Genetika I. - Genomika

Hogyan változtak az elképzelések a genom tartalmáról, a szerveződési komplexitás és génszám közti összefüggésről?

Genomok szerkezete és variabilitása

ELTE TTK Genetikai Tanszék

2018.09.25.

Humán genom: „... valahol a csirke és a szőlő között?"

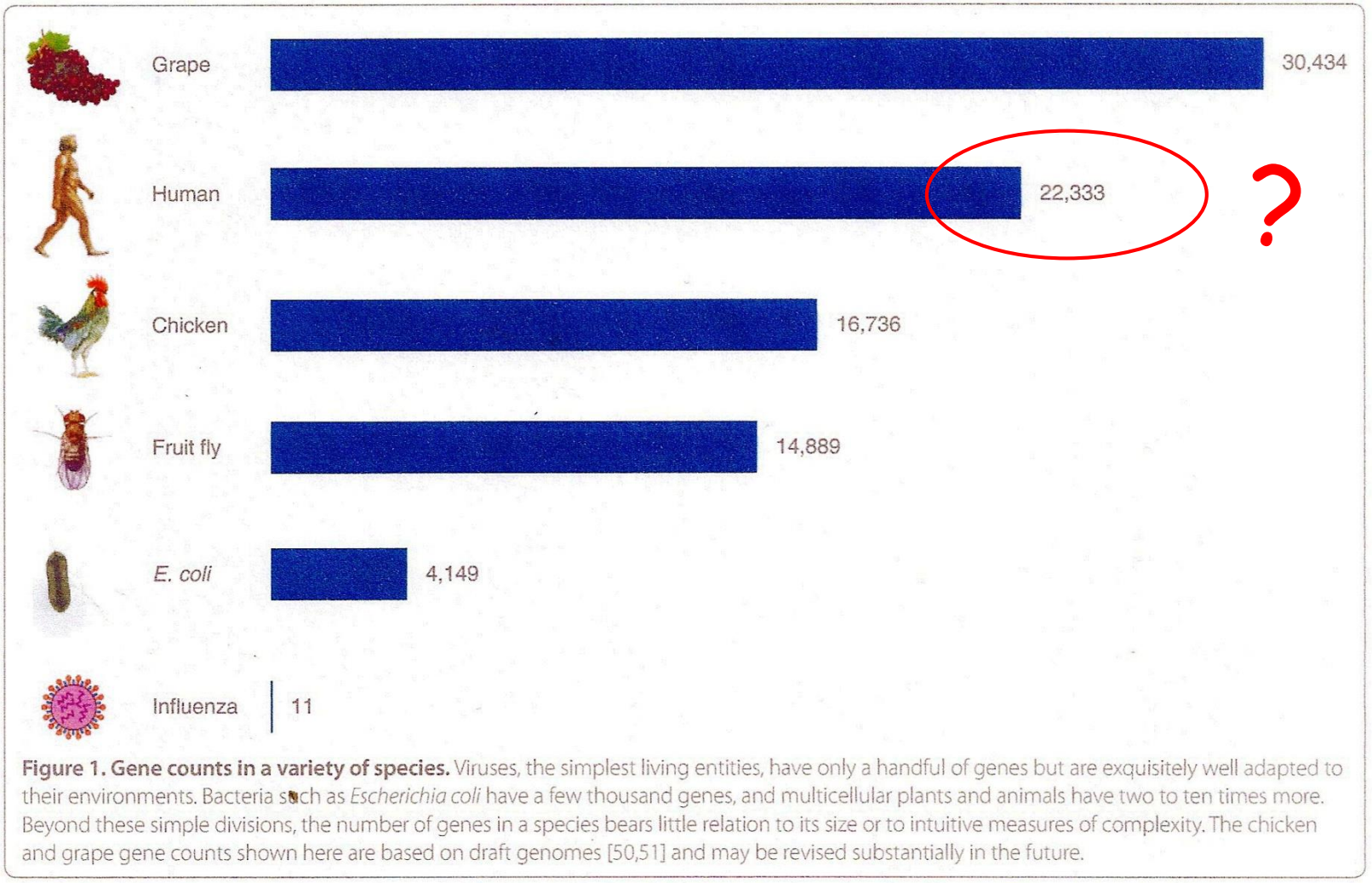


Figure 1. Gene counts in a variety of species. Viruses, the simplest living entities, have only a handful of genes but are exquisitely well adapted to their environments. Bacteria such as *Escherichia coli* have a few thousand genes, and multicellular plants and animals have two to ten times more. Beyond these simple divisions, the number of genes in a species bears little relation to its size or to intuitive measures of complexity. The chicken and grape gene counts shown here are based on draft genomes [50,51] and may be revised substantially in the future.

(Pertea & Salzberg, Genome Biology 2010, 11:206)

Első becslések a genom méretéről és a gének számáról

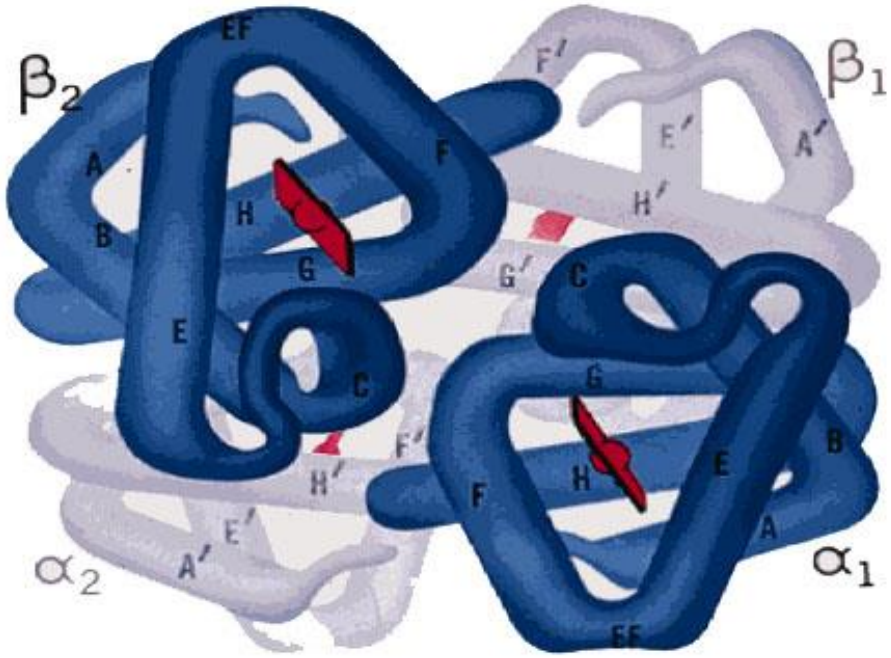
1964: F. Vogel (Heidelberg)

-Hemoglobin α és β lánc

-leegyszerűsített feltevések

-Haploid genom: 3×10^9 bp

-Gének száma: 6,7 millió !



1990: NIH/DOE report on Human Genome Project

- becslés: 100 000 gén, az átlagos gén méret (30 000 bp) alapján

2001, Human Genome Project: csökkenő génszám, növekvő bizonytalanság

A génfogalom fejlődése: mit nevezünk ma egy génnek?

A Gén fogalmának jelentős átalakulása az elmúlt száz esztendőben: protein/RNS kódolás, intron/exon fogalom, szabályozó funkciók, stb.

Gén
- 1950-es évektől
- a kódolási szabályok felismerése
- a DNS azon szakasza, amely egy transzkriptum (mRNS, rRNS, snRNS, tRNS, …) átírásáért felelős genetikai információt tartalmazza

ORF
- *open reading frame* (nyitott leolvasási keret)
- 1990-es évektől (genomika korszaka)
- gének annotálása: bioinformatikai módszerekkel prediktálnak transzkriptumok átírását végző DNS szakaszokat (konzervált szekvencia motívumok alapján)

Egy gén a genetikai állományunk jól körülhatárolt szakasza, mely mRNS-ként átíródik és egy v. több fehérjét kódol. (pl. alternatív splicing: izoformák)

Human Genom Project (HUGO)

- Első kezdeményezés: 1980-as évek eleje
- orvosbiológiai megközelítés, infrastruktúrális beruházások
- Folyamatban lévő genom szekvenálási projektek
- λ-fág, SV40, humán mitokondriális genom
- Genetikai és fizikai térképezések
- Botstein et al., 1980; Sulston et al., 1986;
- DNS-szekvenálási technológia és bioinformatika
- shotgun sequencing, automated sequencing, **ESTs, STSs,**
- NRC Report 1988, US DOE, NIH,
- genetikai és fizikai térkép, parallel projektek modell organizmusokon, technológiai fejlesztések, bioetika

Universal Landmark

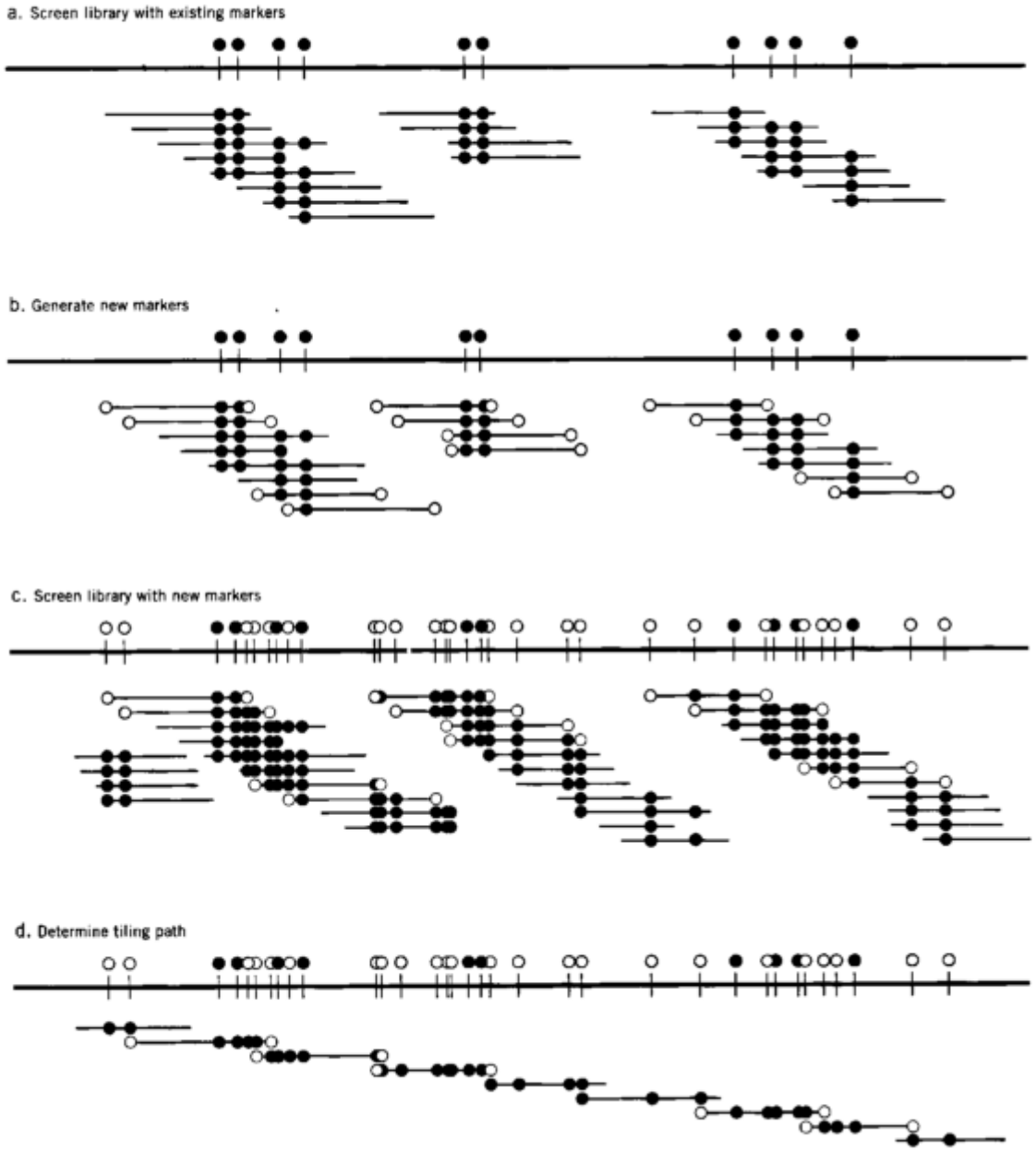
Sequence Tagged Site (STS) 1989

Replaces cloned DNA probe mapping landmarks with PCR assays.

Each STS is uniquely described by a pair of oligonucleotides, a product size, and PCR reaction conditions. Can be stored and distributed electronically.

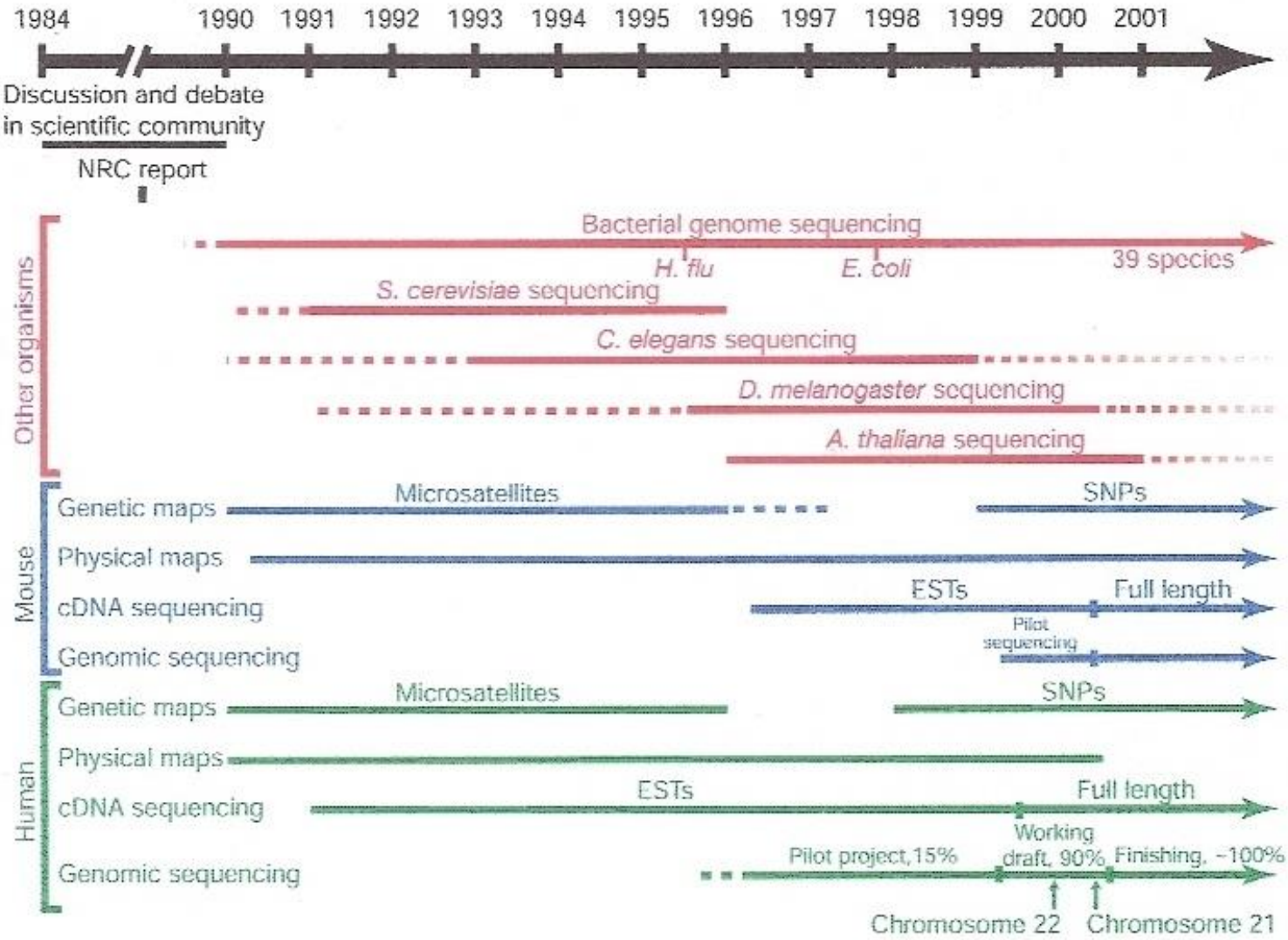
Enables merging of mapping data obtained from many labs using many different methods into a single consensus map of landmarks along a chromosome.

Eliminates the need for huge collections of cloned probe segments upon which prior maps depended.



Clone ends – Clone-based Physical Map

Genom projektek időskálán – a „hőskor”



Humán Genom Projekt
- résztvevők és módszerek

- **HUGO**: Human Genome Organization
- US DOE és NIH, UK MRC és WTSI, CEPH , FMDA, Japán, Európai Közösség (élesztő genom), Németország, Kína
- 1990-1995: genetikai és fizikai térképezés
- betegség gének, fizikai pontok fixálása, modell szervezetek
- large-scale sequencing: két fázisú „shotgun" szekvenálás
- 2001: draft genom szekvencia, 2003: teljes genom szekvencia

- **Celera Genomics**:
- Applied Biosystems., TIGR (C. Venter)
- 1998-2001: „whole genome shotgun"
- ABI PRISM 3700 DNA Analyzer

Technology speeds science. ABI sequencers at Venter Insitute, 2007.

