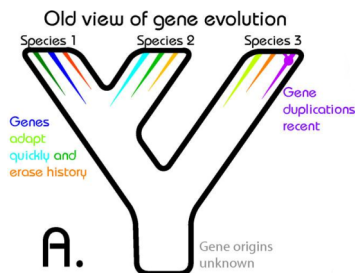


Metazoa evolúció genomikai szemszögből

Varga Máté
(m.varga@ucl.ac.uk)

A homológ gének változó paradigmája

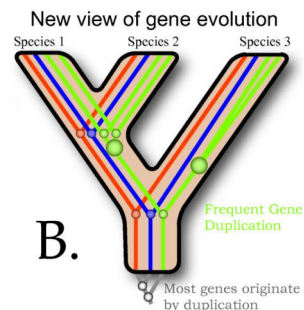


-a Modern Szintézis klasszikus nézete szerint a gének a fajok adaptív jellegei, amelyek gyorsan változnak.

"Much that has been learned about gene physiology makes it evident that the search for homologous genes is quite futile except in very close relatives". – Ernst Mayr

-a hatvanas évektől aztán fokozatosan kiderült, hogy ez nem így van, számos gén közös egészen távoli fajok közt is (hemoglobin, cyt c).

-nyolcvanas-kilencvenes években arra is fény derült, hogy a legkülönbözőbb állatok homológ szerveinek fejlődéséért is homológ gének felelnek! (*eya*, *Hox*, *tinman/nkx2.5*)



(Rose and Oakley (2007) *Biol Direct*)

Az első indirekt genom-összehasonlítások

Table 1. Differences in amino acid sequences of human and chimpanzee polypeptides. Lysozyme, carbonic anhydrase, albumin, and transferrin have been compared immunologically by the microcomplement fixation technique. Amino acid sequences have been determined for the other proteins. Numbers in parentheses indicate references for each protein.

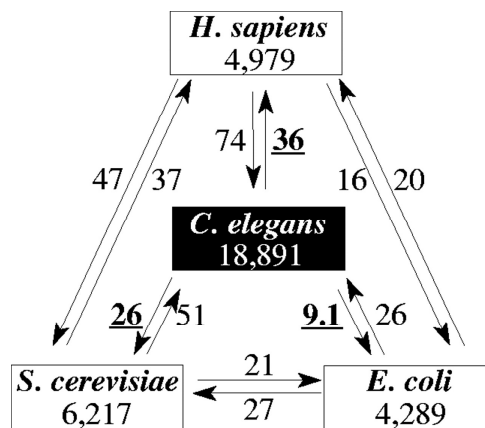
Protein	Amino acid differences	Amino acid sites
Fibrinopeptides A and B (3)	0	30
Cytochrome c (4)	0	104
Lysozyme (13)	~0	130
Hemoglobin α (4)	0	141
Hemoglobin β (4)	0	146
Hemoglobin α_{γ} (5, 6)	0	146
Hemoglobin α_{δ} (5, 6)	0	146
Hemoglobin δ (5, 8)	1	146
Myoglobin (7)	1	153
Carbonic anhydrase (4, 12)	~3	264
Serum albumin (10)	~6	580
Transferrin (11)	~8	647
Total	~19	2633

“A relatively small number of genetic changes in systems controlling the expression of genes may account for the major organismal differences between humans and chimpanzees.”

(King and Wilson (1975) *Science*)

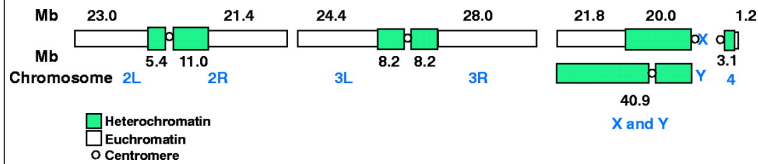
Caenorhabditis elegans

- hat pár kromoszómán 97 Mb, 20,100 prediktált fehérjekódoló gén (plusz kb 16,000 RNS-kódoló gén, de ezekre csak az utóbbi években derült fény)



(The *C. elegans* Sequencing Consortium (1998) *Science*)

Drosophila melanogaster



~180Mb, kb. 13,600 (fehérje kódoló) gén

-a rövidebb *C. elegans* genom több gént tartalmaz => szerkezeti komplexitás és génszám közt nincs közvetlen összefüggés

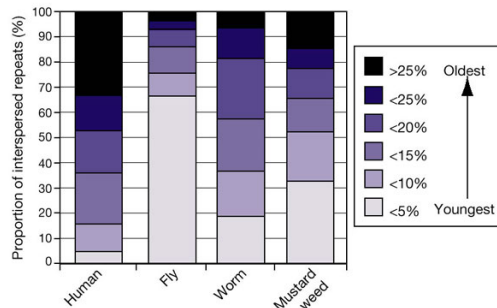
(Adams et al. (2000) *Science*)

Az emberi genom rengeteg, régi ugráló elemet tartalmaz

Table 12 Number and nature of interspersed repeats in eukaryotic genomes

	Human		Fly		Worm		Mustard weed	
	Percentage of bases	Approximate number of families	Percentage of bases	Approximate number of families	Percentage of bases	Approximate number of families	Percentage of bases	Approximate number of families
LINE/SINE	33.40%	6	0.70%	20	0.40%	10	0.50%	10
LTR	8.10%	100	1.60%	50	0.00%	4	4.80%	70
DNA	2.80%	60	0.70%	20	5.30%	80	5.10%	80
Total	44.40%	170	3.10%	90	6.50%	90	10.50%	160

The complete genomes of fly, worm, and chromosomes 2 and 4 of mustard weed (as deposited at ncbi.nlm.nih.gov/genbank/genomes) were screened against the repeats in RepBase Update 5.02 (September 2003) with RepeatMasker at sensitive settings.



