

GENOMIKA

előadások a genetikai állomány szerveződéséről



ELTE TTK Genetikai Tanszék

THE TIMES

THE SUNDAY TIMES

Archive Article

Please enjoy this article from The Times & The Sunday Times archives. For

From [The Sunday Times](#)

June 13, 2010

Genetics to solve why Ozzy Osbourne is still alive

[Jack Grimston](#)

THE mystery of why Ozzy Osbourne is still alive after decades of drug and alcohol abuse may finally be solved.

The 61-year-old former Black Sabbath lead singer — who this week begins his health advice column in *The Sunday Times Magazine* — is to become one of only a few people in the world to have his full genome sequenced.

In addition to giving Osbourne information that could help prevent diseases, it is hoped the results will provide insights into the way drugs are absorbed into the body.

The first full genome was sequenced in 2003 after 13 years of work. Today, analysing a genome takes three months and costs about £27,000.

[EXPLORE HEALTH NEWS](#)

[▶ SWINE FLU](#)

Projects

- [Ultra-barcoding...](#)
- [Whole-Genome Se...](#)
- [The newest disc...](#)
- [Sequencing Ozzy...](#)

Sequencing Ozzy Osbourne



Cofactor Genomics LLC., in conjunction with Knome, constructed genomic DNA libraries and sequenced them, generating approximately 39 Gb of sequence data on a newly installed Applied Biosystems HiSeq 2500 at Carlsbad, CA while Knome, of Cambridge, MA, handled the data analysis.

When the analysis and interpretation of the data was complete at Knome, we went to the UK to present our findings, comparing Ozzy's genome sequence to the 1000 Genomes Project Library of Medicine and human reference genomes. We discovered that Ozzy has several family-specific Haplotypegroup-T and Haplotypegroup-U variants shared by Colbert and Henry "Skip" Gates. Ozzy

has 10 times more Neanderthal DNA than Ozzy.

Other interesting comparisons showed Ozzy is 6 times more likely than the average person to have a dependency to alcohol while showing a lower than average predilection to heroine and nicotine addiction (cigarettes were the first thing he gave up several years ago when he went clean). Based on these results, it is no surprise that he drank several bottles of cognac a day for years. Interestingly, how he was able to handle that amount of alcohol may be explained by a mutation in the regulatory region of his ADH4 gene that metabolizes alcohol. This variation could have allowed him to process the alcohol at a faster rate than the normal person, leading to less health risks.

One of the most interesting findings was Ozzy has two versions of the COMT gene (Catechol-O-methyltransferase) called "warrior" and "worrier". This is an enzyme that degrades dopamine, epinephrine, and norepinephrine. The "warrior" variant has been implicated in increased executive functions such as awareness, planning, organization, self-awareness, and potentially most important for Ozzy, self-regulation. While the "worrier" variant has been implicated in a decrease of these functions. In Ozzy's own words, "I always thought it was just the booze and drugs that made me do crazy things like that, even though I've always been a hypochondriac, and in some ways quite an anxious and insecure person. Maybe it's more to do with my genes. Those two sides of my personality sum me up perfectly. Being a warrior, the crazy bat-eating Prince of Darkness, has made me famous. Being a worrier has kept me alive when some of my dearest friends never made it beyond their mid-twenties."

LETTERS

The complete mitochondrial DNA genome of an unknown hominin from southern Siberia

Johannes Krause¹, Qiaomei Fu¹, Jeffrey M. Good², Bence Viola^{1,3}, Michael V. Shunkov⁴, Anatoli P. Derevianko⁴ & Svante Pääbo¹

With the exception of Neanderthals, from which DNA sequences of numerous individuals have now been determined¹, the number and genetic relationships of other hominin lineages are largely unknown. Here we report a complete mitochondrial (mt) DNA sequence retrieved from a bone excavated in 2008 in Denisova Cave in the Altai Mountains in southern Siberia. It represents a hitherto unknown type of hominin mtDNA that shares a common ancestor with anatomically modern human and Neanderthal mtDNAs about 1.0 million years ago. This indicates that it derives from a hominin migration out of Africa distinct from that of the ancestors of Neanderthals and of modern humans. The stratigraphy of the cave where the bone was found suggests that the Denisova hominin lived close in time and space with Neanderthals as well as with modern humans^{2–4}.

The first hominin group to leave Africa was *Homo erectus* about 1.9 million years (Myr) ago⁵. Archaeological as well as genetic data indicate that at least two groups of hominins left Africa after this event: first, the ancestors of the Neanderthals between 500,000 and 300,000 years ago (500 and 300 kyr ago, respectively), presumably *Homo heidelbergensis* or *Homo rhodesiensis*^{6–9}; and, second, anatomically

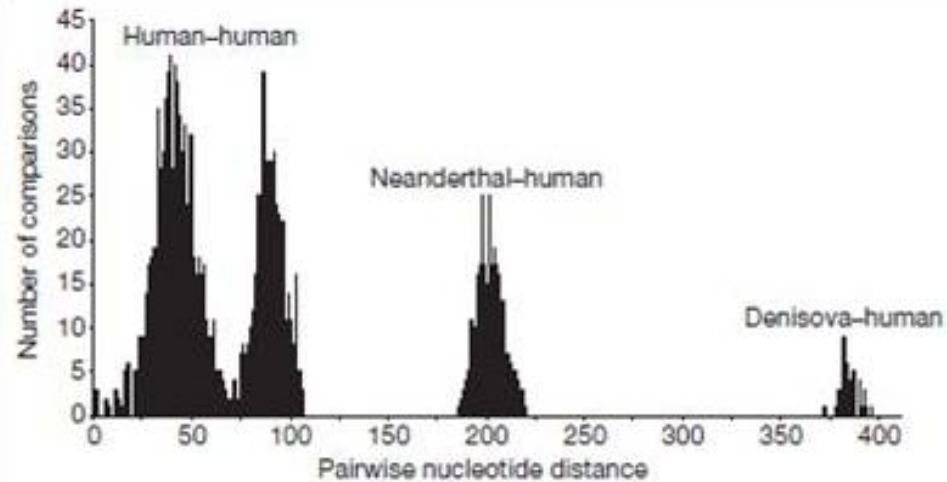


Figure 2 | Distribution of pairwise nucleotide differences. Pairwise nucleotide differences from all pairs of complete mtDNAs from 54 present-day and one Pleistocene modern human, six Neanderthals and the Denisova hominin are shown.

