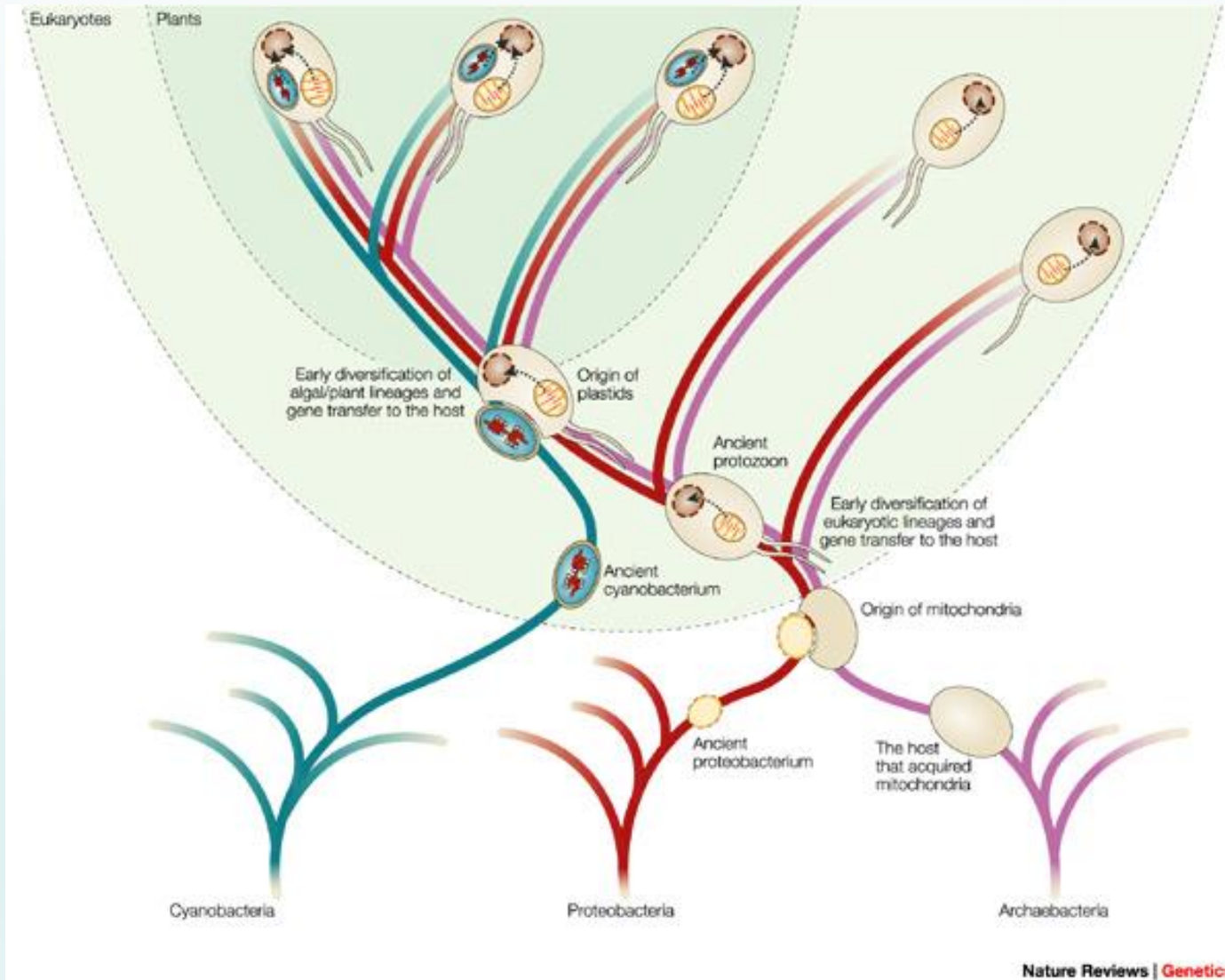
^b Probability of being from European/Asian/African population determined using the STRUCTURE program. The greatest probability, most likely estimate of ancestry, is indicated in bold