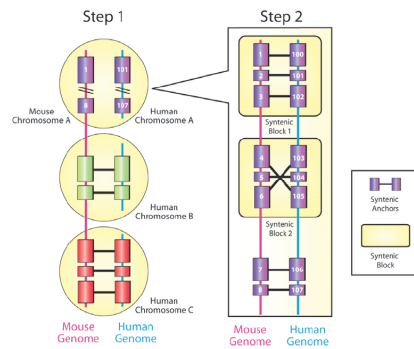
(Lander et al. (2001) *Nature*)

Az egér genom

- az egér genom kb 2.5 GB (vagyis 14%-al kisebb a humán genomnál) -
 a különbség az egér magasabb deléciós rátájának köszönhető

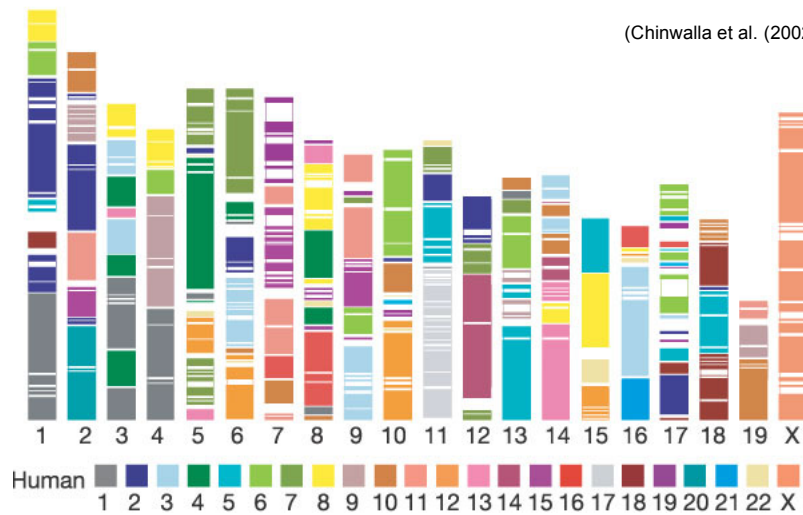
- mind a humán, mind az egér genom kb azonos mennyiségű gént
 tartalmaz (kb. 30,000, akkori számítások szerint), és ezek 80%-a
 megfeleltethető egymásnak

- a genomok közel 90%-a szintenikus blokkokba állítható!



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

Az egér és humán genomok nagyfokú hasonlósága



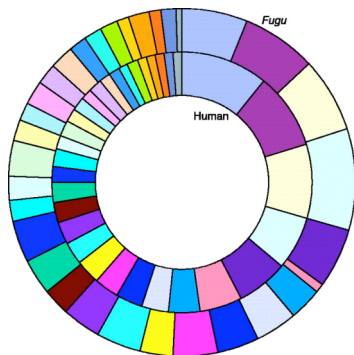
(Chinwalla et al. (2002) *Nature*)

- Minden egyes csík egy >300 kb nagyságú szintenikus blokknak felel meg

A gömbhal (*Takifugu rubripes*)

-a 332.5 Mb hosszú genom kb 31,000 gént kódol (akkori predikciók alapján)

-utóbbi elég jól megegyezik az emberi genomban feltételezett gének számával, de azok majdnem 9x nagyobb helyen helyezkednek el, 2.9 GB-on

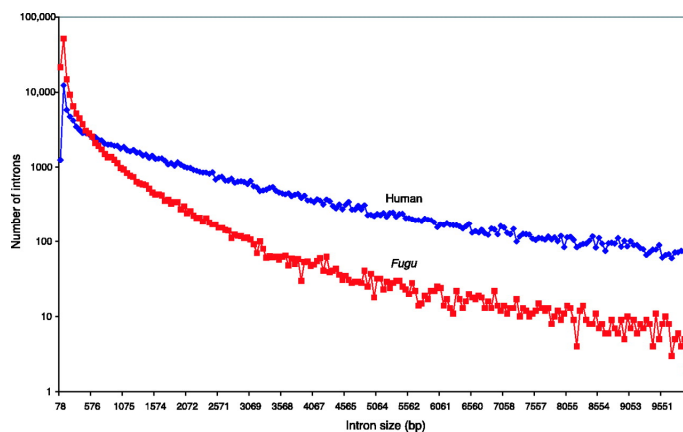


-az egyes fehérje családok is kb. azonos nagyságúak, kivétel a kálium-csatorna alegységek és a Zn-finger típusú transzkripció faktorok képeznek. Előbbiek a *Fugu* genomban vannak többen, utóbbiak pedig az emberiben.

(Aparicio et al. (2002) *Science*)

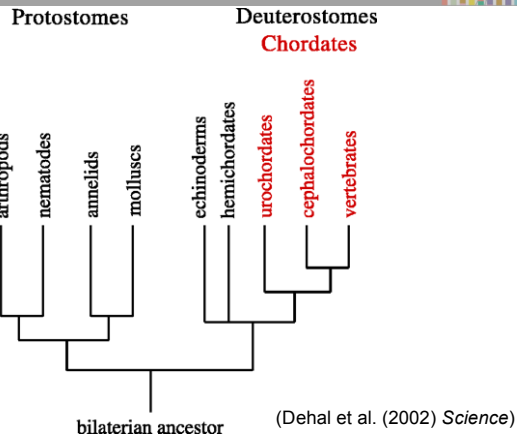
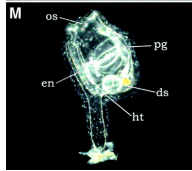
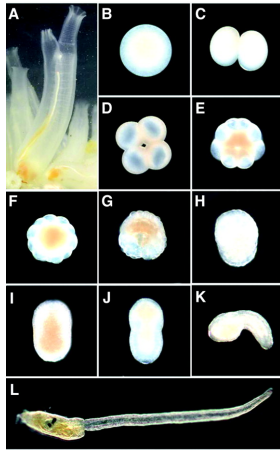
Egy kompakt gerinces genom: kevesebb repetitív szekvenca, rövidebb intronok

- a genom mindössze 2.7%-a repetitív szekvenca, ami *lényegesen* alacsonyabb az emlősöknél talált: 35 - 45%-nál



(Aparicio et al. (2002) *Science*)

A zsákállat (*Ciona intestinalis*)

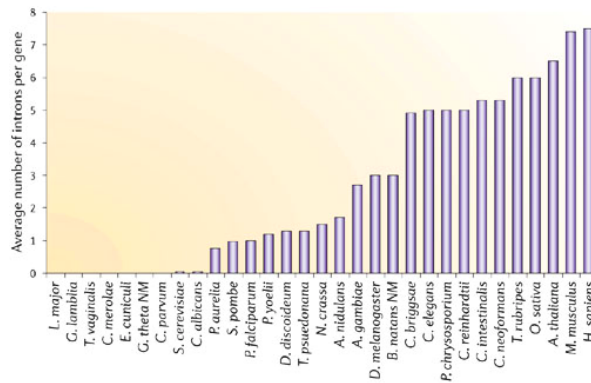


-igen kompakt: ~116.7 Mb, kb. 16,000 gén

-A gének kb 60%-nak létezik Protostom homológja, 20%-nak azonban semelyik más nagyobb csoportban nincs megfelelője

- a *Ciona* gének átlagosan 6.8 exonból állnak (vs. 5 a *Drosophila*-ban és 8.8 embekben) => intronszegény

Korai intronok, vagy késői intronok?

