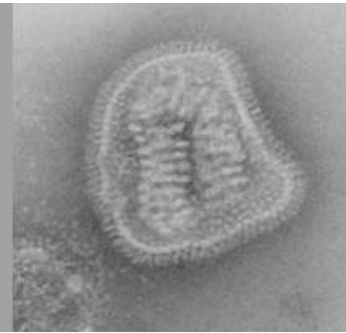




Genomika

Prokarióta és vírus genomok, eukarióta organelleumok



Varga Máté – ELTE Genetikai Tanszék
(mvarga@ttk.elte.hu)

2019.11.18.

Genom méretek

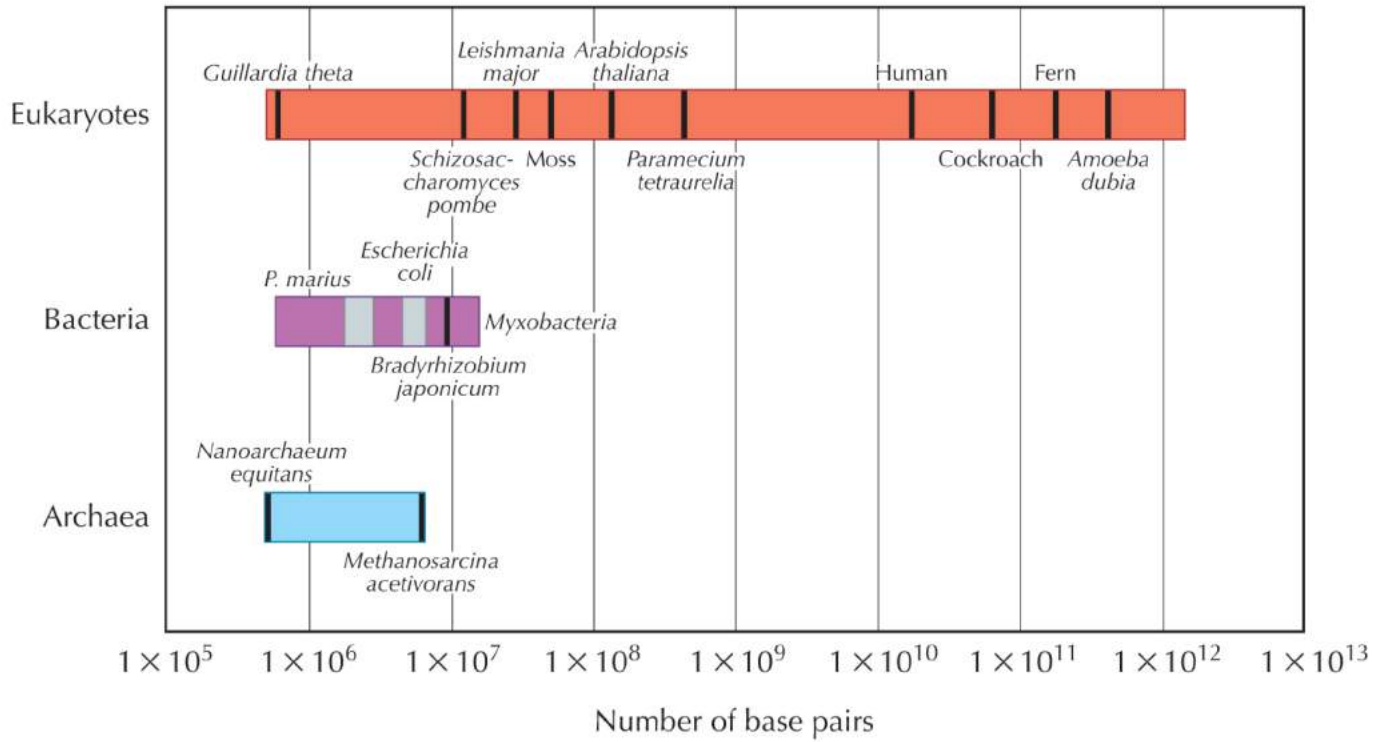
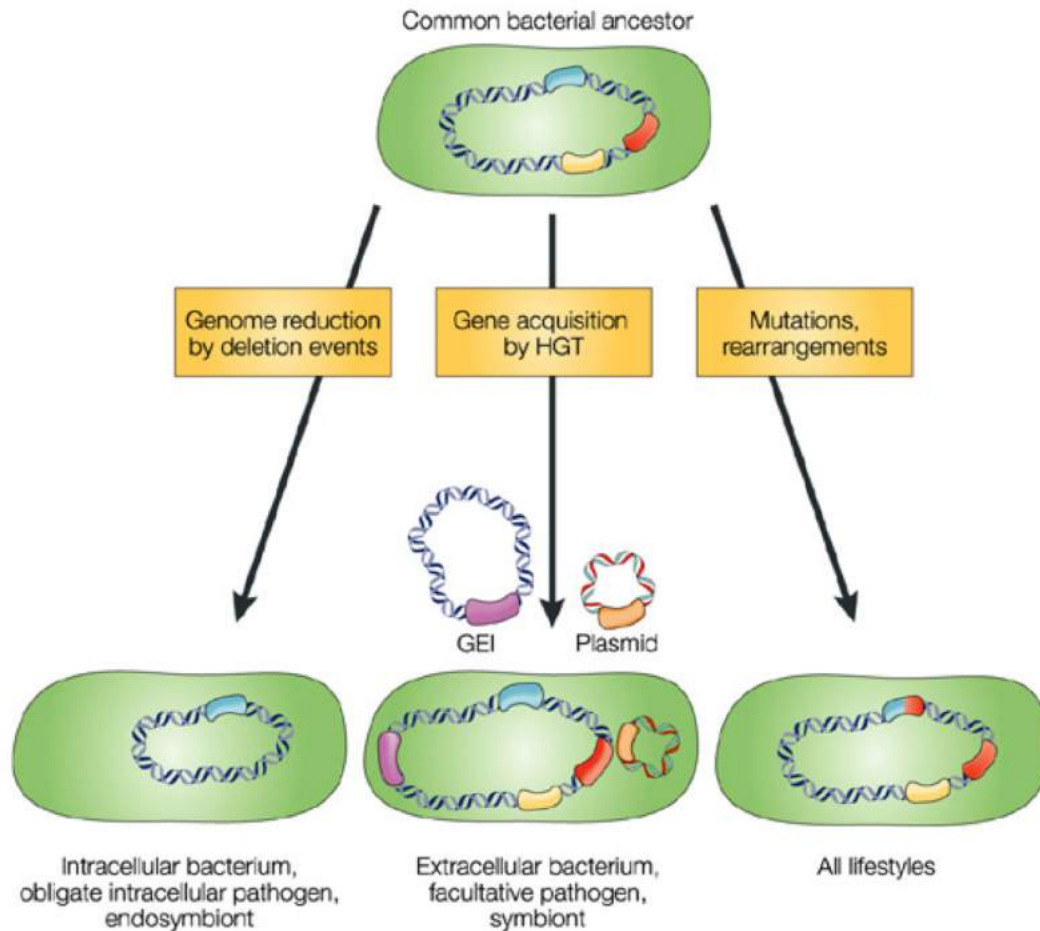


FIGURE 7.1. Genome sizes in the three domains of life. A selection of genome sizes and size ranges from specific groups of organisms is indicated.

7.1, adapted from Bentley S.D. et al., *Annu. Rev. Genet.* **38**: 771–791, © 2004 Annual Reviews, www.annualreviews.org, based on data from DOGS <http://www.cbs.dtu.dk/databases/DOGS/>



A genom struktúrája a baktérium “életvitelét” tükrözi



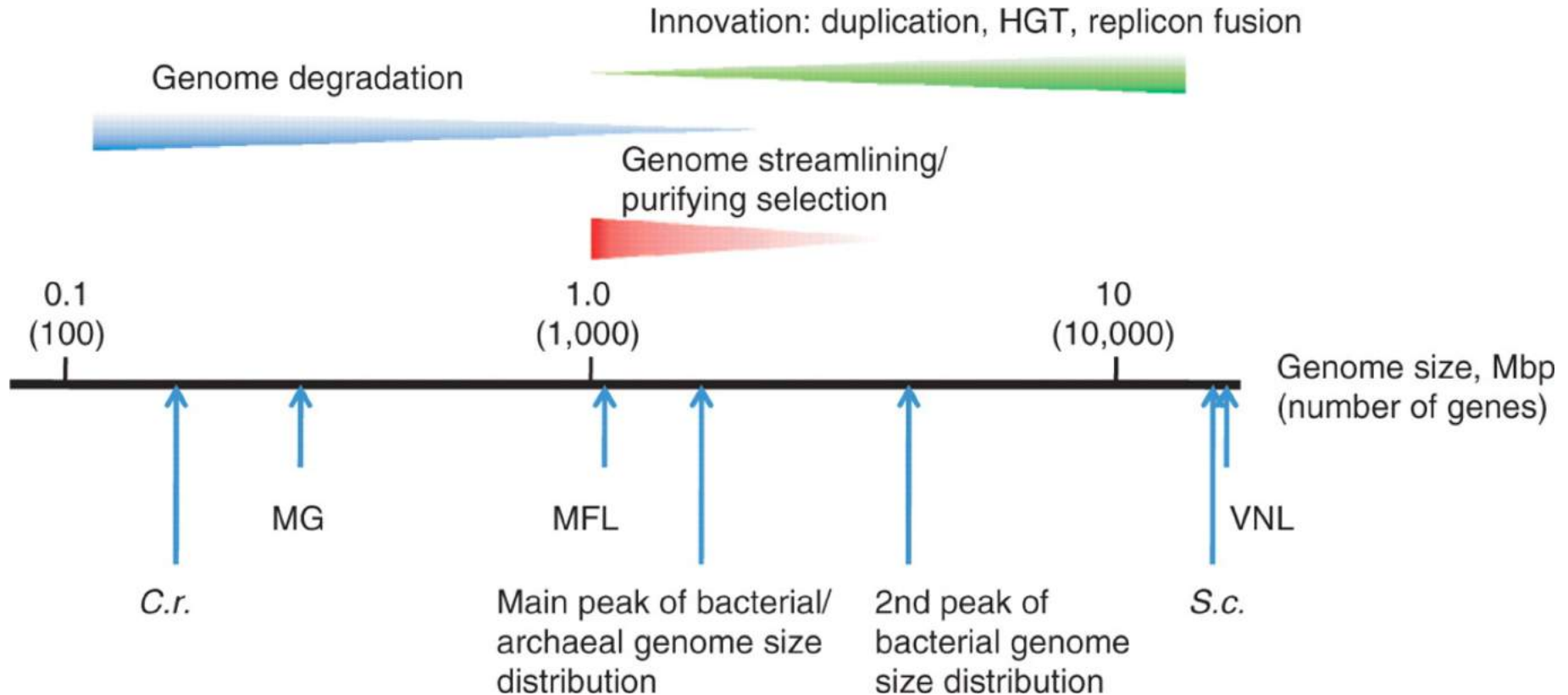
Nature Reviews | Microbiology

-a **stabil környezet** általában genom redukcióhoz vezet (az adott nicheben nem fontos gének elvesznek)

- a **változó életkörülmények** közt található baktériumok genomja nagyobb, mert több eshetőségre kell felkészülniük (különböző körülmények közt más-más gének biztosíthatják a túlélést.)



A bakteriális genom méretét befolyásoló hatások





A stabil környezet genom-redukcióhoz vezet

Table 1 General features of the genomes of *B. pertussis*, *B. parapertussis* and *B. bronchiseptica*

	<i>B. pertussis</i>	<i>B. parapertussis</i>	<i>B. bronchiseptica</i>
Size (bp)	4,086,186	4,773,551	5,338,400
G+C content (%)	67.72	68.10	68.07
Coding sequences	3,816	4,404	5,007
Pseudogenes	358 (9.4%)	220 (5.0%)	18 (0.4%)
Coding density (intact genes)	82.9%	86.6%	91.4%
Coding density (all genes)	91.6%	92.2%	92.0%
Average gene size (bp)	978	987	978
rRNA operons	3	3	3
tRNA	51	53	55
IS481	238	0	0
IS1001	0	22	0
IS1002	6	90	0
IS1663	17	0	0

- az obligát humán patogén *Bordetella pertussis* a kormoszómája 20%-át elvesztette (a *B. bronchiospetica* számos fajt képes megfertőzni)
- a megmaradó géneinek 10%-a pszeudogén
- az inaktiváció egyik hajtóeleme egy IS elem elterjedése

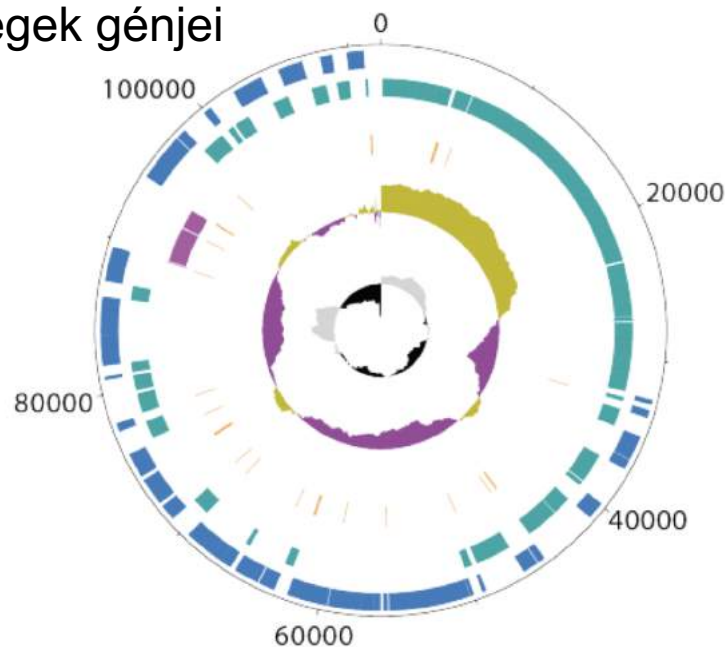
A legkisebb (ismert) bakteriális genom:

Nasuia deltocephalinicola



- a növényi nedvekkel táplálkozó Hemipterák táplálékából hiányzanak az esszenciális aminosavak (10 db), ezeket obligát szimbionták termelik
- a *Sulcia* fajok a legfontosabb ilyen szimbionták, de egyes kabócákban ezek csak 8 aminosavat képesek létrehozni
- a *Sulciáknak* maguknak is igen redukált genomjuk van, a *Macrosteles quadrilineatus* kabóca szimbiontája esetében mindössze 190 Kb, 190 fehérjét kódoló génnel, nincs DNS repair, oxidatív foszforiláció, hiányzanak egyes DNS polimeráz-alegységek génjei

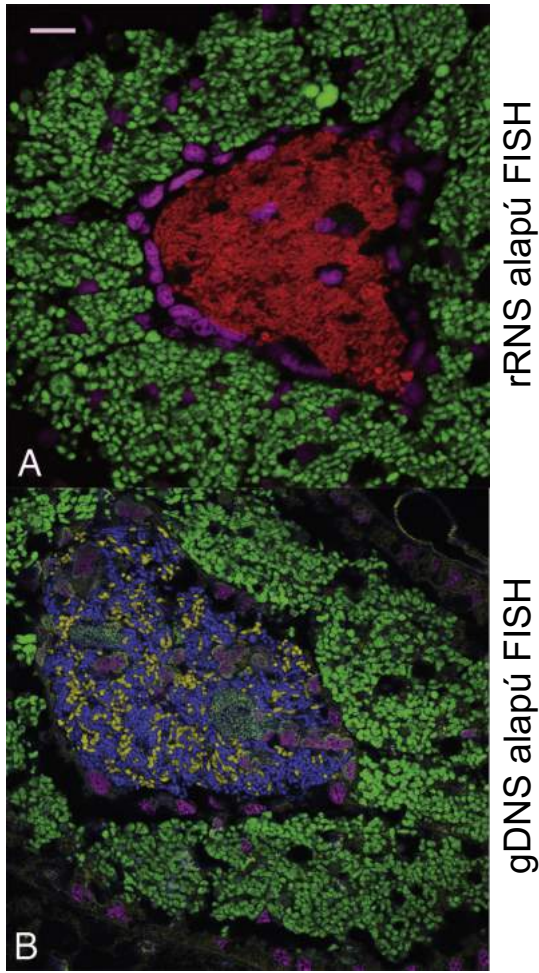
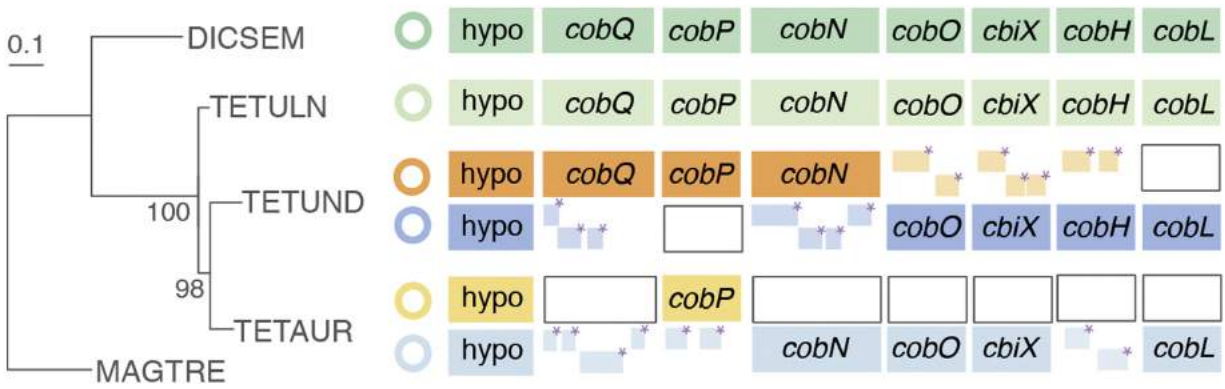
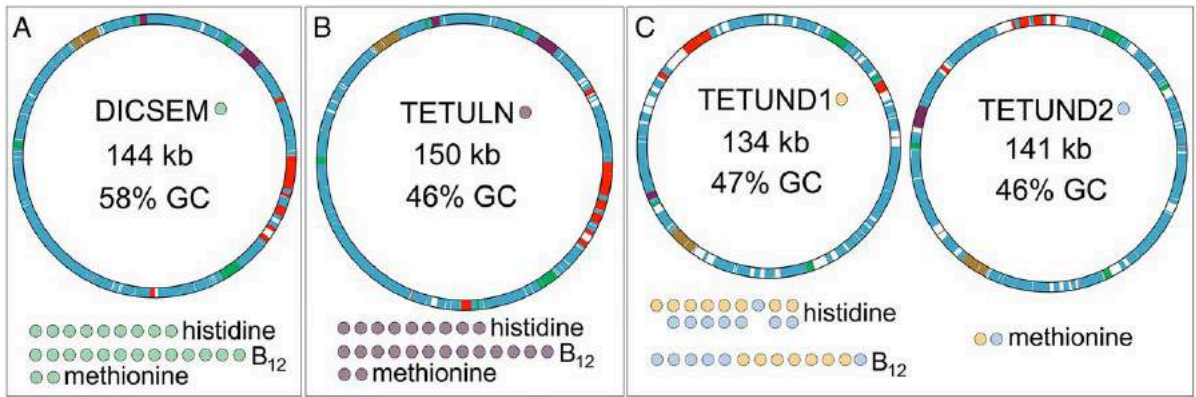
- a legkisebb ismert genom a *M. quadrilineatus* egy másik endoszimbiontája, a *Nasuia deltocephalinicola* nevéhez fűződik
- ez termeli a *Sulcia* által nem biztosított két aminosavat (Met, His) – más útvonal nem is nagyon van épen
- 112 Kb genom, 137 fehérjekódoló gén, alternatív genetikai kód (UAS: STOP -> Trp)



(Bennett and Moran (2013) *Genome Biol Evol*)