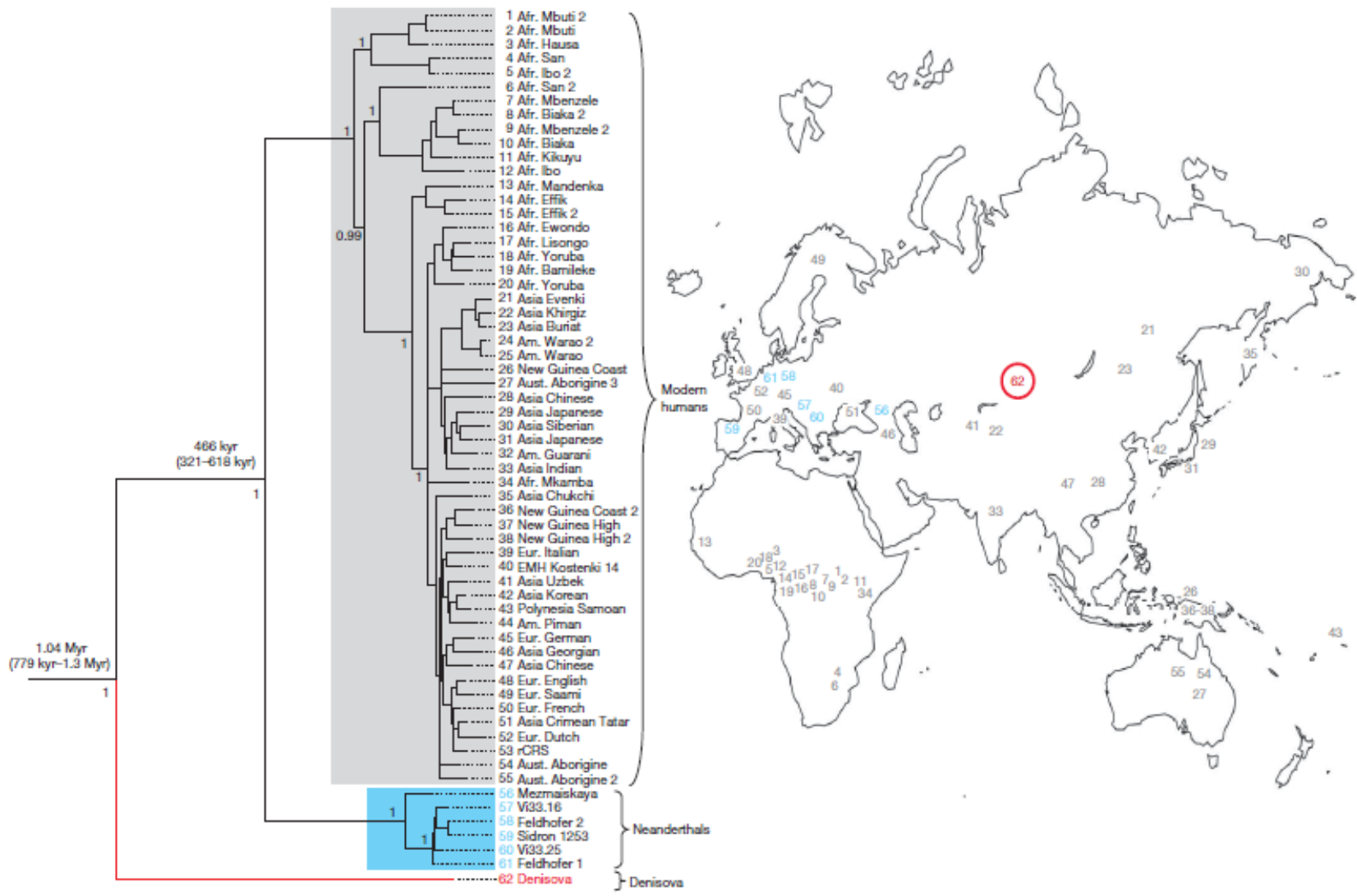


Figure 3 | Phylogenetic tree of complete mtDNAs. The phylogeny was estimated with a Bayesian approach under a GTR+I+ Γ model using 54 present-day and one Pleistocene modern human mtDNA (grey), 6 Neanderthals (blue) and the Denisova hominin (red). The tree is rooted with a chimpanzee and a bonobo mtDNA. Posterior probabilities are given for

each major node. The map shows the geographical origin of the mtDNAs (24, 25, 32, 44 are in the Americas). Note that two partial mtDNAs sequenced from Teshik Tash and Okladnikov Cave in Central Asia fall together with the complete Neanderthal mtDNAs in phylogenies⁴ (not shown).

GENOMIKA - előadás tematika

1. **Szept.15.** Hogyan változtak az elképzelések a genom tartalmáról?
Szervezeti komplexitás és géntartalom (Egyed B)
2. **Szept.22.** Human Genome Project. Genom szekvenálási stratégiák és next generation sequencing (Egyed B)
3. **Szept.29.** A transzkripció szabályozása. Transzkripciós célpontok kimutatása genomi szinten (Varga M)
4. **Okt. 6.** Epigenetika (Varga M)
5. **Okt.13.** Állati genomok: Metazoa evolúció genomikai szemszögből (Varga M)
6. **Okt.20.** Prokarióta és vírus genomszerkezet valamint evolúciójuk (Varga M)
7. **Okt.27.** A földi élet törzsfájának rekonstruálása ritka genomi változások alapján (Ari E)

GENOMIKA - előadás tematika

7. **Nov.10.** A humán genom szerkezete és jellemzése. Gének, szabályozó szekvenciák, mobilis genetikai elemek, pszeudogének (Egyed B)
8. **Nov.17.** Genetikai variabilitás és fenotípus. Variációk a genomban: DNS ujjlenyomat. Asszociációs vizsgálatok (Egyed B)
9. **Nov.24.** Növényi genomok, *GMO* (Kaló P, Gödöllő)
10. **Dec.1.** Génexpressziós vizsgálatok, transzkriptómika (Puskás L, Szeged)
11. **Dec.8.** Ivari kromoszómák: eredet, diverzitás. Y kromoszóma degeneráció. X kromoszóma átrendeződés (Varga M)
12. **Dec.15.** Vizsganap

Ajánlott irodalom, tananyag és követelmények:

The Origins of Genome Architecture

szerző: Michael Lynch

kiadó: Sinauer Associates, Inc. Publishers, 2006

A Primer of Human Genetics

szerző: G. Gibson

kiadó: Sinauer Associates, Inc. Publishers, 2015

The Evolution of the Genome

szerkesztő: T. Ryan Gregory

kiadó: Elsevier Academic Press, 2005

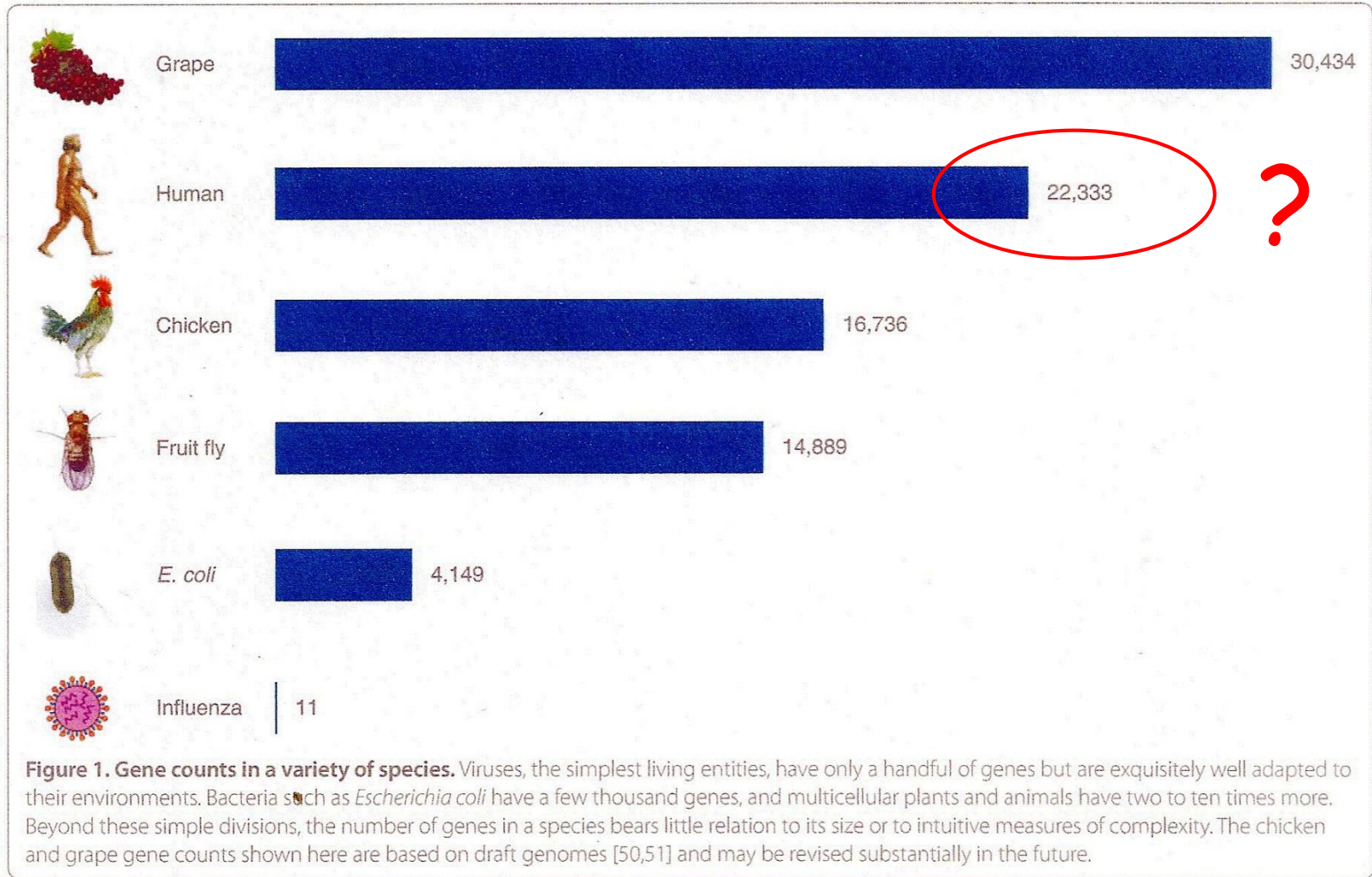
ELTE Genetikai Tsz.: <https://genetics.elte.hu>