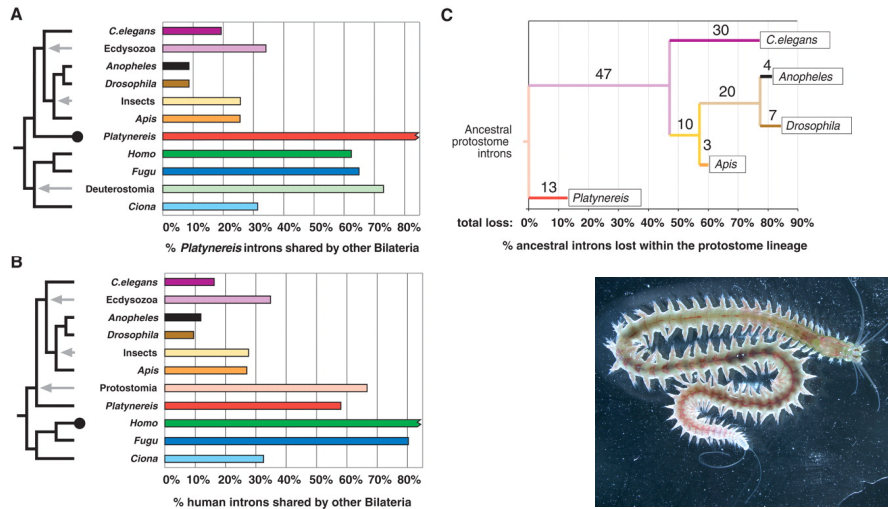


(Roy and Gilbert (2006) *Nat Rev Gen*)

Copyright © 2006 Nature Publishing Group
Nature Reviews | Genetics

- az első genom szekvenálások alapján, az tűnne logikusnak, hogy az Eumetazoa ill. Urbilateria ősök viszonylag kevés intronnal rendelkeztek, és ezek később az Újszájúak (Deuterostomia) vonalán terjedtek el igazán, főleg az emlősökben

Betekintés a *Platynereis* genomba: az intronok ősi eredete



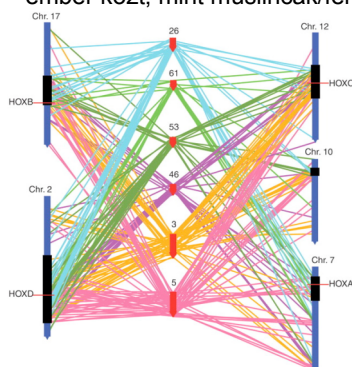
(Raible et al. (2005) *Science*)

A tengeri rózsa (*Nematostella vectensis*)

-a 450Mb nagyságú genom kb 18,000 fehérje kódoló gént tartalmaz

-a genom negyede repetitív szekvencia

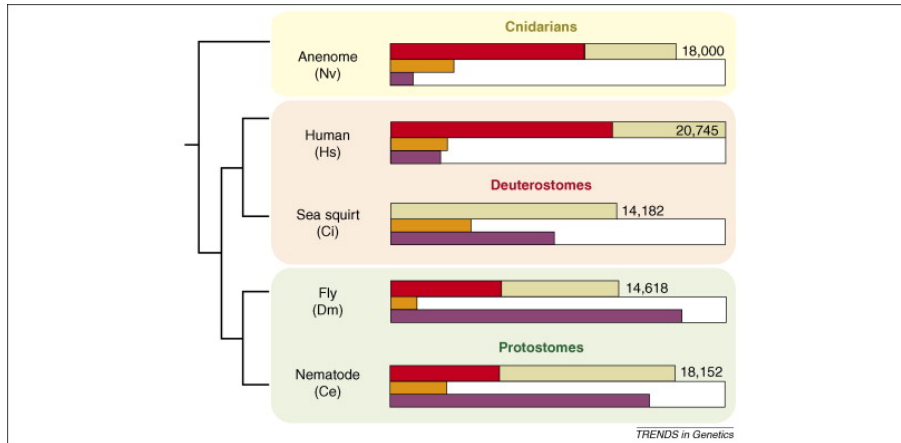
-2x annyi közös gén van egy tengeri rózsa és egy ember közt, mint muslincák/férgek és ember közt



- a 700 millió éves szétválás ellenére még ma is fellelhetők szinténikus blokkok az emberi és a *Nematostella* genom között

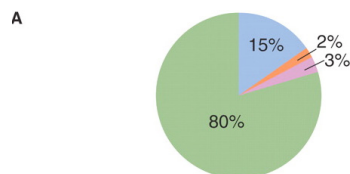
(Putnam et al (2007) *Science*)

Az emberi intronok többsége már a csalánozók és a kétoldali szimmetriájú állatok közös őseiben létezett

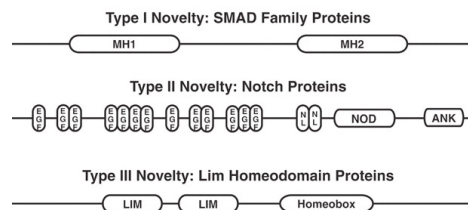


(Miller and Ball (2007) *TiG*)

Az eumetazoa gének eredete



- Type I (completely novel)
- Type II (novel domain)
- Type III (novel pairing)
- Ancient



-az ősi eumetazoa génkészlet kb 80%-a rendelkezik egyértelmű rokonokkal az állatokon kívül

- a maradék 20% eumetazoa "felfedezés":

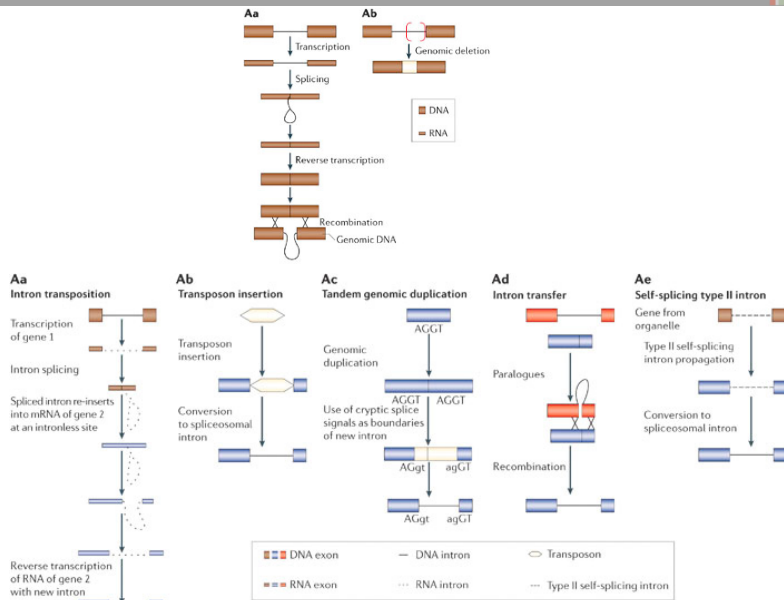
- 15% teljesen új (nem lelhető fel szekvencia hasonlóság nem állat csoportokkal (Type I))

