

# CG020 Genomika

## Přednáška 8

# Struktura a organizace genomů

Markéta Pernisová

**Funkční genomika a proteomika, NCBR**  
Přírodovědecká fakulta, Masarykova univerzita, Brno  
[pernisov@sci.muni.cz](mailto:pernisov@sci.muni.cz)

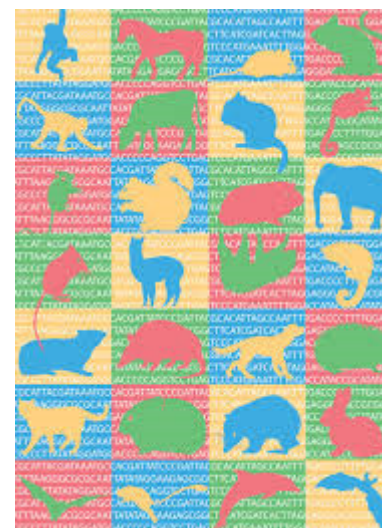


INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

Tato prezentace je spolufinancována  
Evropským sociálním fondem  
a státním rozpočtem České republiky

# Osnova

1. Eukaryotický jaderný genom
2. Genomy prokaryot a eukaryotických organel
3. Genomy virů a mobilní elementy
4. Internetové zdroje
5. Literatura

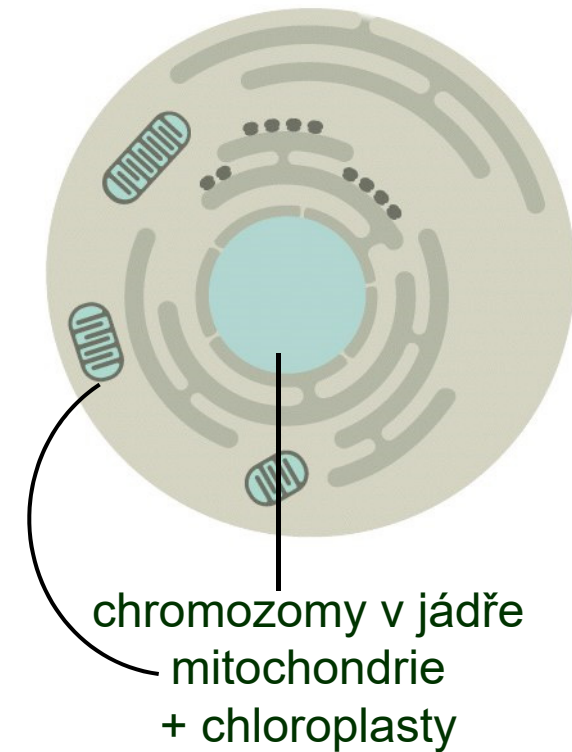
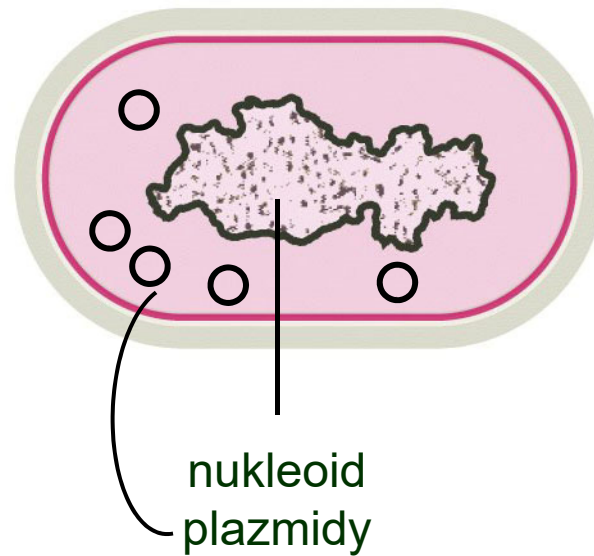
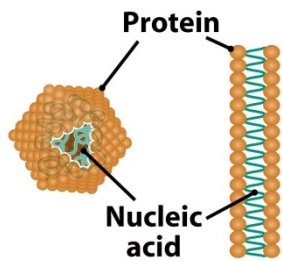


# GENOM

Genom – soubor genetické informace organismu

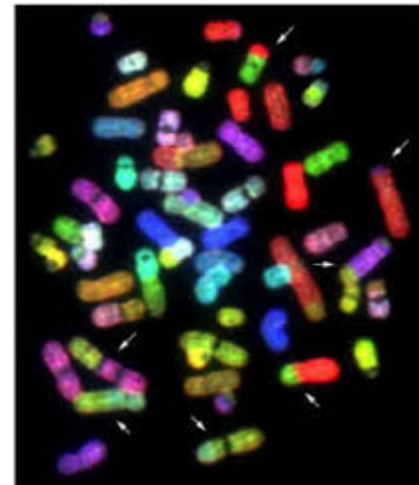
- biologická informace k sestavení a přežití živého jedince

- eukaryotický
- prokaryotický
- viry



# EUKARYOTICKÝ JADERNÝ GENOM

= sada lineárních molekul DNA, bez výjimky,  
minimálně dvě

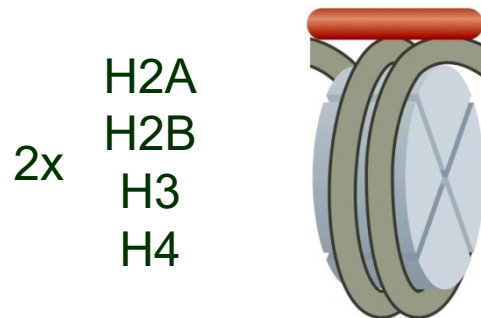




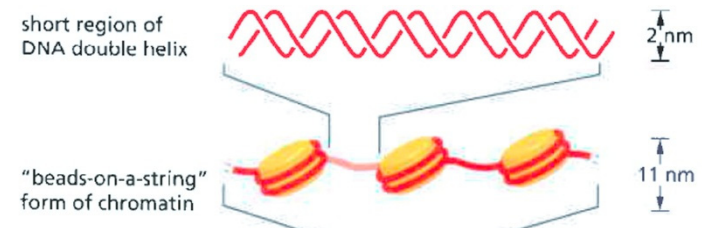
# STRUKTURA CHROMOZOMŮ

DNA + histony = nukleozom

- 140-150 bp
- linker: 50-70 bp
- + spojovací histony



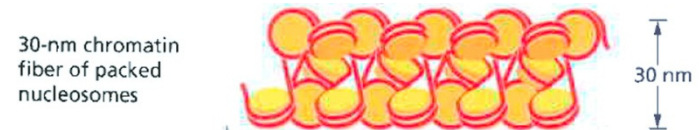
„korálková“ forma chromatinu  
– 11 nm



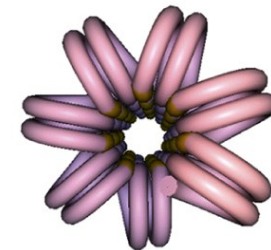
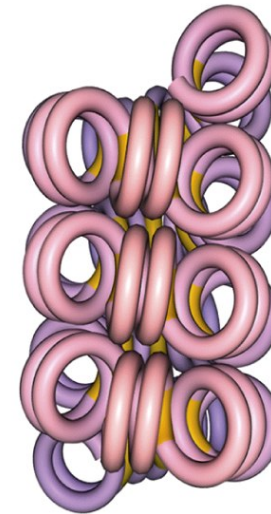
# STRUKTURA CHROMOZOMŮ

30 nm chromatinové vlákno  
interfázní chromozomy

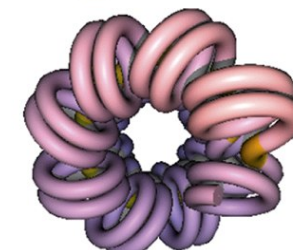
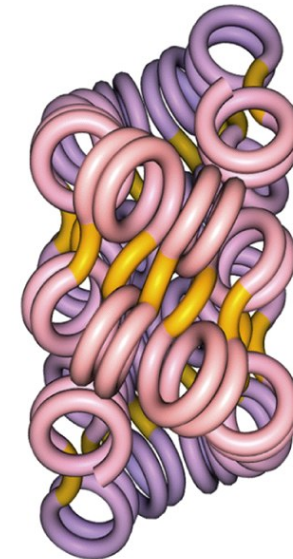
- několik teorií, 2 modely:
  - solenoidový model – pouze spojovací histony (např. H1)
  - helikální model – spojovací histony + konce histonů tvořících jádro nukleozomu – chemické modifikace konců histonů otevírají 30 nm vlákno



**(A) The solenoid model**

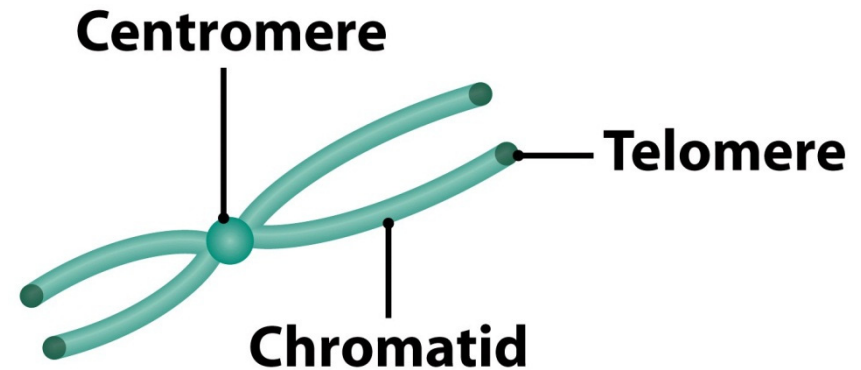
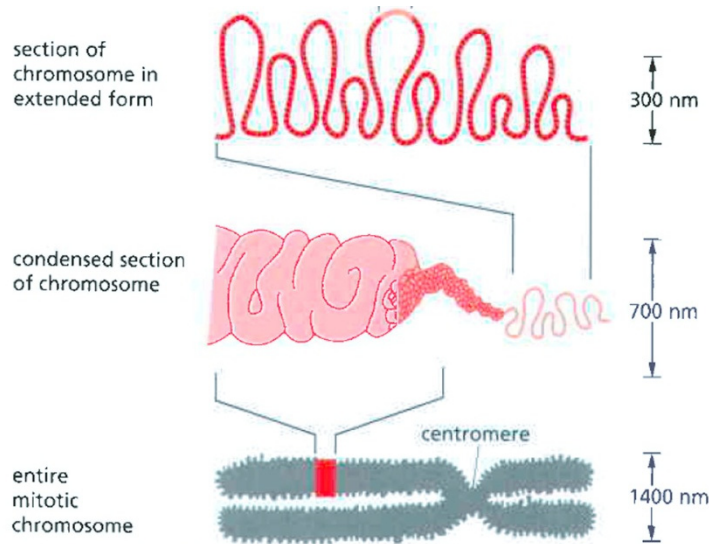


**(B) The helical ribbon model**

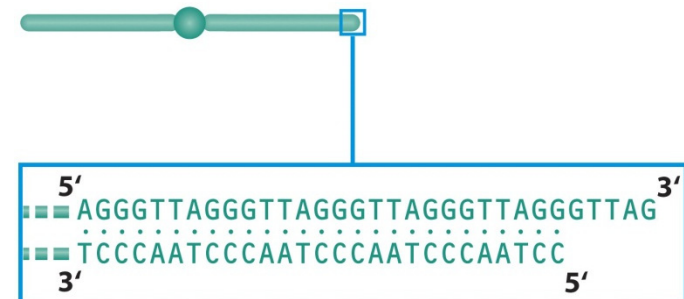


# STRUKTURA CHROMOZOMŮ

Kondenzované metafázní chromozomy: 1400 nm  
jedna chromatida: 700 nm



Centromery, telomery –  
repetitivní sekvence



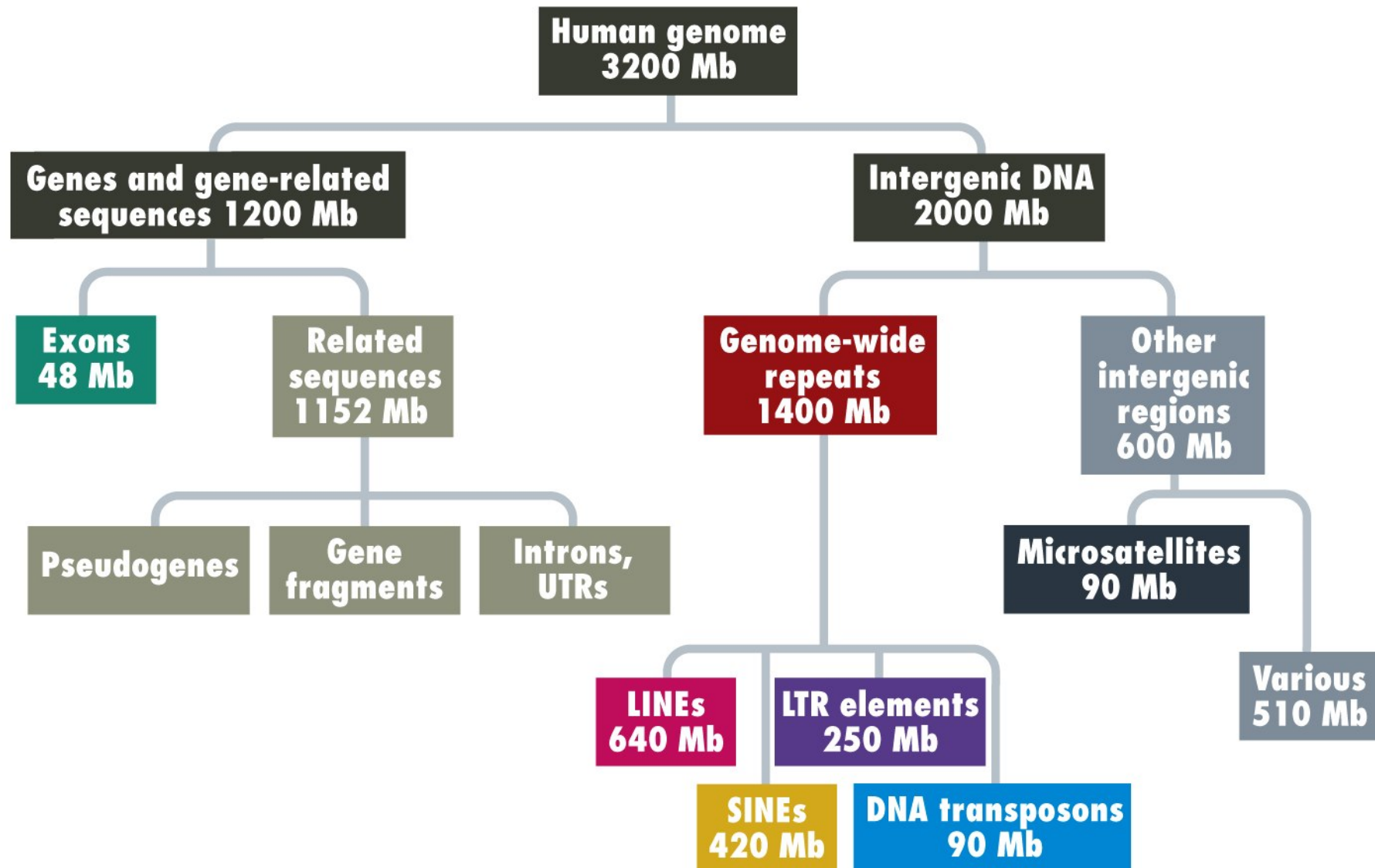


# NETYPICKÉ CHROMOZOMY

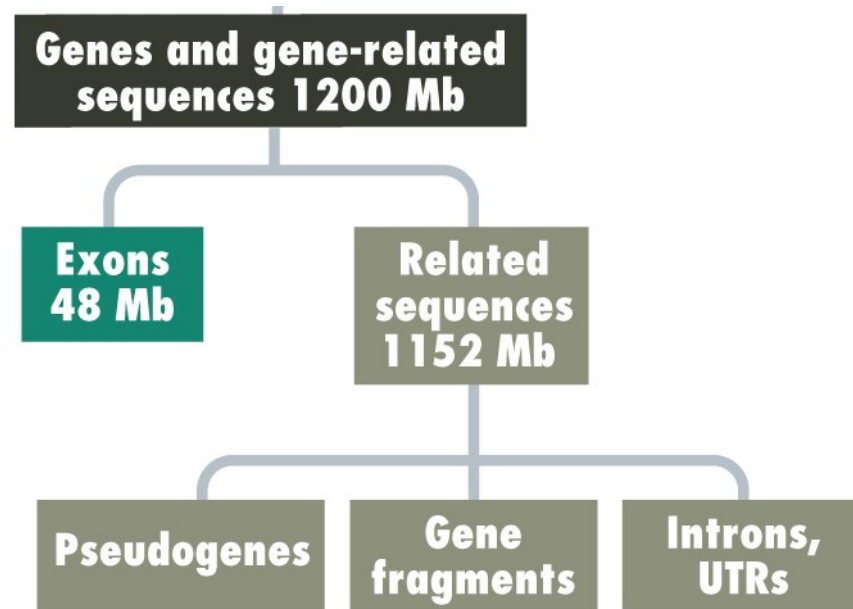
- Minichromozomy
  - krátké, vysoká hustota genů (až 6x)
  - např. kuře
  - přístup: „*de novo*“, „top-down“
- B chromozomy
  - individuální, nevyskytují se u celé populace
  - fragmenty normálních chromozomů, zřejmě pozůstatek chyb v průběhu dělení jádra
  - běžné u rostlin, snižují životaschopnost
  - i houby, hmyz, živočichové
- Holocentrické chromozomy
  - ne jedna centromera, ale několik podobných struktur podél chromozomu
  - např. *Caenorhabditis elegans*



