

Eduard Kejnovský + Roman Hobza

# EVOLUČNÍ GENOMIKA

## III. EVOLUCE GENOMŮ



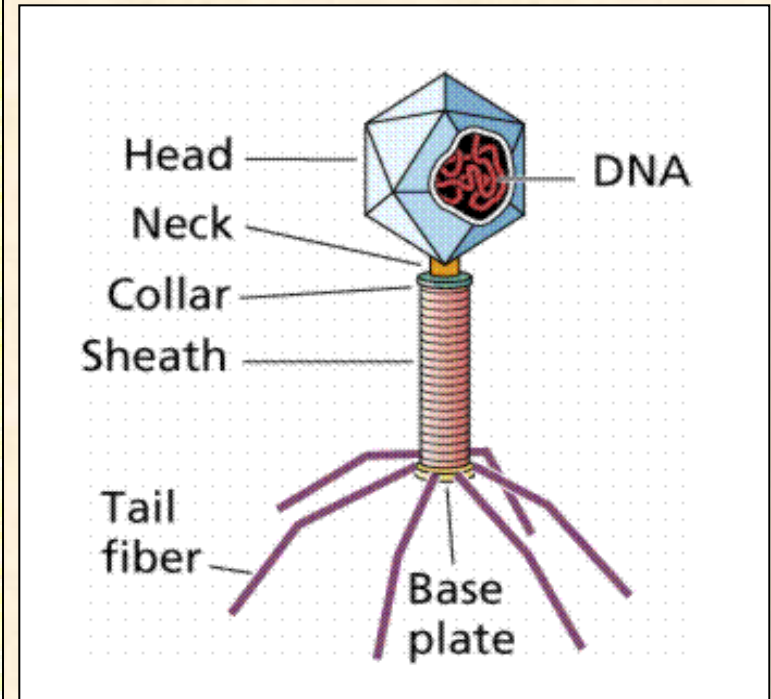
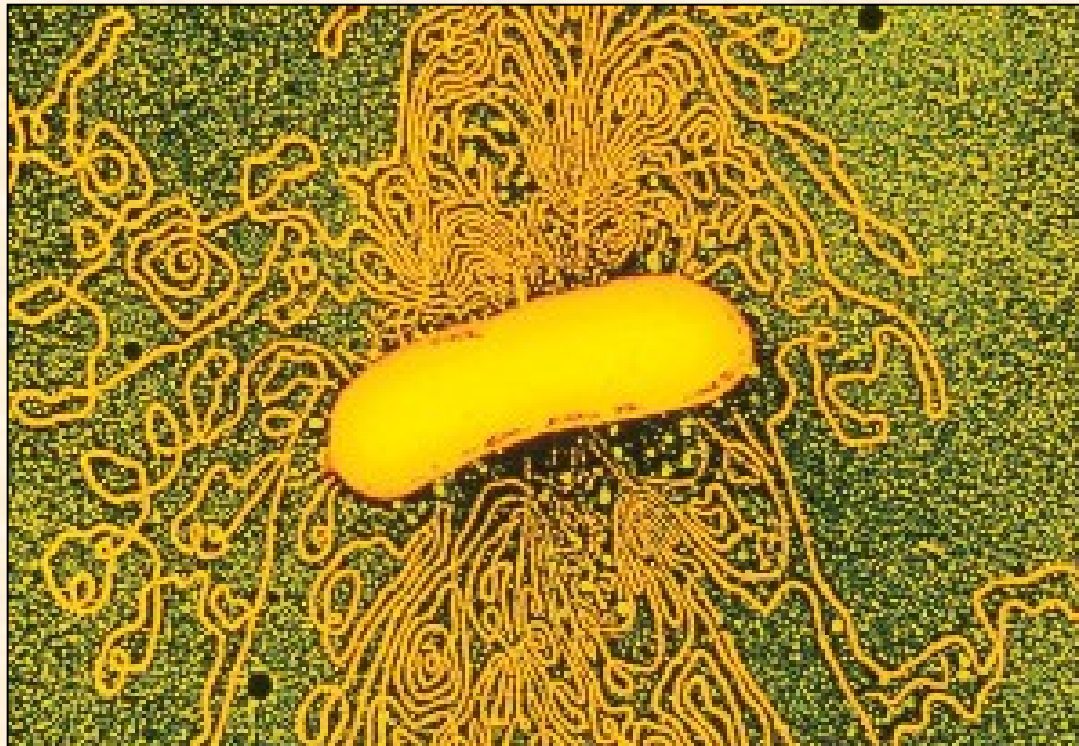
# OBSAH

1. Velikost genomu
2. Extrémní genomy
3. Změny ve velikosti genomu
4. Nekódující DNA a velikost genomu
5. Topografie genomu
6. Genomové projekty

# VELIKOST GENOMU

# Genomy bakterií a virů

Genom = celková  
genetická informace  
buňky nebo viru



## Bakterie:

~ 0.5 - 9 mil bp ~ tisíce genů

- genomy malé, kompaktní
- efektivní využití

## Virus (bakteriofág)

~ 50 kb ~ stovky genů

10 000x menší  
než zrnko soli

V průběhu evoluce se genomy zvětšují

# Obří viry - megaviry, mimiviry

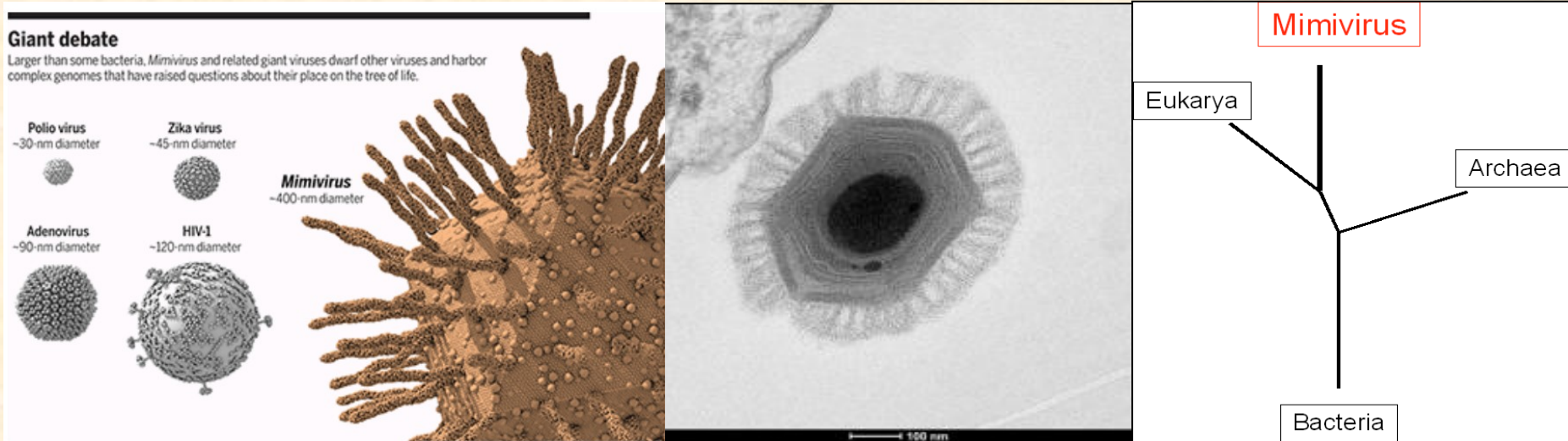
Velikost **genomu** srovnatelná s bakteriemi - *Pandoravirus salinus* (genom 2.5Mbp)

Velikost **kapsidy** srovnatelná s bakteriemi - *Pithovirus sibericum* (velký 1.5 $\mu$ m)

- izolovány 1992 z améby (*Acanthamoeba*)
- kódují ribozómy a kapsidy
- metabolické geny (~2500 genů pro proteiny)
- 10% repetitivní DNA
- jen částečná závislost na hostiteli (proteosyntéza)

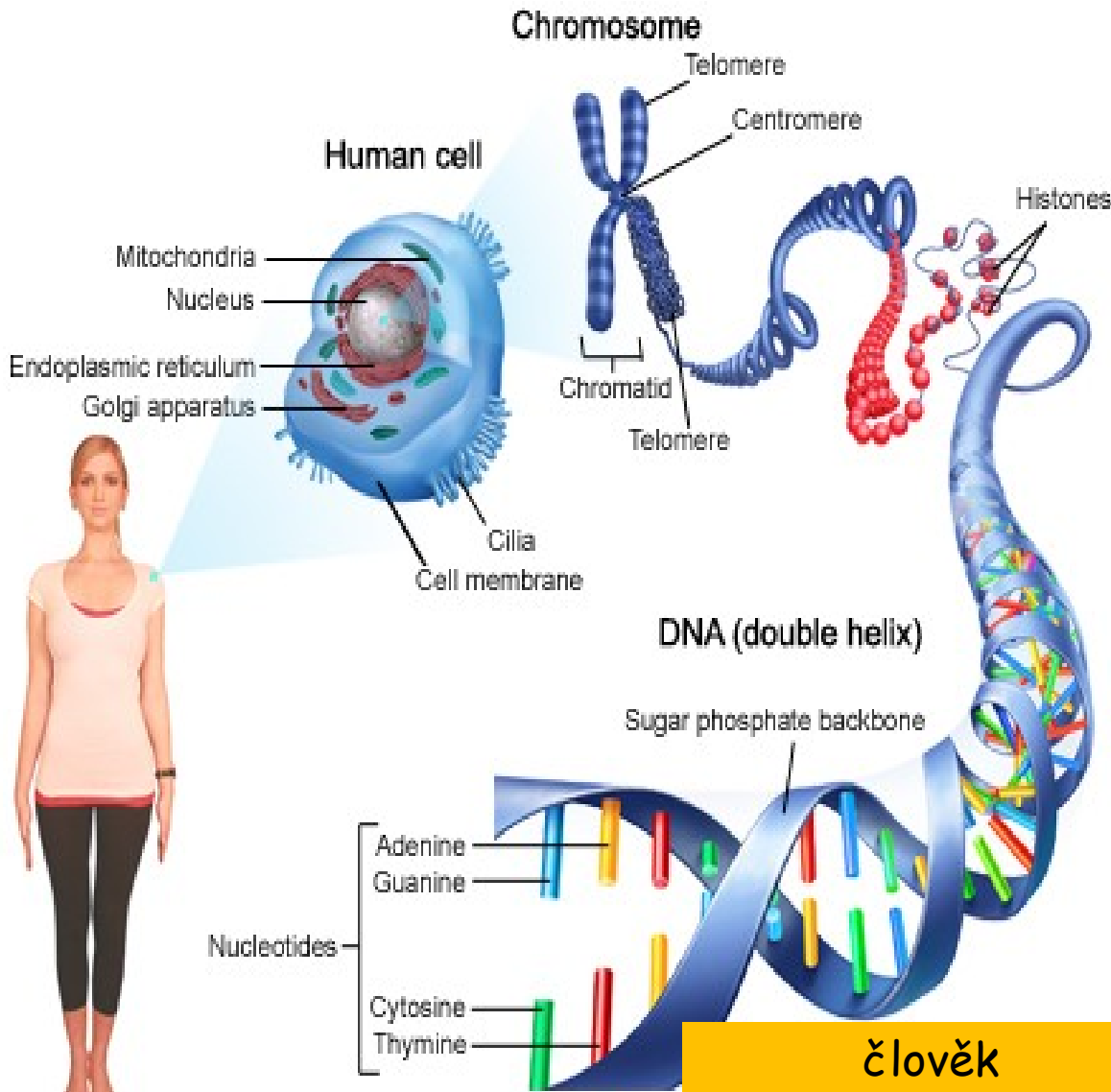
**Virofágy:** parazitují na megavirech, *Sputnik* 18kb/21 genů

*Co bylo dříve - buňka nebo mimivirus?*

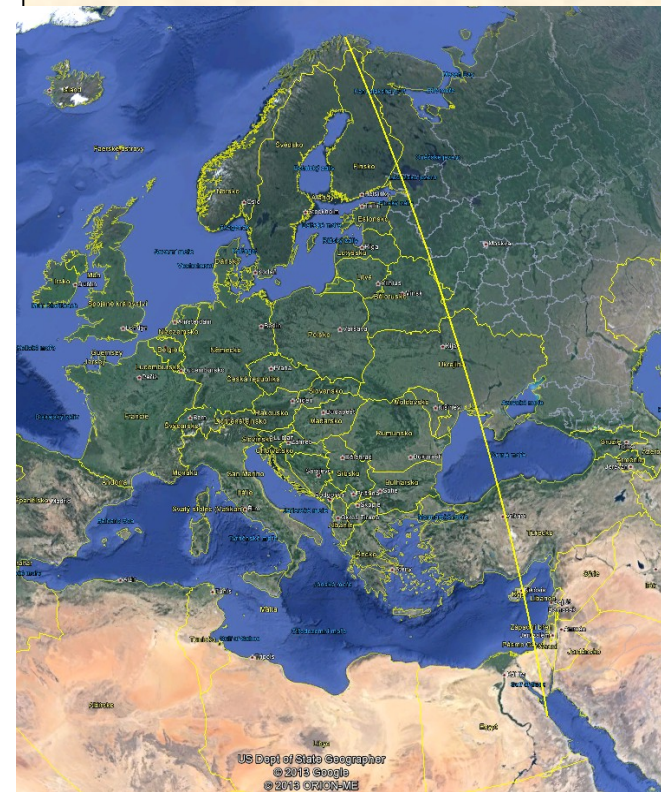




# Vyšší organizmy: kde se nachází DNA?



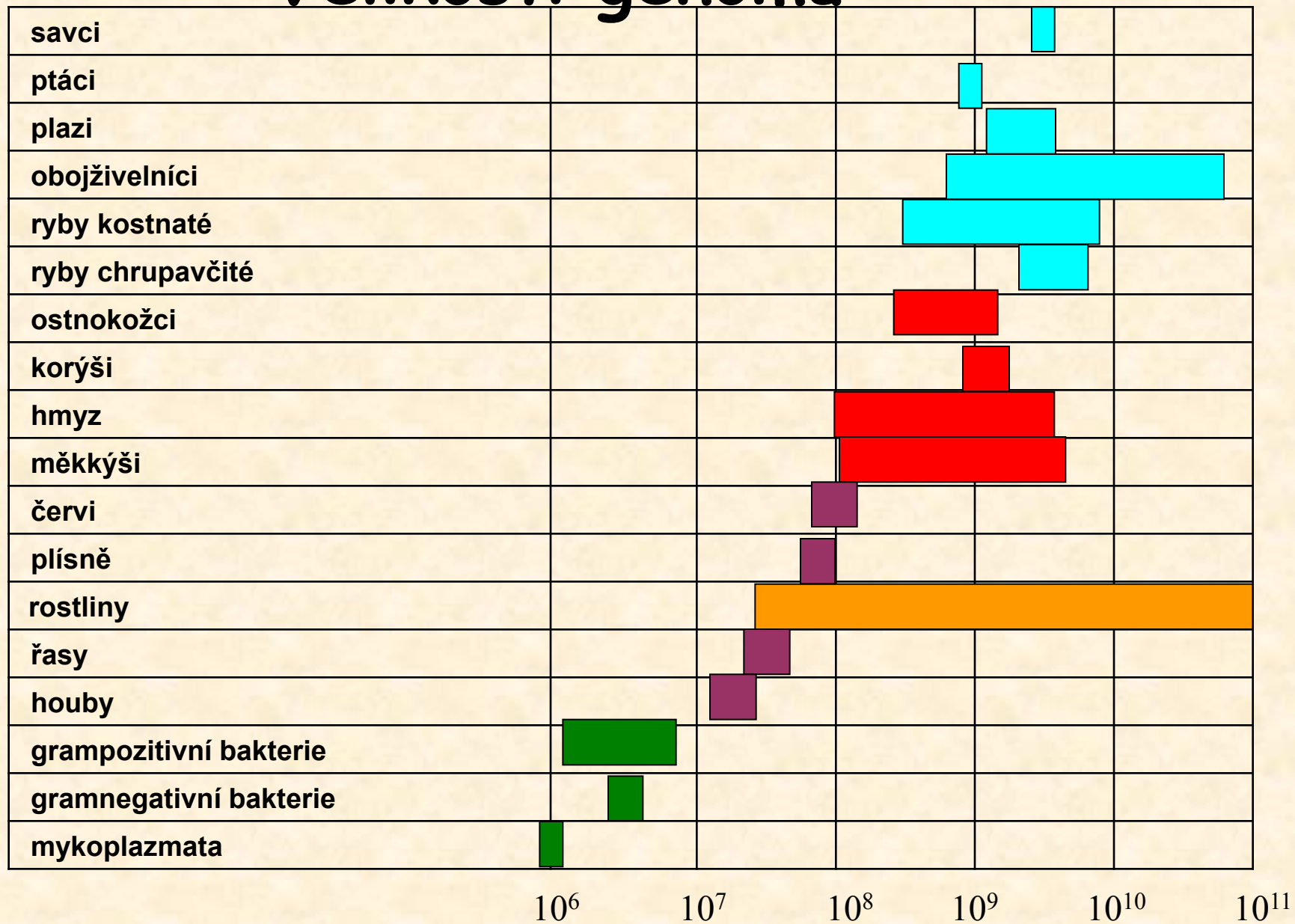
Genom = celková genetická informace buňky nebo viru



člověk  
~ 3 000 000 000 bp

2nm x 1m  
1cm -> 5000km

# Velikosti genomů



# Velikosti genomů a paradox hodnoty C



180 Mb

*Drosophila melanogaster*

100x

Klíčem k řešení paradoxu hodnoty C jsou opakující se úseky DNA (repetice)



18,000 Mb

*Podisma pedestris*

**GENOM** = celková genetická informace organismu (geny i nengenové oblasti/repetice)

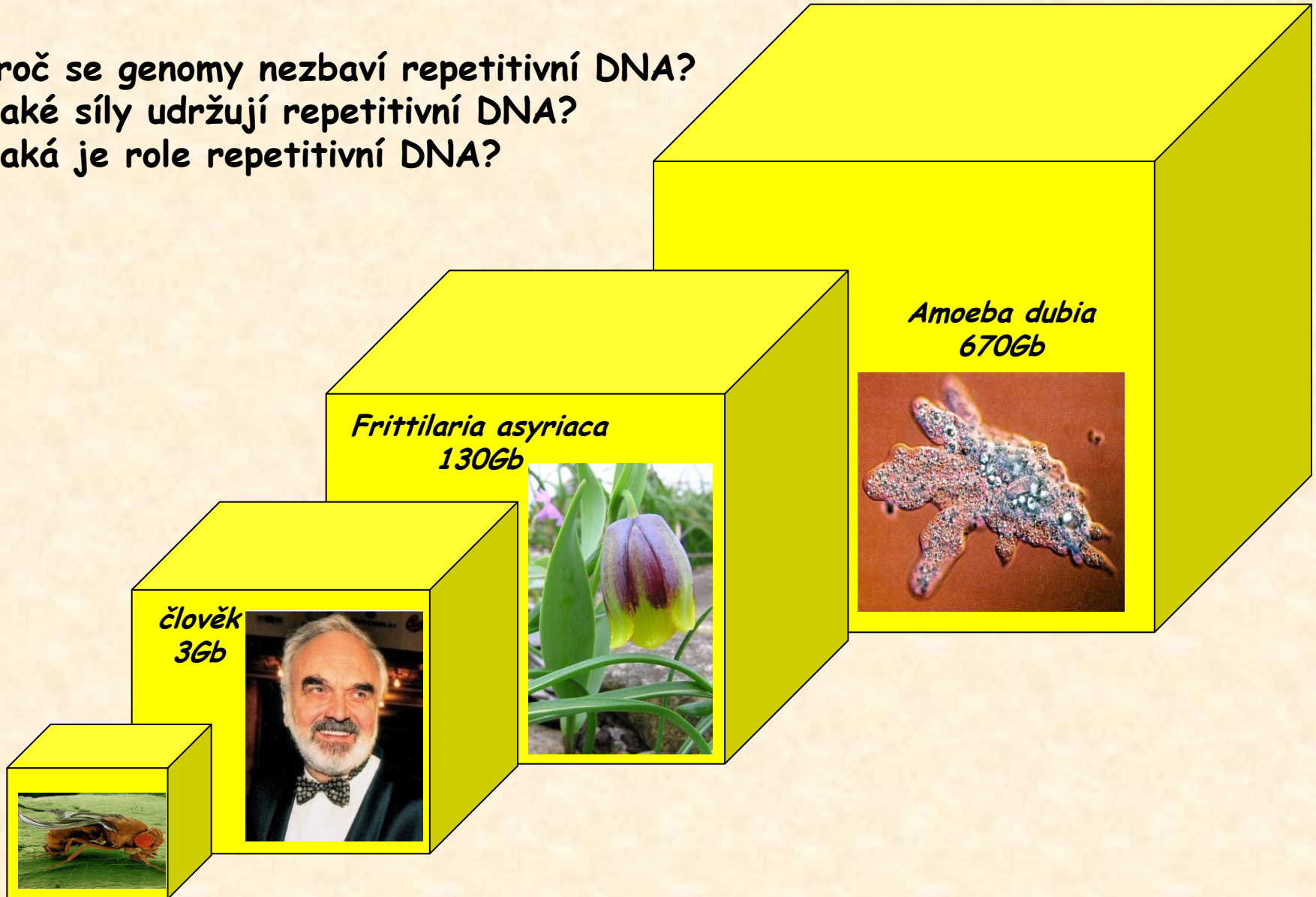
**Paradox hodnoty C** = velikost genomu organismu **není v korelaci** s komplexitou organismu např. jednobuněčná měňavka (*Amoeba dubia*) má 200x větší genom než člověk