- 2% olyan gének, amelyek egyes doménjei már léteznek más csoportokban is, mások azonban újak (Type II)

- 3% olyan gének, amelyek minden doménje létezett korábban, de nem ebben a kombinációban (Type III)

(Putnam et al. (2007) *Science*)

Az intronvesztés és intron születés feltételezett mechanizmusai



(Roy and Gilbert (2006) *Nat Rev Gen*)

A soksejtűség eredete egy gallérs ostoros genomján át (*Monosiga brevicollis*)

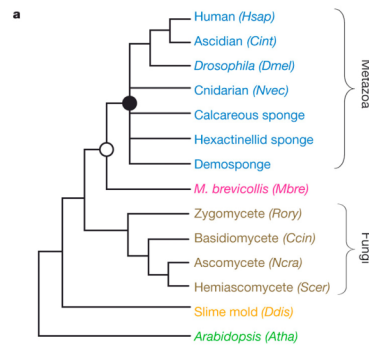
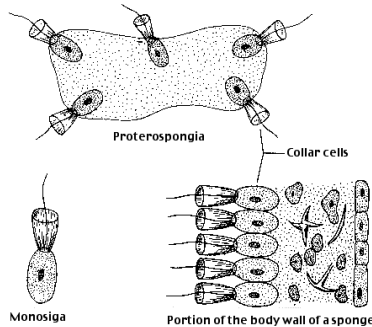


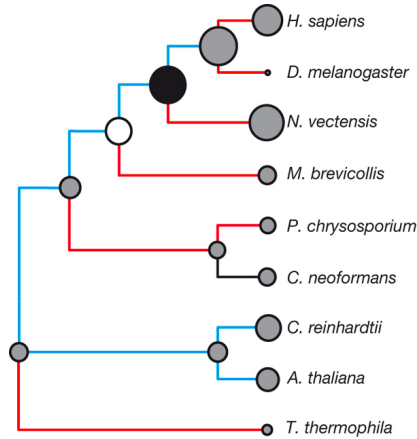
Table 1 | *M. brevicollis* genome properties in a phylogenetic context

	Metazoa				Choanoflagellates	Fungi		Dictyostelium	Plants
	<i>Hsap</i>	<i>Cint</i>	<i>Dmel</i>	<i>Nvec</i>	<i>Mbre</i>	<i>Ccin</i>	<i>Nra</i>	<i>Ddis</i>	<i>Atha</i>
Genome size (Mb)	2,900	160	180	357	42	38	39	34	125
Total number of genes	23,224	14,182	14,601	18,000	9,196	13,544	9,826	13,607	27,273
Mean gene size (bp)	27,000	4,585	5,247	6,264	3,004	1,679	1,528	1,756	2,287
Mean intron density (introns per gene)	7.7	6.8	4.9	5.8	6.6	4.4	1.8	1.9	4.4
Mean intron length (bp)	3,365	477	1,192	903	174	75	136	146	164
Gene density (kb per gene)	127.9	11.9	13.2	19.8	4.5	2.7	4.0	2.5	4.5

Species names follow the four-letter convention from Fig. 1.

(King et al. (2008) *Nature*)

Intron-evolúció

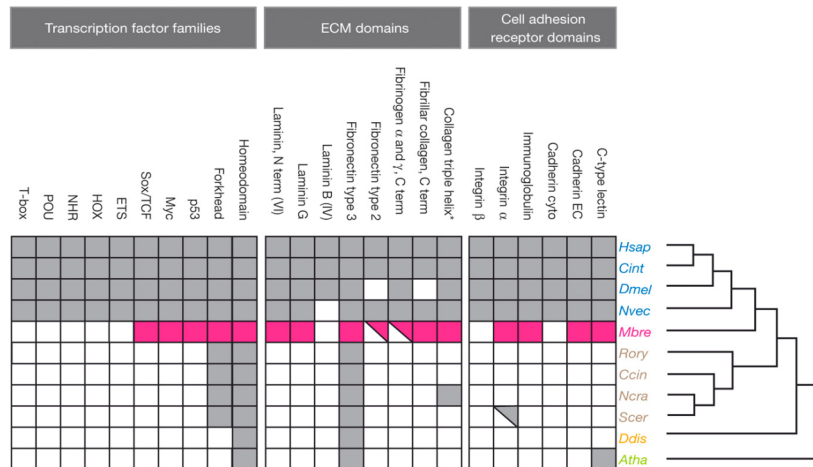


- a körök mérete az intronok számával arányos

- **kék** jelöli azokat a vonalakat, ahol összességében intronok keletkeztek, **piros** azokat ahol elvesződtek, és **fekete** azokat, ahol kb. ua. intron maradt

(King et al. (2008) Nature)

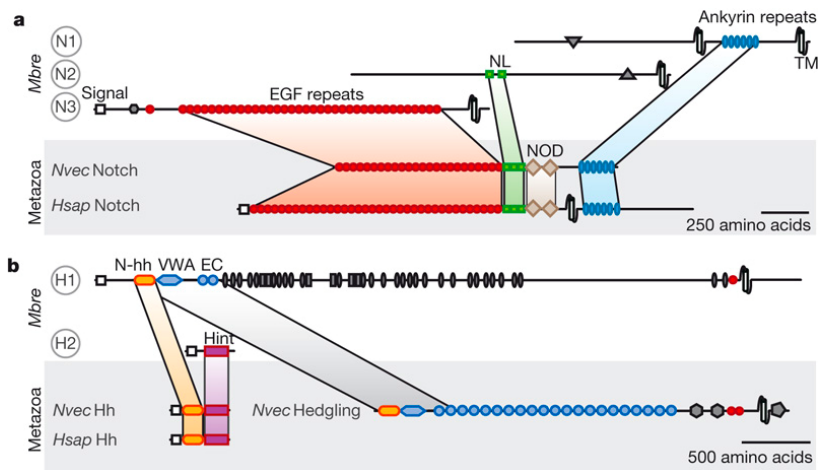
A többsejtűség / sejt-adhézió alapjai már ez egysejtű közös ősből jelen voltak