„shotgun" genom szekvenálási stratégiák

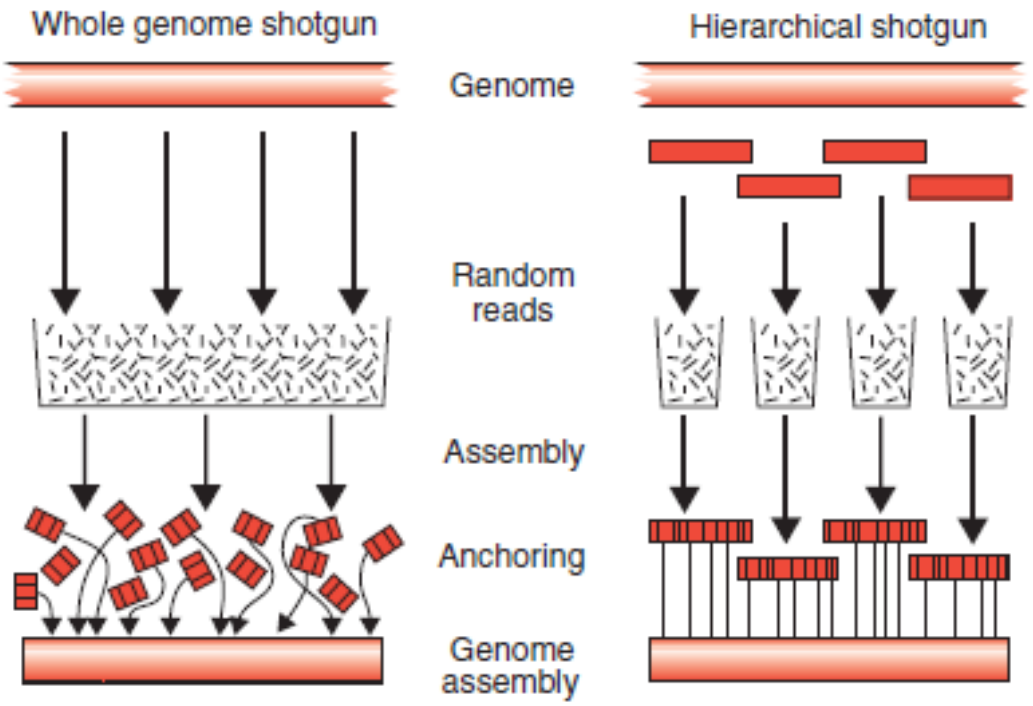
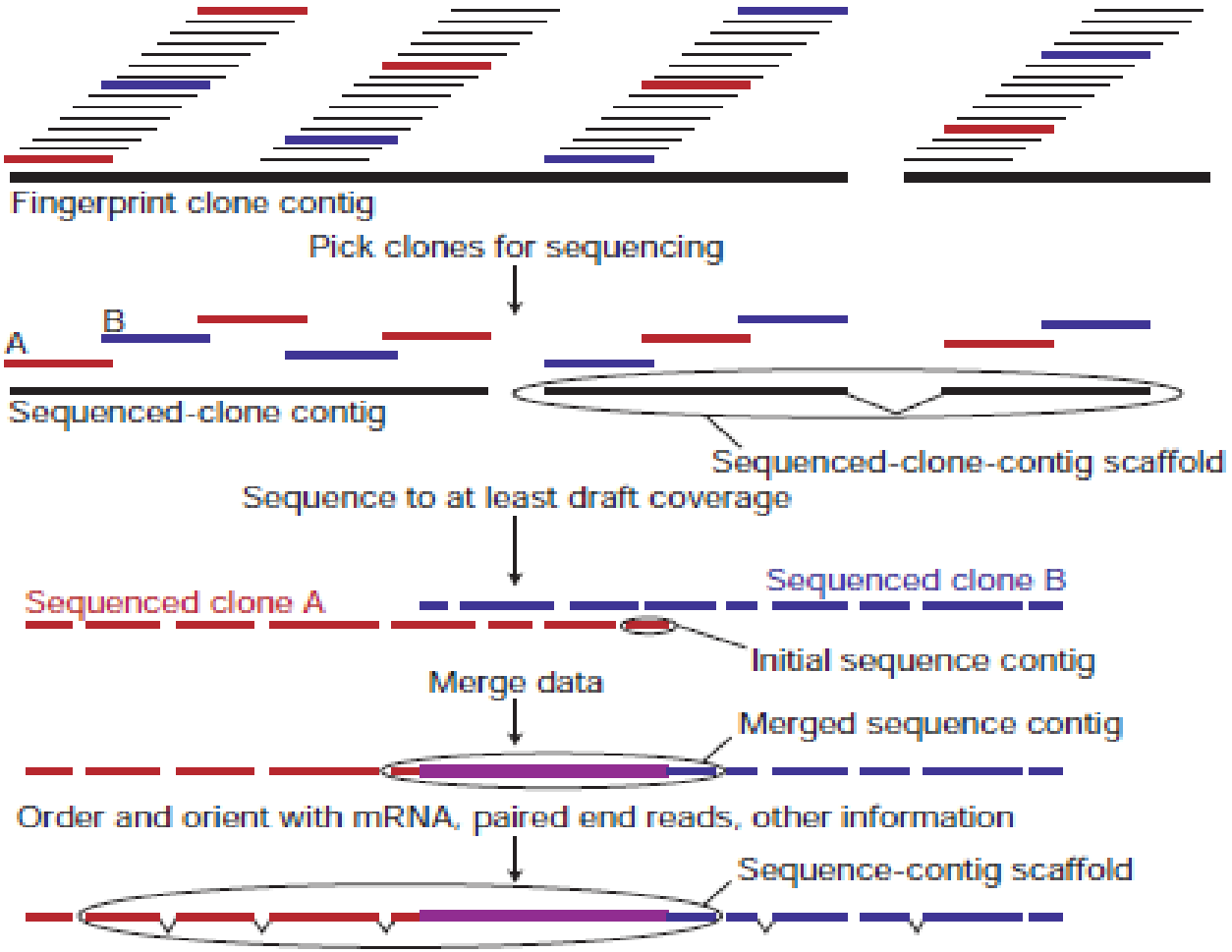


Figure 9.11. Assembling genomic data using the hierarchical and whole genome shotgun approaches. Adapted from Waterston, Lander and Sulston (2002), with permission

RJ Reece: Analysis of Genes and Genomes, 2004

Genom szekvenciaváz összeállítása



International Human Genome Sequencing Consortium: Initial sequencing and analysis of the human genome, Nature 409, 860 (2001)

Teljes genom összeszerelés

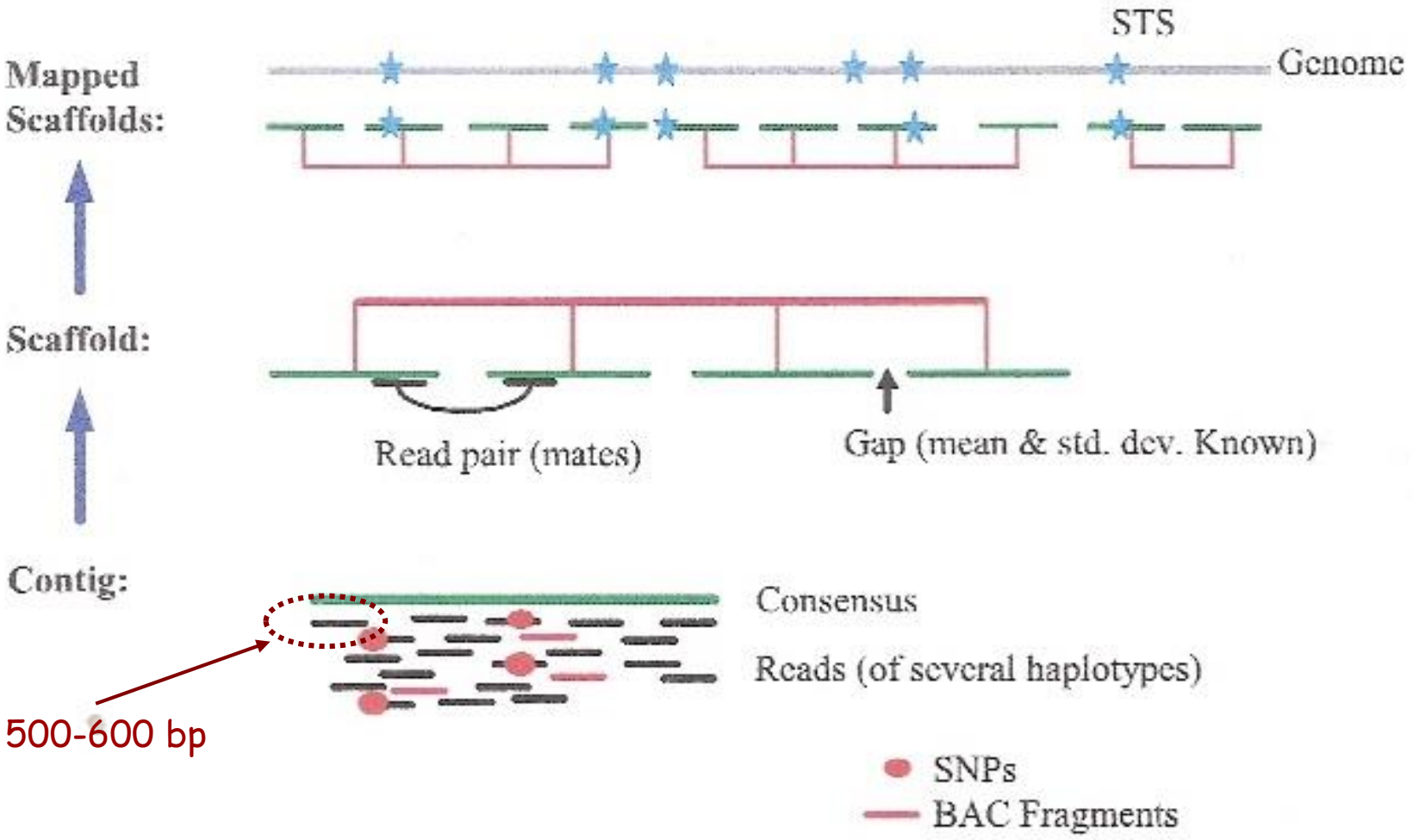


Fig. 3. Anatomy of whole-genome assembly. Overlapping shredded bactig fragments (red lines) and internally derived reads from five different individuals (black lines) are combined to produce a contig and a consensus sequence (green line). Contigs are connected into scaffolds (red) by using mate pair information. Scaffolds are then mapped to the genome (gray line) with STS (blue star) physical map information.

JC Venter, et al.: The Sequence of the Human Genome, Science 291, 1304 (2001)

STS genom térképezés

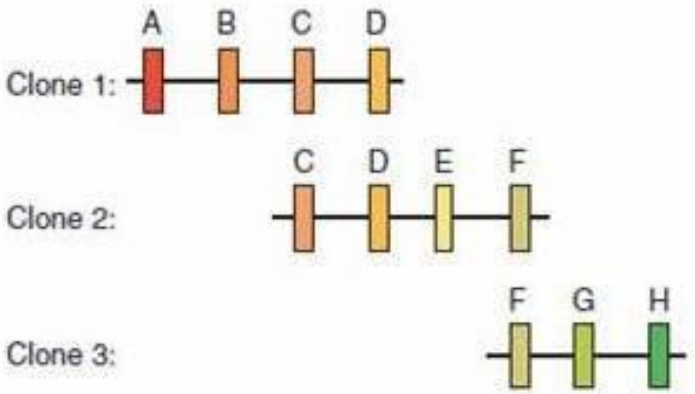


Figure 9.5. Aligning clones by STS mapping. Each clone contains several STSs. Clone 1 has four (A, B, C and D). Clone 2 also contains STSs C and D. Therefore clones 1 and 2 overlap with each other

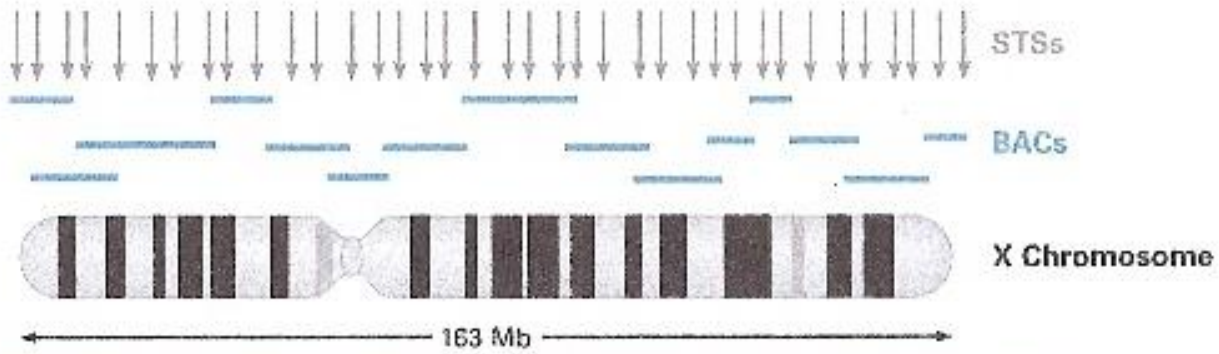
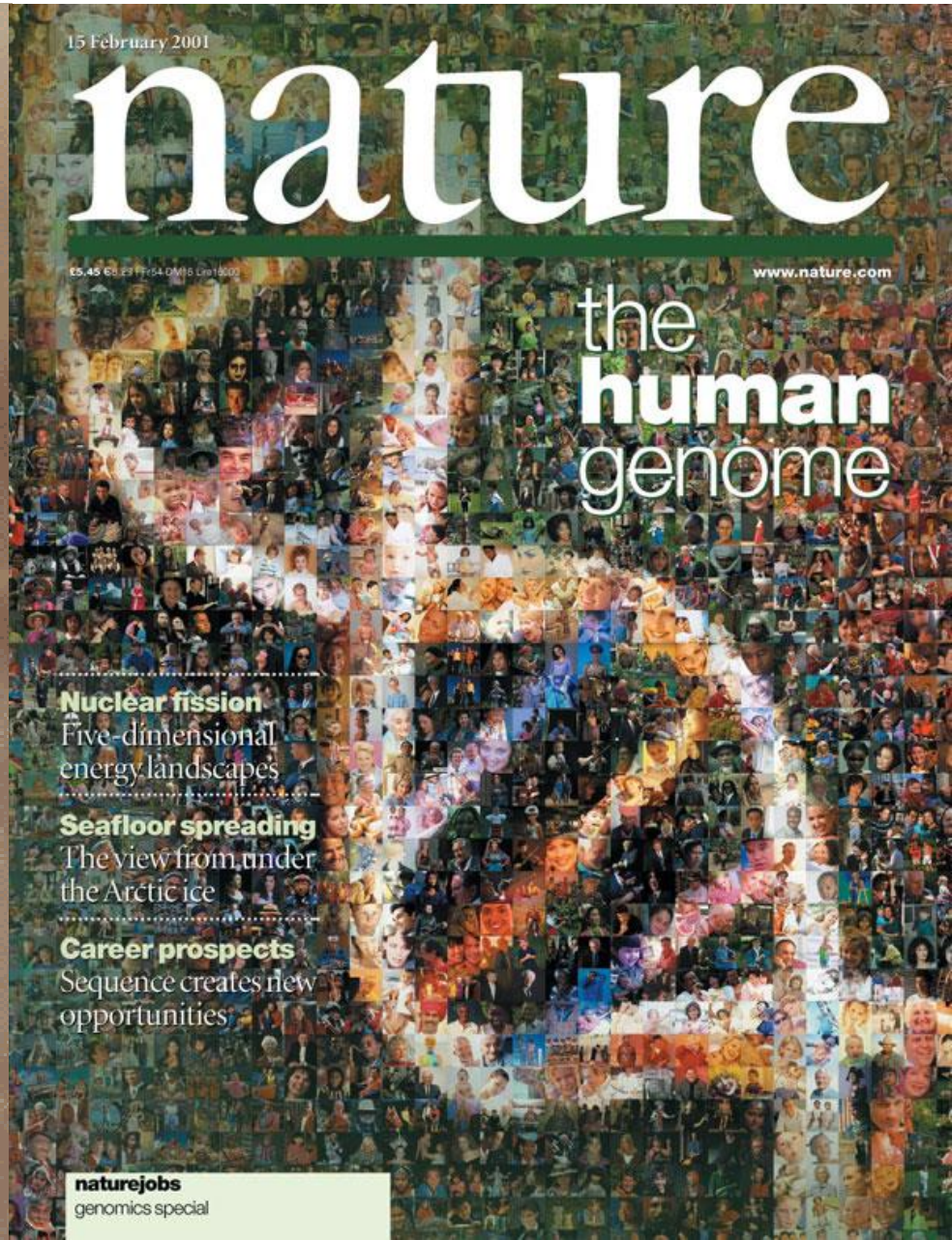


FIGURE 1.3 • Relationships of chromosomes to genome sequencing markers. The X chromosome is about 163 Mb in length. In this diagram, there are 16 overlapping BAC clones that span the entire length. In reality, 1,408 BACs were needed to span the X chromosome. Arrows (top) mark STSs scattered throughout the chromosome and on overlapping BACs.

Humán Genom Projekt



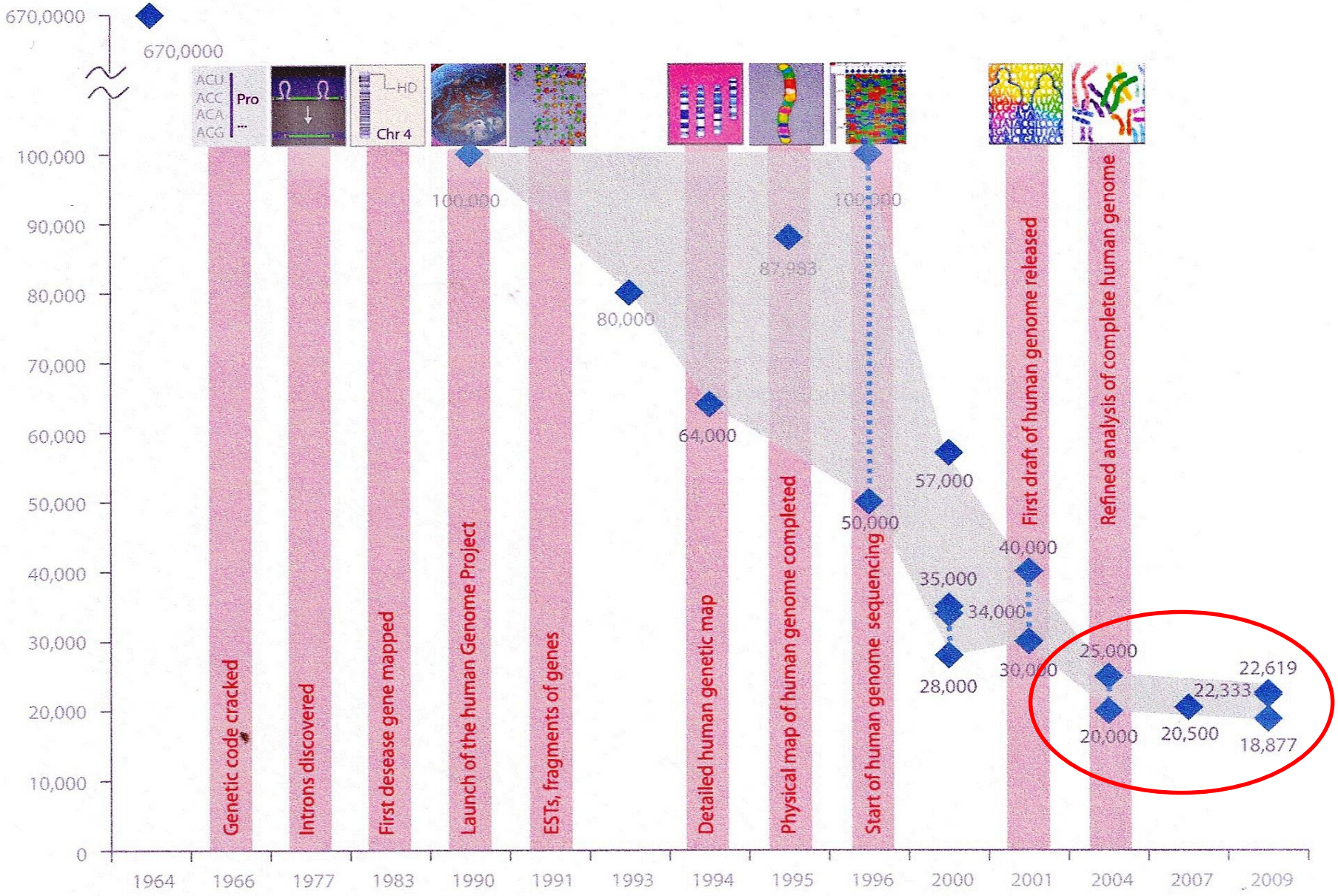


Figure 2. The trend of human gene number counts together with human genome-related milestones. Individual estimates of the human gene count are shown as blue diamonds. The range of estimates at different times is shown by the two vertical blue dotted lines. Note how this range has narrowed in recent years.

Hol tartunk most?