**Genomy jsou repetitivní a dynamické**

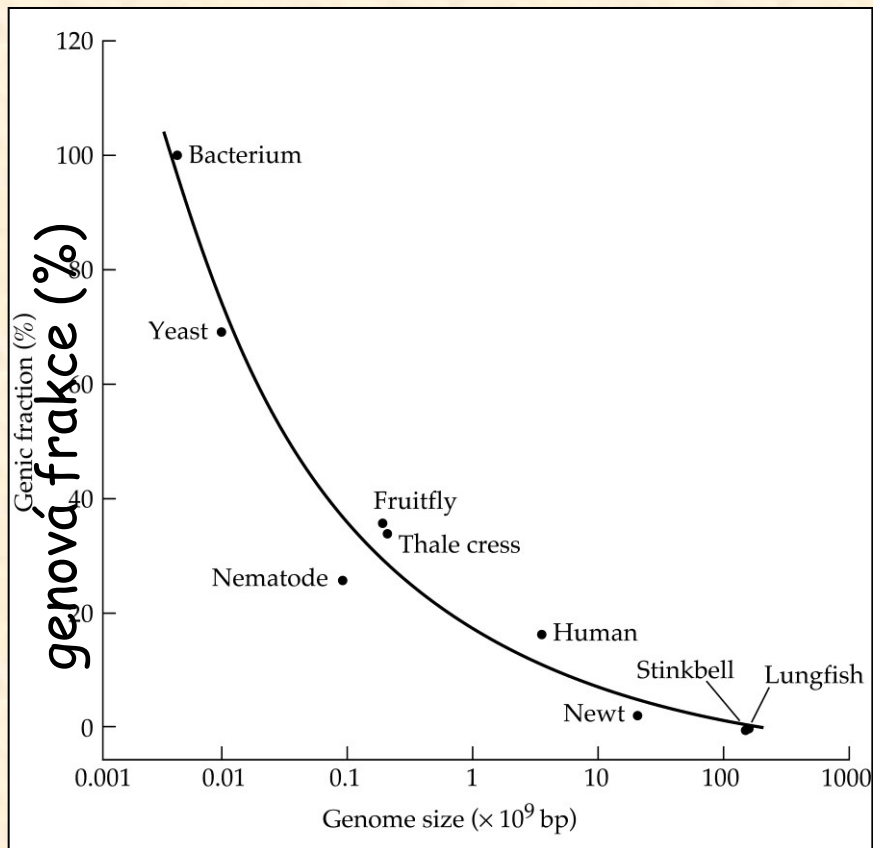


# Velikosti genomů a paradox hodnoty C

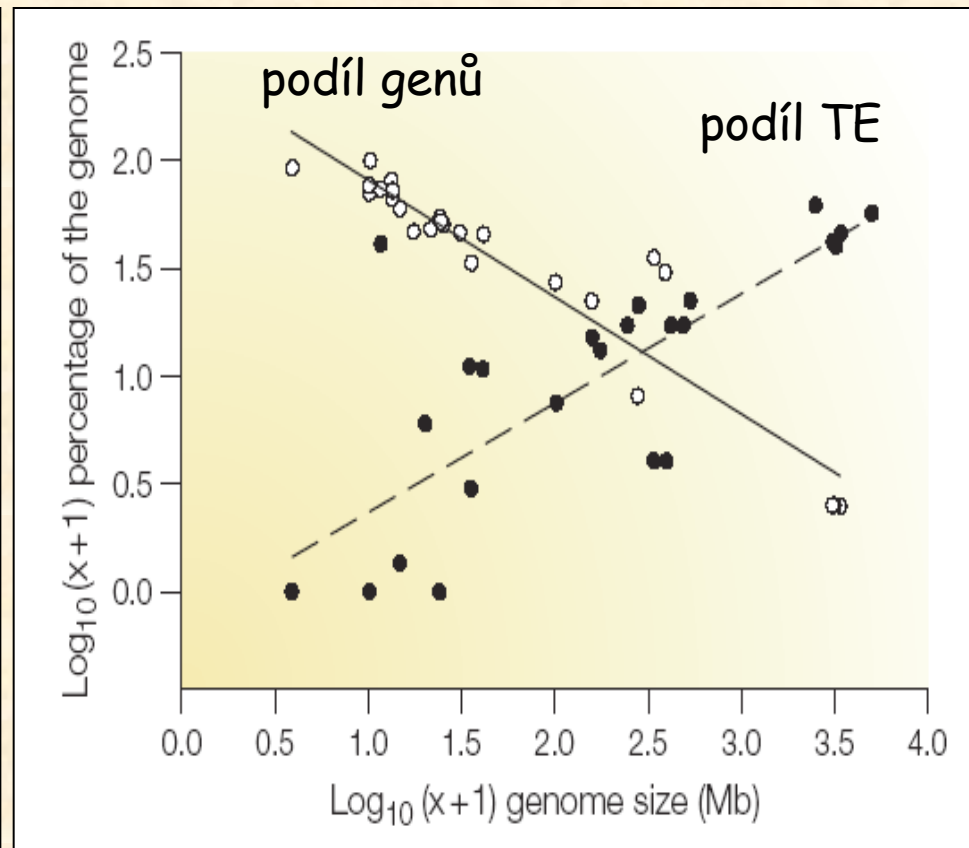
Proč se genomy nezbaví repetitivní DNA?  
Jaké síly udržují repetitivní DNA?  
Jaká je role repetitivní DNA?



# Klíč k řešení paradoxu hodnoty C se nachází v negenových oblastech



velikost genomu



velikost genomu

# Hlavní komponenty eukaryotického genomu

## Kódující části genů:

- u prokaryot tvoří většinu genomu
- u eukaryot méně, člověk 24 000 genů - 1.5%

## Introny:

- původně považovány za příčinu C-paradoxu,
- tvoří většinu genů

## Pseudogeny:

- klasické, retropseudogeny,
- 19000 člověk, 14000 myš, 51 kur, 33 kvasinka, 176 drosophila

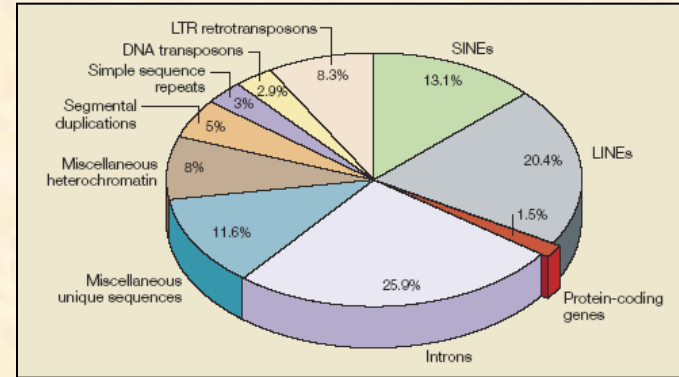
## Mobilní elementy:

- LTR, nonLTR - SINE, LINE, DNA transposony- MITE

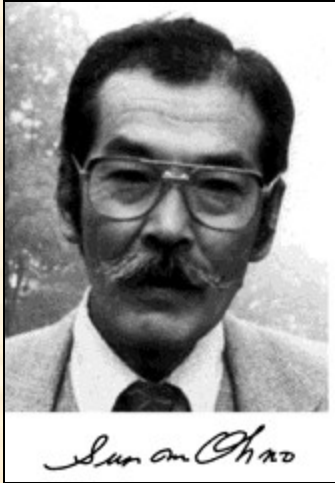
## Satelitní DNA:

## Numt, Nupt:

- inserce promiskuitní DNA



# Evolve genomů - citáty



„The majority of the genome consisted of now-extinct genes“

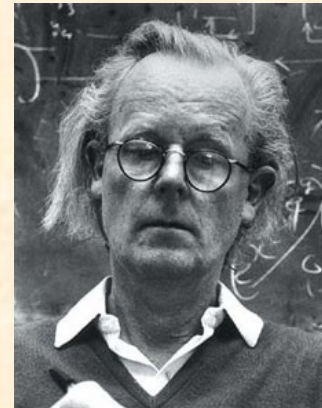
- *Susumu Ohno*

“Triumphs as well as failures of nature's past experiments appear to be contained in our genome“

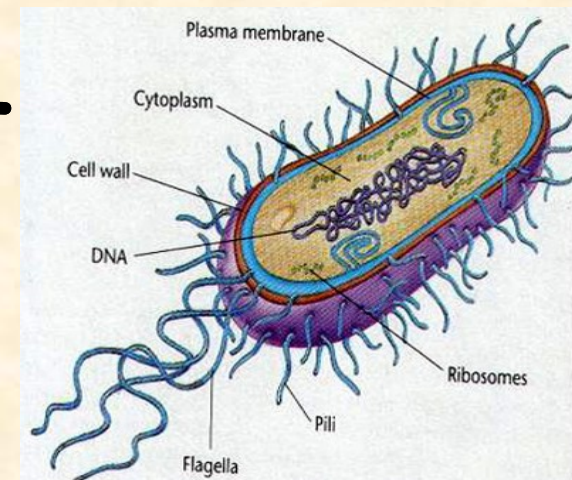
- *Susumu Ohno*

„The evolution of the large-scale features of the genome is one of the most difficult, perhaps the most difficult, question in evolutionary biology“

- *Maynard Smith*



# Velikosti genomů prokaryot



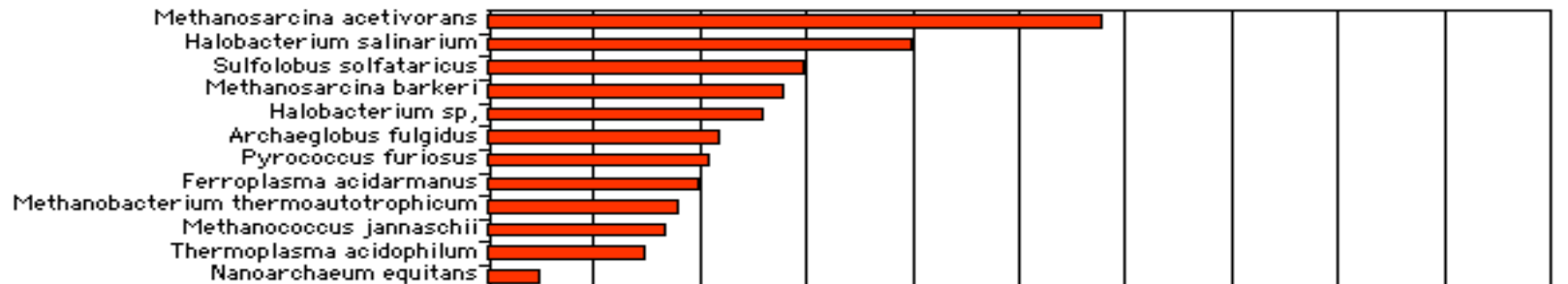
- **Bakteriální genom:  $6 \times 10^5 - > 10^7$  (=0.6-10Mb)**
  - nejmenší známý: *Mycoplasma genitalium*  
(480 genů kódujících proteiny, 3 rRNA geny, 37 tRNA genů)
- **Procesy ovlivňující velikost bakteriálního genomu:**
  - genová duplikace, malé delece a inserce, transpozice, horizontální přenos, ztráta genů v parazitických liniích, atd.

**Velikost genomů  
prokaryot je  
zhruba úměrná  
počtu genů**

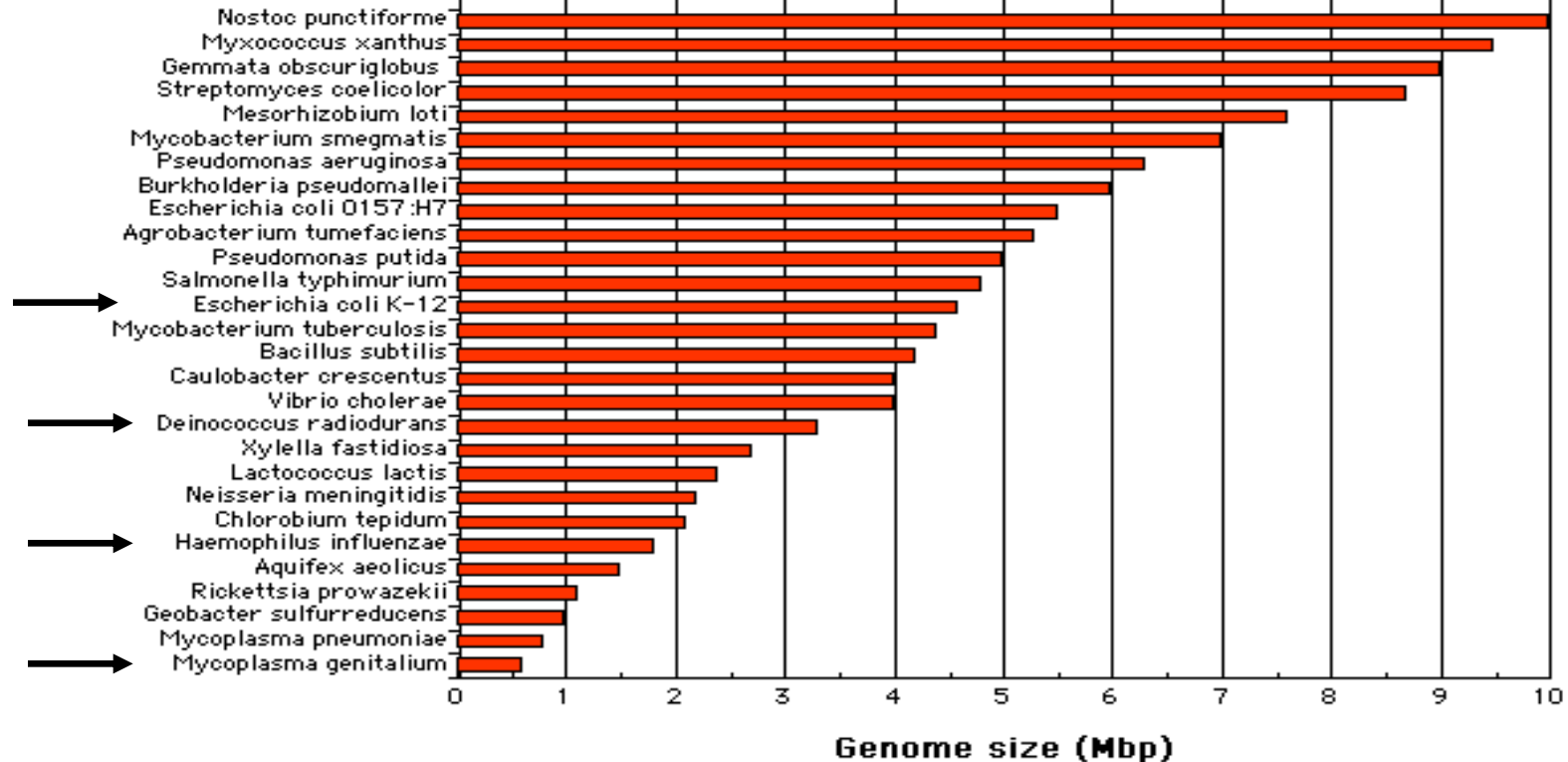


# Velikosti genomů prokaryot

## Archaea:



## Bacteria:



# EXTRÉMNÍ GENOMY

# „Životní minimum“ bakterie

## Život s pouhými 271 geny

### Strategie:

1. odstranění či inaktivace genů (top-down)
2. syntéza minimálního genomu (bottom-up)



Scientists at TIGR Uncover the

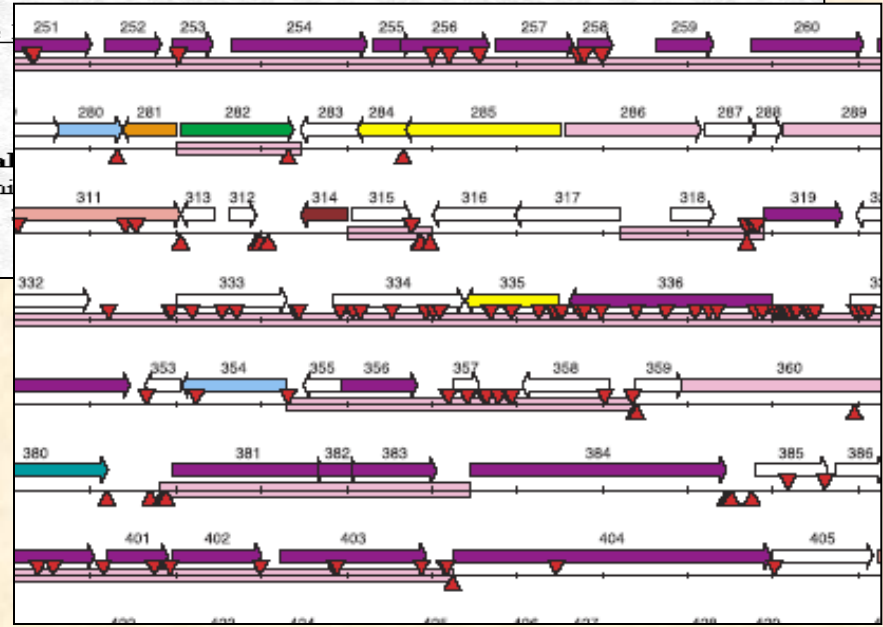
### Press Release

## Global Transposon Mutagenesis and a Minimal Mycoplasma Genome

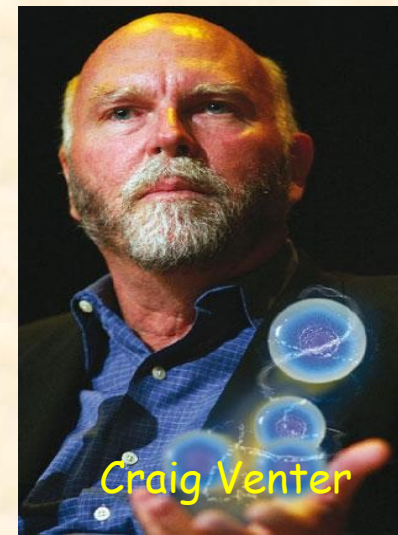
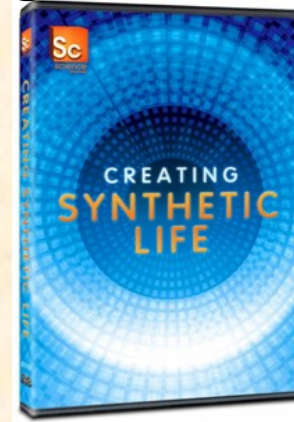
Clyde A. Hutchison III,<sup>1,2\*</sup> Scott N. Peterson,<sup>1†‡</sup> Steven R. Gill,<sup>1</sup>  
Robin T. Cline,<sup>1</sup> Owen White,<sup>1</sup> Claire M. Fraser,<sup>1</sup>  
Hamilton O. Smith,<sup>1‡</sup> J. Craig Venter<sup>1‡§</sup>

*Mycoplasma genitalium* with 517 genes has the smallest gene complement of any independently replicating cell so far identified. Global transposon mutagenesis was used to identify nonessential genes in an effort to learn whether the naturally occurring gene complement is a true minimal genome under laboratory growth conditions. The positions of 2209 transposon insertions in the completely sequenced genomes of *M. genitalium* and its close relative *M. pneumoniae* were determined by sequencing across the junction of the transposon and the genomic DNA. These junctions defined 1354 distinct sites of insertion that were not lethal. The analysis suggests that 265 to 350 of the 480 protein-coding genes of *M. genitalium* are essential under laboratory growth conditions, including about 100 genes of unknown function.