- igaz, az egyes sejtadhéziós és ECM domének sajátos elrendeződésben jelennek meg a *Monosiga* genomában, igazi ortológokat csak ritkán találni

(King et al. (2008) Nature)

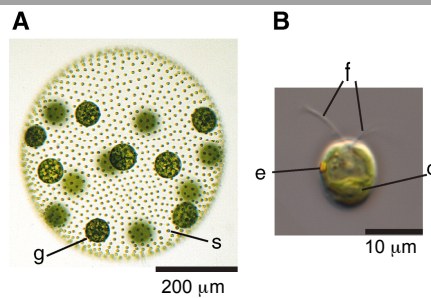
A többsejtű jelátviteli útvonalak eredete: domén-shuffling



- a WNT, TGF β , JAK/STAT útvonalaknak ugyanakkor gyakorlatilag nincs nyoma

(King et al. (2008) *Nature*)

A többsejtség eredete 2. – a *Volvox* genom

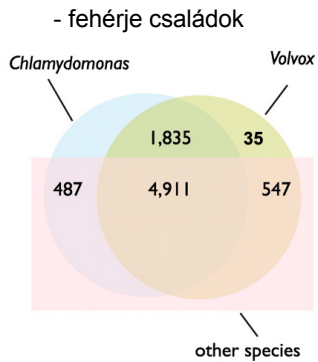


Species	Genome size (Mbp)	Number of chromosomes	% G and C	Protein-coding loci	% coding	% of genes with introns	Introns per gene	Median intron length (bp)
<i>V. carteri</i>	138	14*	56	14,520	18.0	92	7.05	358
<i>C. reinhardtii</i>	118	17	64	14,516	16.3	91	7.4	174

- a 17%-os genom növekedés nagy része a repetitív szekvenciák felszaporodásának köszönhető

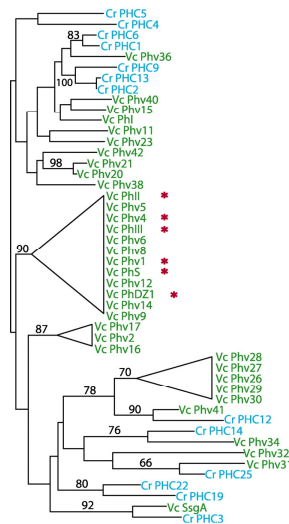
(Prochnik et al. (2010) *Science*)

A többséjtség eredete 2. – a *Volvox* genom

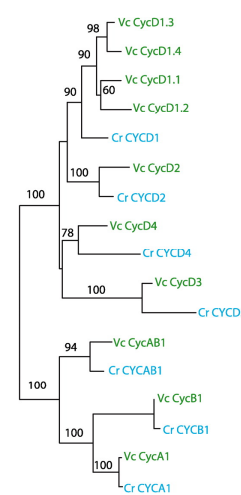


=> Az igazán fontos változások feltehetőleg itt is a szabályozó szekvenciákban következtek be!

A ECM (pherophorins)

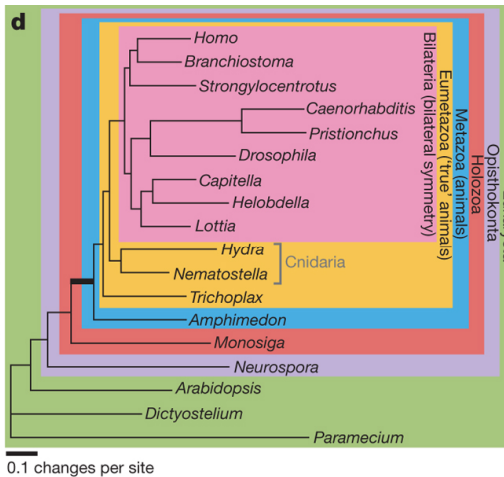
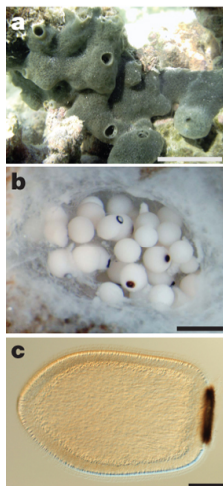


B sejtciklus



(Prochnik et al. (2010) Science)

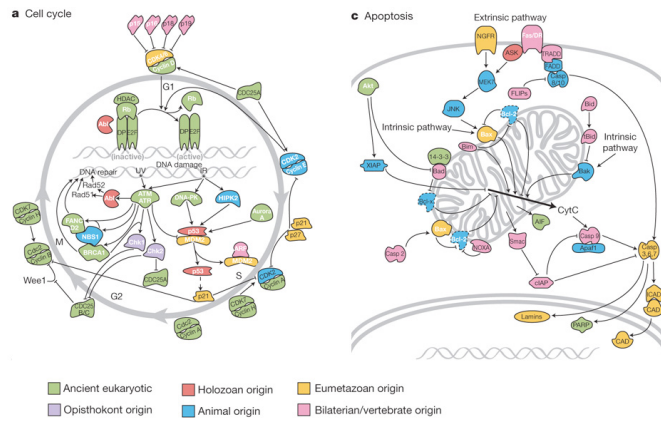
A szivacs genom (*Amphimedon queenslandica*)



- kb. 30,000 fehérje kódoló génje van, amelyek 63%-ának létezik más állatban megfelelője
- a valószínűsített ősi metazoa intronok 84%-a megtalálható

(Srivastava et al. (2010) Nature)

A sejtciklus gének ősisége vs az apoptózis gének fiatalsága

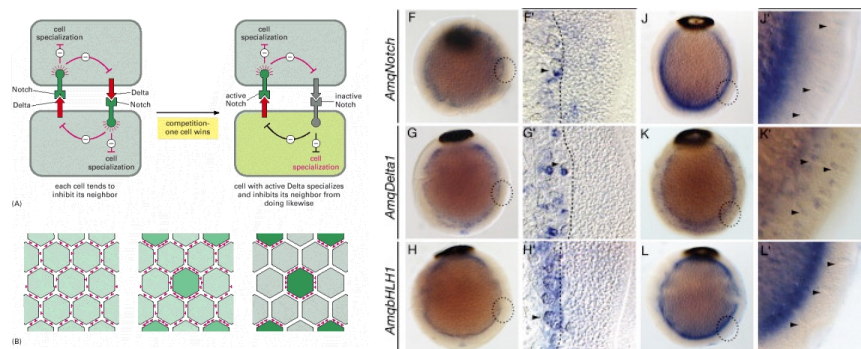


- míg a sejtciklus gének többsége ősi eukariótáktól származtatható, a programozott sejthalál állati "találmány"