user: genetika2017

pw: genetika2017

Kollokvium: írásbeli vizsga december 15., esetleg később is

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

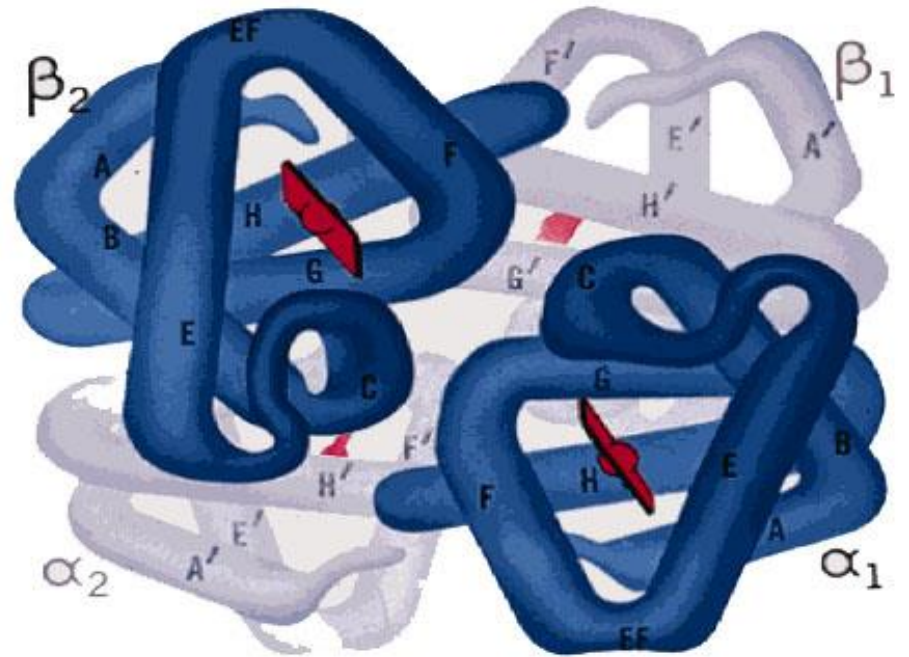


„ ... valahol a csirke és a szőlő között” (Pertea & Salzberg, *Genome Biology* 2010, 11:206)

Első becslések a humán genom méretéről, a gének számáról

1964: F. Vogel (Heidelberg)

- Hemoglobin α és β lánc
- leegyszerűsített feltevések
- Humán 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

Mit nevezünk génnek, hogyan definiálható?

A Gén fogalmának jelentős átalakulása az elmúlt száz esztendőben:

- protein/RNS kódolás, intron/exon fogalom, szabályozó funkciók, stb.
- disztinkció a funkcióban

Jelenlegi definíciónk:

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

Minek nevezzük?

- pl. fehérjét nem kódoló RNS gének (pl. miRNAs, snRNAs)

Automata DNS szekvenálás és „Computer Biology”

ESTs: mRNS poly(A)3' ends → RT-PCR → cDNS könyvtár ('90-)

300 cDNS könyvtár 37 különböző szövetmintából: ~ 87.983 szekvencia (1995)

Adams MD, et al., Nature (1995): → kb. 100 000 gén (NIH/DOE)

ESTs alapján a génszám a 90-es évek végére: 35 000 - 57 000 (CpG islands)

Hogyan találjuk meg a géneket? - Bioinformatikai módszerek:

- protein kódoló szekvencia homológok alapján
- de novo prediktorok szignál motívumok alapján (pl. Genscan: 45.000 gén)
- összehasonlító szoftverek, konzervált mintázatok (pl. Twinscan: 25.600 gén)
- statisztikai modellek (GH Markov Model, CRF: conditional random fields)
- hibás *de novo* predikciók, false positives: pszeudogének
- JIGSAW, Gnomon (NCBI, Ensembl): integratív metodika (2005-)

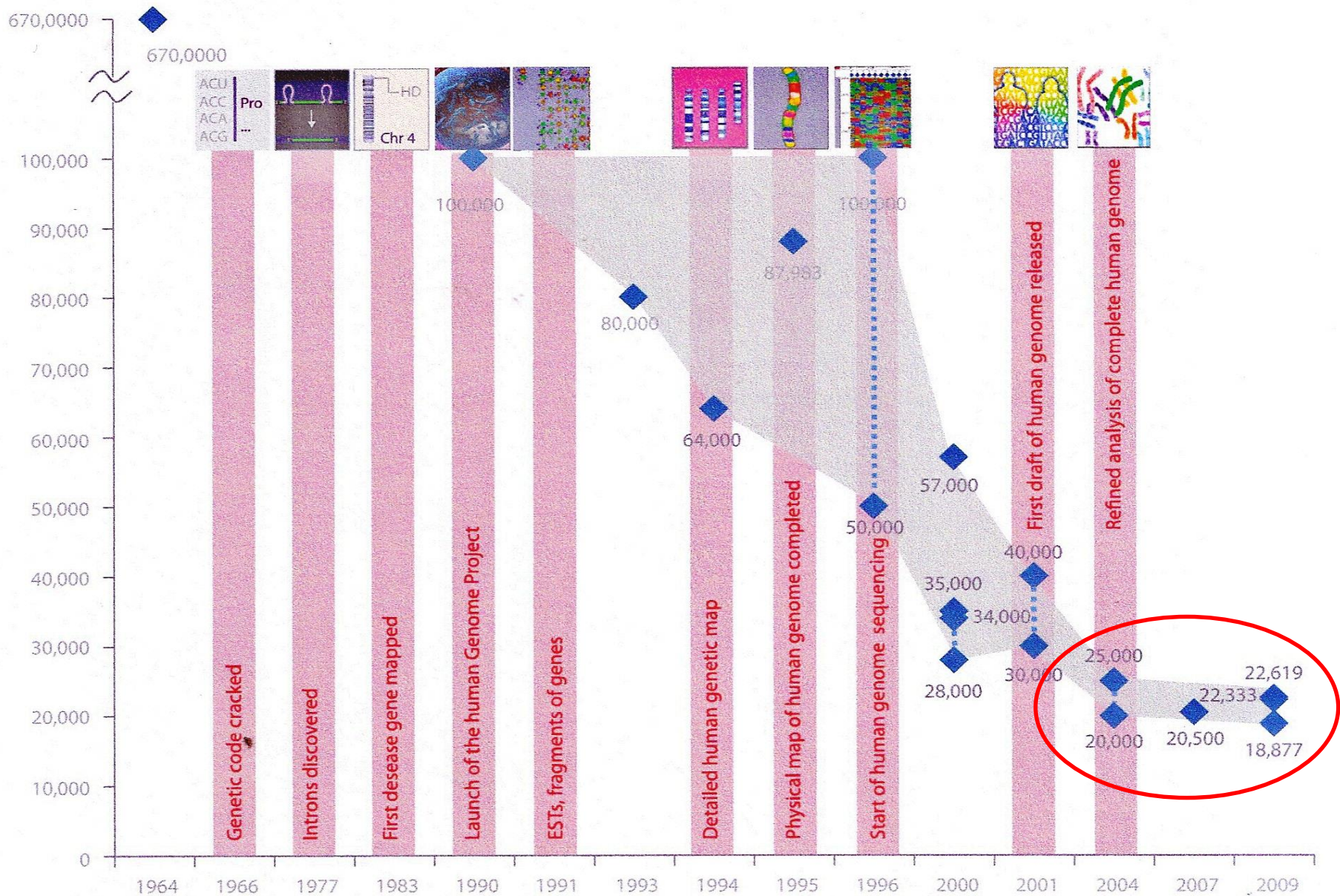


Figure 2. The trend of human gene number counts together with human genome-related milestones. Individual estimates of the human gene count are shown as blue diamonds. The range of estimates at different times is shown by the two vertical blue dotted lines. Note how this range has narrowed in recent 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/CcidsBrowse.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 > 1000 regions)
- 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: non-coding szekvenciákból, ca. 18 új gén?