2001, Human Genome Consortium: 30 000 – 40 000 protein kódoló gén

Celera Consortium: 26 500 „erős" + 12 000 „gyenge" bizonyíték

2004, Human Genome Consortium: 20 000 – 25 000 gén

- kevesebb mint az Arabidopsis → szervezeti komplexitás?

2010, Ensembl: 22 619 / NCBI: 22 333 protein kódoló gén

CCDS: 18 173 (http://www.ncbi.nlm.nih.gov/CCDS/CcdsBrowse.cgi)

fals pozitívak: retrotranszpozonok, pszeudogének, „orphan" DNS

2016.09.12.: CCDS: 18 892 genes > 1 CCDS ID: 7 576

Új gének és gén átrendeződések

- CGH analízisek: rokon fajok között kb. azonos génszám
- *de novo* gén keletkezés: génduplikáció és specializáció
- génszám eltérések egyének között: segmental duplications
- large-scale copy number polymorphisms (CNVs)
- emberi „pángenom": változatok rasszok, csoportok között.

(Li R, et al., 2010, Nat Biotechnol, 28:57-63)

- kb. 40 Mb új szekvencia, + 1,3 %
- *de novo* eredet: új humán gének? (Knowles and McLysaght, 2009)

Emerging novel gene sequences

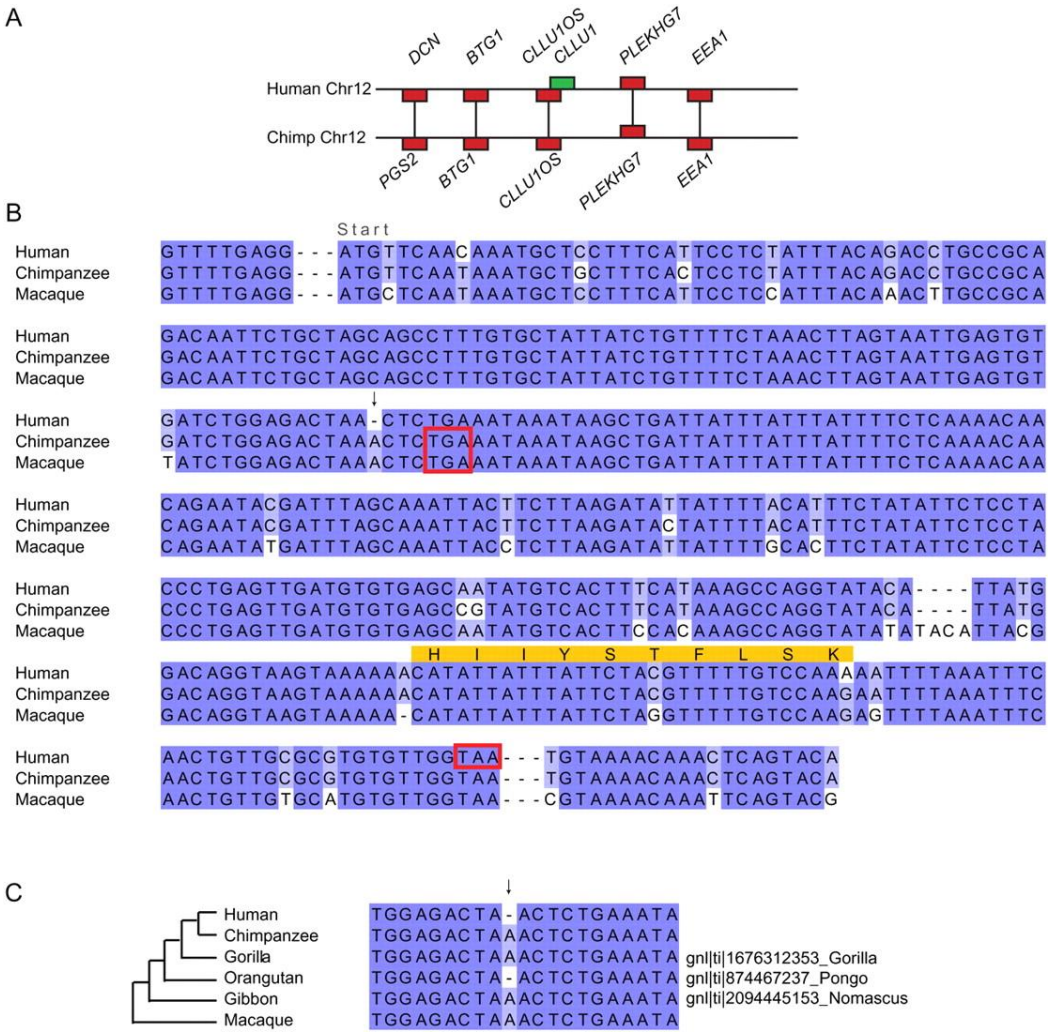
Table 1. Novel human protein-coding genes and supporting evidence.

Gene name	Ensembl ID	Length (codons)	Longest chimp ORF[a]	Expression support and tissue[b]	Primate shared disablers[c]	Other major sequence differences	Presence of enabler in other human complete genome sequences[d]	HapMap SNPs
CLLU1	ENSG00000205056	121	42	EST/cDNA: Blood (AJ845165, AJ845166); UniGene: Blood, embryonic tissue, eye, lymph, lymph node, muscle, pharynx, tonsil (Hs.339918)	1-bp indel[e]	Macaque: 4- and 1-bp indels	Sequence available and enabler conserved in all	1 syn.; 1 nonsyn.
C22orf45	ENSG00000178803	159	87 (25 amino acids align with human sequence)	EST/cDNA: Kidney, other (AX747284, AK091970, DA635985); ArrayExpress: Sperm, lung (E-GEOD-6872, E-GEOD-3020)	Premature stop codon	Chimp: 1-bp indel; Macaque: lacks ATG start codon; 4-bp indel	Reverse strand is available and conserved in Venter	1 nonsyn.
DNAH10OS	ENSG00000204626	163	90 (75 amino acids align with human sequence)	EST/cDNA: Hippocampus (AK127211); UniGene: Blood, embryonic tissue, eye, lymph, lymph node, muscle, pharynx, tonsil (Hs.339918)	10-bp indel	Chimp: 2- and 1-bp indels; Macaque: lacks ATG start codon; 13-, 8-, 1-, and 1-bp indels	Reverse strand is available and conserved in Venter, Watson and HuAA	1 syn.; 1 nonsyn.

[a]Length in codons of longest in-frame (alignable) ORF starting from any ATG in the region.
[b]Type of data/database is listed followed by tissue information with database identifiers in parentheses. Underlined accession numbers are full-length, spliced cDNA.
[c]Shared disablers are sequence differences shared by chimp, gorilla, orangutan, gibbon, and macaque that eliminate the capacity to produce a protein similar to the human protein.
[d]Independently sequenced whole genomes: Venter, Watson, HuAA, HuBB, HuCC, HuDD, and HuFF. All data are listed where available.
[e]Not shared with orangutan.

Knowles D G , McLysaght A Genome Res. 2009;19:1752-1759

Sequence changes in the origin of CLLU1 from noncoding DNA. (A) Region of conserved synteny between human and chimp chromosomes 12.



Knowles D G , McLysaght A Genome Res. 2009;19:1752-1759

TABLE 3.1 Approximate fractional composition of the human genome

TYPE OF DNA	FRACTION
Coding exons	0.008
Internal introns	0.308
5′ Untranslated regions	
Exons	0.045
Introns	0.002
3′ Untranslated regions	
Exons	0.006
Introns	0.001
Intergenic DNA	0.683
Conserved noncoding DNA	0.016
Pseudogenes	0.007
Mobile genetic elements	0.446

Note: Derived from various references given in the text. Intergenic DNA is all DNA except coding exons and internal introns. The fractions do not sum to one because mobile elements, pseudogenes, and transcription factor binding sites reside in introns, UTRs, and/or intergenic DNA.

The Origins of Genome Architecture, Lynch, 2006

TABLE 3.2 Haploid genome size, number of protein-coding genes, and average number of nucleotides per gene for some well-characterized eukaryotic genomes

	GENOME SIZE (MB)	GENE NUMBER	KILOBASES/GENE		
			TOTAL	CODING	NON-CODING
Unicellular species					
Encephalitozoon cuniculi	2.90	1997	1.45	1.01	0.44
Saccharomyces cerevisiae	12.05	6213	1.94	1.44	0.50
Schizosaccharomyces pombe	13.80	4824	2.86	1.43	1.43
Cyanidioschyzon merolae	16.52	5331	3.10	1.55	1.55
Cryptococcus neoformans	19.05	6572	2.89	1.62	1.27
Plasmodium falciparum	22.85	5268	4.34	2.29	2.05
Entamoeba histolytica	23.75	9938	2.39	1.14	1.25
Leishmania major	33.60	8600	3.91	2.15	1.76
Thalassiosira pseudonana	34.50	11242	3.07	0.99	2.08
Trypanosoma spp.	39.20	10000	3.92	1.96	1.96
Oligocellular species					
Ustilago maydis	19.68	6572	2.99	1.84	1.15
Aspergillus nidulans	30.07	9541	3.15	1.57	1.58
Dictyostelium discoideum	34.00	9000	3.78	2.45	1.33
Neurospora crassa	38.64	10082	3.83	1.44	2.39
Land plants					
Arabidopsis thaliana	125.00	25498	4.90	1.80	3.10
Oryza sativa	466.00	60256	7.73	1.18	6.55
Lotus japonicus	472.00	26000	18.15	1.35	16.80
Animals					
Caenorhabditis elegans	100.26	21200	4.73	1.25	3.48
Drosophila melanogaster	137.00	16000	8.56	1.66	6.90
Ciona intestinalis	156.00	16000	9.75	0.95	8.80
Anopheles gambiae	278.00	13683	20.32	1.64	18.68
Fugu rubripes	365.00	38000	9.61	0.93	8.68
Bombyx mori	428.70	18510	23.16	1.66	21.50
Gallus gallus	1050.00	21500	48.84	1.44	47.40
Mus musculus	2500.00	24000	83.33	1.30	82.03
Homo sapiens	2900.00	24000	96.67	1.33	95.36

Source: Lynch 2006a.

Gének száma

vs.

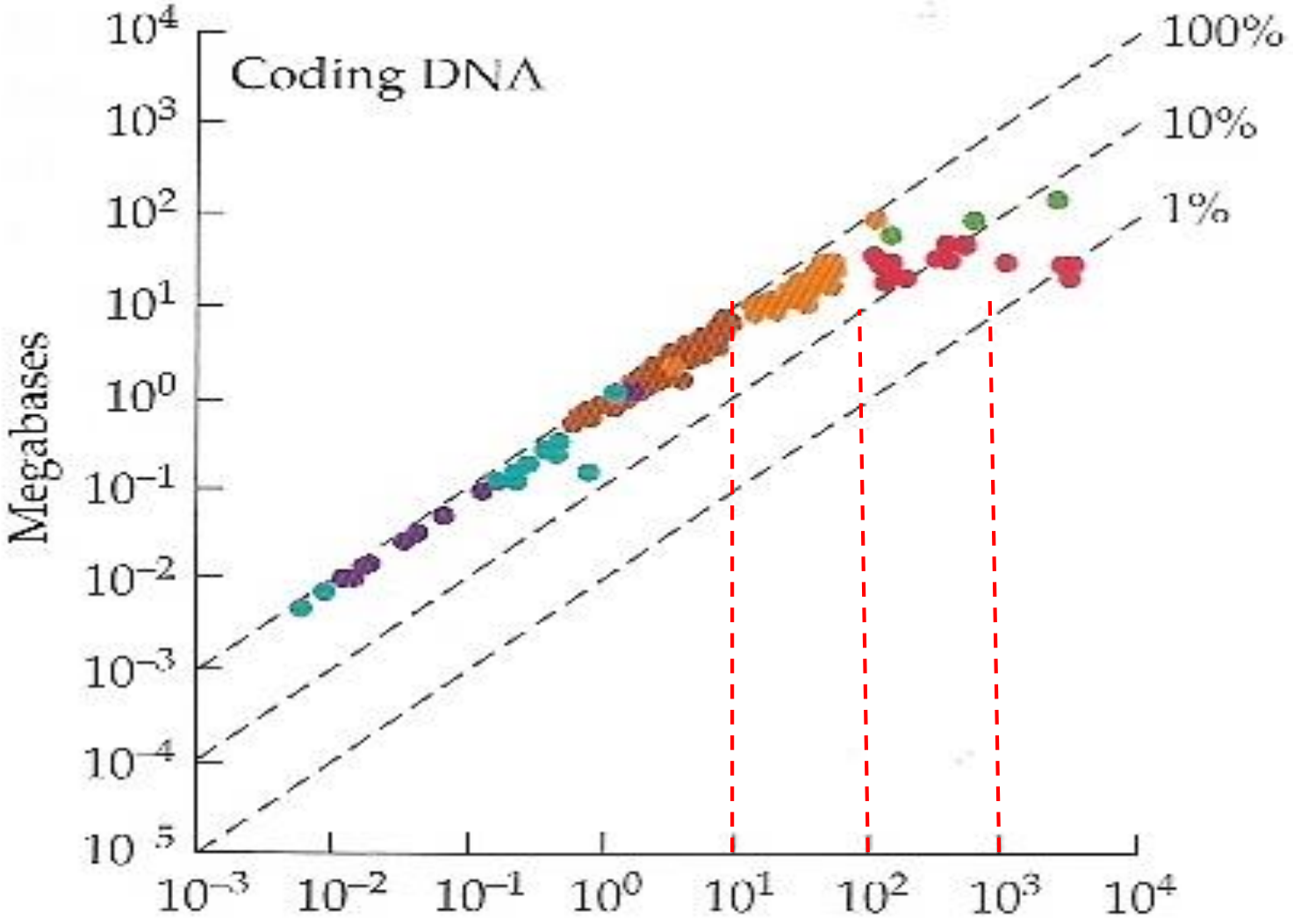
Kódoló szekvenciák hossza

Genom méret

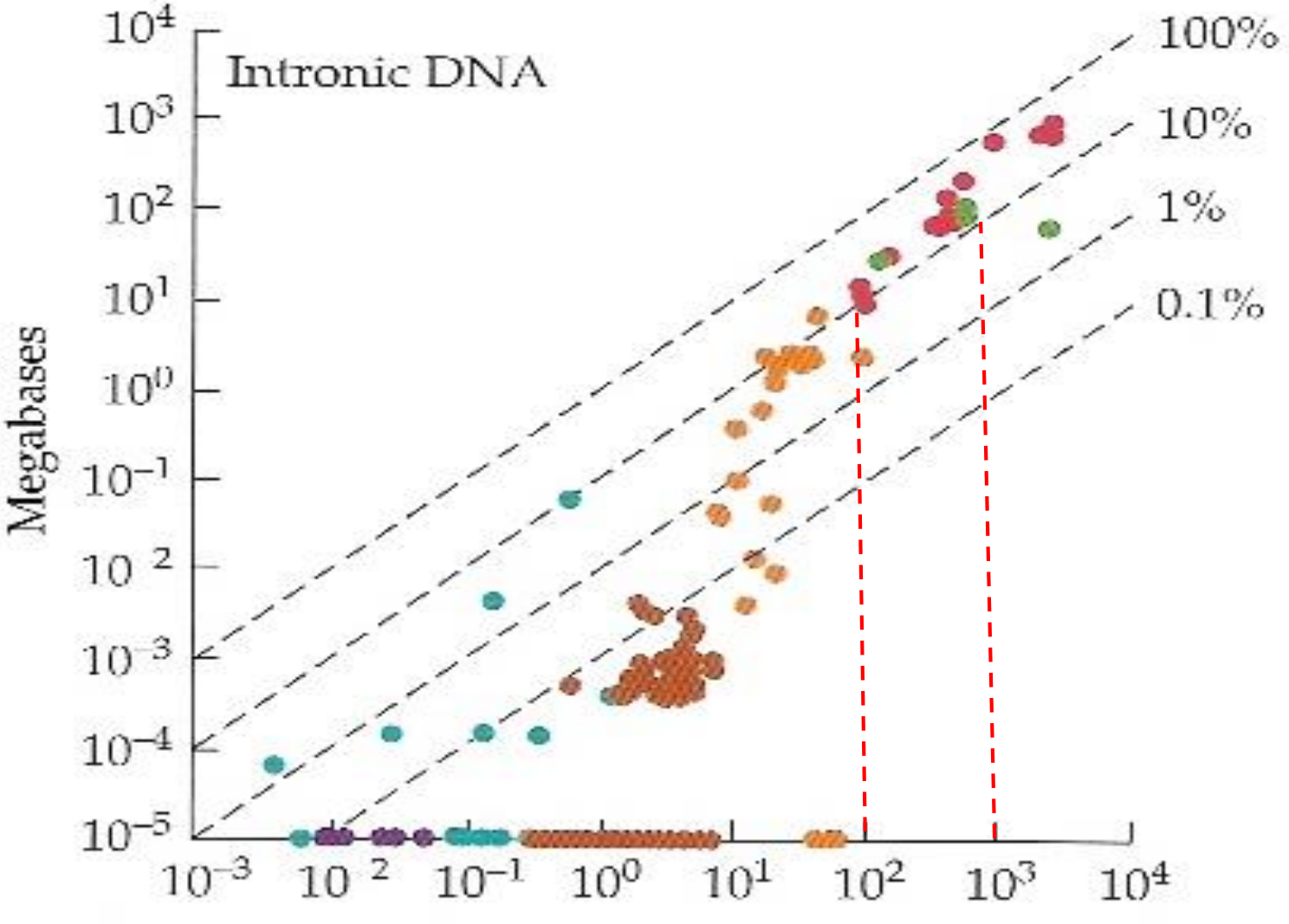
vs.