(Srivastava et al. (2010) *Nature*)

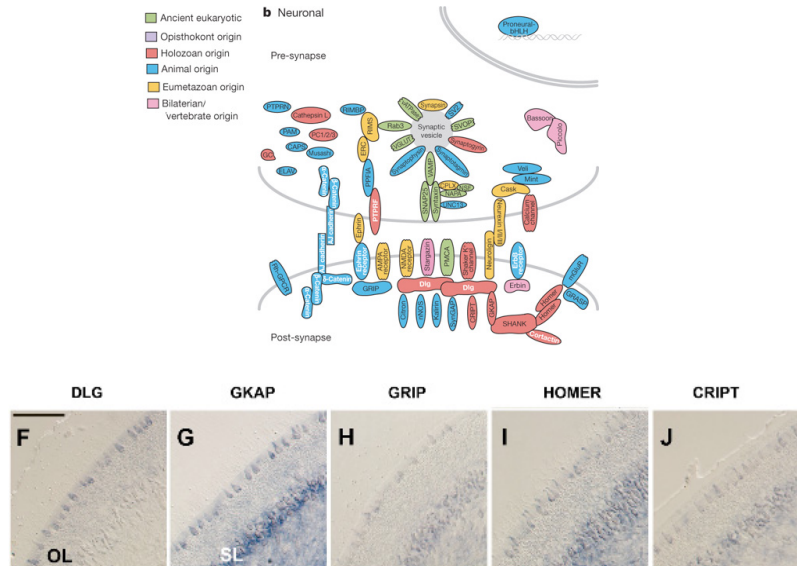
Az idegrendszer kezdetei

- az idegrendszer differenciációjáért felelős Notch-Delta jelátviteli útvonal elemei a szívacsokban jelennek meg és egy lárvális sejttípusban fejlődnek ki



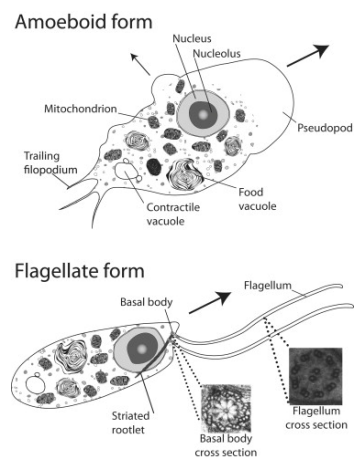
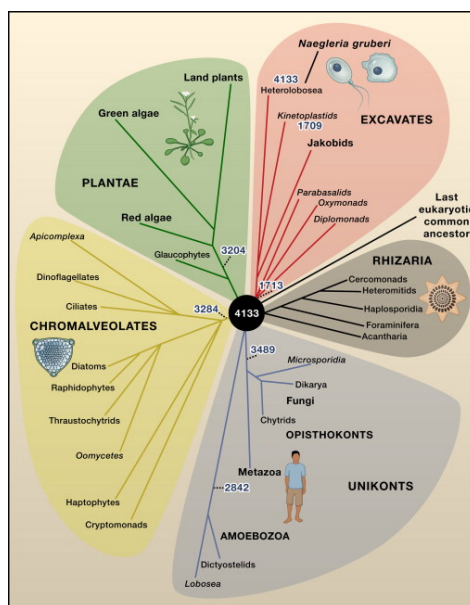
(Richards et al. (2008) *Curr Bio*)

Az idegrendszer kezdetei: a posztszinaptikus komplex elemei már jelen vannak a szivacsokban



(Sakarya et al., (2007) *PLoS One*; Srivastava et al. (2010) *Nature*)

Az ősi eukarióta sejt, egy különös protistán keresztül (*Naegleria gruberi*)



(Fritz-Laylin et al. (2010) *Cell*)

Az ősi eukarióta sejt, egy külön protistán keresztül (*Naegleria gruberi*)

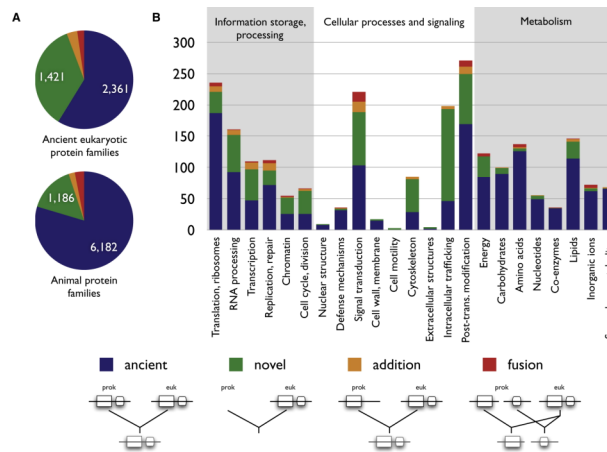
- a *Naegleria* képes aerob és anaerob metabolizmusra is
- ostoros és amöboid formája közt mindössze másfél óra alatt képes váltani

Species	Genome Size (Mbp)	No. Chromosomes	%GC	Protein-Coding Loci	% Coding	% Genes w/ Introns	Introns per Gene	Median Intron Length (bp)
<i>Naegleria</i>	41	>= 12	33	15,727	57.8	36	0.7	60
Human	2851	23	41	23,328	1.2	83	7.8	20,383
<i>Neurospora</i>	40	7	54	10,107	36.4	80	1.7	72
<i>Dictyostelium</i>	34	6	22	13,574	62.2	68	1.3	236
<i>Arabidopsis</i>	140.1	5	36	26,541	23.7	80	4.4	55
<i>Chlamydomonas</i>	121	17	64	14,516	16.3	91	7.4	174
<i>T. brucei</i>	26.1	>100	46	9152	52.6	~0 (1 total)	ND	ND
<i>Giardia</i>	11.7	5	49	6480	71.4	~0 (4 total)	ND	ND

- a *Naegleria* intronok közül számos ortológ pozícióban van más eukarióták intronjaival
- a genom mindössze 5.1%-a repetitív szekvencia

(Fritz-Laylin et al. (2010) *Cell*)

Az ősi eukarióta sejt, egy külön protistán keresztül (*Naegleria gruberi*)