# ORGANIZACE GENOMU



# GENY A PŘÍBUZNÉ SEKVENCE



# GENY

## Geny

- UTR – nepřekládaná oblast genu
- Introny – sestřih
- Exony – funkční produkty

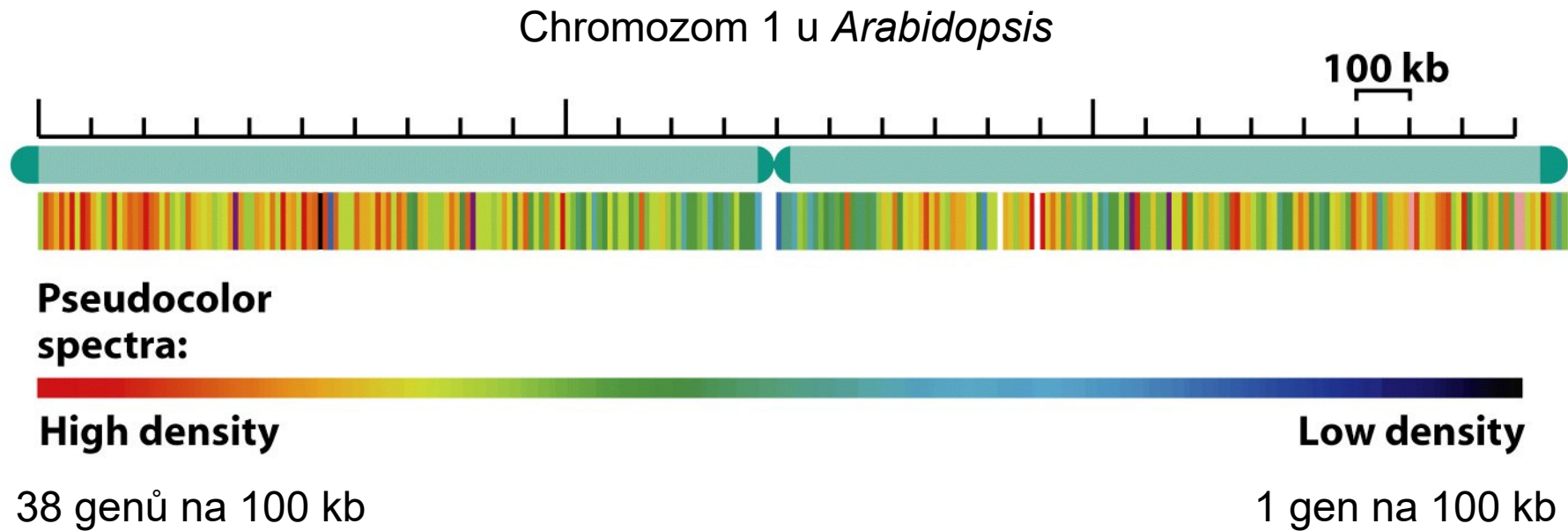
Genové rodiny - skupiny genů se stejnou nebo podobnou sekvencí

- jednoduché
  - vznikly pravděpodobně duplikací genů
  - např. geny pro lidskou ribozomální RNA:
    - 2000 genů pro 5S rRNA – všechny v jedné skupině lokalizované na chromozomu 1
    - 280 kopií genů pro 28S, 5.8S a 18S rRNA – pět skupin po 50-70 kopiích na pěti chromozomech
- komplexní
  - podobné sekvence (ale ne stejné), odlišné vlastnosti
  - např. savčí globiny – exprimované v různých vývojových stádiích



# ROZLOŽENÍ GENŮ NA CHROMOZOMU

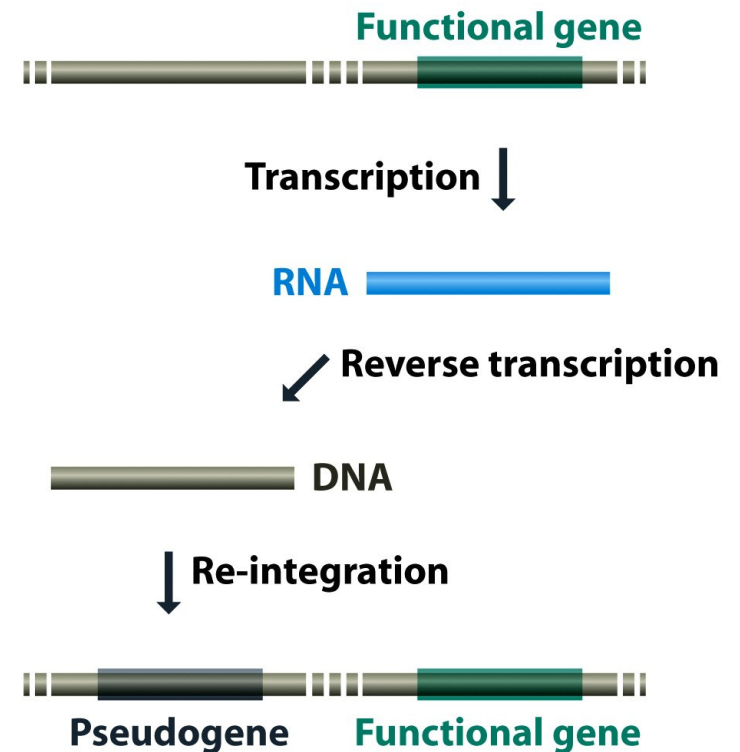
- Nehomogenní
- Geny i v centromere, nižší hustota



Člověk: 1-64 genů na 100 kb

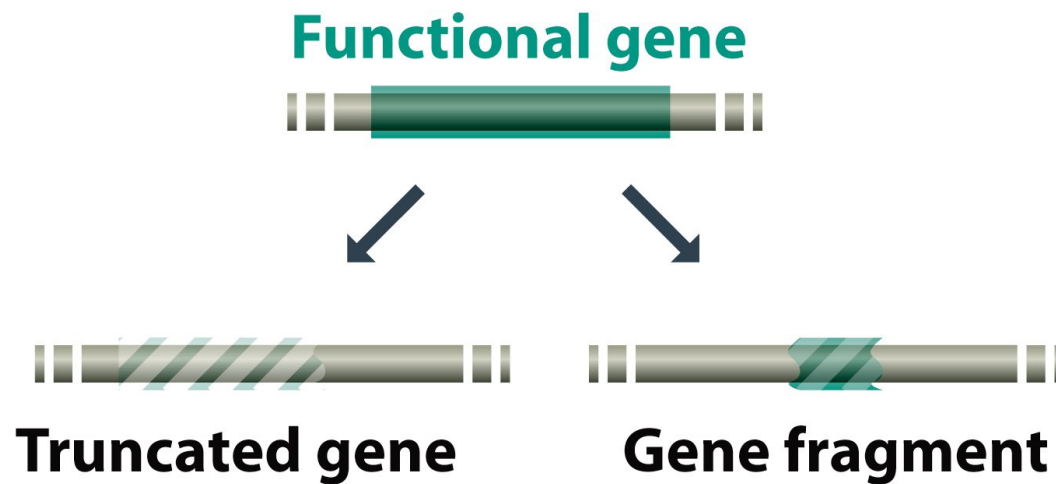
# PSEUDOGENY

- evoluční relikv
- 2 skupiny
  - konvenční – vznikly mutací
    - částečně funkční nebo nefunkční
  - procesované – odvozené z mRNA
    - nemají introny
    - chybí jim regulační sekvence před genem
    - nefunkční



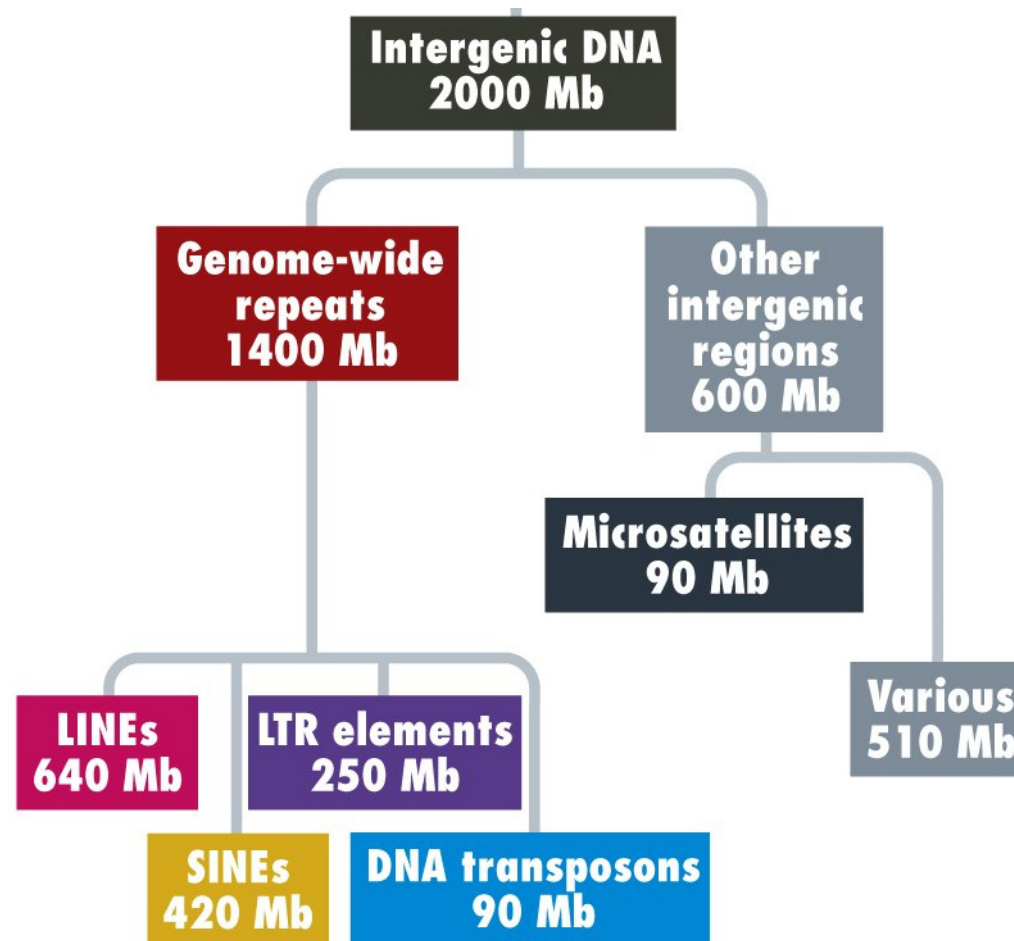
# FRAGMENTY GENŮ

- Zkrácené geny
- Genové fragmenty

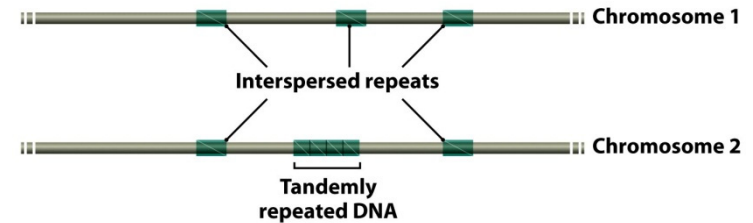


# INTERGENOVÁ DNA

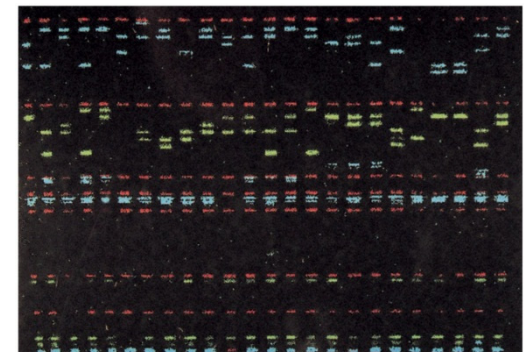
- „junk“ (= veteš, haraburdí) DNA – není pravda



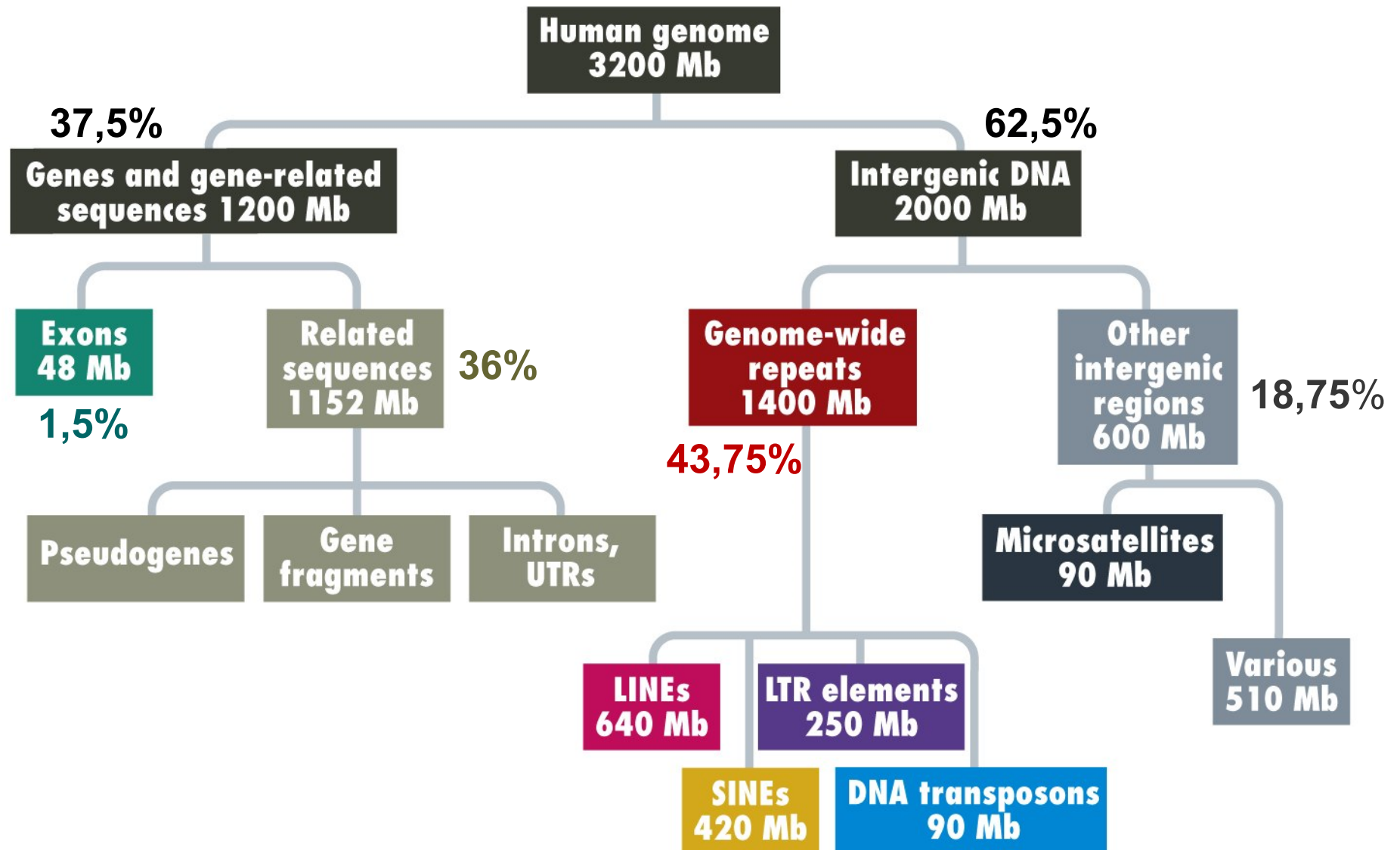
# REPETICE



- Rozptýlené repetice
  - vznikají transpozicí
    - LINEs (long interspersed nuclear elements) – nad 300 bp
    - SINEs (short interspersed nuclear elements) – přibližně 300 bp
- Tandemové repetice - satelity
  - LTR (long terminal repeat) elementy nebo v centromerách
- Minisatelity („variable number of tandem repeats“ - VNTRs)
  - jednotka repetice do 25 bp, délka do 20 kb
  - telomery
- Mikrosatelity („simple tandem repeats“ - STRs)
  - jednotka repetice do 13 bp, délka do 150 bp
  - vznikají chybami při replikaci genomu
  - funkce není známá
  - využití: genetické profilování