Nem-kódoló szekvenciák hossza

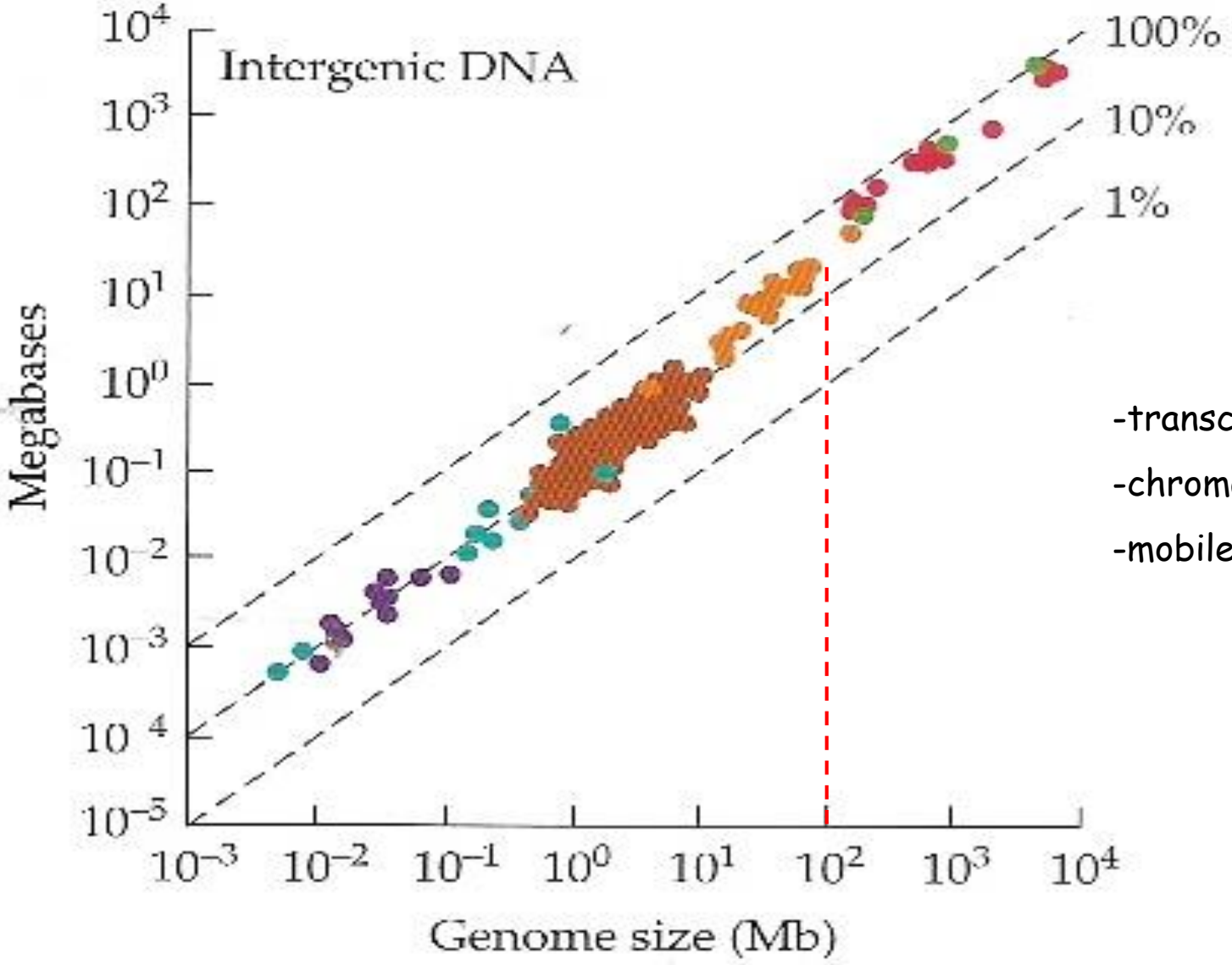
Genom méret vs. kódoló szekvenciák



Genom méret vs. intronok



Genom méret vs. intergénikus DNS



-transcription sites

-chromatin package

-mobile elements

Genom méret és szervezeti komplexitás

- The C-value Paradox: haploid genome size/cell
- Prokarióta: 350-8000 gén, 0.5 – 9 Mb genom
- Multicelluláris Eukarióta: > 13.000 gén, > 100 Mb genom
- Noncoding DNA expanzió (intronok, mobilis elemek, pseudogének)
- Organizmus mérete vs. sejttípusok száma – pozitív korreláció
- Génszám / genom méret vs. multicellularitás / szervezeti komplexitás

Van korreláció? Nem a genom mérettől v. génszámtól függ, hanem ahogy a gének működnek (transzkripciós szabályozás, alternatív splicing, stb.)

Genomok evolúciója rRNA szekvenciák alapján

Woese and Fox, 1977

Woese et al., 1990

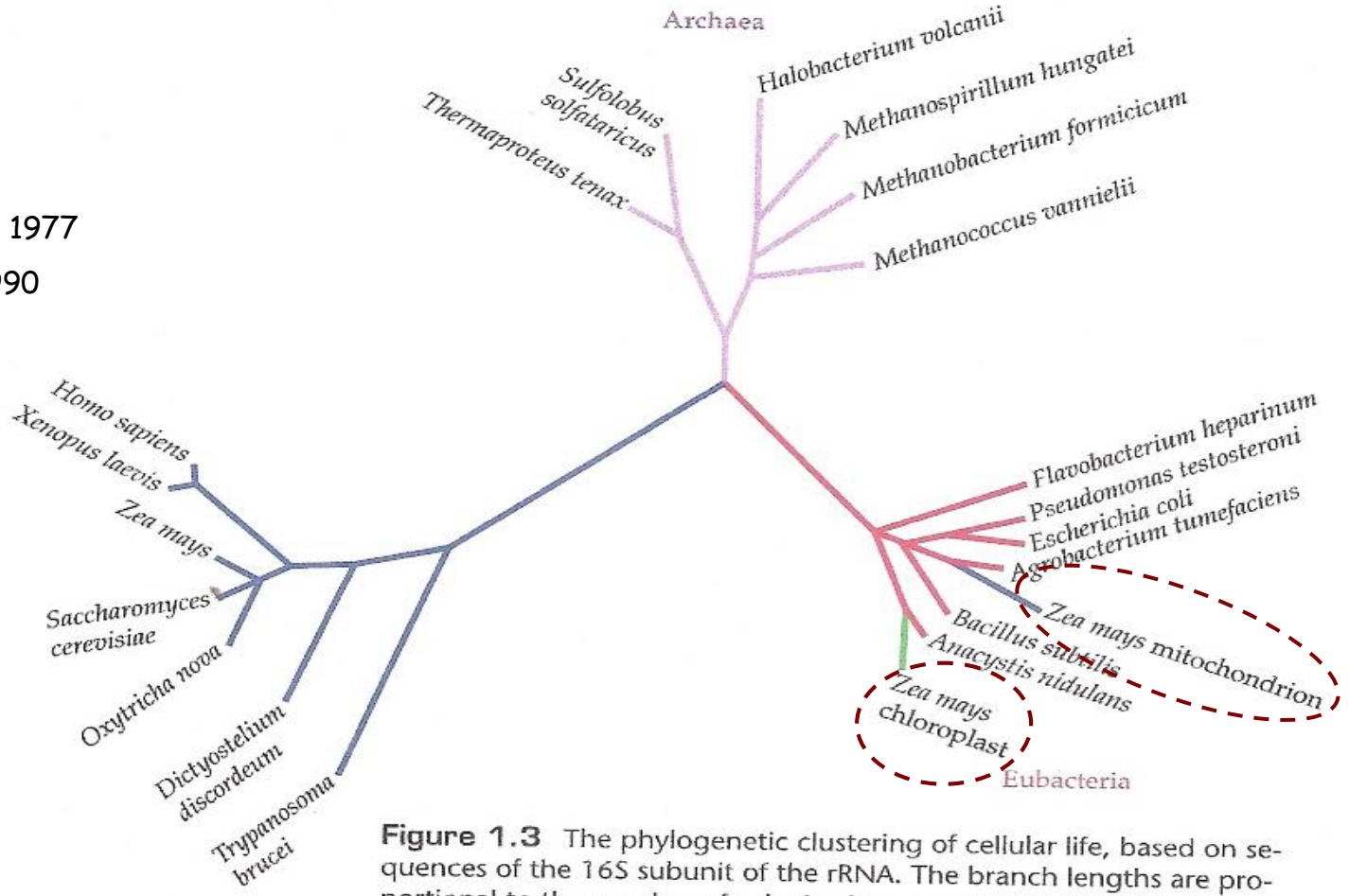


Figure 1.3 The phylogenetic clustering of cellular life, based on sequences of the 16S subunit of the rRNA. The branch lengths are proportional to the number of substitutions per site. Although the exact relationships of some species within groups have not withstood further scrutiny, the distinct nature of the three major domains is well accepted. The presence of mitochondrial and chloroplast sequences in the eubacterial lineage provides compelling evidence for the eubacterial ancestry of these organelles. The tree is unrooted, as the position of the most recent common ancestor of the three major groups is not identified. (Modified from Pace et al. 1986.)

Az Eukarióta genom eredete: archea-eubacteria kiméra?

transzkripció és transzláció: **Archea**

housekeeping funkciók: **Eubacteria**

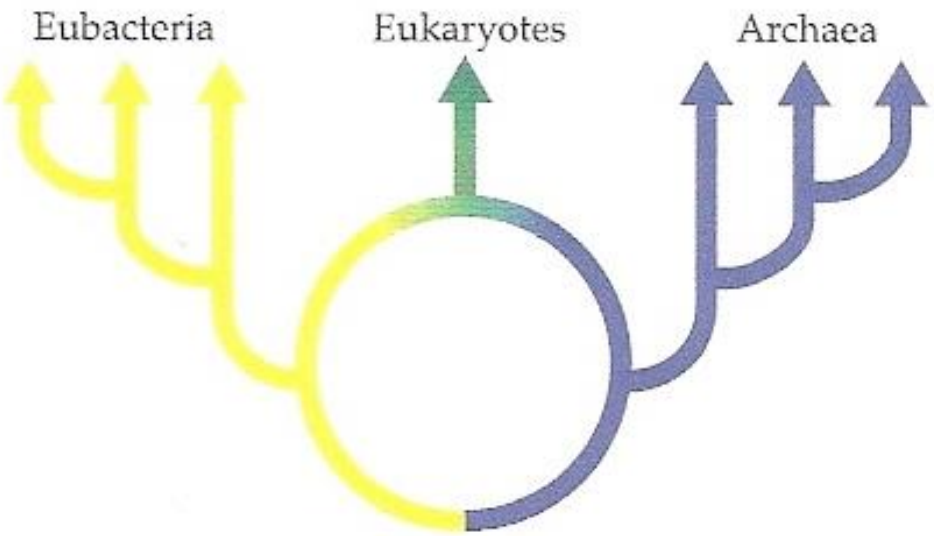
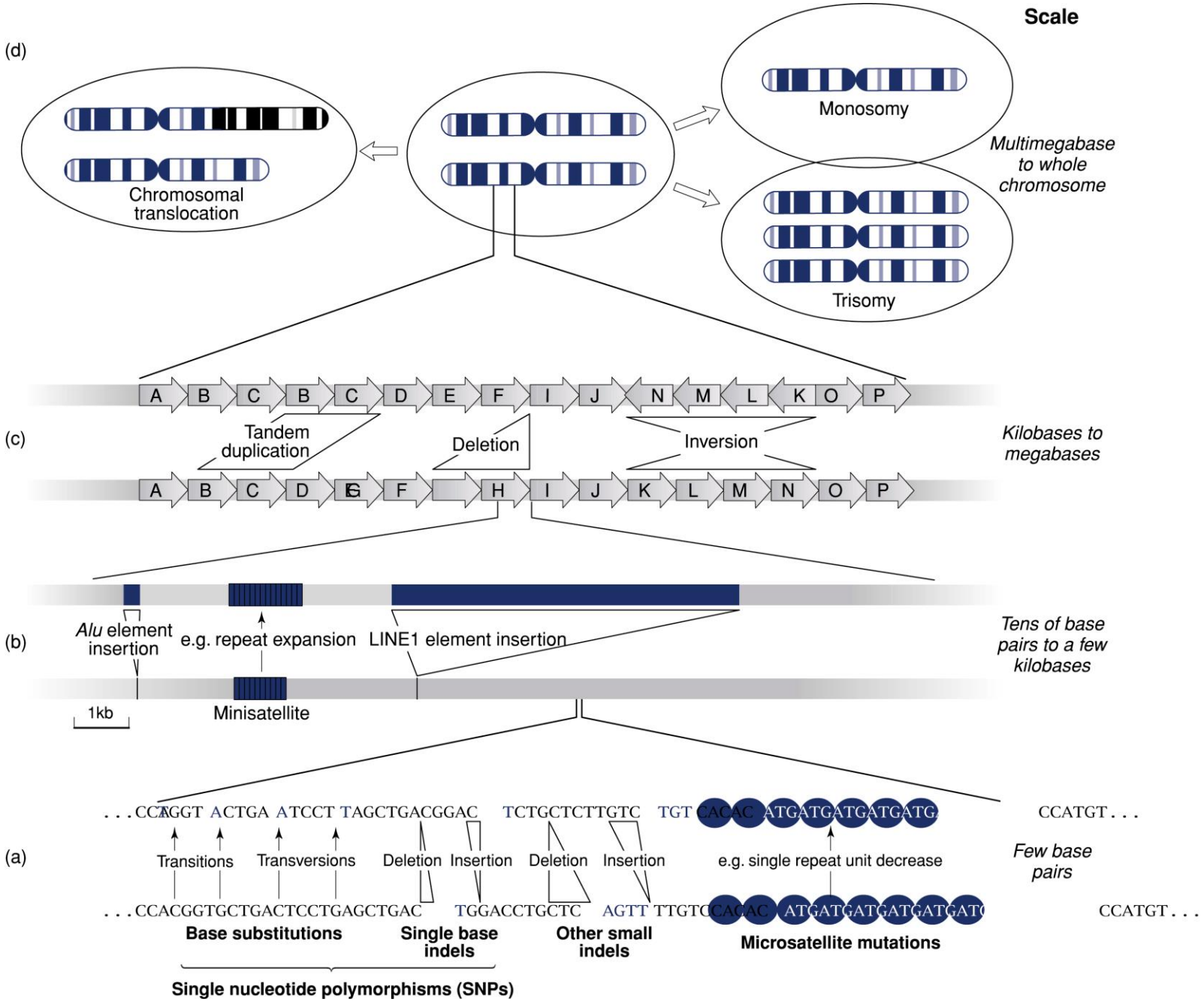


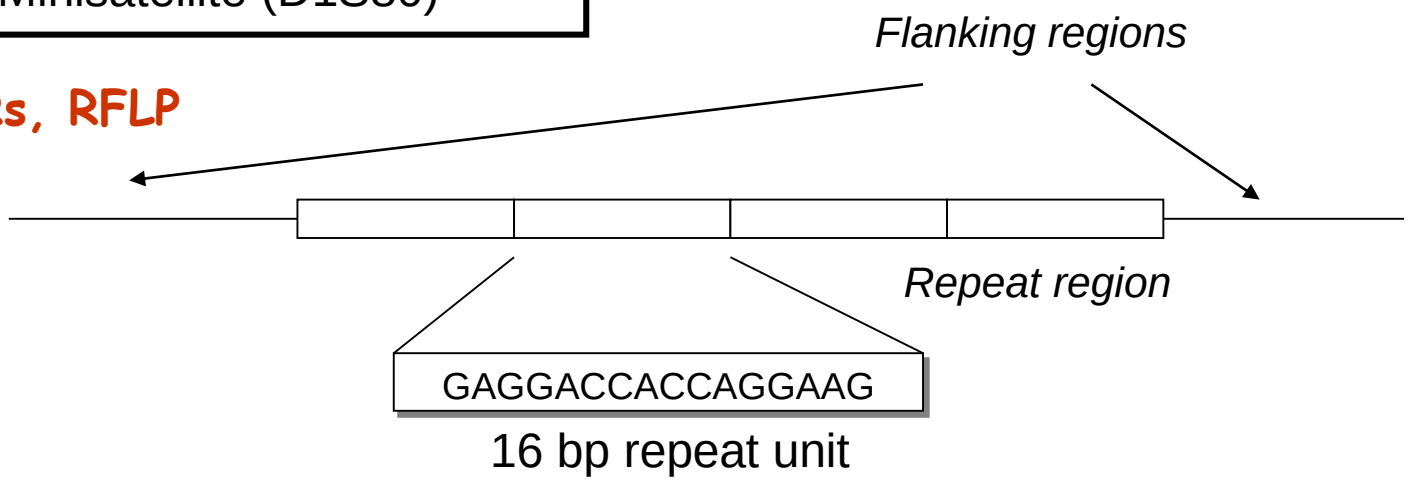
Figure 1.7 The "ring-of-life" hypothesis for the origin of eukaryotes. Yellow and blue lineages denote branches in the phylogenetic trees for eubacteria and archaea, respectively. Members of two such lineages fused to form the eukaryotic domain (green). (Modified from Rivera and Lake 2004.)



Human Evolutionary Genetics, Jobling, 2004

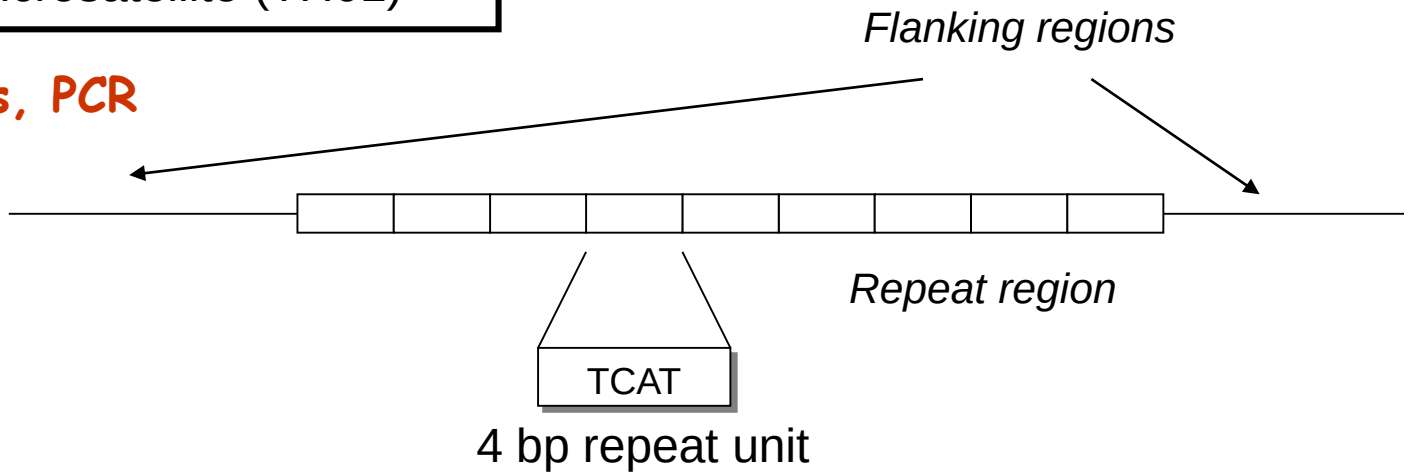
Minisatellite (D1S80)

