

# Definice genu

= základní jednotka genetické informace zapsaná v NK

## Podle šíře definice:

1. všechny sekvence DNA potřebné k syntéze proteinu nebo RNA, tedy i regulační a signální sekvence (nejširší)
2. transkribované sekvence (nezahrnuje regulační oblasti)
3. úseky přímo kódující peptid (nejužší) nebo pořadí bází ve funkčních molekulách RNA

## Typy genů (širší definice):

Geny strukturní, geny pro RNA a geny-regulační sekvence

# Historie konceptu genu

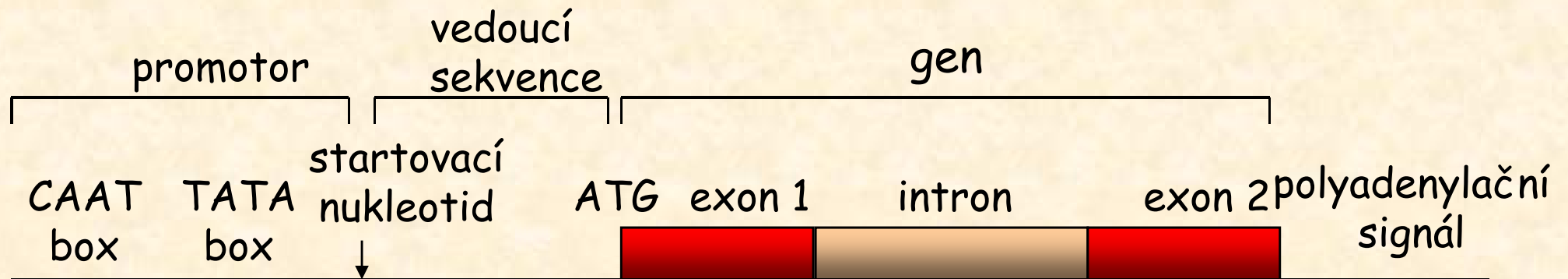
- **Mendel, Bateson**: faktory určující vlastnosti
- **Morgan**: geny jsou uspořádány lineárně, vazbové skupiny
- **Johansen (1909)**: zavedl pojem „gen“
- **Griffith (1928)**: transformační experimenty
- **Avery, MacLeod a McCarthy (1944)**: substancí zodpovědnou za transformaci je DNA
- **Hershey a Chaseová (1952)**: genetickou informaci nese DNA
- pád Caspersonovy teorie - proteinové hypotéze genu/  
tetranukleotidová teorie DNA
- **Beadle a Tatum (1941)**: „one gene - one enzyme“
- **Watson a Crick (1953)**: struktura DNA
- **Crick (1958)**: centrální dogma MB a teorie proteosyntézy
- **Meselson a Stahl (1958)** semikonzervativní replikace
- **Jacob a Monod (1961)**: operonová teorie
- **Nirenberg, Khorana, Ochoa (1966)** genetický kód

# Každý gen vzniká z genu

- geny jsou si podobné, duplikace a postupná divergence genů
- genové rodiny a nadrodiny, genealogické stromy
- počet genů u eukaryot: 10 000 - 40 000
- počet základních modulů malý: stovky-max tisíce vzájemně nepříbuzných exonů, nejmenší jsou genové moduly

# STRUKTURA GENU

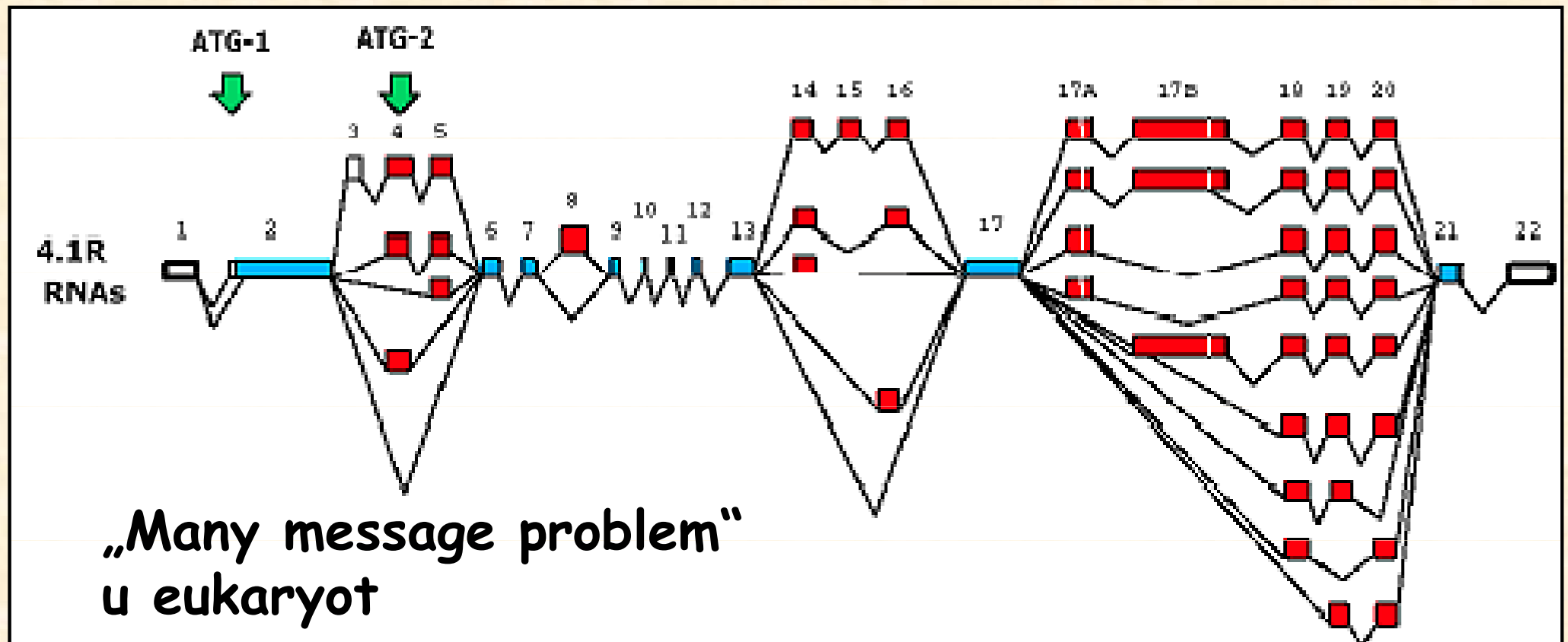
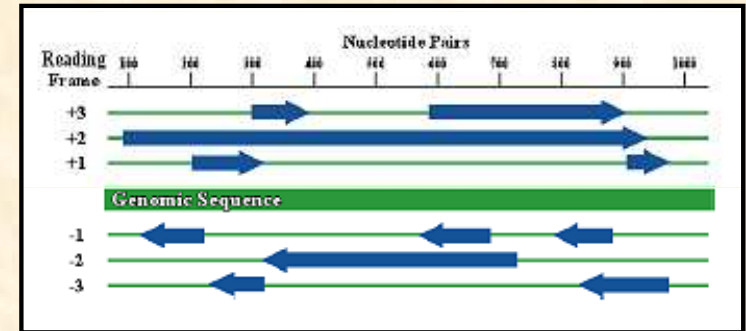
# Obecné schéma eukaryotického genu



# Jak se hledají geny?

- otevřené čtecí rámce (ORF)
- obsah a distribuce nukleotidů, „genové rysy“
- používání kodonů
- hranice exon-intron
- promotory, regulační sekvence
- homologie v databázích, EST

ORF:

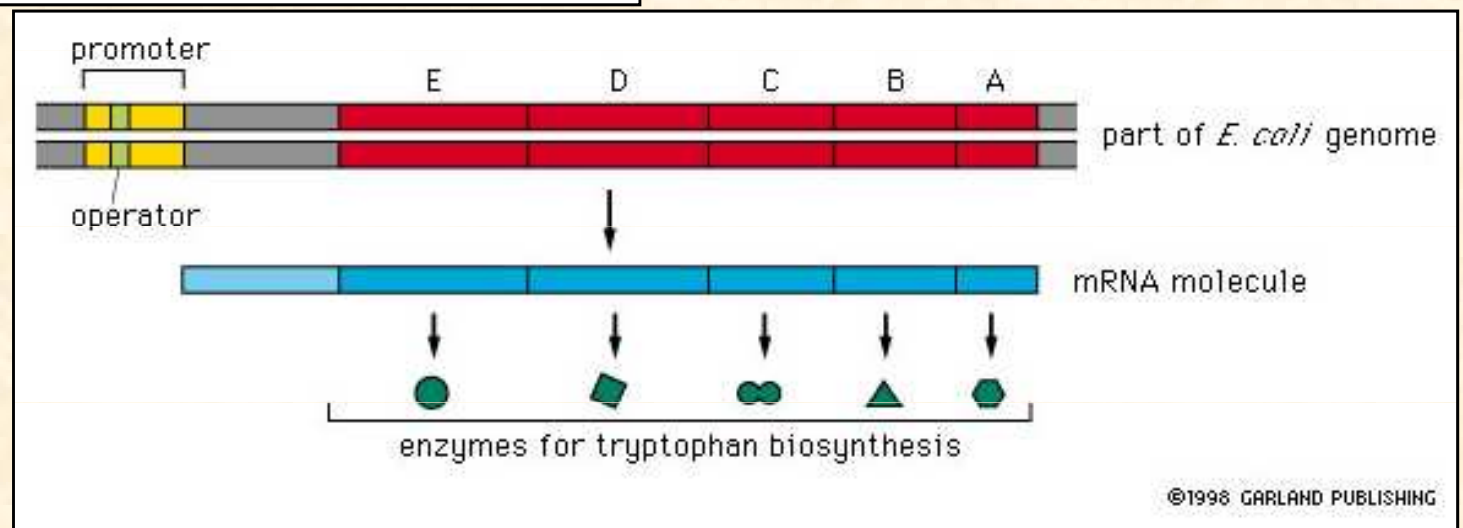
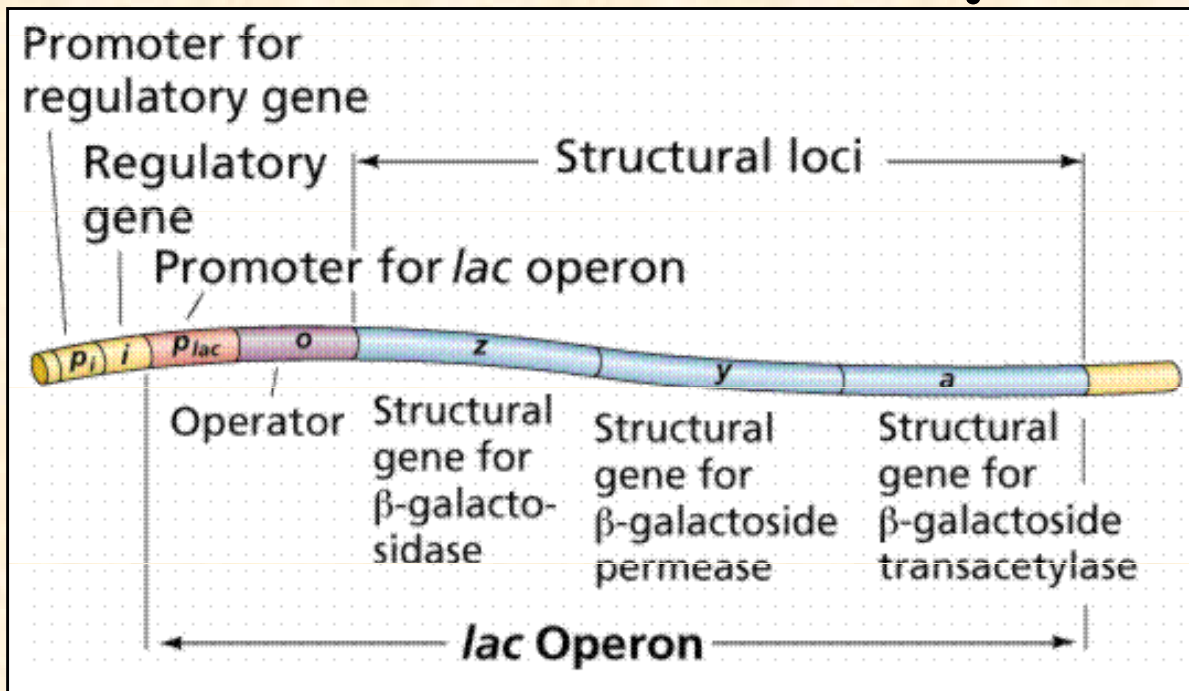


# Hledání genů u prokaryot

- ORF (otevřené čtecí rámce)
  - start kodon je následován nejméně 60 AK, poté stop kodon
  - homologie se známými ORF
- Signální sekvence
  - Transkripce - konsensus promotorové a terminační sekvencí
  - Translace - vazebné místo na ribozóm: Shine-Dalgarnova sekvence
- Rozdíly v obsahu bazí mezi kódující a nekódující sekvencemi DNA
  - obsah GC, tzv. codon bias