Komplementer genomredukció szimbiontákban

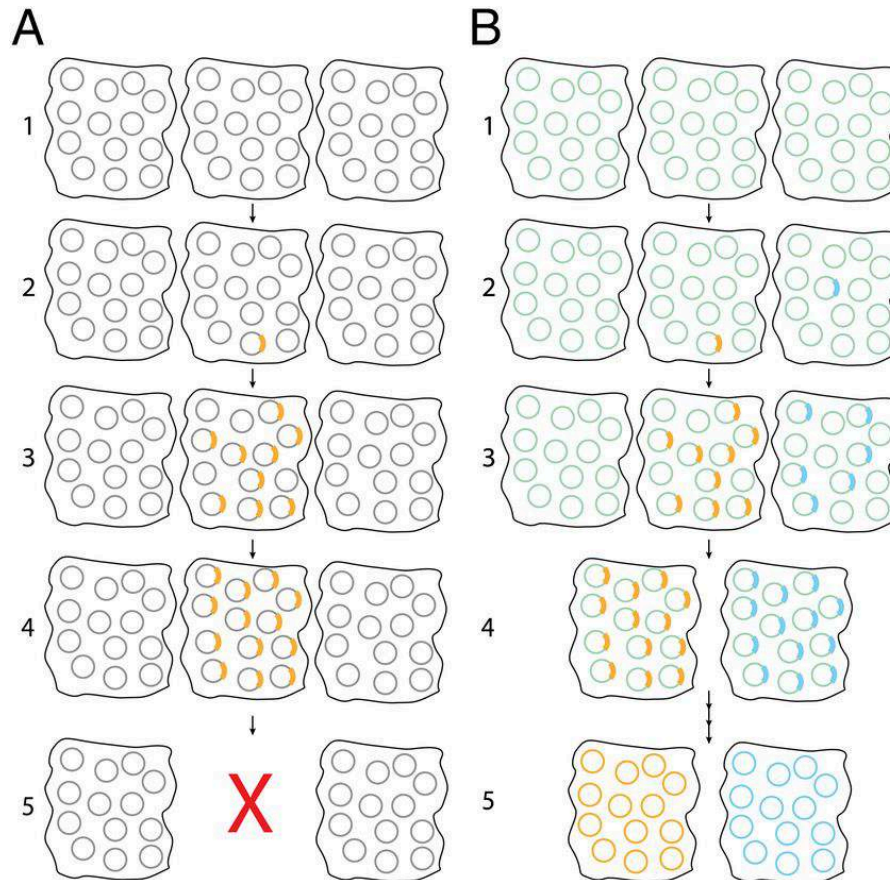


- más kabócáktól eltérően a *Tettigades undata*-ból izolált *Hodgkinia* genomok két kromoszómán szerelhetők csak össze
- ezek a genomok komplementer módon tartalmaznak anyagcsereutakat
- csak genom-alapú festéssel különböztethetők meg, rRNS-alapúval nem

(Van Leuven et al. 2014 *Cell*, Campbell et al. 2014 *PNAS*)



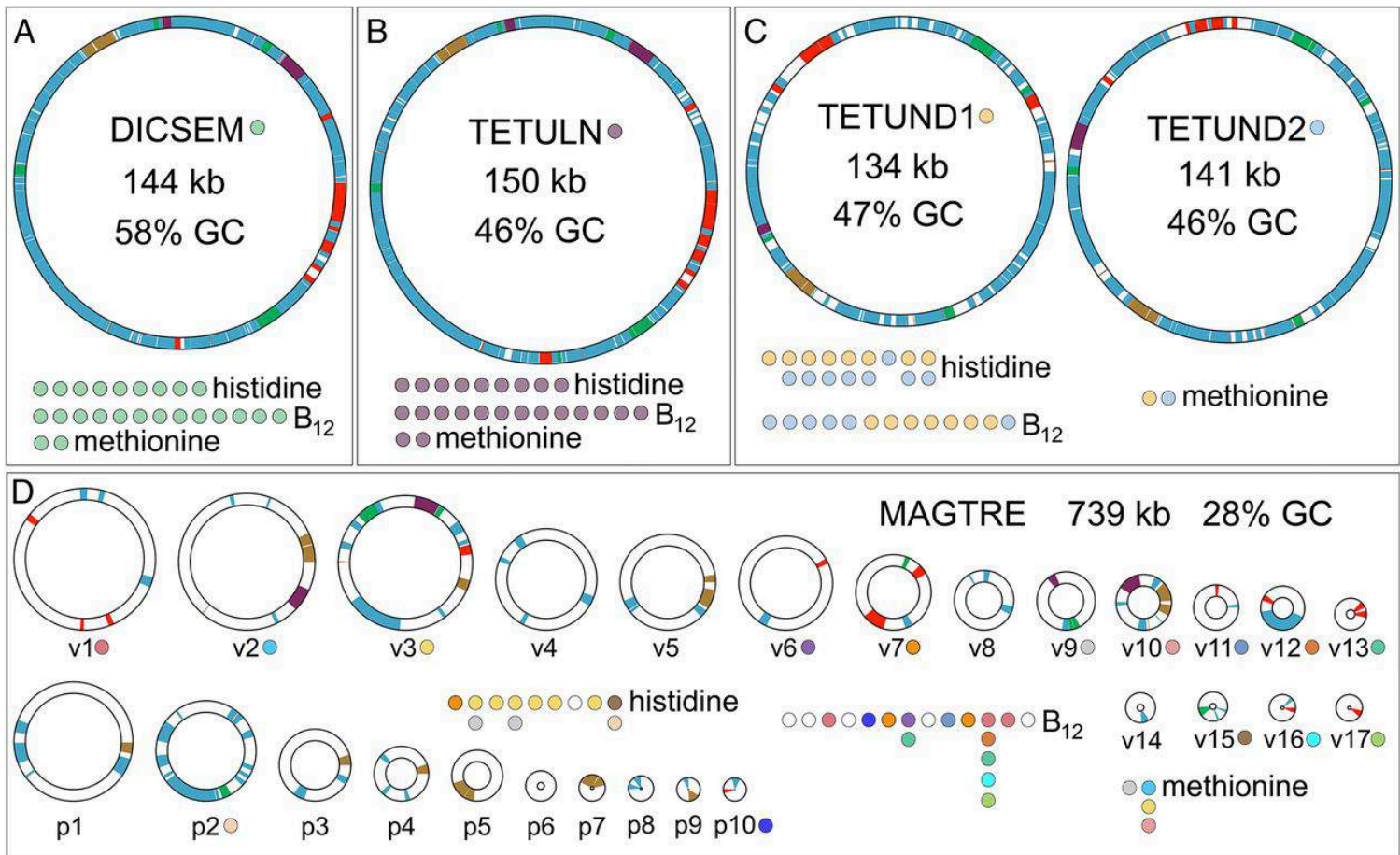
Komplementer genomredukció szimbiontákban



- A: bizonyos szimbiontákban (pl. *Sulcia*) megjelenhetnek inaktiváló mutációk (2) és a drift révén ezek viszonylag gyakoriakká válhatnak (4), de végül a szelekció kiszórja őket.
- B: *Hodgkinia* esetében komplementer inaktiváló-mutációk jelenhetnek meg, ezek elterjednek, és ha véletlenül az ősi vad-genotípus elvesződik, ezek egymásra lesznek teljesen utalva



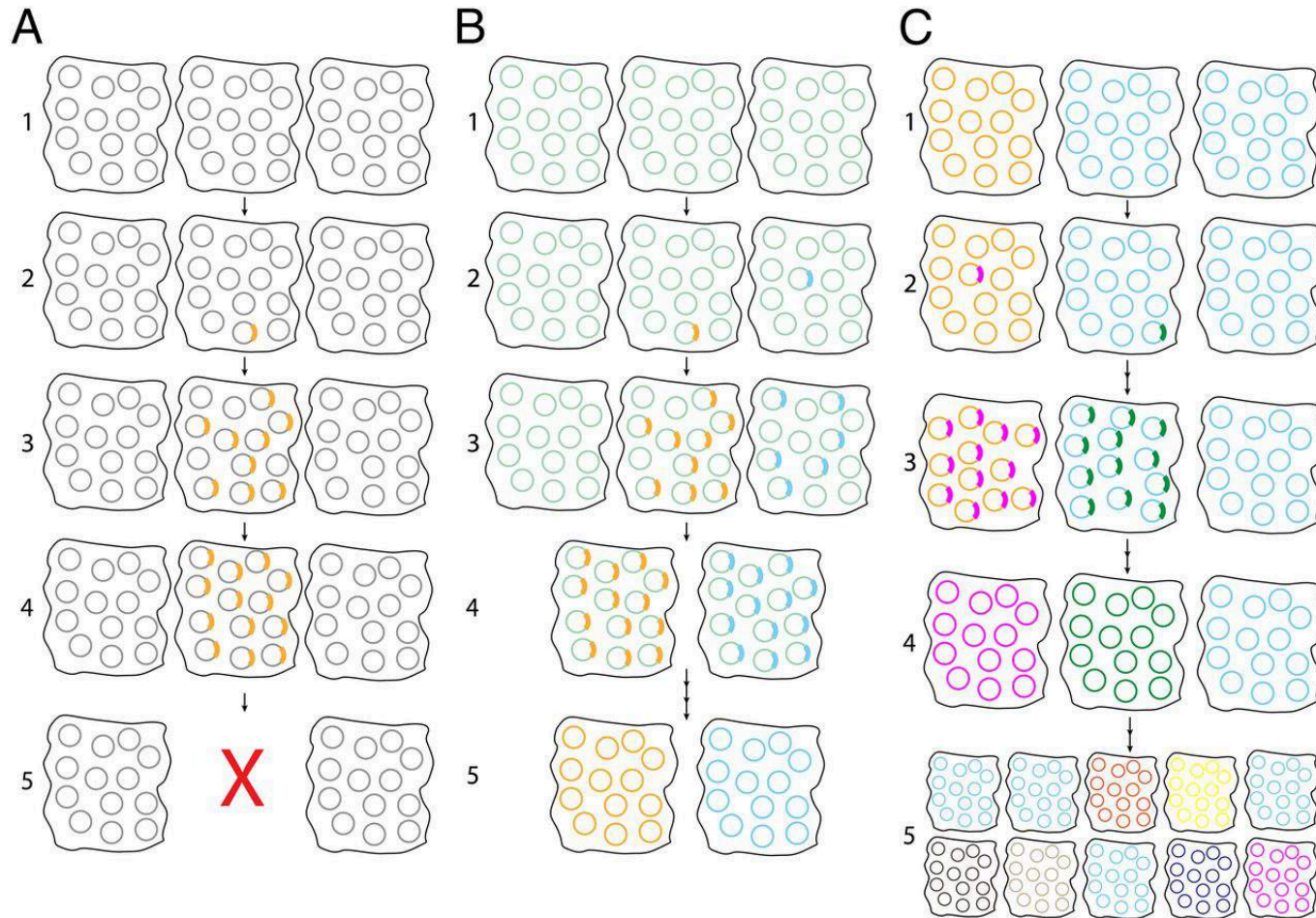
Komplementer genomredukció szimbiontáknban



- A *Magicicada tredecim* (MAGTRE) esetében több mint tucatnyi (min. 17) körkörös, komplementer *Hodgkinia* "scaffold" szerelhető össze a szekvenálási adatokból. Ezek közül több egész biztosan különböző *Hodgkinia* sejtekben van



Komplementer genomredukció szimbiontákban

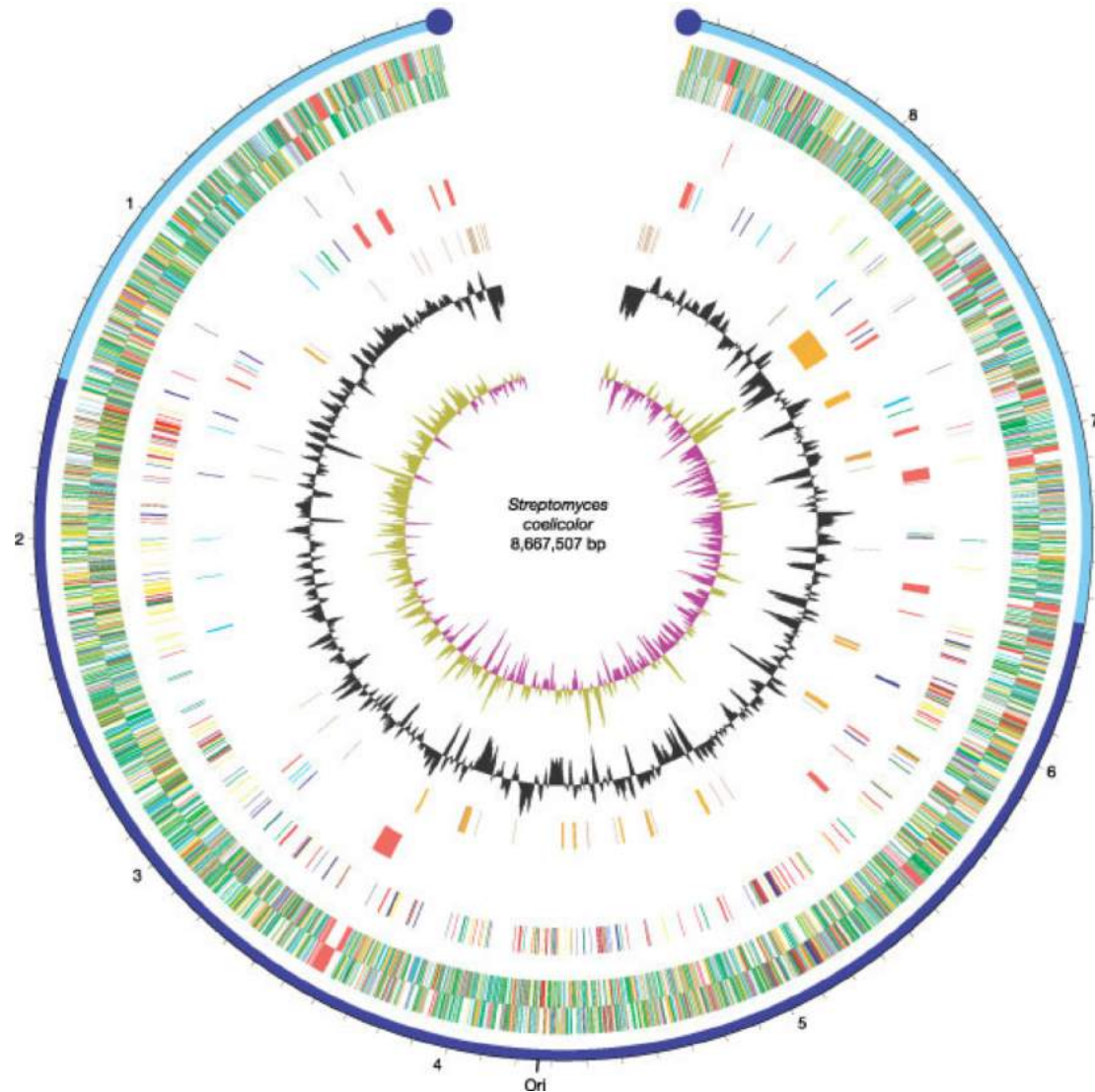


- A *Magicicada tredecim* esetében a megjelent új genotípusok további genotípusokká válhatnak szét. A sokszintű függőségi kapcsolatok miatt (a *Hodgkiniák*, meg a kabóca is függnek egymástól) könnyen lehet, hogy valamelyik *Hodgkinia* klón elvesződése a teljes “ökoszisztéma” pusztulásához vezet.



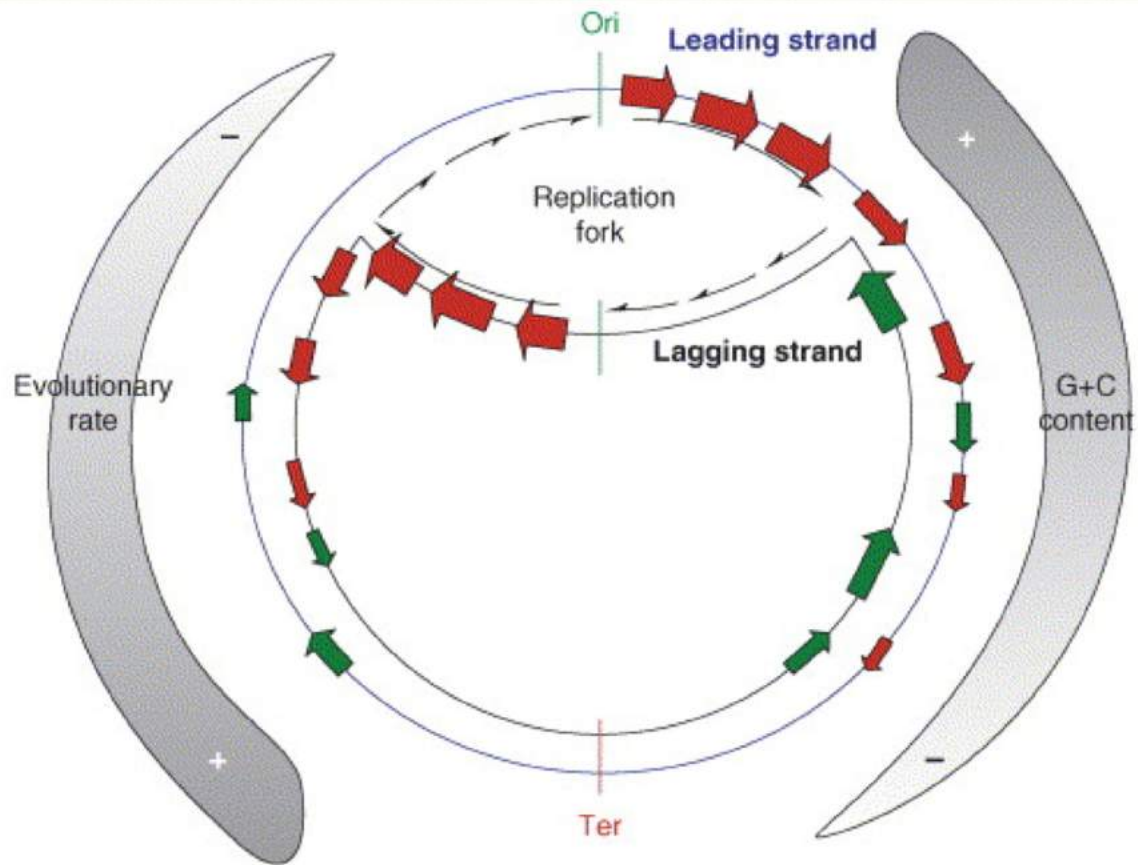
A változó környezet nagy genomot igényel

- a talajban lakó *Streptomyces coelicolor* genomja kb 8.7 Mb és ~7800 fehérjét kódol
- az esszenciális gének a lineáris kromoszóma közepén helyezkednek el
- a nem esszenciális, "készenléti" gének a karokon találhatóak



(Bentley et al. (2002) *Nature*)