(Knowles and McLysaght, 2009, Genome Res)

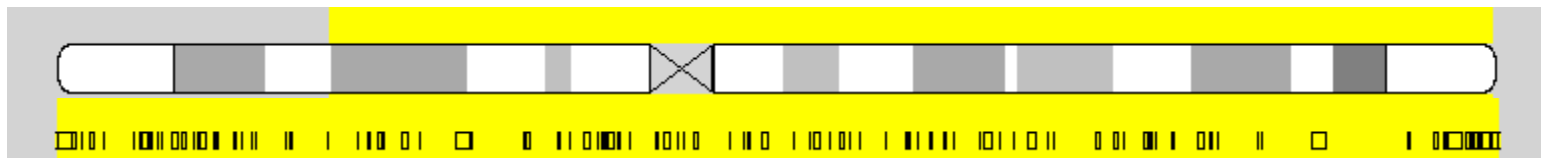
Copy Number Variation (CNV)

Kópia-szám variabilitás

A diploid szervezetek alapesetben minden génből két másolatot hordoznak (homológ párok). Az emberi genom vizsgálata során felismerték, hogy hosszú (ált. több Kb vagy Mb) DNS szakaszok előfordulhatnak kettőnél több példányban is. Ezeket copy-number variation (CNV)-nak nevezték el. Az egyes egyének között a CNV mintázat különböző lehet.

Kb. 300 emberen végzett vizsgálatban 1447 CNV-t mutató genomikus szakaszt azonosítottak, ez kb. a genom 12%-át fedi le.

Sikerült néhány CNV-t betegségekkel kapcsolatba hozni. Pl. a prosztataraák betegség az UGT2B17 gén kópia szám változataival hozták kapcsolatba. Vagy a HIV fertőzéssel szembeni ellenálló képesség a CCL3L1 gén több mint két példányával kapcsolatos.



Az ember 20. kromoszómáján kimutatott CNV-k helyzete és kiterjedése

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
<i>CLLU1</i>	ENSG00000205056	121	42	EST/cDNA: Blood (<u>AJ845165</u> , <u>AJ845166</u>); 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.
<i>C22orf45</i>	ENSG00000178803	159	87 (25 amino acids align with human sequence)	EST/cDNA: Kidney, other (<u>AX747284</u> , <u>AK091970</u> , <u>DA635985</u>); 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.
<i>DNAH10OS</i>	ENSG00000204626	163	90 (75 amino acids align with human sequence)	EST/cDNA: Hippocampus (<u>AK127211</u>); 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.

^aLength in codons of longest in-frame (alignable) ORF starting from any ATG in the region.

^bType of data/database is listed followed by tissue information with database identifiers in parentheses. Underlined accession numbers are full-length, spliced cDNA.

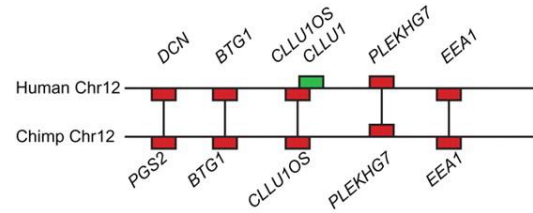
^cShared disablers are sequence differences shared by chimp, gorilla, orangutan, gibbon, and macaque that eliminate the capacity to produce a protein similar to the human protein.

^dIndependently sequenced whole genomes: Venter, Watson, HuAA, HuBB, HuCC, HuDD, and HuFF. All data are listed where available.

^eNot shared with orangutan.

Sequence changes in the origin of CLLU1 from noncoding DNA

A



B

Start

Human
Chimpanzee
Macaque

```
GTTTGGAGG - - - ATGTTCAAC AAATGCTCCTTTCATTCCTCTATTTACAGACC TGCCGCA
GTTTGGAGG - - - ATGTTCAAT AAATGCTGCTTTCAC TCCCTATTTACAGACCTGCCGCA
GTTTGGAGG - - - ATGCTCAAT AAATGCTCCTTTCATTCCTCATTACAAAGCTTGCCGCA
```

Human
Chimpanzee
Macaque

```
GACAATTC TGCTAGCAGCC TTTGTGCTATTATCTGTTTTCTAAACTTAGTAATTGAGTGT
GACAATTC TGCTAGCAGCC TTTGTGCTATTATCTGTTTTCTAAACTTAGTAATTGAGTGT
GACAATTC TGCTAGCAGCC TTTGTGCTATTATCTGTTTTCTAAACTTAGTAATTGAGTGT
```

Human
Chimpanzee
Macaque

```
GATCTGGAGACTAA - CTCTGAAATAAATAAGCTGATTATTTATTTATTTTCTCAAAACAA
GATCTGGAGACTAAACTCTGAAATAAATAAGCTGATTATTTATTTATTTTCTCAAAACAA
TATCTGGAGACTAAACTCTGAAATAAATAAGCTGATTATTTATTTATTTTCTCAAAACAA
```

Human
Chimpanzee
Macaque

```
CAGAATACGATTTAGCAAATTACTTCTTAAGATAT TATTTACATTTCTATATTTCTCCTA
CAGAATACGATTTAGCAAATTACTTCTTAAGATACTATTTACATTTCTATATTTCTCCTA
CAGAATA TGATTTAGCAAATTACTTCTTAAGATAT TATTTGCAC TTCTATATTTCTCCTA
```

Human
Chimpanzee
Macaque

```
CCCTGAGTTGATGTGTGAGCAATATGTCACCTTTCATAAAGCCAGGTATACA - - - TTATG
CCCTGAGTTGATGTGTGAGCCGATATGTCACCTTTCATAAAGCCAGGTATACA - - - TTATG
CCCTGAGTTGATGTGTGAGCAATATGTCACCTTCCACAAGCCAGGTATATATACATTACG
```

Human
Chimpanzee
Macaque

```
          H I I Y S T F L S K
GACAGGTAAGTAAAAAACATATTTATTTATTTACGTTTTTGTCCAAAAATTTTAAATTTCT
GACAGGTAAGTAAAAAACATATTTATTTATTTACGTTTTTGTCCAAAGAAATTTTAAATTTCT
GACAGGTAAGTAAAAAACATATTTATTTATTTACGTTTTTGTCCAAAGAGTTTTTAAATTTCT
```

Human
Chimpanzee
Macaque

```
AACTGTTGCGCGTGTGTTGGTAA - - - TGTA AAACAAACTCAGTACA
AACTGTTGCGCGTGTGTTGGTAA - - - TGTA AAACAAACTCAGTACA
AACTGTTG TGCATGTGTTGGTAA - - - CGTA AAACAAACTCAGTACG
```

C

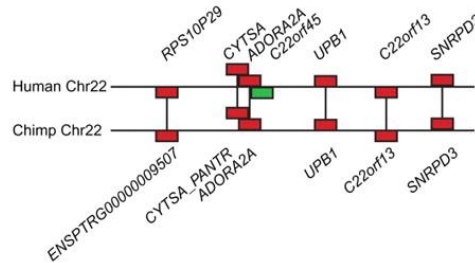


Knowles D G , McLysaght A Genome Res. 2009;19:1752-1759



Sequence changes in the origin of C22orf45 from noncoding DNA

A



B

Start

Human
Chimpanzee
Macaque

CCAG - - - GACATGAGGG - - - ATGGAGCAGGAC TGGCAACC TGGAGAGGAAGTC ACTCC TG
 CCAG - - - GACATGAGGG - - - ATGGAGCAGGAC TGGCAACC TGGAGAGGAAGTC ACTCC TG
 TCAG - - - GACATGAGGG - - - ACGGAGCAGGAT TGGCAACC TGGAGAGGAAGTC AGTCC TG

Human
Chimpanzee
Macaque

GTCC TGAGCCCTGTTTCAAAGGGCCAGGC TCCCTG - TACCCCAT TGTCATGTGACAGAG
 GTCC TGAGCCCTGTTTCAAAGGGCCAGGC TCCCTG T TACCCCAT TGTCATGTGACAGAG
 GTCC TGAGCCCTGTTTCAAAGGGT CAGGC TCCCTG - TACCCCAT TGTCATGTGACGAG

Human
Chimpanzee
Macaque

CTCAAACACACAGACCCCAACTTTCCCTCCAACTCCAAATGCTGTGGCACC TC AAGTGGC
 CTC AAACACACAGACCCCAACTTTCCCTCCAACTCCAAATGCTGTGGCACC TC AAGTGGC
 CTC AAACA - - - GACCCCAACTTTCCCTCCAACTCCAAATGCTGTGAGCACC TC AAGTGGC