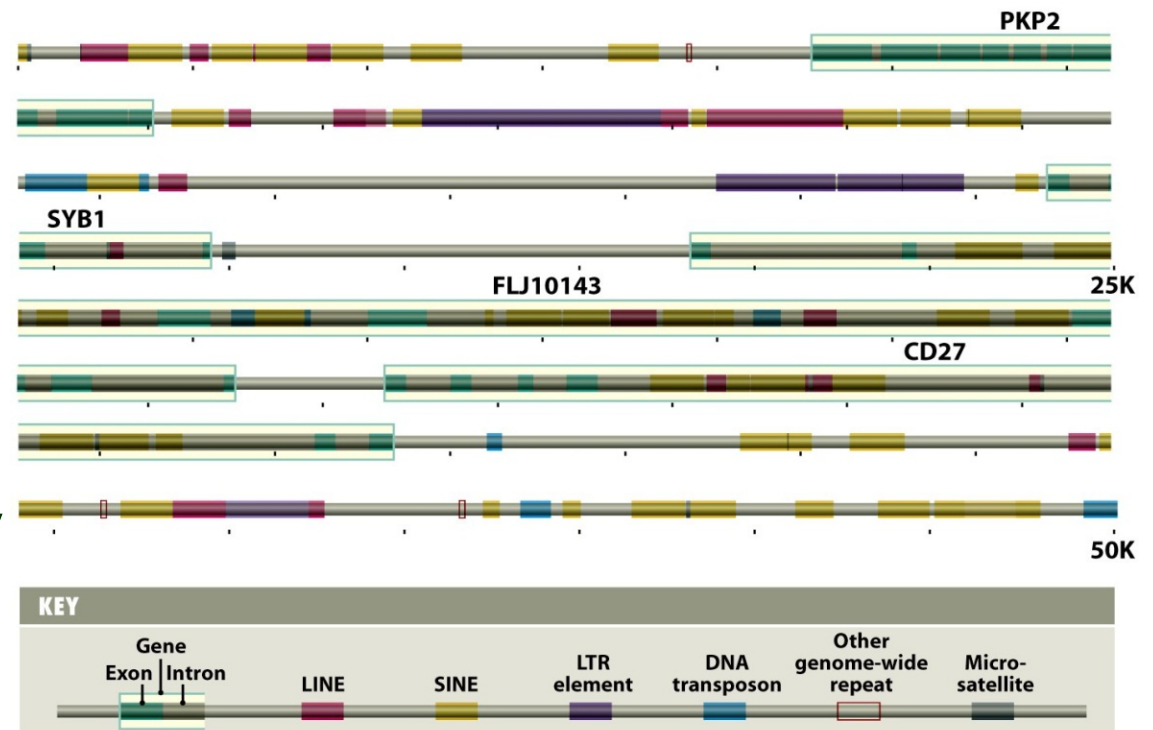
# ORGANIZACE LIDSKÉHO GENOMU



# ORGANIZACE JADERNÉHO GENOMU

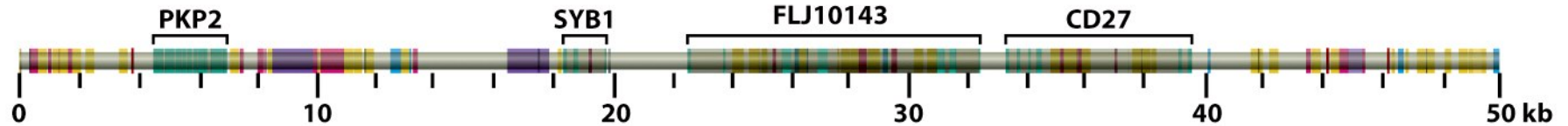
Lidský genom – 50 kb

- 4 geny
- 88 repetice
  - LINEs
  - SINEs
  - LTRs
  - DNA transpozony
- 7 mikrosatelitů (z toho 4 v intronech)
- 30% nekódující DNA bez repetice a známé funkce

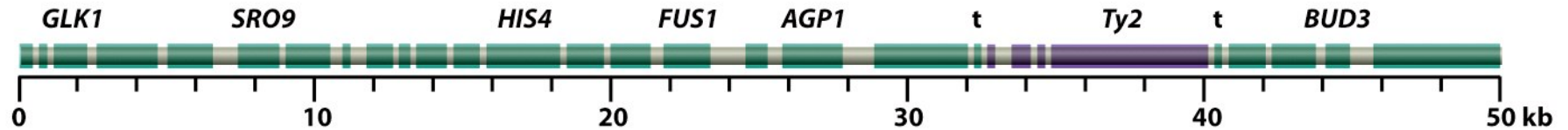


# ORGANIZACE GENOMŮ

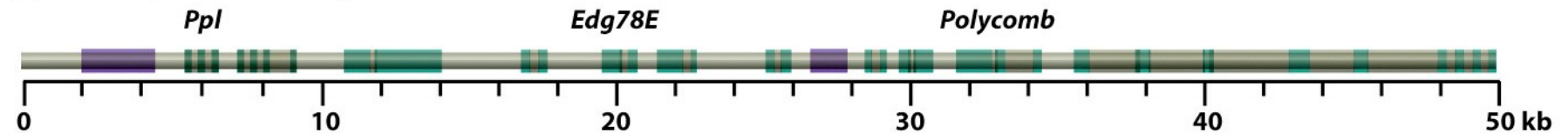
(A) Human



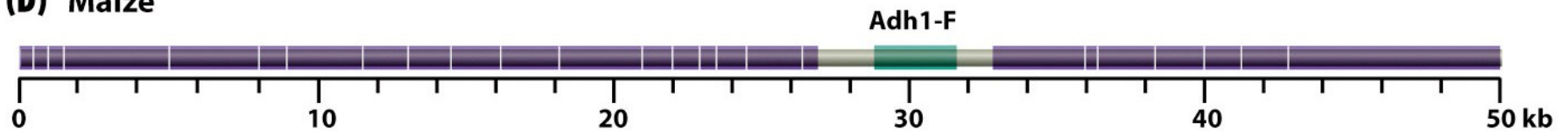
(B) *Saccharomyces cerevisiae*



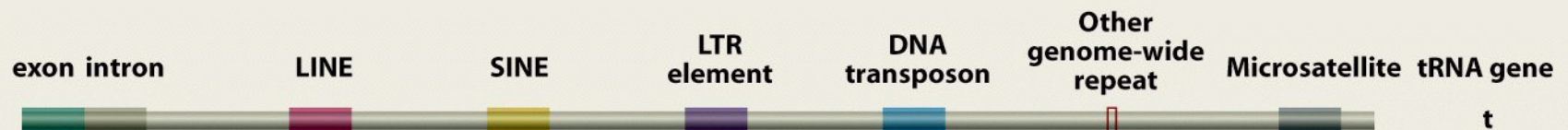
(C) *Drosophila melanogaster*



(D) Maize



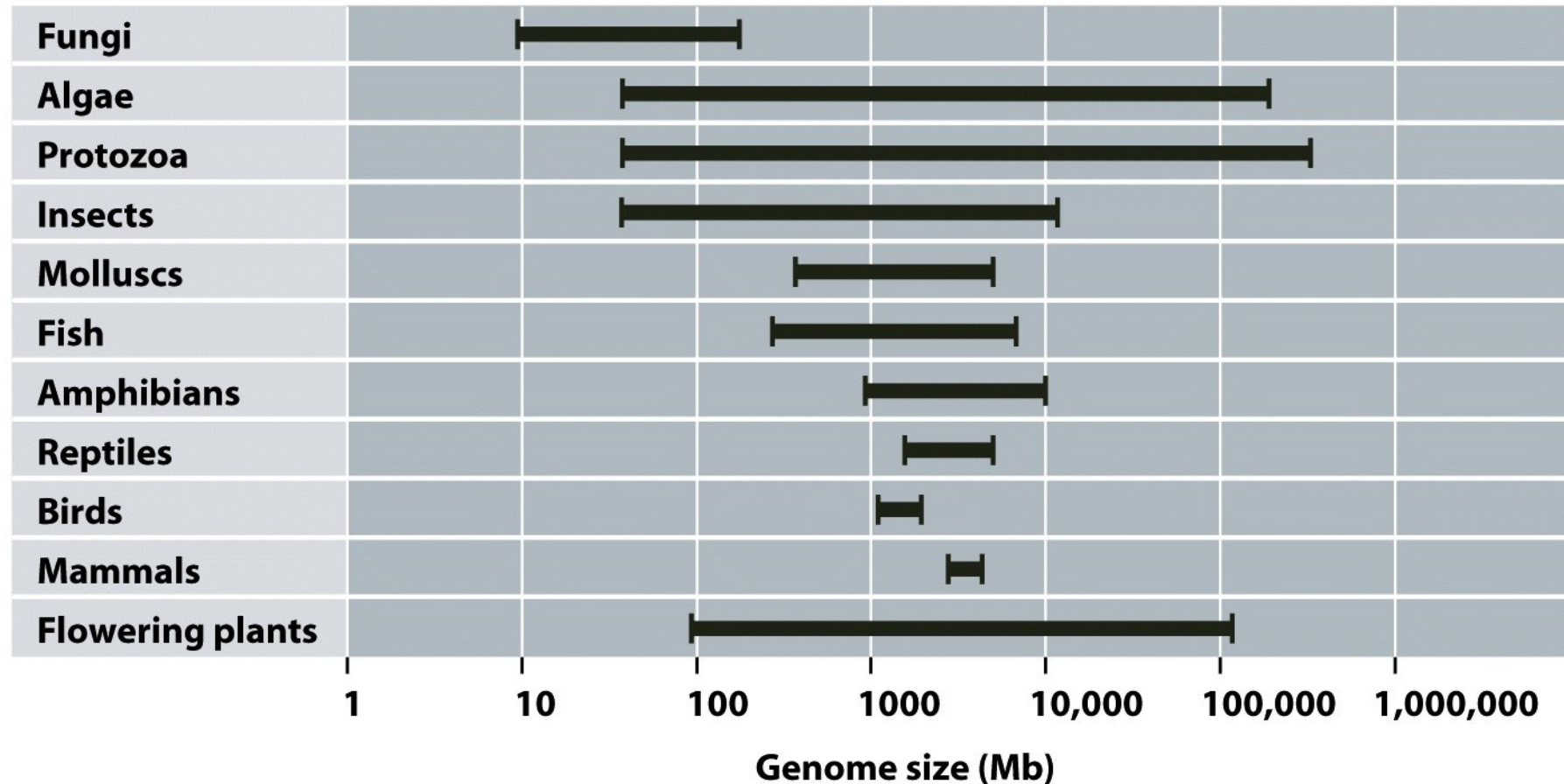
## KEY





# ORGANIZACE GENOMŮ

- C-value paradox (C-value enigma) – velikost genomu nekoreluje se složitostí organismu



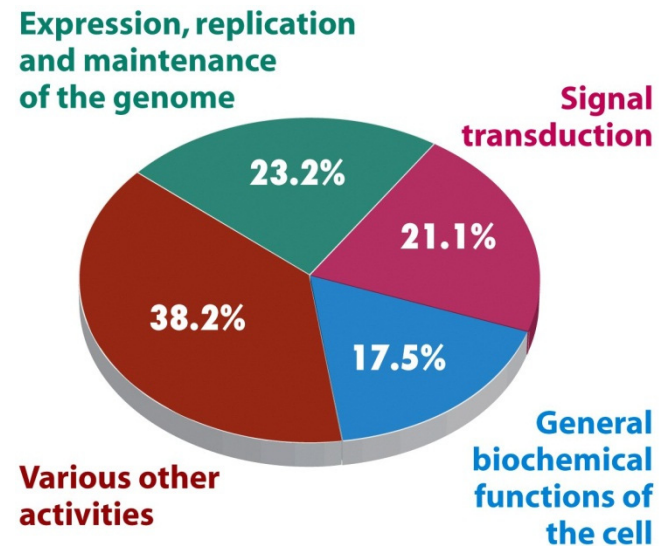
# ORGANIZACE GENOMŮ

- Velikost genomu a počet genů nekoreluje

Species	Size of genome (Mb)	Approximate number of genes
<i>Saccharomyces cerevisiae</i> (budding yeast)	12.1	6100
<i>Schizosaccharomyces pombe</i> (fission yeast)	12.5	4900
<i>Caenorhabditis elegans</i> (nematode worm)	97	19,000
<i>Arabidopsis thaliana</i> (plant)	125	25,500
<i>Drosophila melanogaster</i> (fruit fly)	180	13,600
<i>Oryza sativa</i> (rice)	466	40,000
<i>Gallus gallus</i> (chicken)	1200	20,000–23,000
<i>Homo sapiens</i> (human)	3200	30,000–40,000