A mitokondriális genom



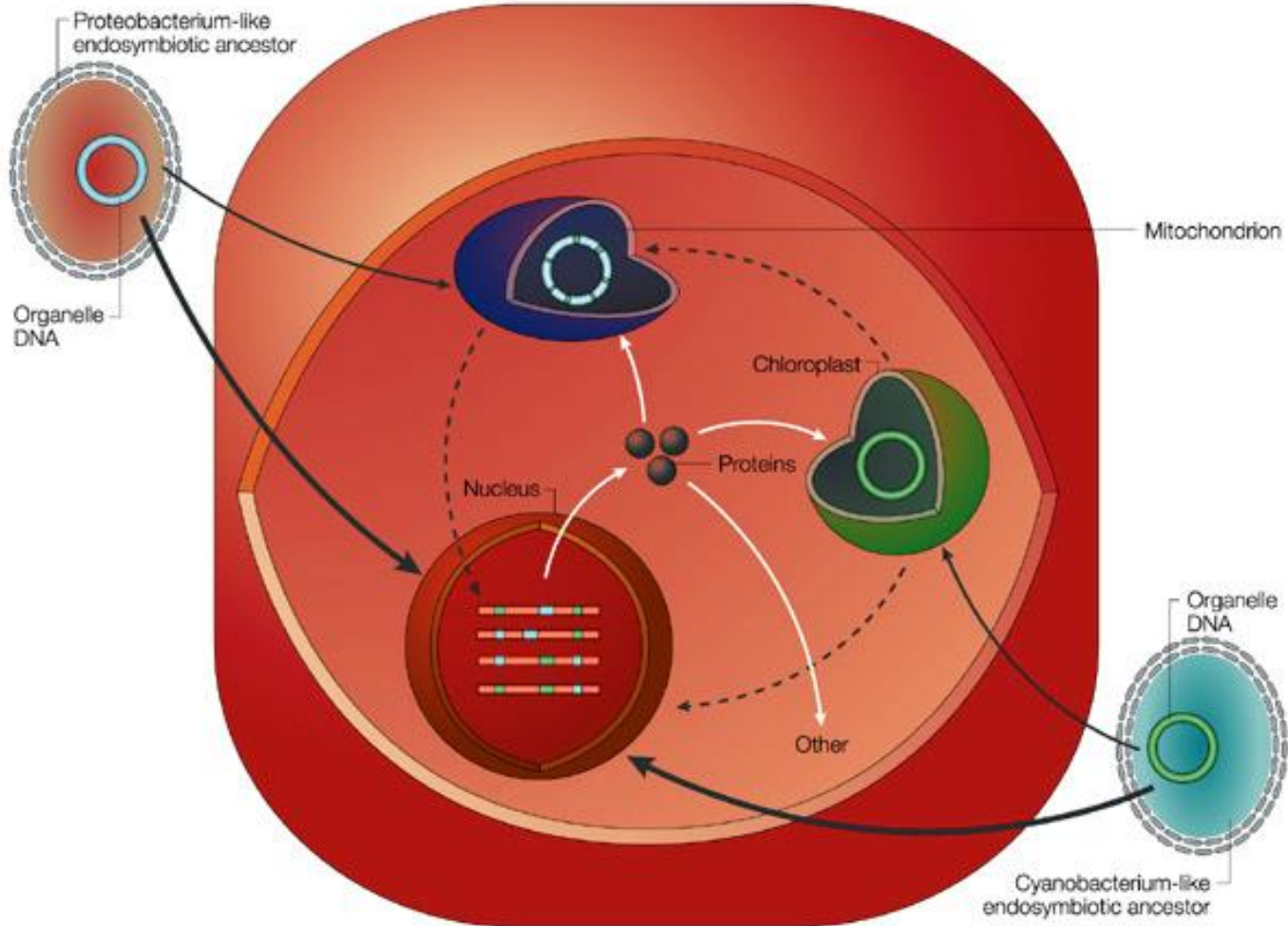
**Genomredukció endoszimbiotikus
gén-transzferrel**

Endoszimbiotikus gén-transzfer



(Timmis et al., 2004, *Nat Rev Gen*)