# Geny prokaryot jsou uspořádány do operonů

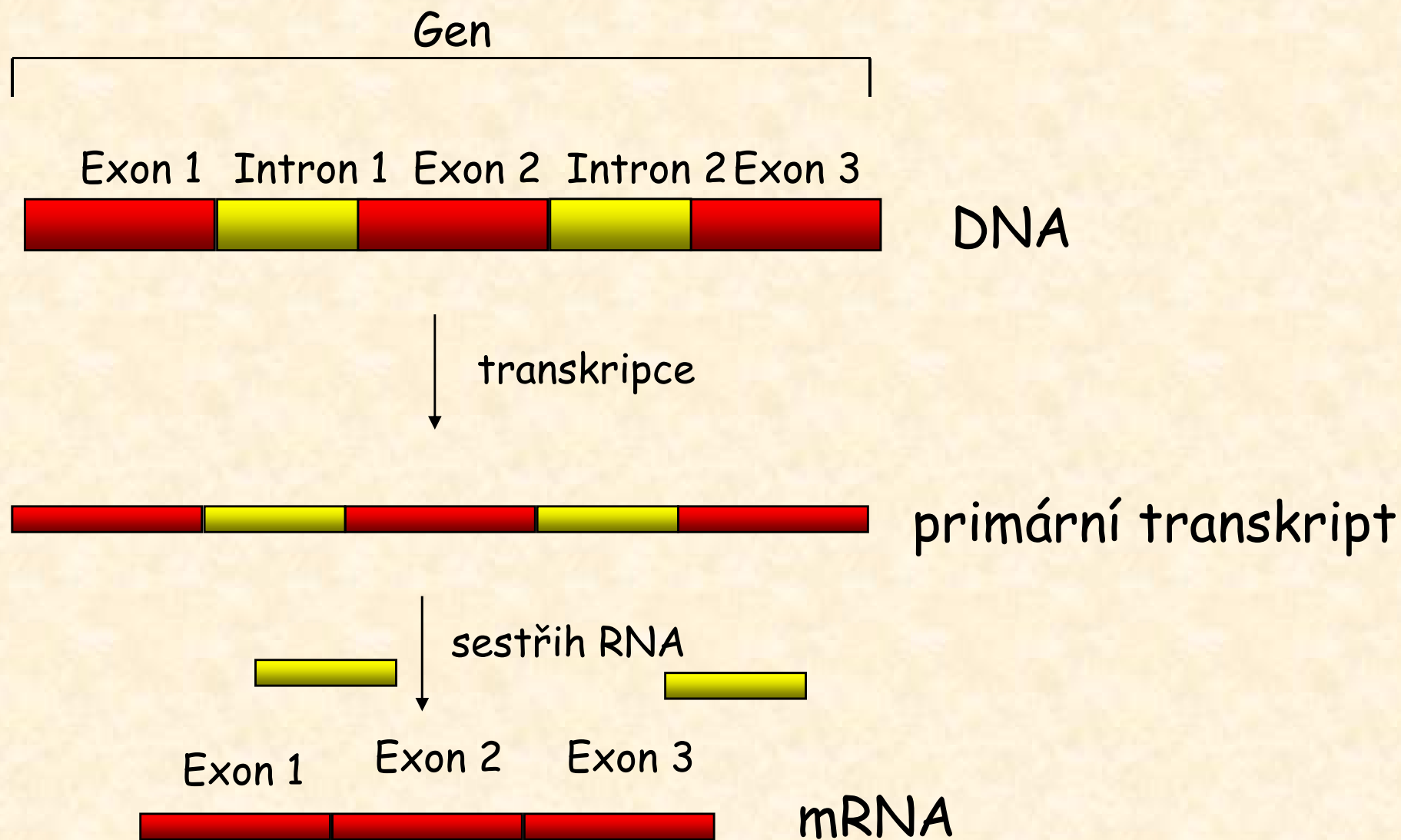




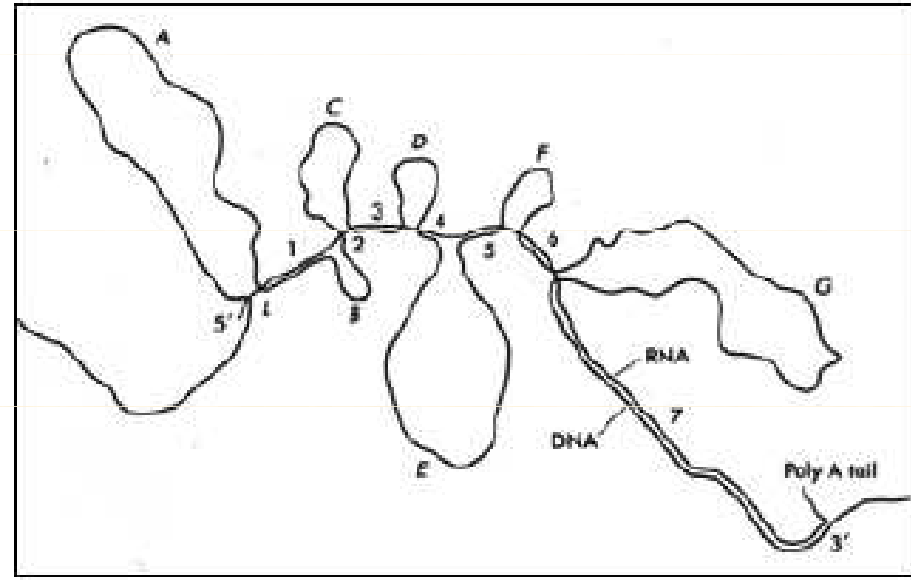
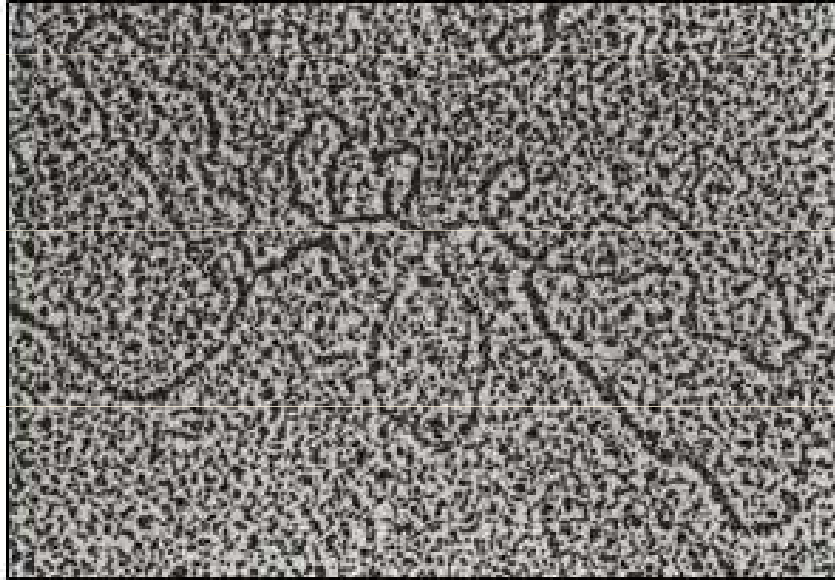
# Komplikace u eukaryot

- **Složené geny (split genes)**
  - introny a exony,
  - obratlovci - délka genu 30kb/1-2kb je kódující
  - např. gen pro dystrophin 2.4 Mb, desítky exonů, introny až 32kb
- **Velké genomy** - u rostlin až 110 000 Mb (*Fritillaria assyriaca*)
- **Většina DNA je nekódující**
  - introny, regulační oblasti, "junk" DNA
  - asi 3% kódující (člověk)
- **Složitá regulace genové exprese** (modifikace chromatinu, metylace DNA, RNAi, alternativní sestřih)
- **Regulační sekvence mohou být daleko od start kodonu**

# Složení gen



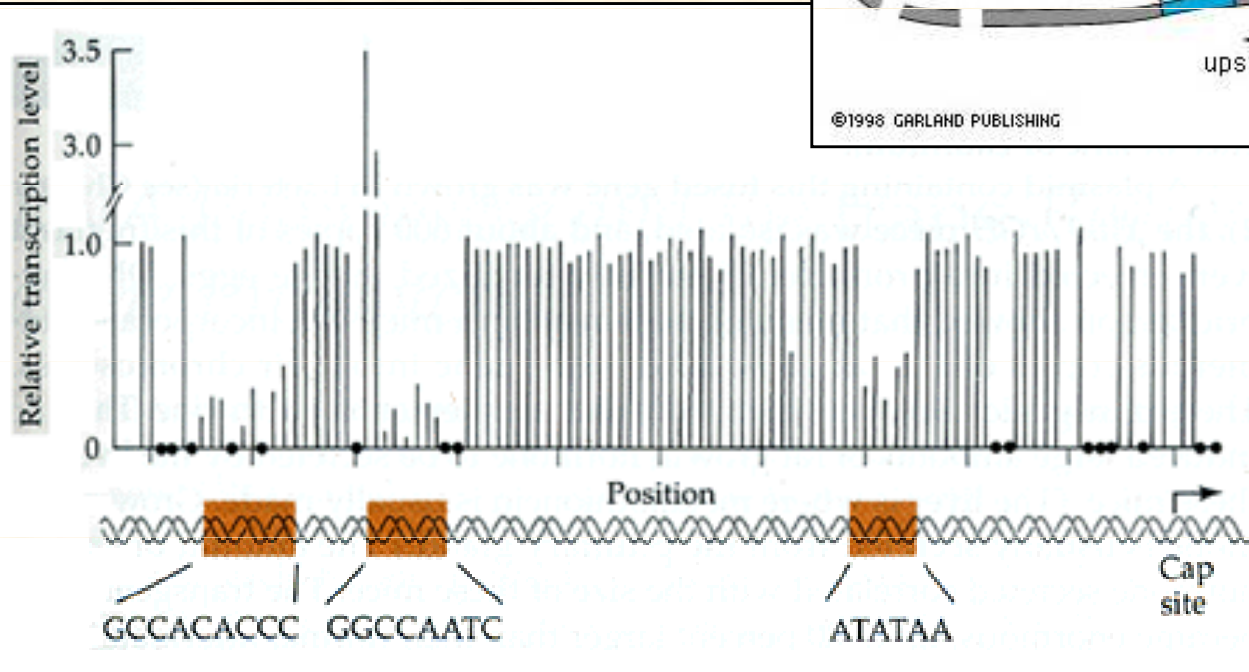
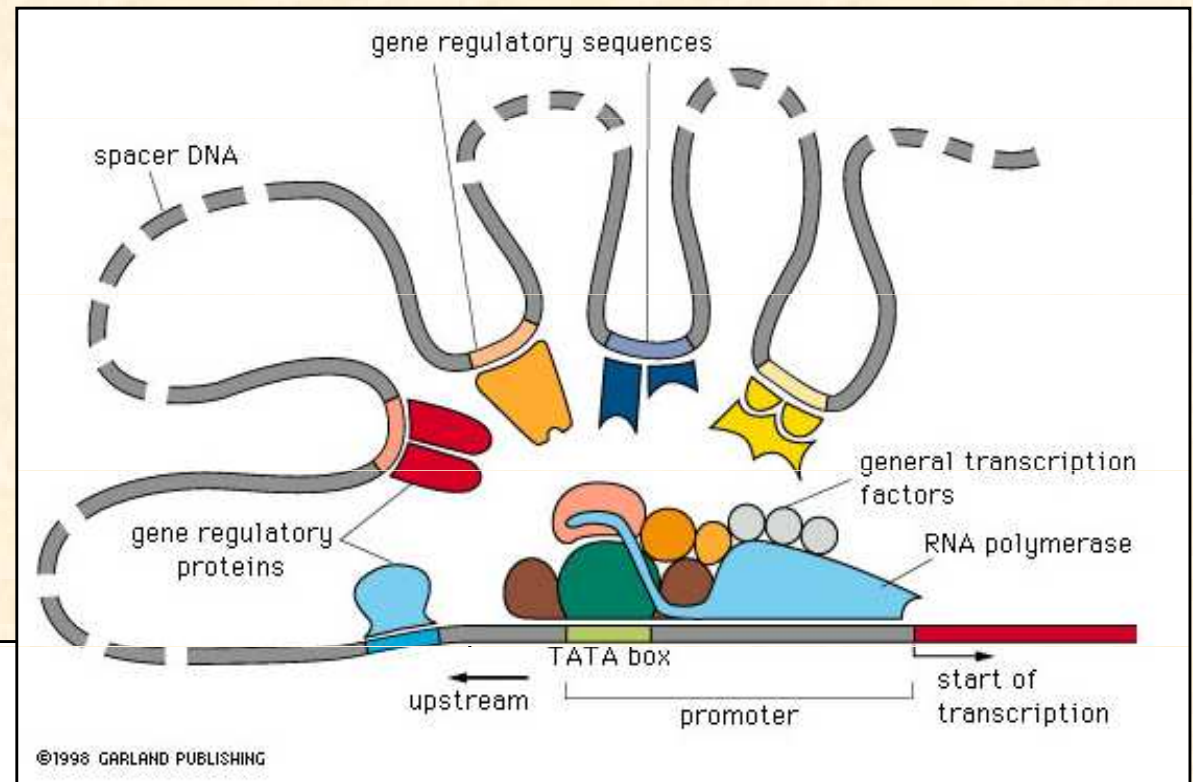
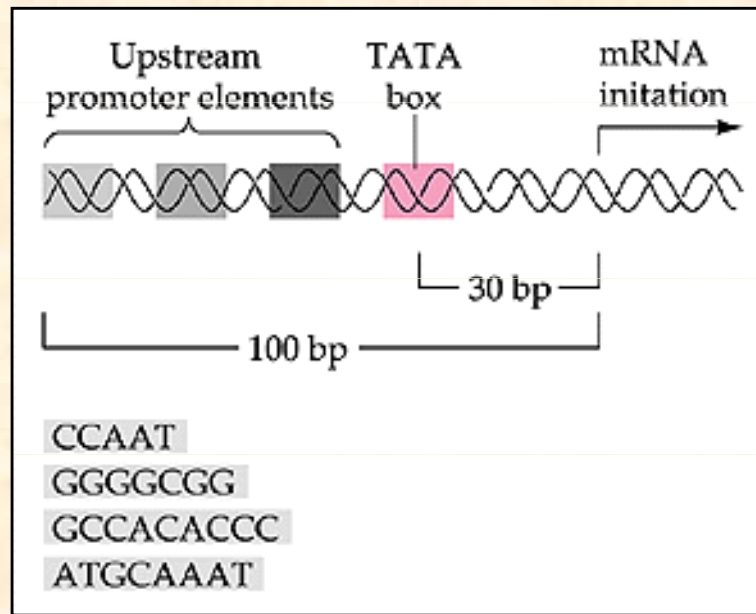
# Struktura typického eukaryotického genu



Ovalbumin gene from chicken

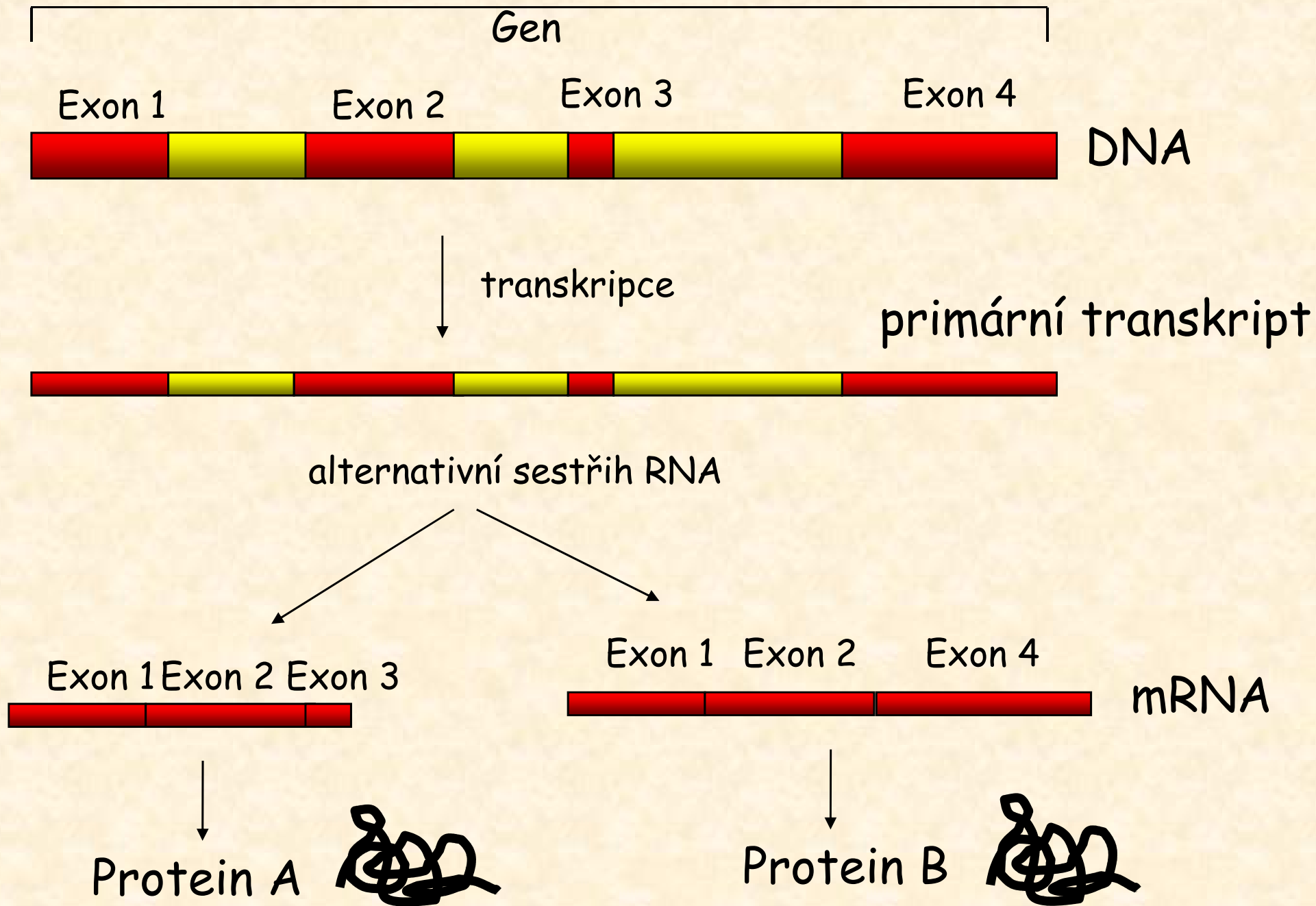
exon - expressed sequence, intron - intervening sequence