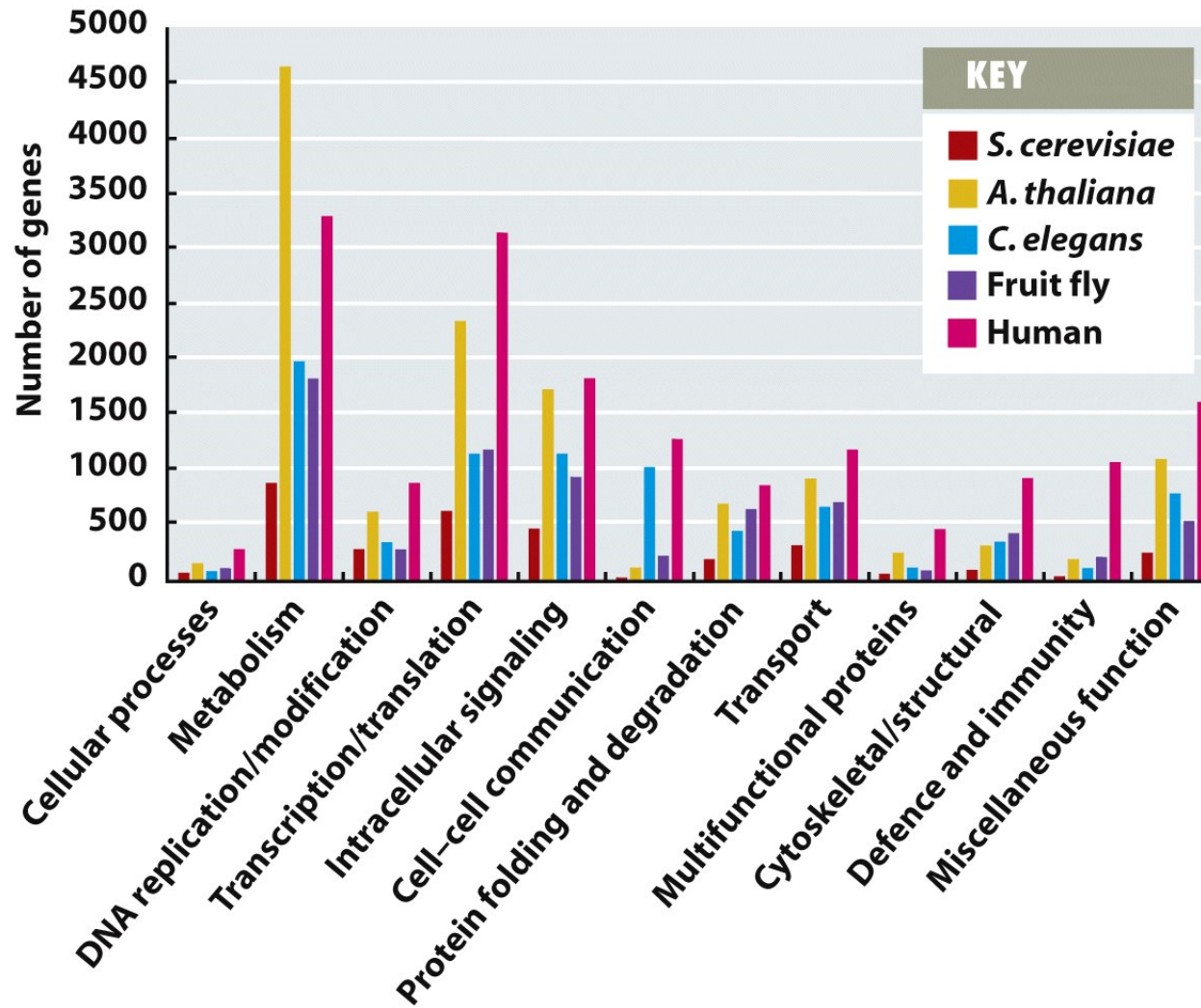
# KATALOG GENŮ

- Organizmy se sekvenovaným genomem
- Lidský katalog genů:



- co nevyčteme: Co dělá člověka člověkem?

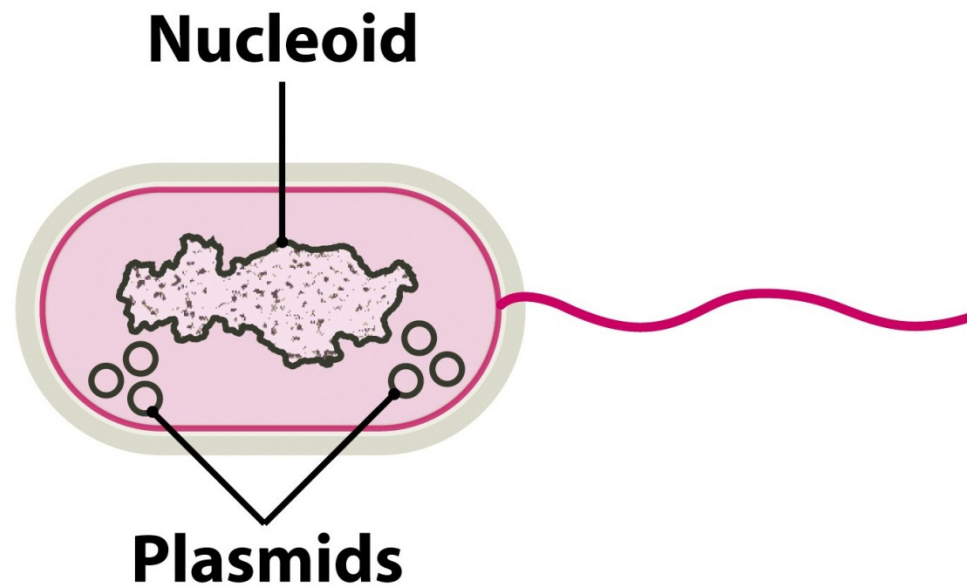
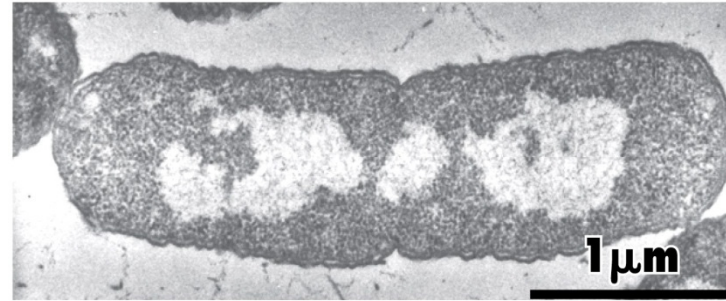
# KATALOG GENŮ



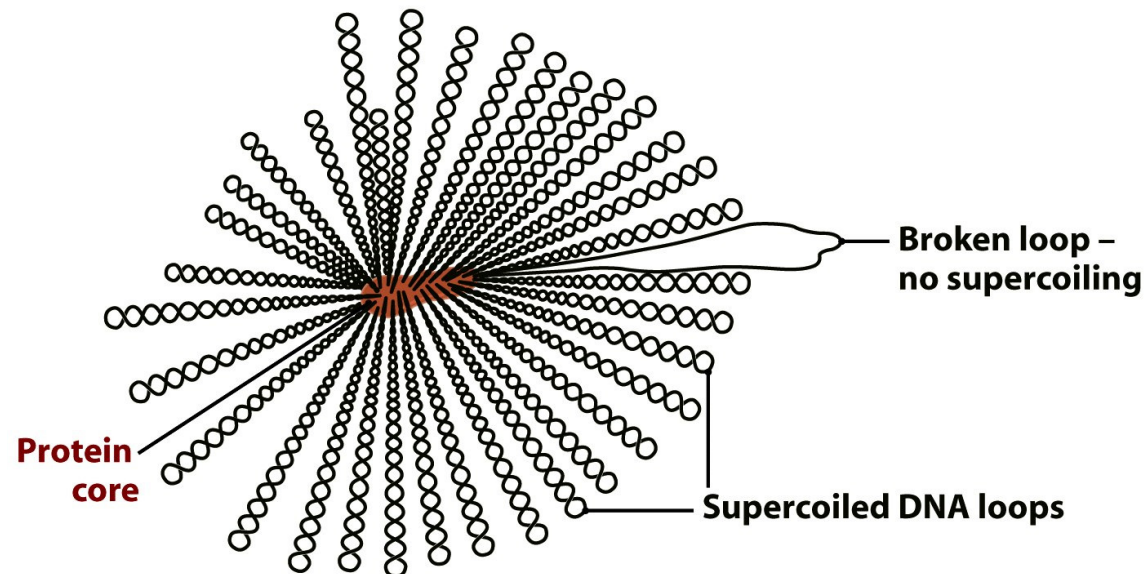
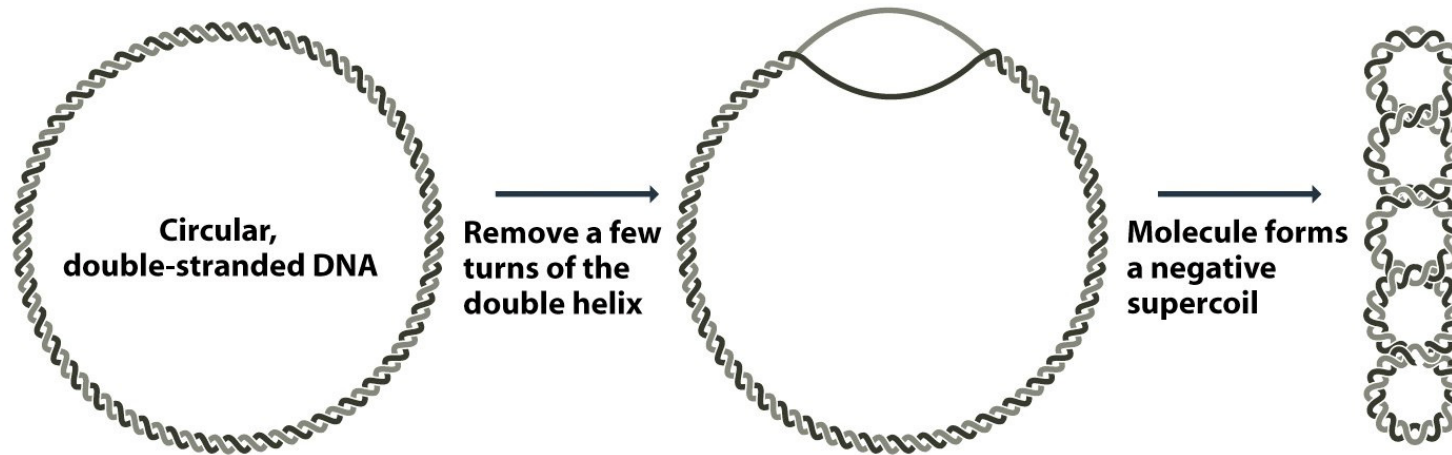
# **PROKARYOTICKÝ GENOM**

# PROKARYOTICKÝ GENOM

- Prokaryota
  - bacteria
  - archaea



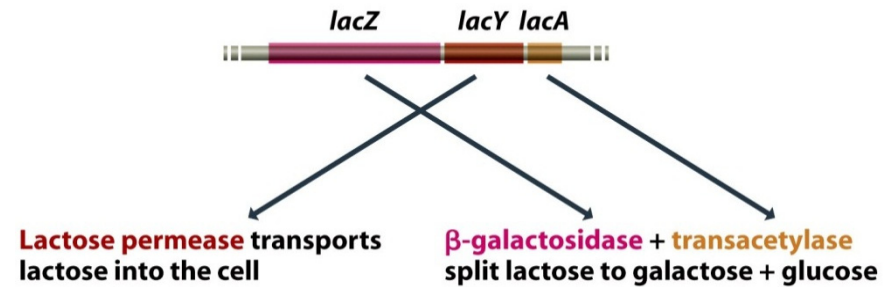
# PROKARYOTICKÝ GENOM



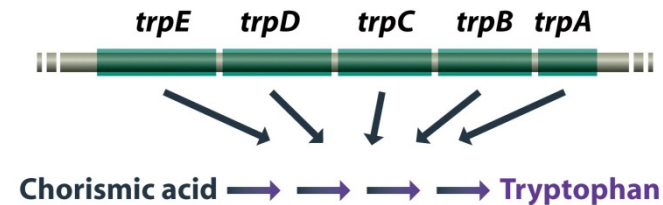


# OPERON

- laktózový operon
  - využití laktózy



- tryptofanový operon
  - stejná biochemická dráha



- *Methanococcus jannaschii* (archaea) and *Aquifex aeolicus* (bakterie)
  - rozdílné funkce





# PLAZMIDY

- dodatečná genetická informace
- adaptace na podmínky prostředí
- zvýhodnění nositele
- některé se včleňují do hlavního genomu

Type of plasmid	Gene functions	Examples
Resistance	Antibiotic resistance	Rbk of <i>Escherichia coli</i> and other bacteria
Fertility	Conjugation and DNA transfer between bacteria	F of <i>E. coli</i>
Killer	Synthesis of toxins that kill other bacteria	Col of <i>E. coli</i> , for colicin production
Degradative	Enzymes for metabolism of unusual molecules	TOL of <i>Pseudomonas putida</i> , for toluene metabolism
Virulence	Pathogenicity	Ti of <i>Agrobacterium tumefaciens</i> , conferring the ability to cause crown gall disease on dicotyledonous plants

# PROKARYOTICKÝ GENOM

Species	Genome organization		
	DNA molecules	Size (Mb)	Number of genes
<i>Escherichia coli</i> K12	One circular molecule	4.639	4405
<i>Vibrio cholerae</i> El Tor N16961	Two circular molecules		
	Main chromosome	2.961	2770
	Megaplasmid	1.073	1115
<i>Deinococcus radiodurans</i> R1	Four circular molecules		
	Chromosome 1	2.649	2633
	Chromosome 2	0.412	369
	Megaplasmid	0.177	145
	Plasmid	0.046	40
<i>Borrelia burgdorferi</i> B31	Seven or eight circular molecules, eleven linear molecules		
	Linear chromosome	0.911	853
	Circular plasmid cp9	0.009	12
	Circular plasmid cp26	0.026	29
	Circular plasmid cp32*	0.032	Not known
	Linear plasmid lp17	0.017	25
	Linear plasmid lp25	0.024	32
	Linear plasmid lp28-1	0.027	32
	Linear plasmid lp28-2	0.030	34
	Linear plasmid lp28-3	0.029	41
	Linear plasmid lp28-4	0.027	43
	Linear plasmid lp36	0.037	54
	Linear plasmid lp38	0.039	52
	Linear plasmid lp54	0.054	76
Linear plasmid lp56	0.056	Not known	

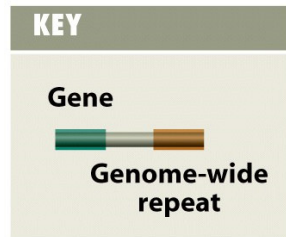
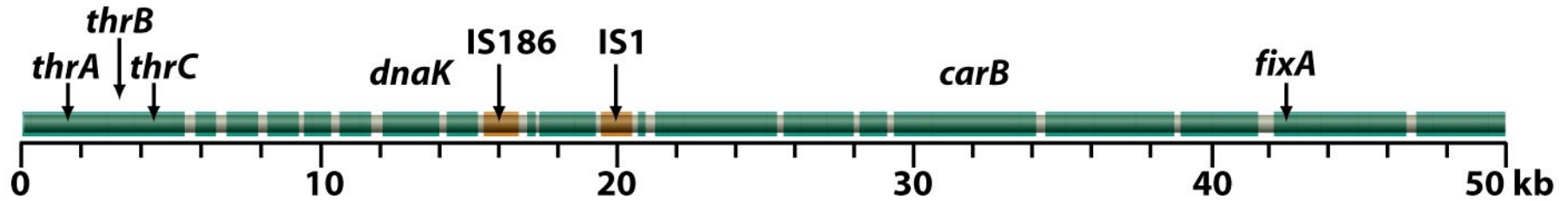
# VELIKOST PROKARYOTICKÉHO GENOMU

- většinou velikost genomu koreluje s počtem genů
- průměr: 950 genů na 1Mb

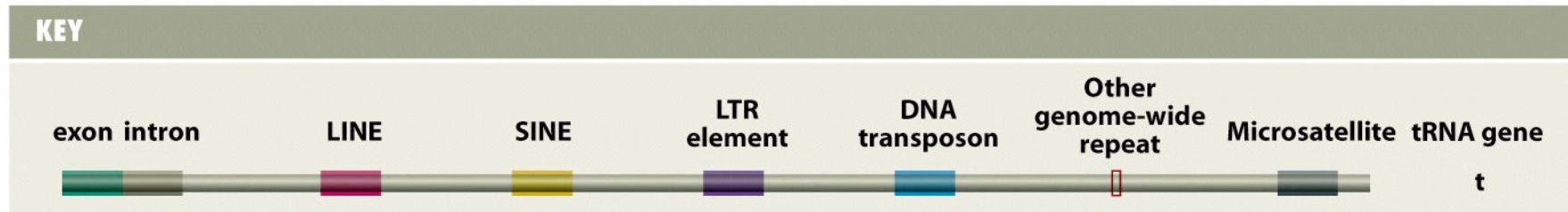
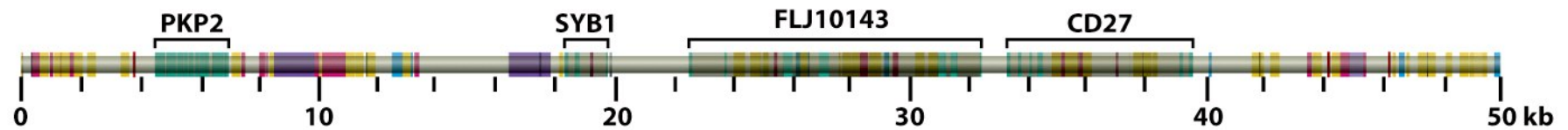
Species	Size of genome (Mb)	Approximate number of genes
<b>Bacteria</b>		
<i>Mycoplasma genitalium</i>	0.58	500
<i>Streptococcus pneumoniae</i>	2.16	2300
<i>Vibrio cholerae</i> El Tor N16961	4.03	4000
<i>Mycobacterium tuberculosis</i> H37Rv	4.41	4000
<i>Escherichia coli</i> K12	4.64	4400
<i>Yersinia pestis</i> CO92	4.65	4100
<i>Pseudomonas aeruginosa</i> PA01	6.26	5700
<b>Archaea</b>		
<i>Methanococcus jannaschii</i>	1.66	1750
<i>Archaeoglobus fulgidus</i>	2.18	2500