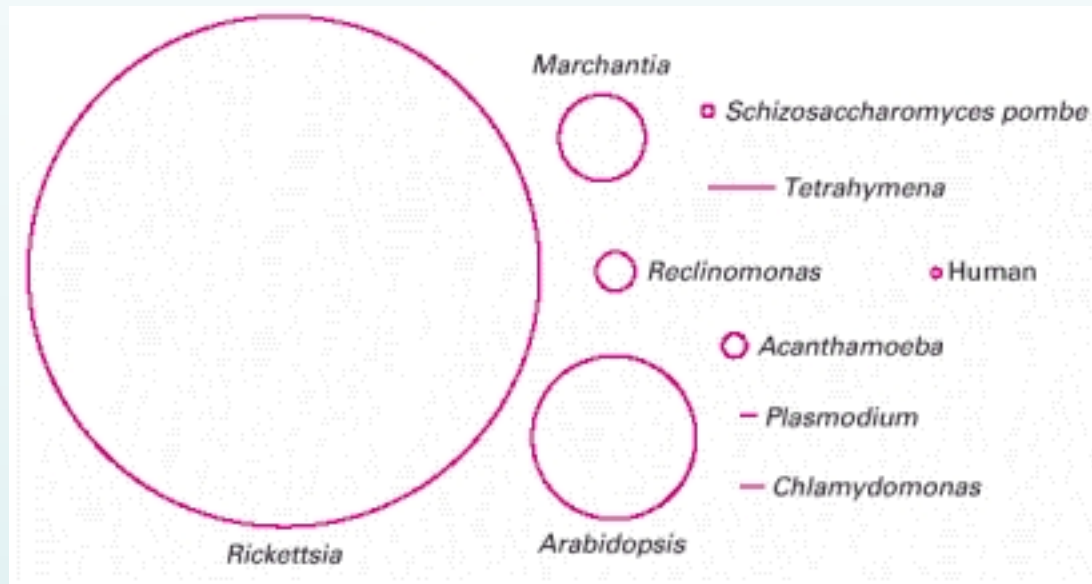
Endoszimbiotikus gén-transzfer



Nature Reviews | Genetics

(Timmis et al., 2004, *Nat Rev Gen*)

A mitokondriális genom mérete különböző csoportokban



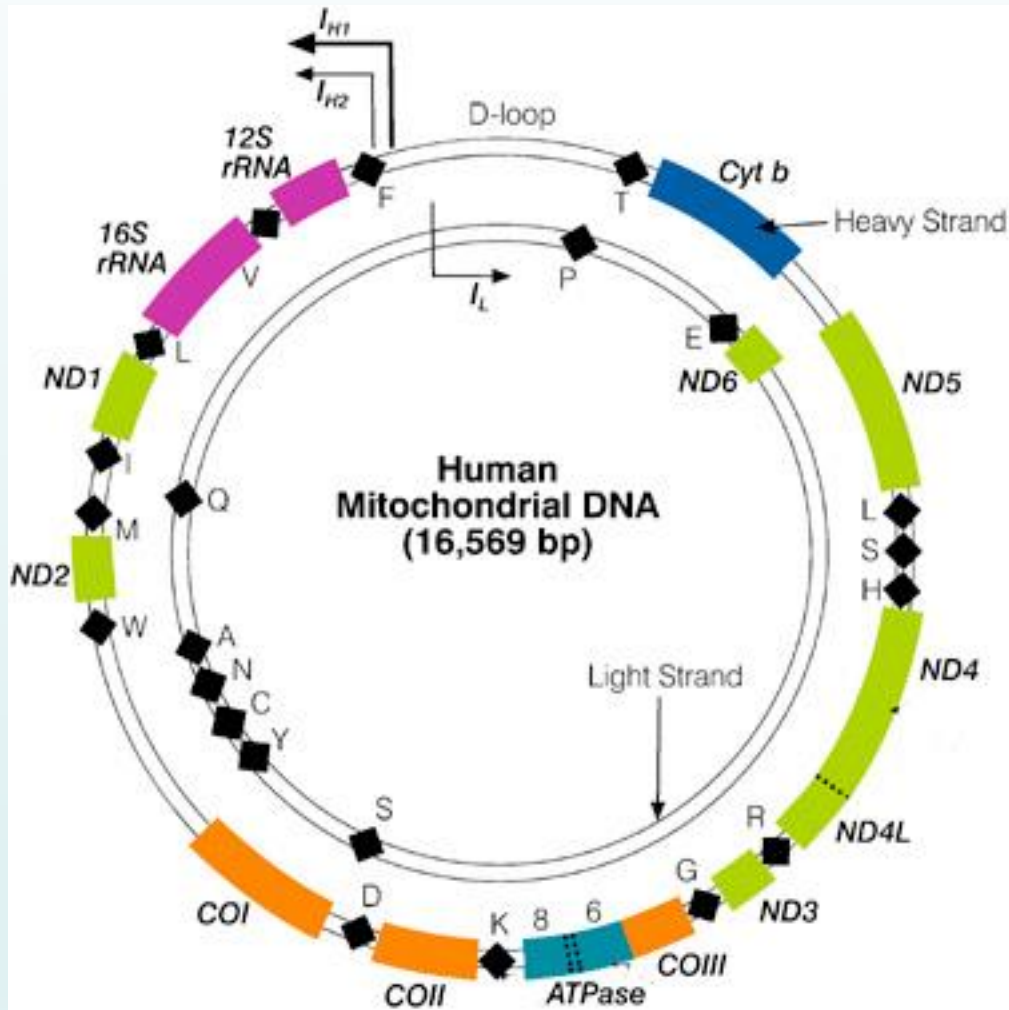
- a mitokondriális genomok mérete 6000 bp (*Plasmodium falciparum*) és 300,000 bp (egyres növények) között változik