Global Hutchison C.M.,

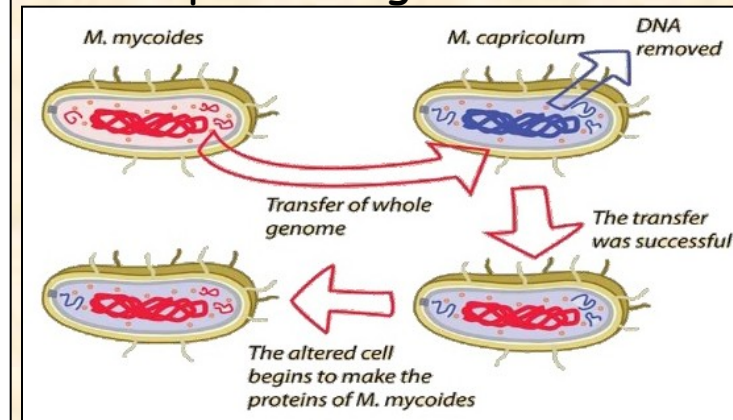


# Syntetický život: Pouhých 271 genů



Craig Venter

## Transplantace genomu:



*Genomy umíme nejen číst ale také psát*

## *Mycoplasma laboratorium* (Synthia)

- nahradili genom *M. capricolum* uměle nasyntetizovanou DNA *M. mycoides*
- nový software přebudoval původní hardware, vodoznaky v umělé DNA



# Největší prokaryotické genomy

## **Pseudomonas aeruginosa (bakterie):**

- 5500 genů
- přes 6 Mb
- přirozeně rezistentní k antibiotikům (ochranný obal)
- R-faktor, žije ve společenství jiných bakterií, konjugace
- lidský patogen (kožní n., močové, dýchací a trávicí cesty)



## **Nostoc punctiforme (sinice):**

- 7432 ORF
- 8.9 Mb
- **repetice**, transpozony
- fotoautotrofní, také fakultativně heterotrofní
- možnost symbiózy s rostlinami i houbami

Repetice <i>N.p.</i>	Sites in genome
AATGAC <sub>n</sub> (STRR2)	
AATGACA	69
AATGACT	63
AATGACC	39
AATCCC (STRR4)	
AATCCC	41
AATGCC	37
AATTACG (STRR5)	45
AdTCCCC (STRR1)	
ATTCCCC	39
AATCCCC	19
AGTCCCC	15
AGCAGGGG (STRR6)	29
AAAATTC (STTR7)	13



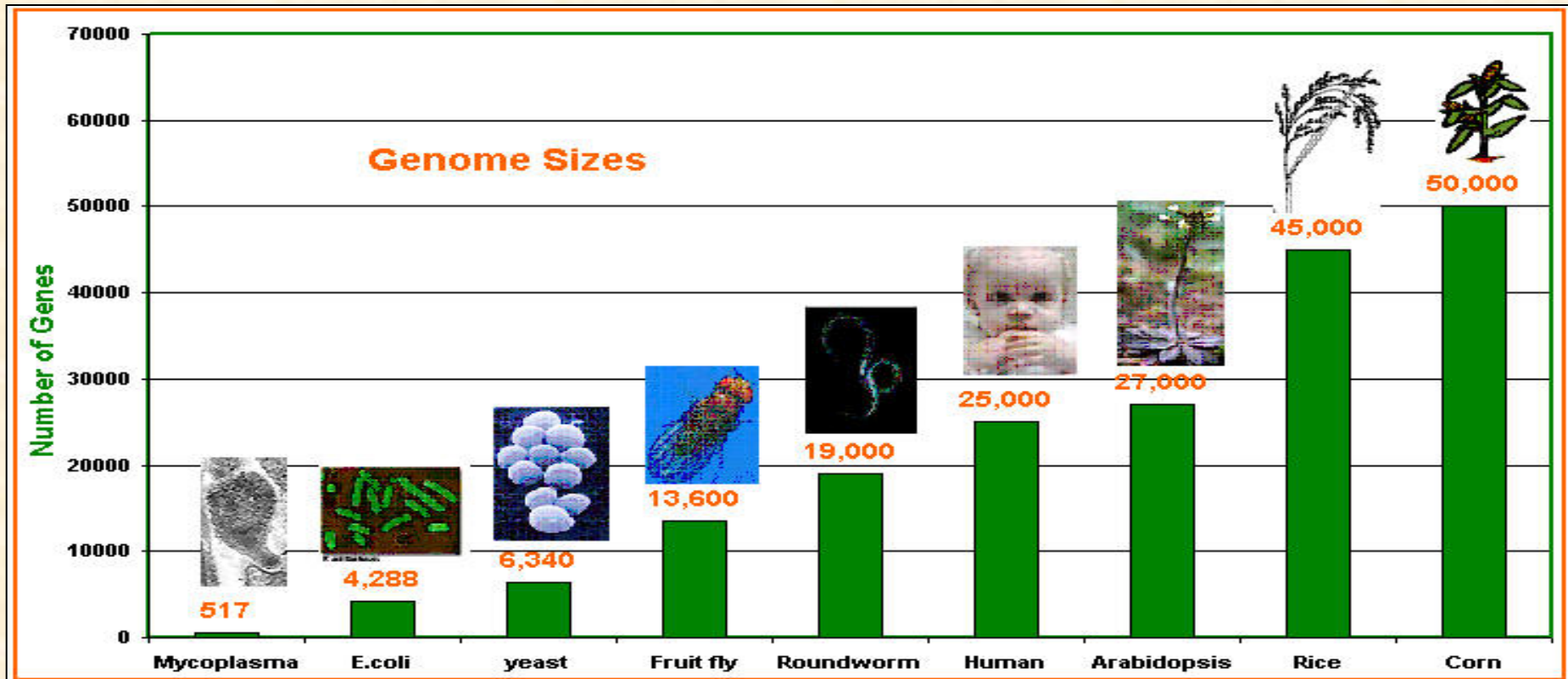
# Velikosti genomů eukaryot

Eukaryotický genom:  $8.8 \times 10^6 \sim 6.9 \times 10^{11}$   
(8.8Mb - 670Gb)

Nejmenší známý: *Saccharomyces cerevisiae*

Největší známý: *Amoeba dubia*

Velikosti eukaryotických genomů NEjsou úměrné počtu genů nebo komplexitě organismu

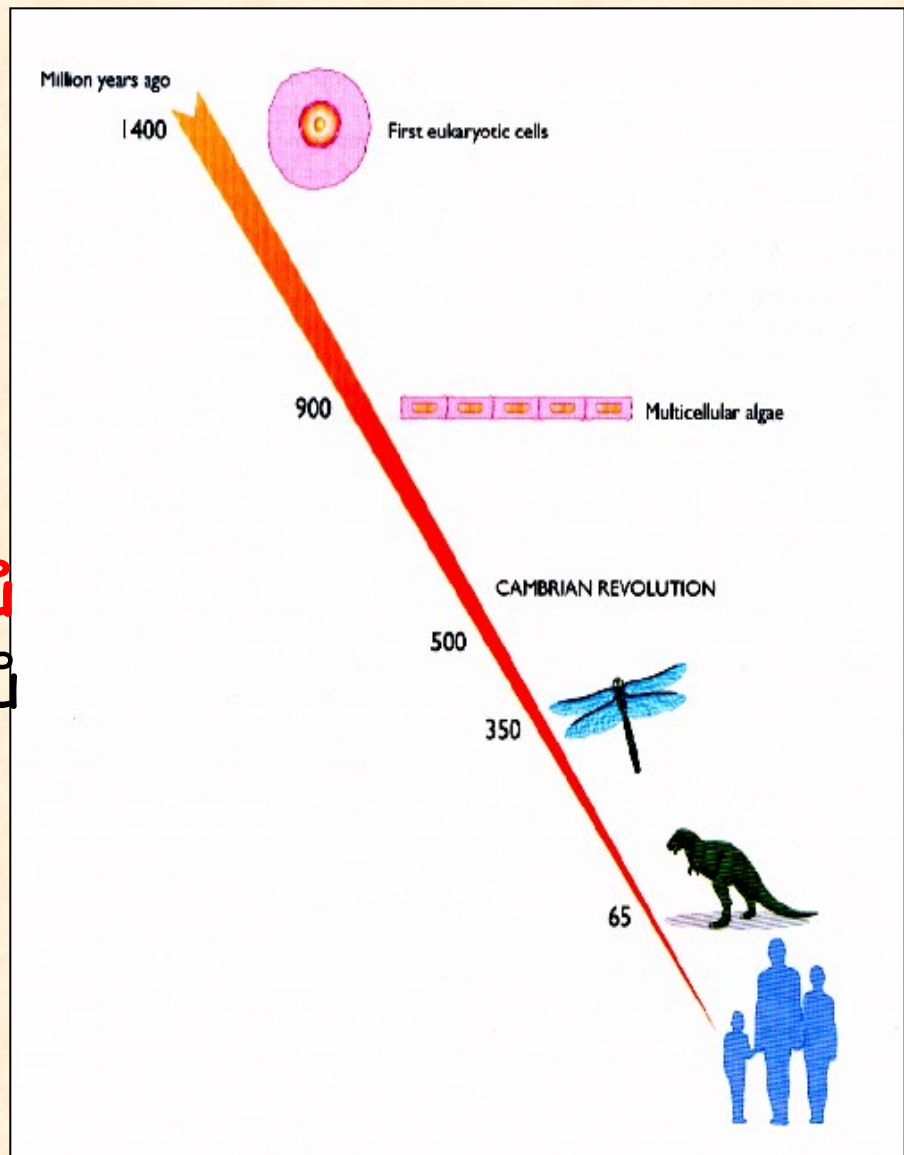


Organism	Genome size (base pairs)	Note
Virus, Bacteriophage MS2	3,569	First sequenced RNA-genome <sup>[2]</sup>
Virus, SV40	5,224	<sup>[3]</sup>
Virus, Phage $\Phi$ -X174;	5,386	First sequenced DNA-genome <sup>[4]</sup>
Virus, Phage $\lambda$	50,000	
Bacterium, <i>Haemophilus influenzae</i>	1,830,000	First genome of living organism, July 1995 <sup>[5]</sup>
Bacterium, <i>Carsonella ruddii</i>	160,000	Smallest non-viral genome. <sup>[6]</sup>
Bacterium, <i>Buchnera aphidicola</i>	600,000	
Bacterium, <i>Wigglesworthia glossinidia</i>	700,000	
Bacterium, <i>Escherichia coli</i>	4,000,000	<sup>[7]</sup>
Amoeba, <i>Amoeba dubia</i>	670,000,000,000	Largest known genome. <sup>[8]</sup>
Plant, <i>Arabidopsis thaliana</i>	157,000,000	First plant genome sequenced, Dec 2000. <sup>[9]</sup>
Plant, <i>Genlisea margaretae</i>	63,400,000	Smallest recorded <a href="#">flowering plant</a> genome, 2006. <sup>[9]</sup>
Plant, <i>Fritillaria assyrica</i>	130,000,000,000	
Plant, <i>Populus trichocarpa</i>	480,000,000	First tree genome, Sept 2006
Yeast, <i>Saccharomyces cerevisiae</i>	20,000,000	<sup>[10]</sup>
Fungus, <i>Aspergillus nidulans</i>	30,000,000	
Nematode, <i>Caenorhabditis elegans</i>	98,000,000	First multicellular animal genome, December 1998 <sup>[11]</sup>
Insect, <i>Drosophila melanogaster</i> aka Fruit Fly	130,000,000	<sup>[12]</sup>
Insect, <i>Bombyx mori</i> aka Silk Moth	530,000,000	
Insect, <i>Apis mellifera</i> aka Honey Bee	1,770,000,000	
Fish, <i>Tetraodon nigroviridis</i> , type of Puffer fish	385,000,000	Smallest vertebrate genome known
Mammal, <i>Homo sapiens</i>	3,200,000,000	
Fish, <i>Protopterus aethiopicus</i> aka <a href="#">Marbled lungfish</a>	130,000,000,000	Largest vertebrate genome known

# ZMĚNY VE VELIKOSTI GENOMU

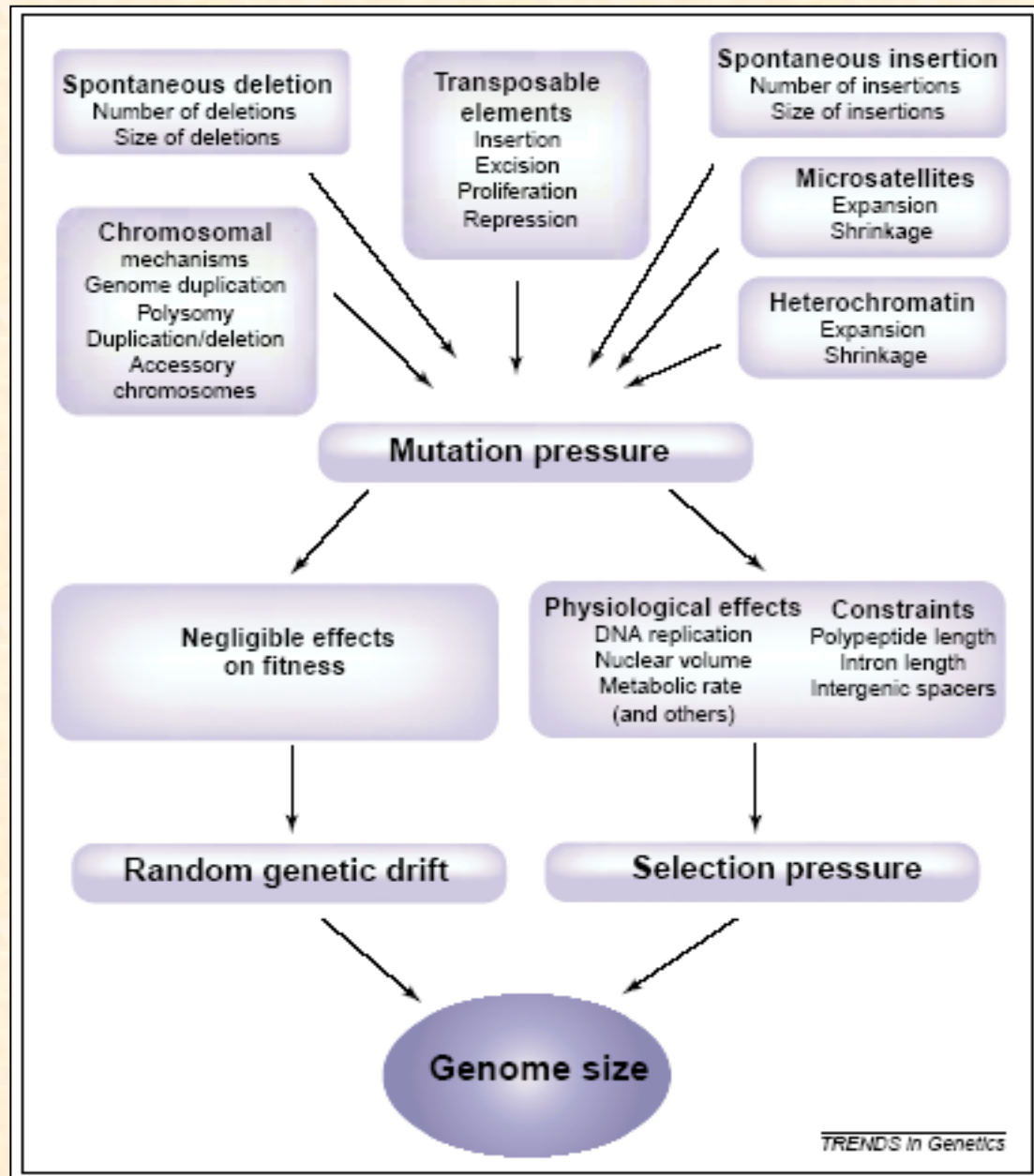
# Zvětšení genomu

- celkové zvětšení: polyploidizace (duplikace celého **genomu**)
- duplikace části genomu, zmnožení počtu **chromosomů**
- duplikace **genů** a skupin genů
  
- amplifikace **transpozonů**
- inserce **virové DNA**
- inserce **organelové DNA**
- expanze (mikro)**satelitů**



Vzrůstající komplexita živých forem byla doprovázena vzrůstem velikosti genomů a počtu genů

# Evoluční síly ovlivňující velikost genomu





# Změny ve velikosti genomů: plynulé nebo skoky?

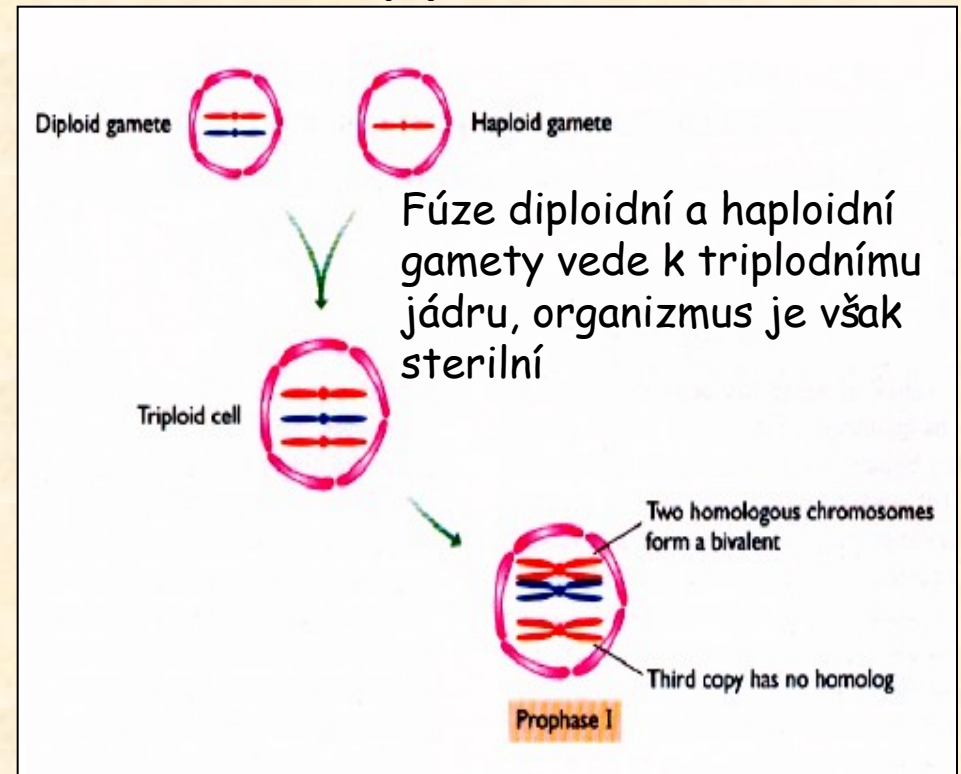
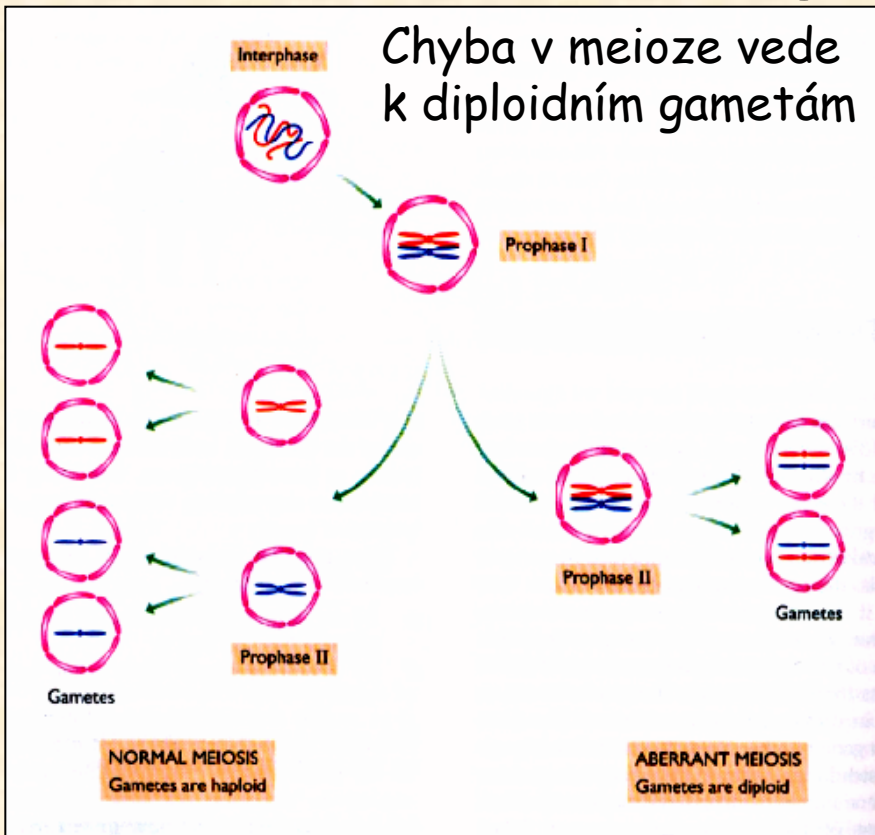
## Plynulost:

- savci, ptáci, kostnaté ryby
- postupná akumulace nebo delece, malé segmenty DNA
- normální distribuce velikostí genomů v těchto skupinách
- podíl jedinečných sekvencí konstantní, mění se spolu se změnou velikosti genomu, je dán velikostí intronů

## Skoky (kvantové změny):

- bezobratlí, rostliny (50% nahosemenných je polyploidní), vyjimečně obratlovci
- polyplodizace, přídavná replikace některého chromosomu, kryptopolyploidizace (koordinovaná replikace všech repetitivních sekvencí v genomu - *Chironomus*)
- kvantové rozdíly mezi blízkými příbuznými, velké rozdíly ve velikostech genomů

# Růst velikosti genomu: Polyploidizace



**Autopolyploidie:** multiplikace jedné základní sady chromosomů, jeden organizmus, chyba při meioze (růže...)

**Allopolyploidie:** kombinace geneticky odlišných sad chromosomů (pšenice...)

**Kryptopolyploidie:** dávná polyploidizace, organizmus se jeví jako diploidní, přestavby - delece, translokace..., kvasínka, obratlovci, obilniny, odhalení až sekvenací