# Struktura promotoru



mutace v kritických místech blokují transkripci

# Alternativní sestřih

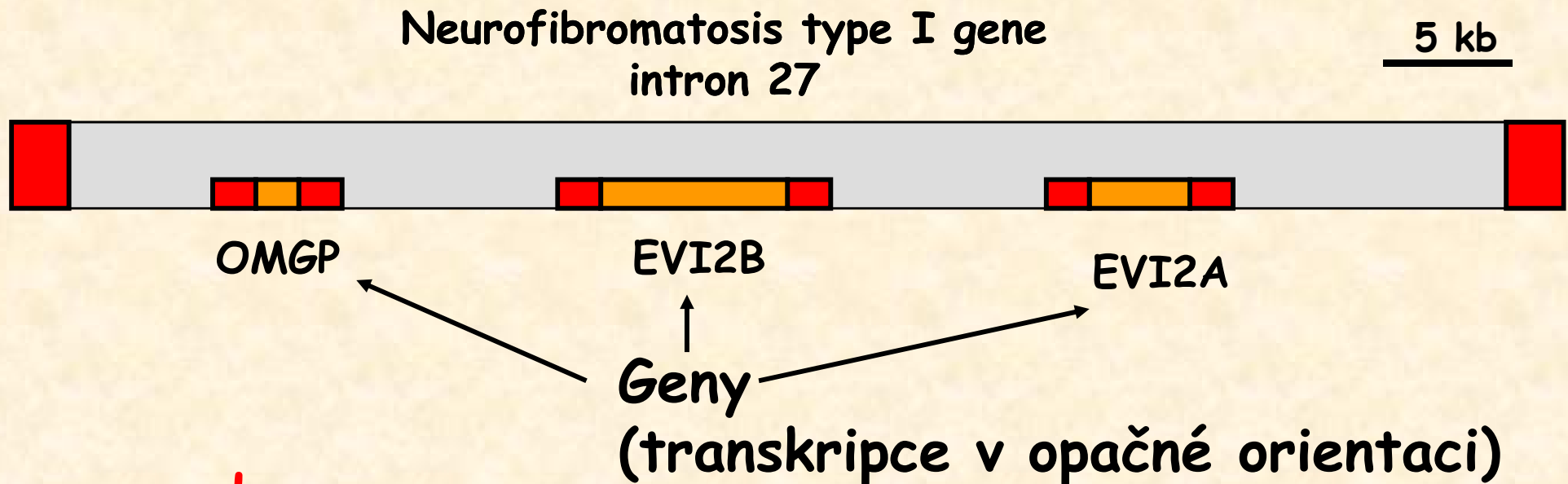


# Geny v genech a jiné podivnosti

- **překrývající se geny:**

met val .... Gen A  
GTTTATGGTA  
val tyr gly ..... Gen B

- geny **uvnitř** jiných genů:



- **pseudogeny:**



# Pseudogeny

## Definice:

- sekvence podobná genu, nekóduje funkční produkt
- nefunkční relikt původně funkčního genu

## Problém definice:

- pseudogen může plnit důležitou funkci (nekódující RNA, regulační sekvence, stabilita RNA svého homologa),
- komplikují molbiol. studie

**An expressed pseudogene regulates the messenger-RNA stability of its homologous coding gene**

## Vznik:

- (a) duplikace a degenerace jedné kopie - „non-processed“
- (b) retrotransposice - „processed“ pseudogen
- člověk má 19 000 pseudogenů, pravidlo 50:50



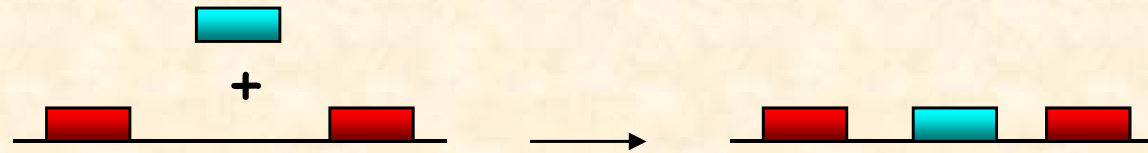
# Periodicita sekvencí uvnitř genů

- Krátké úseky 8-20bp se v genech často opakují
- Periodicita vznikla před vznikem genetického kódu a odráží původní mechanismus autoreprodukce krátkých úseků NK
- Periodicita vznikla v době existence genetického kódu a odráží jeho vlastnosti. Její význam:
  - obrana proti častému výskytu **stop-kodonů** - 300bp/0.82% bez stop
  - vznik nových genů **posunovou** mutací - ORF ve druhém čtecím rámcí
  - obrana proti častému zahajování **transkripce** mimo geny
  - nerovnoměrné používání **synonymních** kodonů, kodony a hojnost tRNA, nikoli maximalizace ale optimalizace rychlosti syntézy proteinů, pro každý protein jiná, - důsledek periodicity

# VZNIK NOVÝCH GENŮ

# Vznik nových genů

(a) Přeskupování exonů:



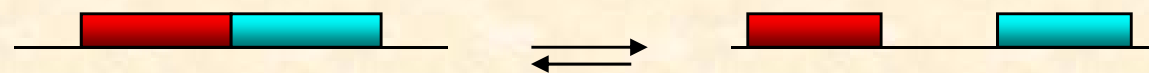
(b) Duplikace genů:



(c) Retrotranspozice:

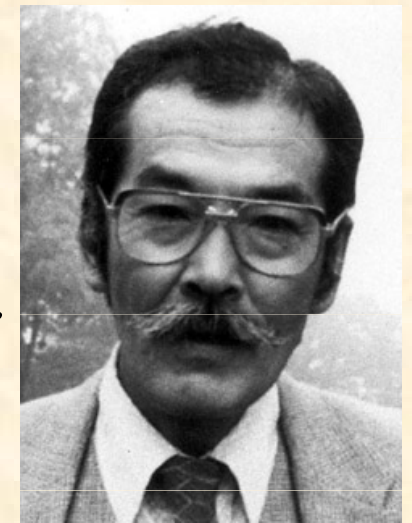


(d) Fúze a štěpení genů:



# Evoluce genovou duplikací

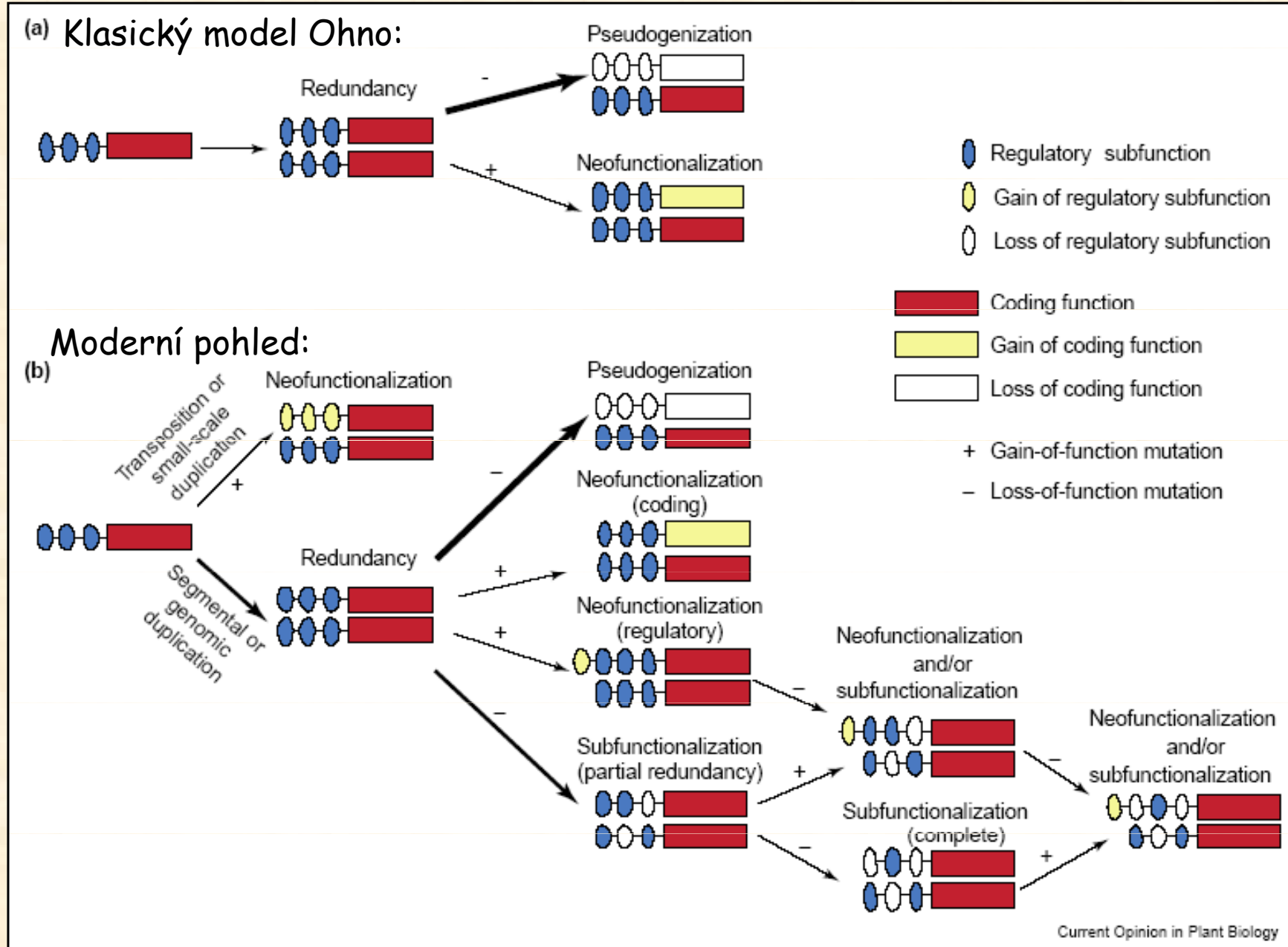
- duplikace je základem diverzifikace
- zrod nových genů u rostlin, kvasinky a drosophily je 10x pomalejší než u *C. elegans*
- poločas rozpadu genů delší u rostlin, duplikáty přetrvávají, mechanismy retence duplikátů?



(Ohno, 1970)

- disperzní x tandemové kopie - rychlost asymetrické evoluce, často u rostlin zůstávají v tandemu
- v nerekombinujících oblastech - rychlejší evoluce
- **Duplikace části genu:**  
duplikace domén/vnitřní části genu → zvýšení funkce nebo nová funkce prostřednictvím nových kombinací
- **Duplikace celého genu (genová rodina)**  
stejná kopie: zvýšení dávky genu,  
rozdílnění kopií: nové funkce
- **Duplikace klastru genů**

# Genová duplikace: pseudogenizace, neofunkcionalizace, subfunkcionalizace



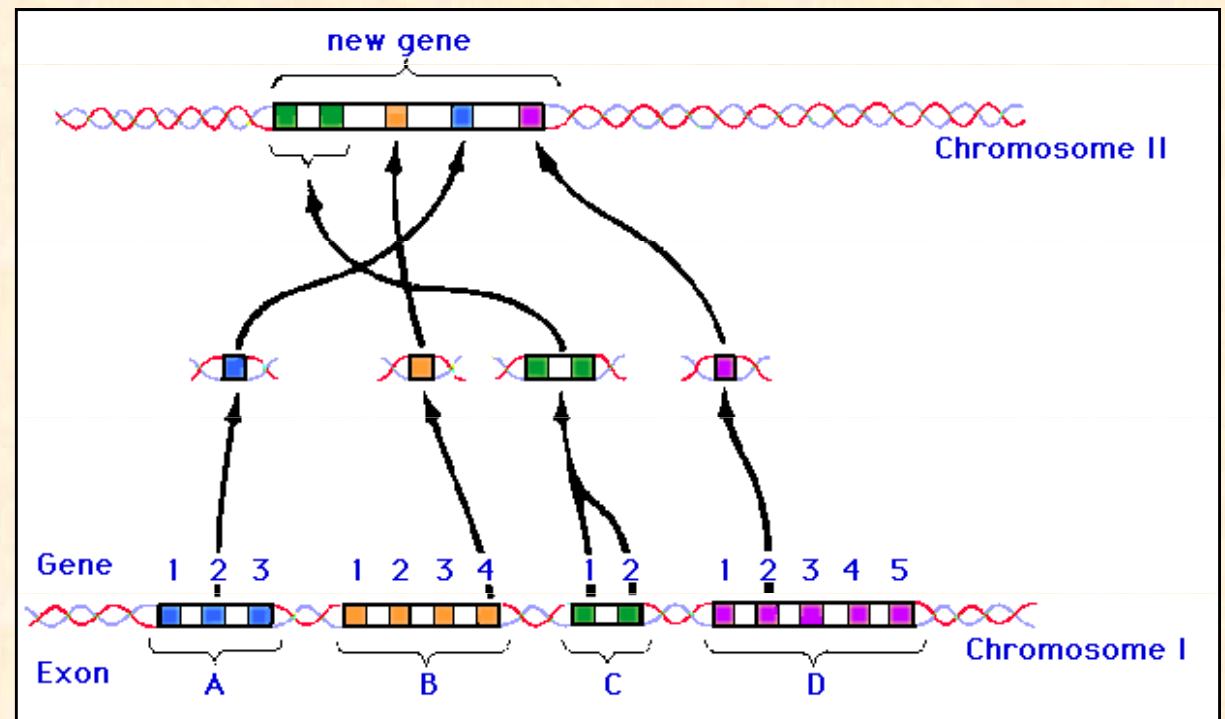
# Původ nových genů: Přeskupování exonů (exon shuffling)

- exony různých genů jsou spojeny dohromady za vzniku nového genu
- exon může být duplikován za vzniku nové exon-intronové struktury
- kombinace domén různých proteinů - mozaikový protein