- az eukarióta fehérje családok 40%-a új a baktériumokhoz képest

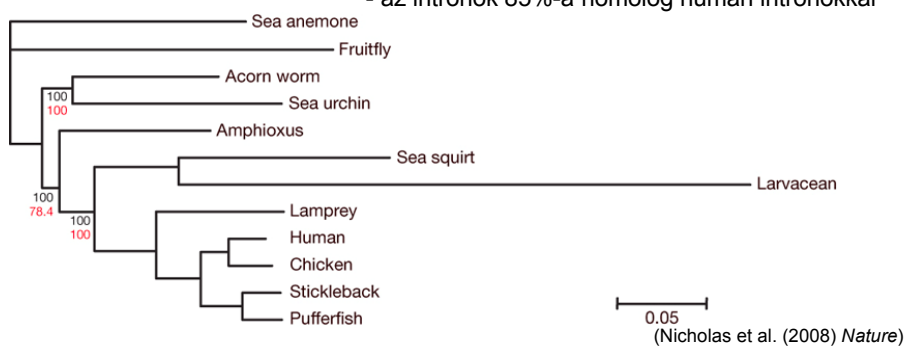
- a Metazóák - más eukarióták viszonylatban csak 20% ez az arány

(Fritz-Laylin et al. (2010) *Cell*)

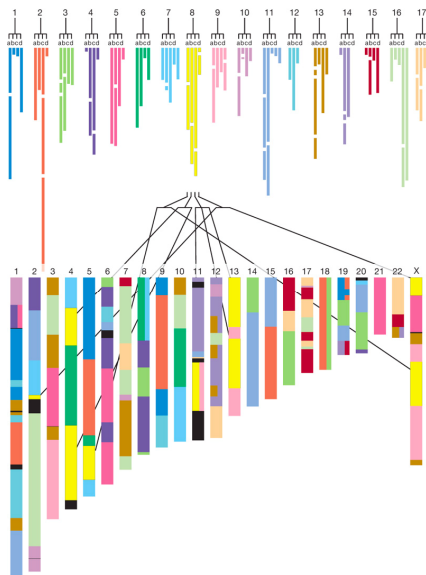
Az *Amphioxus* genom és a gerinces genom eredete



- a közös ős 550 millió éve élt
- ~520 Mb, 19 kromoszómán (de csak 17 fizikai scaffold összeszerelése sikerült)
- 21,900 fehérje kódoló lókuszt
- a genom 30%-a transzpozon eredetű
- az intronok 85%-a homológ humán intronokkal



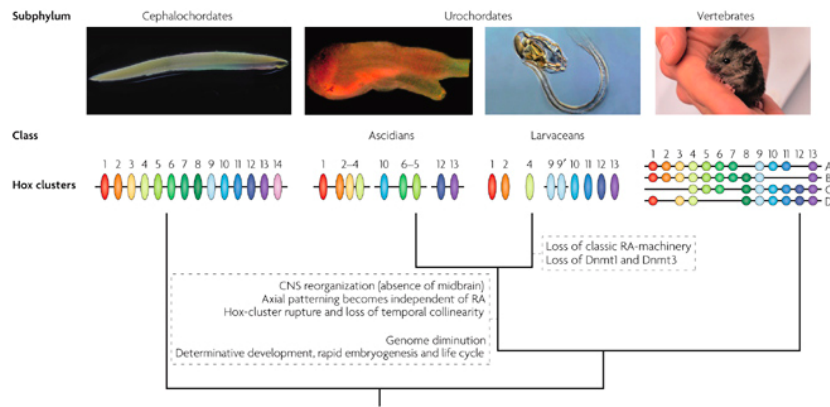
Korai genom duplikációk a gerincesek evolúciója során



- a nagyfokú szinténia lehetővé teszi, hogy jól szemléltethessük a gerincesek evolúciójának kezdetén történt genomi duplikációkat: a legtöbb *Amphioxus* genom régióknak négy gerinces megfelelője van

(Nicholas et al. (2008) Nature)

Korai genom duplikációk a gerincesek evolúciója során: a Hox-cluster eredete



Nature Reviews | Genetics

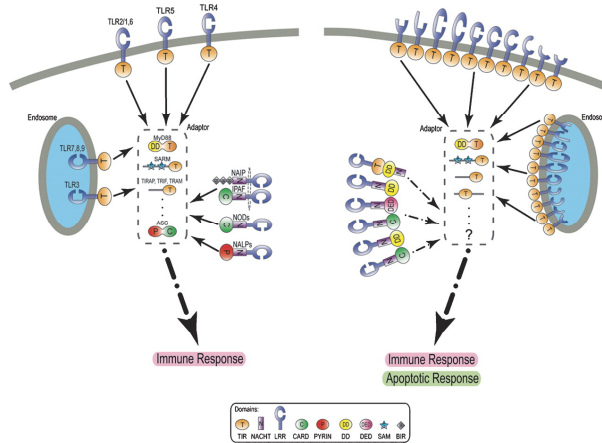
- a korai genom duplikációk egyik legjobb példája a Hox-luster esete
- az *Amphioxus* genomban egyetlen, de teljes cluster van jelen
- az előgerinchúrosok esetében a specializált életmód a cluster fragmentációjával járt

(Cañestro et al. (2007) *Nat Rev Gen*; Holland et al. (2008) *Genome Res*)

A természetes immunitás evolúciója *Amphioxus*-ban

	Innate immunity	Adaptive immunity
Specificity	For structures shared by classes of microbes ("molecular patterns") Different microbes Identical mannose receptors	For structural detail of microbial molecules (antigens); may recognize non-microbial antigens Different microbes Distinct antibody molecules
Receptors	Encoded in germline; limited diversity LPS receptor N-formyl methionyl receptor Mannose receptor Scavenger receptor	Encoded by genes produced by somatic recombination of gene segments; greater diversity TCR Ig
Distribution of receptors	Non-clonal: identical receptors on all cells of the same lineage	Clonal: clones of lymphocytes with distinct specificities express different receptors
Discrimination of self and non-self	Yes; host cells are not recognized or they may express molecules that prevent innate immune reactions	Yes; based on selection against self-reactive lymphocytes; may be imperfect (giving rise to autoimmunity)

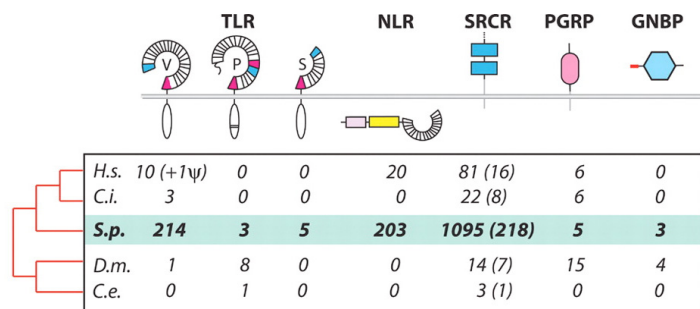
A természetes immunitás evolúciója *Amphioxus*-ban



- két-háromszor annyi *Toll-receptor* gén, mint gerincesekben
- apoptózisban részt vevő fehérje-családok hasonló szintű expansziója is valószínűleg az immunitással függ össze.