VNTRs, RFLP

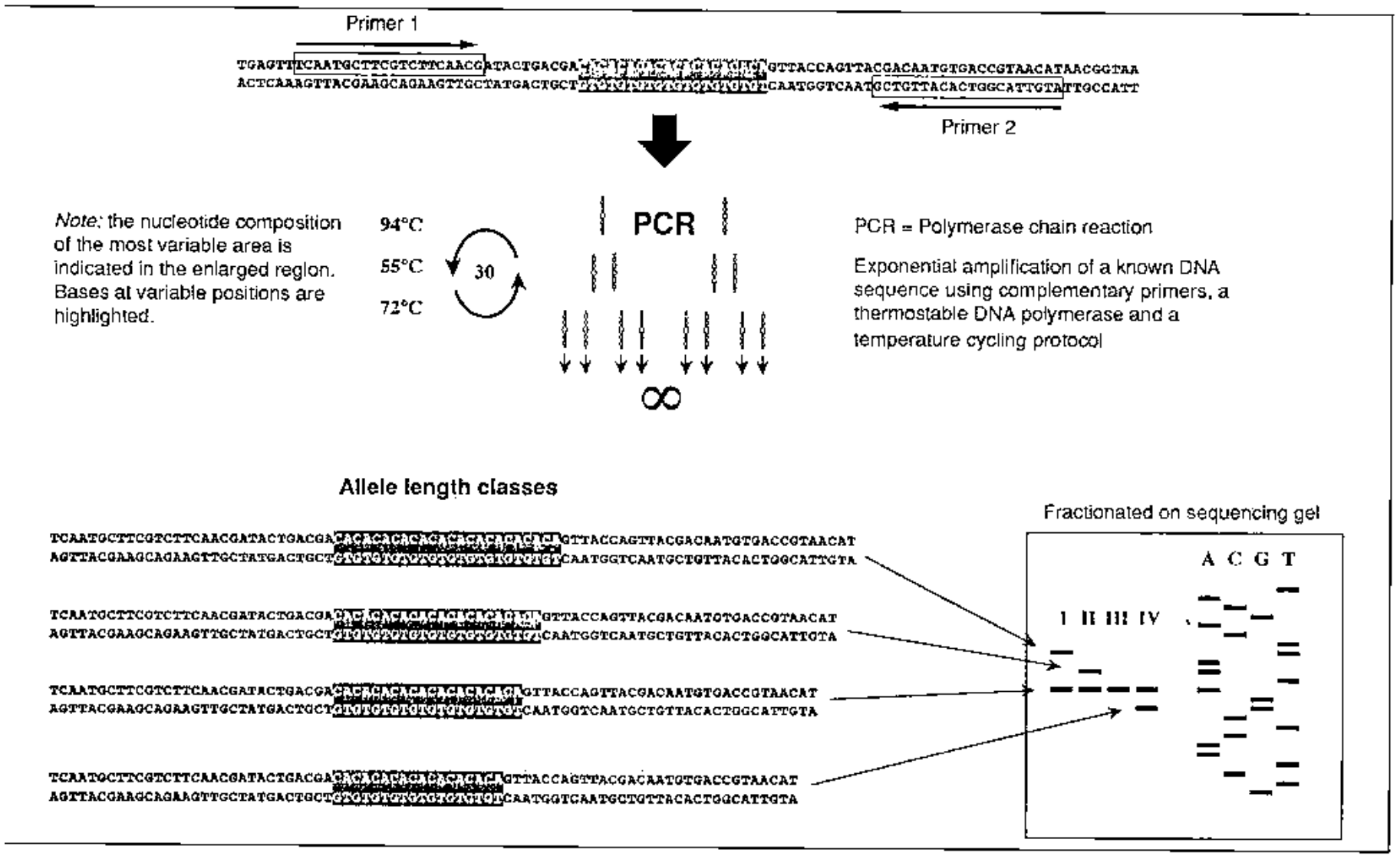


Microsatellite (TH01)

STRs, PCR



Mikroszatellita genotipizálás PCR-rel



STR allélek genotipizálása: multiallélek

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

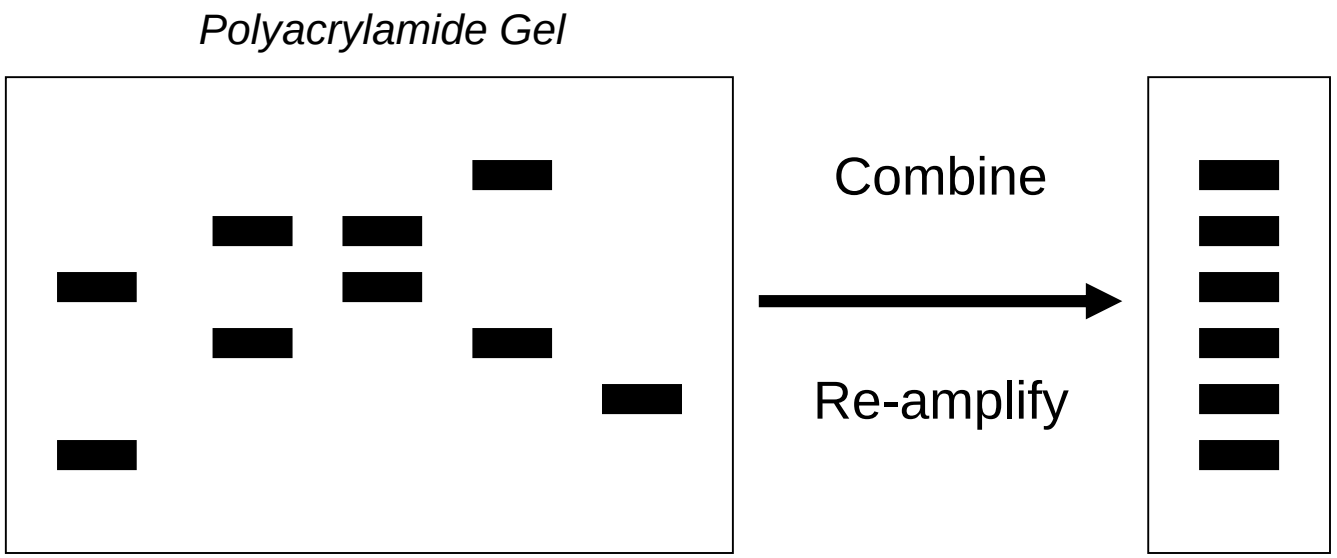
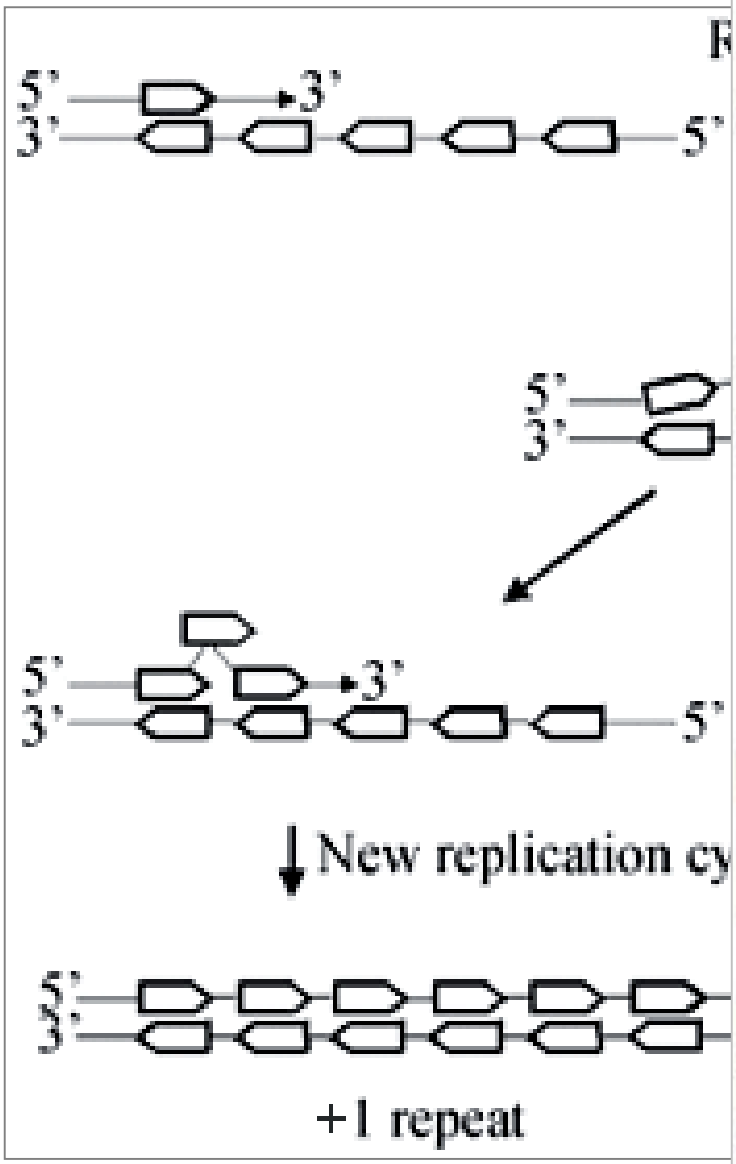
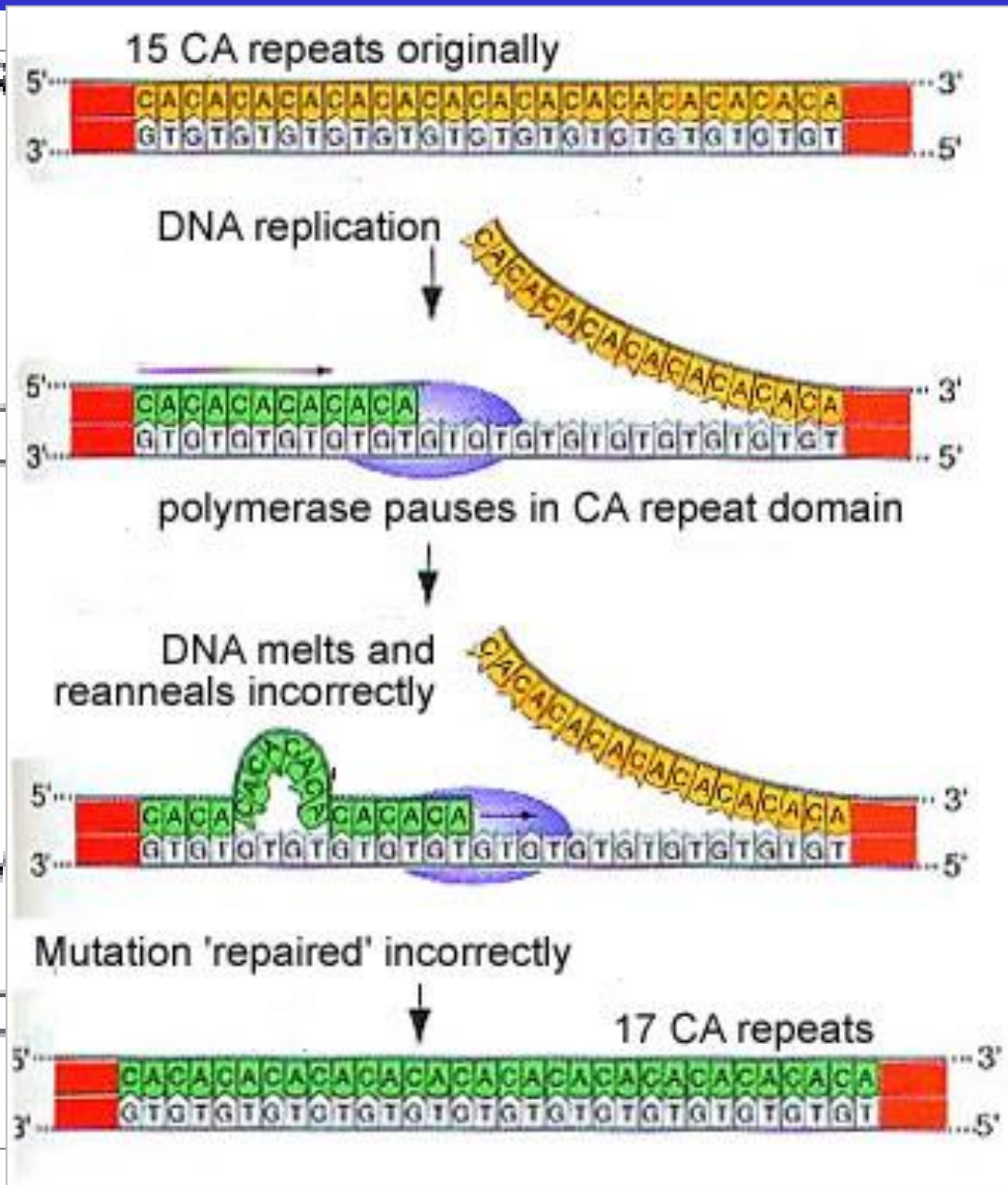


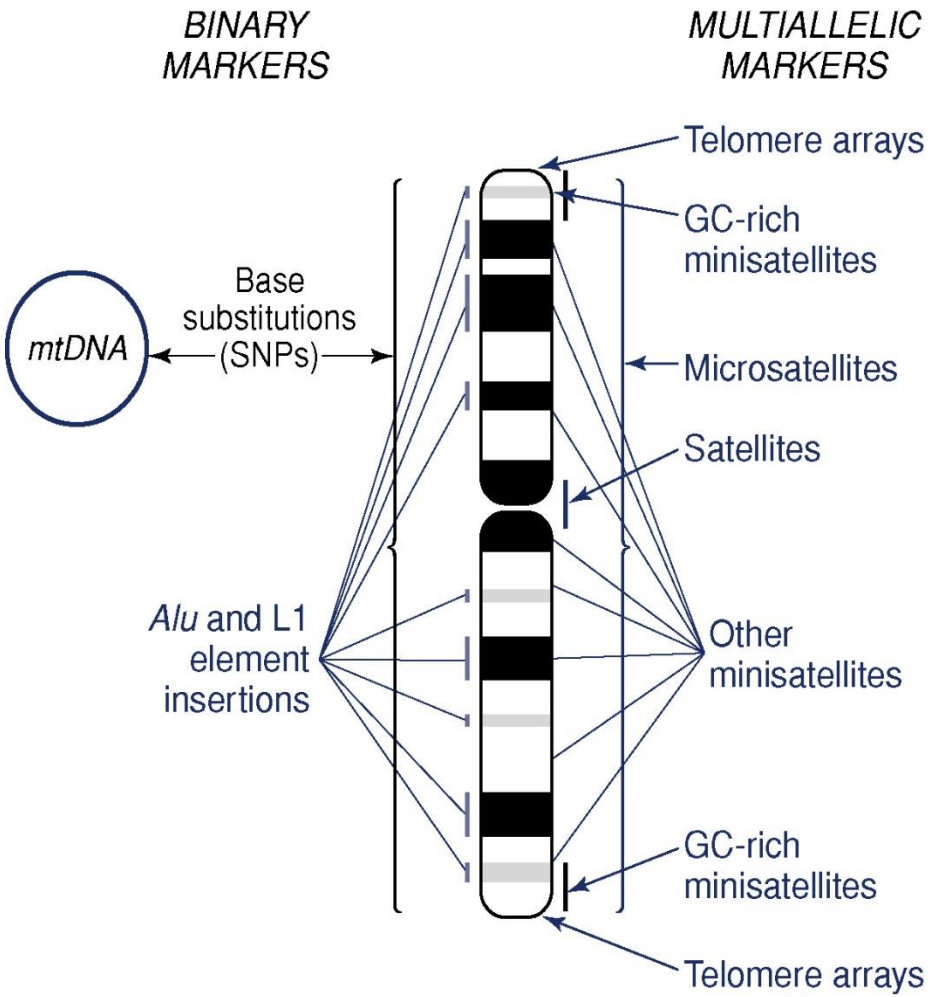
Figure 5.3, J.M. Butler (2005) *Forensic DNA Typing*, 2nd Edition © 2005 Elsevier Academic Press