## Mechanismy:

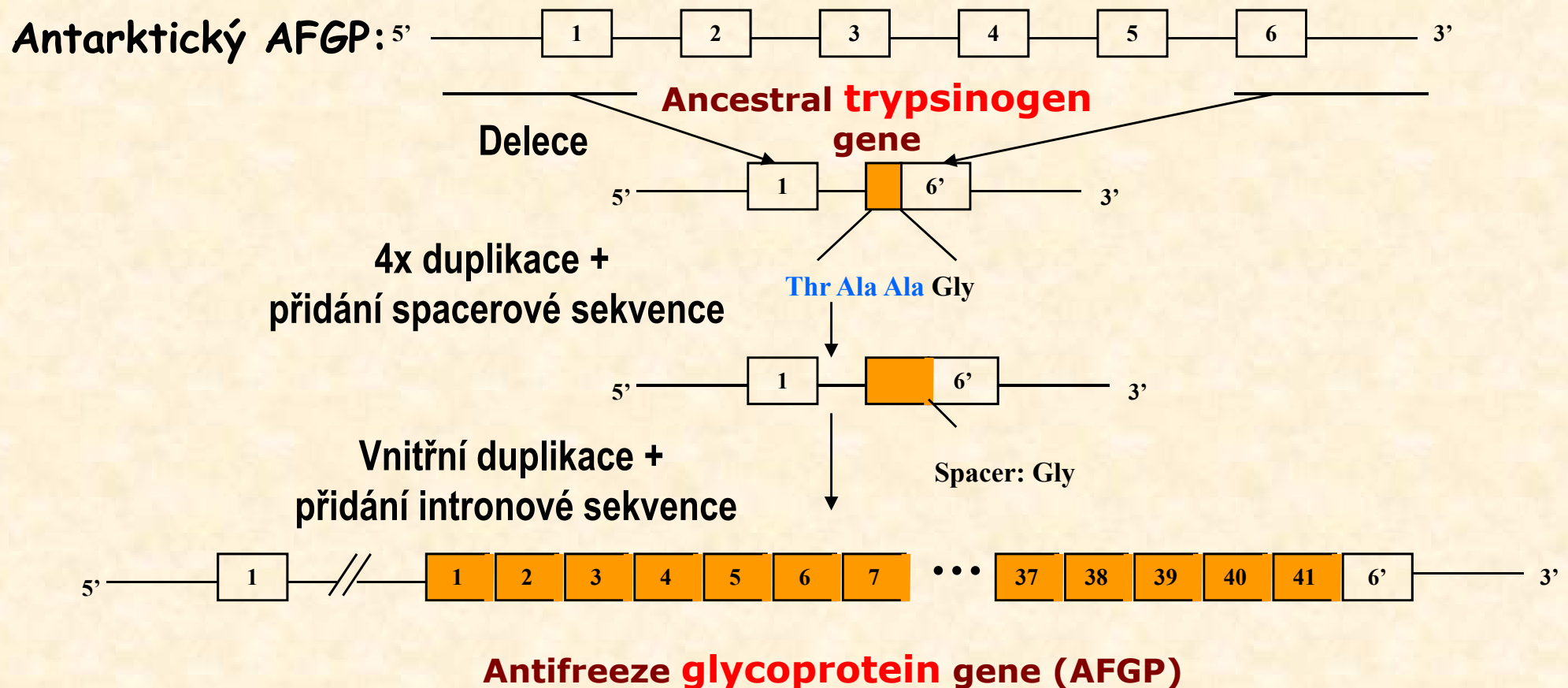
Ektopická rekombinace

Nelegitimní rekombinace





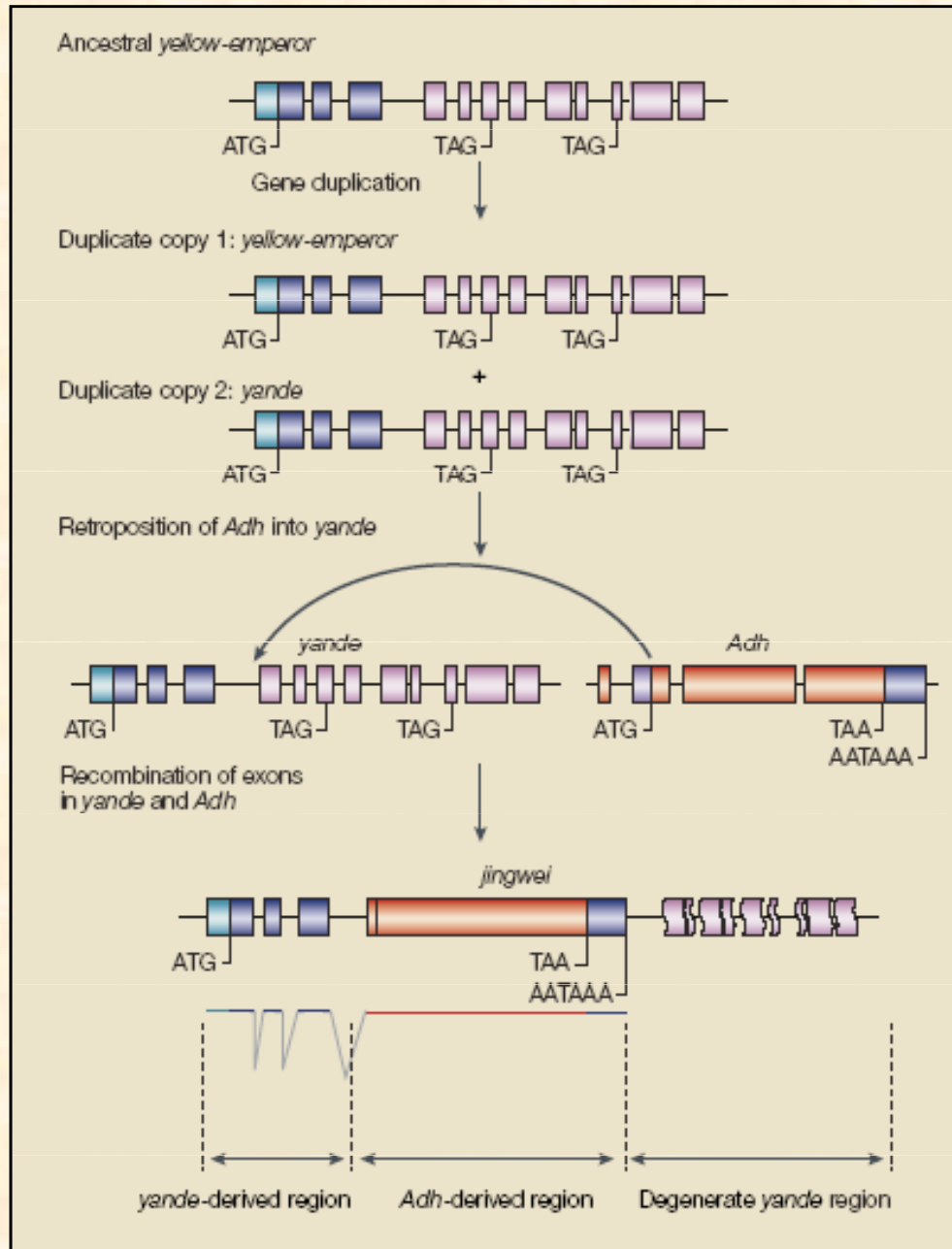
# Vznik nového genu na příkladu AFGP



- brání zmrznutí tělních tekutin, růstu krystalků ledu
- vznikl před 10 mil let, první zamrznutí polárních oblastí
- vznik z trypsinogenu, zachován 5' a 3' konce (sekrece)
- amplifikace (Thr-Ala-Ala)<sub>n</sub>, kde n=4 až n=55
- vznik 4x nezávisle, konvergentní evoluce - antarktický a arktický



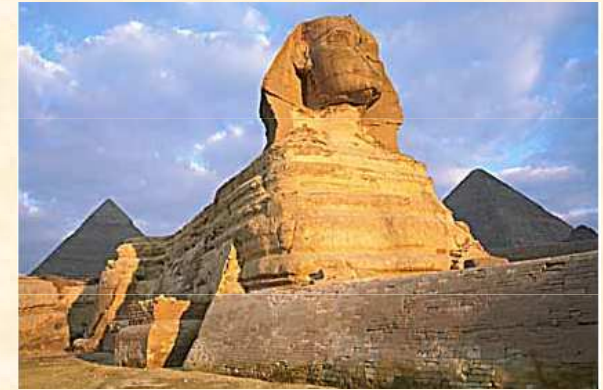
# Původ genu *Jingwei*



- vznik před 2 mil let, drosophila
- základem yellow emperor
- duplikace a retro-včlenění *Adh*
- *Adh* terminační signál
- degenerace exonů na 3'konci
- nová kombinace exonů

- pohádka o princezně *Jingwei*:  
reinkarnace utonulé princezny v krásného ptáka podobně jako odhalení fungujícího genu v původně objeveném pseudogenu

# Původ genu Sphinx (spx)



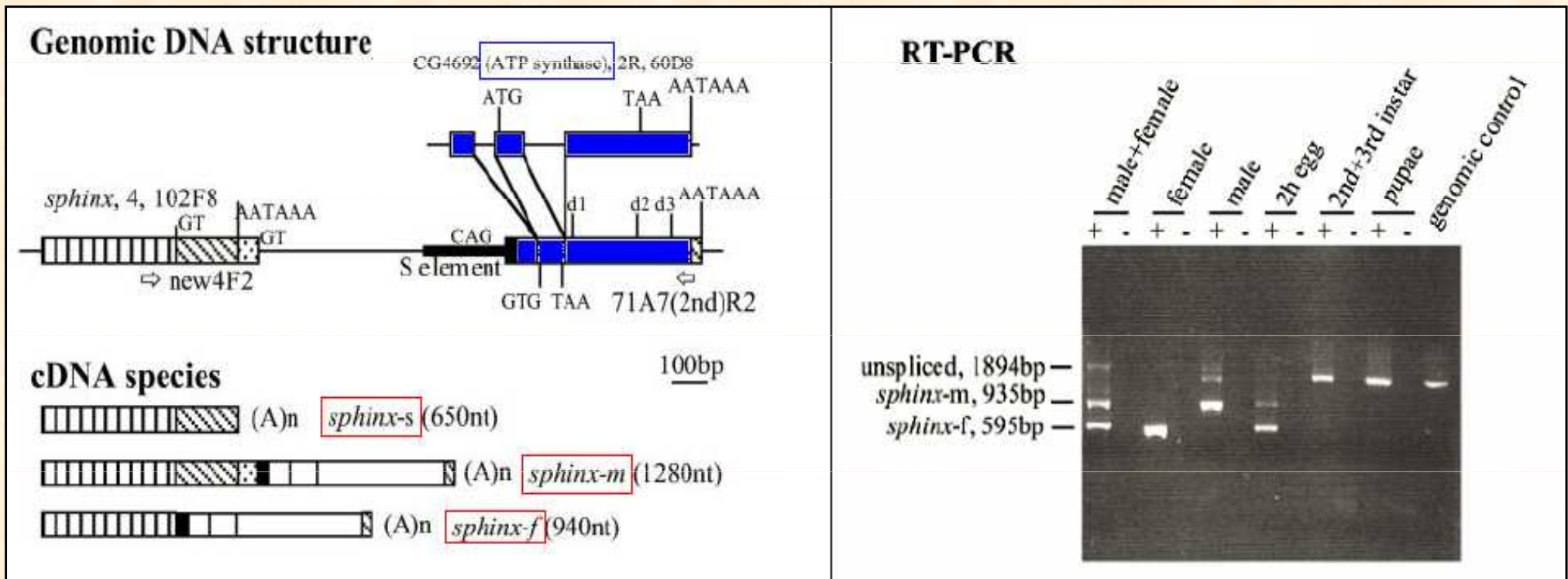
## Origin of *sphinx*, a young chimeric RNA gene in *Drosophila melanogaster*

Wen Wang\*, Frédéric G. Brunet\*, Eviatar Nevo†, and Manyuan Long\*\*

\*Department of Ecology and Evolution, University of Chicago, 1101 East 57th Street, Chicago, IL 60637; and †Institute of Evolution, University of Haifa, Mount Carmel, Haifa 31905, Israel

Contributed by Eviatar Nevo, February 4, 2002

- *spx* vznikl inzercí reverzně transkribovaného genu ATP syntázového řetězce F do genu pro RNA na chromosomu IV
- po oddělení *D. melanogaster* před 2-3 miliony let



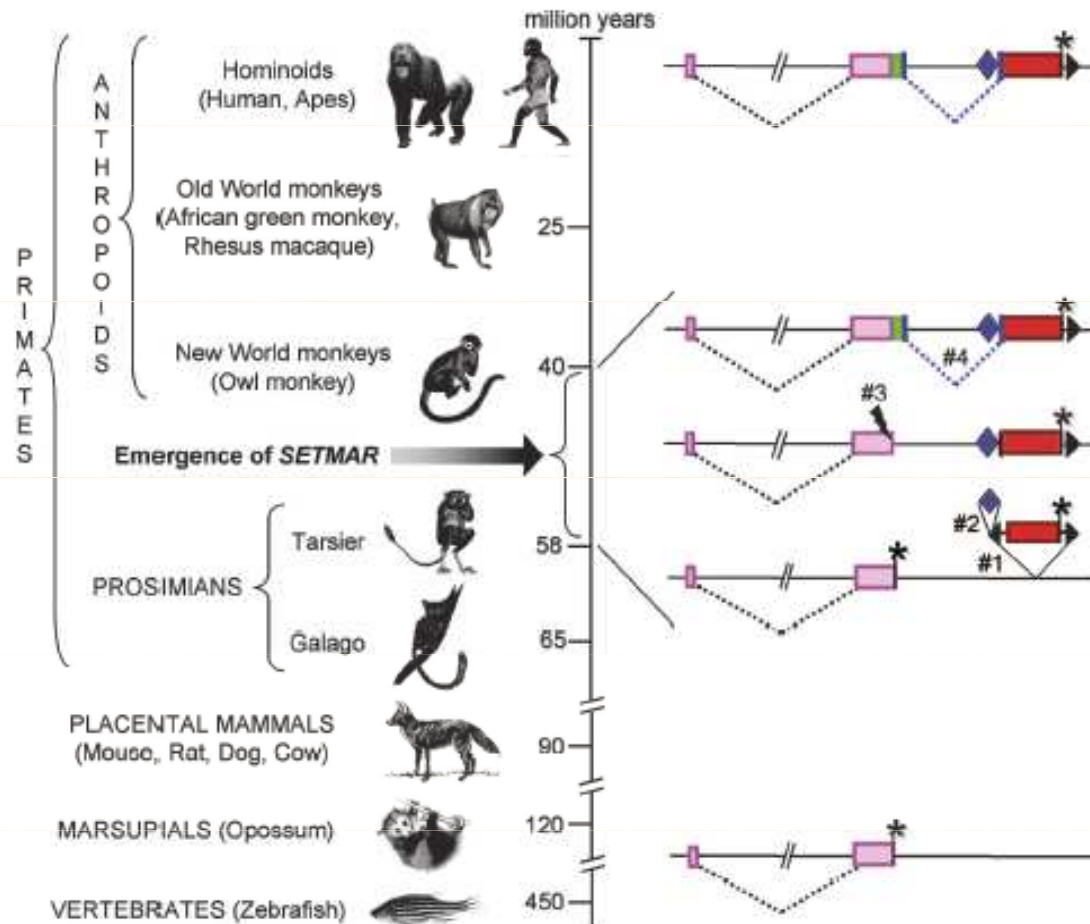
# Původ genu SETMAR - „recyklace“ transposonu

Birth of a chimeric primate gene by capture of the transposase gene from a mobile element

Richard Cordaux\*, Swalpa Udit†, Mark A. Batzer\*, and Cédric Feschotte\*\*

Histon  
metyltransferáza

+  
transpozáza



- zrušení stop
- vznik nového stop
- exonizace
- degenerace TIRu
- vznik intronu
- DNA vazebná doména Tn zachována
- TIR místa v genomu