Human
Chimpanzee
Macaque

TGGAACAGGAT TGGCACGGGC TGCAGCCAT ACC TGGGAC TGGAGGTTCC TCC TGC ACCCAG
 TGG AACAGGAT TGGCACGGGC TGCAGCCAT ACC TGGGAC TGGAGGTTCC TCC TGC ACCCAG
 TGG AACAGGAT TGGCACAGGC TGCAGCCAT ACC TGGGAC TGGAGGTTCC TCC TGC ACCCAG

Human
Chimpanzee
Macaque

CAGGCCCTTTTGCCCTACTAGGAGCC TGGGAATGGAGCAT TGAACACAGAAGCAGGAGGA
 CAGGCCCTTTTGCCCTACTAGGAGCC TGGGAATGGAGCAT TGAACACAGAAGCAGGAGGA
 CAGGCCCTTTTGCGTCTACTAGGAGCC T TGGGAATGGAGCAT TGAACACAGAAGCAGGAGGA

Human
Chimpanzee
Macaque

GGAAGGAGAGAGCAGAG - CCAGAAACCC TGCAGCAACGGAGGGCC TGCAGCAGC TGGAGA
 GGAAGGAGAGAGCAGAGCCAGAAACCC TGCAGCAACGGAGGGCC TGCAGCAGC TGGAGA
 GGAAGGAGAGAGCAGAGACCAGAGACCC TGCAGCAAAATGGAGGGCC TGCAGCAGC TGGAGA

Human
Chimpanzee
Macaque

GGGCCGAGTCCCTCCC AAGCCCC TGC TTTCC ATGAGC ACTTGCCAGGCAGCCATTCACAA
 GGGCTGAGTCCCTCCC AAGCCCC TGC TTTCC ATGGGGCAC TTGCCAGGCAGCCATTCACAA
 GGGCTGAGTCCCTCCC AAGCCCC TGC TTTCC ATGGGGCAC TTGCCAGGCAGCCATTCACAA

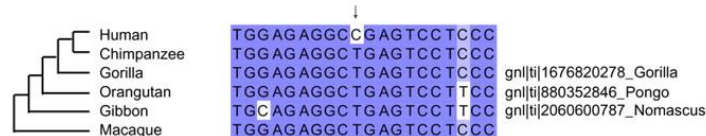
Human
Chimpanzee
Macaque

AGTGTGTCGTTGGCAGGGATGCC ACCAGACCAGC TC TCC TGGC ACC ATCC TGGCCACACT
 AGTGTGTTGTTGGCAGGGATGCC ACCAGACCAGC TC TCC TGGC ACC ATCC TGGCCACACT
 AGCATGTCC TGGCAGGGATGCC ACCAGACCAGC TC TCC TGGC ACC ATC TCTGGCCACACT

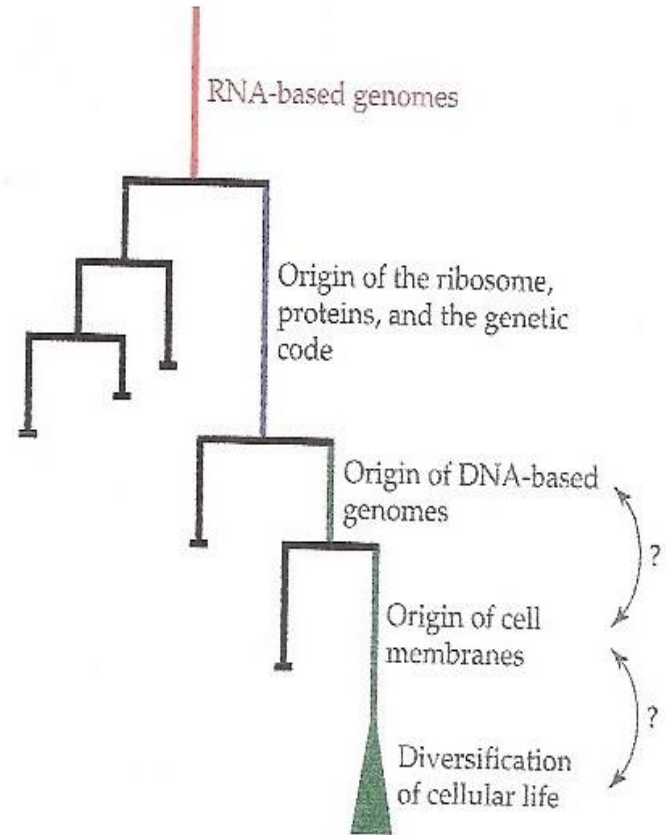
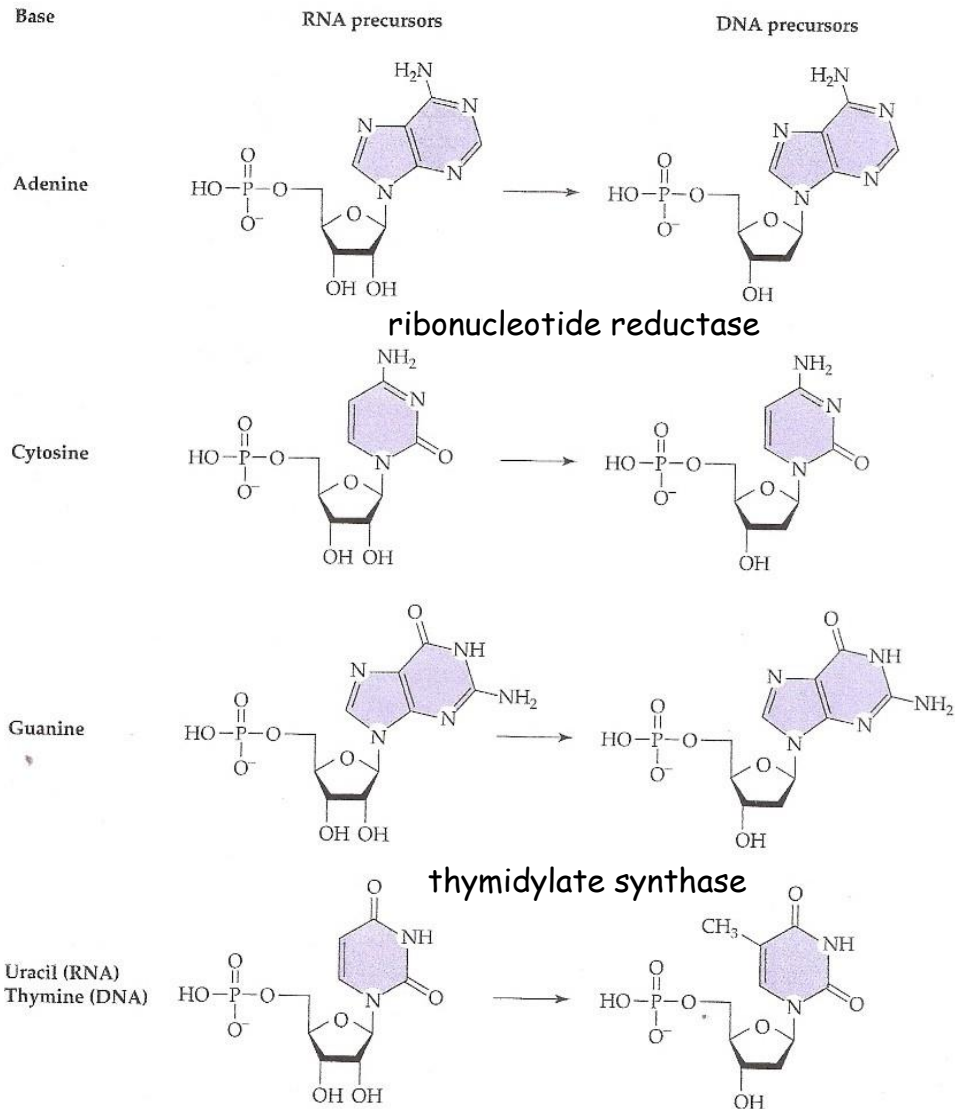
Human
Chimpanzee
Macaque

CAAGGAACACAGTTATCCC TGA - - - TGC TCTGGC
 CAAGGAACACAGTTATCCC TGA - - - TGC TCTGGC
 CAAGGAACACAGTTATCCC TGA - - - TGTTC TGGC