- többségük cirkuláris, de akad lineáris is

- az állatok (Eumetazoa) mtDNS mérete eléggé stabil, kb. 16,500 bp

- a referenciaként használt *Rickettsia* genom kb 1.1 millió bp hosszú

A humán (emlős) mitokondriális genom



- dupla szálú, cirkuláris DNS molekula, ami mindkét szálán kódol:
H - „heavy” és
L - „light” szálak

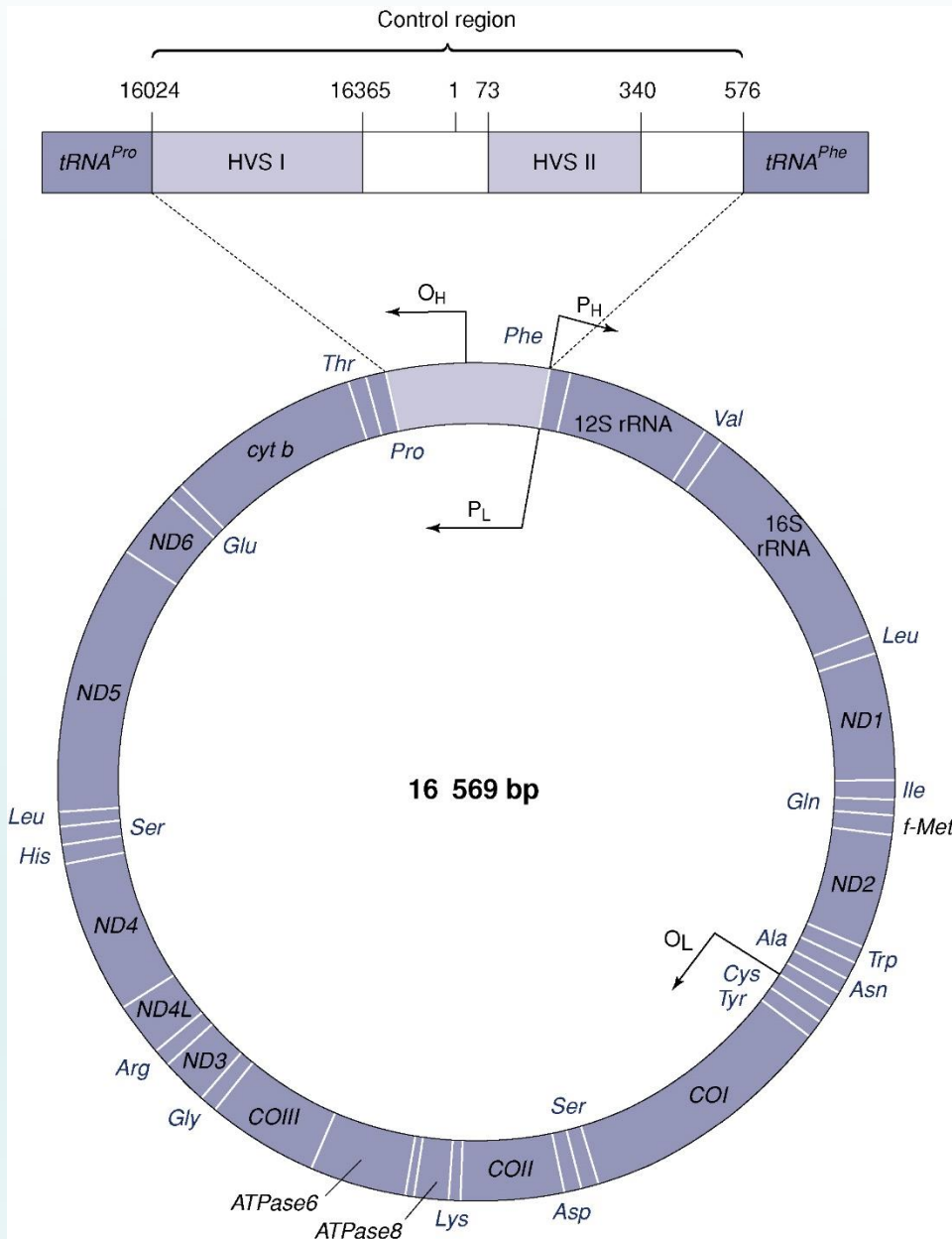
- emberek esetében 16,569 bp hosszú

- 37 gént kódol