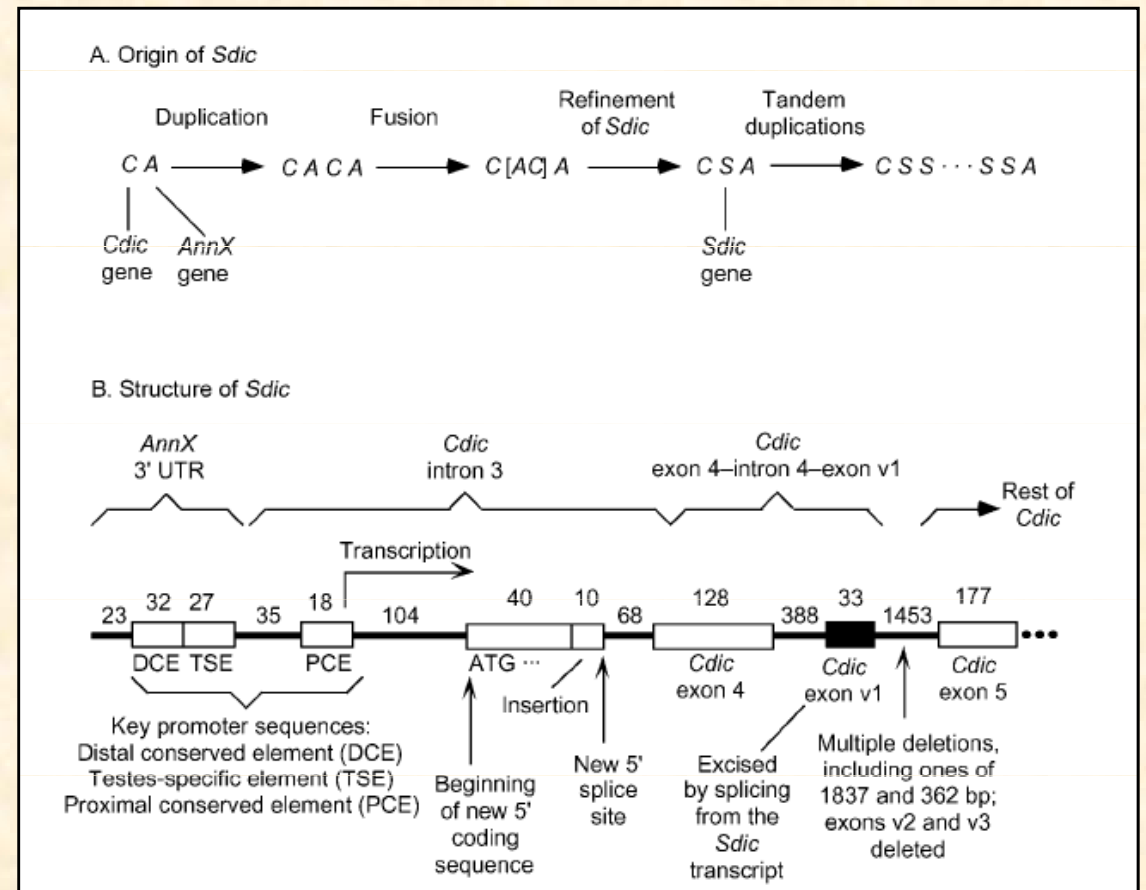
- 50 mil let

Fig. 1. Milestones leading to the birth of *SETMAR*. The structure of the *SETMAR* locus (Right) and a simplified chronology of the divergence time of the species examined relative to hominoid primates (Left) are shown. Pink boxes represent the two *SET* exons, which are separated by a single intron (interrupted black line) and form a “*SET*-only” gene whose structure is conserved in all nonanthropoid species examined and terminated with a stop codon (\*) located at a homologous position (except in cow; see Fig. 2a). The *Hsmar1* transposon (event 1) was inserted in the primate lineage, after the split between tarsier and anthropoids, but before the divergence of extant anthropoid lineages. The transposon is shown here with its TIRs (black triangles) and transposase coding sequence (red box). The secondary *Alu5x* insertion within the TIR of *Hsmar1* (event 2) is represented as a blue diamond. The position of the deletion removing the stop codon of the “*SET*-only” gene (event 3) is indicated as a lightning bolt. The *de novo* conversion from noncoding to exonic sequence is shown in green, the creation of the second intron is represented as a dashed blue line (event 4), and the splice sites are shown as thick blue lines.

# Původ nových genů: Vznik genu de novo?

Gen *Sdic* (*D. melanogaster*): přeměna intronu v exon

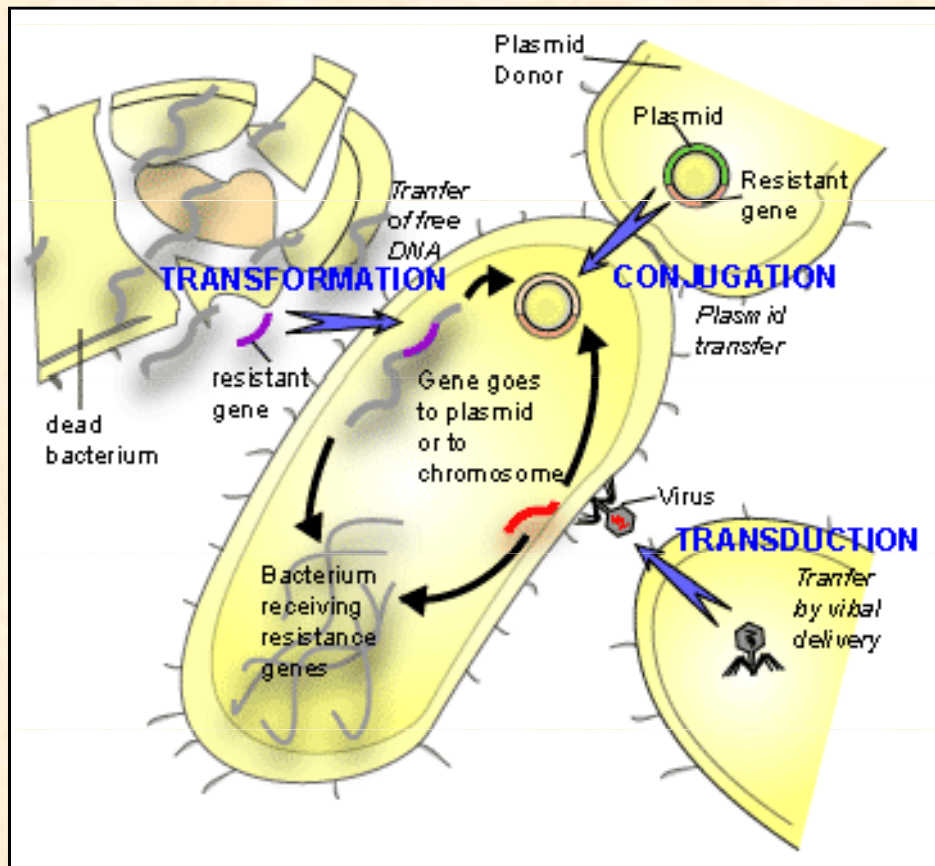
- duplikace a fúze genů *Cdic* + *AnnX*
- vznik *Sdic*, nachází se v deseti kopiích mezi rodičovskými geny, nové promotory - tkáňově-specif. exprese
- tandemová duplikace





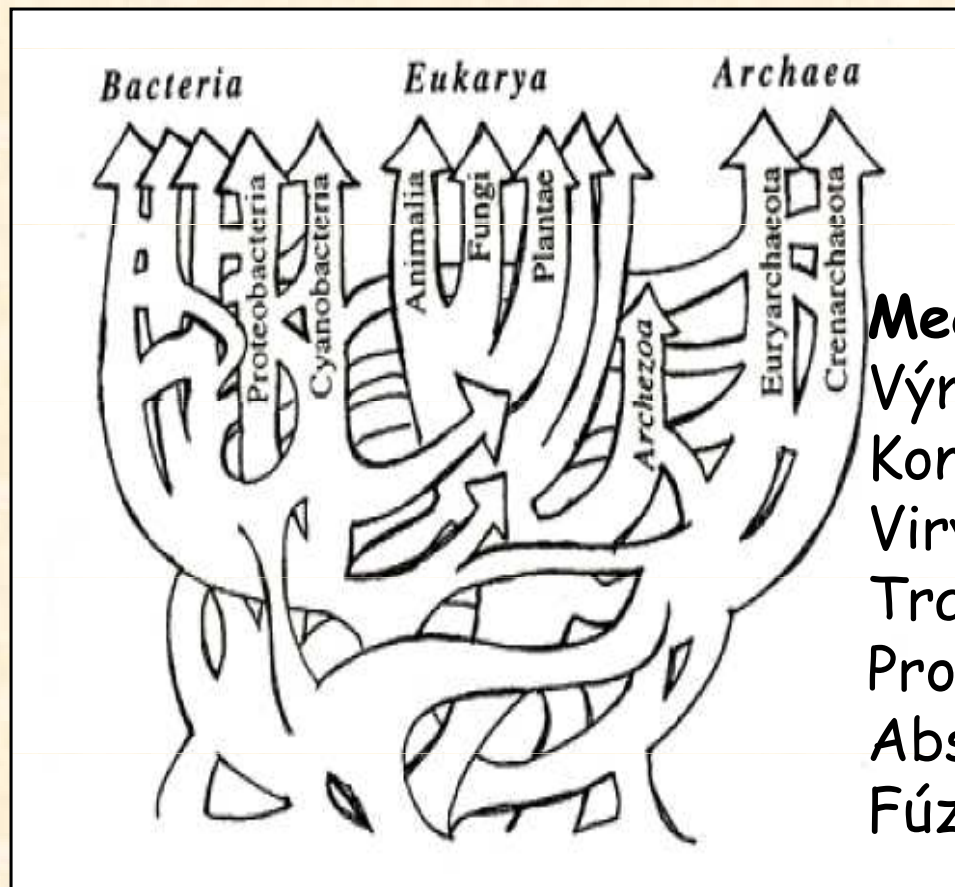
# Původ nových genů: Horizontální přenos

- vertikální (sexualita) a horizontální přenos (mezi druhy)
- konjugace, transdukce a transformace. Endosymbióza
- šíření rezistence k antibiotikům
- vnitrobuněčný parazitismus (Wolbachia)



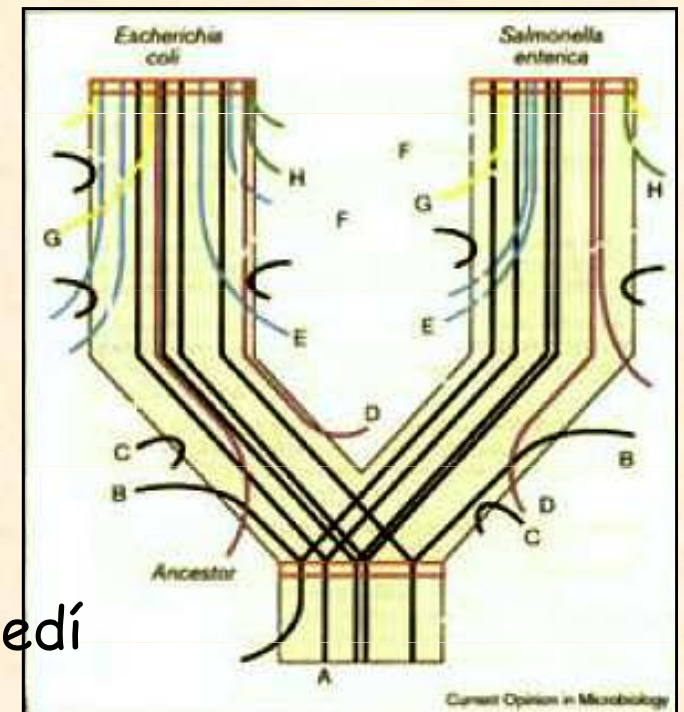
# Horizontální genový přenos

- přírodní genetické inženýrství je časté, dokonce i mezi evolučně vzdálenými taxony
- GMO organizmy si budou vyměňovat geny s ne-GMO
- dnes málo informací o horizontálním přenosu a jeho regulaci
- místo stromu života lépe mluvit o pavučině (síti, webu) života

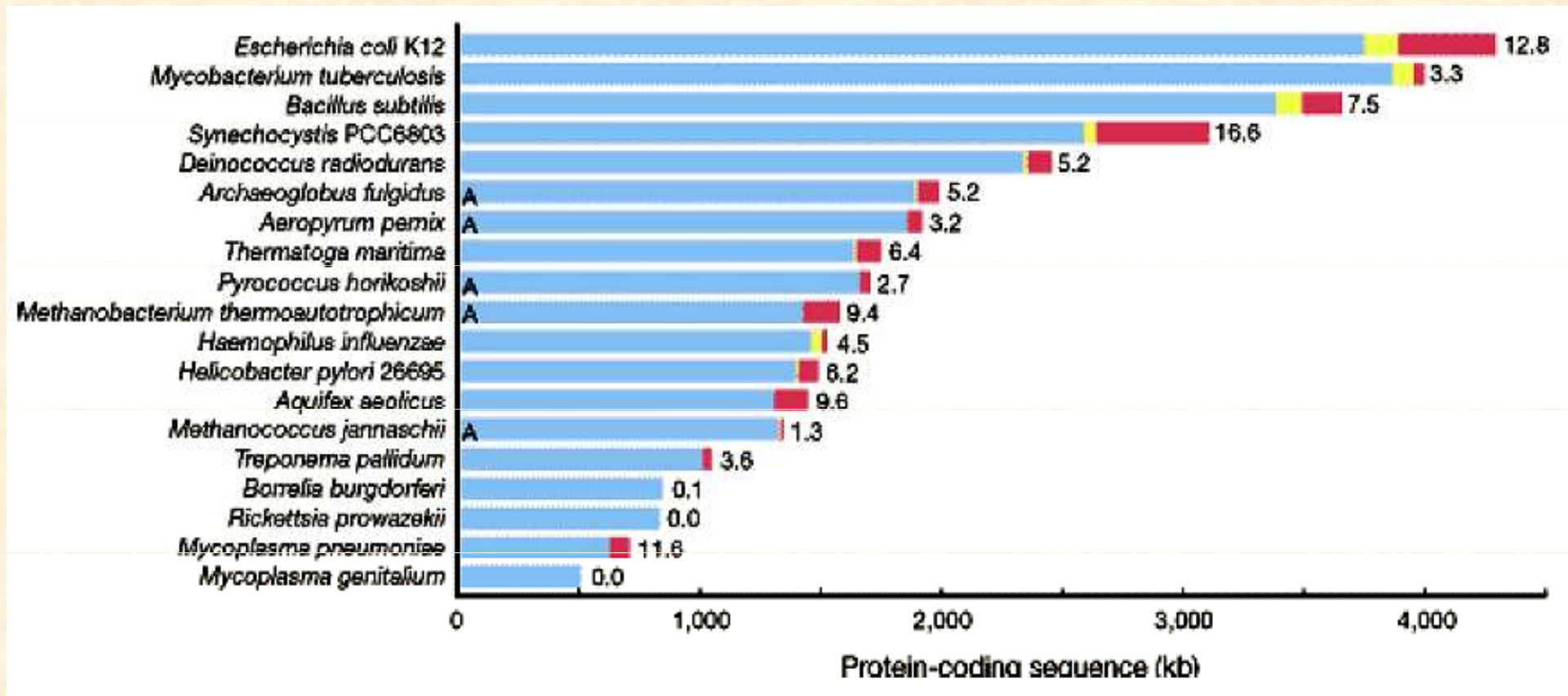


## Mechanismy:

- Výměna plazmidů
- Konjugace
- Viry
- Transposony
- Promiskuitní DNA
- Absorbce z prostředí
- Fúze buněk



# Horizontální genový přenos



## Metody studia

*přímé:*

Subtraktivní hybridizace

Microarrays

*nepřímé:*

Zastoupení kodonů (codon bias)