C



Eukarióta genom eredete: RNA world



2'-OH instabilitás / Mutáció: C » U

különböző útvonalak a membrán lipid szintézisben archea és eubacteria között:

- izoprén éther vs. zsírsav észter

Genomok evolúciója rRNA szekvenciák alapján

Woese and Fox, 1977
Woese et al., 1990

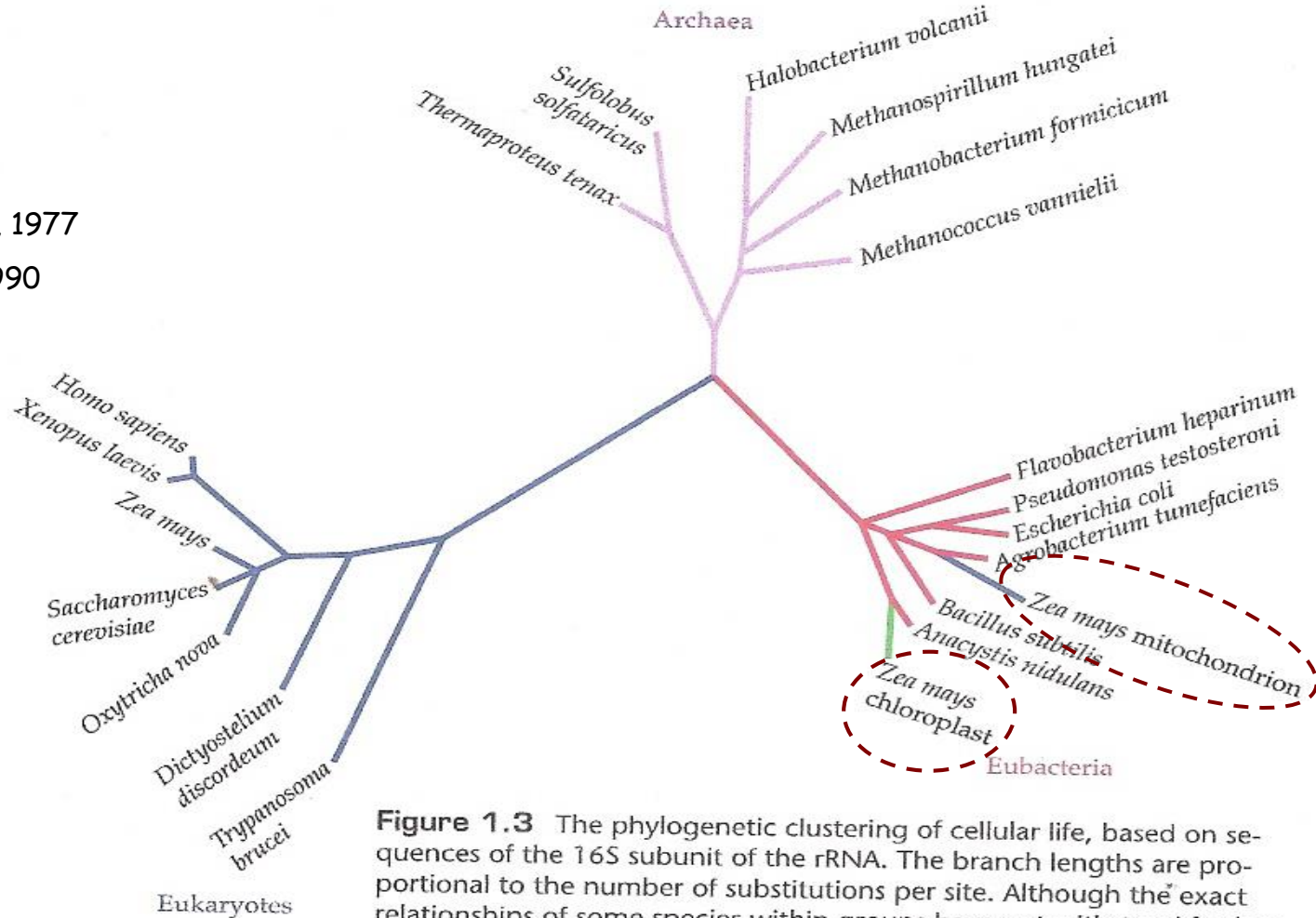


Figure 1.3 The phylogenetic clustering of cellular life, based on sequences of the 16S subunit of the rRNA. The branch lengths are proportional to the number of substitutions per site. Although the exact relationships of some species within groups have not withstood further scrutiny, the distinct nature 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.)

Genom evolúció duplikált gén szekvenciák alapján

ATPase membrán
duplikált alegységek:

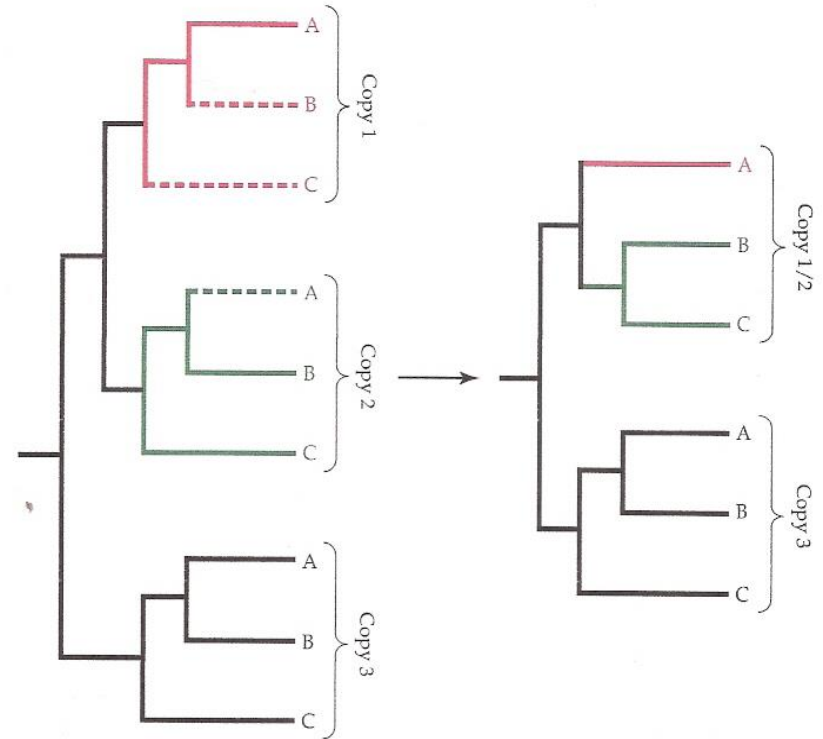
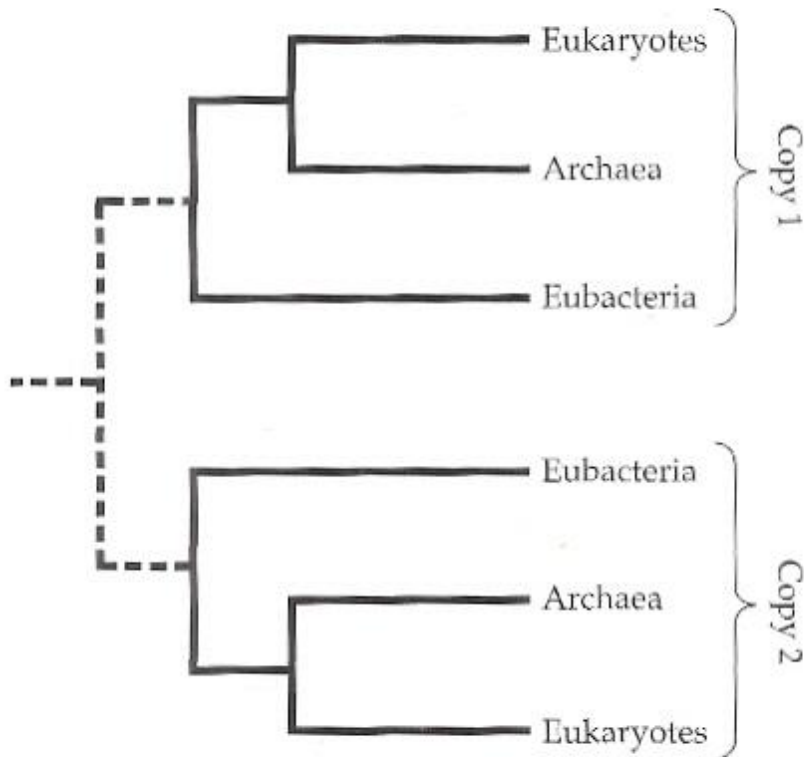


Figure 1.5 *Left:* Here the ancestral species had three gene copies (the first duplication led to copy 3 and the ancestor to 1 and 2; and the second duplication created copies 1 and 2). Copy 1 was lost in species B and C, and copy 2 was lost in species A (dashed lines). *Right:* Using just the surviving gene copies, a conflicting signal is obtained on the phylogenetic relationships of species A, B, and C, even though the overall topology of extant gene relationships is correct. The top cluster incorrectly implies a phylogeny in which species B and C are grouped together (as a consequence of an incorrect mixture of copy 1 and 2 genes), whereas the bottom cluster correctly groups A and B.