A bakteriális genom evolúcióját befolyásoló belső tényezők

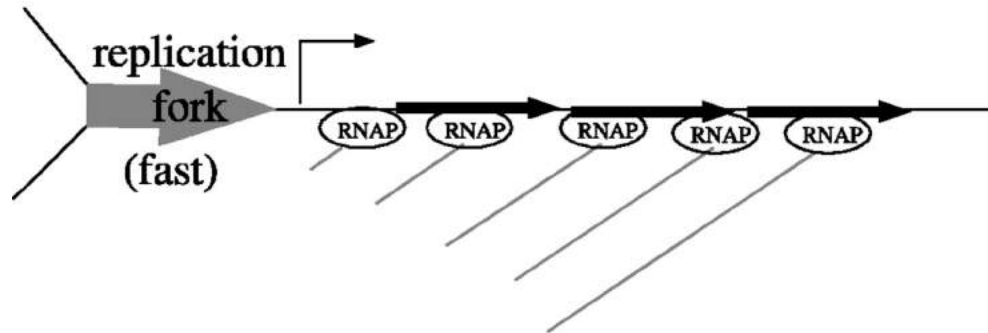


TRENDS in Microbiology

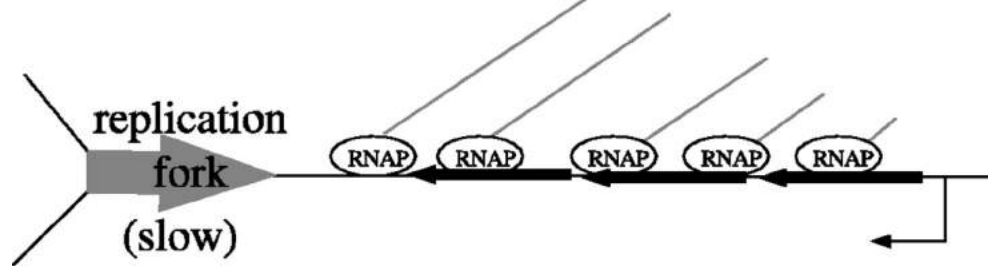


Replikáció vs. transzkripció: a szelekciós nyomás eredete

A. Operon on leading strand



B. Operon on lagging strand

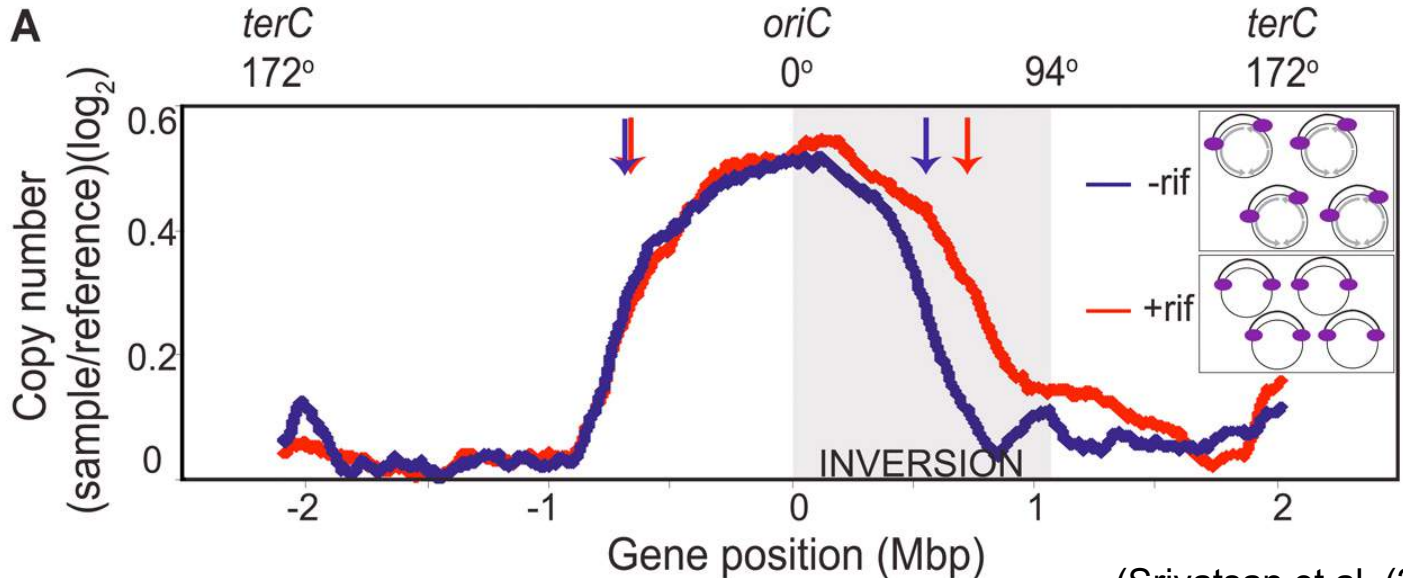
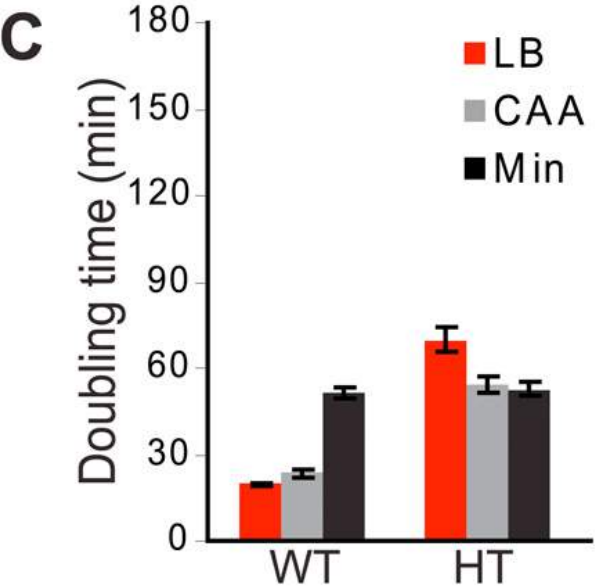
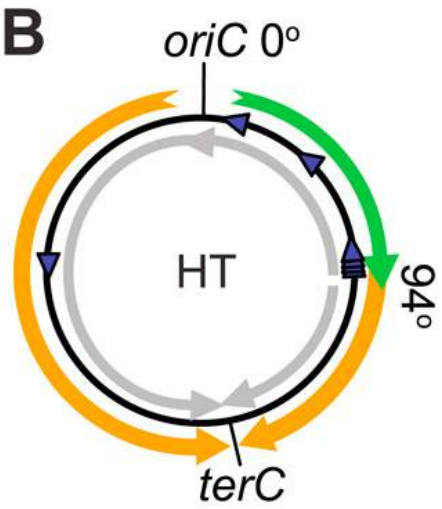
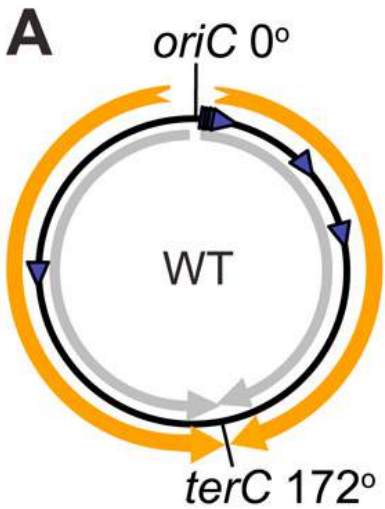


-a “leading strand”-en végighaladó replikációs villa megszakítja az operonban a transzkripciót, de az azonnal újraindulhat, amint a villa áthaladt a transzkripció starton

- a “lagging strand” esetében, az ellentétes irányultság miatt mindez sokkal tovább tart, aminek magas expressziót igénylő gének esetében jelentősége lehet



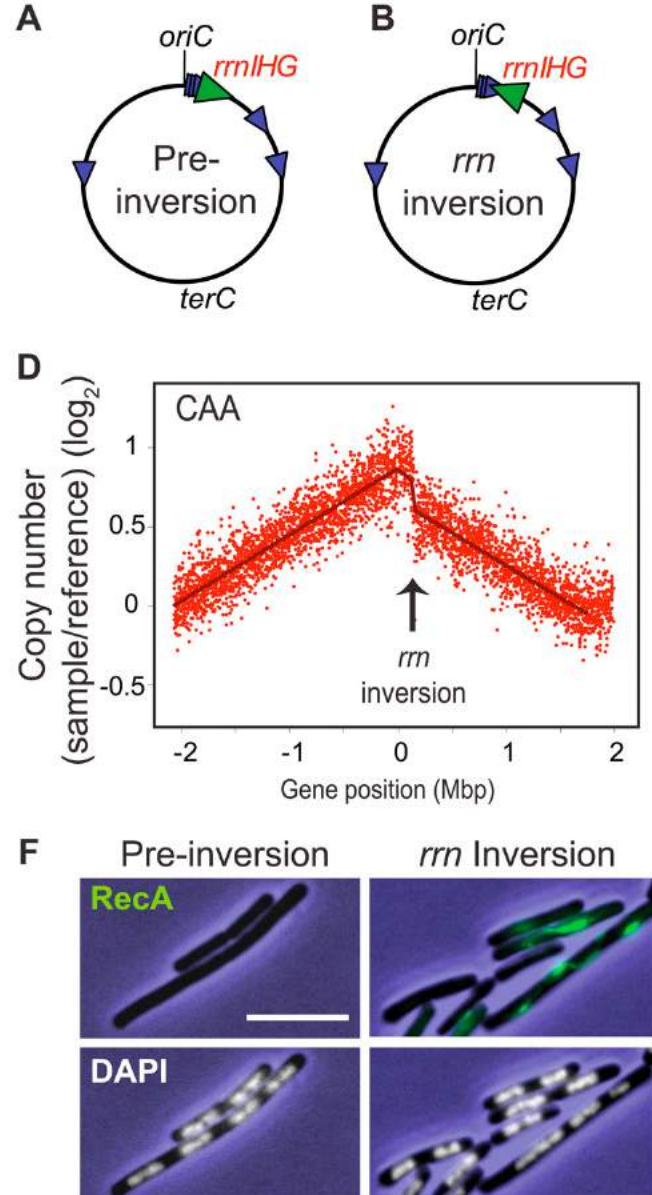
Replikáció és transzkripció ütközése lassabb növekedéshez vezet



(Srivatsan et al. (2010) PLoS Genet)



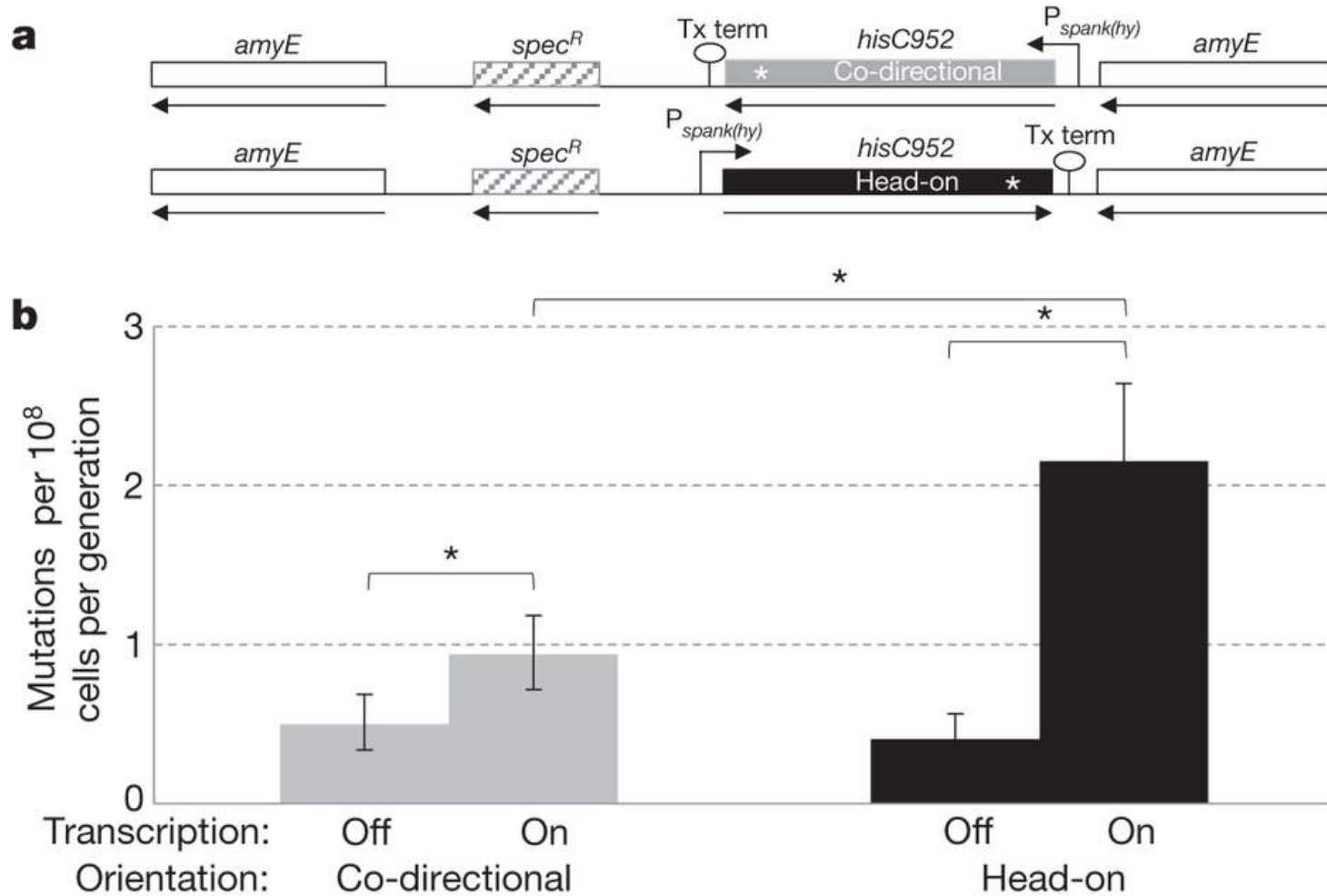
A magas expressziójú gének fordított orientációja a DNS-repair bekapcsolásához vezet



(Srivatsan et al. (2010) *PLoS Genet*)

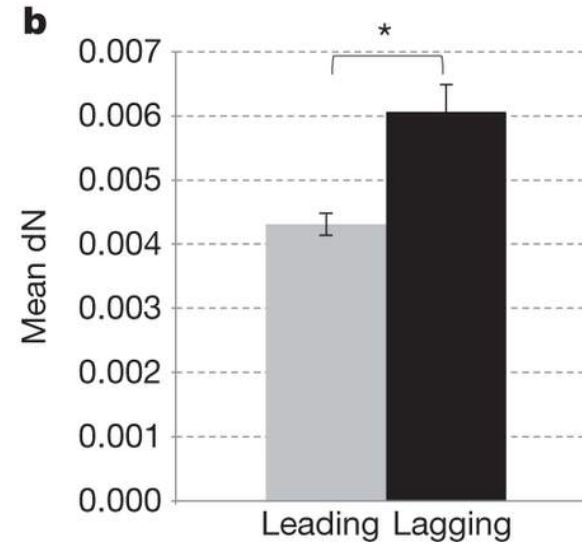
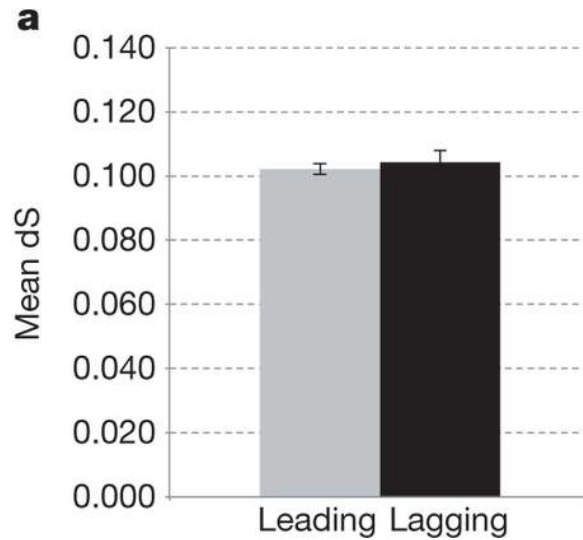


A fordított orientáció magasabb mutációs rátához vezethet

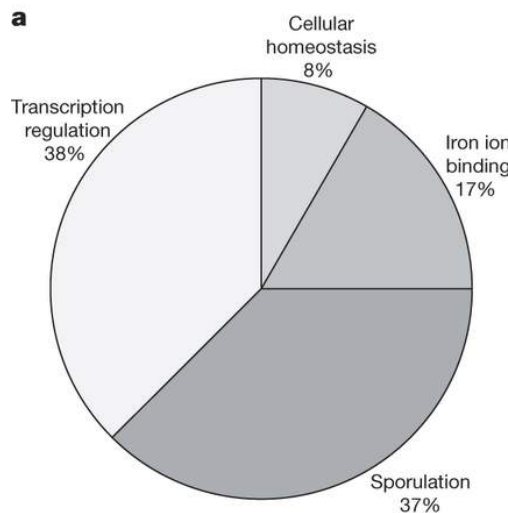




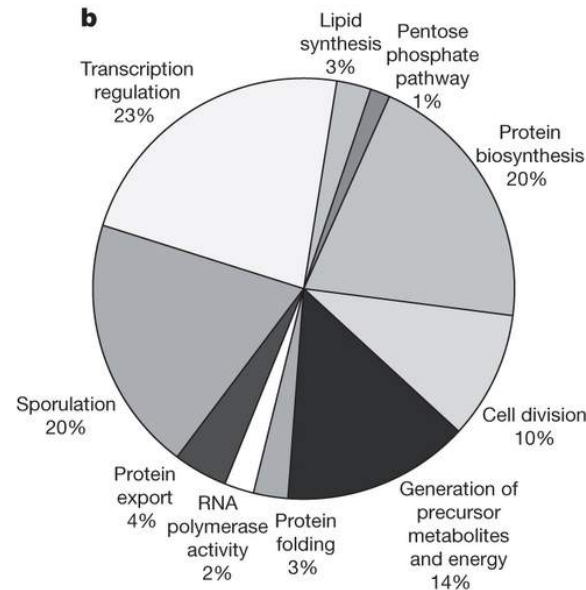
A lagging strand magasabb mutációs rátája előny is lehet gyors változást preferáló gének számára



lagging strand



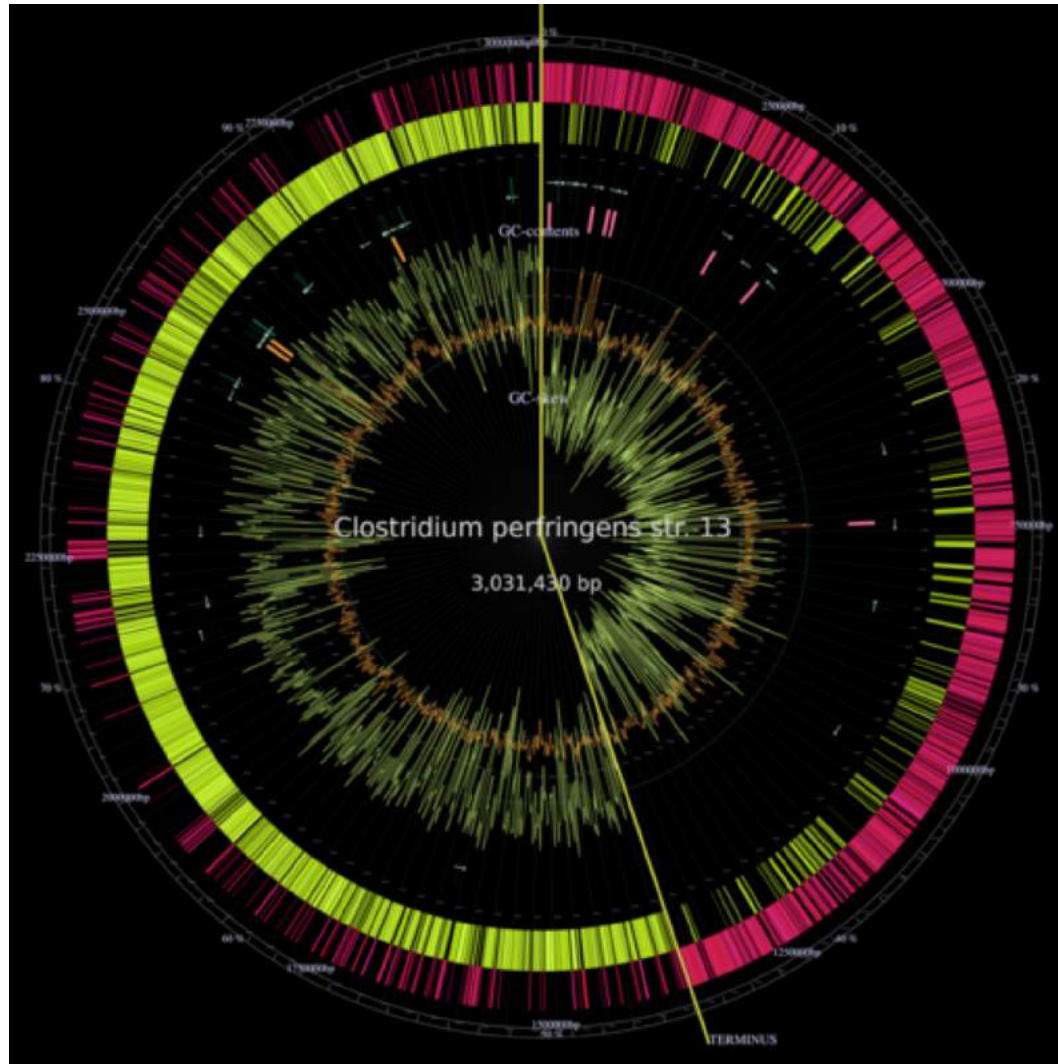
leading strand



(Paul et al. (2013) *Nature*)



“GC-skew” a *Clostridium perfringens* példáján át



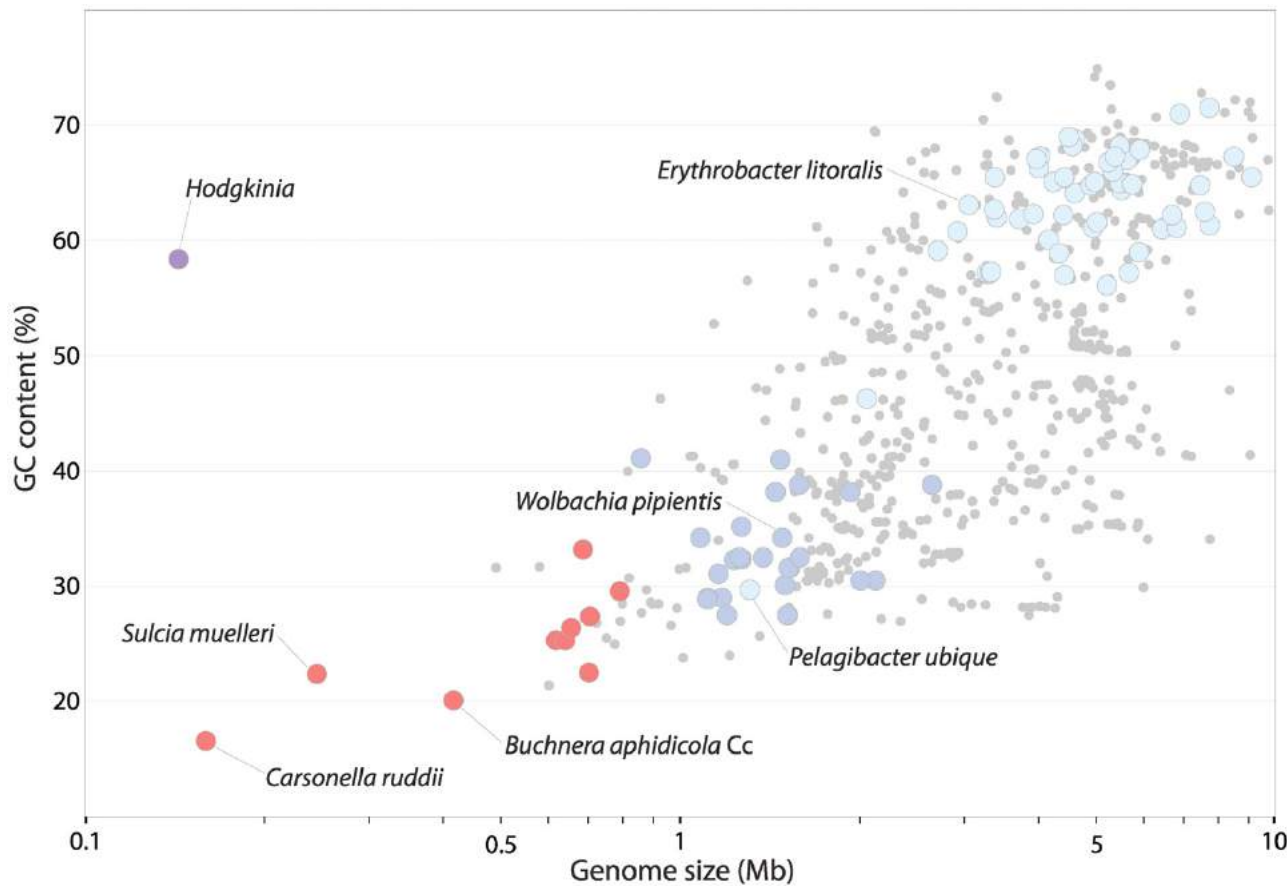
- “GC skew” a relatív C többletet írja le: $(C-G)/(C+G)$.