# Polyploidní plodiny

- tetraploidi: kukuřice, bavlna, brambor, zelí



- hexaploidi: pšenice, chrysanéma

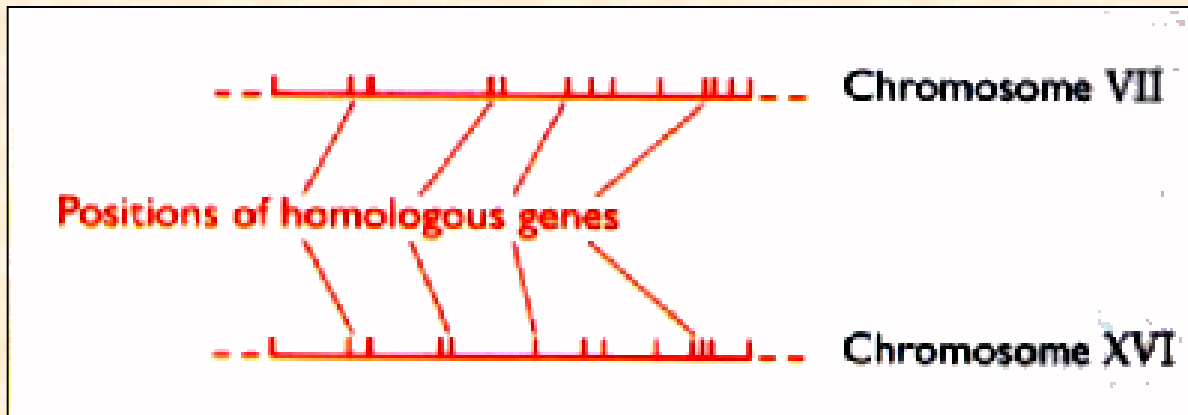


- oktoploidi: jahodník



# Genom kvasinky: Polyploidizace nebo duplikace segmentů?

- 54 nepřekrývajících se duplikovaných oblastí
- 50 z nich si zachovalo stejnou orientaci vzhledem k centroméře
- nebyla nalezena žádná triplikovaná oblast, přičemž 7 oblastí by bylo očekáváno dle Poissonova rozložení
- příbuzné druhy bez duplikace - *Kluveromyces waltii* a *Ashbya gossypii*
- 457 genových párů - 17% se vyvíjí rychleji než jejich homology u *K. waltii*



Duplikace jsou výsledkem tetraploidie spíše než postupných regionálních duplikací

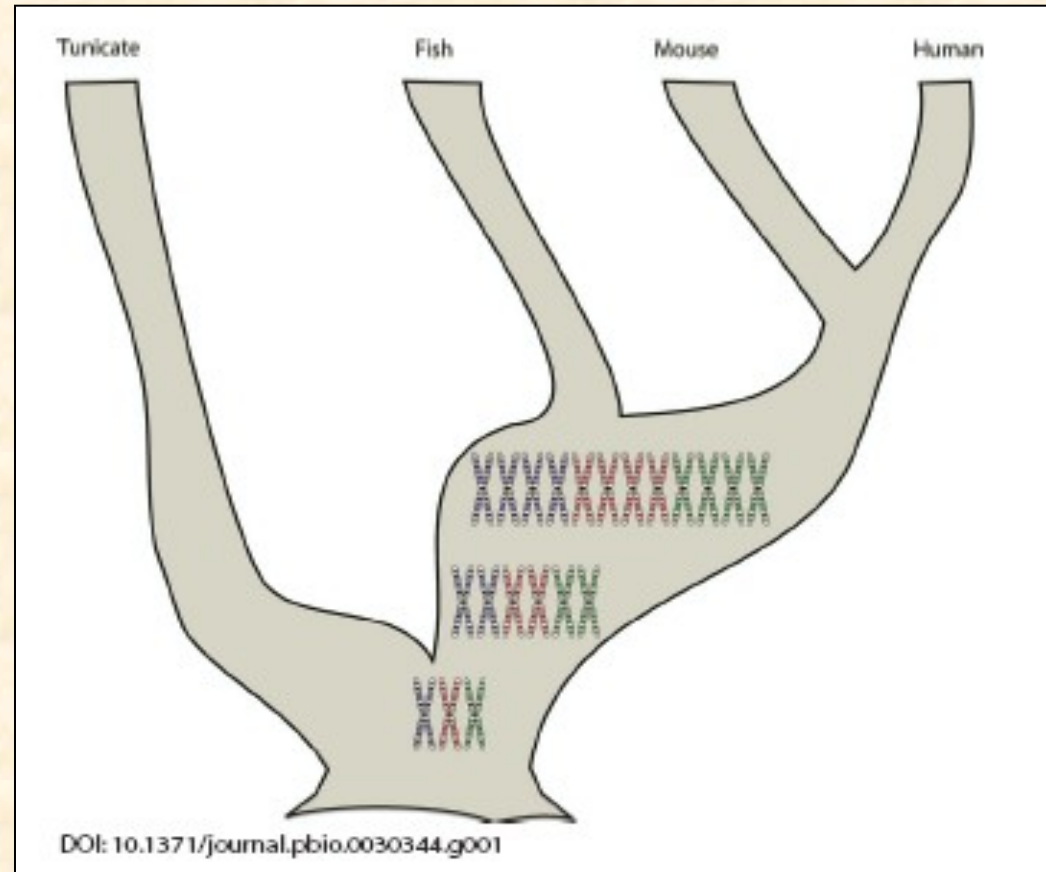


# Polyploidizace v linii obratlovců

S. Ohno:

- dvě genomové duplikace u obratlovců - hypotéza „2R“ (2 rounds)
- některé genové rodiny mají 4 členy u obratlovců a jednu u bezobratlých
- srovnání ježovky (bezobratlý) s obratlovcem (člověkem, myší a rybou Fugu)
- v genomu člověka dlouhé segmenty ve čtyřech kopiích podél všech chromosomů

Duplikace genomu před 450 mil let napomohla vzrůstu komplexity a diverzifikaci obratlovců

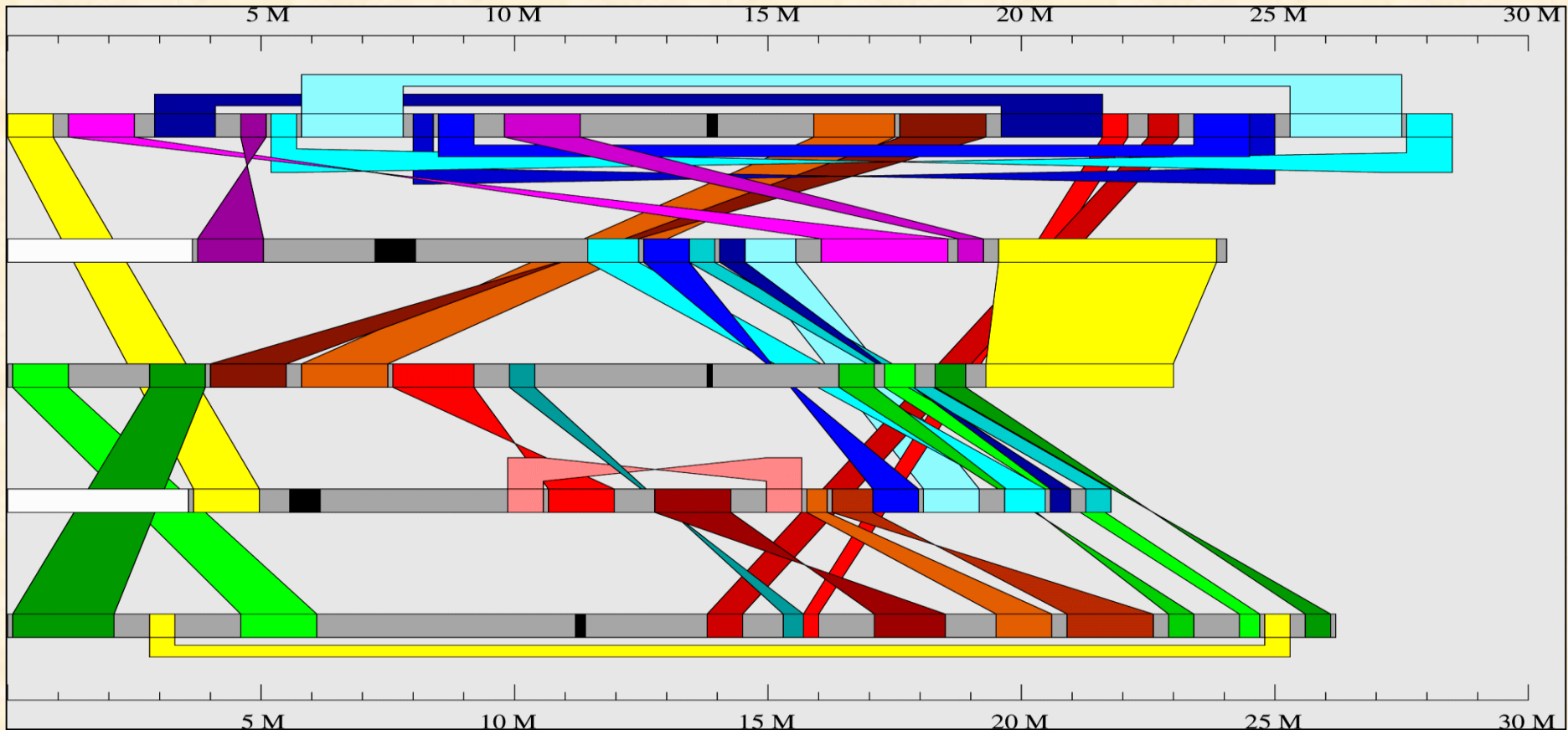




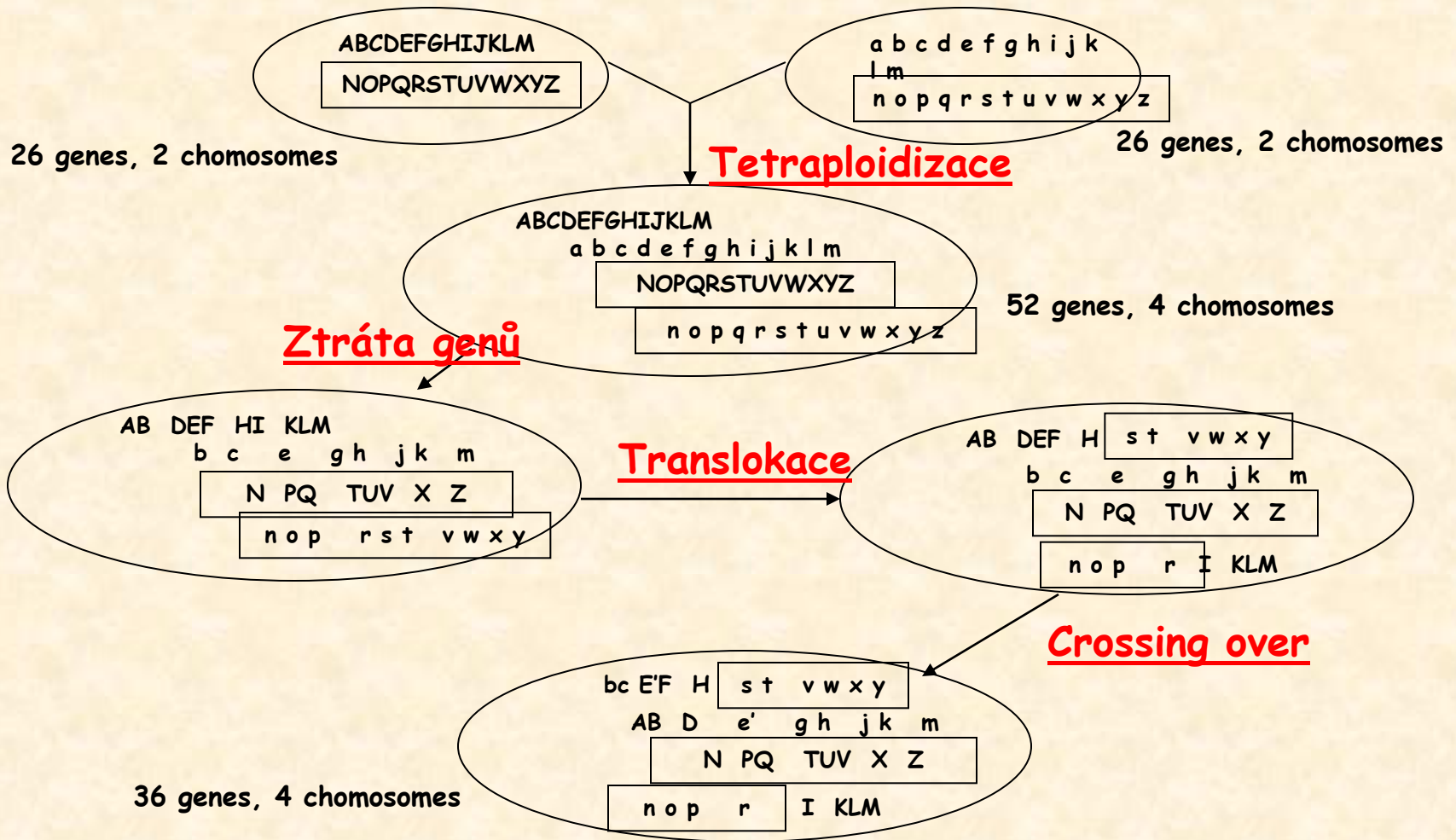
# Polyploidizace u rostlin

## Arabidopsis:

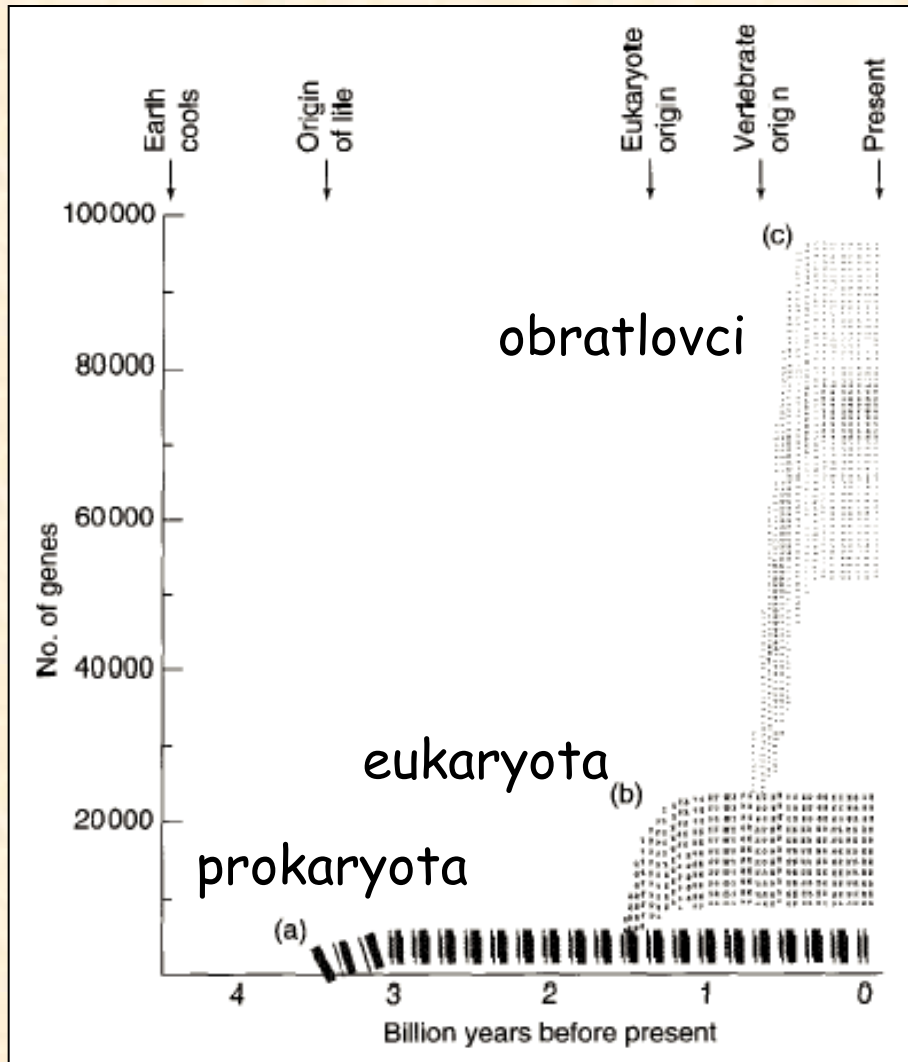
- duplikace před 200 a 80 mil. let
- možná i více než 4 genomové duplikace
- asi 60% genomu leží v duplikovaných segmentech
- 50% genů v nich je konzervováno



# Počet genů a evoluce duplikovaného genomu



# Vzrůst počtu genů, potlačení šumu a biologická komplexita



## Prokaryota/eukaryota:

Potlačení šumu - **separace** transkripce a translace, jaderná membrána a histony

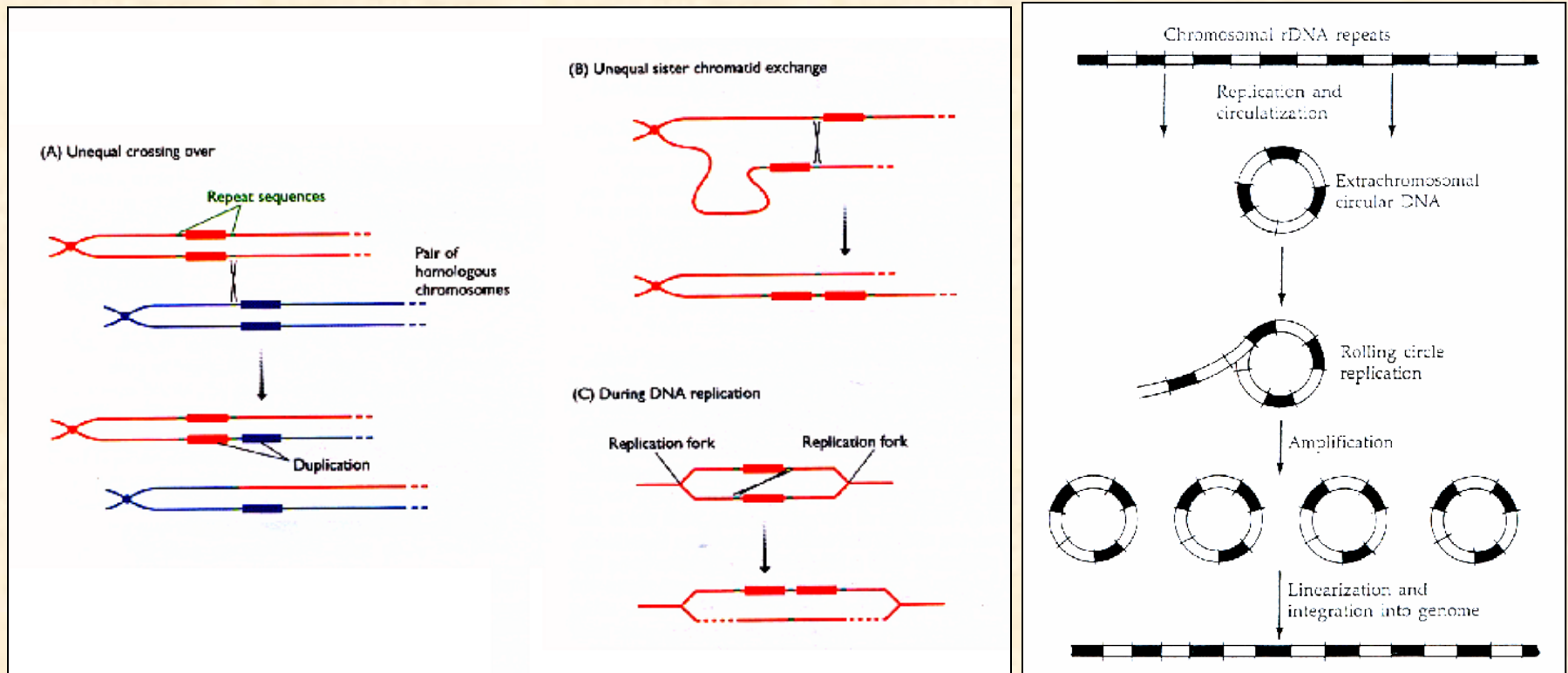
## Obratlovci:

Potlačení šumu - **metylace** genů, mobilních elementů a duplikátů

**RNA interference**  
(smallRNAs)

# Mechanismy zvětšení genomů

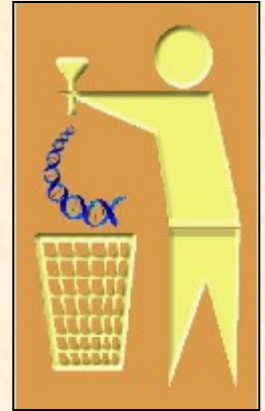
- Duplikativní (retro)transpozice
- Nerovnoměrný (unequal) crossing-over
- Replikační klouzání (Replication slippage)
- Genová amplifikace (rolling circle replication)



# NEKÓDUJÍCÍ DNA a VELIKOST GENOMU



# Teorie úlohy nekódující DNA: adaptivní role vs. sobecká DNA



- Jaké evoluční síly produkují „zbytečnou“ DNA?
- Jaká je role „zbytečné“ DNA?
- Proč selekce toleruje „zbytečnou“ DNA?