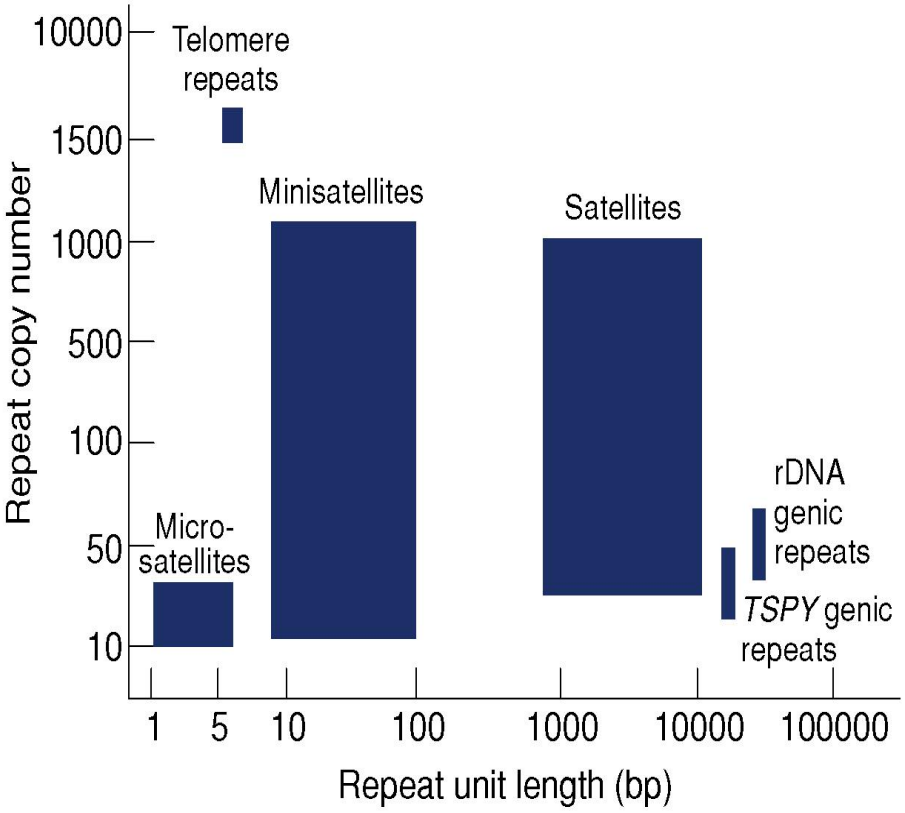
Mikroszatellita evolúció





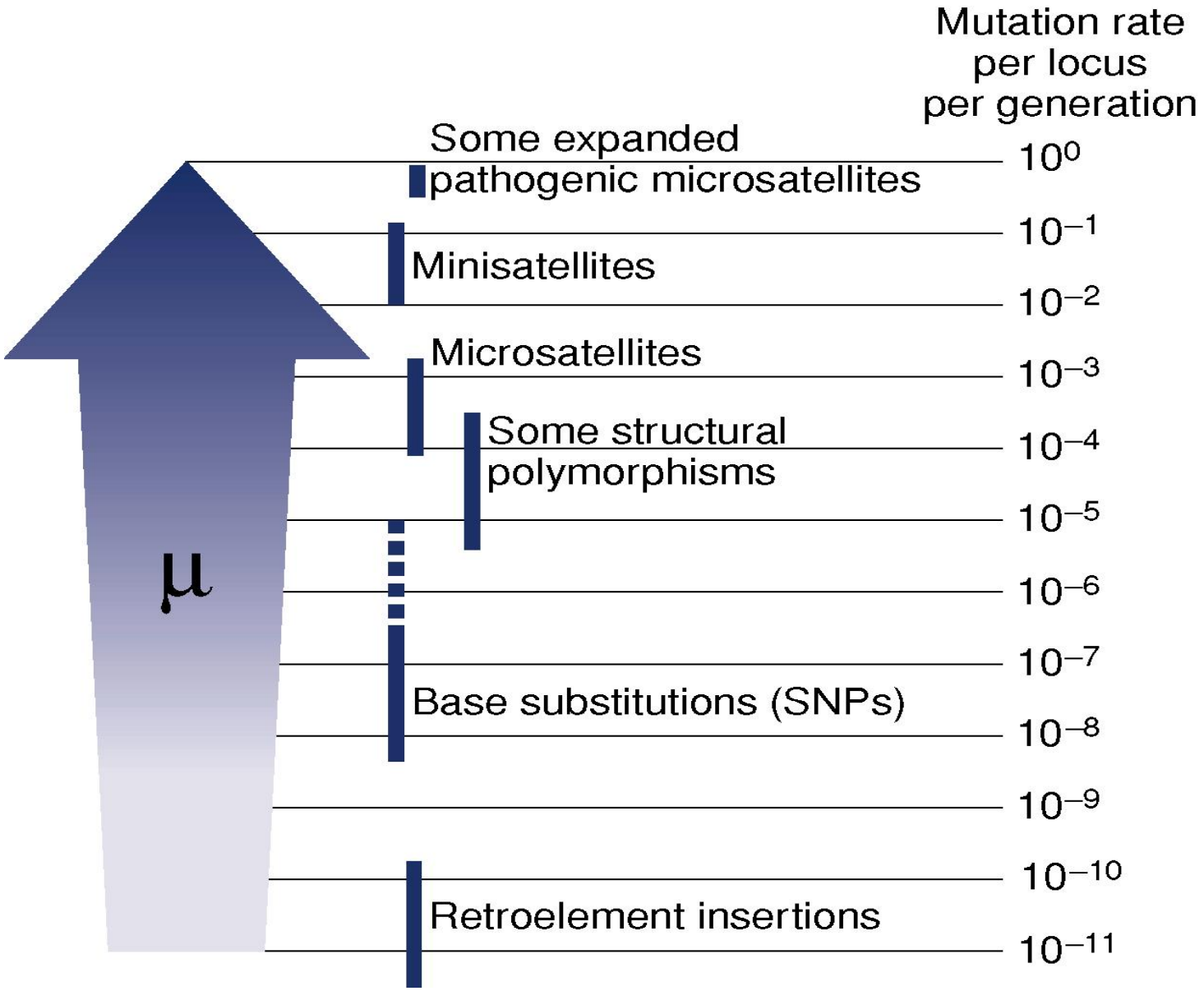
Ismétlődő elemek a genomban



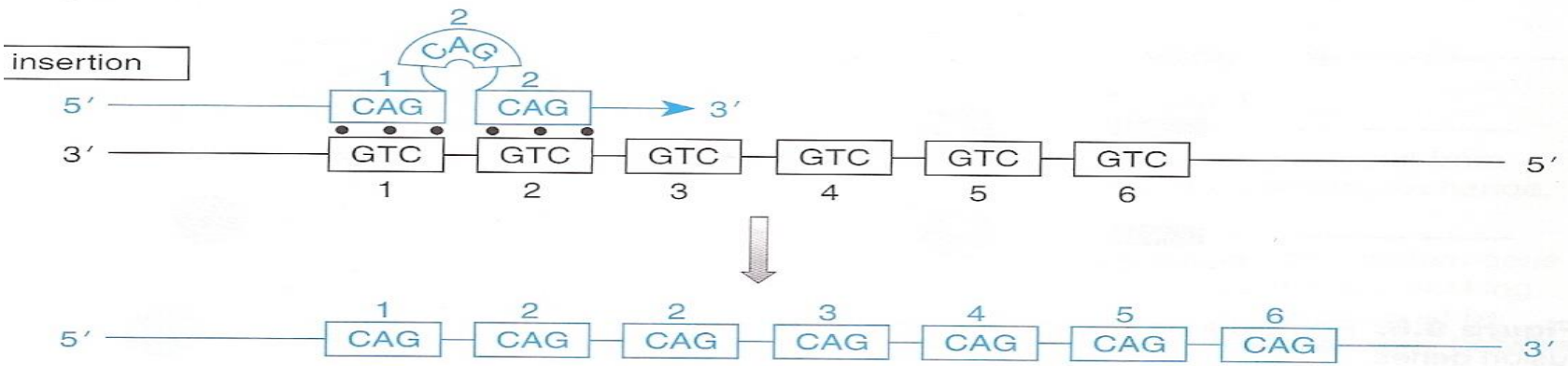


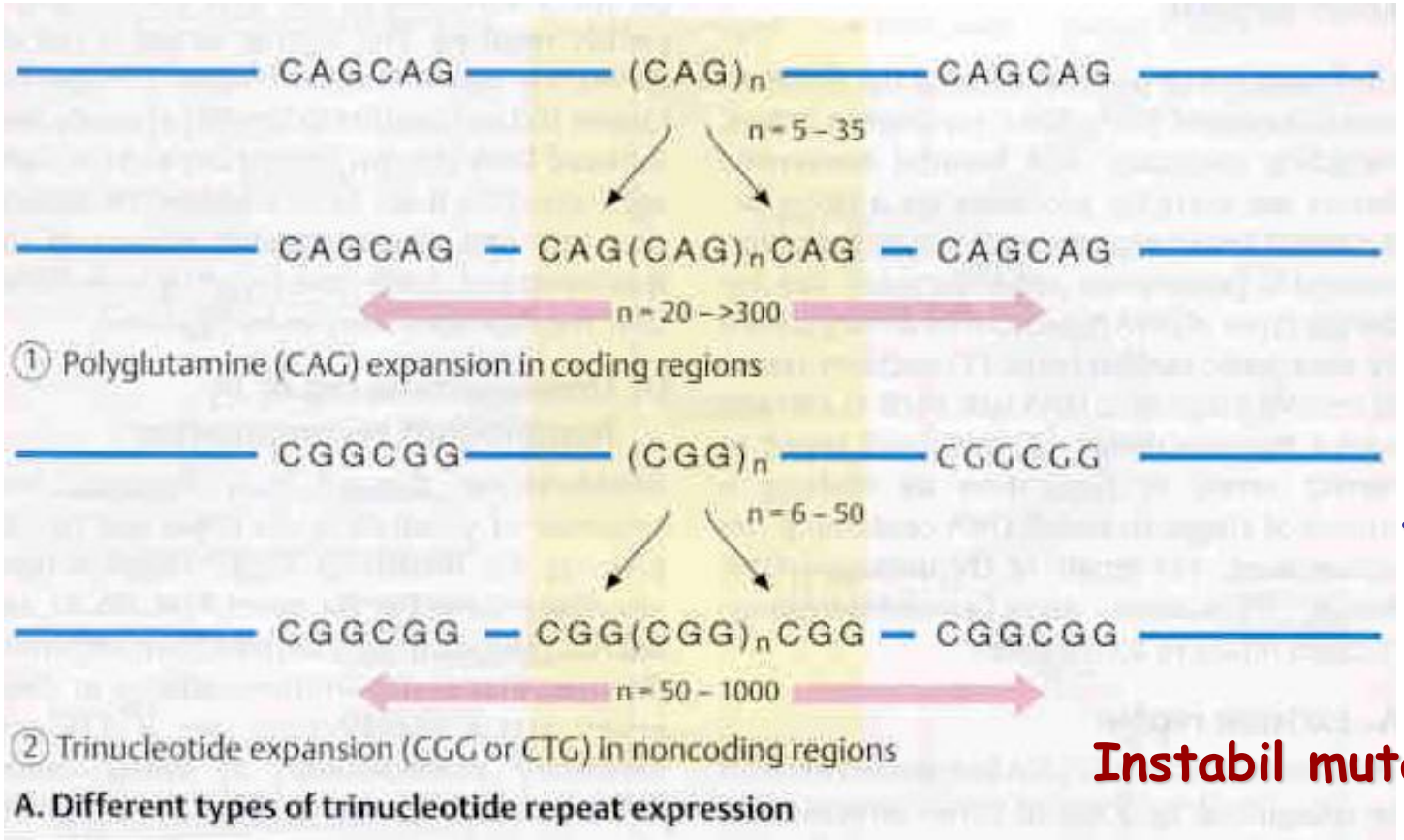
Human Evolutionary Genetics, Jobling, 2004

Polimorf genomi szekvenciák mutációs rátája (μ)



Trinukleotid repeat expanzió I.



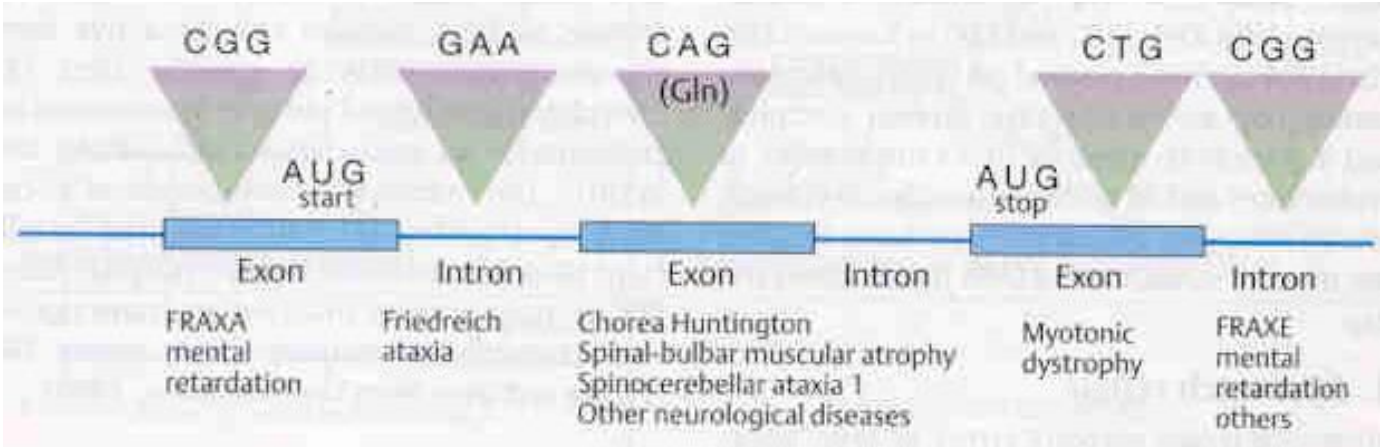


Type I

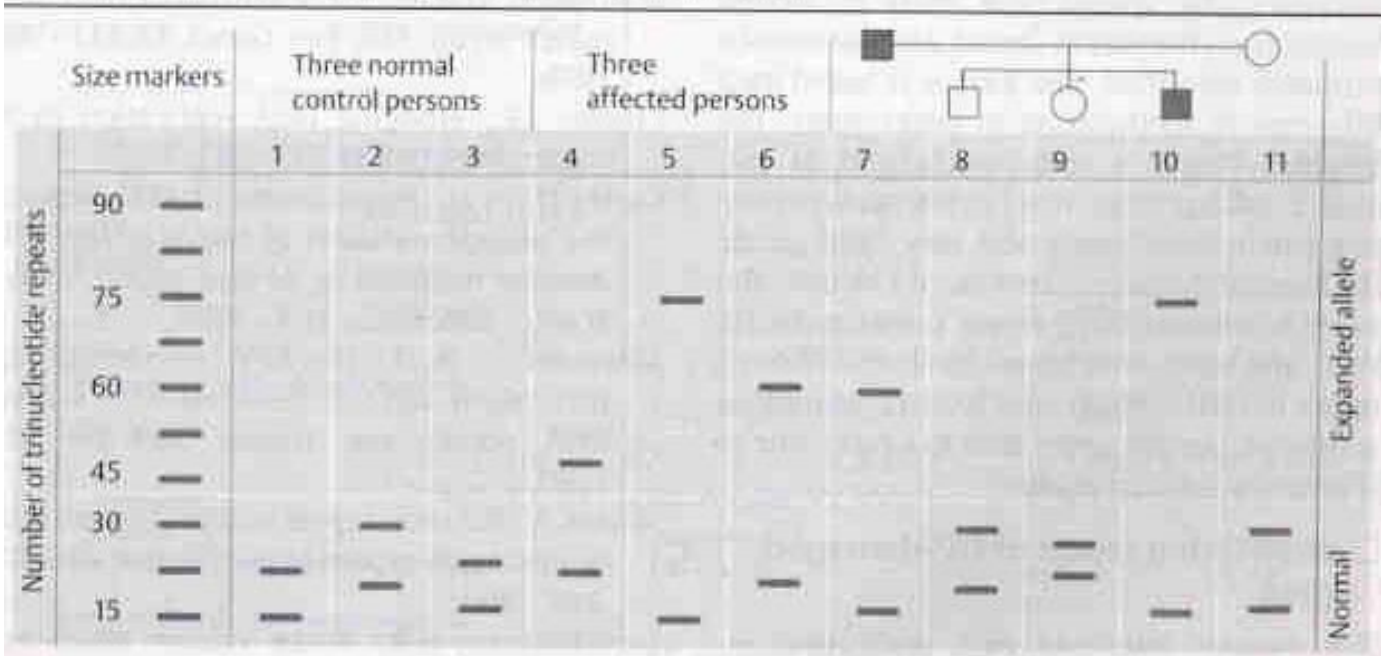
Type II

Instabil mutáció!

Trinukleotid repeat expanzió II.



B. Unstable trinucleotide repeats in different diseases



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

Repeat expanziók okozta defektusok

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

Fragilis X

Huntington disease

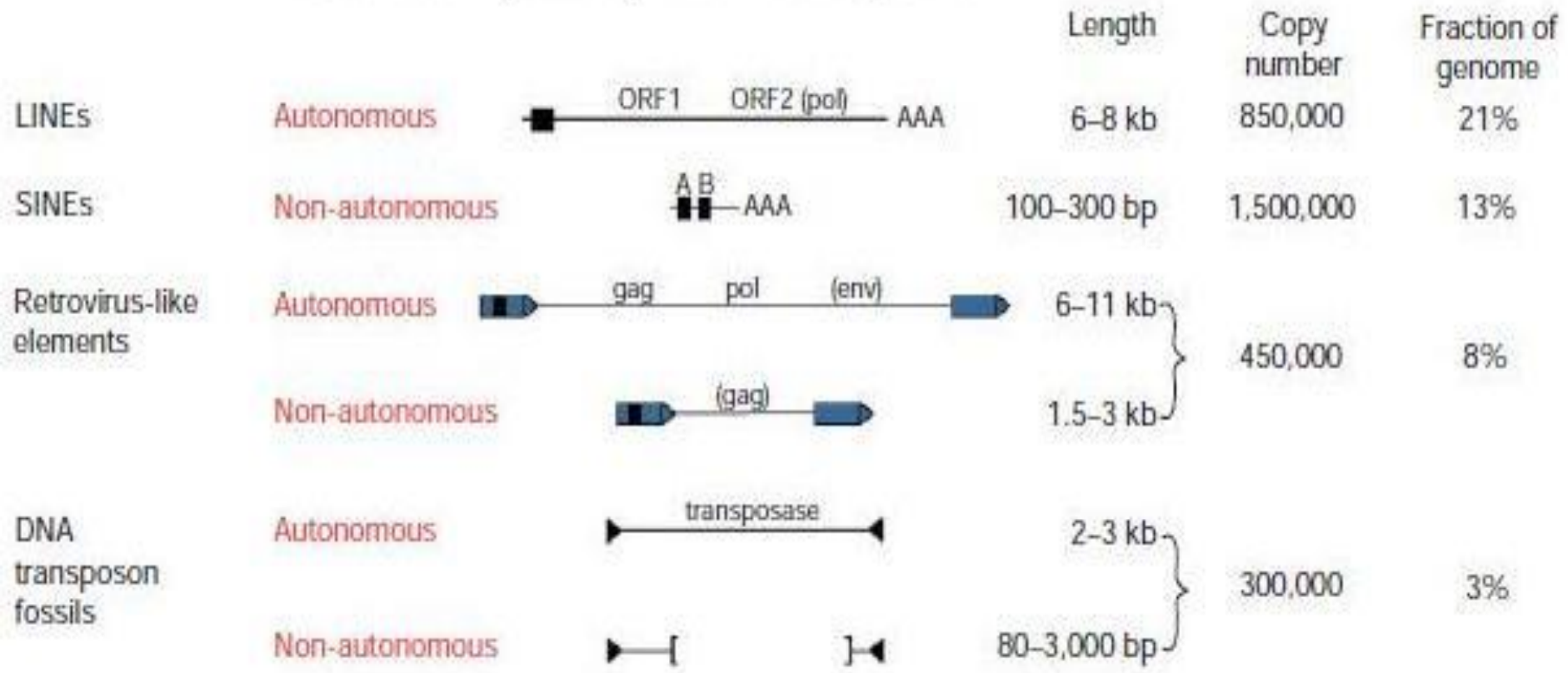
Myotonic dystrophy

Friedrich ataxia

stb.

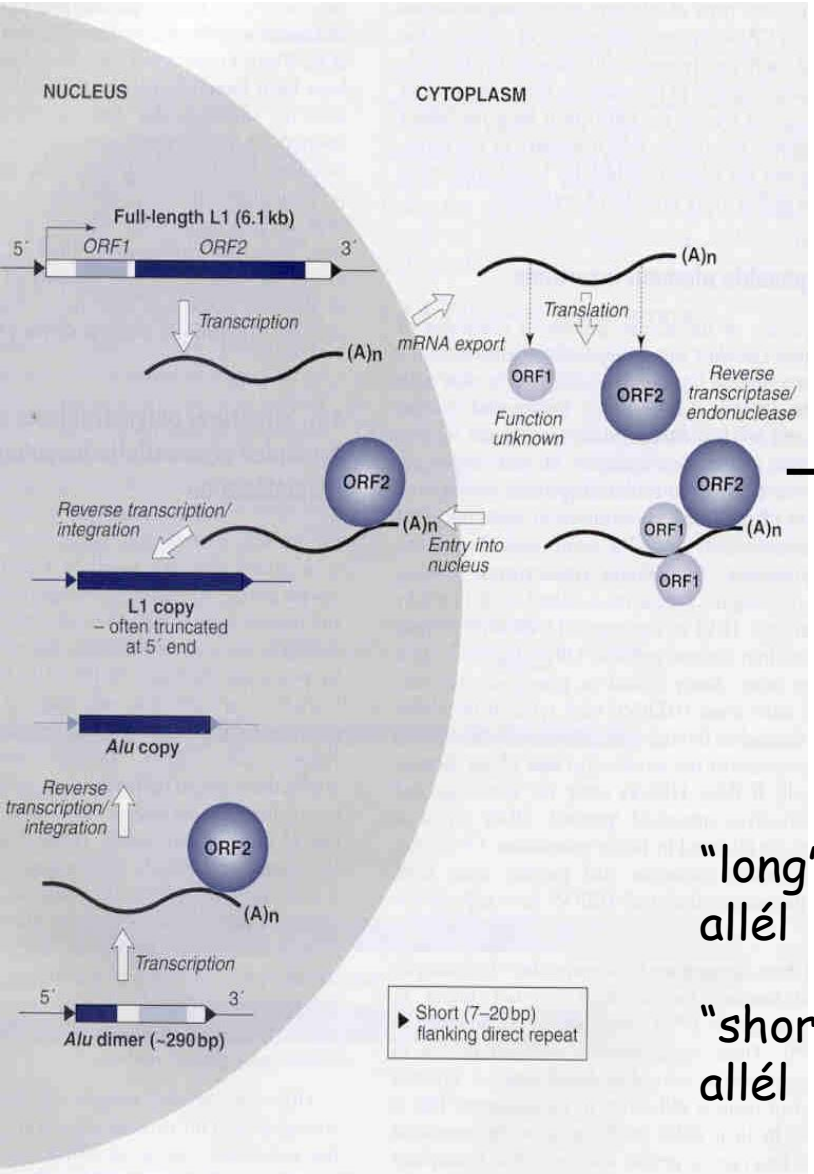
Mobilis genetikai elemek a humán genomban

Classes of interspersed repeat in the human genome

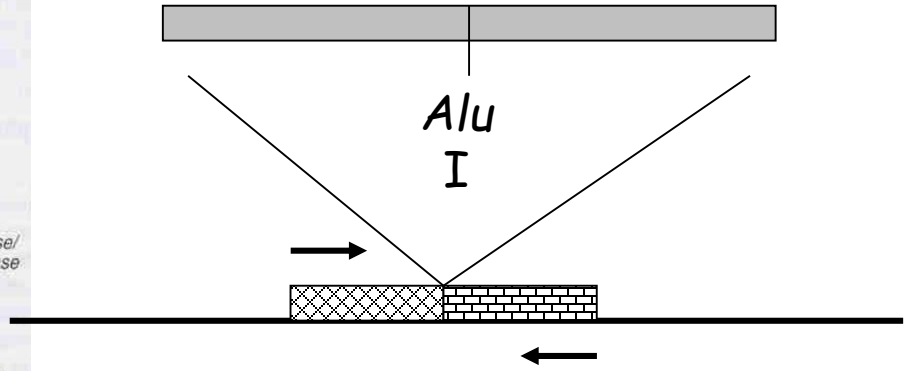


„copy-and-paste v. cut-and-paste"