- mivel a bakteriális genomokban a replikáció jobban szereti a leading strandben a G-t, ezért a replikáció kezdete és vége könnyen megállapítható a GC skew érték alapján

- a *Clostridium* genom ennek az extrém példája



GC tartalom vs. genom méret



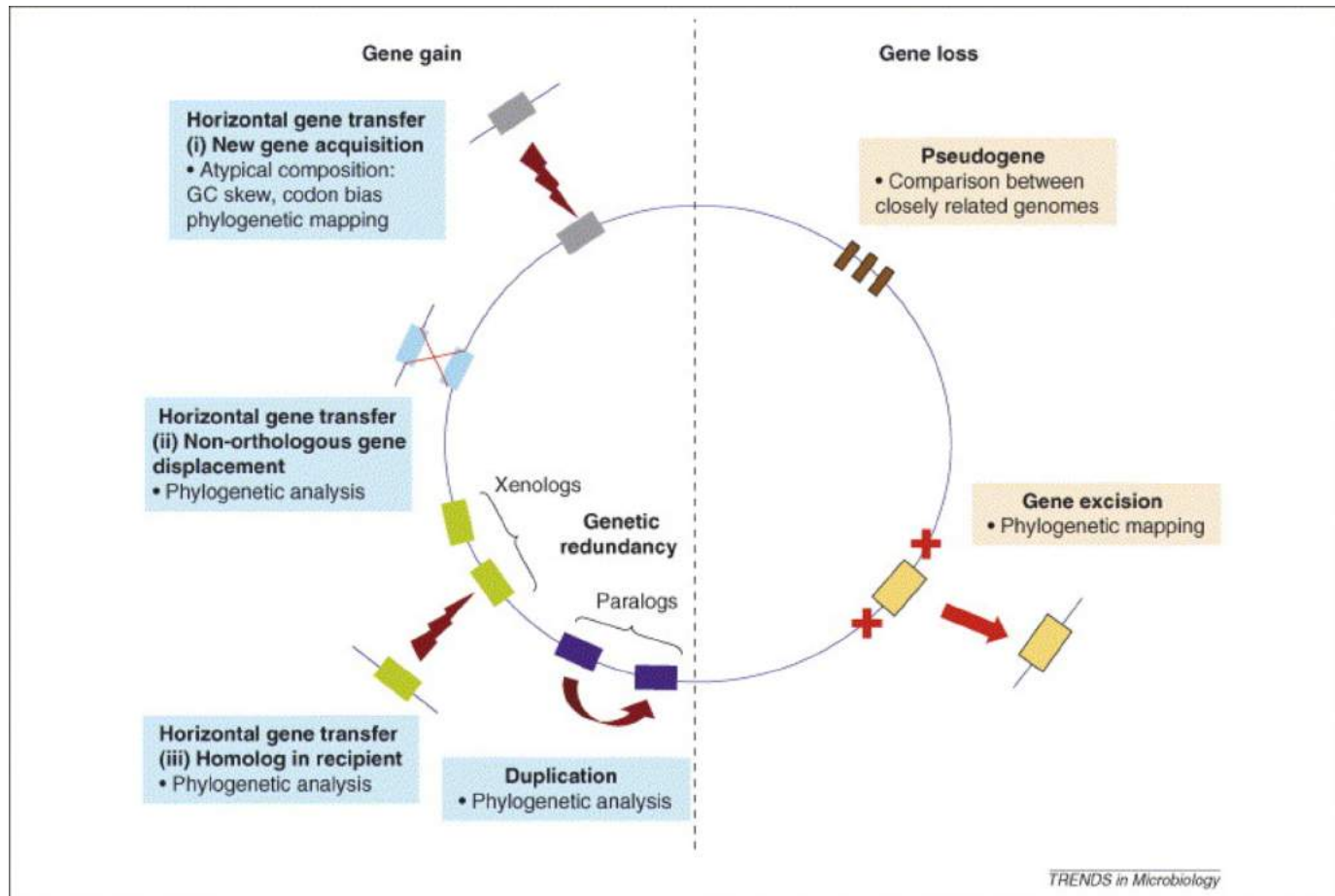
(McCutcheon et al. (2009) *PLoS Genet*)

Két lehetséges magyarázat:

- **energetikai okok:** a GTP és CTP szintézise több energiát fogyaszt, így a kis genomú (parazita) fajok ebben is optimalizálnak
- **mutációs okok:** a kis genomú prokarióták gyakran elvesztették a hibajavító enzimeiket, a leggyakoribb mutáció típus pedig C -> T tranzíció

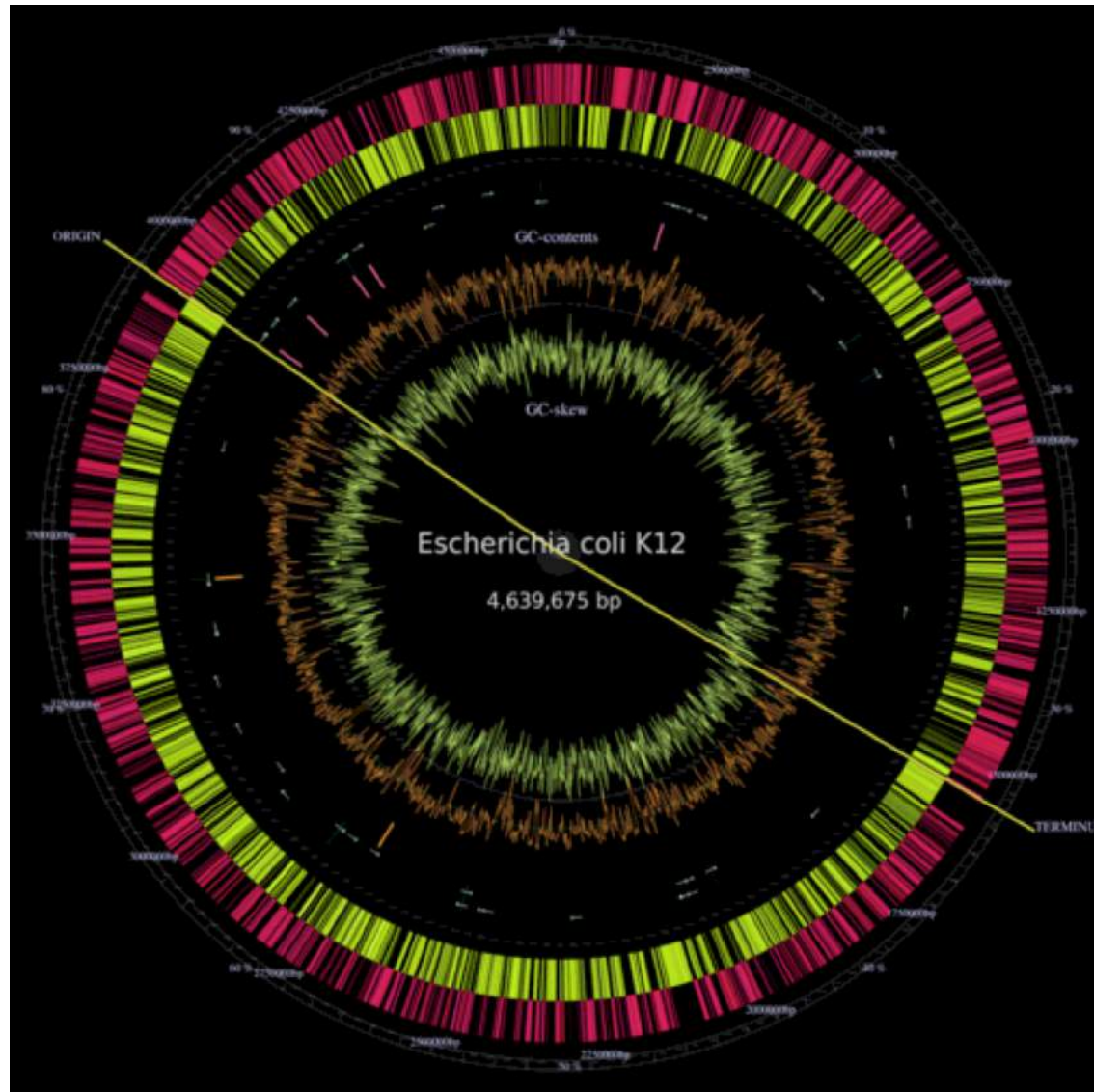


A bakteriális genom dinamikája





Escherichia coli K12

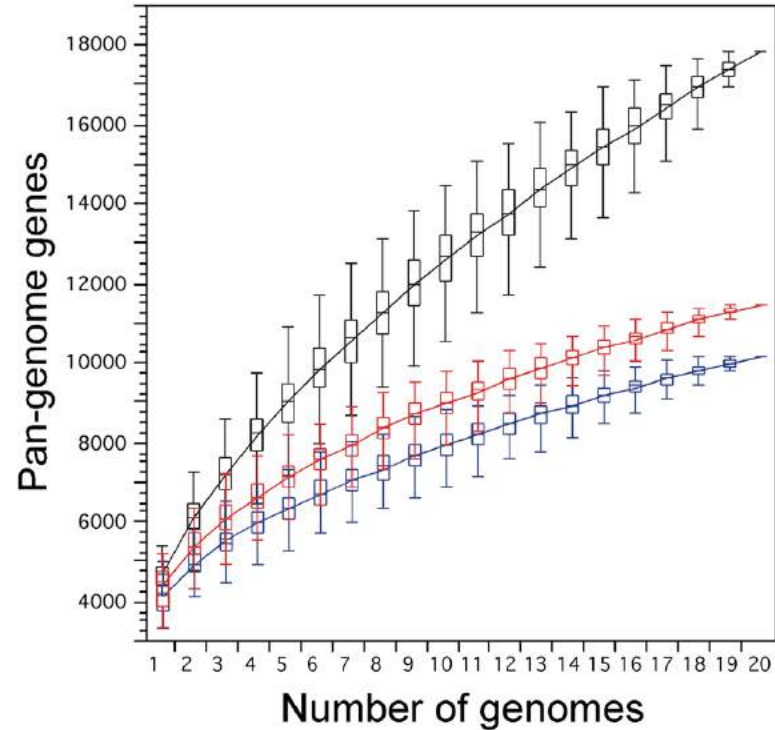
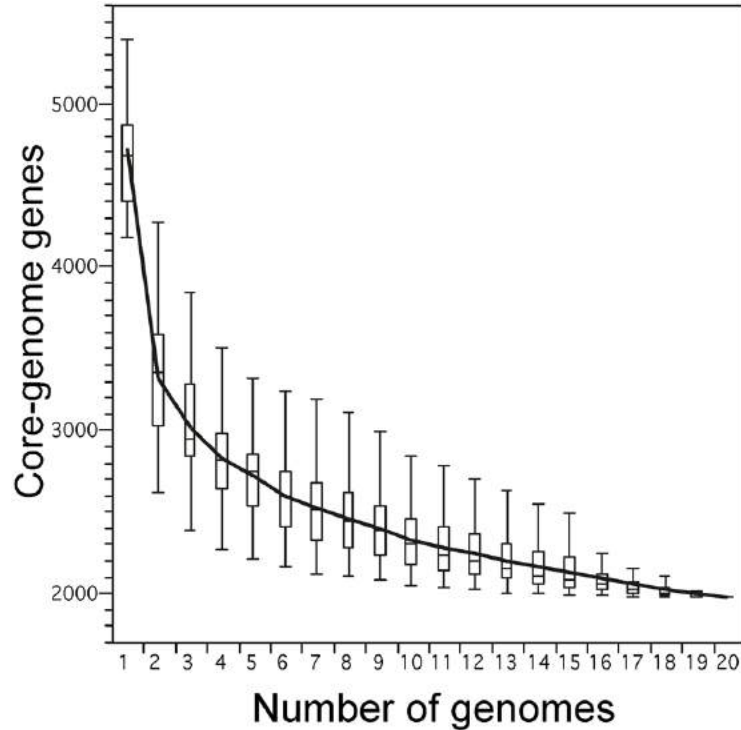


- 4.6 Mb

- 4288 fehérje kódoló gén (kb. harmadának ismeretlen a funkciója)

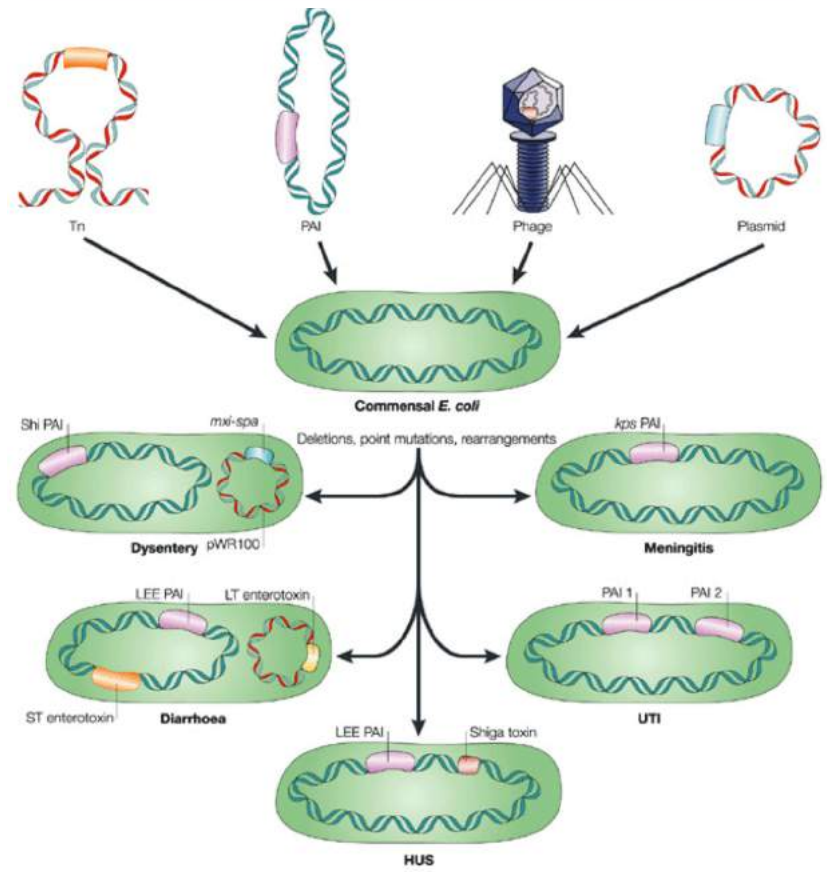
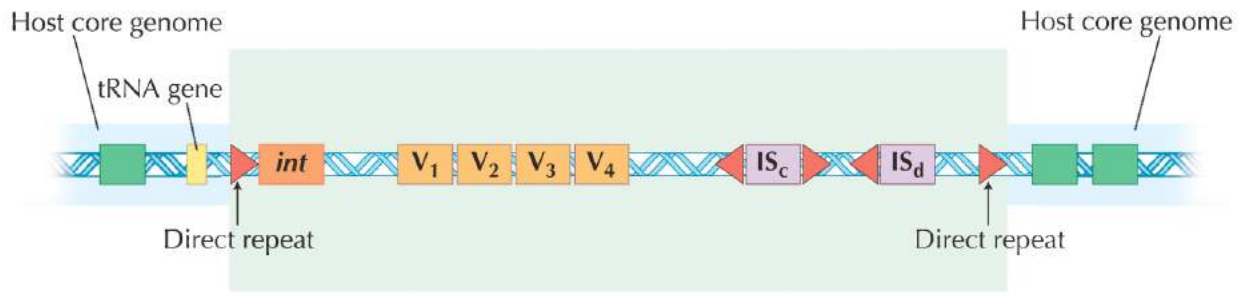


Escherichia coli - pan-genome vs core-genome



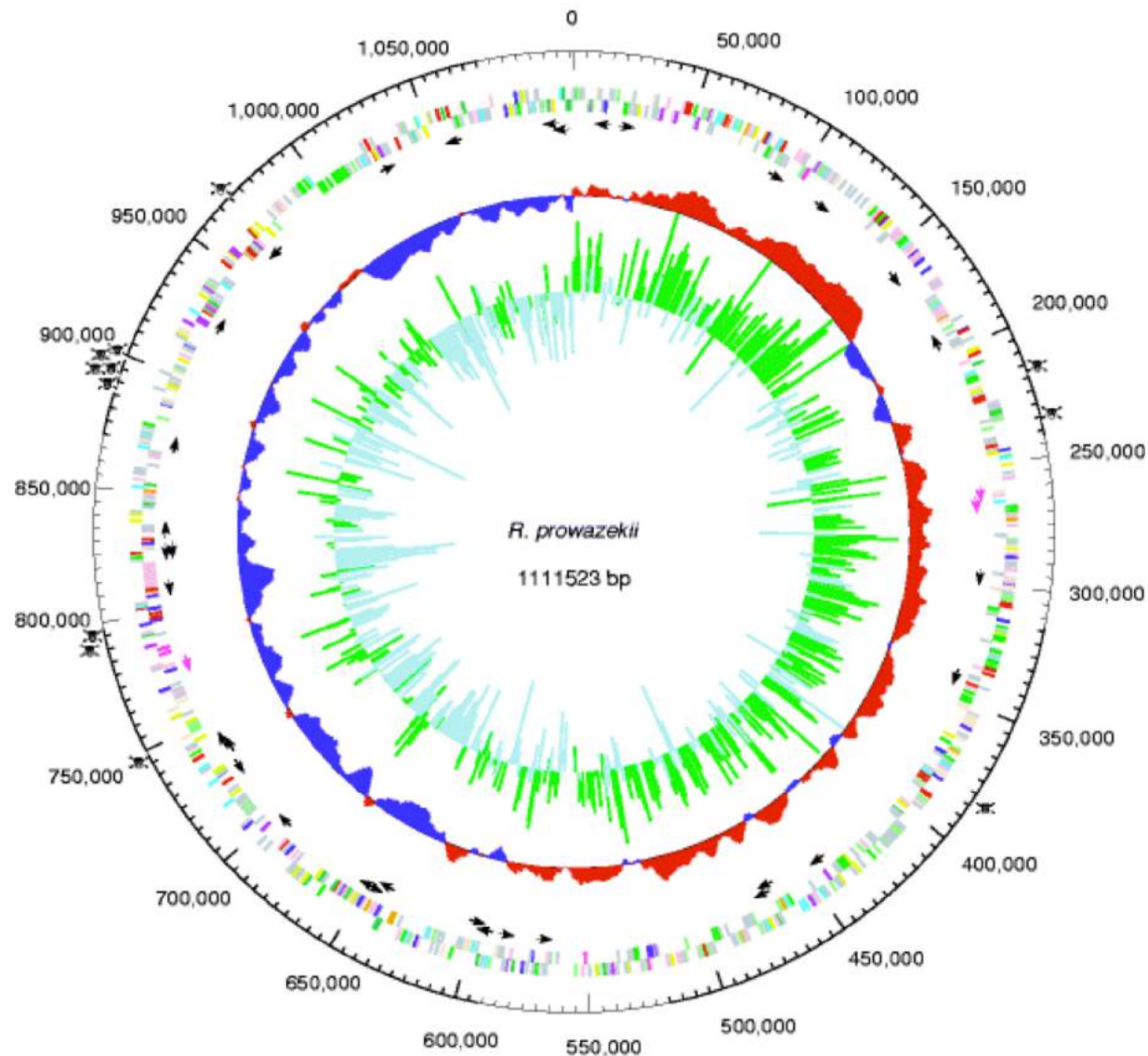
- több *E. coli* genom megszekvenálása után azonban nyilvánvaló, hogy az esszenciális gének száma csak töredéke a K12-ben találtaknak
- mai becslések szerint az *E. coli* alap (core) genom <1900 génből áll, az ún. pán-genomban található gének száma (az összes gén ami előfordul valamilyen *E.coli*-ban) pedig 17,000 felett van