## Adaptivní role:

- vliv nadbytečné DNA na fenotyp
- vliv na velikost jádra a buňky
- ochrana kódujících sekvencí před mutacemi
- pufrování koncentrace regulačních proteinů

## Sobecká DNA (junk DNA):

- jsou to jen parazitické sekvence, mobilní elementy, fixace genetickým driftem, velikost genomu je tolerovatelné maximum závisující na ekologických a vývojových potřebách organismu

# Genomová obezita u rostlin - jednosměrný proces?

Teorie sobecké DNA (Dawkins, 1976)  
- jediným cílem DNA je se replikovat

## Dva modely:

- pouze růst genomů
- růst i zmenšování genomů

## Limity růstu genomů:

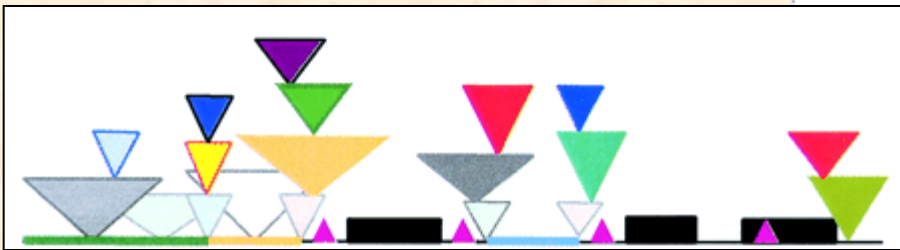
- **fitness** hostitele
- mechanismy potlačující amplifikaci - **metylace**
- mechanismy odstraňování repeticí - **rekombinace**
  - BARE - 42x více soloLTR než vnitřních částí (dávná aktivita)
  - rekombinace uvnitř elementů i mezi elementy
  - selekce na kratší LTR, solo LTR nejdelší



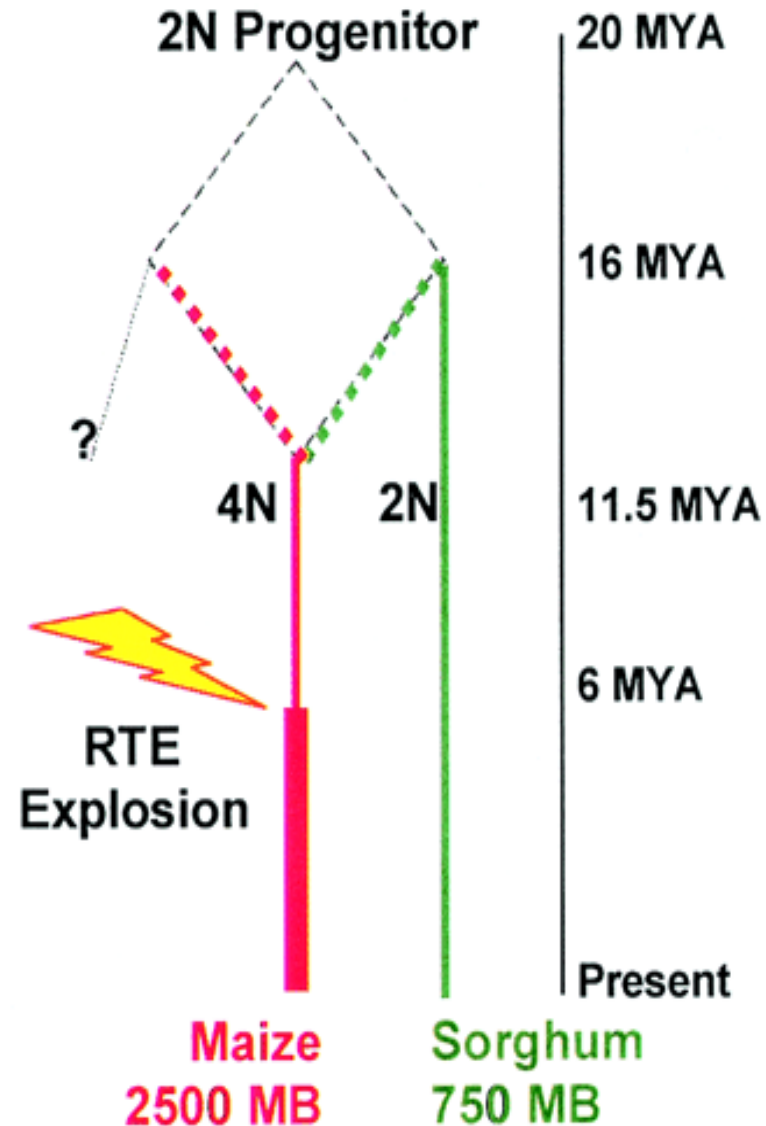
Illustration by James Collins

# Big Bang v genomu kukuřice

- polyploidizace
- amplifikace TE (6 mil let)
- inserce do mezigenových oblastí



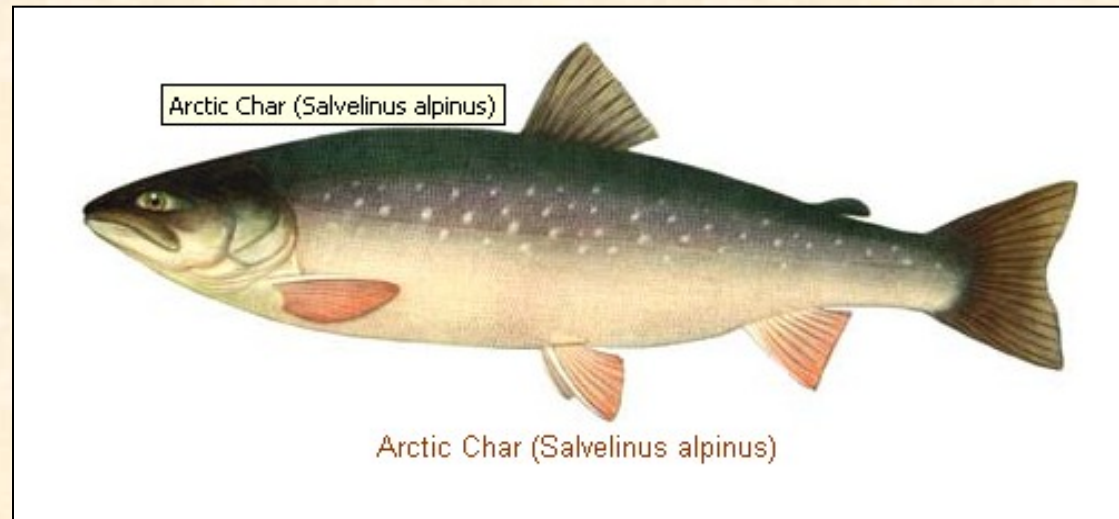
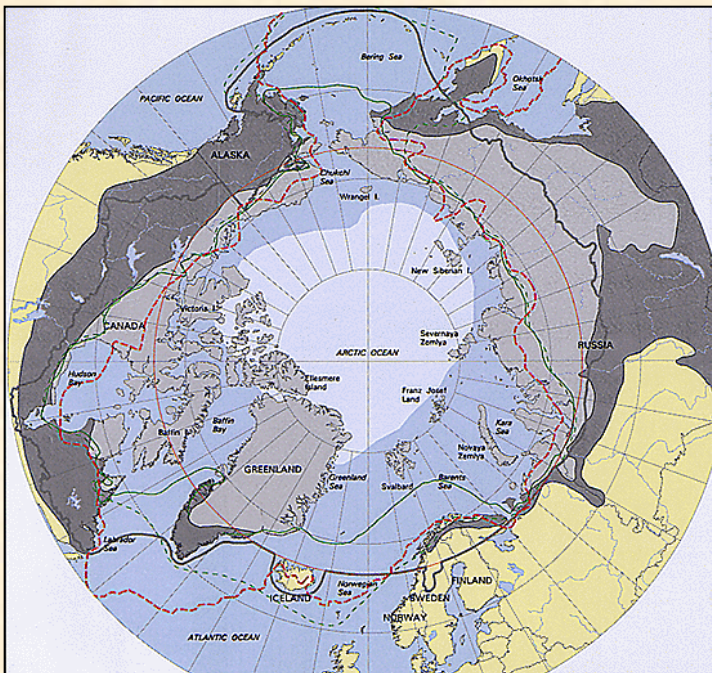
Commentary PNAS 98, 8163-8164 (2001)  
**Gene galaxies in the maize genome**  
Virginia Walbot\* and Dmitri A. Petrov  
Department of Biological Sciences, 385 Serra Mall, Stanford University, Stanford, CA 94305-5020





# Závislost velikosti genomu na zeměpisné šířce a nadmořské výšce

- korelace mezi velikostí genomu a **teplotním režimem**
- větší genomy nebo **polyploidii**:
  - **arktické** lososovité ryby
  - zooplankton arktických jezer (*Daphnia*, *Bosmina*)
  - rostliny v polárních oblastech
  - populace v teplých oblastech jsou diploidní



# Obsah DNA je proměnlivý i v rámci jedince

## Zvýšení obsahu DNA - endopolyploidie a polytenie:

- drosophila - polytenní chromosomy ve slinných žlázách
- *Daphnia* - tkáňově-specifické rozdíly v ploidii, 2-2048C, vliv na morfologii hlavy indukovanou predátorem
- *Bombyx mori* - 1 000 000 ploidní buňky žláz - hedvábí
- korelace ploidie a velikosti buňky

## Snížení obsahu DNA v somatických buňkách:

- nematoda, bičíkovci, dvoukřídlí,
- např. *Cyclops strenuus* - somatické b. mají 5% obsahu DNA oproti oplozenému vajíčku
- vysvětlení z hlediska **sobecké DNA** - delece DNA ze slepé somatické linie a ne ze zárodečné linie
- změna obsahu DNA během života jako odpověď na stimuly
- Lamarckismus



*Daphnia pulex*



*Bombyx mori*



*Cyclops strenuus*



# Vliv velikosti genomu na fenotyp

Velikost genomu koreluje s:

+

- velikostí jádra
- velikostí buňky (nucleotypic effect)
- dobou mitózy a meiózy
- minimální generační dobou
- velikostí semen
- odpovědí letniček vůči  $CO_2$
- dobou vývoje embrya u mloků



-

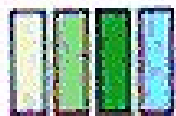
- rychlostí bazálního metabolismu u obratlovců (negativní korelace)  
(malý genom ptáků a netopýrů - rychlý metabolismus při letu,  
velký genom ryb - estivace za hypoxických podmínek)
- morfologickou komplexitou mozků u žab a mloků (negativní korelace)

# TOPOGRAFIE GENOMU

# Uspořádání genů v genomech



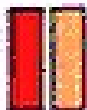
## KEY



Tandem repeats



Centromere-associated tandem repeats



Telomeric and sub-telomeric repeats



Dispersed *Ty1/copia* retroelements and microsatellites

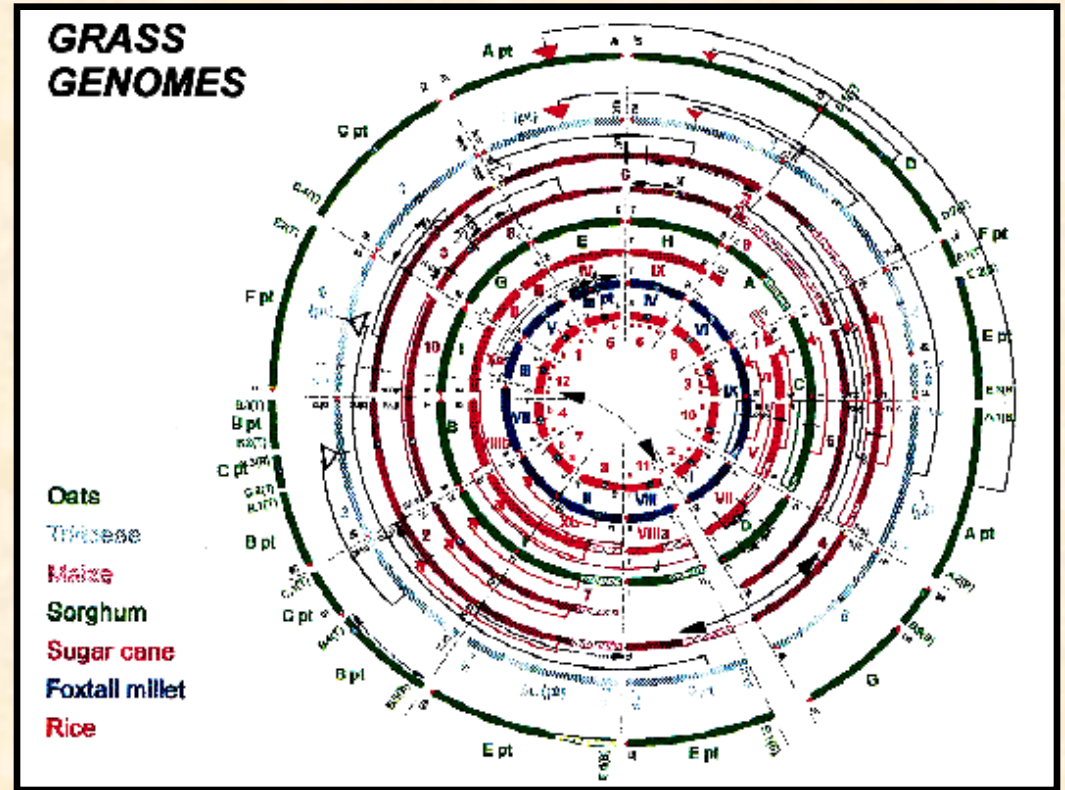
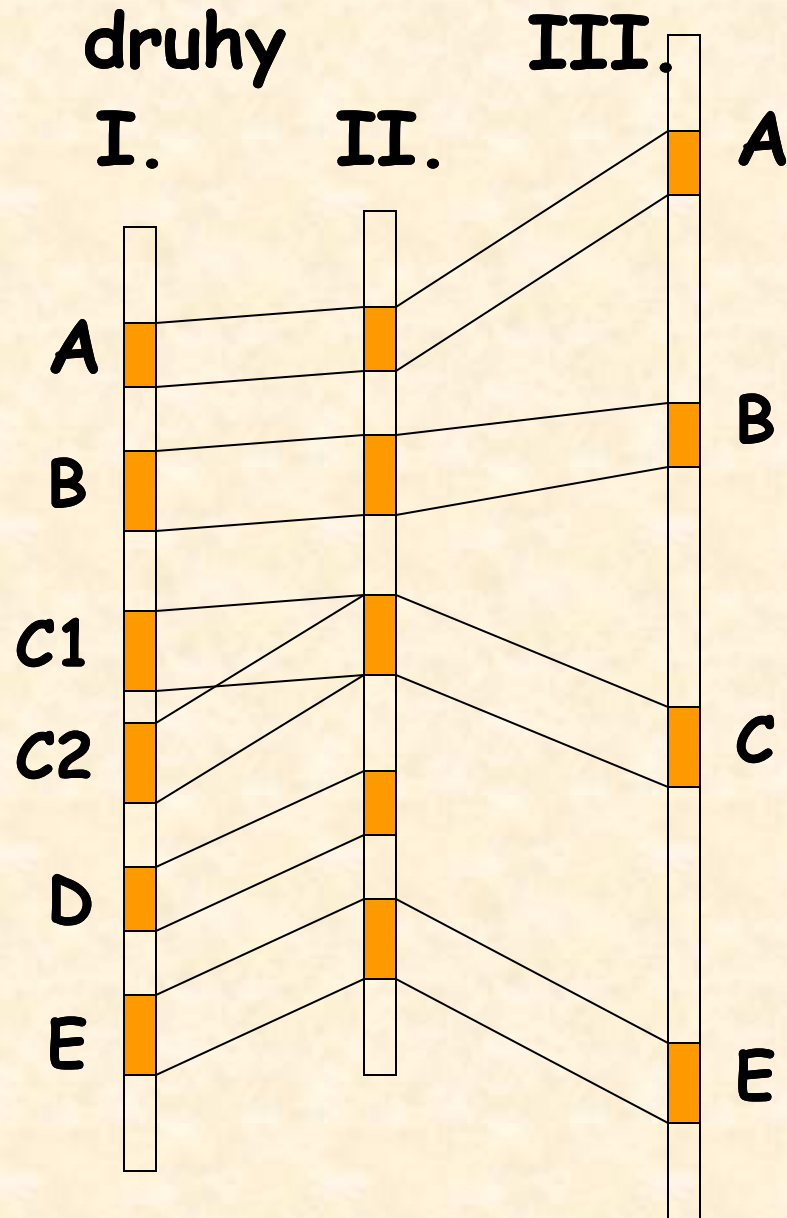


LINEs



Single and low-copy sequences including genes

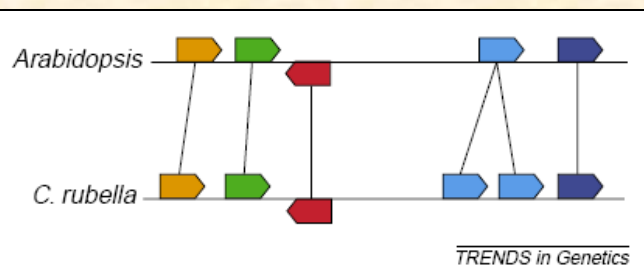
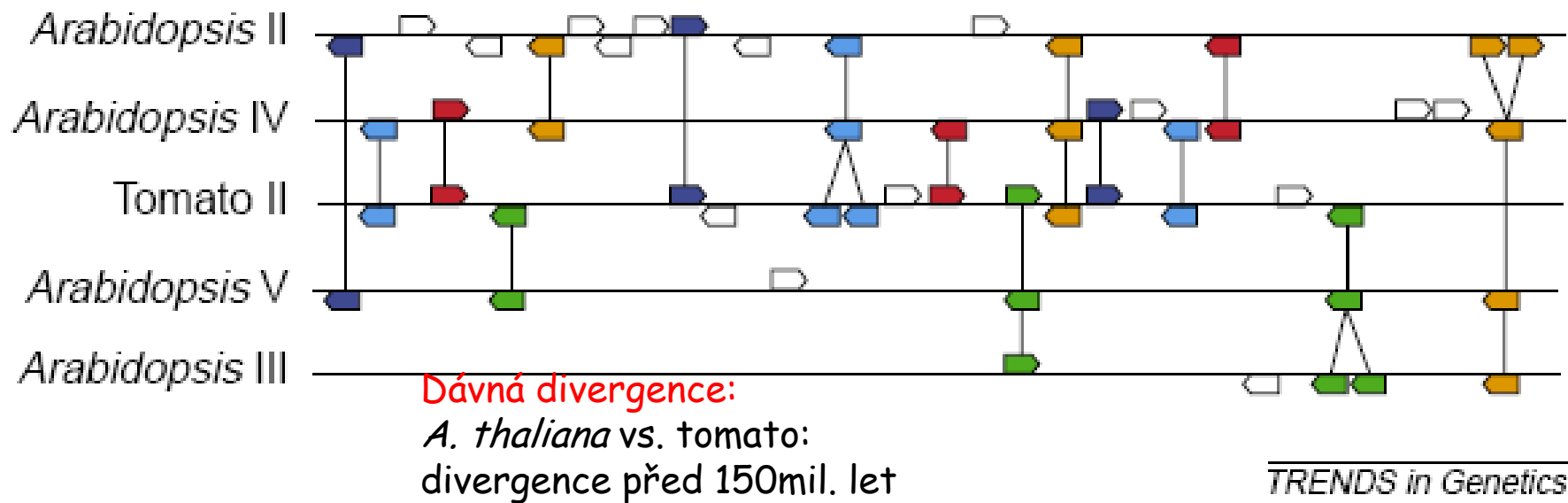
# Syntenie - konzervativita pořadí genů



- konzervativita genů
- rozdílné mezigenové sekvence

# Makro- vs. mikro- kolinearita

- konzervativita v dlouhých úsecích
- mikrostruktura více dynamická, mix s jinými geny
- polyploidie urychlí evoluci mikrostruktury
- změny již u prvních generací syntetického alotetraploida u Brassica



**Nedávňá divergence:**  
*A. thaliana* vs. *Capsella rubella*:  
divergence před 6.2-9.8 mil. let

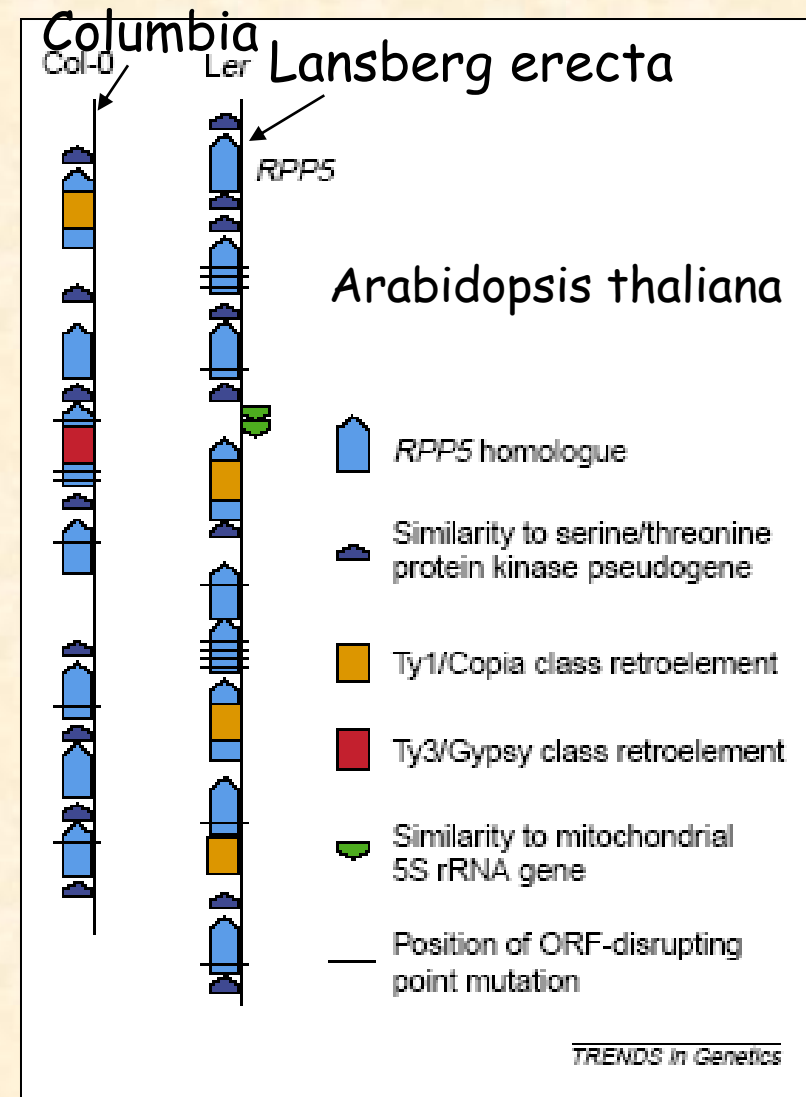
(Bancroft, 2001)