# PROKARYOTA vs. EUKARYOTA

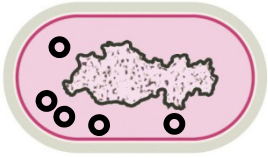
E.coli



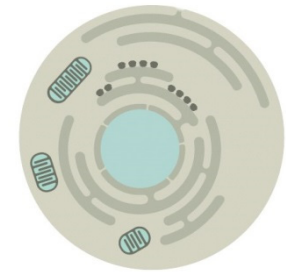
člověk



# PROKARYOTA vs. EUKARYOTA



- nukleoid
- volně v cytoplazmě
- -
- plazmidy
- kompaktní
- většina genomu kódující sekvence
- **operony**
- málo repetitivních sekvencí
- -



- chromozomy
- v jádře
- **introny**
- -
- **C-value paradox**
- většina genomu nekódující sekvence
- -
- velké množství repetitivních sekvencí
- mitochondrie a chloroplasty

# **GENOM ORGANEL**

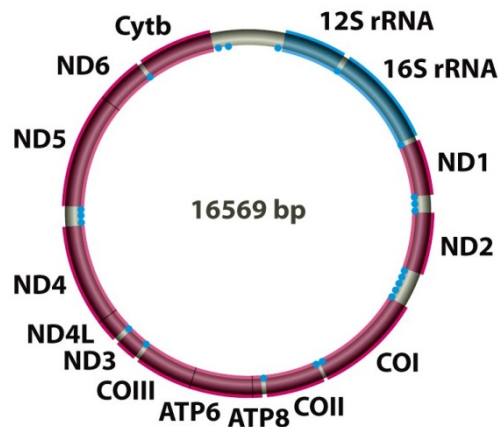
# GENOM ORGANEL

- endosymbiotická teorie vzniku organel
  - volné bakterie
  - symbióza s předchůdci eukaryotických buněk
  - endosymbióza
- 1 až 100 kopií v jedné mitochondrii
- dědičnost – jako jedna kopie ???
- transfer DNA z organel do jádra i mezi organelami
  - *Arabidopsis*
    - mitochondriální genom obsahuje jadernou DNA i chloroplastovou
    - jaderný genom obsahuje části chloroplastové i mitochondriální DNA
  - obratlovci
    - mitochondriální DNA v jaderném genomu

# MITOCHONDRIÁLNÍ GENOM

- cirkulární, ale i lineární (např. Chlamidie)
- 1 mitochondrie – 10 identických molekul = přibližně 8000 v jedné buňce (u člověka)
- rRNA, tRNA, proteiny dýchacího řetězce, ribozomální proteiny, transkripce, translace, transportní proteiny ...

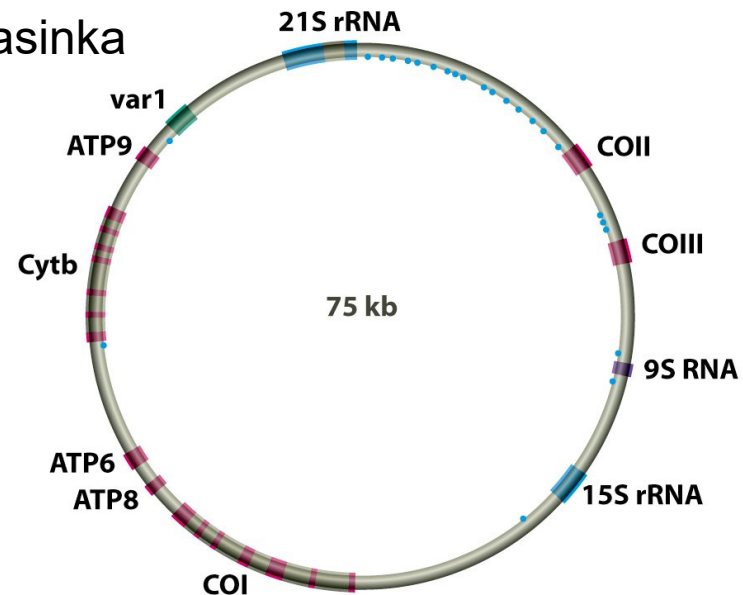
člověk



## KEY

- Respiratory complex gene
- Ribosomal RNA gene
- Transfer RNA gene

kvasinka



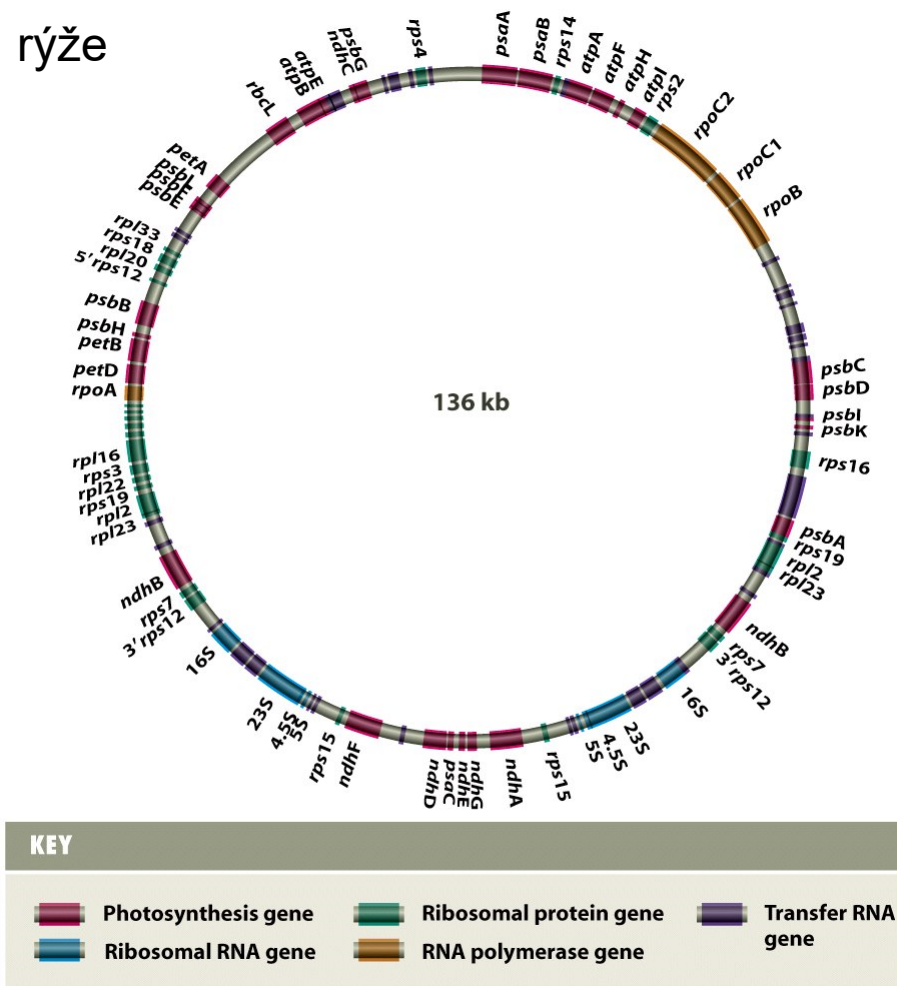
## KEY

- Respiratory complex gene
- Ribosomal RNA gene
- Ribosomal protein gene
- Transfer RNA gene
- Intron
- Other RNA gene



# CHLOROPLASTOVÝ GENOM

- většina stejná sada přibližně 200 genů
- rRNA, tRNA, ribozomální proteiny, proteiny spojené s fotosyntézou ...



# VELIKOST GENOMŮ ORGANEL

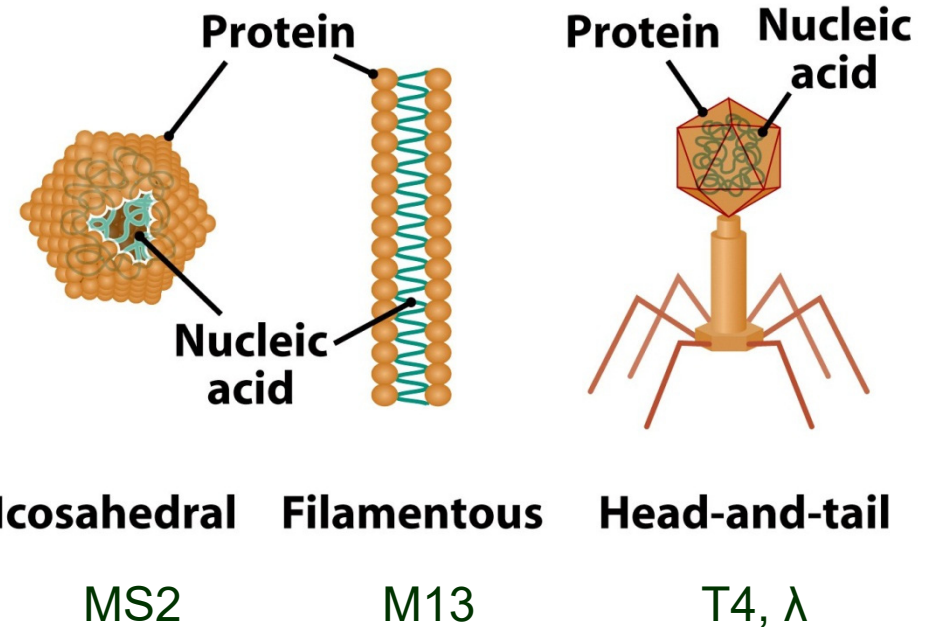
Species	Type of organism	Genome size (kb)
<b>Mitochondrial genomes</b>		
<i>Plasmodium falciparum</i>	Protozoan (malaria parasite)	6
<i>Chlamydomonas reinhardtii</i>	Green alga	16
<i>Mus musculus</i>	Vertebrate (mouse)	16
<i>Homo sapiens</i>	Vertebrate (human)	17
<i>Metridium senile</i>	Invertebrate (sea anemone)	17
<i>Drosophila melanogaster</i>	Invertebrate (fruit fly)	19
<i>Chondrus crispus</i>	Red alga	26
<i>Aspergillus nidulans</i>	Ascomycete fungus	33
<i>Reclinomonas americana</i>	Protozoa	69
<i>Saccharomyces cerevisiae</i>	Yeast	75
<i>Suillus grisellus</i>	Basidiomycete fungus	121
<i>Brassica oleracea</i>	Flowering plant (cabbage)	160
<i>Arabidopsis thaliana</i>	Flowering plant (vetch)	367
<i>Zea mays</i>	Flowering plant (maize)	570
<i>Cucumis melo</i>	Flowering plant (melon)	2500
<b>Chloroplast genomes</b>		
<i>Pisum sativum</i>	Flowering plant (pea)	120
<i>Marchantia polymorpha</i>	Liverwort	121
<i>Oryza sativa</i>	Flowering plant (rice)	136
<i>Nicotiana tabacum</i>	Flowering plant (tobacco)	156
<i>Chlamydomonas reinhardtii</i>	Green alga	195

# GENOM VIRŮ A MOBILNÍ ELEMENTY



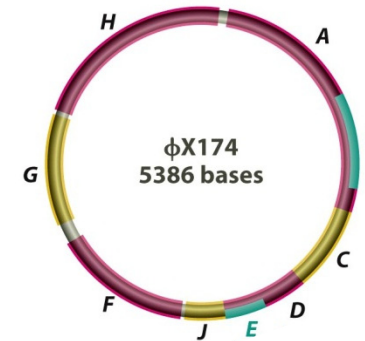
# VIROVÝ GENOM

- virus – nukleoproteinová částice
- závislé na hostiteli = parazité – potřebují ribozomy a translační aparát pro syntézu proteinového obalu viry
  - bakterií: bakteriofágy (fágy)
  - eukaryot
- virový genom
  - DNA nebo RNA
  - cirkulární nebo lineární
  - ss nebo ds
  - segmentovaný nebo nesegmentovaný



# GENOMY FÁGŮ

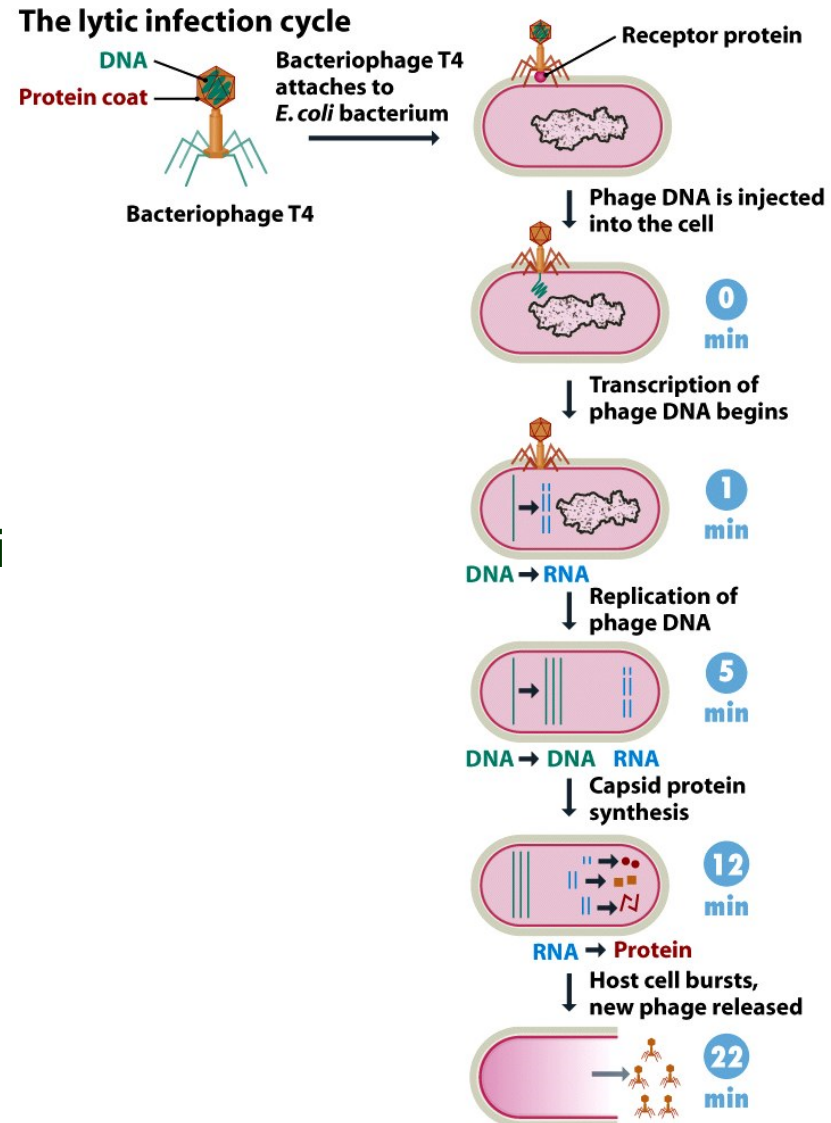
- počet genů: 3-200
- překrývající se geny
- fágy
  - lytické (virulentní), např. T4
  - lyzogenní (mírné), např. fág  $\lambda$