Az Eukarióta genom eredete: archea-eubacteria kiméra?

transzkripció és transzláció: **Archea**

housekeeping funkciók: **Eubacteria**

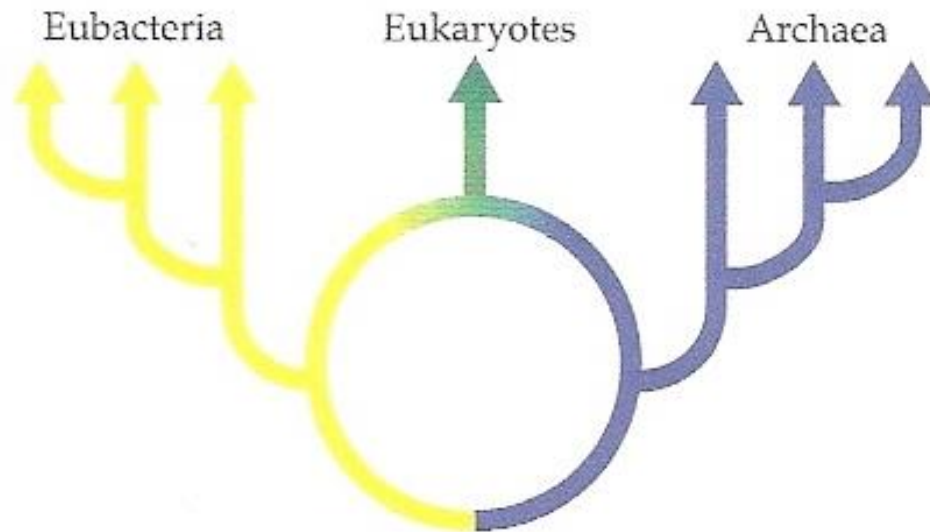


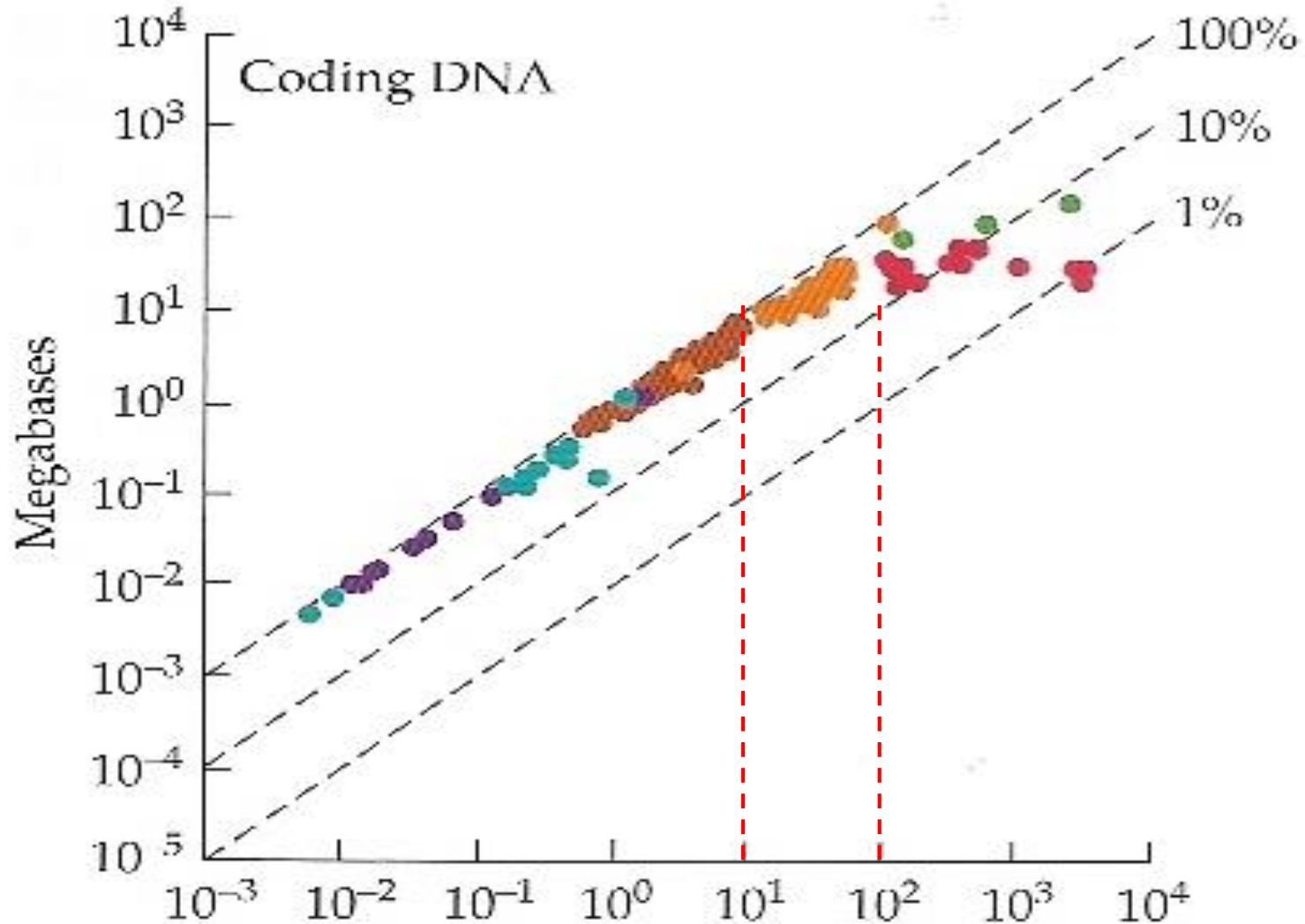
Figure 1.7 The “ring-of-life” hypothesis for the origin of eukaryotes. Yellow and blue lineages denote branches in the phylogenetic trees for eubacteria and archaea, respectively. Members of two such lineages fused to form the eukaryotic domain (green). (Modified from Rivera and Lake 2004.)