# Rychlá evoluce oblasti rezistence k patogenu

(Bancroft, 2001)

- gen „RPP5“ - rezistence k *Perenospora parasitica*
- rozdíly mezi A.t. **ekotypy** *Lansberg erecta* a *Columbia*
- většinou pseudogeny
- tandemové duplikace a nerovnoměrný crossing-over je mechanismem rychlé evoluce



# Počty chromosomů u různých druhů

Common Name	Species	Diploid number	Common Name	Species	Diploid number
<b>Animals (2n)</b>			<b>Plants (2n)</b>		
Human	<i>Homo sapiens</i>	46	Corn	<i>Zea mays</i>	20
Monkey	<i>Macaca mulatta</i>	42	Potato	<i>S. tuberosum</i>	48
Dog	<i>Canis familiaris</i>	78	Green algae	<i>A. mediterranea</i>	20
Cat	<i>Felis domesticus</i>	38			
Mouse	<i>Mus musculus</i>	40	<b>Fungi (2n)</b>		
Frog	<i>Rana pipiens</i>	26	Yeast	<i>S. cerevisiae</i>	32
Fruit fly	<i>Drosophila melanogaster</i>	8	<b>Fungi (1n)</b>		<b>Haploid number</b>
Flatworm	<i>Planaria torva</i>	16	Mold	<i>Penicillium species</i>	4

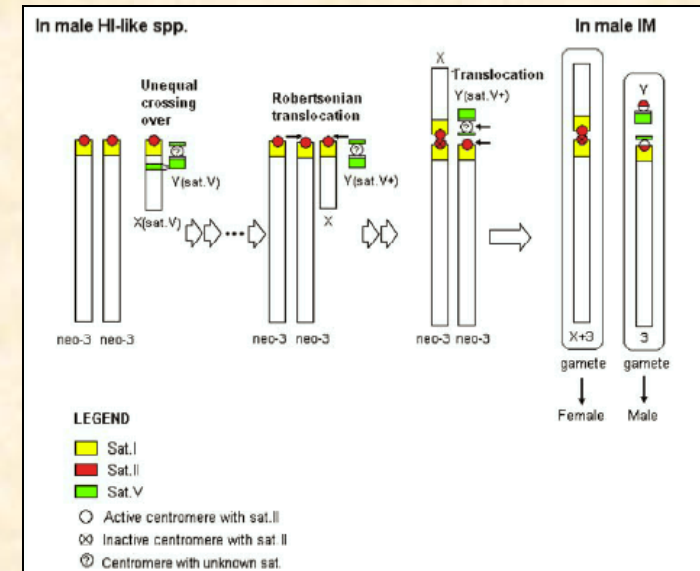
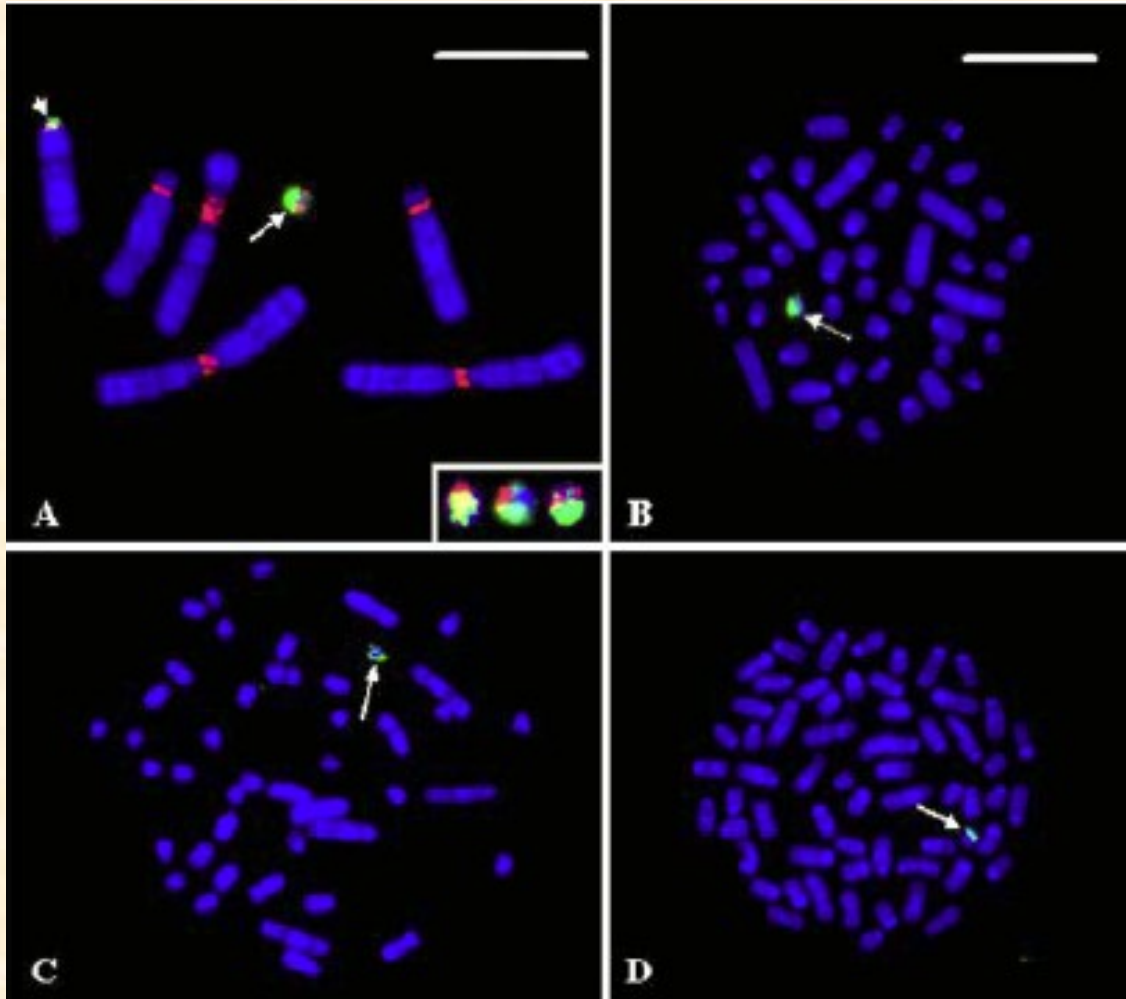
## Extrémisté:

mravenec *Myrmecia pilosula* - 1 pár, samec 1 chromosom  
 kapradina *Ophioglossum reticulatum* - 630 párů

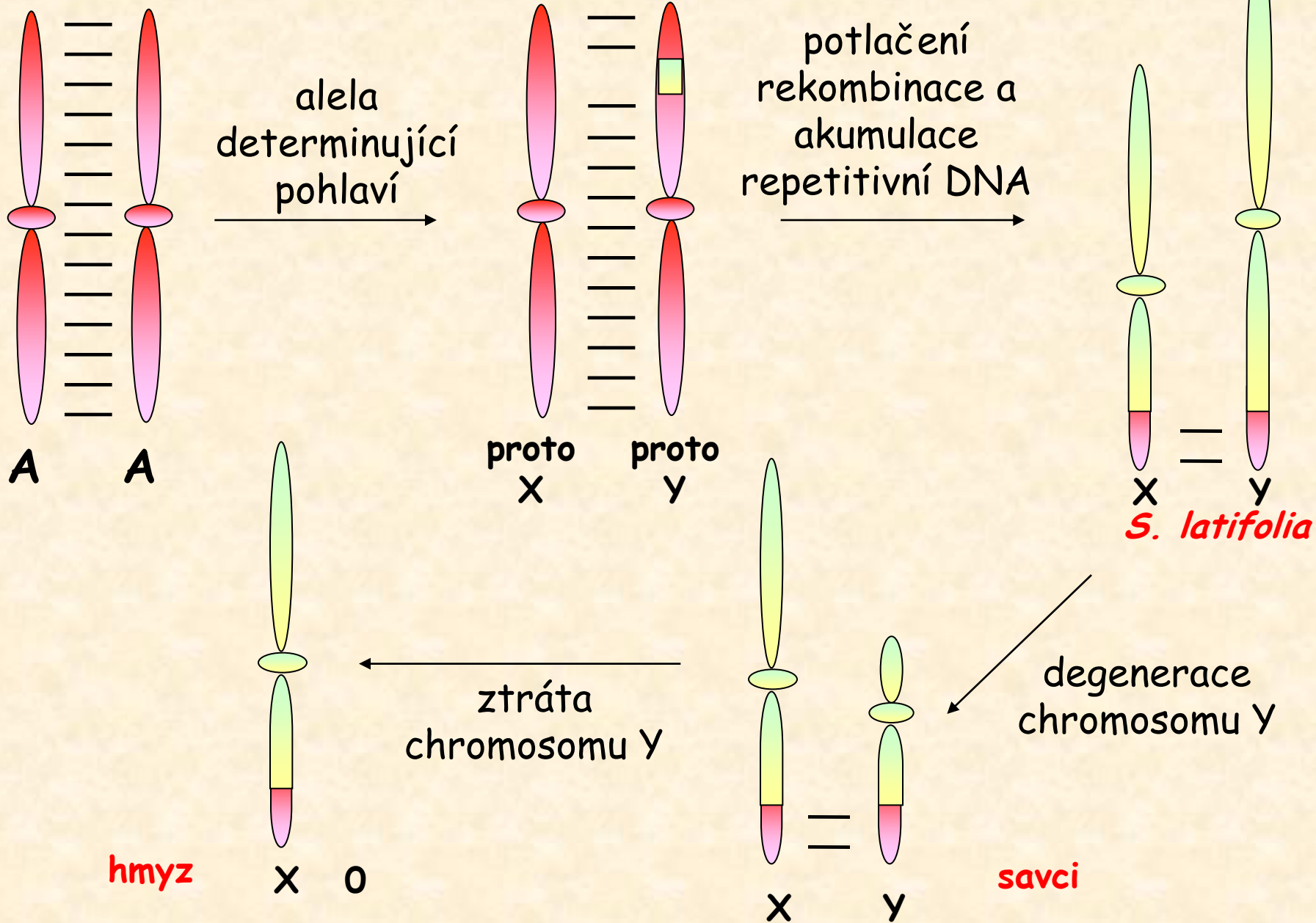


# Různé počty chromosomů u blízce příbuzných druhů jelíneků *Muntiacus*

různé druhy:

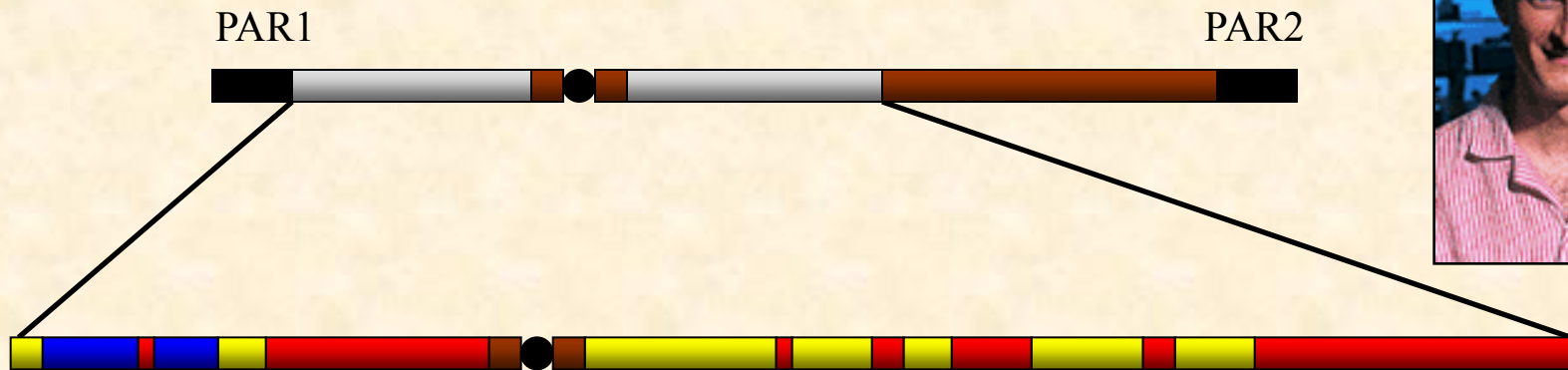
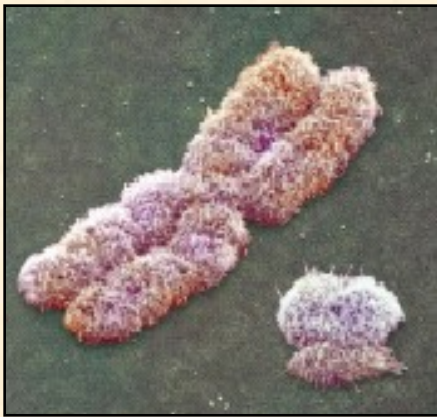


# Evoluce pohlavních chromosomů








# Lidský chromosom Y: „Sál plný zrcadel“

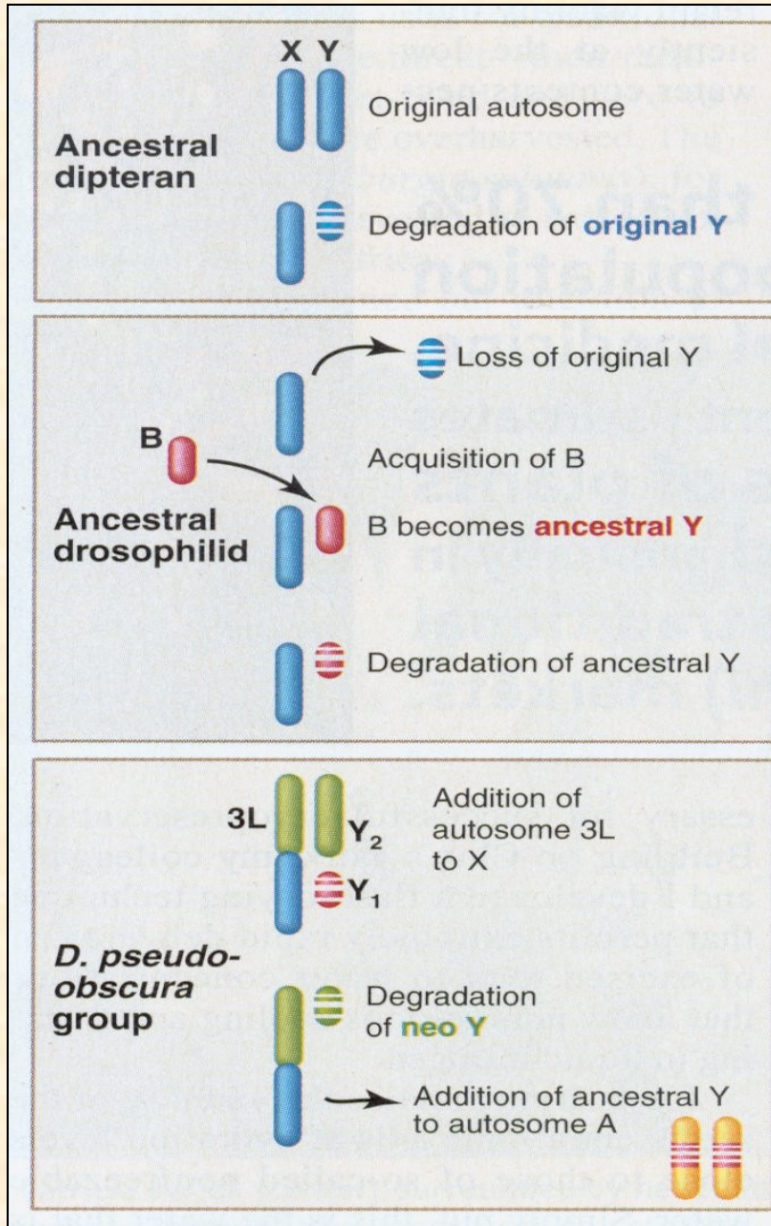


## Palindromy:

- 3 000 000 bází dlouhé
- 99.9% identita

-  Homologní oblasti mezi X a Y
-  Duplikativní přenos z X
-  Palindromy

# Recyklace pohlavních chromosomů u drozofily



První Y chromosom: „**original Y**“:  
- vymizel před více než 60 mil. lety

Druhý Y chromosom: „**ancestral Y**“:  
- párování B chromosomu s X chromosomem ~ před 60 mil let  
- získání užitečných genů z autosomů  
- degenerace

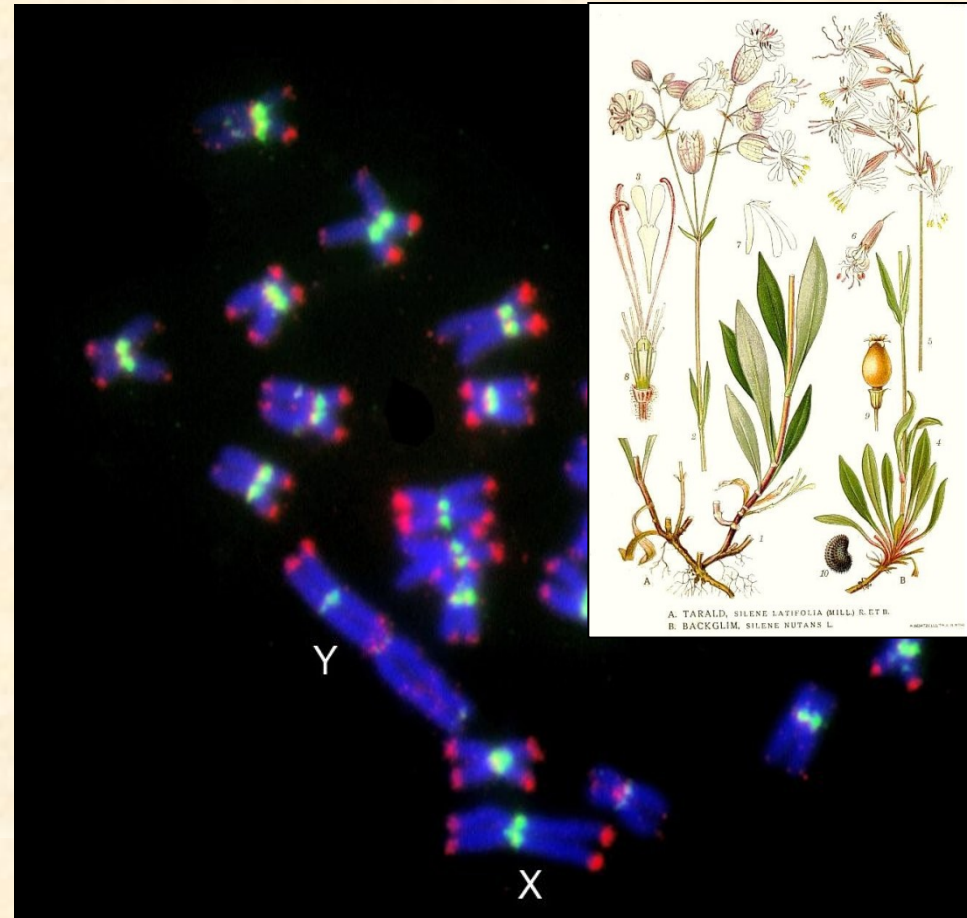
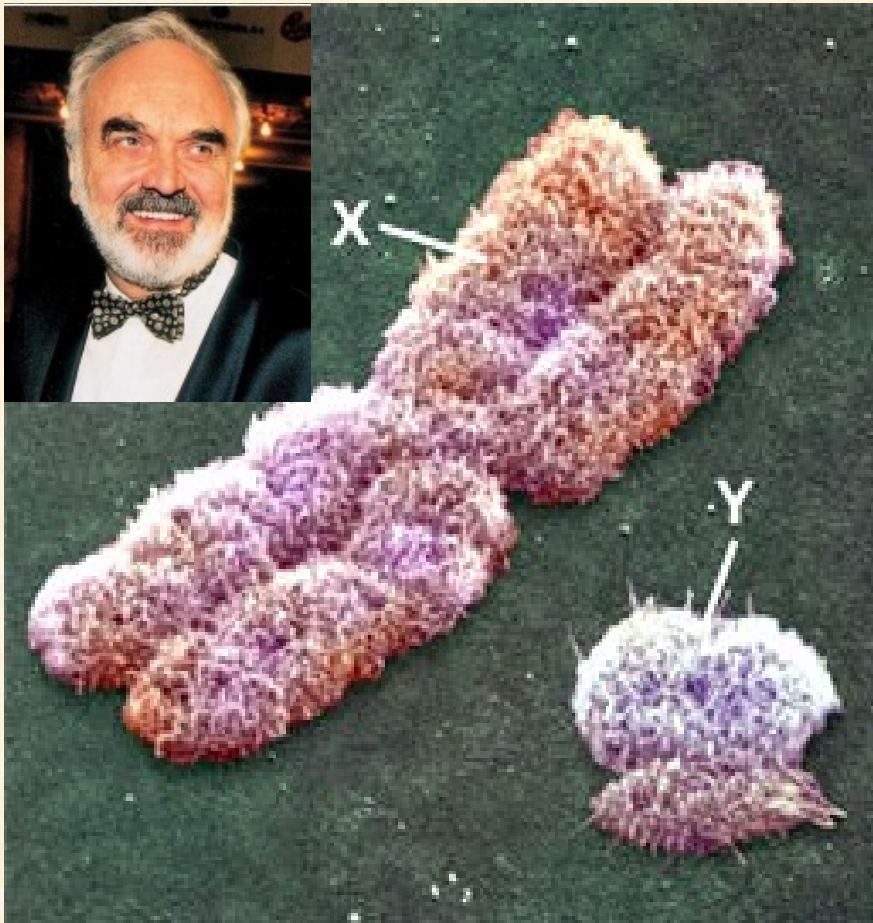
Třetí Y chromosom: „**neo Y**“  
- fúze X s autosomem  
- připojení ancestrálního Y na A, recovery  
- opět degenerace neo-Y