- 13 az oxidatív foszforilációban játszik szerepet, a többi tRNS és rRNS

- minden mitokondriális mátrixban több kópia található

A humán mitokondriális genom



Sejten belüli nagy kópiaszám (~1000)

Maternális öröklésmentet:

- Rekombináció hiánya

Emelkedett mutációs ráta:

- DNA védelem és repair

Kódoló Szakaszok: 37 gén

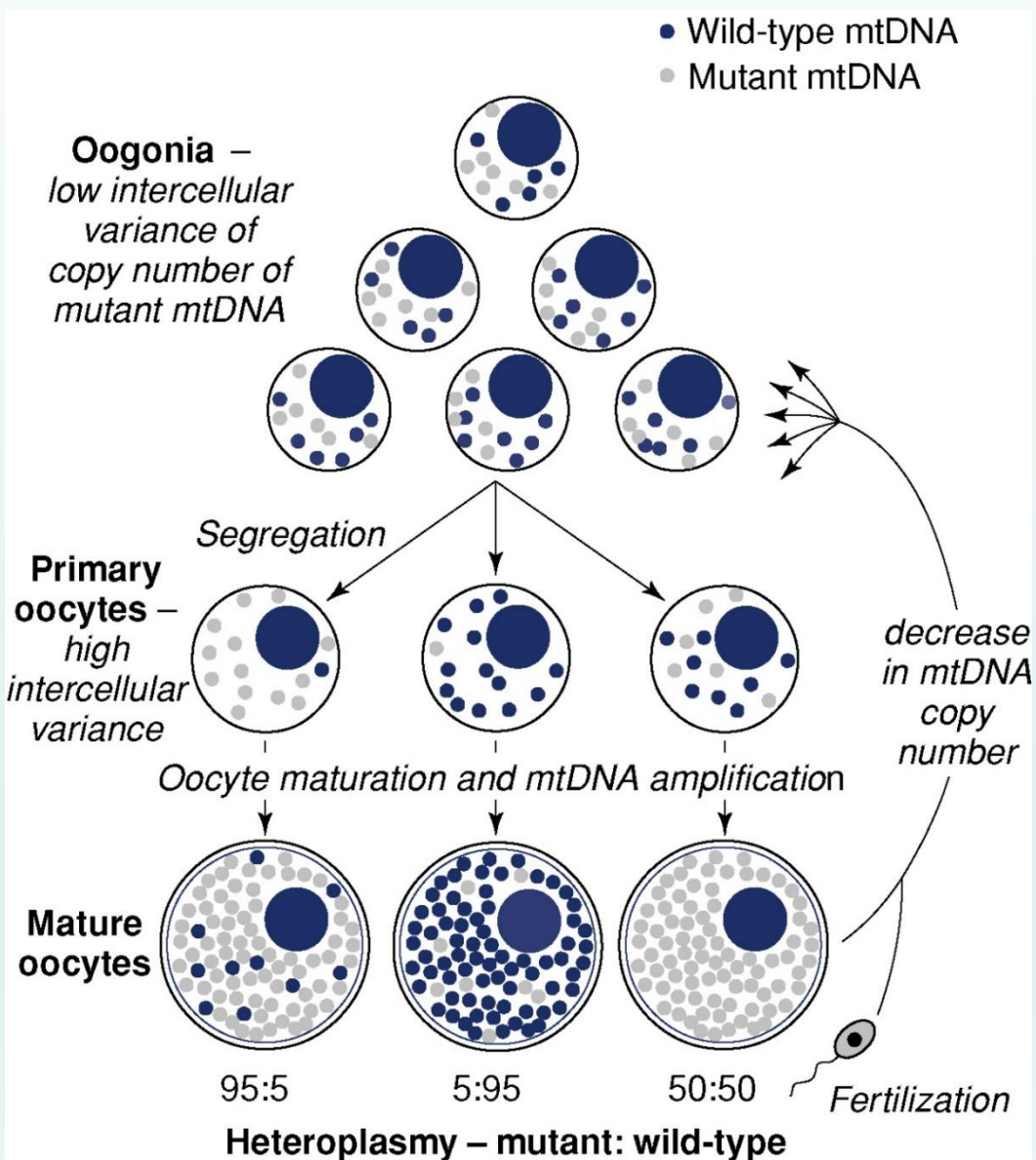
Kontroll Régió v. D-loop (~1120 bp):