Eukarióta genom versus prokarióta

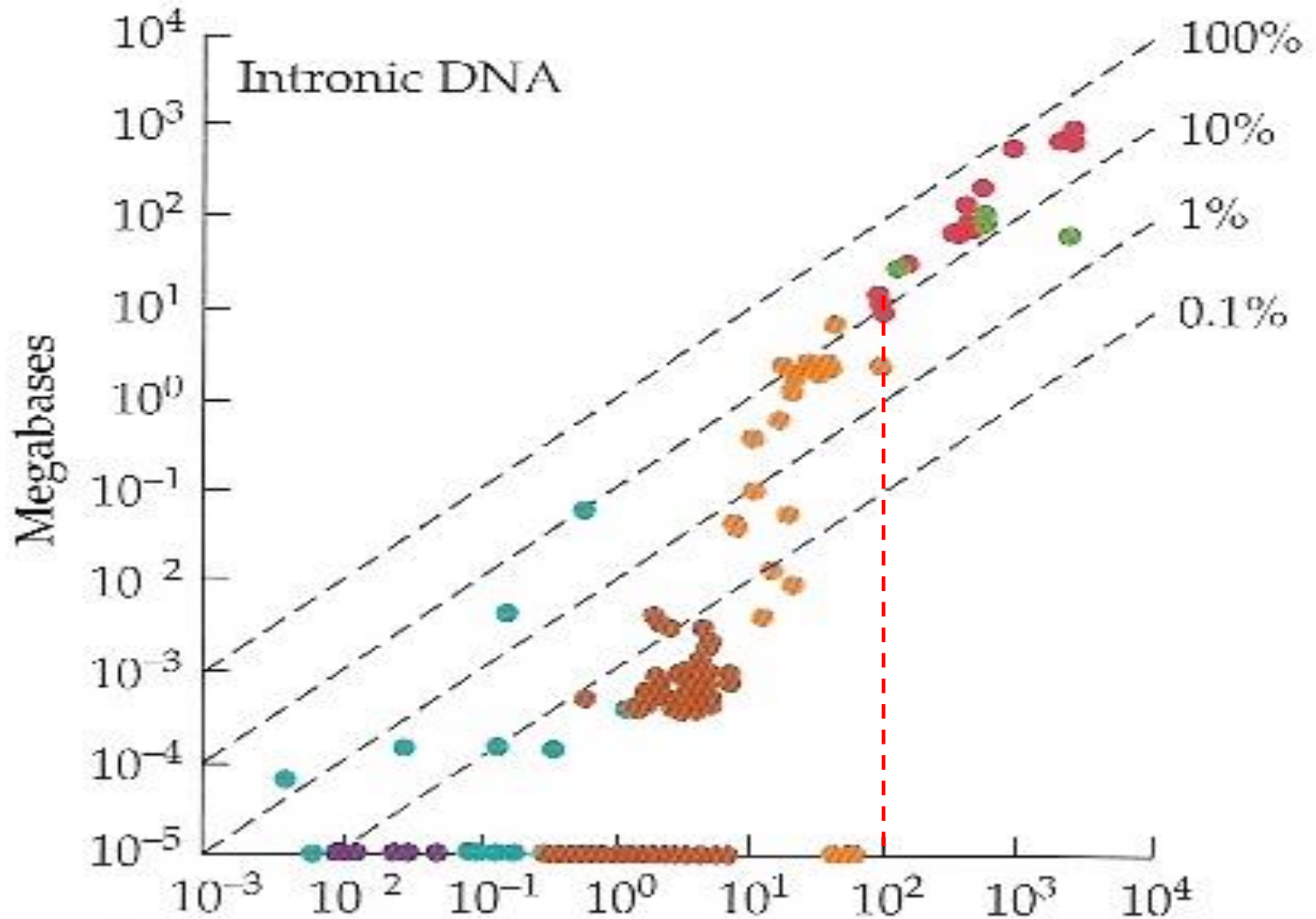
TABLE 1.1 Some of the features that set eukaryotic genomes apart from those of prokaryotes, and their exceptions

EUKARYOTES	PROKARYOTES
Presence of a nuclear membrane	Also present in the Planktomyces
Organelles derived from endosymbionts	Also present in the β -proteobacteria
Cytoskeleton and vesicle transport machinery	Tubulin-related proteins, but not microtubules
<i>Trans</i> -splicing	Absent
Introns in protein-coding genes, and a complex spliceosomal apparatus for excising them	Rare self-splicing introns, but almost never in coding DNA
Expansion of the untranslated regions of transcripts	Untranslated regions are generally very short
Addition of poly(A) tails to all mRNAs	Rare and nonessential polyadenylation of transcripts
Translation initiation by scanning for start codon	Ribosome binds directly to a Shine-Dalgarno sequence
Messenger RNA surveillance	The nonsense-mediated decay pathway is absent
Multiple linear chromosomes capped with telomeres	Single linear chromosomes in a few eubacteria
Mitosis and meiosis	Absent
Expansion in gene number	The largest prokaryotic genomes contain more genes than the smallest eukaryotic genomes
Expansion of cell size and number	A few have very large cell sizes (e.g., <i>Thiomargarita</i>), and several produce multiple cell types

Genom méret vs. kódoló szekvenciák



Genom méret vs. intronok



Genom méret vs. intergénikus DNS

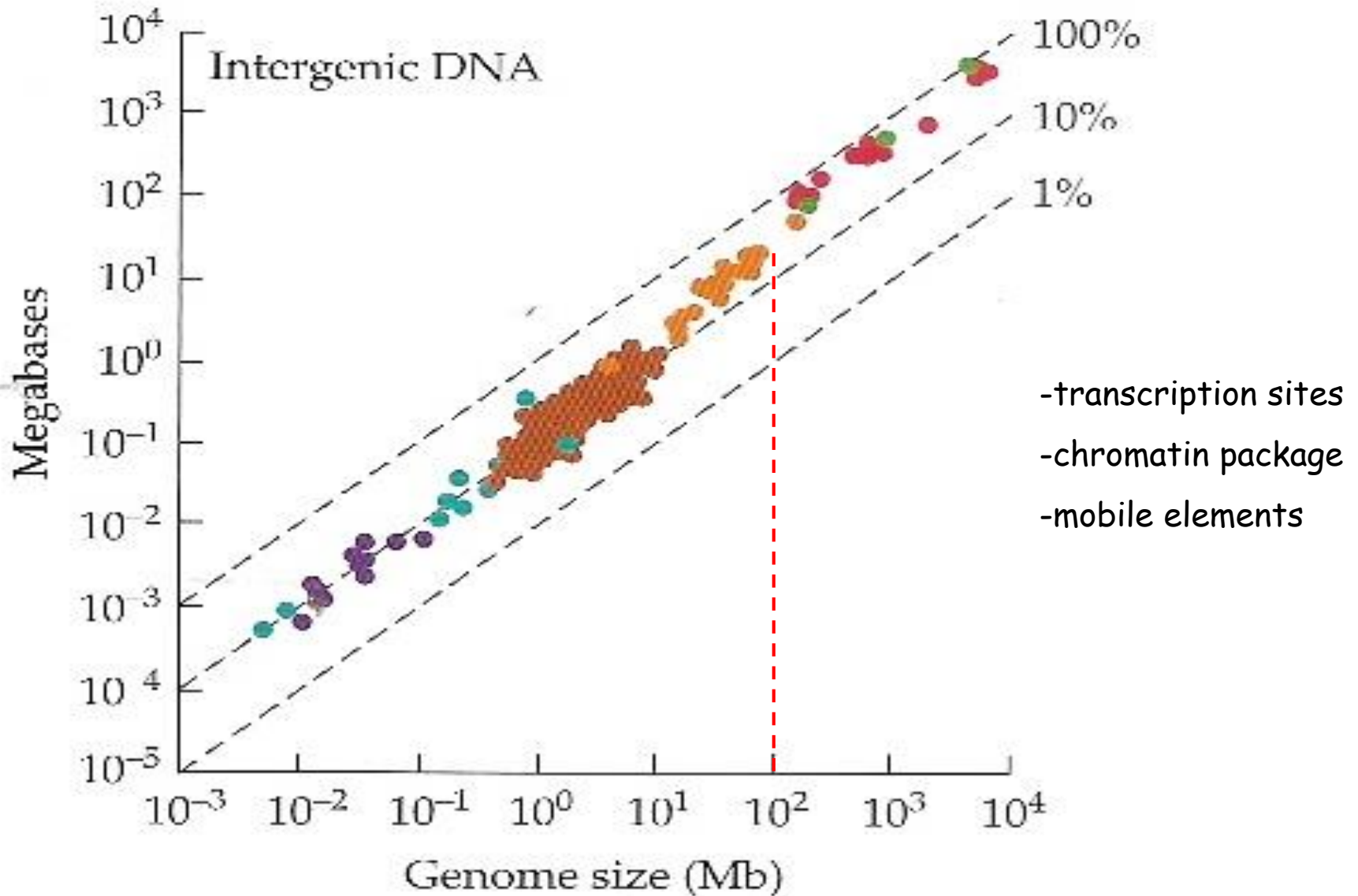


TABLE 3.2 Haploid genome size, number of protein-coding genes, and average number of nucleotides per gene for some well-characterized eukaryotic genomes

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

- WGC: rekurrens mutációk teljes egyedi genomok összehasonlítása révén (pl. Y-kromoszóma), divergens kromoszómák
- 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 / szerkezeti komplexitás

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

Mi következik ebből?

- világos átmenet az egyszerűbb vírus, prokarióta, egysejtű szervezetektől a többsejtű organizmusok irányában a genomméret, génszám, inergénikus és intron szakaszok, stb. vonatkozásában.
- nincsenek hirtelen diszkontinuus szakaszok a genomméret és genetikai tartalom függvényében az alapvetően különböző szervezetek csoportjai között (C-value paradox: variábilis genom)
- indirekt bizonyíték arra, hogy a genom szerveződés elsősorban nem a sejt szerkezeti és a szervezeti komplexitásának következménye.