Patogenitás szigetek





Rickettsia prowazekii



- obligát intracelluláris parazita, a tífusz okozója

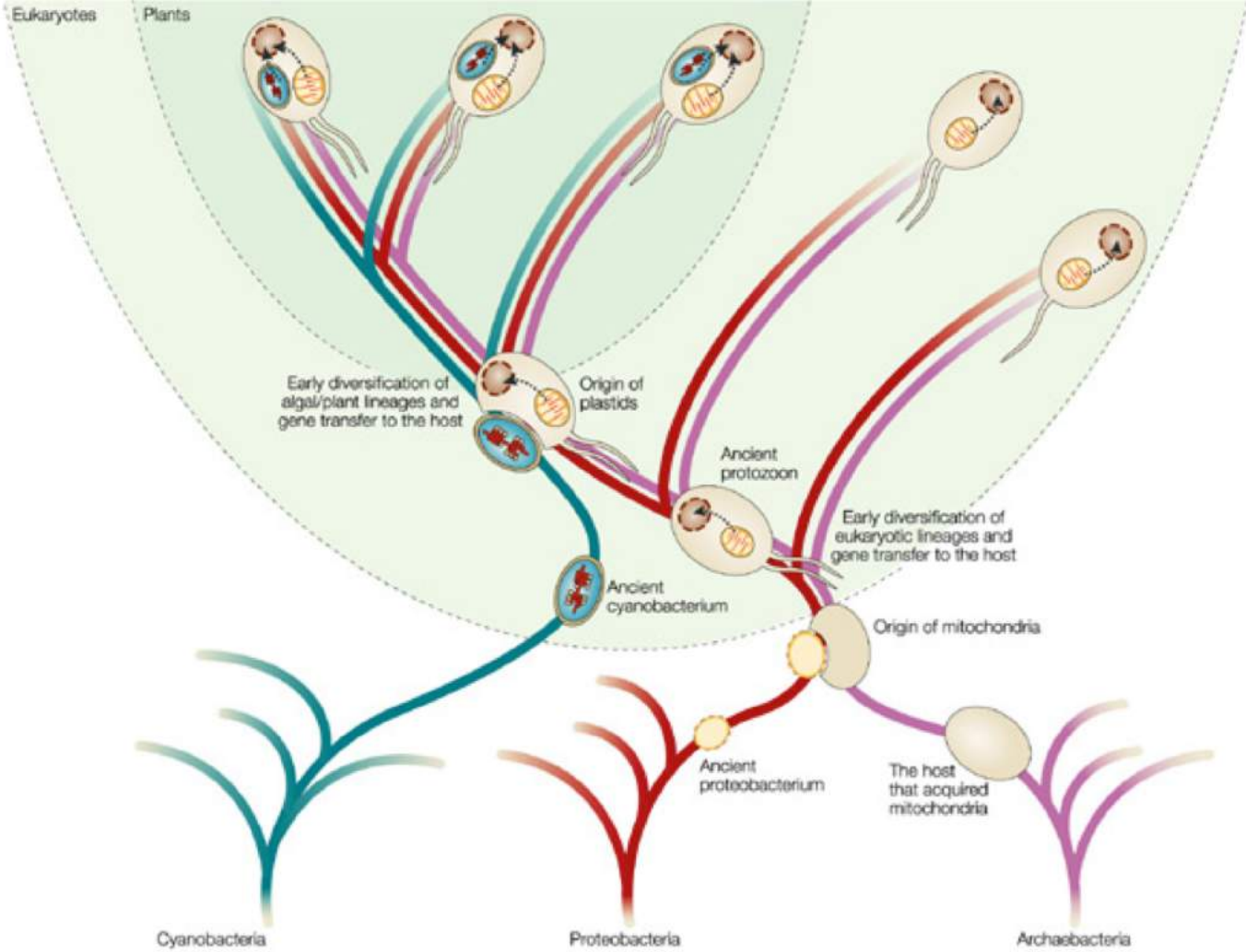
- a genom szokatlanul nagy része (24%) tartalmaz nem kódoló szekvenciákat

- ezek minden jel szerint pszeudogének, amelyek fokozatosan egyre több mutációt gyűjtenek össze

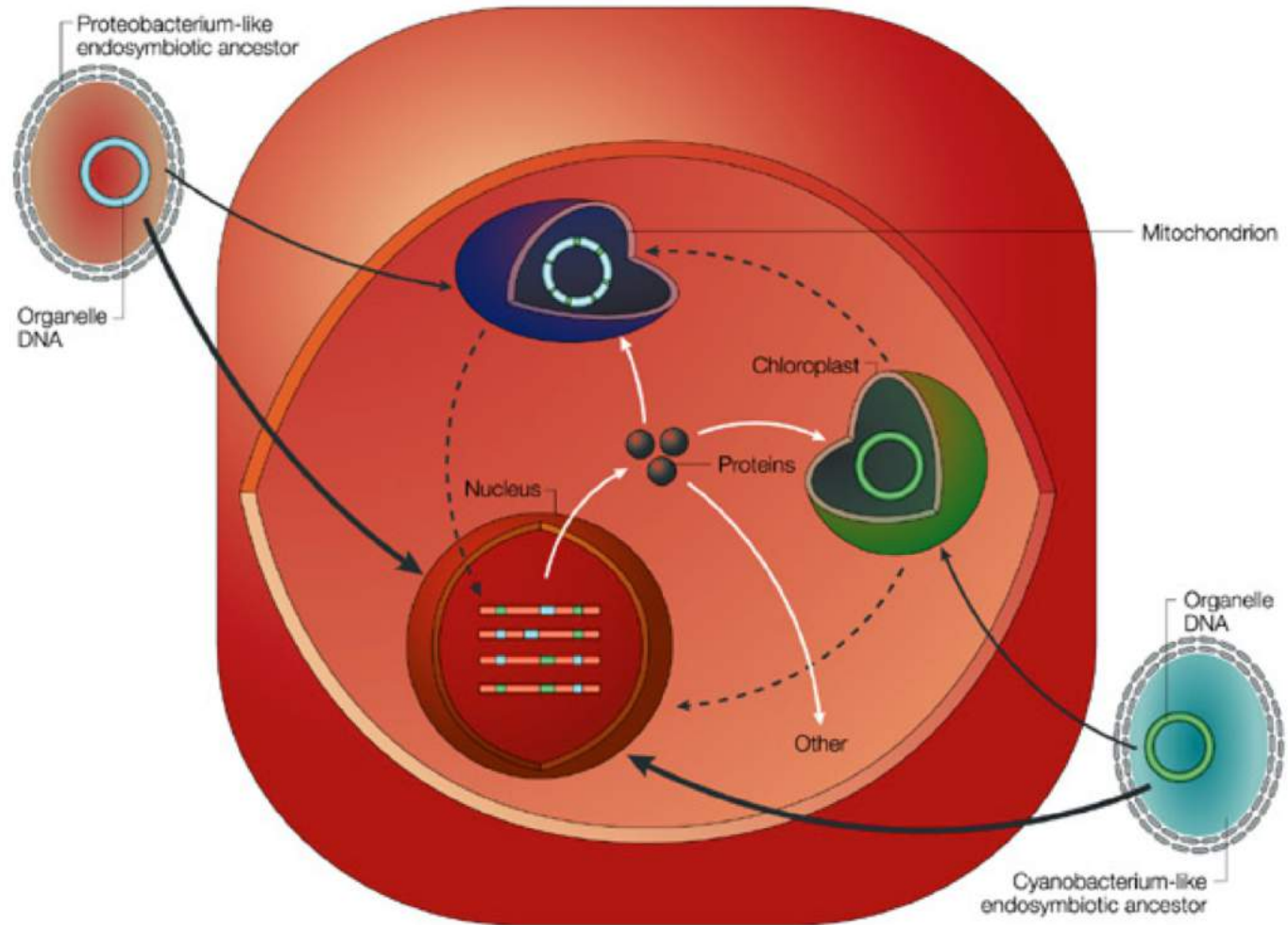
- a *Rickettsia* az α -proteobaktériumokhoz tartozik, akárcsak a mitokondriumok őse, így genomai degradációja informatív a mitokondriális genom evolúciójáról



Endoszimbiotikus gén-transzfer



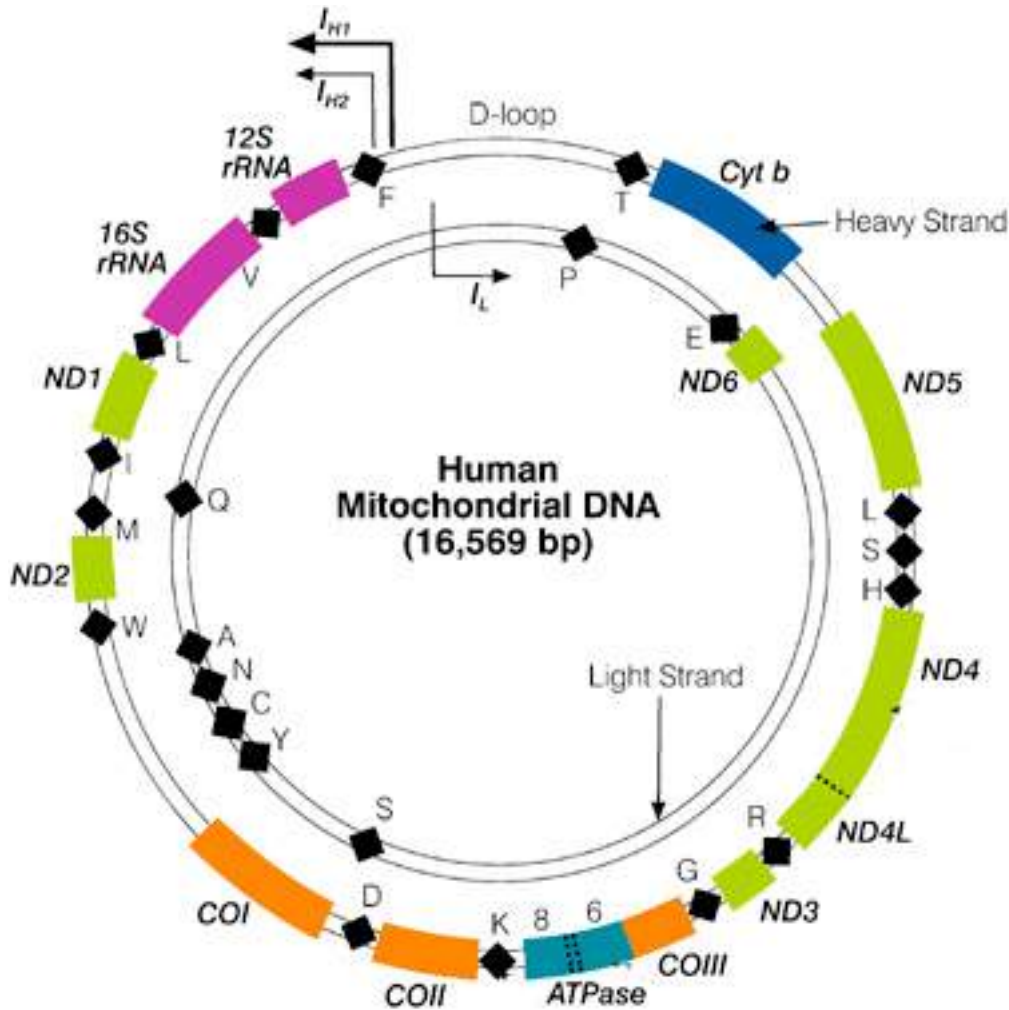
Endoszimbiotikus gén-transzfer



Nature Reviews | Genetics

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

A mitokondriális genom



- dupla szálú, cirkuláris DNS molekula, ami mindkét szálán kódol

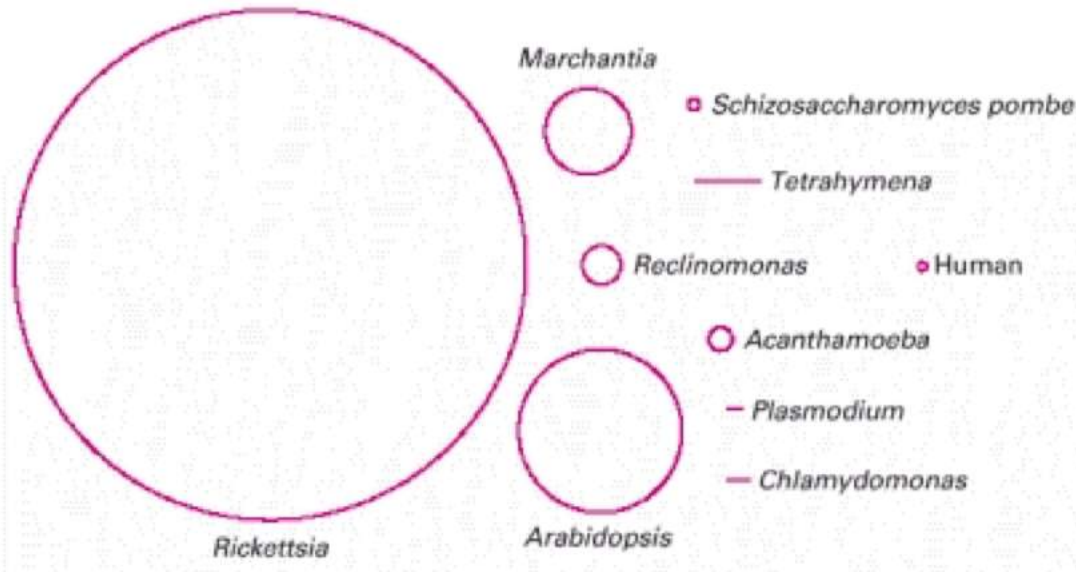
- emberek esetében 16,569 bp hosszú

- 37 gént kódol

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

- minden mitokondriális mátrixban több kópia található

A mitokondriális genom mérete különböző csoportokban



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

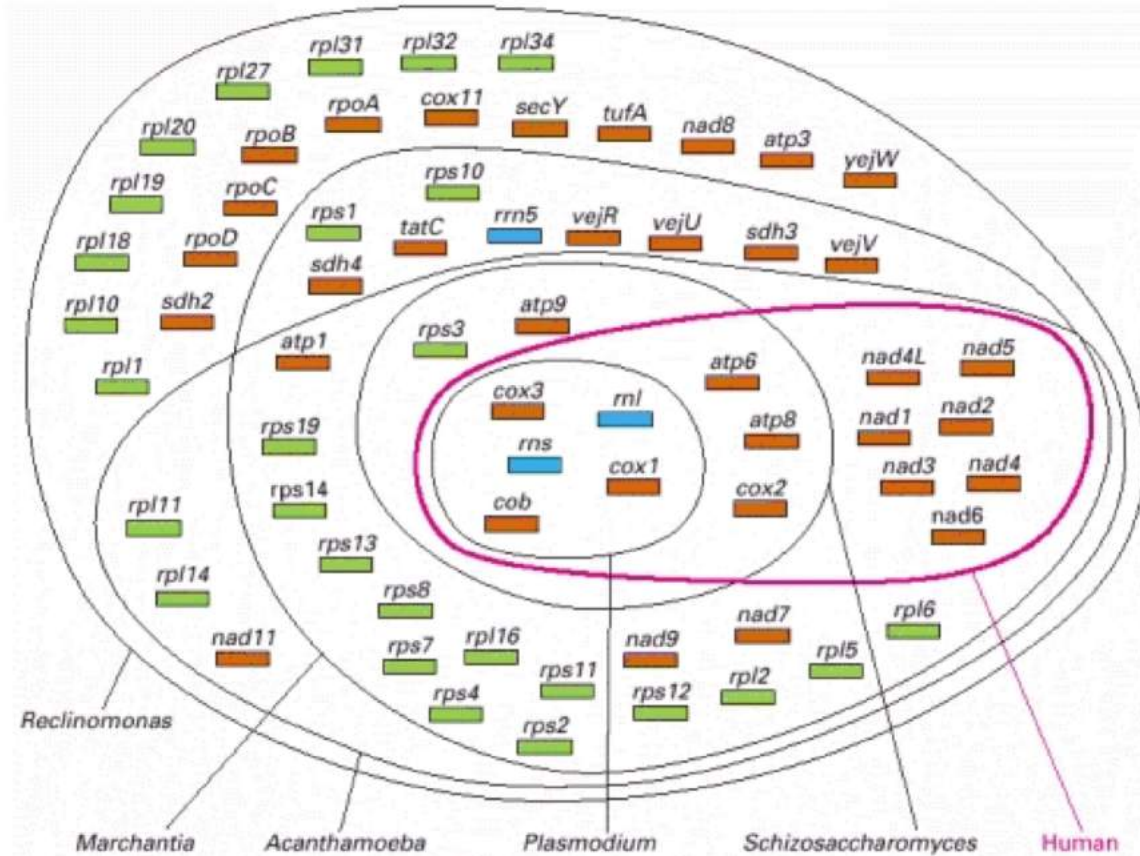
A mitokondriális genom tartalma különböző csoportokban



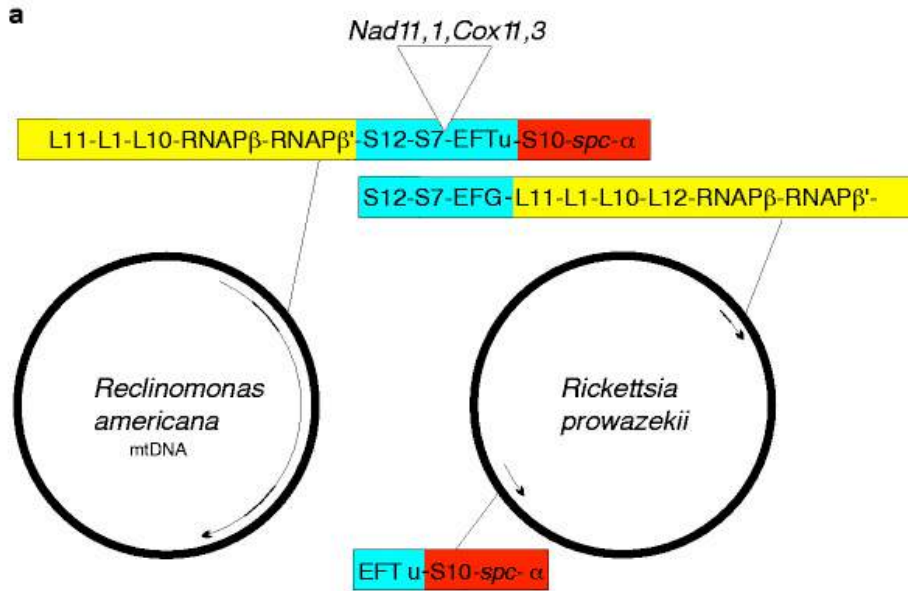
- a kevésbé komplex mitokondriális genomok, általában az összetettebb genomokban levő gének részalmazait kódolják

- öt olyan gén van, ami minden mtDNS-ben fellelhető: *cob*, *cox1*, *cox3*, *rns*, *rnl*

- hova lett a többi gén?