Mobilis elemek: biallélikus hossz-polimorfizmus



Human *Alu* Repeat
(~300 bp)

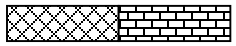


Kétféle alléltípus

"long" (+) allél

400 bp

"short" (-) allél



100 bp

Current Biology 19, 1453–1457, September 15, 2009 ©2009 Elsevier Ltd All rights reserved DOI 10.1016/j.cub.2009.07.032

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

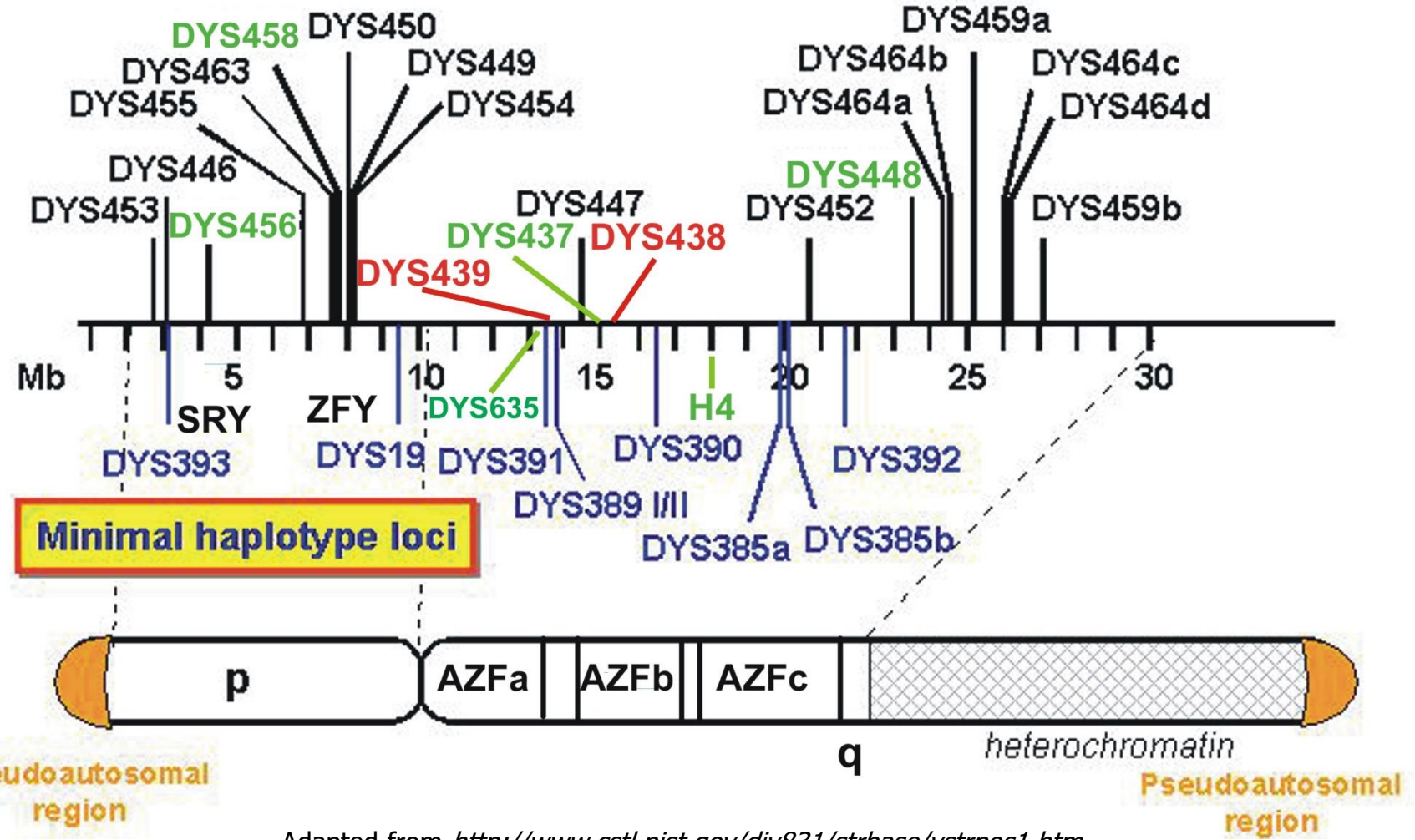
ABSTRACT

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.

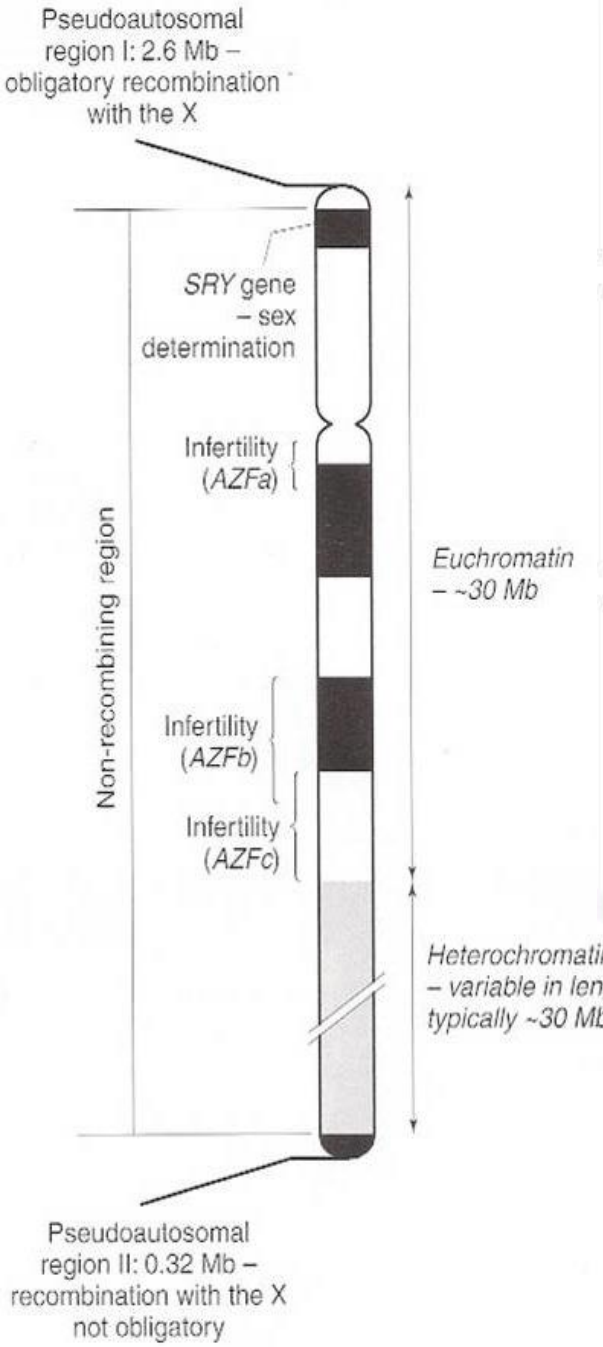
Y STR Positions along Y Chromosome

Extended haplotype loci

ABI AmpF/STR Yfiler loci



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



Sample Number	M175	M5	M122	PN31	LLY22G	M134	M7	M113	M121	M159	M164	B_DYS388I	B_DYS389II	B_DYS390	B_DYS456	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)	DYS 583(B)	DYS579 (G)	DYS525 (Y)	DYS531 (B)	DYS488 (G)	DYS589 (Y)	DYS578(B)	DYS580(G1)	DYS636(Y)	DYS590(B)	DYS533(G1)	DYS617(Y)	DYS594(G2)	DYS505(B)	DYS641(G)	DYS638(Y)	DYS476(B)	DYS492(G)	DYS540(Y)	DYS537	DYS405S1(G)	DYS568(Y)	DYS459(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)	DYS567(Y)	DYS494(B)	DYS578(G)	DYS565(Y)	DYS481(G)	DYS576(B)	DYF390S1(G)	DYS569(Y)	DYS618(G)	DYS511(Y)	DYS643(B)	DYS556(G)	DYS573(Y)	DYS530(B)	DYS491(G)	DYS549(Y)	DYS640(G)	DYS554(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

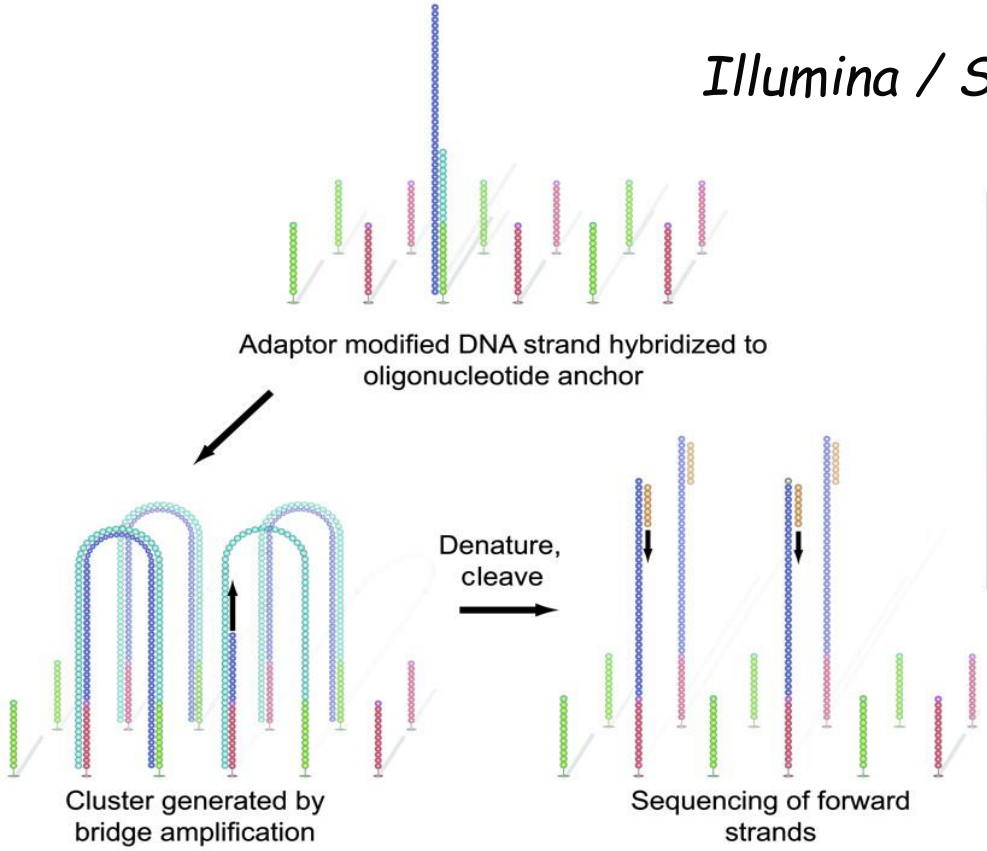
Table S1. Y-SNP and Y-STR haplotypes of the DFNY1-66 and DFNY1-101 chromosomes

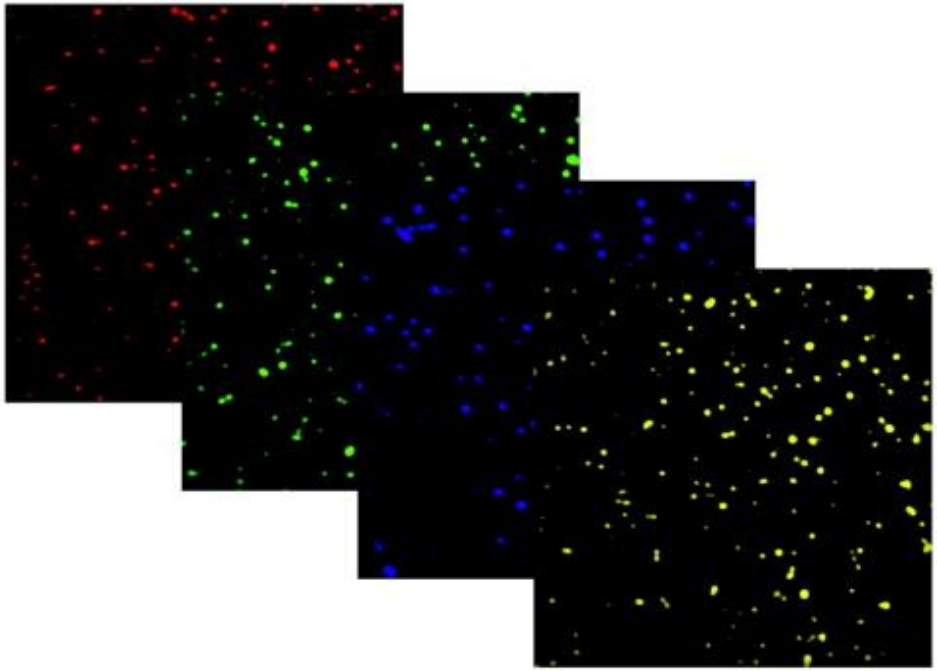
Megegyező Y kromoszómás haplotípus
67 mikroszatellita és 11 SNP markeren

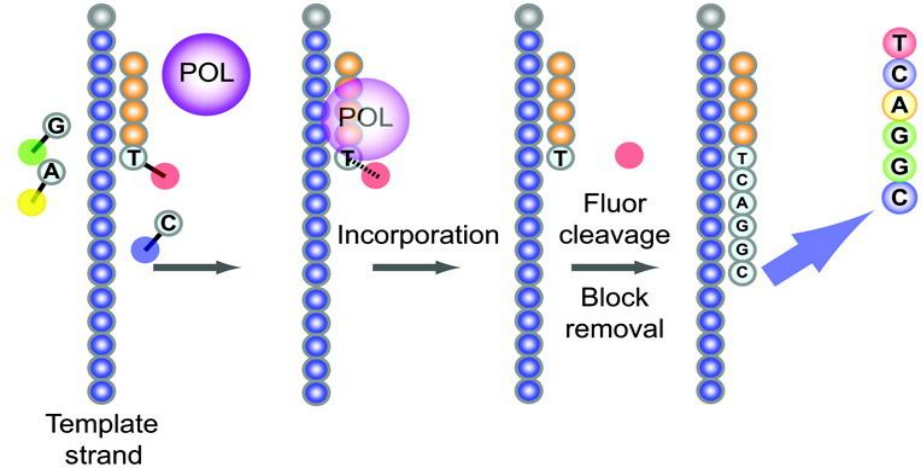
-generációs távolság: 13 generáció

-markerek lokalizációja: eukromatin

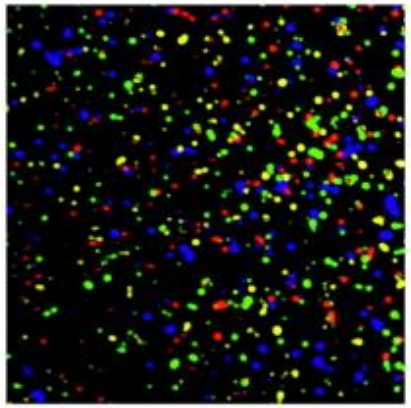
Illumina / Solexa NGS genomszekvenálás







Sequencing by reversible dye terminators



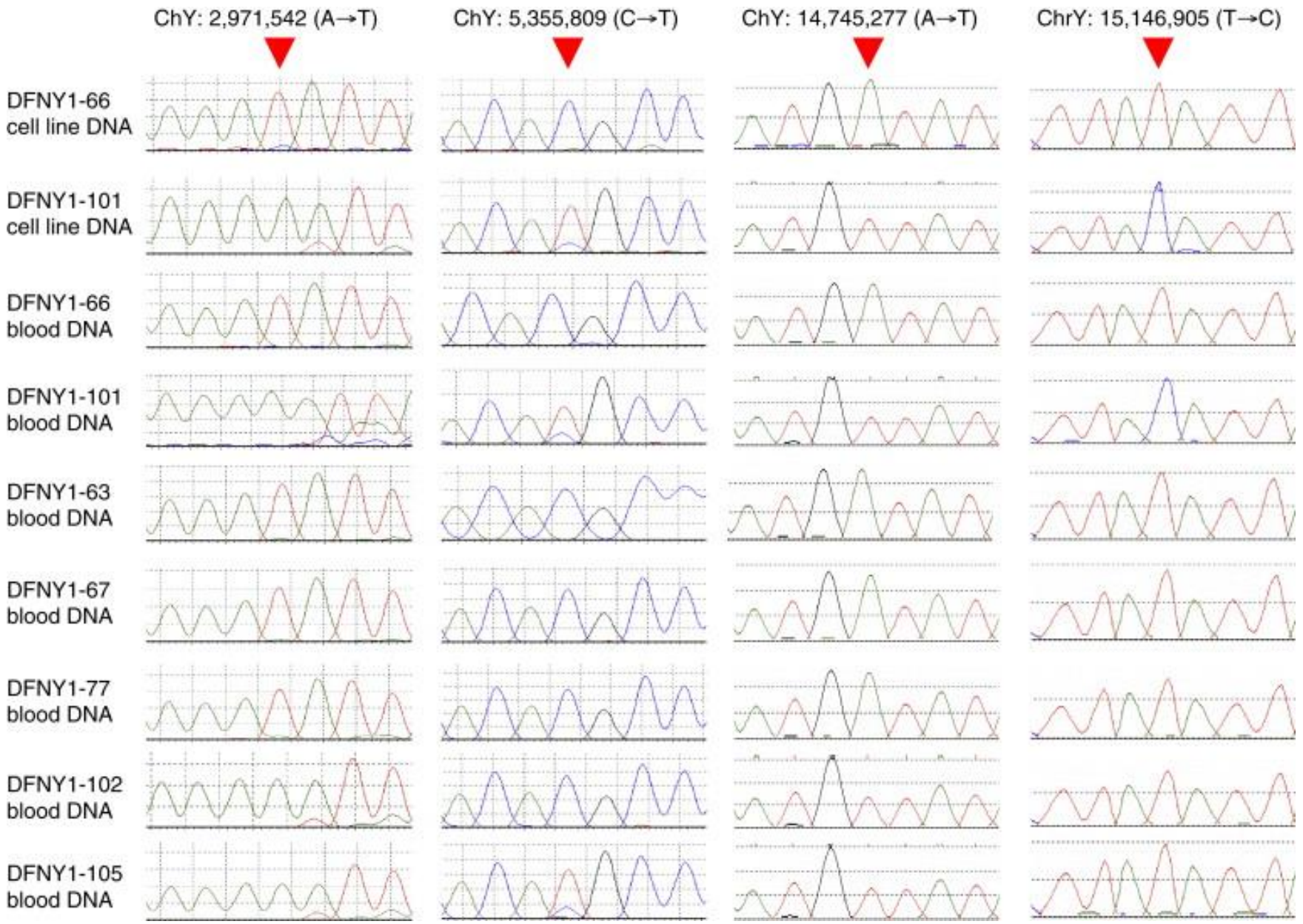
Y kromoszómális eukromatin kandidáns pontmutációk