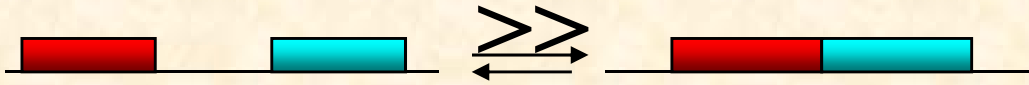
GC obsah

Konzervativní pořadí genů

Vysoká homologie se vzdáleným druhem



# Původ nových genů: Štěpení a fúze genů



Větší genom - více fúzí

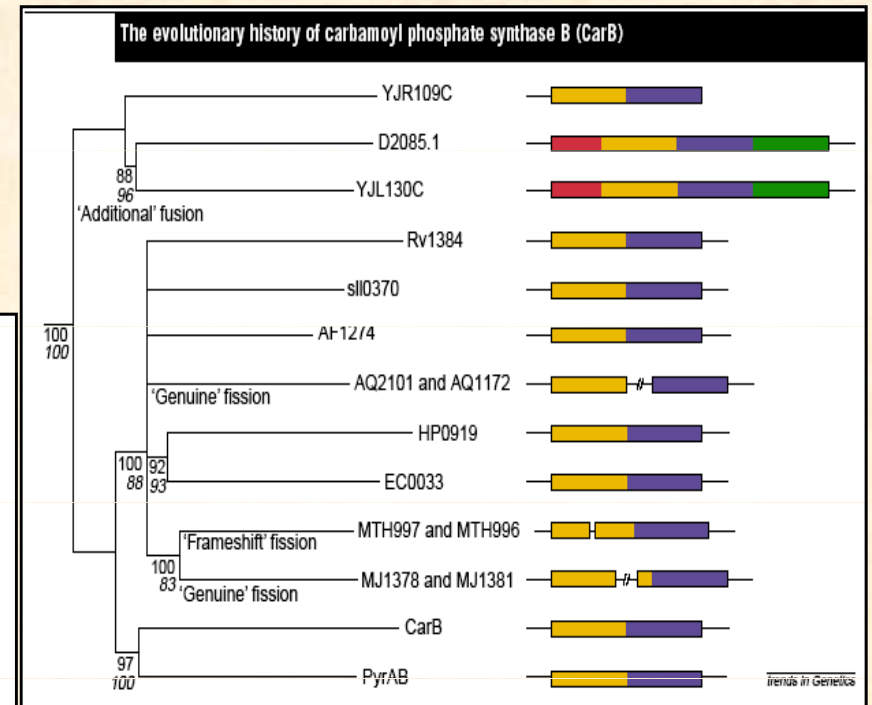
Number of gene organizations resulting from fission and fusion

Species <sup>c</sup>	Genome size <sup>a</sup>	Fusion	Fission		
			Total	Genuine <sup>b</sup>	Frameshift <sup>b</sup>
<i>Mycoplasma genitalium</i>	468	2	2	1	1
<i>Mycoplasma pneumoniae</i>	677	2	1	0	1
<i>Rickettsia prowazekii</i>	834	6	2	0	2
<i>Borrelia burgdorferi</i>	850	3	1	1	0
<i>Chlamydia trachomatis</i>	876	8	0	0	0
<i>Treponema pallidum</i>	1031	6	0	0	0
<i>Aquifex aeolicus</i>	1522	12	13	8	5
<i>Helicobacter pylori</i> 26695	1590	9	0	0	0
<i>Haemophilus influenzae</i>	1717	18	13	3	10
<i>Methanococcus jannaschii</i>	1735	12	7	5	2
<i>Methanobacterium thermoautotrophicum</i>	1871	16	18	5	13
<i>Pyrococcus horikoshii</i>	2061	4	3	3	0
<i>Archaeoglobus fulgidus</i>	2407	19	9	8	1
<i>Synechocystis</i> PCC6803	3168	24	4	4	0
<i>Mycobacterium tuberculosis</i>	3924	36	4	1	3
<i>Bacillus subtilis</i>	4100	19	1	1	0
<i>Escherichia coli</i>	4290	33	10	2	8

<sup>a</sup>Genome size in number of predicted genes.

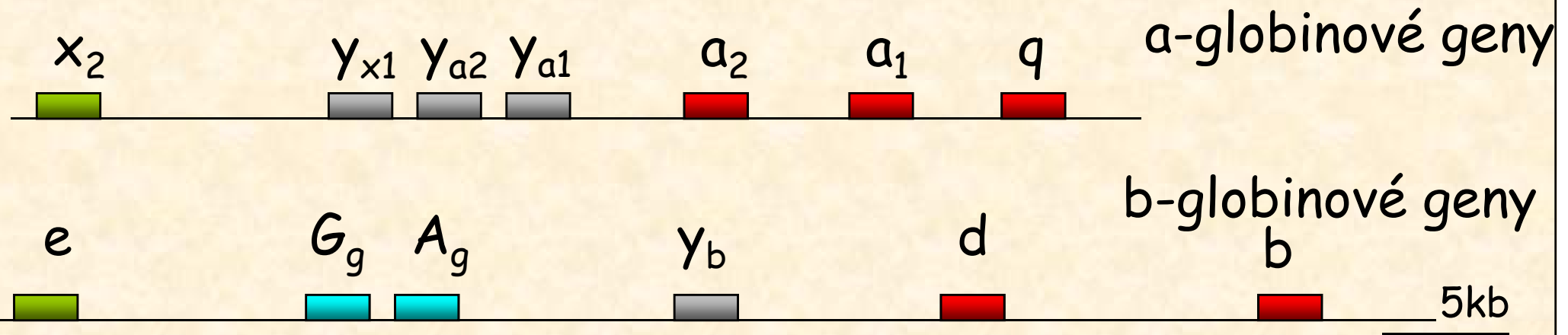
<sup>b</sup>For a definition of the subdivision in 'genuine' and 'frameshift', see text.

<sup>c</sup>Thermophilic species are shown in bold.

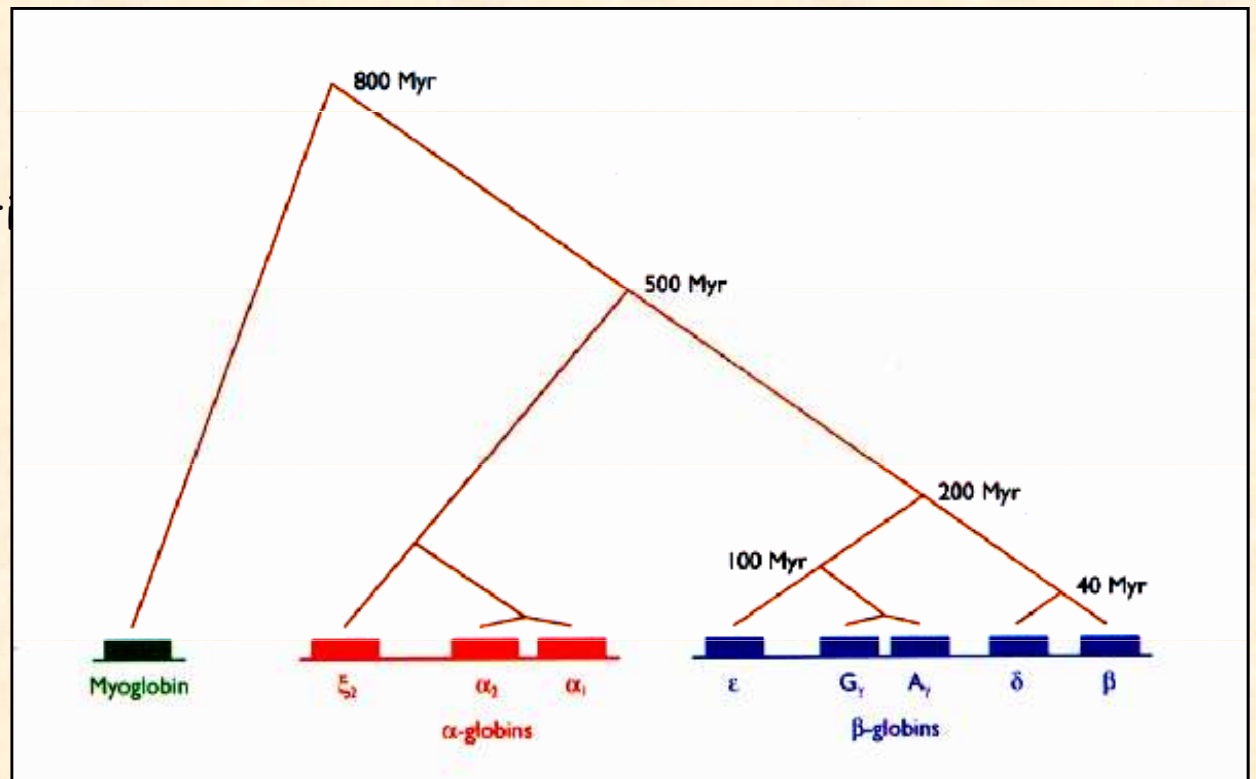


- častější fúze než štěpení
- štěpení u termofilů

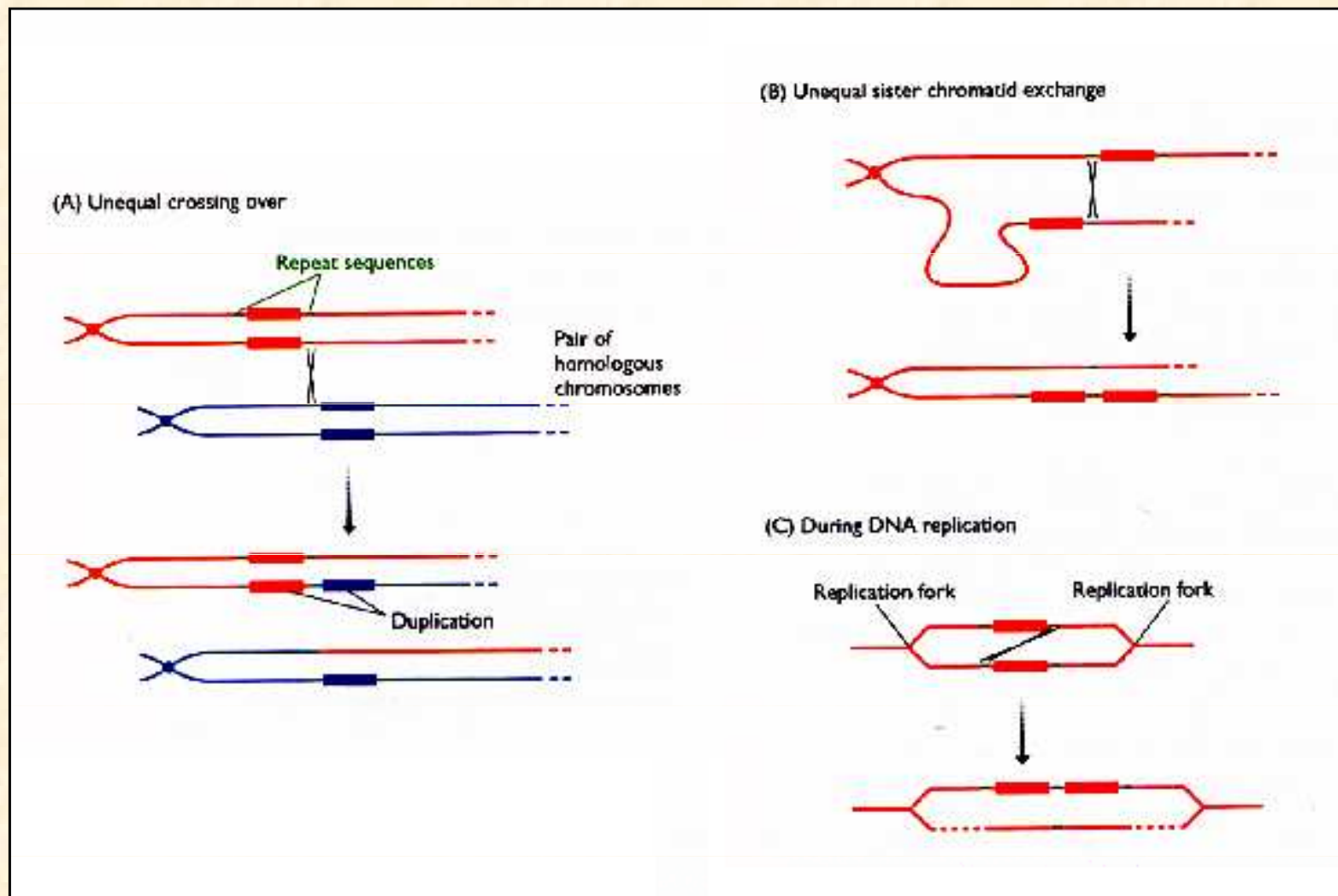
# Globinová genová rodina



- Geny exprimované v embryu
- Geny exprimované v plodu
- Geny exprimované v dospělosti
- Pseudogeny

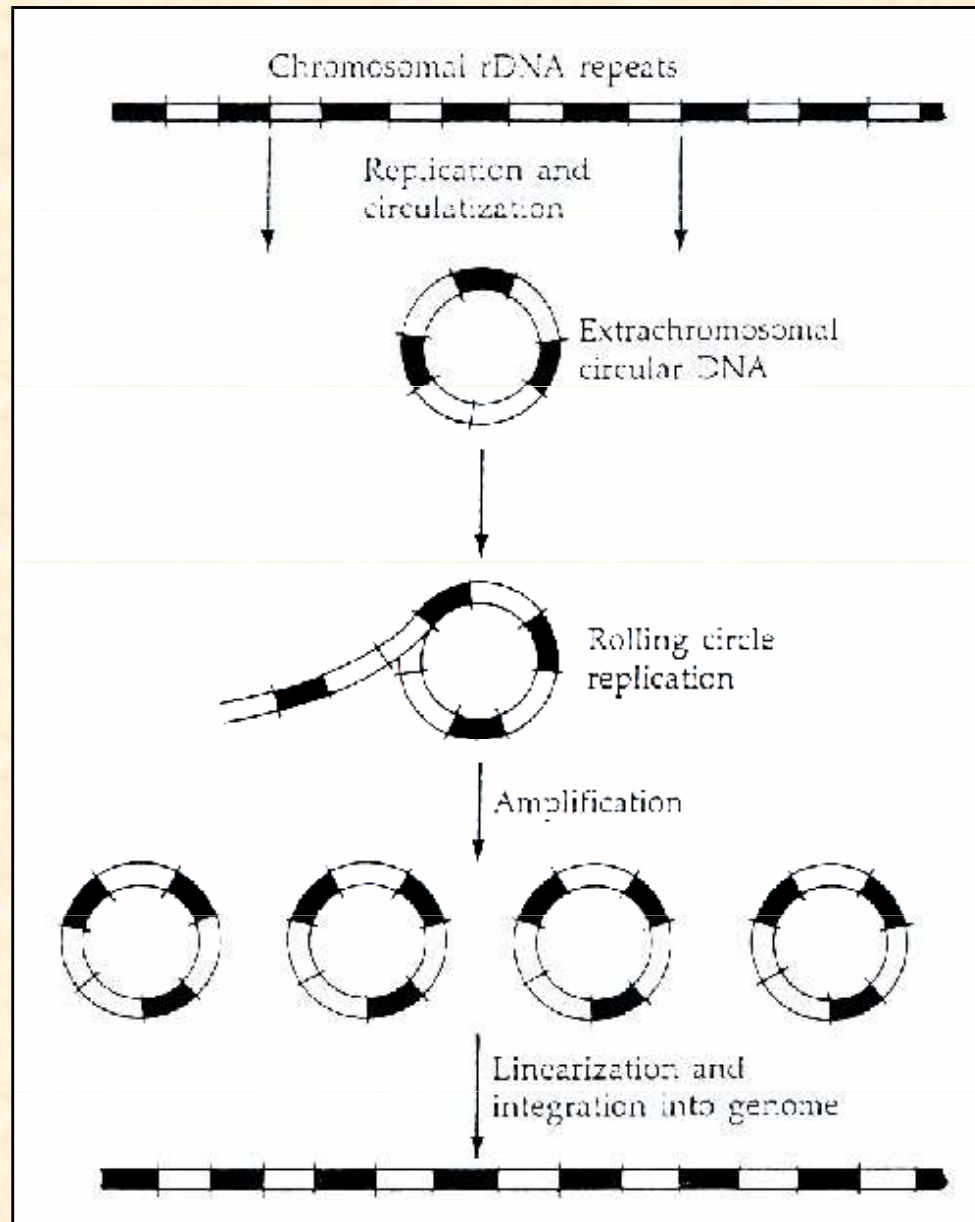


# Mechanismy duplikace genů

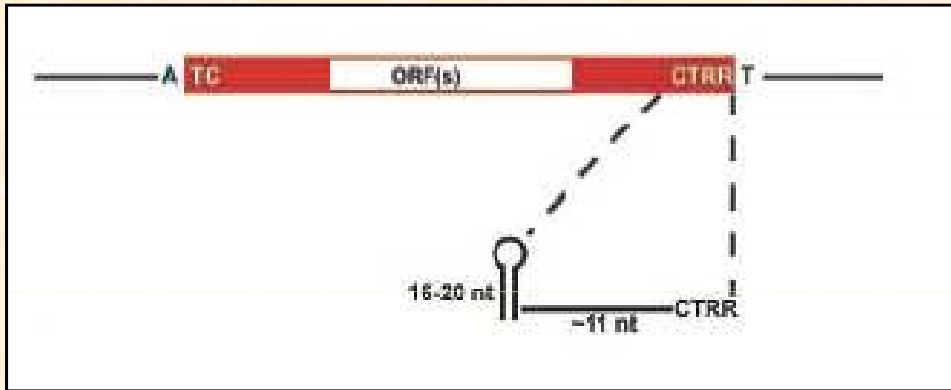


1. nerovnoměrný crossing-over (různé chromosomy)
2. nerovnoměrná výměna mezi sesterskými chromatidami
3. duplikace při replikaci