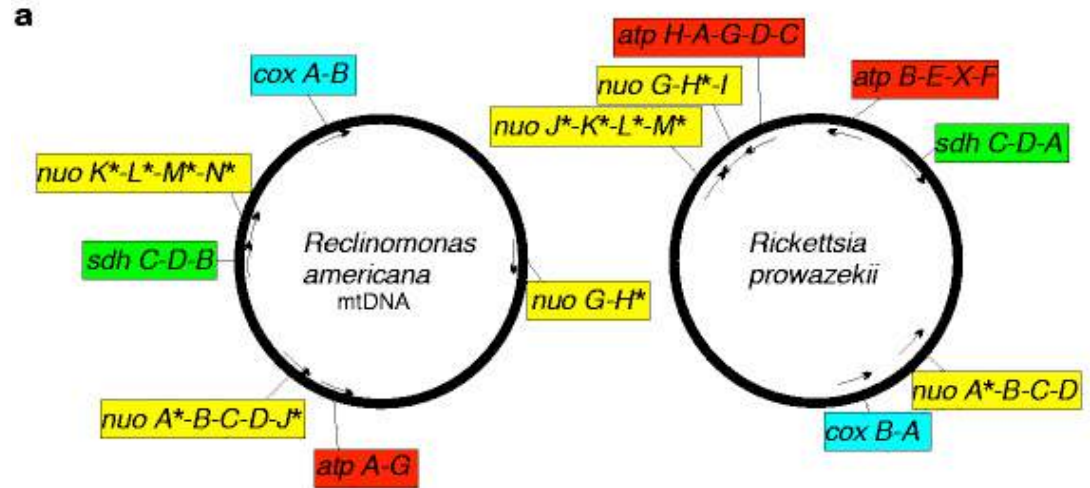


A mitokondriális genom és a *Rickettsia* genom szerkezete



- riboszomális fehérjék

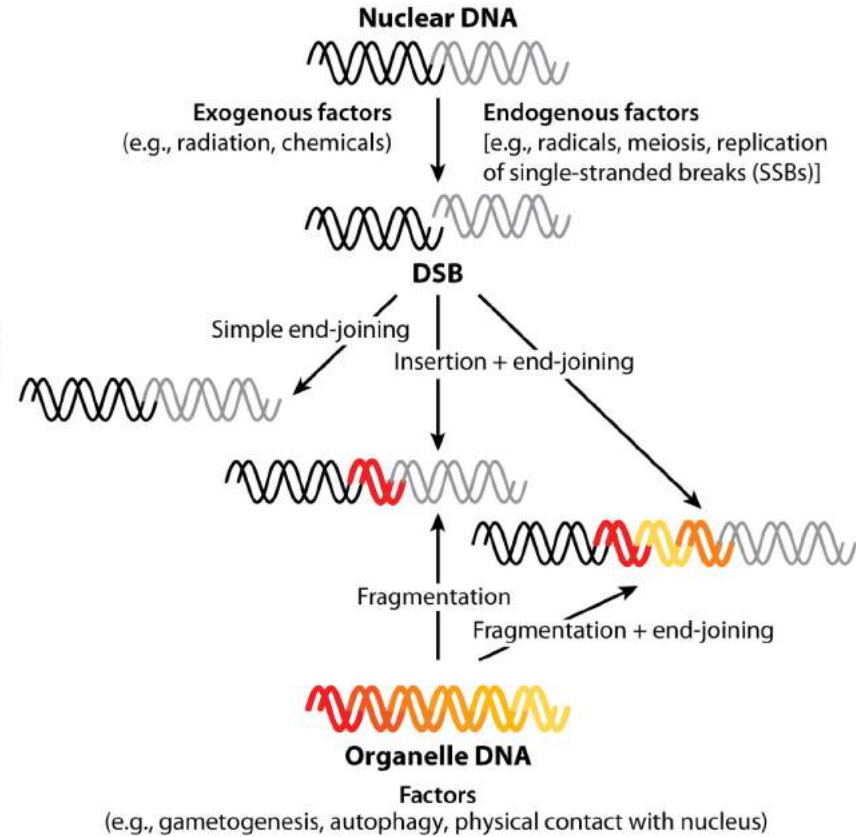
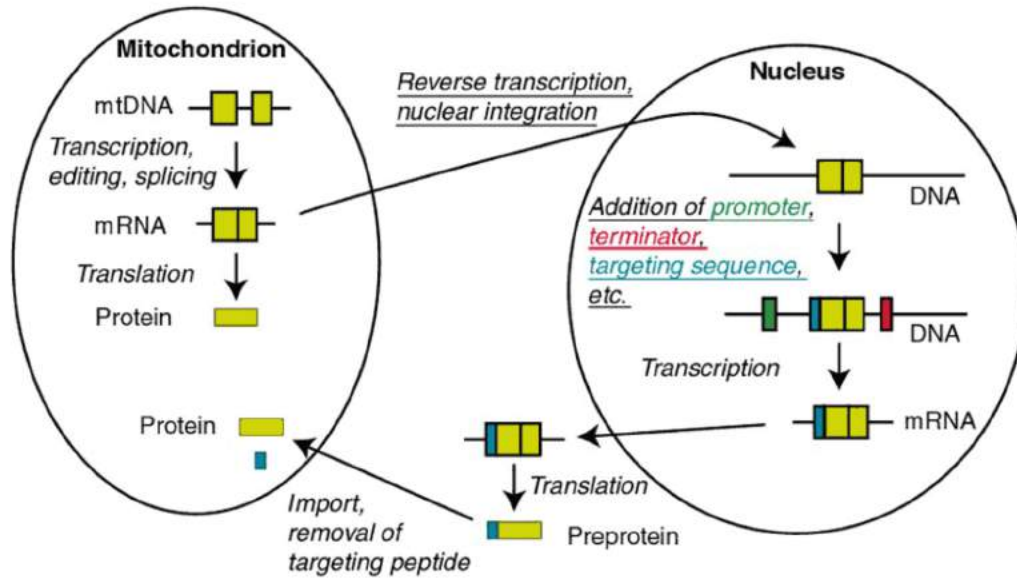
- az ATP szintéziséért felelős gének (az ATP szintézis azonos módon történik a *Rickettsiában*, mint a mitokondriumban.)



(Andersson et al. (1998) *Nature*)

Complex I ■ Complex IV ■
 Complex II ■ Complex V ■

A gén-transzfer feltételezett mechanizmusai



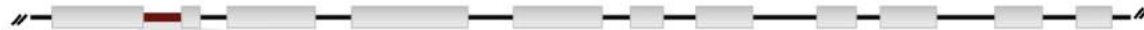
AR Kleine T, et al. 2009.
Annu. Rev. Plant Biol. 60:115–38



mtDNS, mint intron alapanyag

Bigelowiella natans

alpha subunit guanine nucleotide binding protein gene 1



Intron 1 = 74 bp

gDNA...	aatcg	G	TATC	CGG	ATTTTC	ATAGTCAACAG	CACAAAC	CCACCATCA	T	TAGAACGATGAT	TATG	GTTCTACTGATCACAG	ggaaa...	
cDNA...	aatcg												ggaaa...	
mtDNA...	TAGGA	T	TATC	TGG	TTTTCA	CAGTCAACAGT	TACAAAC	T	CACCATCA	A	TAGAACGATGA	CTAT	ATTCTACTGATCACAA	AGACA...

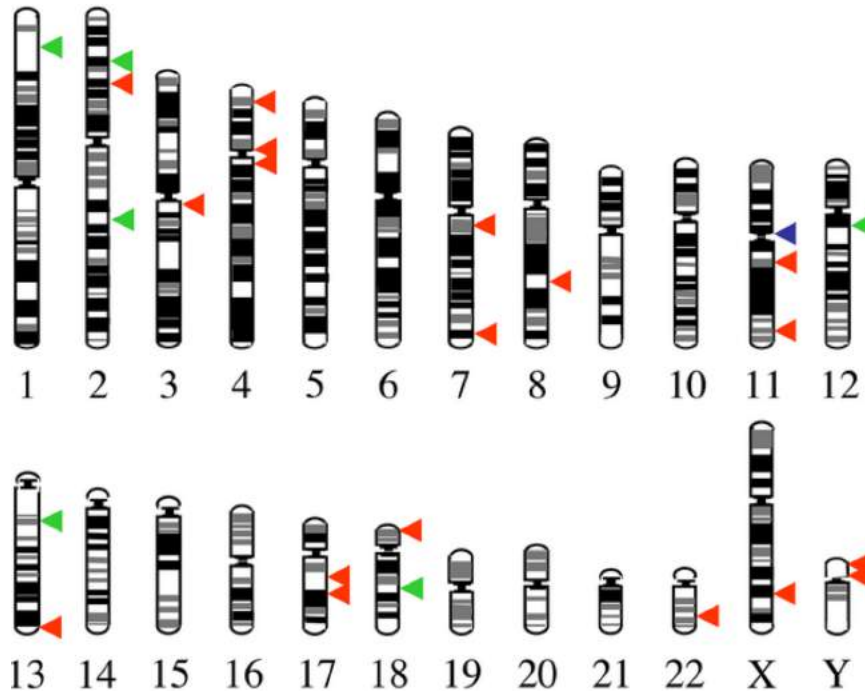
Current Biology

- egy egysejtű algafaj egyik fehérjekódoló génjének első intronja 86%-ban megegyezik a *cox1* mitokondriális gén egy szakaszával
- a splice akceptor és donor nukelotidok csak később alakultak ki, feltehetően kezdetben a szekvenca gyenge splice-site-ként működött

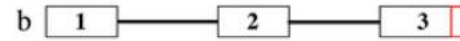
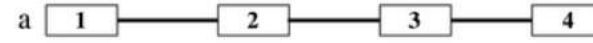
(Curtis and Archibald (2010) *Curr Biol*)



mtDNS, mint intron alapanyag



hypothetical protein



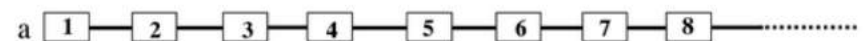
NUMT 12-89

hypothetical protein



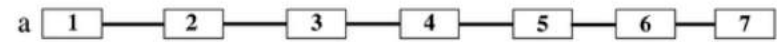
NUMT 17-653

hypothetical protein



NUMT 5-8781

protein Q8N7L5



NUMT 1-74

- csak az emberi genomban 27 darab specifikus NUMT található - ezek az utóbbi 4-6 millió évben keletkeztek és fixálódtak
- többségük feltételezett intronokba integrálódott

A mtDNS transláció kódja bizonyos csoportokban különbözik az “univerzális kódtól”



CODON	"UNIVERSAL" CODE	MITOCHONDRIAL CODES			
		MAMMALS	INVERTEBRATES	YEASTS	PLANTS
UGA	STOP	<i>Trp</i>	<i>Trp</i>	<i>Trp</i>	STOP
AUA	Ile	<i>Met</i>	<i>Met</i>	<i>Met</i>	Ile
CUA	Leu	Leu	Leu	<i>Thr</i>	Leu
AGA } AGG }	Arg	<i>STOP</i>	<i>Ser</i>	Arg	Arg

**Italics and color shading indicate that the code differs from the "Universal" code.*

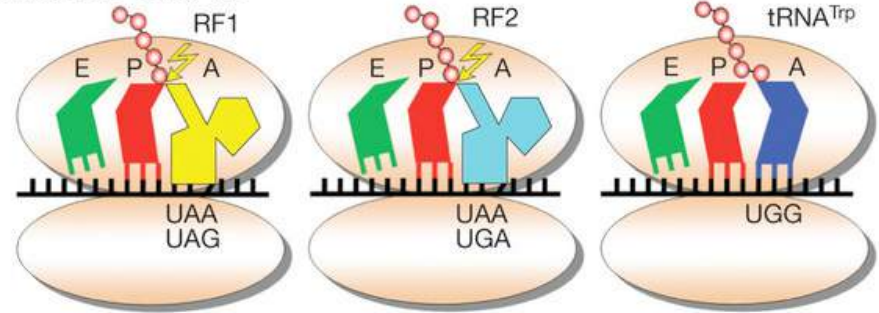
- növényekben és a legnagyobb mtDNS genomot hordozó protozoa *Reclinomonas* fajokban az mtDNS kódja is “univerzális”
- a STOP -> Trp változás hasonlít az egyes baktériumokban leírtakhoz
- valószínűleg az mtDNS-ben kódolt kisszámú gén jobban tolerálja egy-egy ritka kodon megváltozását

Alternatív genetikai kód kialakulása egy parazita baktériumban



	DnaE (335)	RpoB (711)	RpoC (131)
<i>Hodgkinia</i>	SDFTL.AKAHN	VAFMC.NGFNY	PVVHA.FHGSA
<i>Mloti</i>	ADFIKWAKAQQ	VAFMPWNGYNY	PVAHIWFLKSL
<i>Cces</i>	SDFIKWGKAHG	VAFMPWNGYNF	PVAHIWFLKSL
<i>Pdeni</i>	ADFIKWAKEHN	VAFMPWNGYNY	PVAHIWFLKSL
<i>Rrubr</i>	ADFIQWAKDAD	VAFMPWNGYNF	PVAHIWFMKSL
<i>Elito</i>	ADFIQWAKDHG	VAFMPWNGYNY	PVAHIWFLKSL
<i>Pubiq</i>	SDYIKWAKNND	VAFMPWQGYNF	PVAHIWFLKSL
<i>Rrick</i>	SDFIKWSKKEG	VAFLPWNGYNF	PVAHIWFLKSL
<i>Ecoli</i>	MEFIQWSKDNG	VAFMPWNGYNF	PTAHIWFLKSL
<i>Nmeni</i>	QDFINWAKTHG	IAFMPWNGYNY	PVAHIWFLKSL
<i>Gmeta</i>	ADFINWAKDHG	VAFMPWGGYNF	PVAHIWFLKSL

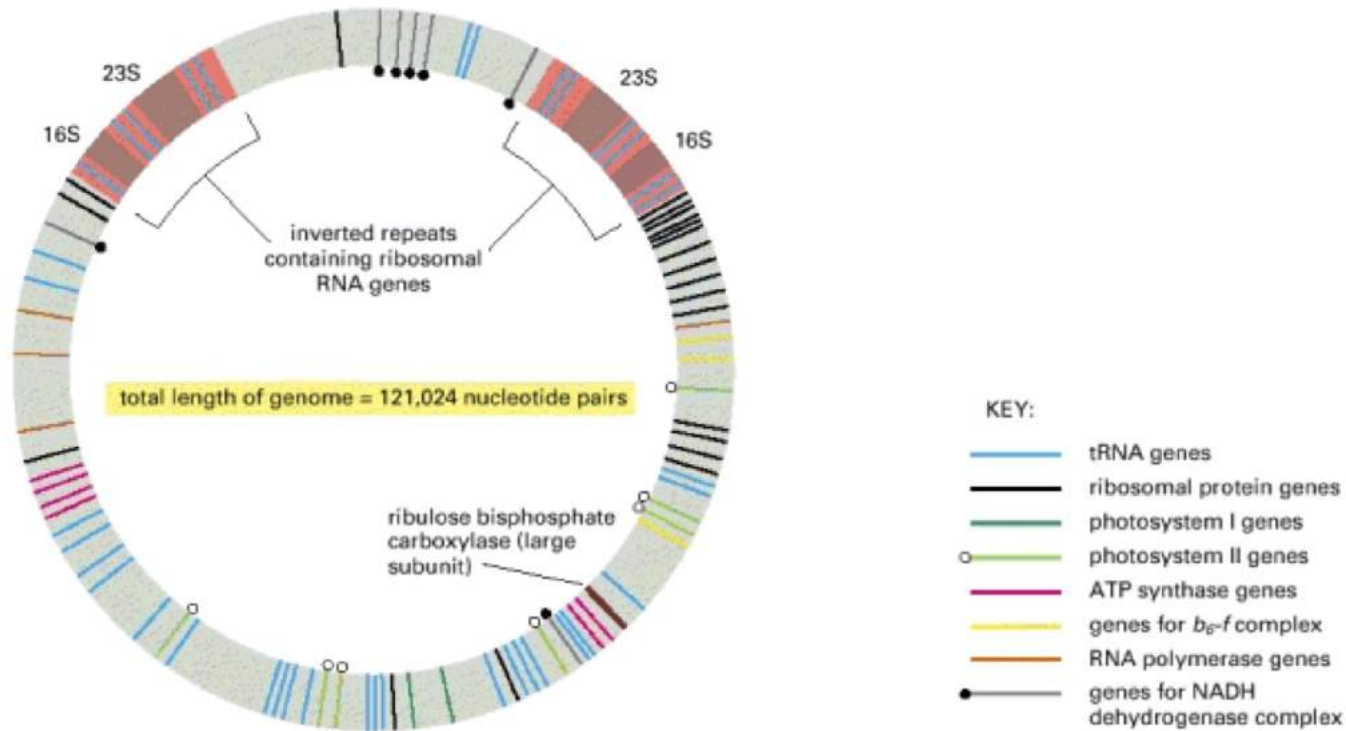
UAA, stop UGA, stop
UAG, stop UGG, Trp



	tRNA-Trp anticodon	release factors	UGA encodes
initial state	CCA	RF1 RF2	STOP
1 mutation of tRNA-Trp gene			
some readthrough of UGA	*CCA	RF1 RF2	STOP Trp
2 loss of Release Factor 2 (RF2)			
only UAA and UAG read as stop	*CCA	RF1	Trp
3 mutation of tRNA-Trp anticodon			
UGA, UGG both read by wobble rules	UCA	RF1	Trp
4 genomic codon adaptation			
new UAA and UAG stops generated; some UGG codons changed to UGA	UCA	RF1	Trp

(McCutcheon et al. (2009)
PLoS Genet)

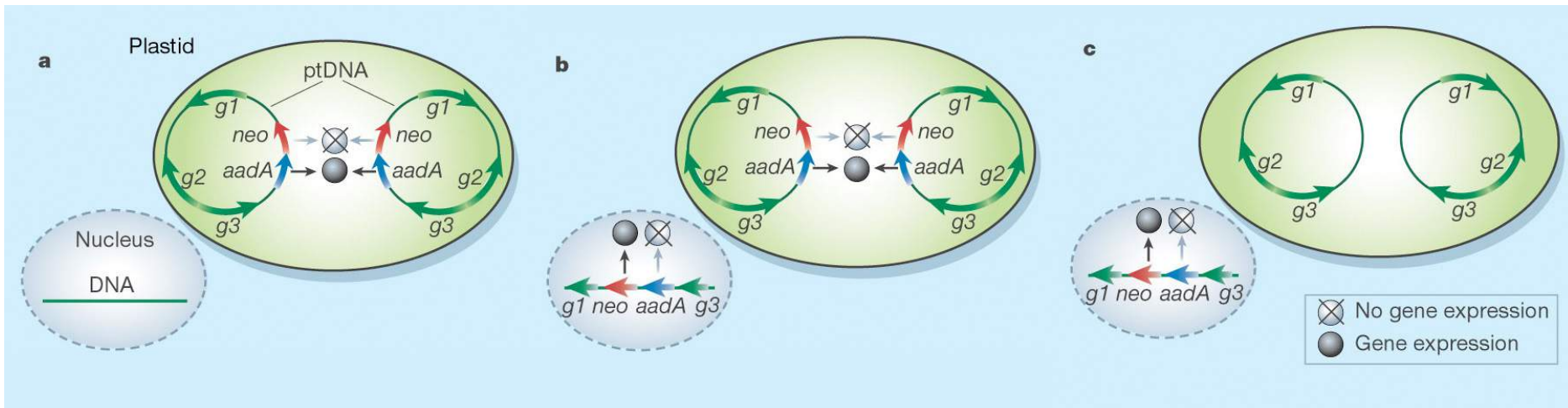
A kloroplasztisz genom



- dupla szálú, cirkuláris DNS molekula, ami mindkét szálán kódol
- a transzkripciót szabályozó elemek szinte azonosak bakteriális homológ szekvenciákkal



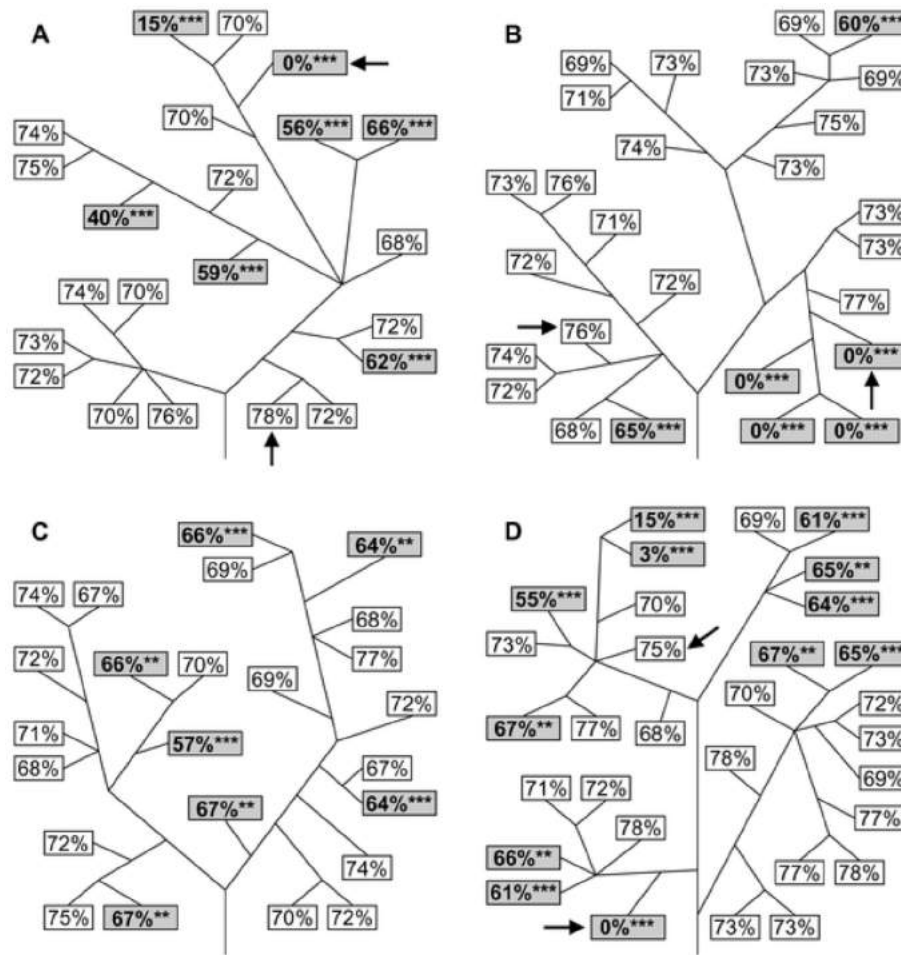
Plasztisz-sejtmag DNS transzfer



- genetikai screen plasztisz-sejtmag DNS transzfer gyakoriságának vizsgálatára:
- a spectinomycin rezisztenciát hordozó gén (*aadA*) egy bakteriális promóter mögött van (így az a plasztiszban is aktív), míg a neomycin rezisztencia gén egy eukarióta promóter mögött van, így az csak a sejtmagban aktív
 - szomatikus sejtekben (levélben) 1 : 5 millió volt az átugrás gyakorisága, míg pollen sejtekben 1 : 16,000 (a különbség egyik oka az lehet, hogy pollenképződés közben lebomlik a plasztisz, és a felszabaduló DNS nagyobb eséllyel juthat a sejtmagba)