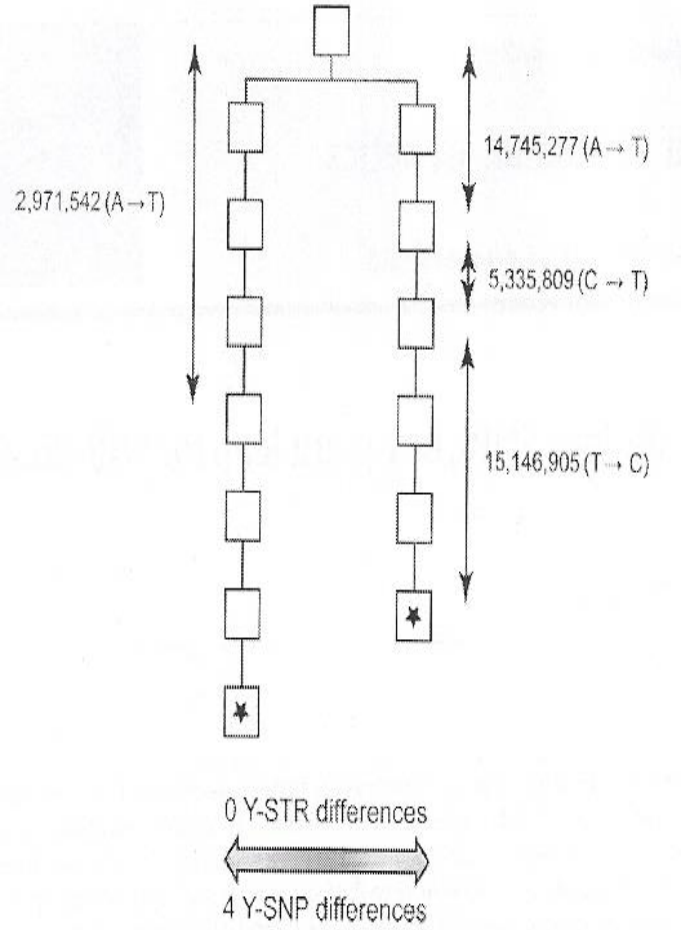
Table 2. Details of the Filtered Candidate Mutations

		DFNY1_101 Pileup		DFNY1_66 Pileup		Confirmation	
Chromosome Coordinate	Base	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	cCccCcccCCCCC	13	TTTTTTTTTTttT	Yes	No
chrY:4,980,623	T	5	ggggg	7	TtTTTTT	Yes, het	No
chrY:5,355,809*	C	12	TtTTTTTTTtTt	9	cCccccCcC	Yes	Yes
chrY:6,555,594	G	13	TgTttTTtTTtTT	12	GGGGGgGGgGGG	No	
chrY:7,381,330	G	7	cCcCCCc	12	GGGGGgGGgGGG	No	
chrY:12,063,011	C	5	gggGG	8	ccccCCCC	Yes	No
chrY:14,745,277*	A	9	TtTTtTttt	6	aaAaAa	Yes	Yes
chrY:15,126,873	T	7	cccCccc	8	tttTttTT	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	TTtttTTt	Yes	No
chrY:2,971,542*	A	4	aAAA	14	tTTtTTttTtttT	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	tttTTTTTTt	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	aAAgggggggggg	11	aaaAaAaaAAa	No	
chrY:13,833,351	C	17	cCccCCggccCcCcccc	16	CCcCcCCcCttCtttc	No	
chrY:14,573,532	A	21	GAAAAaaAaAAaAaaAAaAAg	5	AAggg	No	
chrY:15,375,202	G	4	GGGg	4	TTTT	No	

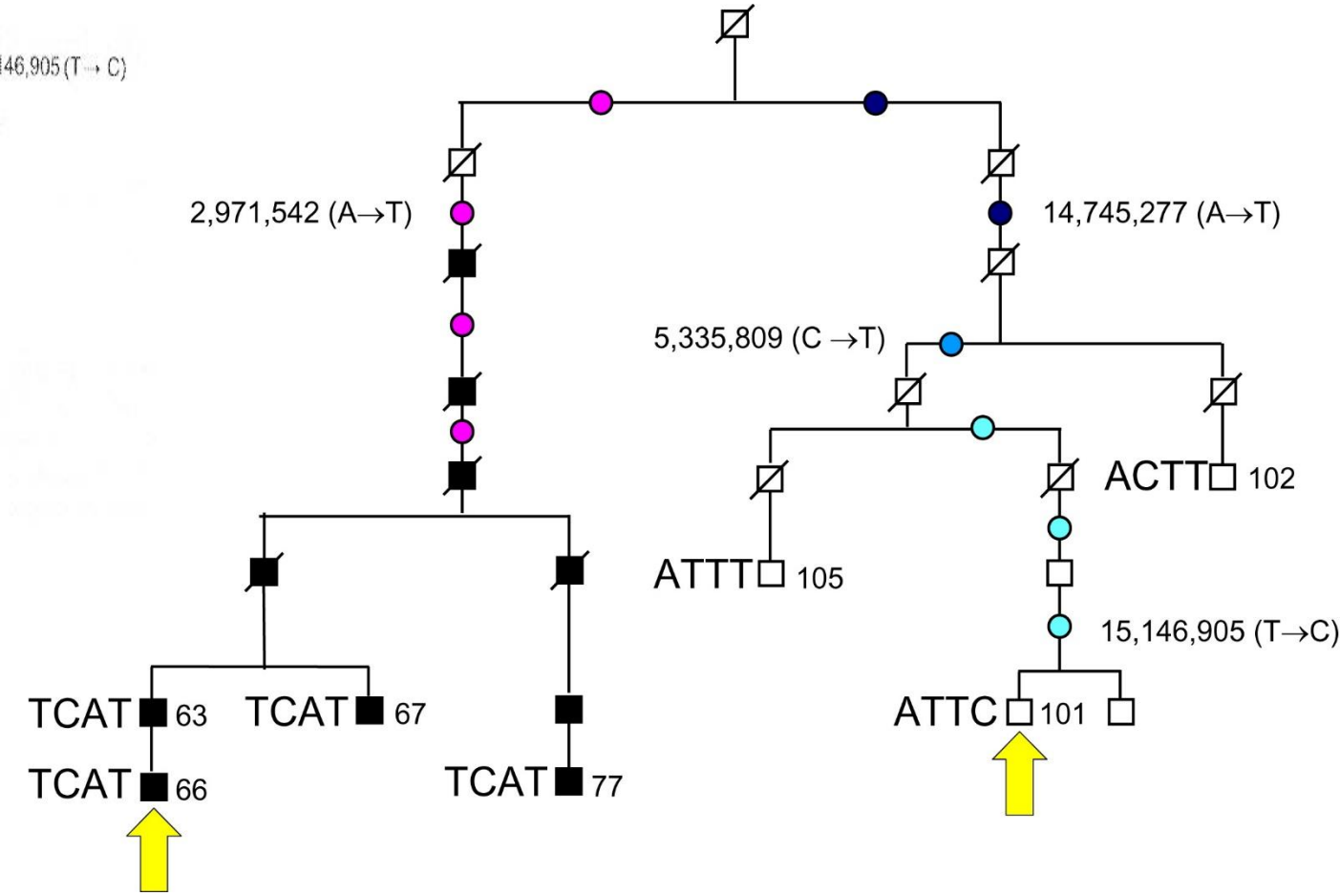
An asterisk denotes mutations that were confirmed in blood DNA.

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

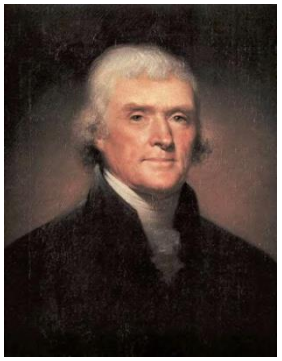

ChY: 2,971,542 (A→T)
ChY: 5,355,809 (C→T)
ChY: 14,745,277 (A→T)
ChrY: 15,146,905 (T→C)
DFNY1-66 cell line DNA
DFNY1-101 cell line DNA
DFNY1-66 blood DNA
DFNY1-101 blood DNA
DFNY1-63 blood DNA
DFNY1-67 blood DNA
DFNY1-77 blood DNA
DFNY1-102 blood DNA
DFNY1-105 blood DNA



de novo szubsztitúciós mutációk az Y kromoszómás eukromatinban multigenerációs pedigrében



Genetic History



Thomas Jefferson II

Field Jefferson

Peter Jefferson

President Thomas Jefferson

?

Eston Hemings

Same Y Haplotype

John Weeks Jefferson born in 1946

Thomas Woodson

Different Y Haplotype

Jefferson Y Haplotype

Jefferson Y Haplotype

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

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

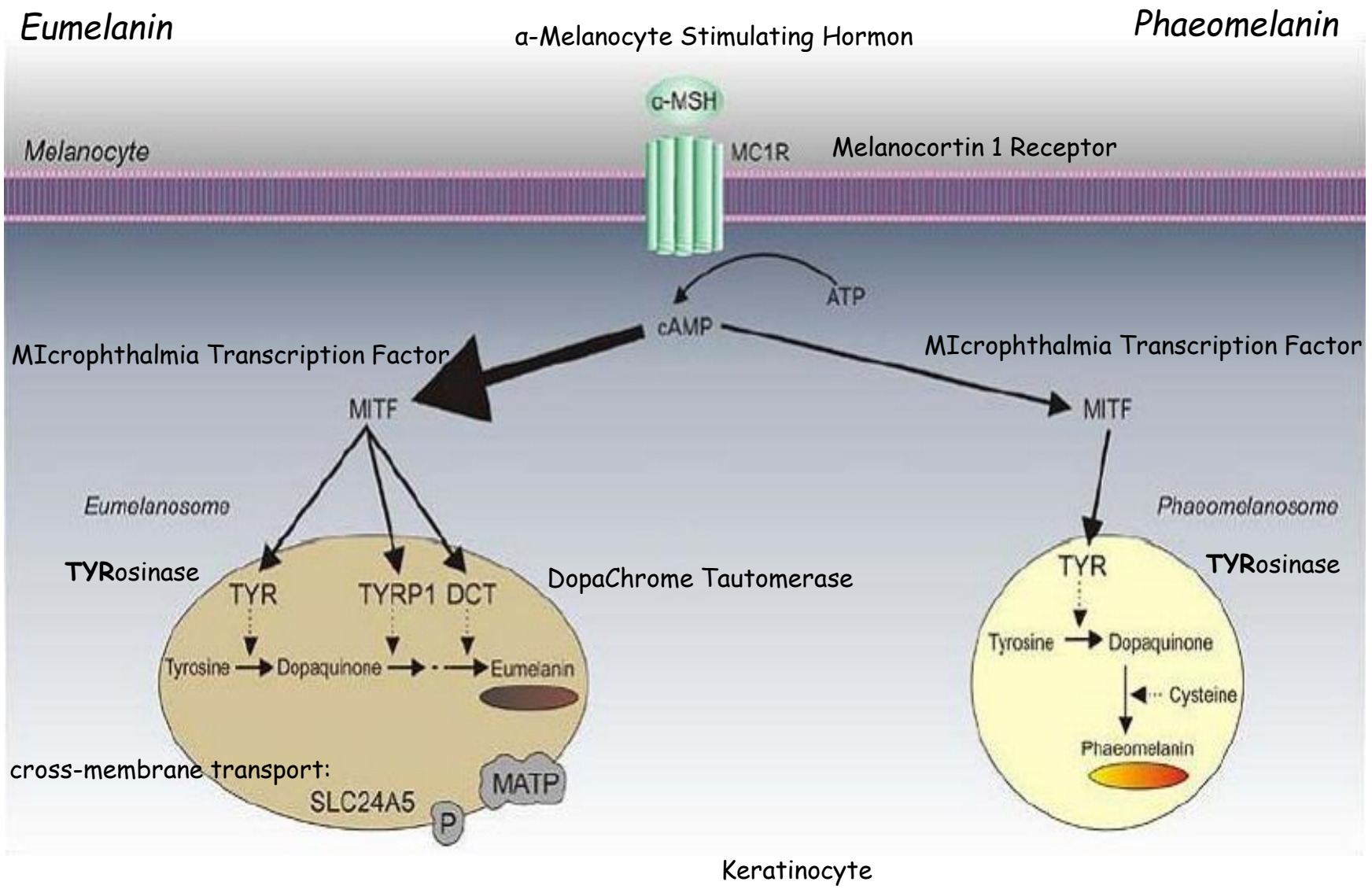
TABLE 12.2 Categories of SNP Markers (See Budowle & van Daal 2008, Butler et al. 2008).

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 melanogenezis biokémiai útvonal



Bouakaze et al., Int J Legal Med (2009) 123:315

SNPs - pigmentációs gének

- ASIP (aguti): 3'UTR 8818A - MSH antagonista - 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
MC1R	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
HERC2	15q13	Unknown	rs12913832	A/G	Non-coding, intron 86
OCA2	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
SLC45A2	5p13.3	MATP: membrane-associated transporter protein	rs16891982	C/G	ns coding, c.1122C>G, p.F374L
SLC24A5	15q21.1	SLC24A5 (or NCKX5): solute carrier family 24, member 5; potassium-dependent sodium–| calcium ion exchanger	rs1426654	G/A	ns coding, p.A111T
DCT	13q32	DCT or TYRP2/TRP-2: dopachrome tautomerase or tyrosinase-related protein-2	rs2031526	G/A	Non-coding, intronic

ns non-synonymous

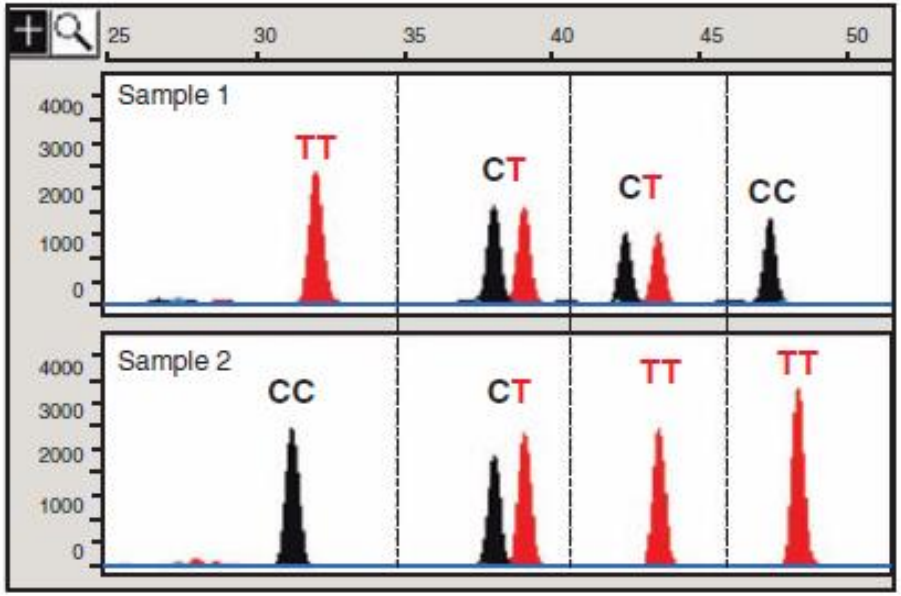
[b] Reference SNP ID refer to the reference sequence identifier given to the SNP in the dbSNP database

Bouakaze et al., Int J Legal Med (2009) 123:315

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**

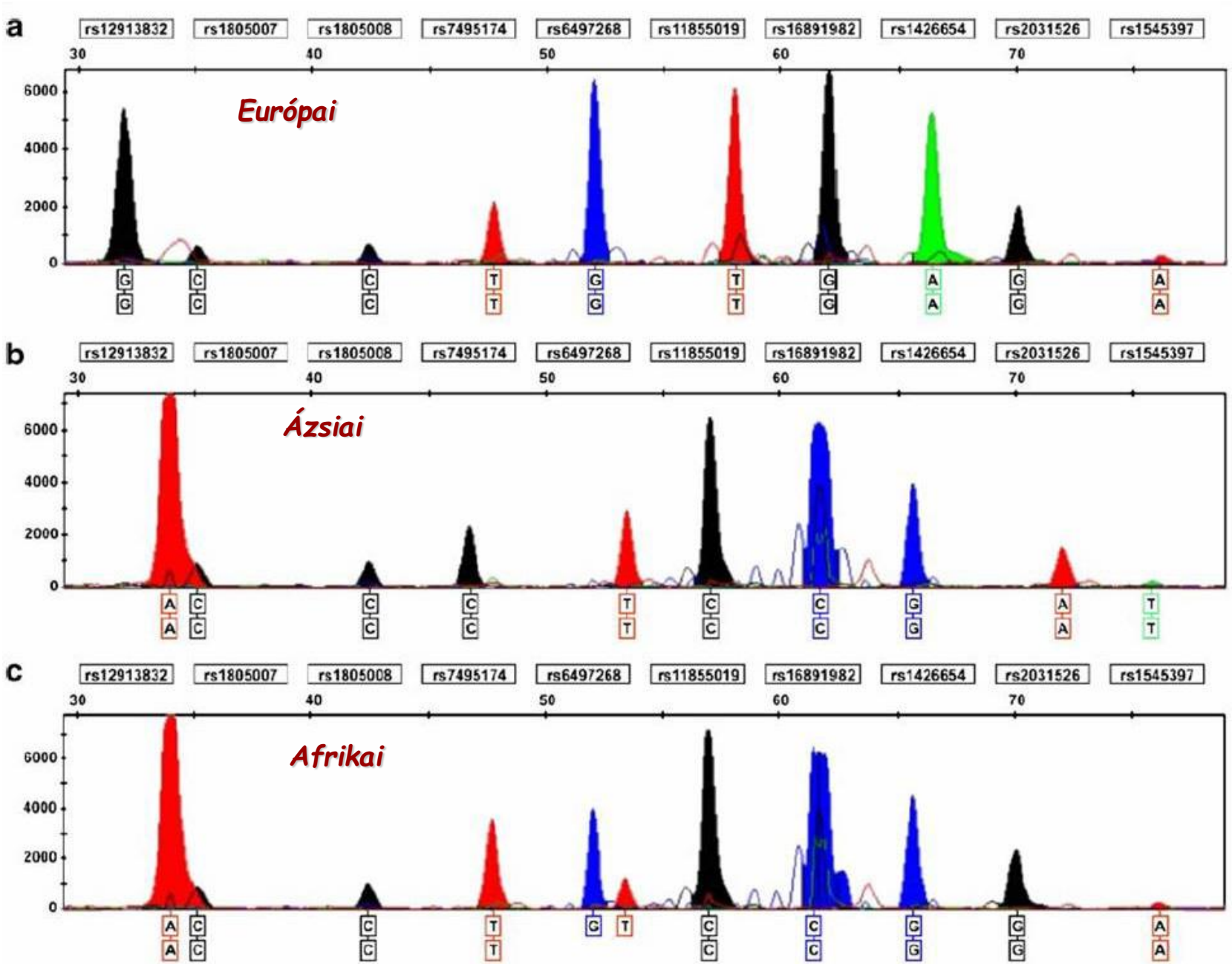
(TTTTT)–(TTTTT)–(TTTTT)–**primer3** (chromosome 14)-**ddC/ddT**

(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

10 pigmentációs gén SNPs genotipizálás (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