# Amplifikace genů mechanismem otáčivé kružnice



# Napomáhají Helitrony vzniku nových genů?



## Rolling-circle transposons in eukaryotes

Vladimir V. Kapitonov\* and Jerzy Jurka

Genetic Information Research Institute, 2081 Landings Drive, Mountain View, CA 94043

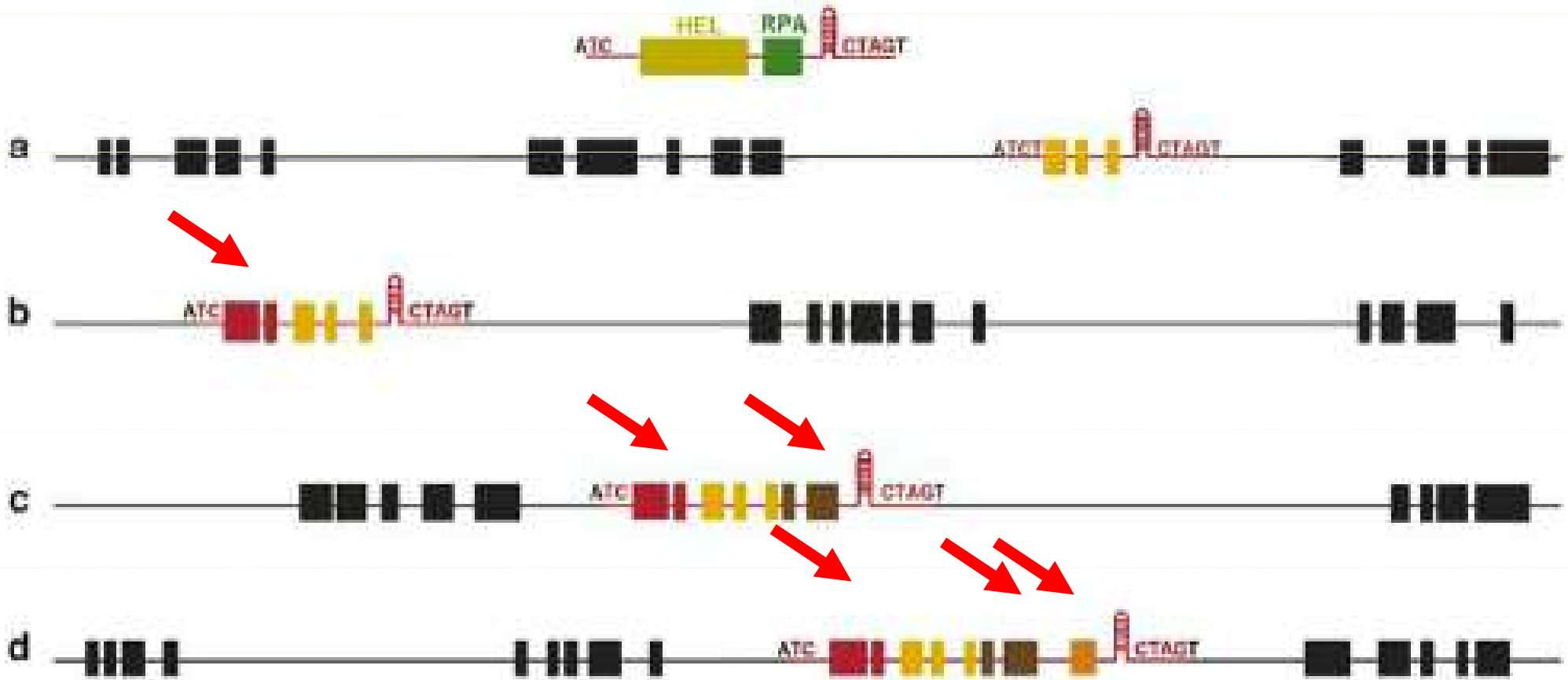
Communicated by Margaret G. Kidwell, University of Arizona, Tucson, AZ, May 29, 2001 (received for review April 10, 2001)

All eukaryotic DNA transposons reported so far belong to a single category of elements transposed by the so-called "cut-and-paste" mechanism. Here, we report a previously unknown category of eukaryotic DNA transposons, *Helitron*, which transpose by rolling-circle replication. Autonomous *Helitrons* encode a 5'-to-3' DNA helicase and nuclease/ligase similar to those encoded by known rolling-circle replicons. *Helitron*-like transposons have conservative 5'-TC and CTRR-3' termini and do not have terminal inverted repeats. They contain 16- to 20-bp hairpins separated by 10-12 nucleotides from the 3'-end and transpose precisely between the 5'-A and T-3', with no modifications of the AT target sites. Together with their multiple diverged nonautonomous descendants, *Helitrons* constitute ~2% of both the *Arabidopsis thaliana* and *Caenorhabditis elegans* genomes and also colonize the *Oriza sativa* genome. Sequence conservation suggests that *Helitrons* continue to be transposed.

and best illustrated by a recent study of *Sleeping Beauty*, a Tc1-like transposon from fish (13), reconstructed from its inactive copies and demonstrated to be transpositionally active in a test tube. Another much more ancient example is a PiggyBac-like DNA transposon, *Looper*, discovered in the human genome [V.V.K. and J.J., *Rebase Update* (1998) [www.girinst.org/Rebase\\_Update.html](http://www.girinst.org/Rebase_Update.html)], whose consensus sequence is based on a multiple alignment of the inactive copies, which are ~100 million years old. All genomic copies of *Looper* are mutated to the extent that no traces of its transposase could be detected at the sequence level. However, the transposase re-emerged from the virtual background noise after reconstructing the consensus sequence.

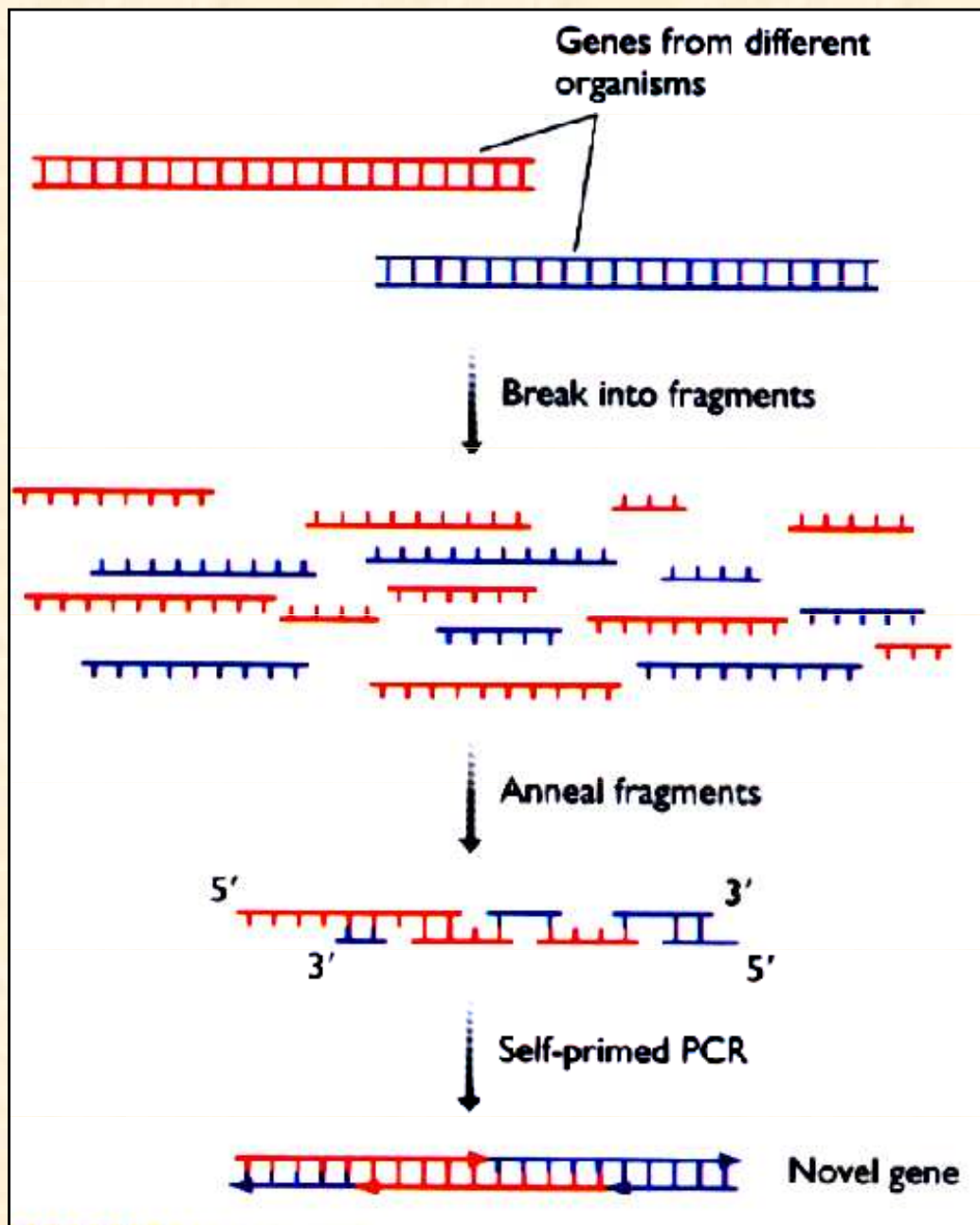
### Materials and Methods

**Computational Analysis.** TEs reported in the manuscript were identified by running DNA sequences of prospective TEs against





# Řízená evoluce a design nových genů



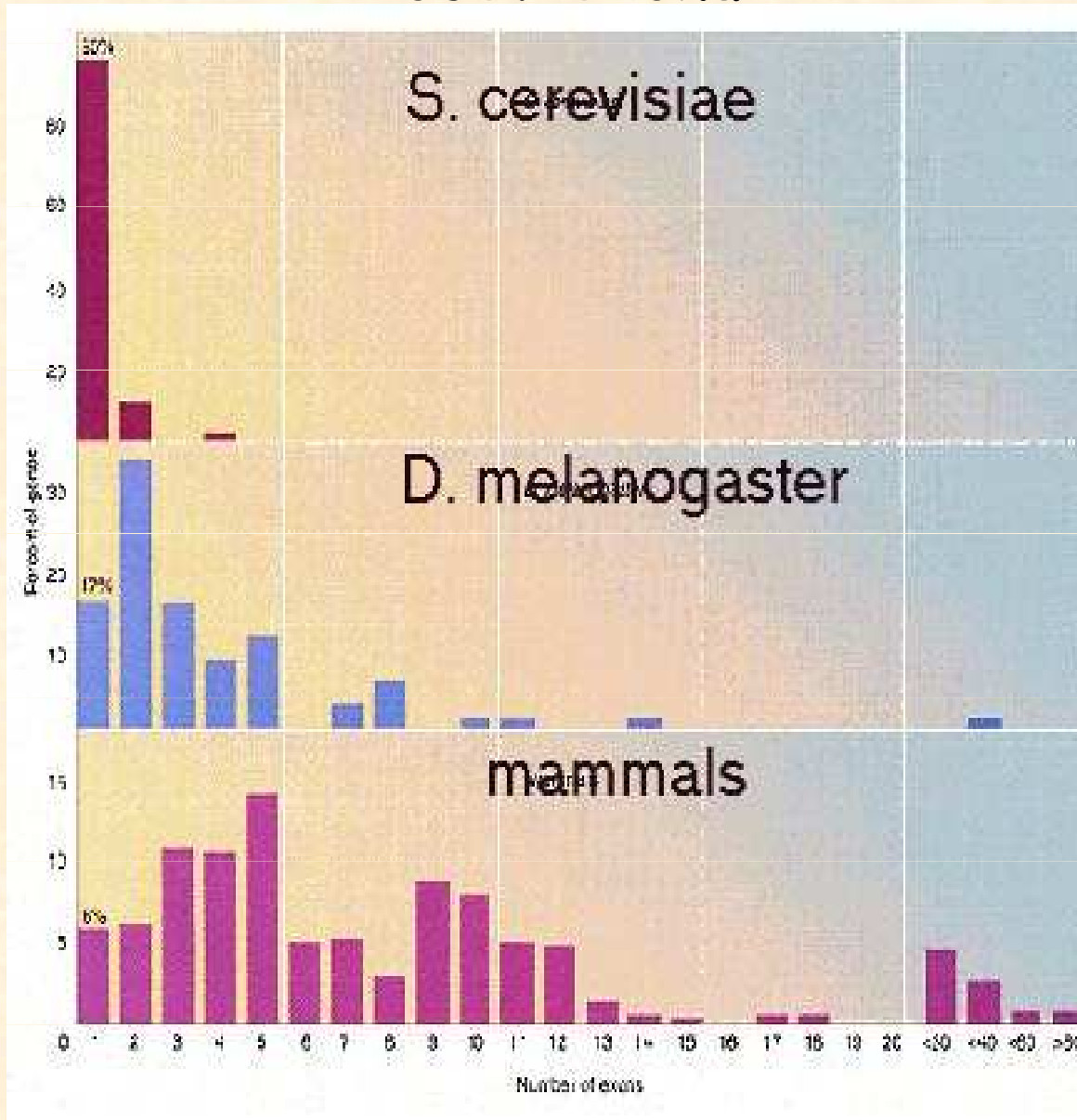
- tvorba rekombinantních proteinů
- mix klonovaných genů
- fragmentace a self-priming PCR
- cephalosporinase enzym
- rezistence k moxalactamu
- Citrobacter, Enterobacter aj.
- 0.75ug/ml --- 200ug/ml (250x)