Phage	Host	Capsid structure	Genome structure	Genome size (kb)	Number of genes
$\lambda$	<i>Escherichia coli</i>	Head-and-tail	Double-stranded linear DNA	49.5	48
$\phi$ X174	<i>E. coli</i>	Icosahedral	Single-stranded circular DNA	5.4	11
f6	<i>Pseudomonas phaseolicola</i>	Icosahedral	Double-stranded segmented linear RNA	2.9, 4.0, 6.4	13
M13	<i>E. coli</i>	Filamentous	Single-stranded circular DNA	6.4	10
MS2	<i>E. coli</i>	Icosahedral	Single-stranded linear RNA	3.6	3
PM2	<i>Pseudomonas aeruginosa</i>	Icosahedral	Double-stranded linear DNA	10.0	approx. 21
SPO1	<i>Bacillus subtilis</i>	Head-and-tail	Double-stranded linear DNA	150	100+
T2, T4, T6	<i>E. coli</i>	Head-and-tail	Double-stranded linear DNA	166	150+
T7	<i>E. coli</i>	Head-and-tail	Double-stranded linear DNA	39.9	55+

# LYTICKÁ INFEKCE

- = virulentní, produktivní
- např. T4
- dochází k lyzi buněk
- doba latence – 22 minut
  - čas potřebný k reprodukci fága

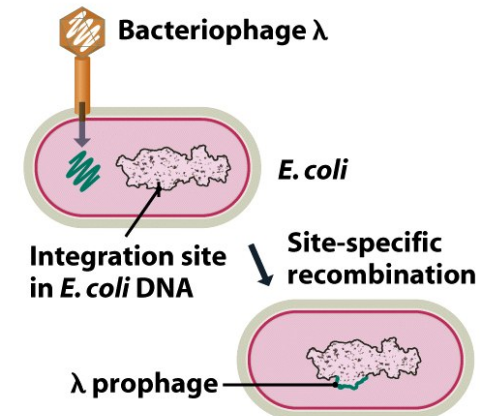




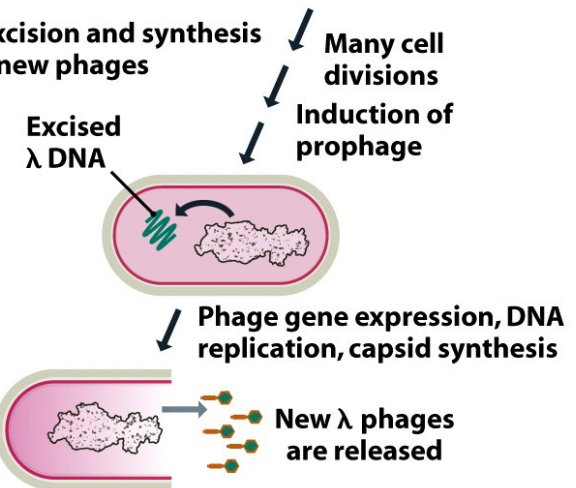
# LYZOGENNÍ INFEKCE

- = mírná, skrytá
- např. fág  $\lambda$
- okamžitě po proniknutí fágové DNA - integrace virového genomu do genomu hostitele místně specifickou rekombinací – profág
- indukce excize profága – chemické nebo fyzikální faktory – zřejmě vázané na poškození DNA

(A) Integration into the host DNA



(B) Excision and synthesis of new phages





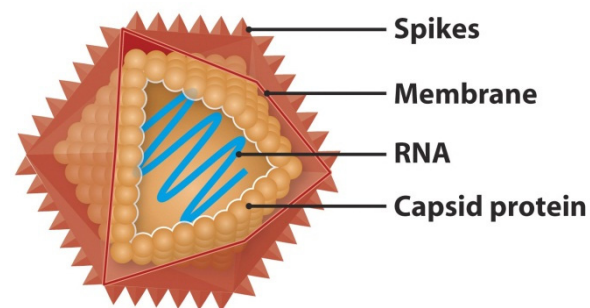
# VIRY EUKARYOT

- genomy variabilní
  - DNA, RNA; ds, ss; cirkulární, lineární; segmentované nesegmentované
- velikost genomů: 1,5-240 kb

Virus	Host	Genome structure	Genome size (kb)	Number of genes
Adenovirus	Mammals	Double-stranded linear DNA	36.0	30
Hepatitis B	Mammals	Partly double-stranded circular DNA	3.2	4
Influenza virus	Mammals	Single-stranded segmented linear RNA	22.0	12
Parvovirus	Mammals	Single-stranded linear DNA	1.6	5
Poliovirus	Mammals	Single-stranded linear RNA	7.6	8
Reovirus	Mammals	Double-stranded segmented linear RNA	22.5	22
Retroviruses	Mammals, birds	Single-stranded linear RNA	6.0–9.0	3
SV40	Monkeys	Double-stranded circular DNA	5.0	5
Tobacco mosaic virus	Plants	Single-stranded linear RNA	6.4	6
Vaccinia virus	Mammals	Double-stranded circular DNA	240	240

# VIRY EUKARYOT

- kapsid – ikozaedr nebo filamentální
- lipidická membrána – odvozená z hostitelské buňky
- rostlinné viry – většinou RNA
- lytická i lyzogenní infekce
- příklad: virové retroelementy
  - retroviry – RNA genom
  - pararetroviry – DNA genom



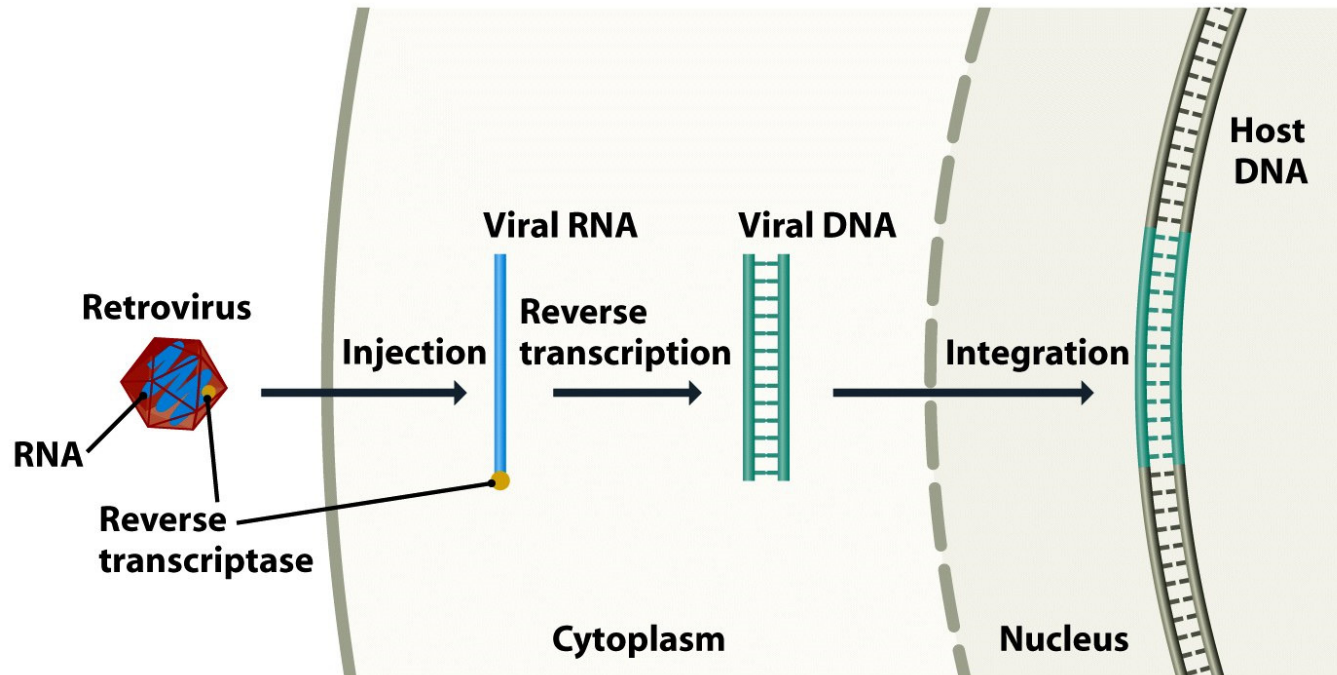
# RETROVIRY

- genom retrovirů – každý ze tří genů kóduje polyproteiny, které jsou po translaci štěpeny na dva nebo více funkčních produktů
  - *gag* - proteiny kapsidu (= group antigens)
  - *pol* – reverzní transkriptáza, integráza, proteáza
  - *env* – proteiny kapsidu (= envelope)
  - LTR – důležité regulační oblasti pro transkripci a replikaci



# RETROVIRY

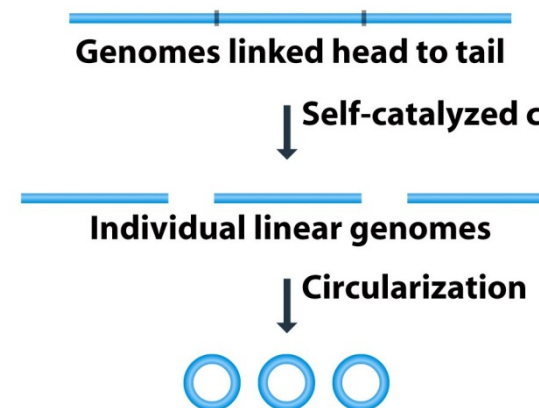
- začlenění retrovirového genomu do genomu hostitele



# VIRUSOIDY A VIROIDY

- satelitní RNA, virusoid – zejména v rostlinách
  - RNA molekula, 320-400 bází
    - satelitní RNA - nekóduje vlastní kapsid, využívá genom pomocného viru
    - virusoid – kóduje vlastní kapsid
- viroid
  - RNA molekula, 240-375 bází, žádné geny, nemá kapsid = nahá RNA
- replikace pomocí enzymů hostitele nebo pomocného viru
- schopnost samosestřihu
- zřejmě evolučně souvisí se sestřihem RNA

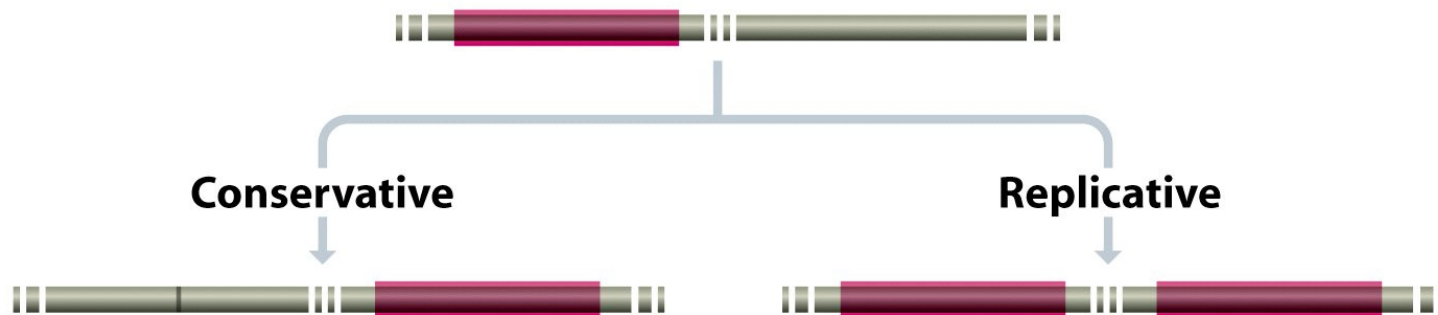
## Self-catalyzed cleavage of viroid and virusoid RNAs



# MOBILNÍ ELEMENTY

# MOBILNÍ ELEMENTY

- = transpozony
- transpozice – schopnost segmentu DNA přesouvat se v genomu z místa na místo
  - konzervativní
  - replikativní
- pomocí rekombinace



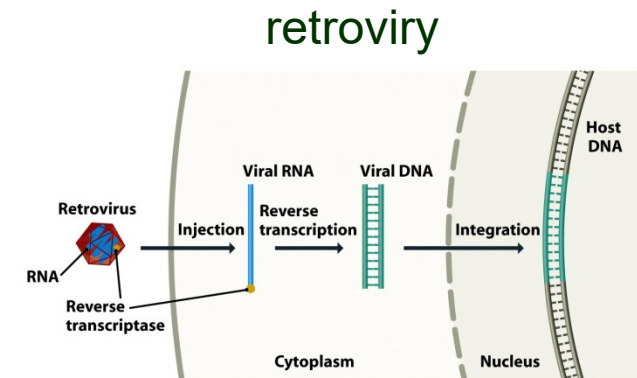
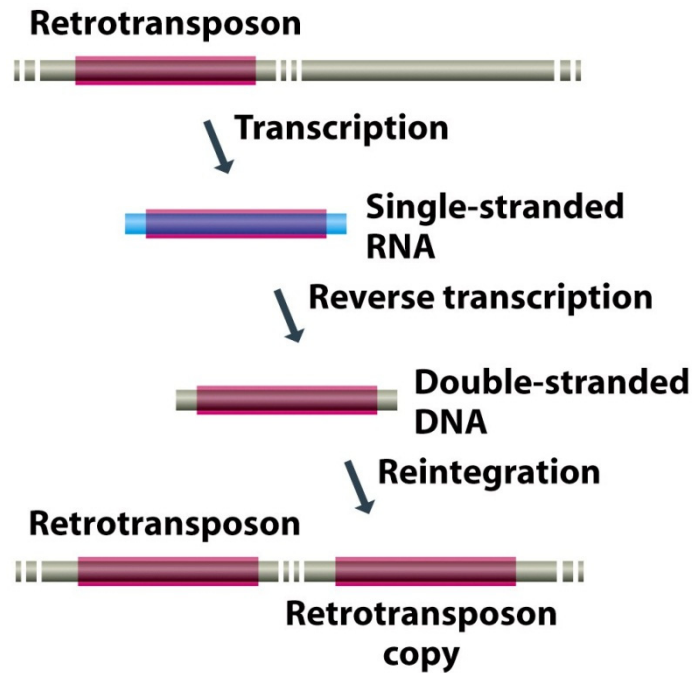


# MOBILNÍ ELEMENTY

- RNA transpozony
  - retrotranspozony s LTR
  - retrotranspozony bez LTR
- DNA transpozony
  - v prokaryotických genomech
    - Inzerční sekvence (IS)
    - Kompozitní transpozony
    - Tn3-typ transpozony
    - Transpozibilní fágy
    - ...
  - v eukaryotických genomech
    - Ac/Ds
    - Spm
    - ...