# Silenka širolistá: model pro studium mladých pohlavních chromosomů

Degenerace chromosomu Y již v časných fázích jeho evoluce

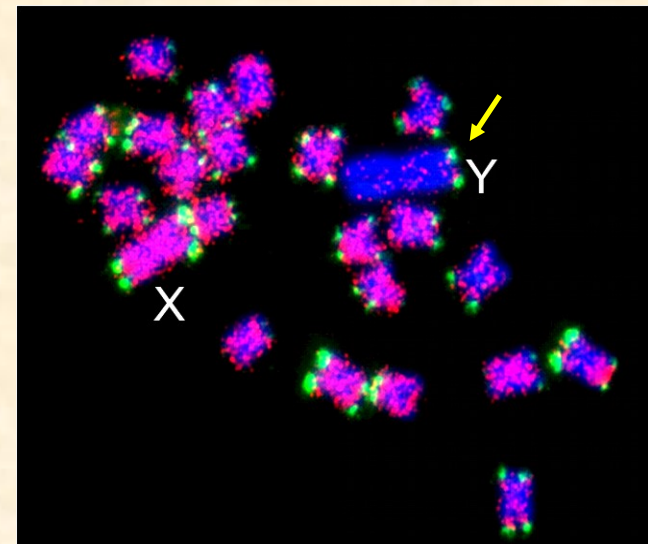
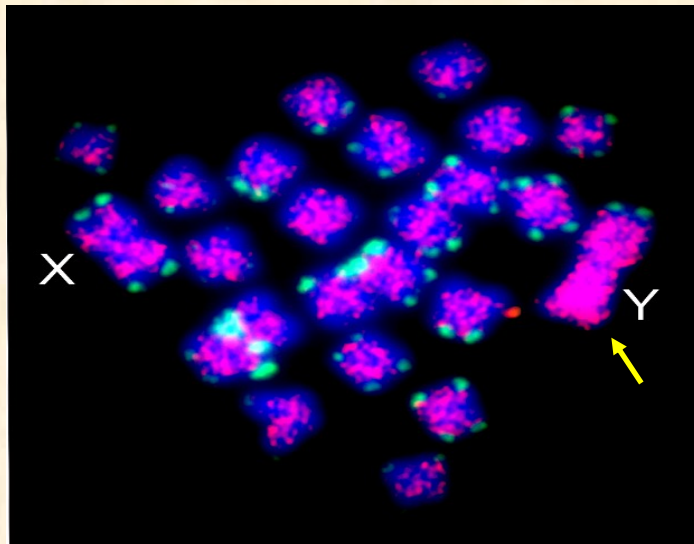
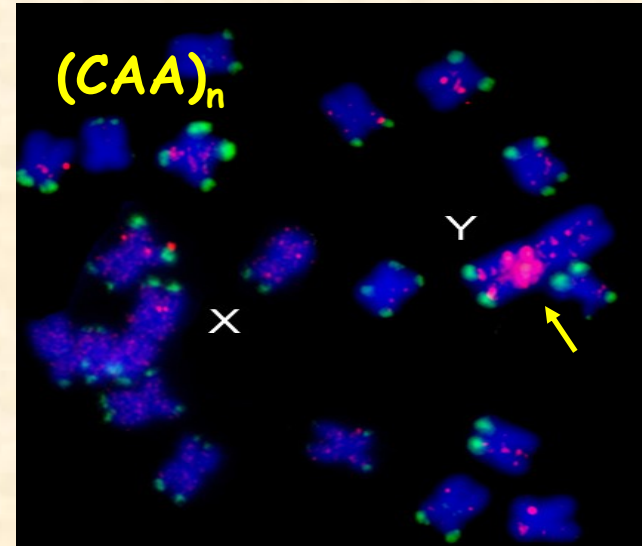
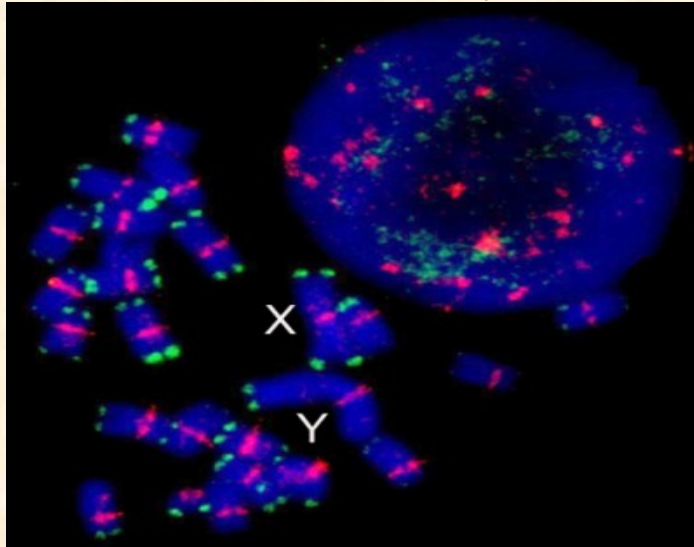
člověk (*Homo sapiens*): 300 mil let

silenka širolistá (*Silene latifolia*): 10 mil let





# Akumulace repeticí na mladém chromosomu Y



Chromosome Research (2008) 16:961–976  
DOI: 10.1007/s10577-008-1254-2

© Springer 2008

Survey of repetitive sequences in *Silene latifolia* with respect to their distribution on sex chromosomes

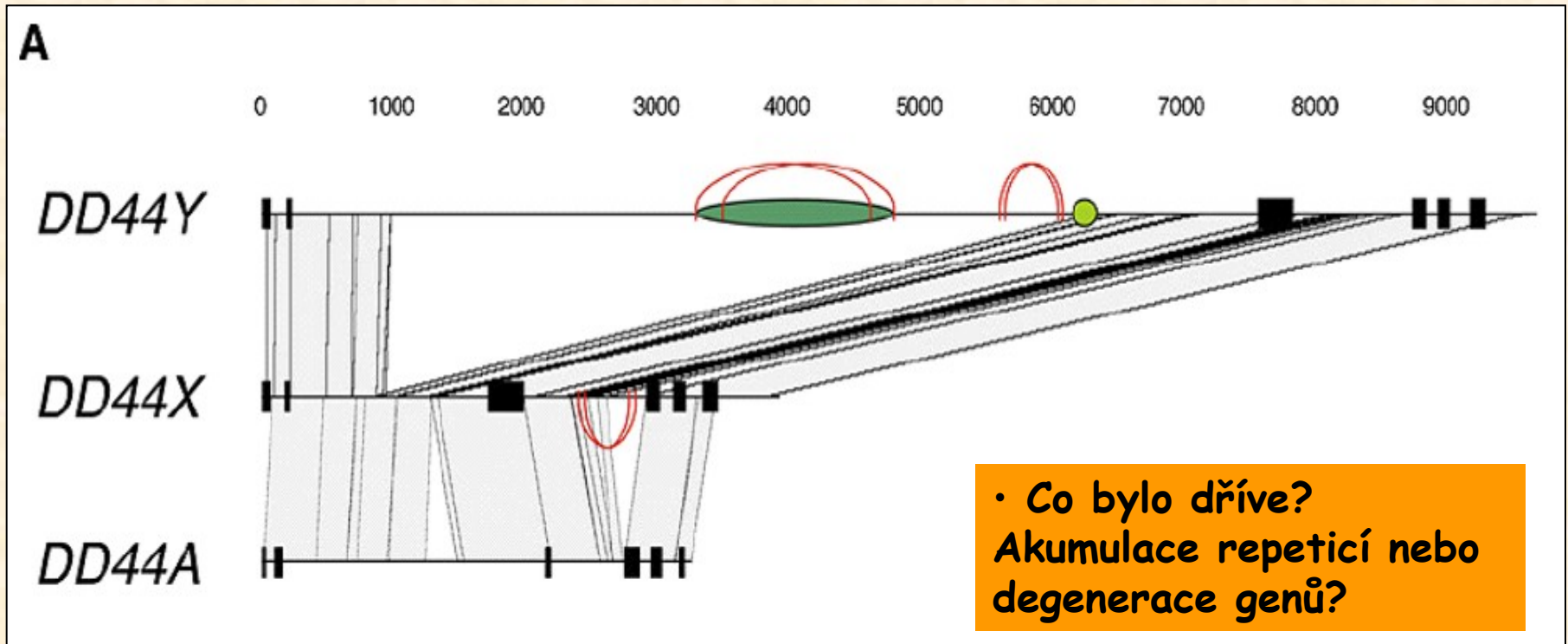
Tomas Cermak<sup>1</sup>, Zdenek Kubat<sup>1</sup>, Roman Hobza<sup>1</sup>, Andrea Koblikova<sup>2</sup>, Alex Widmer<sup>3</sup>, Jiri Macas<sup>2</sup>, Boris Vyskot<sup>1</sup> & Eduard Kejnovsky<sup>1\*</sup>

Microsatellite accumulation on the Y chromosome in *Silene latifolia*

Zdenek Kubat, Roman Hobza, Boris Vyskot, and Eduard Kejnovsky



# Degenerace genů na chromosomu Y: Více repetitivní DNA v genech na chromosomu Y



Current Biology 18, 545–549, April 8, 2008 ©2008 Elsevier Ltd All rights reserved DOI 10.1016/j.cub.2008.03.023

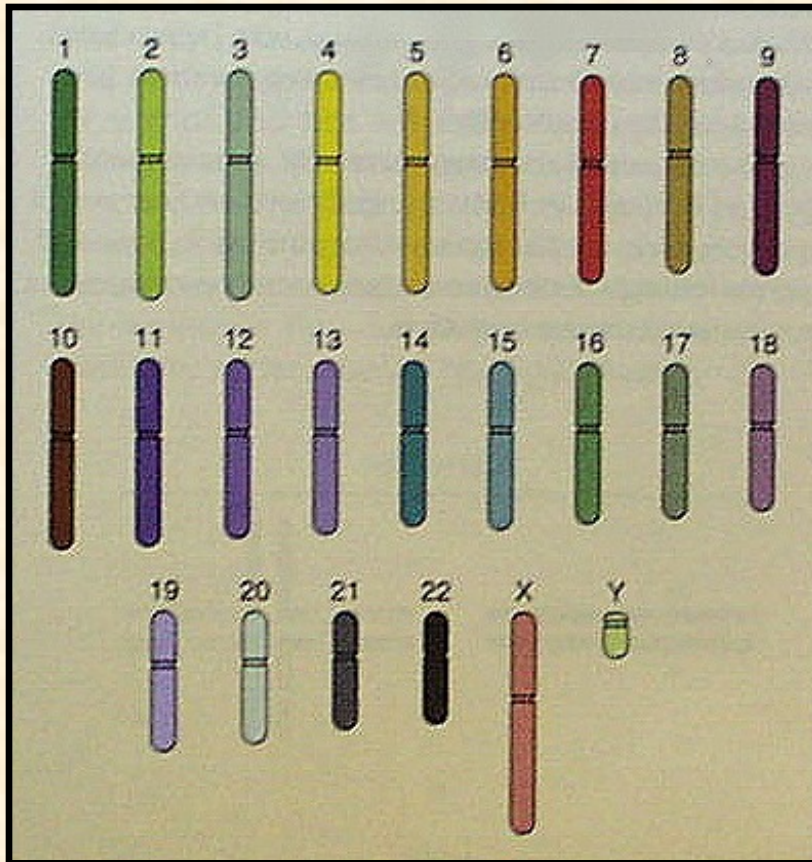
Report

## Evidence for Degeneration of the Y Chromosome in the Dioecious Plant *Silene latifolia*

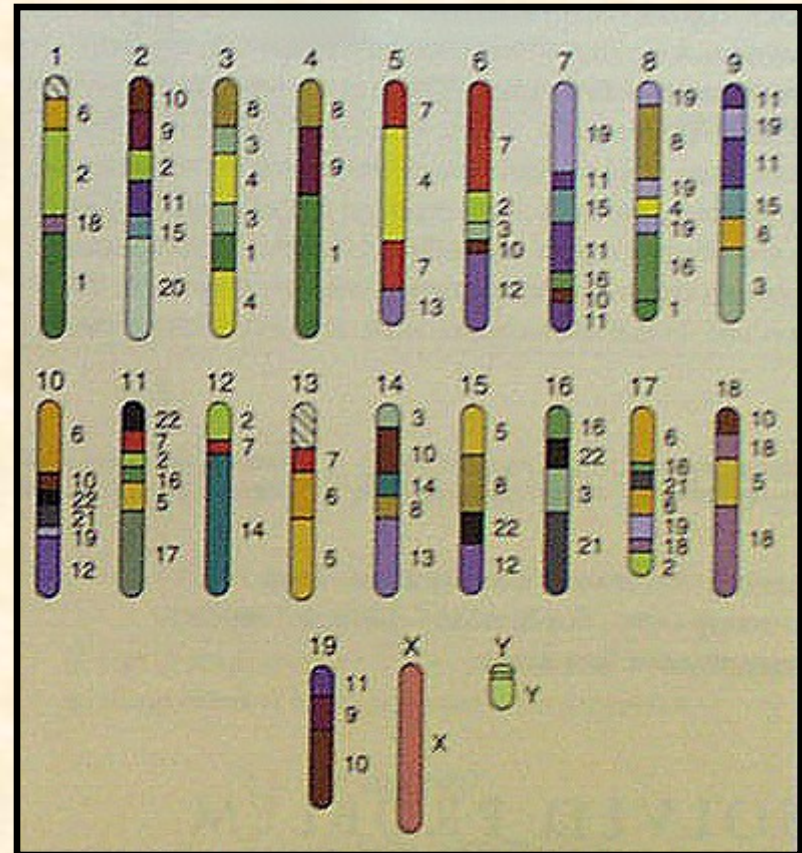
Gabriel A.B. Marais,<sup>1,\*</sup> Michael Nicolas,<sup>2,6</sup> Roberta Bergero,<sup>3</sup>  
Pierre Chambrier,<sup>2</sup> Eduard Kejnovsky,<sup>4</sup> Françoise Monéger,<sup>2</sup>  
Roman Hobza,<sup>4,5</sup> Alex Widmer,<sup>5</sup> and Deborah Charlesworth<sup>3</sup>

# Srovnání lidských a myších chromosomů

člověk



myš



Od evoluční divergence došlo k četným přestavbám, které umístily bloky genů do různých kombinací.

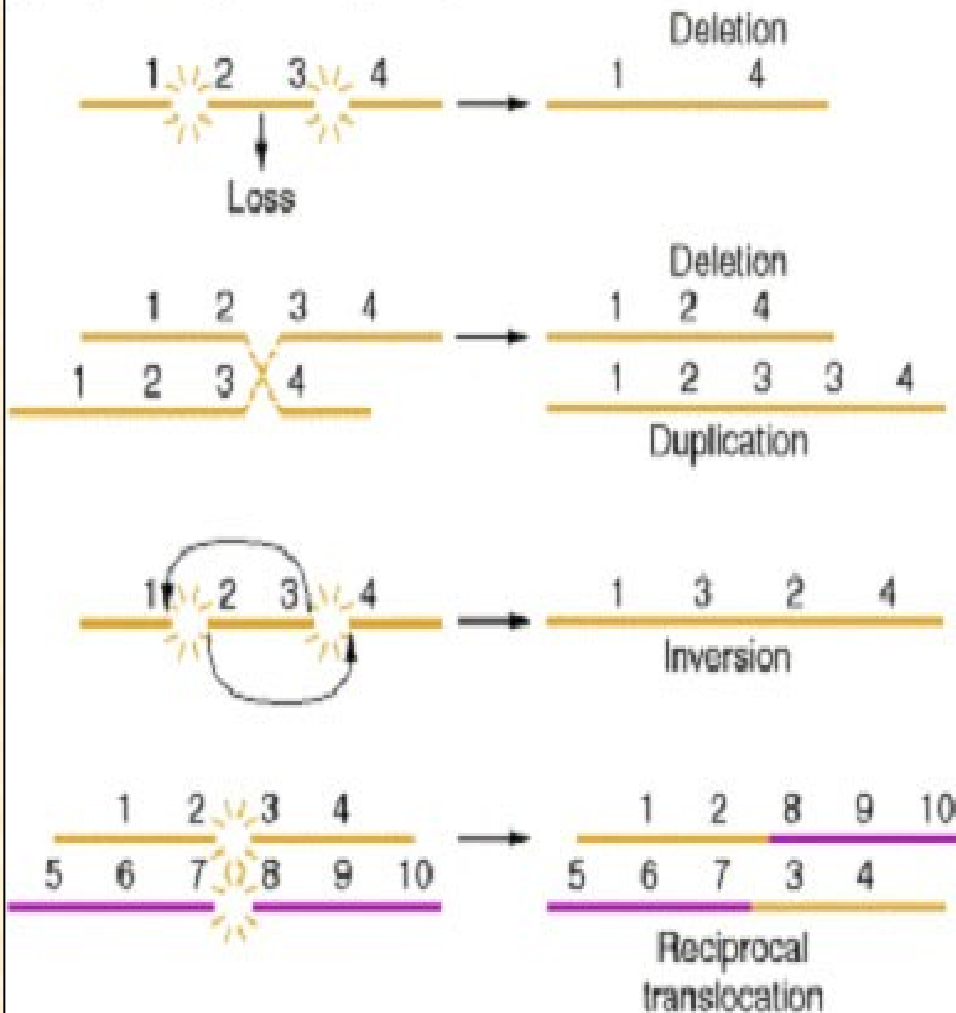
# Chromosomové přestavby

zlom a znovuspojení (del., dupl., inv.)