(Holland et al. (2008) *Genome Res*)

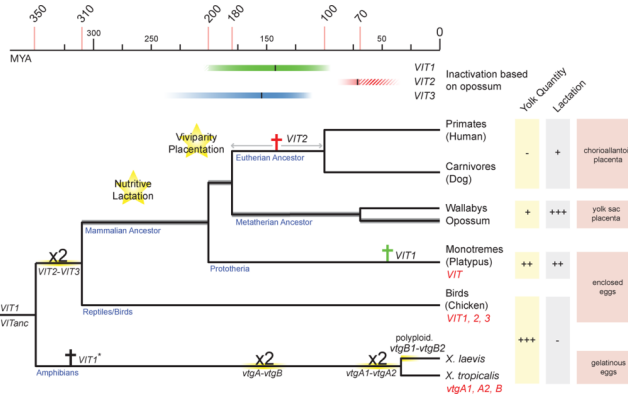
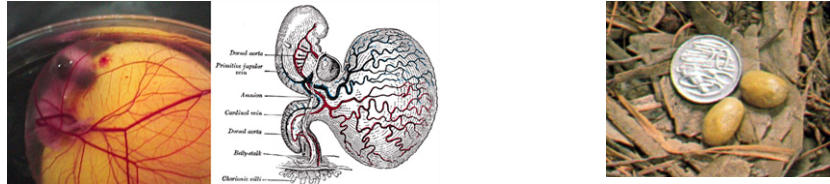
A természetes immunitás (konvergens) evolúciója tengeri sünökben (*S. purpuratus*)



- a tengeri sünök genomjában levő gének 4-5%-a az immunitásban játszik szerepet

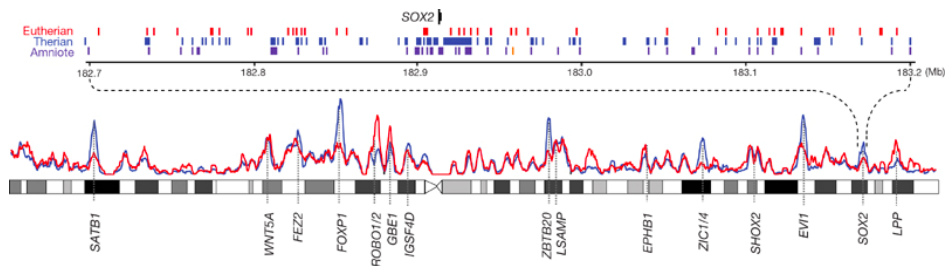
(Rast et al. (2006) *Science*)

A laktációra való áttérés a szikanyagot kódoló gének elsorvadásával jár



(Brawand et al. (2008) PLoS Biol)

Emlős szabályozó szekvenciák eredete – az oposszum genom (*Monodelphis domestica*)

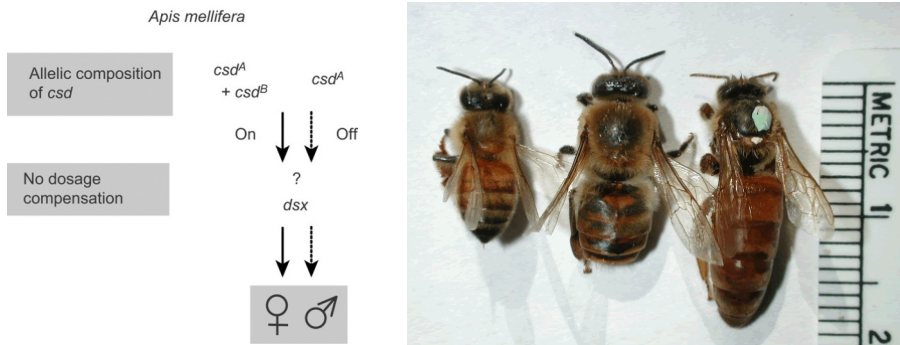


-a magzatburkos állatokban (Amniota) található CNE-k fele, a méhlepényes emlősökben levők 35%-a 204 nagy clusterben van

- ezek kb 240, lassan változó, fejlődésben kulcsfontosságú gént vesznek körül
=> pleiotróp gének szabályozásának finomhangolása fontos evolúciós változásokkal járhat

(Mikkelsen et al. (2007) Nature))

A fejlődés epigenetikai szabályozása: a méh (*Apis mellifera*) genom



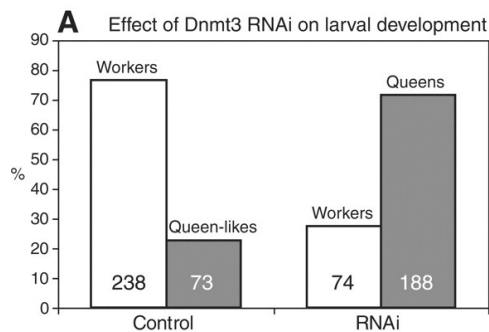
- a haploid/hemizigóta állatok hímek lesznek, a diploidok nőstények
- a királynő és a dolgozók genotípusa azonban azonos, az *egyetlen* különbség köztük a korai táplálékukban rejlik: a méhpempővel (royal jelly) táplált nőstényekből királynő lesz
- hogyan alakulhat ki azonos genotípusból ennyire eltérő fenotípus?

(Weinstock et al. (2006) *Nature!*)

Metiláció és kaszt-alapú társadalom

	Total	Methylated in Queens	Methylated in Workers	Methylated in Both Castes
CG	10,030,209	69,064	68,222	54,312
CHG	8,673,113	14	130	0
CHH	45,072,611	561	3,019*	0

The thresholds used for methylation calls are detailed in the Methylation Assessment section.
 *Nearly all of the 3,019 CHH that were inferred to be methylated in worker brains on the basis of Solexa reads were found to be not methylated by an additional sequencing of selected amplicons using the 454 technology.
 doi:10.1371/journal.pbio.1000506.t001



(Kucharski et al. (2008) *Science*; Lyko et al. (2010) *PLoS Biol*)

Metiláció és kaszt-alapú társadalom