# RETROTRANSPOZONY

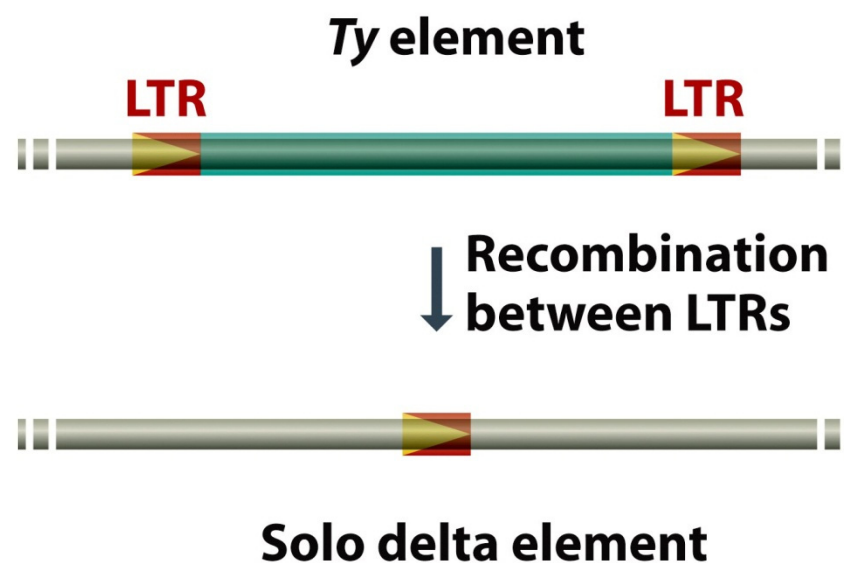
- transpozice přes RNA meziprodukt



- retrotranspozony
  - obsahující LTR sekvence
  - bez LTR sekvence

# RETROTRANSPOZONY s LTR

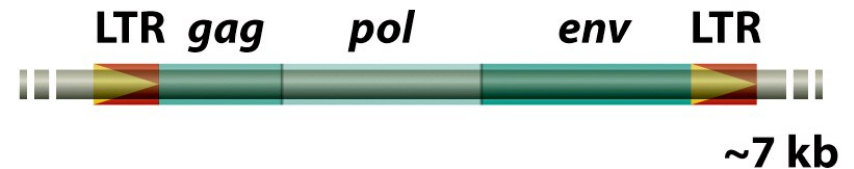
- *Ty* element
  - první objevený
  - v kvasinkách
  - 6.3 kb, 25-35 kopií
- „delta“ element
  - LTR sekvence
  - 330 bp
  - přibližně 100 kopií



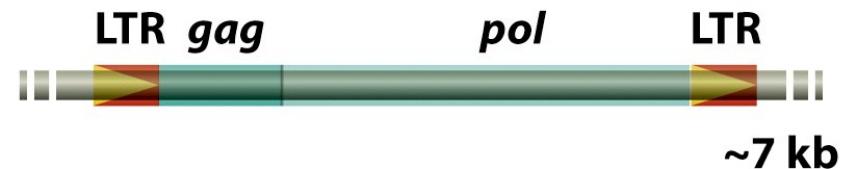
# RETROTRANSPOZONY s LTR

- *Ty1/copia*
  - nejčastější
  - chybí *env* gen
  - neschopen tvořit infekční virové částice - neschopen se dostat ven z buňky
  - schopen tvořit částice podobné virům (VLP)
- *Ty3/gypsy*
  - ekvivalent *env*
  - některé schopny tvořit infekční viry
- endogenní retroviry (ERV)
  - u člověka a savců

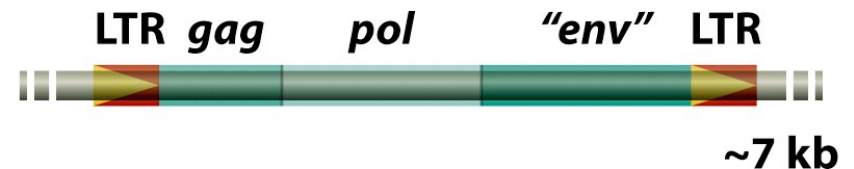
## (A) Viral retroelement



## (B) *Ty1/copia* retroelement

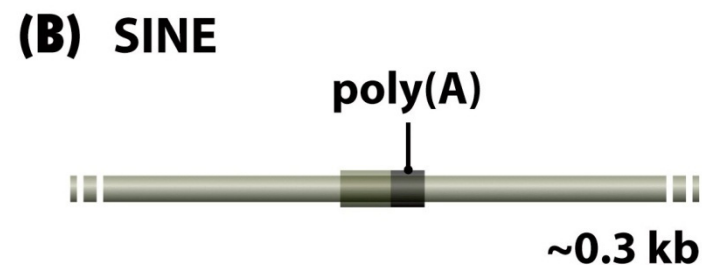
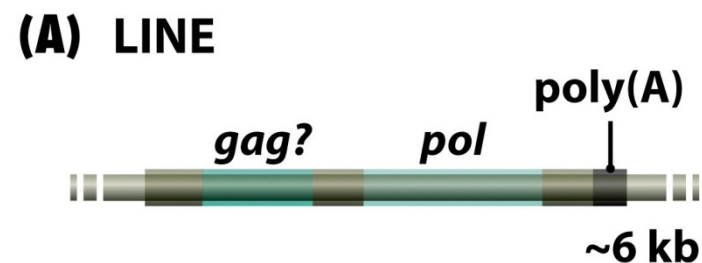


## (C) *Ty3/gypsy* retroelement



# RETROTRANSPOZONY bez LTR

- retropozony
  - LINEs (long interspersed nuclear elements)
    - *pol* gen
    - funkční reverzní transkriptáza
  - SINEs (short interspersed nuclear elements)
    - 100-400 bp
    - žádný gen
    - „půjčují“ si reverzní transkriptázu od LINE
    - např. Alu



# DNA TRANSPOZONY U PROKARYOT

- nepotřebují RNA meziprodukt
- méně časté než retrotranspozony
- IS – inzerční sekvence
  - konzervativní i replikativní transpozice
- Kompozitní transpozon
- Tn3 typ
  - nemá IS
  - replikativní transpozice
- Transpozibilní fág
  - replikativní transpozice

(A) Insertion sequence



(B) Composite transposon



(C) Tn3-type transposon

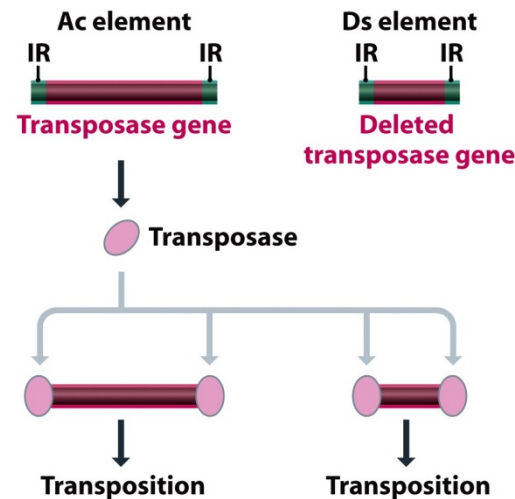


(D) Transposable phage



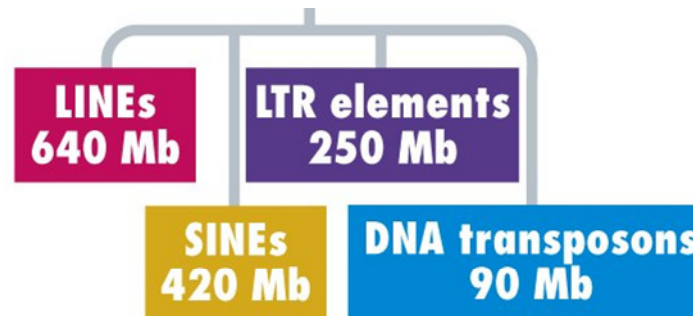
# DNA TRANSPOZONY U EUKARYOT

- Lidský genom
  - 350 000 transpozonů
  - invertované koncové repetice (ITR)
  - gen pro transponázu
  - většinou nefunkční
- Kukuřice
  - Ac/Ds elementy
  - Spm element
- *Drosophila*
  - P element





# MOBILNÍ ELEMENTY V LIDSKÉM GENOMU



Class	Family	Approximate number of copies	Fraction of genome (%)
SINE	Alu	1,200,000	10.7
	MIR	450,000	2.5
	MIR3	85,000	0.4
LINE	LINE-1	600,000	17.3
	LINE-2	370,000	3.3
	LINE-3	44,000	0.3
LTR retroelements	ERV	240,000	4.7
	MaLR	285,000	3.8
DNA transposons	MER-1	213,000	1.4
	MER-2	68,000	1.0
	Others	60,000	0.4

# **INTERNETOVÉ ZDROJE A LITERATURA**

# GOLD – Genomes OnLine Database

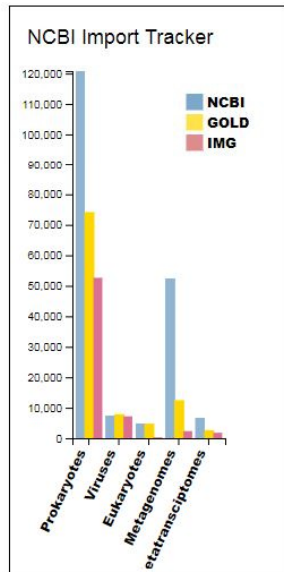


[JGI HOME](#) [LOG IN](#)

[Home](#) [Search](#) [Distribution Graphs](#) [Biogeographical Metadata](#) [Statistics](#) [GOLD Usage Policy](#) [Team](#) [Help](#) [News](#)

Studies	30 773
Biosamples	26 244
Sequencing Projects	161 667
Analysis Projects	124 972
Organisms	284 060

Excel Data file  
File last generated: 08 Nov, 2017



## Welcome to the Genomes OnLine Database

**GOLD**: Genomes Online Database, is a World Wide Web resource for comprehensive access to information regarding genome and metagenome sequencing projects, and their associated metadata, around the world.

GOLD Release v.6

### 1. Register

Register your project information and Metadata in the Genomes Online Database

Register

### 2. Annotate

Annotate your microbial genome or metagenome with IMG/ER or IMG/MER

Annotate

### 3. Publish

Publish your genome or metagenome in open access standards-supportive journal.

Publish

<b>Studies</b> <hr style="border: 1px solid yellow;"/> Metagenomic <a href="#">1 079</a> Non-Metagenomic <a href="#">29 537</a>	<b>Biosamples</b> <hr style="border: 1px solid purple;"/> Classification Ecosystems Host-associated <a href="#">8 628</a> Engineered <a href="#">3 334</a> Environmental <a href="#">14 265</a>	<b>Sequencing Projects</b> <hr style="border: 1px solid red;"/> Complete Projects <a href="#">11 825</a> Permanent Drafts <a href="#">86 079</a> Incomplete Projects <a href="#">60 754</a> Targeted Projects <a href="#">1 227</a>	<b>Analysis Projects</b> <hr style="border: 1px solid blue;"/> Genome Analysis <a href="#">88 037</a> Metagenome Analysis <a href="#">15 316</a> Metagenome - Cell Enrichment <a href="#">935</a> Metagenome - Single Particle Sort <a href="#">2 721</a> Metagenome - Assembled Genome (MAG) <a href="#">5 067</a> Metatranscriptome Analysis <a href="#">2 723</a> Combined Assembly <a href="#">136</a> Single Cell - Screened (SAG) <a href="#">2 153</a> Single Cell - Unscreened (SAG) <a href="#">1 050</a> Transcriptome Analysis <a href="#">207</a>
<b>Special Projects</b> <hr style="border: 1px solid orange;"/> Type Strain Projects <a href="#">6 028</a> Strains at Genbank <a href="#">4 786</a> GEBA Projects <a href="#">3 151</a> HMP Projects <a href="#">2 913</a>	<b>Projects with Genbank Data</b> <hr style="border: 1px solid yellow;"/> Seq. Projects <a href="#">83 650</a> Archaeal Projects <a href="#">766</a> Bacterial Projects <a href="#">70 110</a> Eukaryal Projects <a href="#">4 269</a> Viral Projects <a href="#">8 505</a>	<b>JGI Projects</b> <hr style="border: 1px solid blue;"/> JGI Studies <a href="#">1 251</a> JGI Biosamples <a href="#">12 712</a> JGI Sequencing Projects <a href="#">69 308</a> JGI Analysis Projects <a href="#">32 123</a>	<b>Organisms</b> <hr style="border: 1px solid green;"/> Organisms <a href="#">282 609</a> Archaea <a href="#">2 416</a> Bacteria <a href="#">250 332</a> Eukarya <a href="#">20 956</a> Viruses <a href="#">8 876</a> Bacterial Type Strains <a href="#">10 695</a> Archaeal Type Strains <a href="#">416</a>

verze z 9.11.2017

<https://gold.jgi-psf.org/index>

# GOLD – Genomes OnLine Database



**GOLD**

GENOMES ONLINE DATABASE

[JGI HOME](#) [LOG IN](#)

[Home](#) [Search](#) [Distribution Graphs](#) [Biogeographical Metadata](#) [Statistics](#) [GOLD Usage Policy](#) [Team](#) [Help](#) [News](#)

Studies	<a href="#">30 773</a>
Biosamples	<a href="#">26 244</a>
Sequencing Projects	<a href="#">161 667</a>
Analysis Projects	<a href="#">124 972</a>
Organisms	<a href="#">284 060</a>

Current Filters:: None Set

GOLD Study ID 	Study Name 	Add Date 
<a href="#">Gs0132309</a>	Genome sequencing of <i>Corynebacterium glutamicum</i> TQ2223	2017-11-08
<a href="#">Gs0132308</a>	Genome sequencing of <i>Bacillus subtilis</i> GS 188	2017-11-08
<a href="#">Gs0132307</a>	Genome sequencing of <i>Bacillus thuringiensis</i> BM-BT15426	2017-11-08
<a href="#">Gs0132306</a>	Population dynamics of <i>Staphylococcus aureus</i> in Cystic Fibrosis patients	2017-11-08
<a href="#">Gs0132305</a>	Genome sequencing of <i>Pseudomonas aeruginosa</i> FA-HZ1	2017-11-08
<a href="#">Gs0132304</a>	Genome sequencing of <i>Legionella</i> sp. CDC-D5610	2017-11-08
<a href="#">Gs0132303</a>	Sequencing of plant-associated <i>Rhodococcus</i> strains	2017-11-08
<a href="#">Gs0132302</a>	Genome sequencing of <i>Bosea</i> sp. AS-1	2017-11-08
<a href="#">Gs0132301</a>	Genome sequencing of <i>Lactobacillus paracasei</i> TK1501	2017-11-08
<a href="#">Gs0132300</a>	Genome sequencing of <i>Aeromonas salmonicida</i> S44 Atlantic salmon	2017-11-08
<a href="#">Gs0132299</a>	Genome sequencing of <i>Aeromonas salmonicida</i> S68	2017-11-08
<a href="#">Gs0132298</a>	Genome sequencing of <i>Altererythrobacter</i> sp. C9-11	2017-11-08
<a href="#">Gs0132297</a>	Genome sequencing of <i>Bacillus velezensis</i> 157	2017-11-08
<a href="#">Gs0132296</a>	Genome sequencing of <i>Bacillus cereus</i> C1L	2017-11-08
<a href="#">Gs0132295</a>	Genome sequencing of <i>Campylobacter jejuni</i> 11168H/araE	2017-11-08
<a href="#">Gs0132294</a>	Genome sequencing of <i>Campylobacter jejuni</i> 11168H/lacY	2017-11-08
<a href="#">Gs0132293</a>	Genome sequencing of <i>Macroccoccus caseolyticus</i> IMD0819	2017-11-08
<a href="#">Gs0132292</a>	Genome sequencing of <i>Xanthomonas citri</i> pv. <i>mangiferae</i> indicae XC01	2017-11-08
<a href="#">Gs0132291</a>	Genome sequencing of <i>Prosthecochloris</i> sp. GSB1	2017-11-08
<a href="#">Gs0132290</a>	Genome sequencing of <i>Lactobacillus agilis</i> La3	2017-11-08
<a href="#">Gs0132289</a>	Genome sequencing of <i>Xanthomonas citri</i> pv. <i>malvacearum</i> XcmN1003	2017-11-08
<a href="#">Gs0132288</a>	Genome sequencing of <i>Mycobacterium chimaera</i> ZUERICH-1	2017-11-08
<a href="#">Gs0132287</a>	Genome sequencing of <i>Bacillus cereus</i> D12_2	2017-11-08
<a href="#">Gs0132286</a>	Genome sequencing of <i>Bacillus cereus</i> K8	2017-11-08
<a href="#">Gs0132285</a>	Genome sequencing of <i>Corynebacterium stationis</i> ATCC 21170	2017-11-08

[ 1 - 25 ] of 30,773  Show  results.

# GOLD – Genomes OnLine Database

- verze z roku 2014

**GOLD**  
Genomes Online Database

Home Search Distribution Graphs Biogeographical Metadata Statistics References Team Help News

JGI/IMG Gold User Log in

**Welcome to the Genomes OnLine Database**  
GOLD: Genomes Online Database, is a World Wide Web resource for comprehensive access to information regarding genome and metagenome sequencing projects, and their associated metadata, around the world.