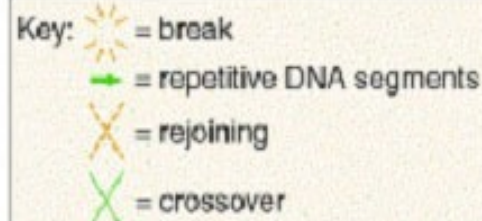
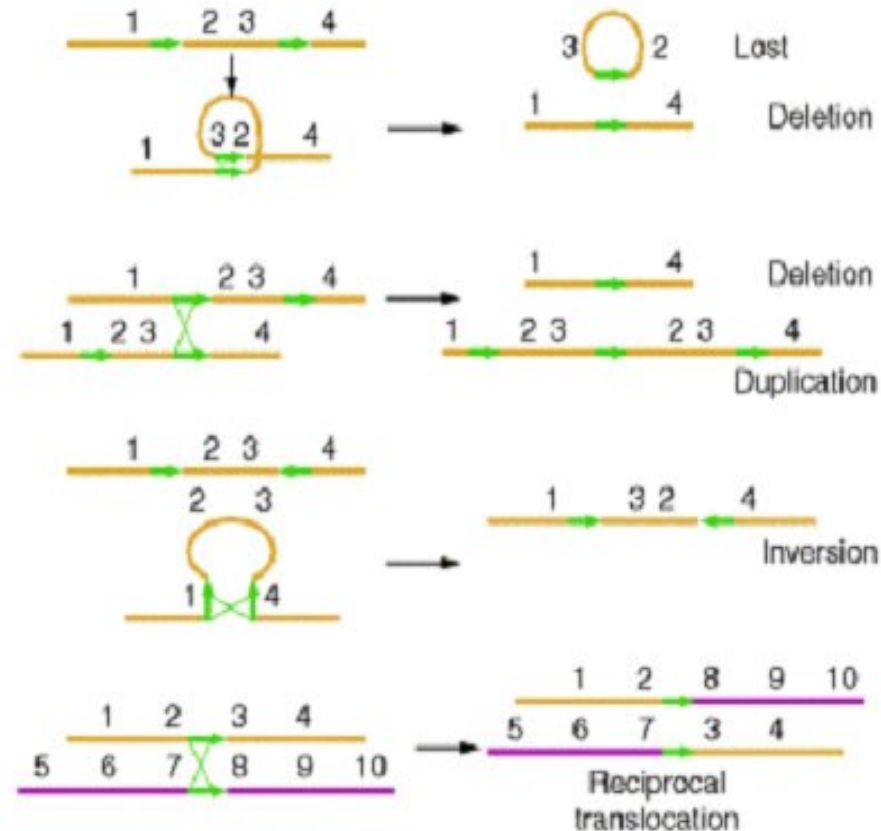
rekombinace mezi repetitivními

## Origins of Chromosomal Rearrangements

### (a) By breakage and rejoining



### (b) By crossing-over between repetitive DNA





# Isochorový model organizace genomu

- Isochory:** bloky genomové DNA (stovky kb-Mb) s charakteristickým GC-obsahem, ostrý přechod (*G. Bernardi*)
- shlukování do skupin s diskrétními GC-obsahy, u člověka typy L1, L2, H1, H2 a H3 (GC nejbohatší, tvoří jen 3% ale obsahuje 25% genů)
  - Mozaiková struktura genomu obratovců a rostlin



## Původ izochor:

1. Výsledek mutací?
2. Výsledek selekce?
3. Genová konverze?

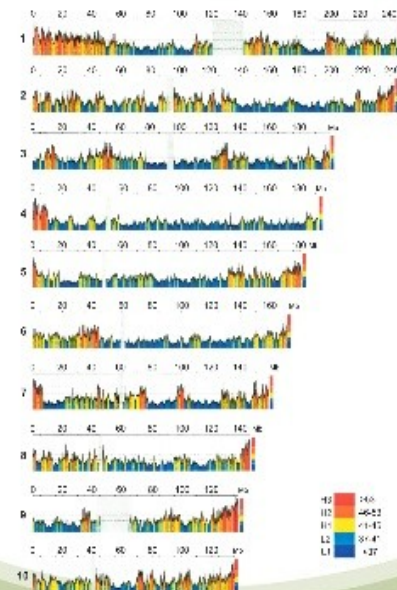
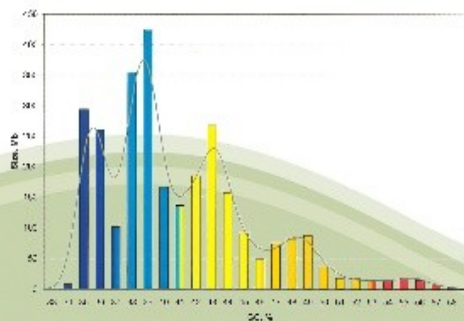


Figure 1 | Large-scale variation in G+C content. a | The classic isochore model. b | G+C content across the three classes of human major histocompatibility (MHC) region of chromosome 6 (data from GenBank); G+C content is plotted as a moving average, and the window size is 100 kb, advanced by 10 kb each step.

## Finding isochores

- Isochores: homogeneous regions of %GC content

- Easy to find with windowed (100kbp) %GC calculation, from sequenced genomes.
- 3200 isochores characterised in the human genome, consistent with 5 levels (L1, L2, H1, H2, H3) found by staining/hybridisation.

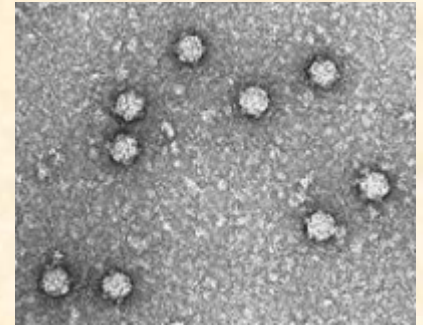


# GENOMOVÉ PROJEKTY



# První genomové projekty - začátek éry genomiky

1977: první kompletní genom (virus) -  
bakteriofág *phiX174* (5 386 bp)  
Sanger et al., Nature 265, 687-695.



1995: první kompletní genom autonomního organismu -  
bakterie *Haemophilus influenzae* (1.38 Mb)  
Fleischmann et al., Science 269, 496-512.



1996: první eukaryotický organismus -  
kvasinka *Saccharomyces cerevisiae*  
Goffeau et al., Science 274, 563-567.



# Genomové projekty



1996: *Saccharomyces cerevisiae* (12 Mb/ 6548 genes)

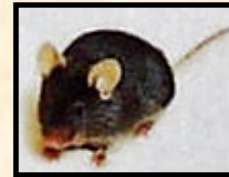


1998: *Caenorhabditis elegans* (97 Mb/19 099 genů)

2000: *Drosophila melanogaster*  
(137 Mb/13 500 genů)



*Mus musculus*  
(3 300 Mb)

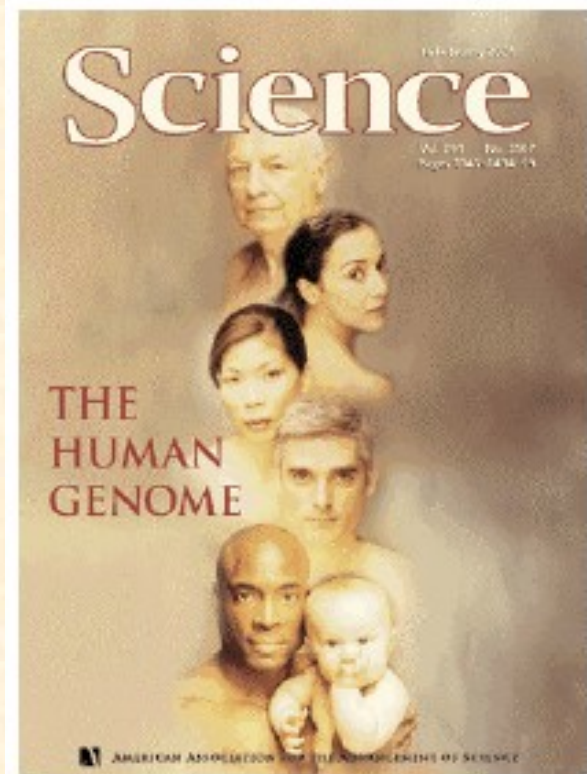


2000: *Arabidopsis thaliana*  
(125 Mb/25 500 genů)










# Sekvenování lidského genomu



# Jak zjistit stav genomových projektů

Genomes OnLine Database (GOLD): <https://gold.jgi.doe.gov/>

<b>Studies</b> Metagenomic <a href="#">1 662</a> Non-Metagenomic <a href="#">36 877</a>	<b>Biosamples</b>  <b>Classification</b> Ecosystems Host-associated <a href="#">30 777</a> Engineered <a href="#">5 175</a> Environmental <a href="#">27 633</a>	<b>Sequencing Projects</b>  Complete Projects <a href="#">16 754</a>  Permanent Drafts <a href="#">174 424</a>  Incomplete Projects <a href="#">88 541</a>  Targeted Projects <a href="#">1 085</a>	<b>Analysis Projects</b> Genome Analysis <a href="#">151 307</a> Metagenome Analysis <a href="#">39 132</a> Metagenome - Cell Enrichment <a href="#">1 674</a> Metagenome - Single Particle Sort <a href="#">4 706</a> Metagenome - Assembled Genome (MAG) <a href="#">10 613</a> Metatranscriptome Analysis <a href="#">6 264</a> Combined Assembly <a href="#">255</a> Single Cell - Screened (SAG) <a href="#">2 167</a> Single Cell - Unscreened (SAG) <a href="#">1 928</a> Transcriptome Analysis <a href="#">528</a>
<b>Special Projects</b> Type Strains with Projects <a href="#">7 760</a> Type Strains with projects at Genbank <a href="#">6 515</a> GEBA Projects <a href="#">3 358</a> HMP Projects <a href="#">2 914</a>	<b>Projects w. Genbank Data</b> Seq. Projects <a href="#">147 013</a> Archaeal Projects <a href="#">920</a> Bacterial Projects <a href="#">133 285</a> Eukaryal Projects <a href="#">4 300</a> Viral Projects <a href="#">8 508</a>	<b>JGI Projects</b> JGI Studies <a href="#">1 383</a> JGI Biosamples <a href="#">22 571</a> JGI Sequencing Projects <a href="#">83 028</a> JGI Analysis Projects <a href="#">43 120</a>	<b>Organisms</b> Organisms <a href="#">360 291</a> Archaea <a href="#">3 210</a> Bacteria <a href="#">318 095</a> Eukarya <a href="#">30 019</a> Viruses <a href="#">8 967</a> Bacterial Type Strains <a href="#">15 559</a> Archaeal Type Strains <a href="#">558</a>

Genbank - sekvence DNA, veřejně přístupná

SRA -sequence read archive - NGS data

SWISS-PROT - sekvence proteinů, popis funkce, struktury domén

dbEST - sekvence cDNA z různých organismů

REBASE - restriktázy a metylázy

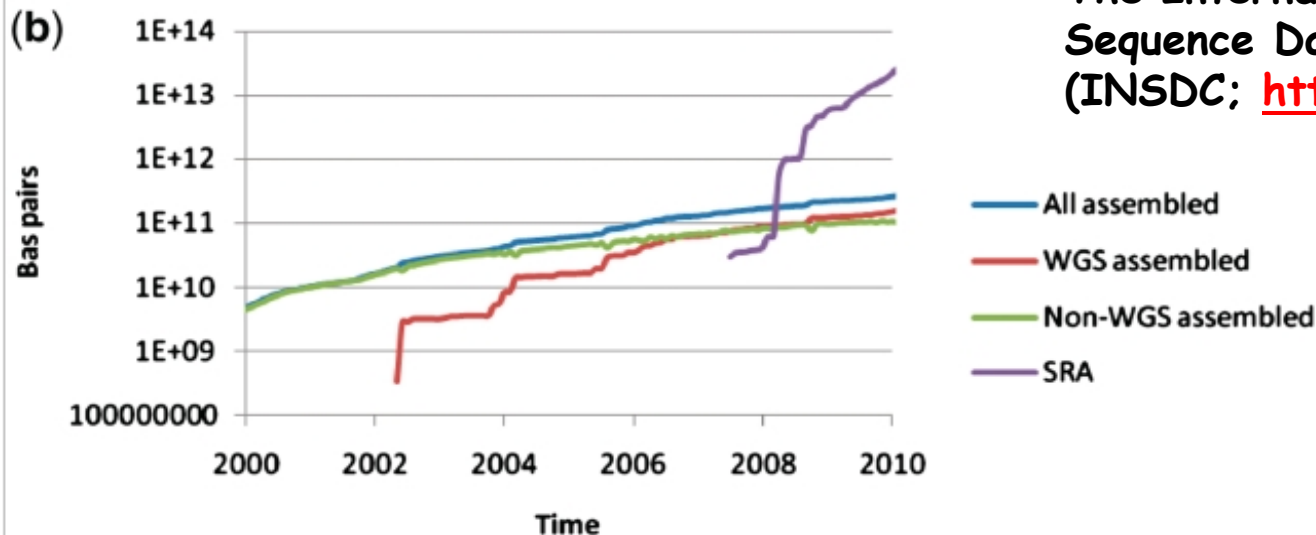
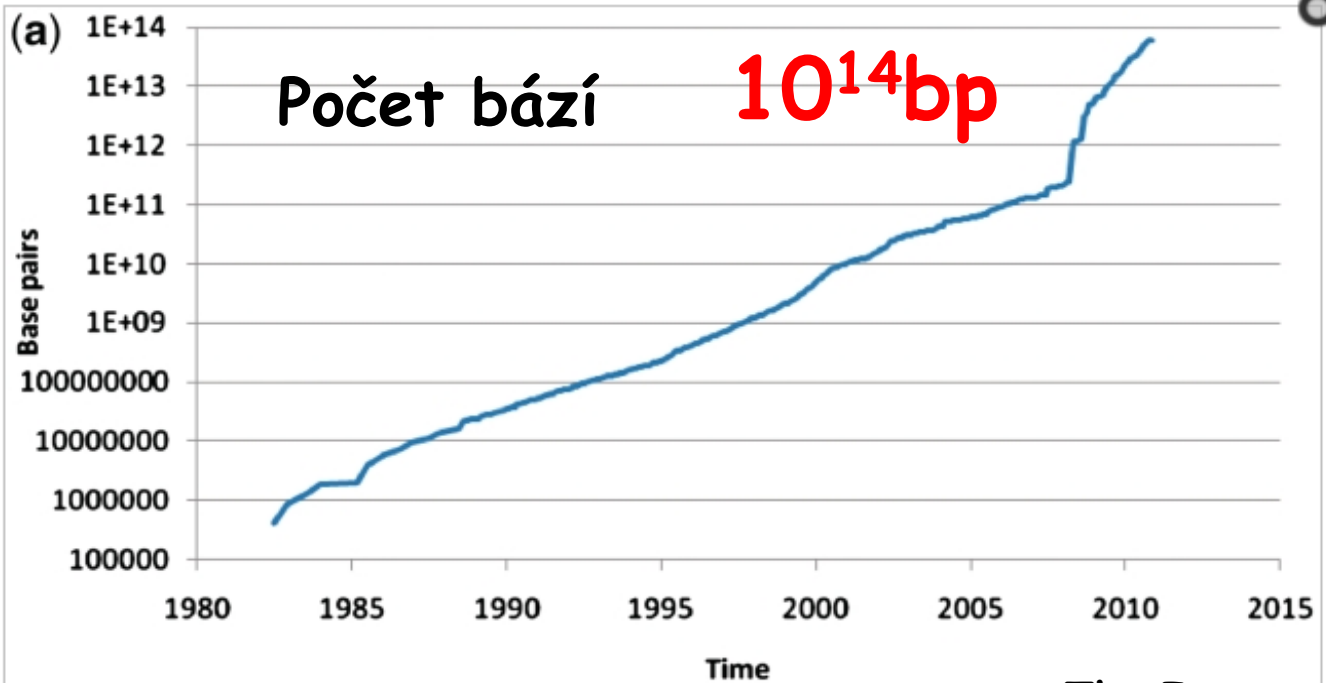
PEDANT - kompletní nebo parciální sekvence genomů

PDB - 3D koordináty makromolekulárních struktur

REPBASE - repetitivní sekvence

7.3.2019

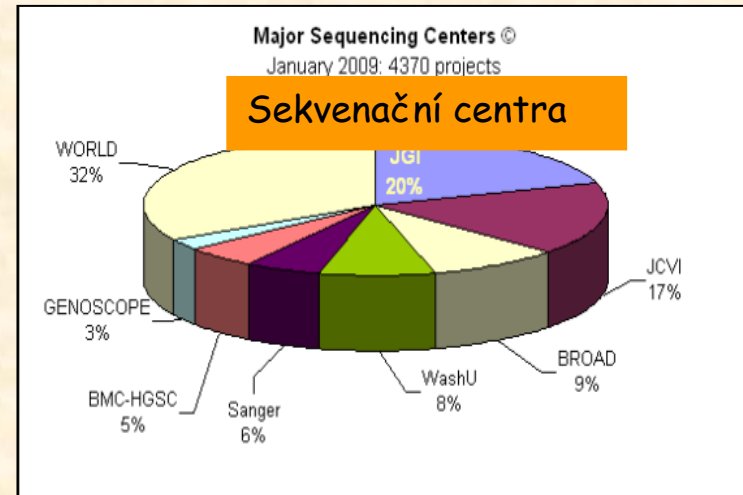
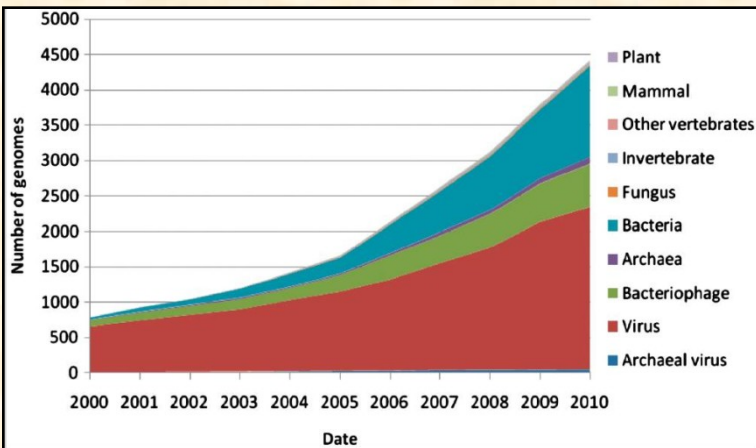
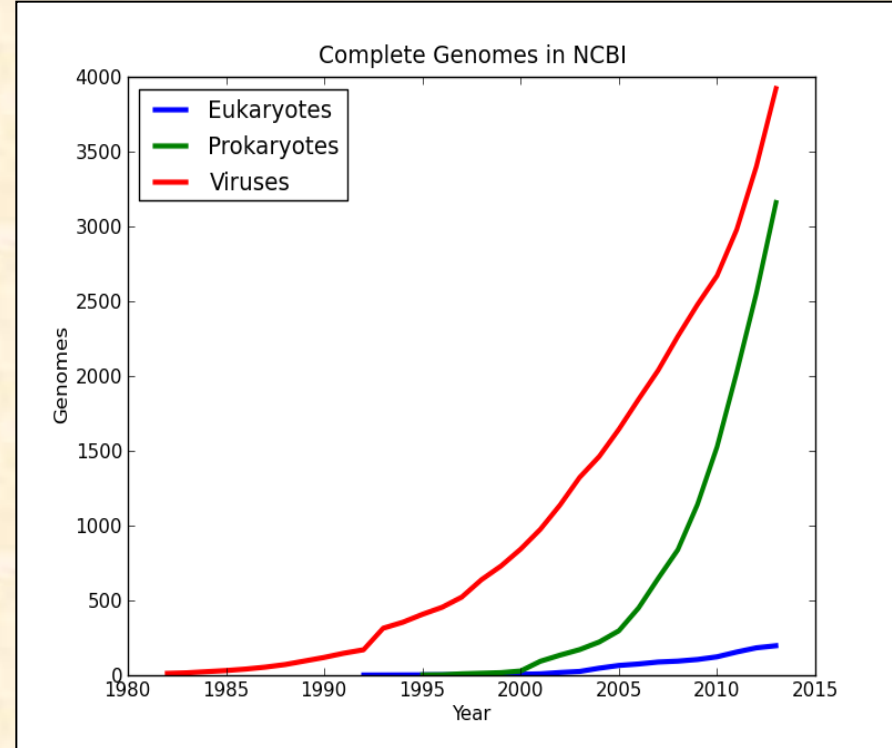
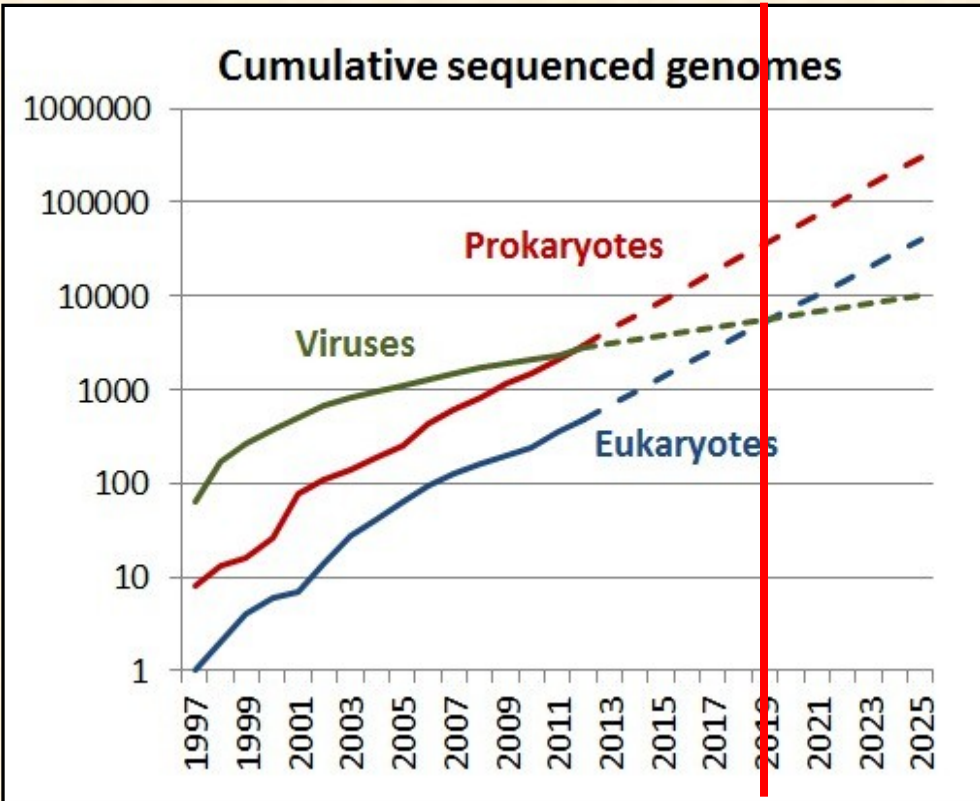
# Objem dat v databázích roste exponenciálně



The International Nucleotide  
Sequence Database Collaboration  
(INSDC; <http://www.insdc.org>)



# Genomové sekvenační projekty (počet genomů)



# Prudký pokles ceny sekvenování genomu (člověk)

Cost per Genome

Které jiné  
zboží zlevnilo  
téměř  
100 000x ???