- Hipervariábilis szakaszok (HVS1, HVS2)

Cambridge referencia szekvencia

Heteroplazmia, mutációs hot-spots

Mitokondriális DNS mutációk: heteroplazmia



Deléció: mitokondriális eredetű betegségek (anyai)

- MITOMAP

Szubsztitúció: ált. neutrális

Mutációs ráta: non-uniform

$10^{-6} - 10^{-7}$ / bp / generation

Mutációs hot-spotok

Citoplazmikus szegregáció:

„bottleneck”
(palacknyak effektus)

Heteroplazmia - homoplazmia:
szövet- és módszer specifikus

Heteroplazmia detektálása

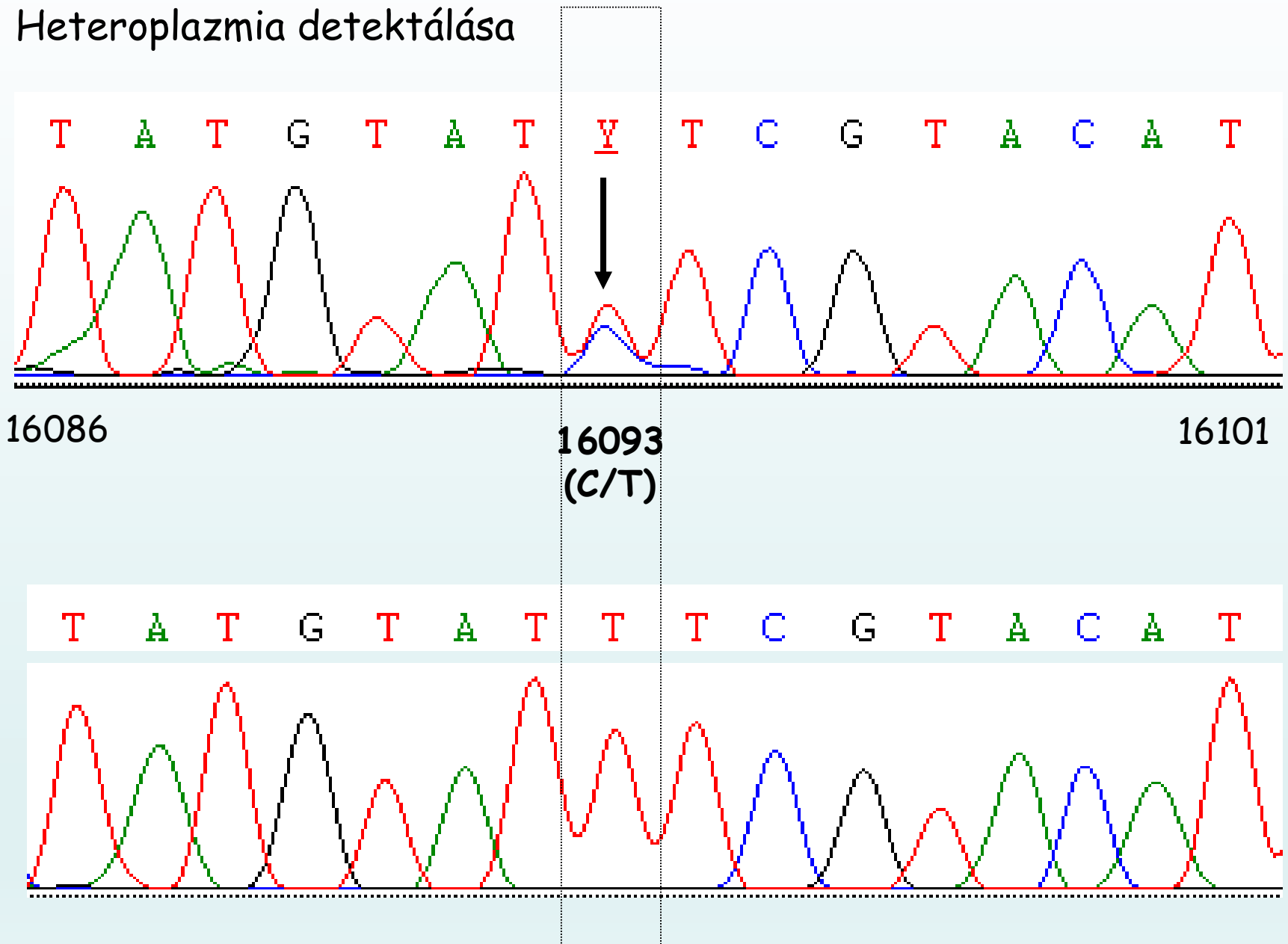


Figure 10.9, J.M. Butler (2005) *Forensic DNA Typing*, 2nd Edition © 2005 Elsevier Academic Press

Mitokondriális DNS haplotípusok

A vizsgált minták mtDNS szekvenciáját a referenciaként használt rCRS-hez illesztjük (jelen ábrán: 16071-16140)

	16080	16100	16110	16120	16130	16140
rCRS	ACCGCTATGT	ATTCGGTACA	TTACTGCCAG	CCACCATGAA	TATTGTACAG	TACCATAAAT
Q	ACCGCTATGT	ATTCGGTACA	TTACTGCCAG	CCACCATGAA	TATTGTACAG	TACCATAAAT
K	ACCGCTATGT	ATTCGGTACA	TTACTGCCAG	CCACCATGAA	TATTGTACAG	TACCATAAAT

16093

16129

A referencia szekvenciától való eltérések pozícióját és nukleotidját jegyezzük le, ami megadja az adott minta mtDNS haplotípusát

Q

16093C

16129A

K

16093C

16129A

Identification of the remains of the Romanov family by DNA analysis

Peter Gill¹, Pavel L. Ivanov², Colin Kimpton¹, Romelle Piercy¹, Nicola Benson¹, Gillian Tully¹, Ian Evett¹, Erika Hagelberg³ & Kevin Sullivan¹

Nine skeletons found in a shallow grave in Ekaterinburg, Russia, in July 1991, were tentatively identified by Russian forensic authorities as the remains of the last Tsar, Tsarina, three of their five children, the Royal Physician and three servants. We have performed DNA based sex testing and short tandem repeat (STR) analysis and confirm that a family group was present in the grave. Analysis of mitochondrial (mt) DNA reveals an exact sequence match between the putative Tsarina and the three children with a living maternal relative. Amplified mtDNA extracted from the remains of the putative Tsarina has been cloned to demonstrate heteroplasmy at a single base within the mtDNA control region. One of these sequences matches two living maternal relatives of the Tsar. We conclude that the DNA evidence supports the hypothesis that the remains are those of the Romanov family.



A Romanovok maradványainak azonosítása molekuláris genetikai módszerrel