**GOLD Release v.5**

Studies <sup>+</sup>	20354
Biosamples <sup>+</sup>	59036
Sequencing Projects <sup>+</sup>	59171
Analysis Projects <sup>+</sup>	44919

[Download Excel Data file](#)

**Studies**

- Metagenomic [545](#)
- Non-Metagenomic [19809](#)

**Biosamples**

[Classification](#)

- Ecosystems
- Host-associated [11813](#)
- Engineered [1655](#)
- Environmental [6740](#)


**Projects**

- Complete Projects [6649](#)
- Permanent Drafts [23552](#)
- Incomplete Projects [26572](#)
- Targeted Projects [1404](#)

**Organisms**

- Organisms [53794](#)
- Archaea [935](#)
- Bacteria [39183](#)
- Eukarya [9250](#)


**1. Register**



Register your project information and Metadata in the Genomes Online Database

[Register](#)


**2. Annotate**



Annotate your microbial genome or metagenome with IMG/ER or IMG/MER

[Annotate](#)

**3. Publish**

 Standards in Genomic Sciences

Publish your genome or metagenome in open access standards-supportive journal.

[Publish](#)



# GOLD – Genomes OnLine Database

- verze z roku 2012

**GOLD** Genomes Online Database

Home

Version 4.0

Last update: 2012-11-20  
Total # of genomes: **18893**

## Welcome to the Genomes OnLine Database

**GOLD:** Genomes Online Database, is a World Wide Web resource for comprehensive access to information regarding genome and metagenome sequencing projects, and their associated metadata, around the world.

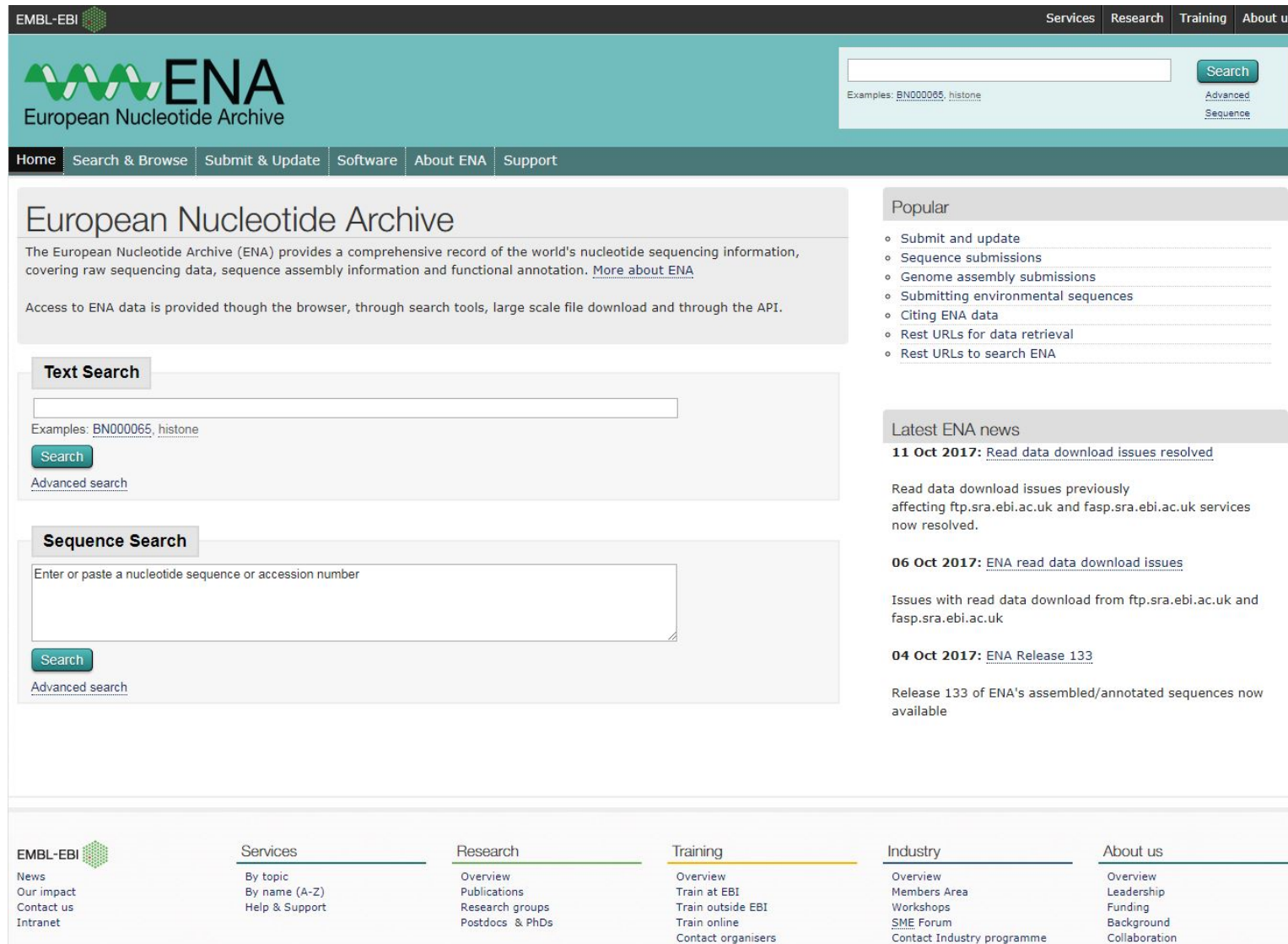
Metagenomes	Isolate Genomes	Genome Distribution
<b>Classification</b> <ul style="list-style-type: none"><li>• <b>Studies:</b> 345</li><li>• <b>Samples:</b> 2145</li></ul>	Complete Projects: <b>3811</b> Incomplete Projects: <b>15038</b> Targeted Projects: <b>1743</b>	<ul style="list-style-type: none"><li>• <a href="#">Project Type</a></li><li>• <a href="#">Sequencing Status</a></li><li>• <a href="#">Phylogenetic</a></li></ul>

1. Register	2. Annotate	3. Publish
<p>Register your project information and Metadata in Genomes Online Database</p> <p><a href="#">Register</a></p>	<p>Annotate your microbial genome or metagenome with IMG/ER or IMG/MER</p> <p><a href="#">Annotate</a></p>	<p>Publish your genome or metagenome in open access standards-supportive journal.</p> <p><a href="#">Publish</a></p>

©2012 The Regents of the University of California  
[Disclaimer](#) | [Credits](#)

U.S. DEPARTMENT OF **ENERGY** Office of Science

# ENA – European Nucleotide Archive



The screenshot shows the ENA website homepage. At the top, there is a navigation bar with 'Services', 'Research', 'Training', and 'About us'. Below this is the ENA logo and a search bar with a 'Search' button. The main content area is divided into several sections: a header for 'European Nucleotide Archive' with a brief description and a link to 'More about ENA'; a 'Text Search' section with a search input field and a 'Search' button; a 'Sequence Search' section with a larger search input field and a 'Search' button; a 'Popular' section with a list of links; and a 'Latest ENA news' section with three news items. At the bottom, there is a footer with 'EMBL-EBI' logo and six columns of links for 'Services', 'Research', 'Training', 'Industry', and 'About us'.

EMBL-EBI

Services Research Training About us

ENA  
European Nucleotide Archive

Search

Examples: BN000065, histone

Advanced Search Sequence

Home Search & Browse Submit & Update Software About ENA Support

## European Nucleotide Archive

The European Nucleotide Archive (ENA) provides a comprehensive record of the world's nucleotide sequencing information, covering raw sequencing data, sequence assembly information and functional annotation. [More about ENA](#)

Access to ENA data is provided through the browser, through search tools, large scale file download and through the API.

### Text Search

Search

Examples: BN000065, histone

Advanced search

### Sequence Search

Enter or paste a nucleotide sequence or accession number

Search

Advanced search

### Popular

- Submit and update
- Sequence submissions
- Genome assembly submissions
- Submitting environmental sequences
- Citing ENA data
- Rest URLs for data retrieval
- Rest URLs to search ENA

### Latest ENA news

**11 Oct 2017:** [Read data download issues resolved](#)

Read data download issues previously affecting ftp.sra.ebi.ac.uk and fasp.sra.ebi.ac.uk services now resolved.

**06 Oct 2017:** [ENA read data download issues](#)

Issues with read data download from ftp.sra.ebi.ac.uk and fasp.sra.ebi.ac.uk

**04 Oct 2017:** [ENA Release 133](#)

Release 133 of ENA's assembled/annotated sequences now available

EMBL-EBI

News  
Our impact  
Contact us  
Intranet

Services  
By topic  
By name (A-Z)  
Help & Support

Research  
Overview  
Publications  
Research groups  
Postdocs & PhDs

Training  
Overview  
Train at EBI  
Train outside EBI  
Train online  
Contact organisers

Industry  
Overview  
Members Area  
Workshops  
SME Forum  
Contact Industry programme


About us  
Overview  
Leadership  
Funding  
Background  
Collaboration

<http://www.ebi.ac.uk/ena>



# ENA – European Nucleotide Archive

EMBL-EBI [Services](#) [Research](#) [Training](#) [About us](#)

 **ENA**  
European Nucleotide Archive

Search:  [Search](#)  
Examples: [BN000066](#), [histone](#) [Advanced](#) [Sequence](#)

[Home](#) [Search & Browse](#) [Submit & Update](#) [Software](#) [About ENA](#) [Support](#)

Search results for *arabidopsis* [Show more data from EMBL-EBI](#)

- Assembly**  
Assembly (710)
  - Assembly (710 results found)**  
GCA\_000001735.1 TAIR10 assembly for Arabidopsis thaliana  
[View all 710 results](#)
- Sequence**  
Sequence (Update) (201)  
Sequence (Release) (3,037,417)
  - Sequence (Update) (201 results found)**  
KU233416 Arabidopsis thaliana isolate Lerik-1 hypothetical protein (AT5G38460) gene, exon 4 and partial cds.  
[View all 201 results](#)
  - Sequence (Release) (3,037,417 results found)**  
AE005172 Arabidopsis thaliana chromosome 1 top arm, complete sequence.  
[View all 3,037,417 results](#)
- Contig set**  
Genome assembly contig set (490)
  - Genome assembly contig set (490 results found)**  
AFMZ01000000 Arabidopsis thaliana, WGS project AFMZ01000000 data  
[View all 490 results](#)
- Coding**  
Coding (Update) (2,640)  
Coding (Release) (2,415,887)
  - Coding (Update) (2,640 results found)**  
APU50275 Arabidopsis thaliana (thale cress) partial amidotransferase 1  
[View all 2,640 results](#)
  - Coding (Release) (2,415,887 results found)**  
AED92956 Arabidopsis thaliana (thale cress) P-type ATPase of Arabidopsis 2  
[View all 2,415,887 results](#)
- Non-coding**  
Non-coding (Release) (110,133)  
Non-coding (Update) (1,260)
  - Non-coding (Release) (110,133 results found)**  
CP002684.1:23655587..23657382:misc\_RNA Arabidopsis thaliana (thale cress) ARI6  
[View all 110,133 results](#)
- Read**  
Experiment (23,798)  
Run (31,225)
- Study**  
Study (2,017)  
Study (Sequence) (4,797)
- Taxon**  
Taxon (55)
- Sample**  
Sample (35,915)
- Submission**  
Submission (Read/Analysis) (78)
- About**  
ENA (2)

# NCBI National Center for Biotechnology Information

NCBI Resources How To Sign in to NCBI

NCBI National Center for Biotechnology Information

All Databases Search

- NCBI Home
- Resource List (A-Z)
- All Resources
- Chemicals & Bioassays
- Data & Software
- DNA & RNA
- Domains & Structures
- Genes & Expression
- Genetics & Medicine
- Genomes & Maps
- Homology
- Literature
- Proteins
- Sequence Analysis
- Taxonomy
- Training & Tutorials
- Variation


## Welcome to NCBI

The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information.

[About the NCBI](#) | [Mission](#) | [Organization](#) | [NCBI News & Blog](#)


### Submit

Deposit data or manuscripts into NCBI databases




### Download

Transfer NCBI data to your computer




### Learn

Find help documents, attend a class or watch a tutorial




### Develop

Use NCBI APIs and code libraries to build applications




### Analyze

Identify an NCBI tool for your data analysis task



### Research

Explore NCBI research and collaborative projects



## Popular Resources

- PubMed
- Bookshelf
- PubMed Central
- PubMed Health
- BLAST
- Nucleotide
- Genome
- SNP
- Gene
- Protein
- PubChem

## NCBI News & Blog

Twelve new NCBI annotations in RefSeq for otter, oyster & more  
07 Nov 2017

In September and October, the NCBI Eukaryotic Genome Annotation  
Cake, Poetry and Success Stories: NCBI Celebrates 10 Years of dbGaP  
06 Nov 2017

For the past decade, dbGaP, the database of Genotypes and Phenotypes  
November 8 NCBI Minute: New API keys for better E-utilities & EDirect access to NCBI data  
02 Nov 2017

On Wednesday, November 8, 2017, we

[More...](#)

<https://www.ncbi.nlm.nih.gov/>

# NCBI

Genome

Genome ▾

Search

[Limits](#) [Advanced](#)

[Help](#)



## Genome

This resource organizes information on genomes including sequences, maps, chromosomes, assemblies, and annotations.

### Using Genome

[Help](#)

[Browse by Organism](#)

[Download / FTP](#)

[Download FAQ](#)

[Submit a genome](#)

### Custom resources

[Human Genome](#)

[Microbes](#)

[Organelles](#)

[Viruses](#)

[Prokaryotic reference genomes](#)

### Other Resources

[Assembly](#)

[BioProject](#)

[BioSample](#)

[Map Viewer](#)

[Genome Data Viewer](#) **NEW**

### Genome Tools

[BLAST the Human Genome](#)

[Microbial Nucleotide BLAST](#)

### Genome Annotation and Analysis

[Eukaryotic Genome Annotation](#)

[Prokaryotic Genome Annotation](#)

[PASC \(Pairwise Sequence Comparison\)](#)

### External Resources

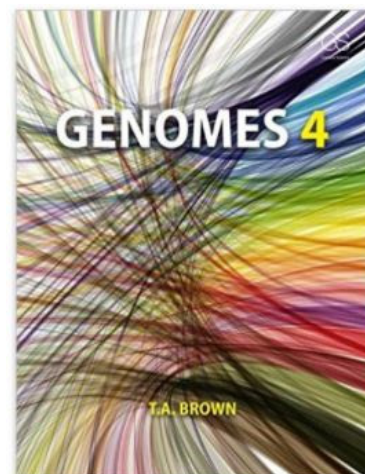
[GOLD - Genomes Online Database](#)

[Bacteria Genomes at Sanger](#)

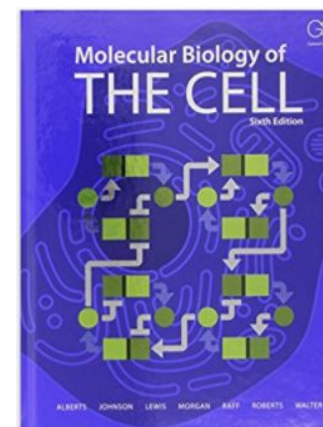
[Ensembl](#)

# LITERATURA

- T.A.Brown: Genomes



- B.Alberts and col.:  
Molecular Biology of the Cell



- G.Gibson and S.V.Muse:  
A Primer of Genome Science



+ internet, odborné články ...

# SHRNUTÍ

- Eukaryotický jaderný genom
  - chromozomy
  - geny
  - intergenové sekvence
  - katalog genů
- Prokaryotický genom
  - nukleoid
  - plazmidy
- Genom mitochondrií a chloroplastů
- Virové genomy
  - viry bakterií – fágy
  - viry eukaryot
- Mobilní elementy
  - RNA transpozony
  - DNA transpozony
- Internetové zdroje a literatura