A NUPT-ok többsége instabil

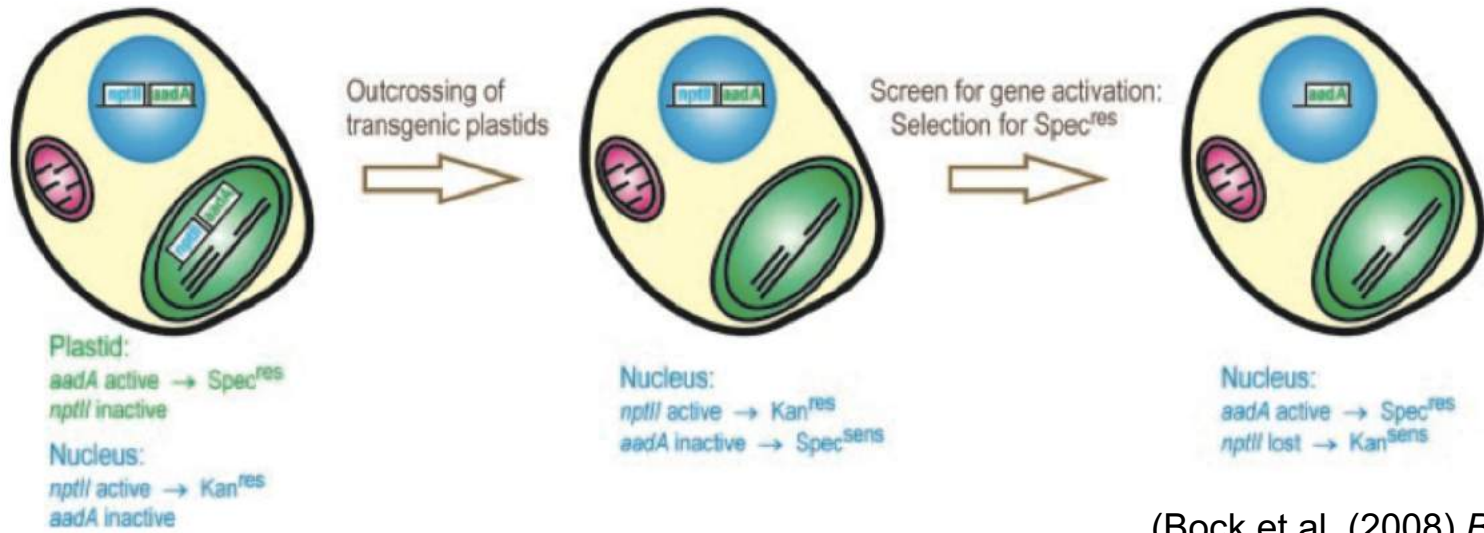


- önmegporzó rendszerben vizsgálva azonban jól látható, hogy a gyakori integrációt gyakori deléció ellensúlyozza
- gyakran egyetlen generáció alatt elvesződik a rezisztencia gén - a dolog pontos mechanizmusa még nem ismert

(Sheppard and Timmins (2009) *PLoS Gen*)



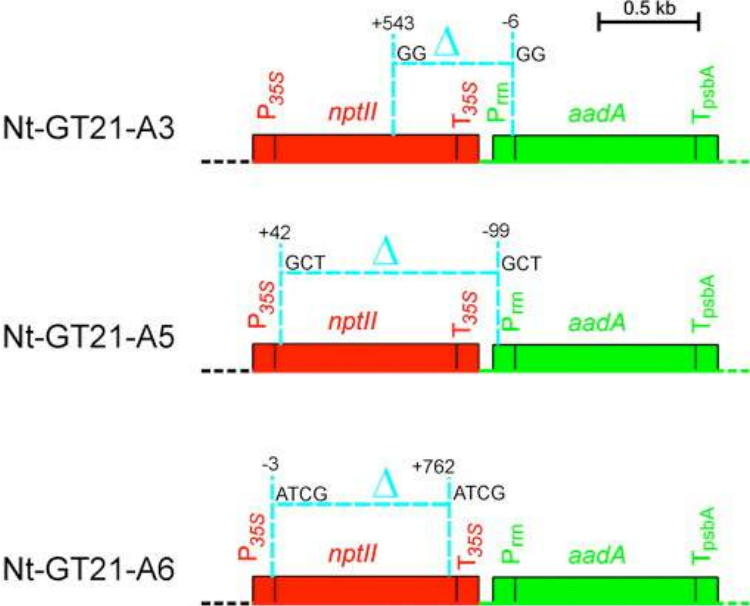
Plasztisz eredetű gének sejtmagi reaktivációja



(Bock et al. (2008) *Bioessays*)

- a plasztisz szekvenicák aktiválódásának gyakorisága kb. azonos nagyságrendű a nukleáris gén-transzfer gyakoriságával

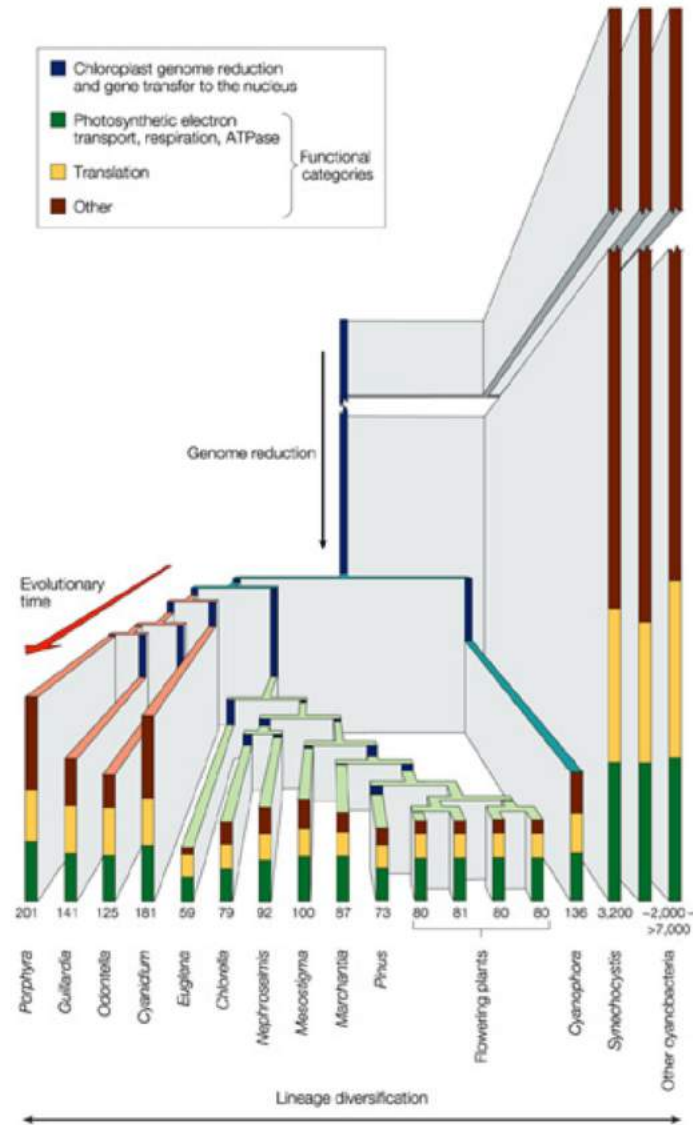
- a sejtmagba kerülő plasztisz-szekvenciák aktivációja upstream promóterek “befogásával” történik



(Stegemann and Bock (2006) *Plant Cell*)



A plasztisz-genom evolúciója



Miért nem transzlokálódott a mitokondriumok és kloroplasztiszok teljes genetikai anyaga a genomba?



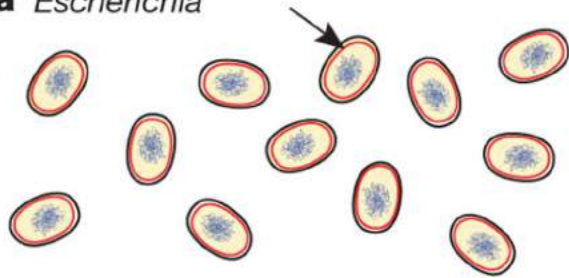
1. Az egyre kisebb genomok alternatív kodonokat hoznak létre, és az új kodontábla által kódolt fehérjék a sejtmagban nem funkcionálisak.
2. Az organelláris DNS-en kódolt gének erősen hidrofóbak, és nem lennének könnyen szállíthatóak a citoplazmán keresztül
3. CORR (COlocation of genes and gene-products for Redox Regulation of gene expression) hipotézis: bizonyos gének átíródását a bionergetikai membránok redox potenciálja szabályozza. Ezek nem kerülhetnek ki az organellumból, mert a szabályozásuk nem megoldható a sejtmagban, így az organelláris genomból való eltűnésük a mitokondrium/plasztisz működésképtelenségét okozná.

(Allen (2003) *Phil Trans R Soc Lond B*)

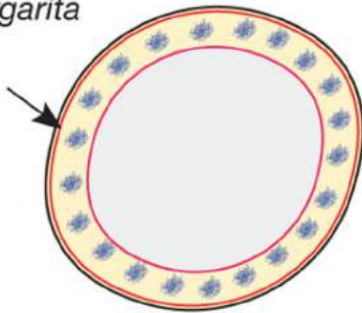
Az eukarióták komplexebb genomját a mitokondriumok teszik lehetővé



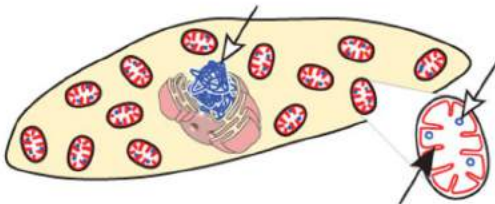
a *Escherichia*



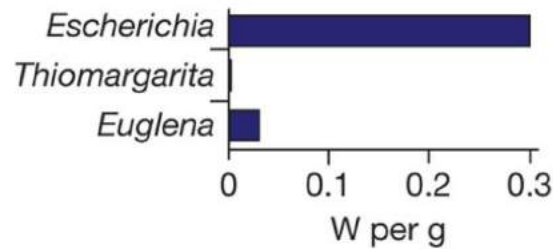
b *Thiomargarita*



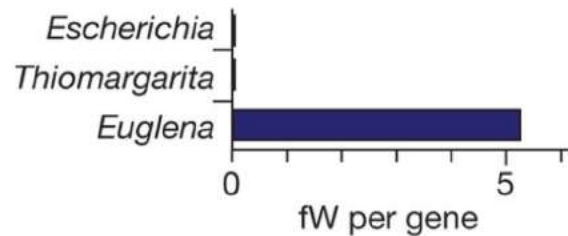
c *Euglena*



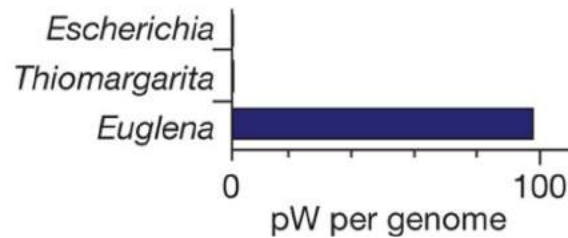
d Power per gram of cells



e Power per gene



f Power per haploid genome

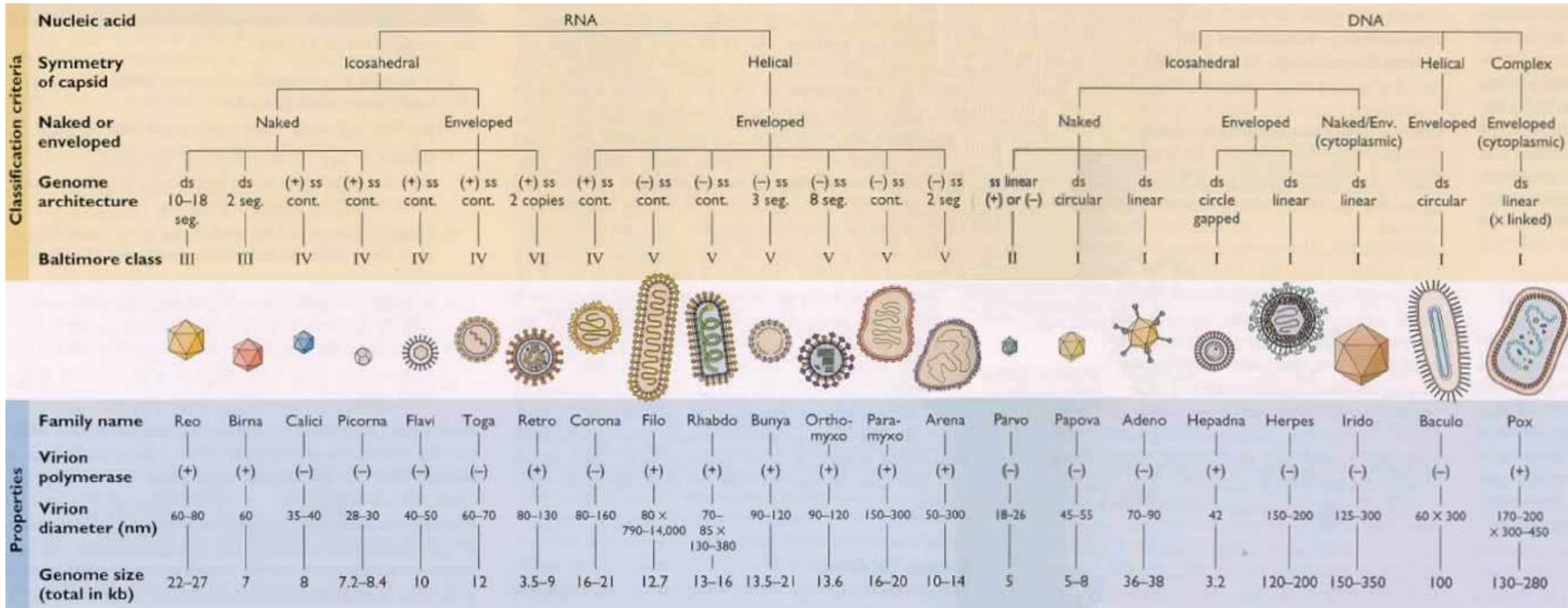


- a prokarióta sejtek energiatermelése limitált, így az egy fehérje termelésére fordítható energiamentiség függ a fehérjék számától

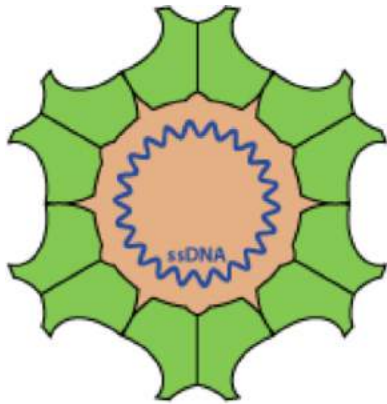
- ez felső határt szab a lehetséges kódoló szekvenciák számának

- eukariótákban a mitokondriumok révén lényegesen megnő a bioenergetikai membránok felülete és a termelt ATP mennyisége az igényekhez igazodik, ez pedig közel 400,000x nagyobb kódolókapacitáshoz vezet

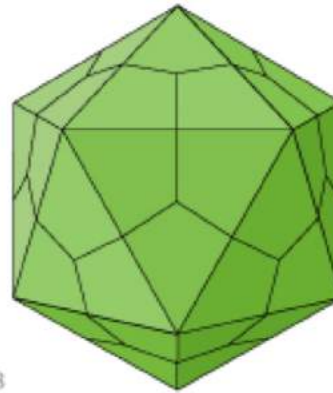
Virális genom típusok



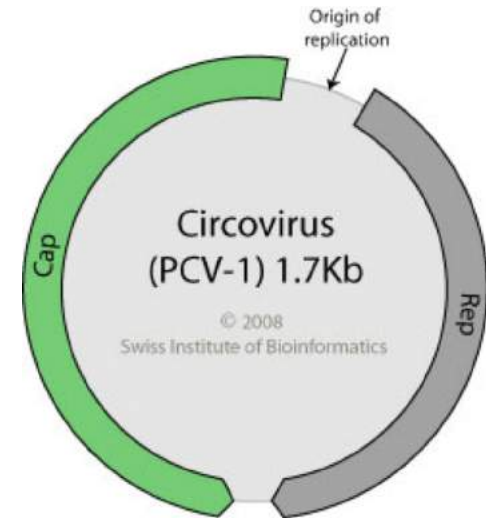
A legkisebb vírusok - *Circocoviridae*



© ViralZone 2008
Swiss Institute of Bioinformatics



T=1

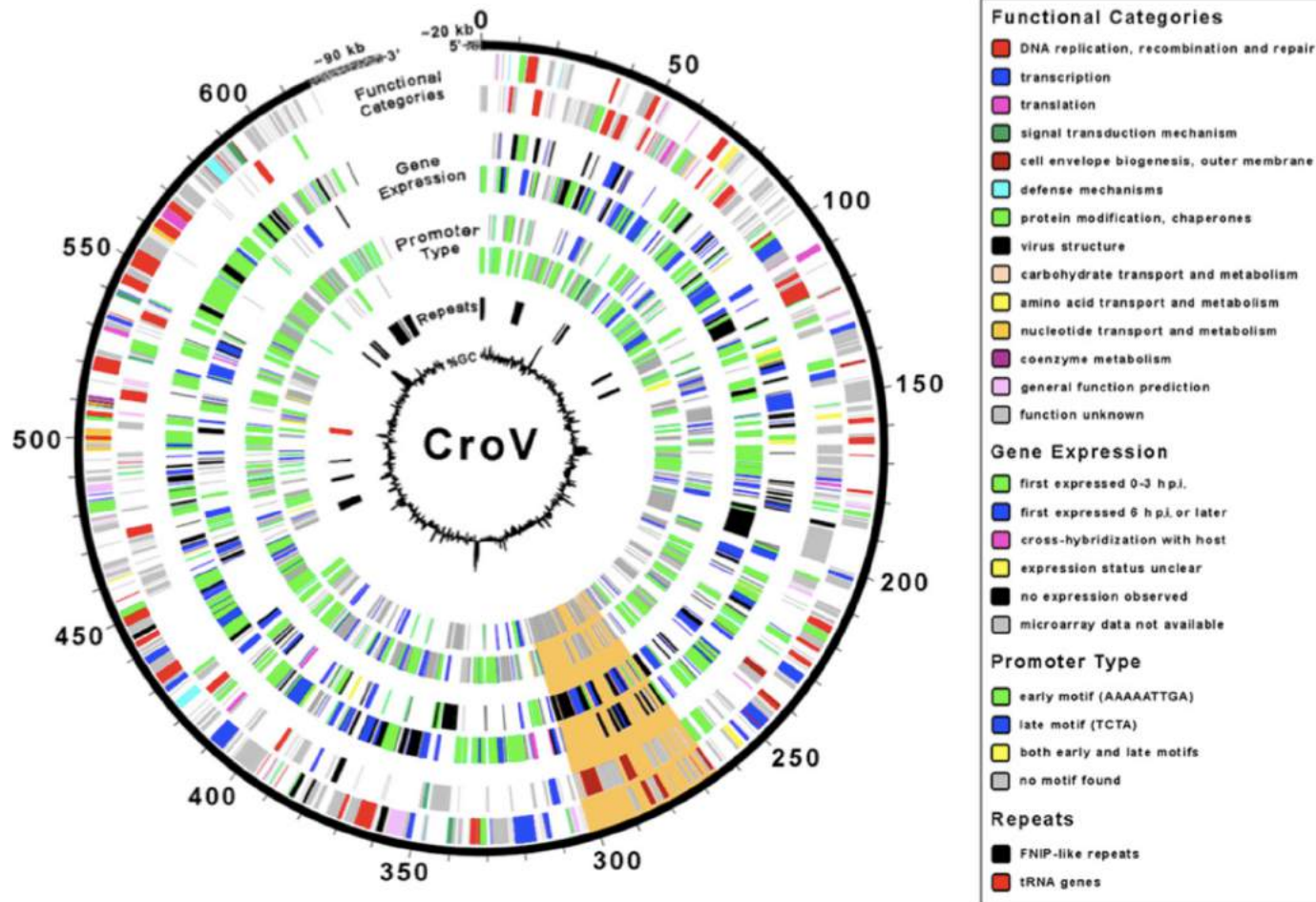


-<2 kb, cirkuláris ssDNS genom, mindössze két fehérjét kódol

- virális élekciklus:

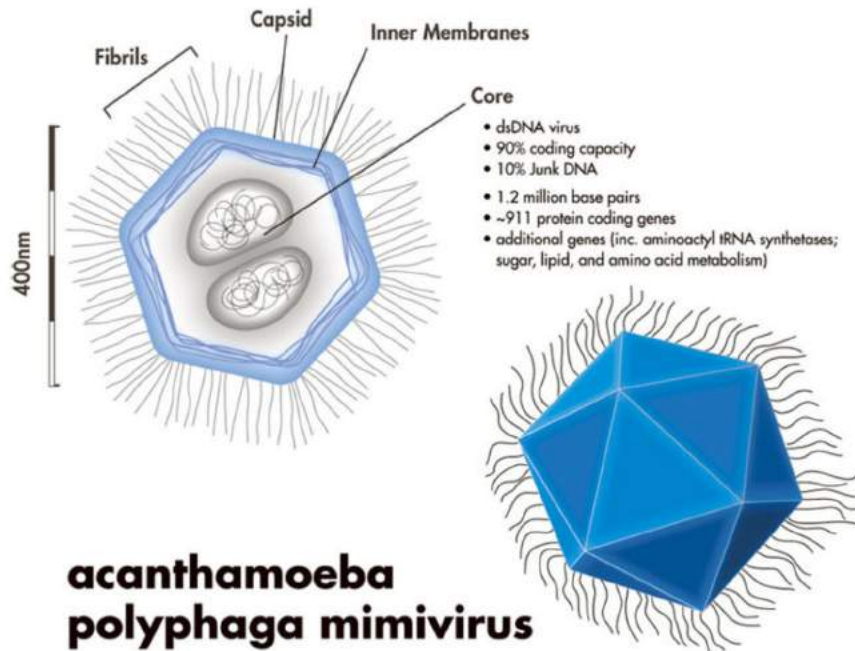
1. Virus penetrates into the host cell.
2. Uncoating, and release of the viral genomic ssDNA into the nucleus.
3. The ssDNA is converted into dsDNA with the participation of cellular factors.
4. viral mRNAs are transcribed and translated to produce viral proteins.
5. Replication may be mediated by a “Rep-like” protein, and would occur by rolling circle
6. These newly synthesized ssDNA can either
 - a) be converted to dsDNA and serve as a template for transcription/replication
 - b) be encapsidated by capsid protein and form virions released from the cell by budding

Óriás vírusok - *Cafeteria roenbergensis*

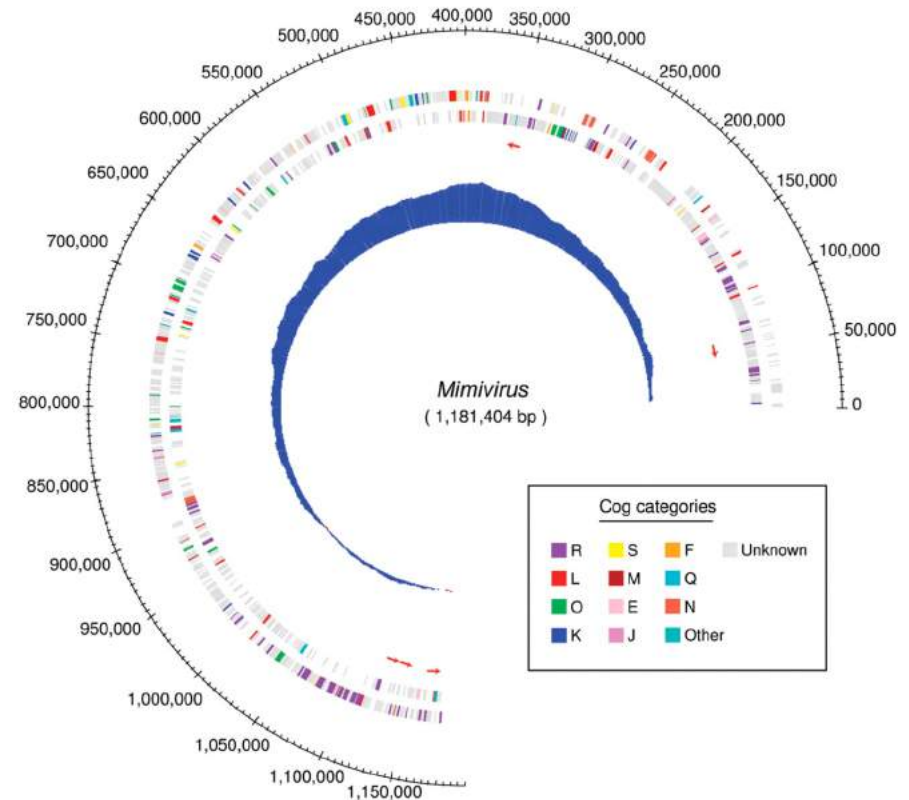


- ~730 kb dsDNA genom, ~550 gén köztük transzlációban szerepet játszó és DNS repair fehérjék is
- a genom kb 5%-a repetitív DNS és a genom egy jókora része bakteriális eredetű!

Óriás vírusok - *Acanthamoeba polyphaga* mimivirus



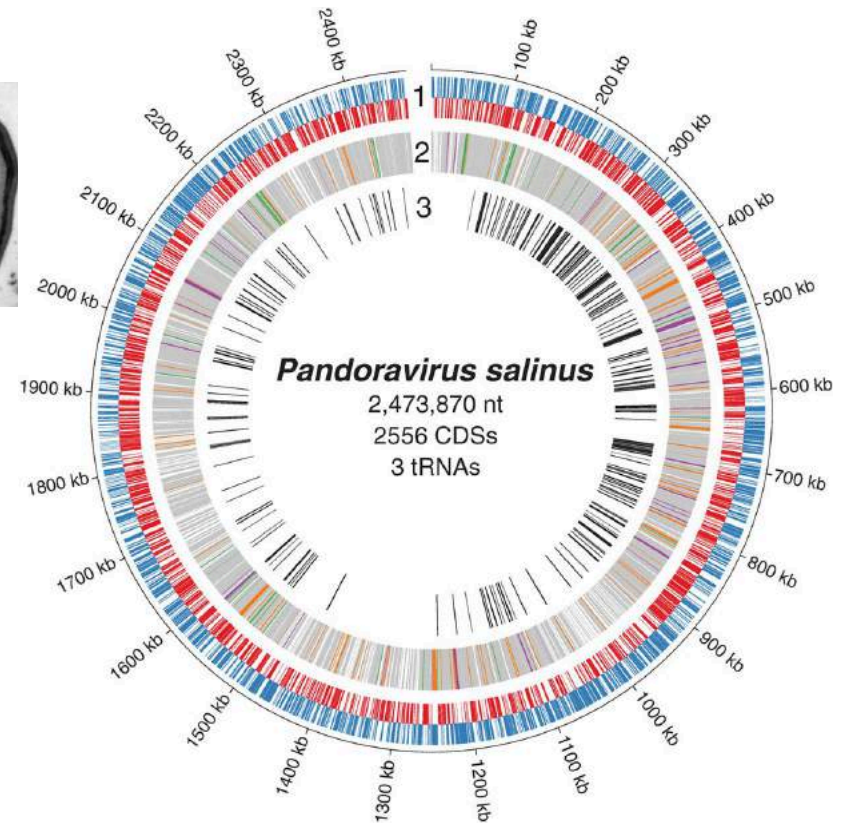
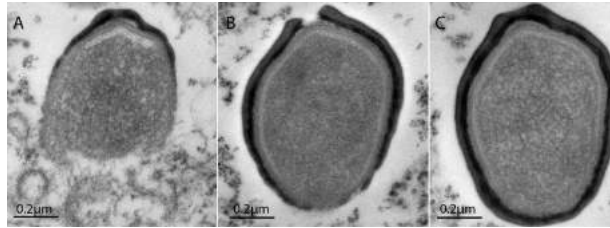
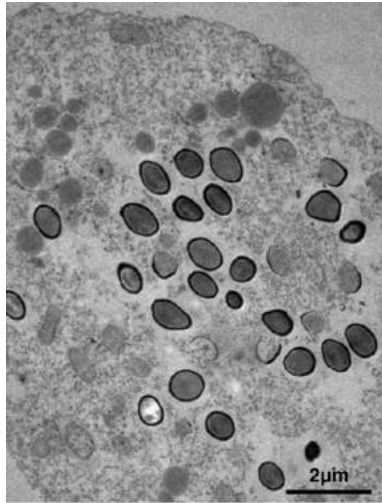
**acanthamoeba
polyphaga mimivirus**



~1.2 Mb dsDNA genom, 981 gén, szintén sok tRNA-szintáz, transzkripció iniciációért, elongációért ill terminációért felelős fehérje, valamint DNA repair enzimek

- a vírusok és a valódi sejtes élőlények közti határ genomikai szempontból nem létező: ezek a vírusok nagyobbak mint egyes patogén prokarióták vírusai, és komplex replikációs apparátust kódolnak

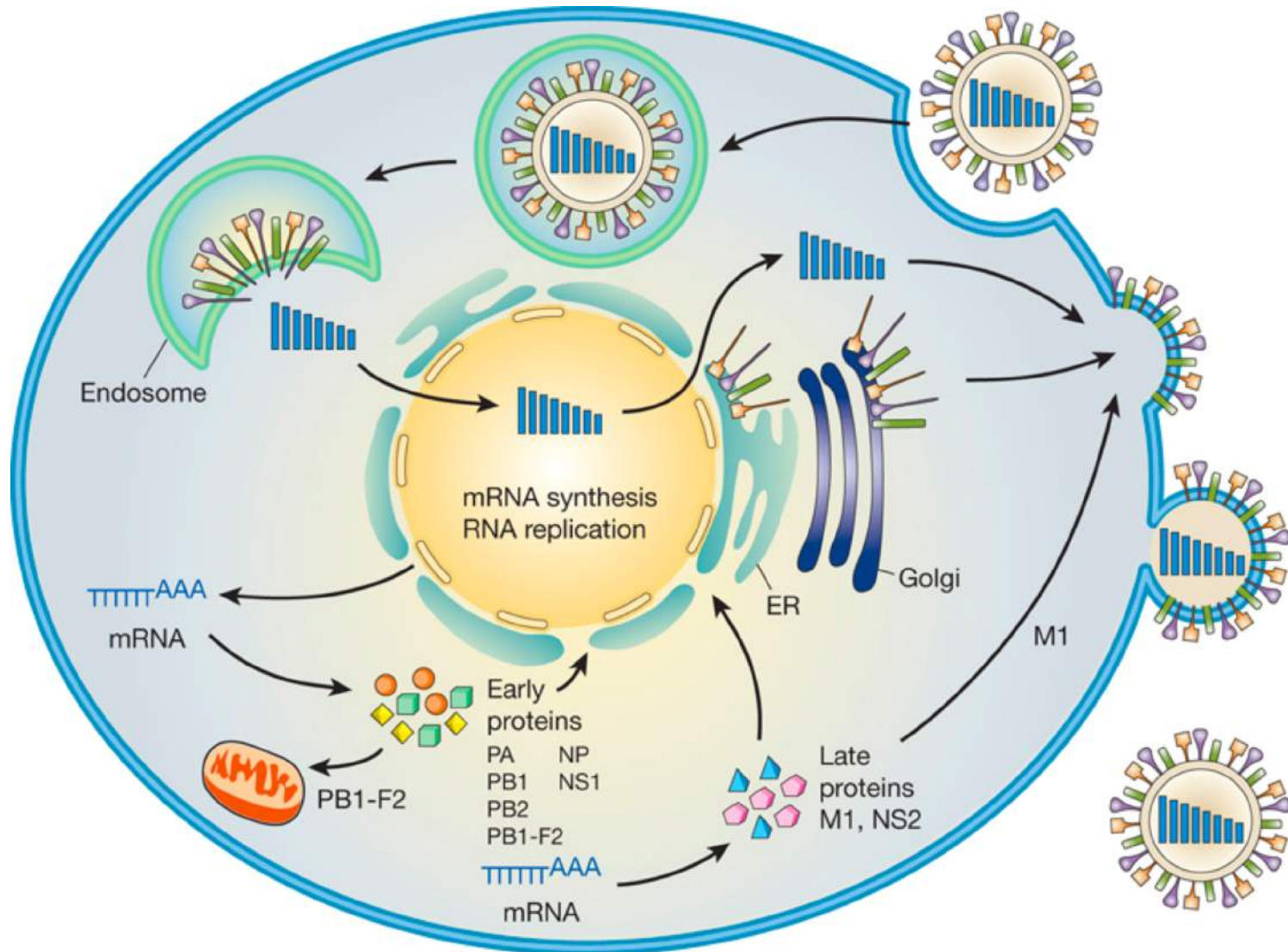
Óriás vírusok – *Pandoravirus salinus*



- az *Acanthamoeba castellanii* vírusa, Chilében fedezték fel, 1 μm hosszú, 0.5 μm széles részecskéket alkot
- ~2.5-2.8 Mb dsDNA genom, 2556 feltételezett fehérjekódoló gén
- de a legtöbbnek (93%) nincs más élőlényben homológja (ami vírusok esetében is szokatlan), így elképzelhető, hogy a transzlációjuk valami különleges szabály szerint zajlik

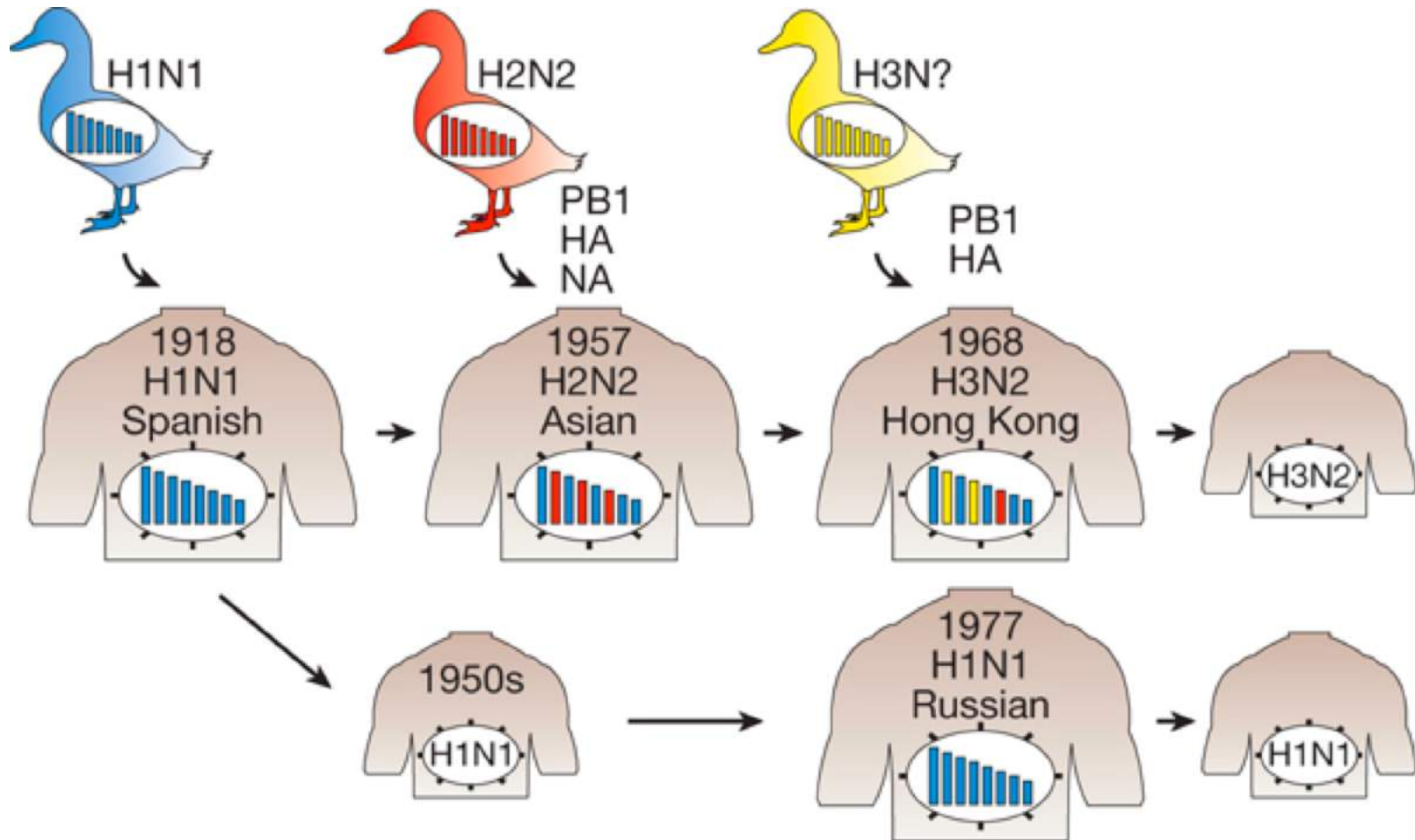
(Philippe et al. (2013) *Science*)

Influenza vírusok élelciklusa



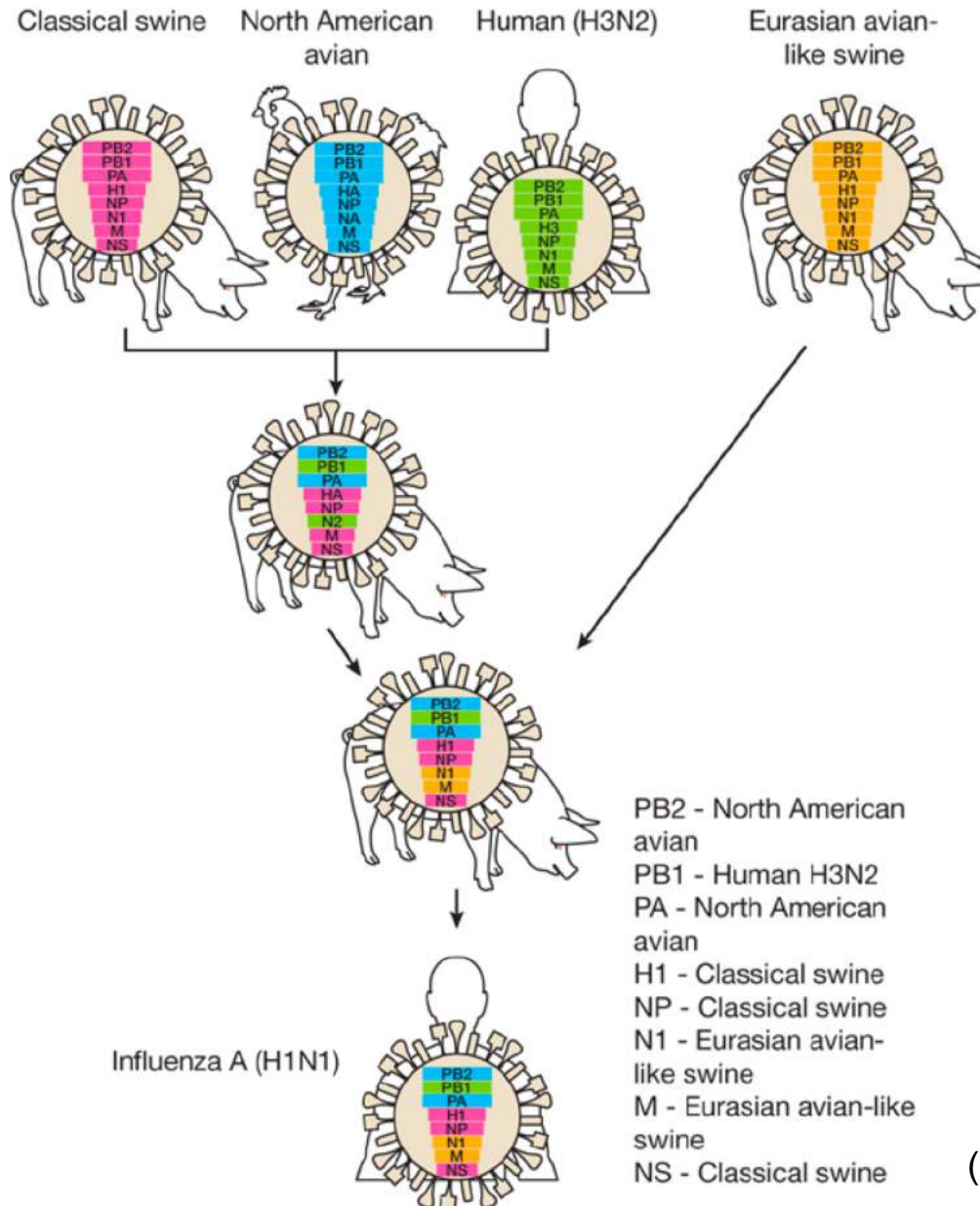
(Neumann et al. (2009) *Nature*)

Influenza vírusok evolúciója reasszortáció által



(Neumann et al. (2009) *Nature*)

Influenza - a 2009 A(H1N1) eredete



(Neumann et al. (2009) *Nature*)