# VELIKOSTI GENÚ

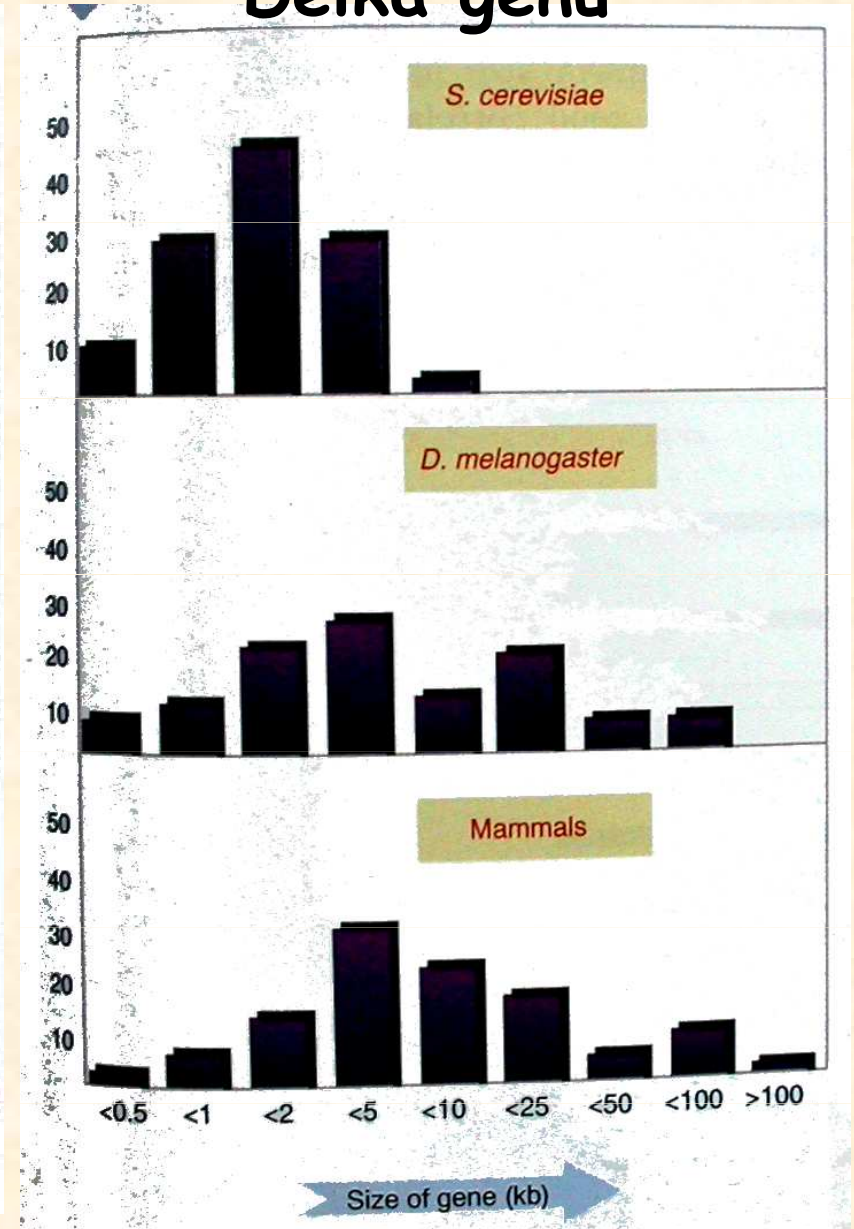


# Počty exonů jsou nejvyšší u savců

## Počet exonů



## Délka genu

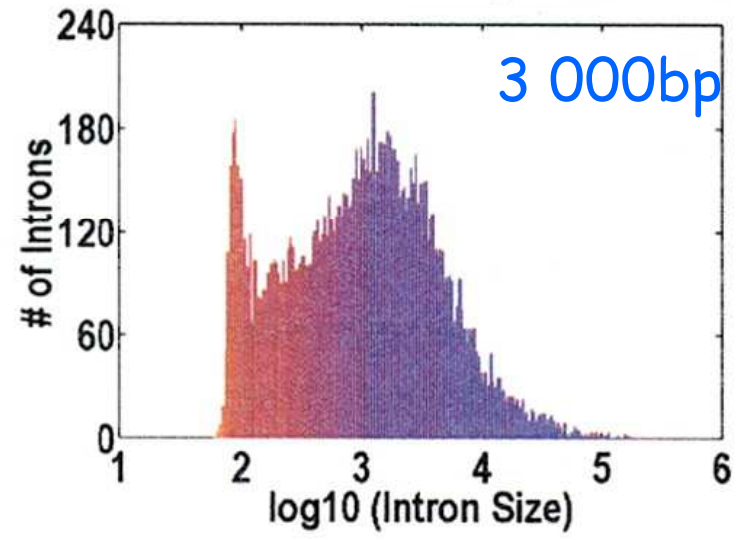


# Velikosti intronů



*Homo sapiens*

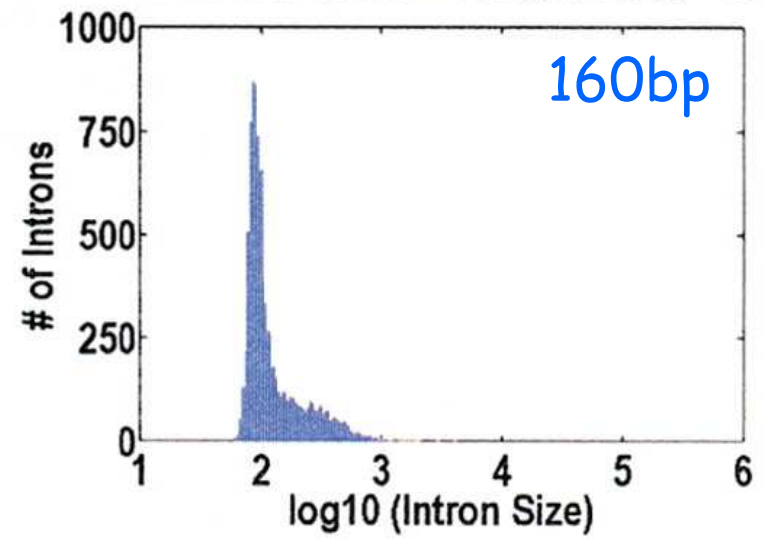
Intron Size: mean = 3116, median = 1044



(a)

*Arabidopsis thaliana*

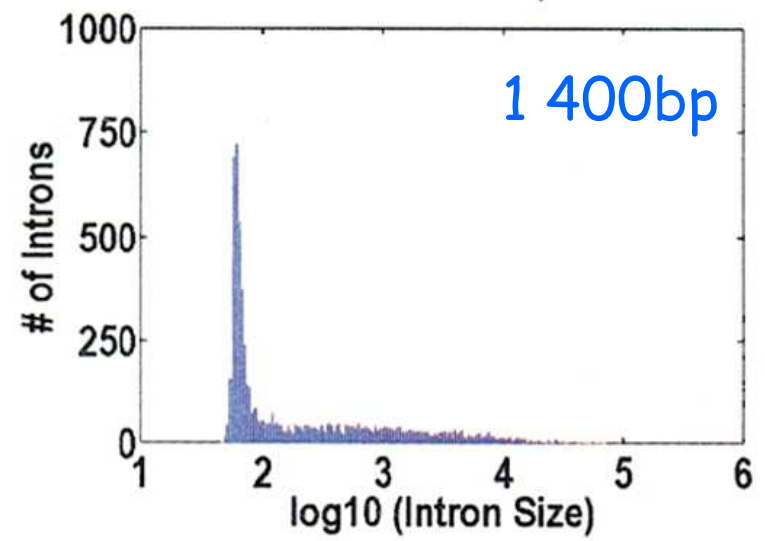
Intron Size: mean = 159.3, median = 98



(b)

*Drosophila melanogaster*

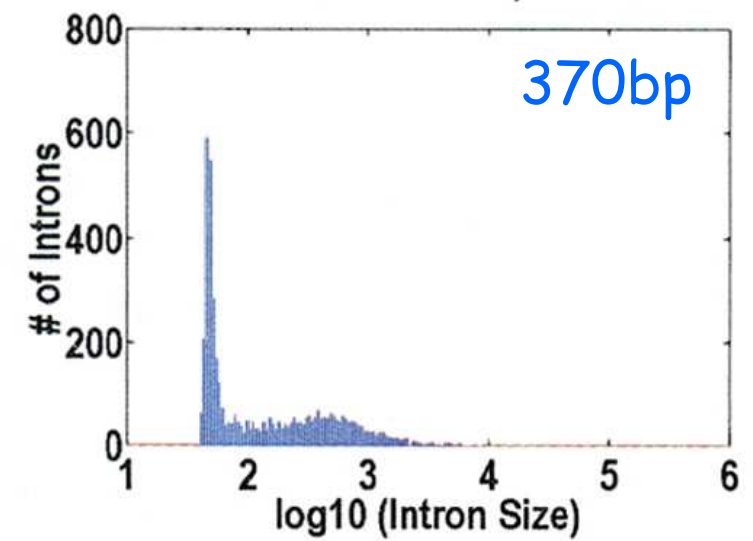
Intron Size: mean = 1411, median = 86



(c)

*Caenorhabditis elegans*

Intron Size: mean = 372, median = 85



(d)

# Nejmenší gen

- peptidy obsahující méně než 10 AK vznikají štěpením delších polypeptidů

## Příklady:

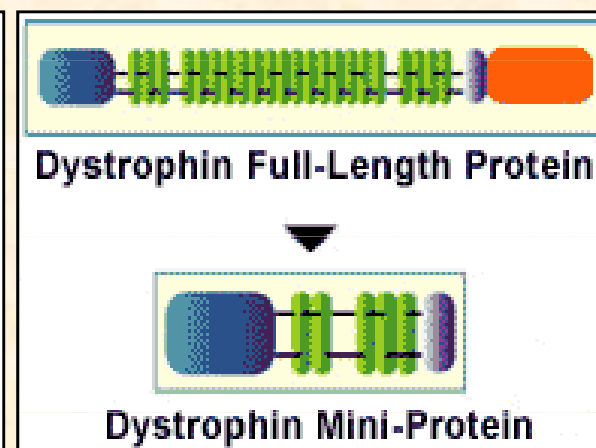
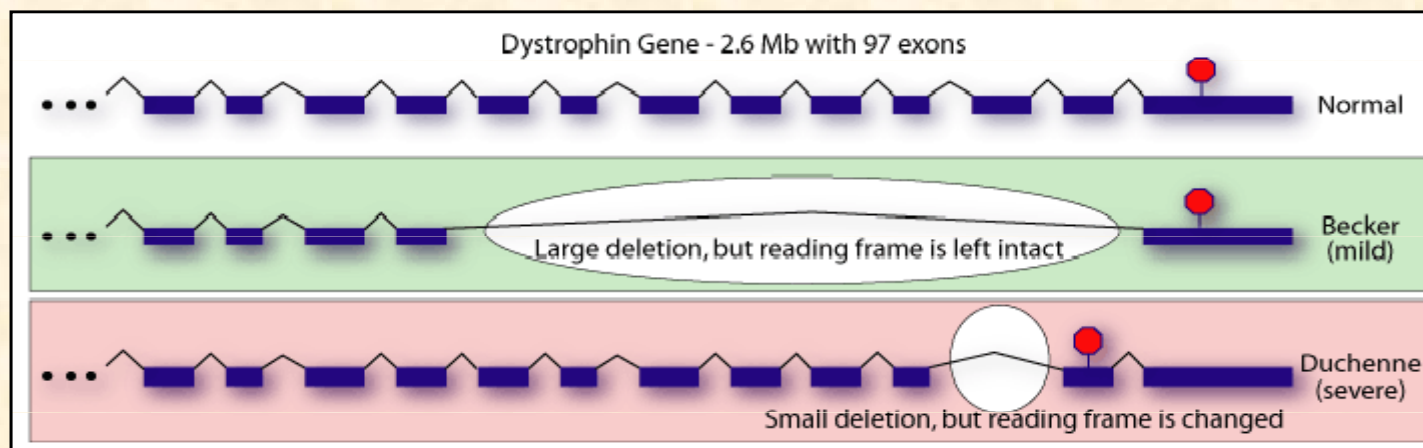
- polypeptidy 26 AK u *Staphylococcus* a *Bacillus subtilis*  
- gen pro peptidový inhibitor pohlavního feromonu *Enterococcus faecalis* má 66bp

## Microcin C7 - nejmenší gen:

- oligopeptid sestávající ze 7 AK - antibioticum microcin C7, přepisován spolu s dalším microcinovým genem, má samostatné vazebné místo na ribozóm umístěné před kodonem pro fMet

# Dystrofinový gen - obří gen

79 exonů, nejdelší známý gen  
8 promotorů, exprese ve svalech a mozku  
2.5 Mb dlouhý (0.1% genomu), 14kb mRNA  
delece: Duchenne MD nebo Becker MD  
Poloha Xp21, 1:3500 u mužů





# FOXP2 - důležitý lidský gen

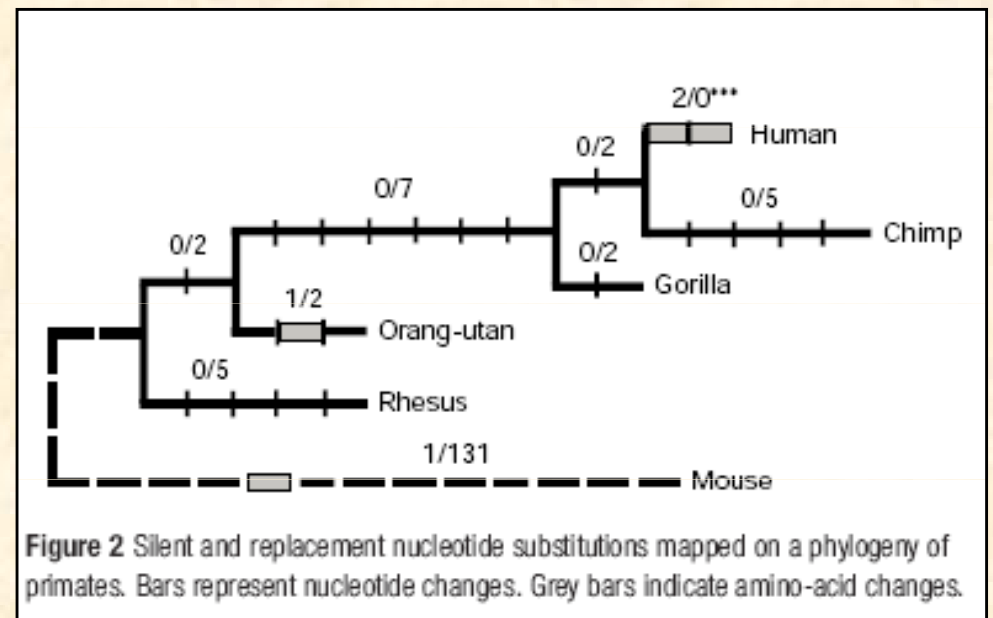
(forkhead box P2)

## FOXP2

- důležitá pro rozvoj řeči, mutace v rodinách s poruchou řeči
- souvislost s chápáním pojmů, gramatika, kreativita, kultura
- existuje i u jiných zvířat,
- nová lidská forma vznikla před 200 000 let
- délka 14 063 bp
- divergenece od šimpanze 0.87%, orangutana 2.5%

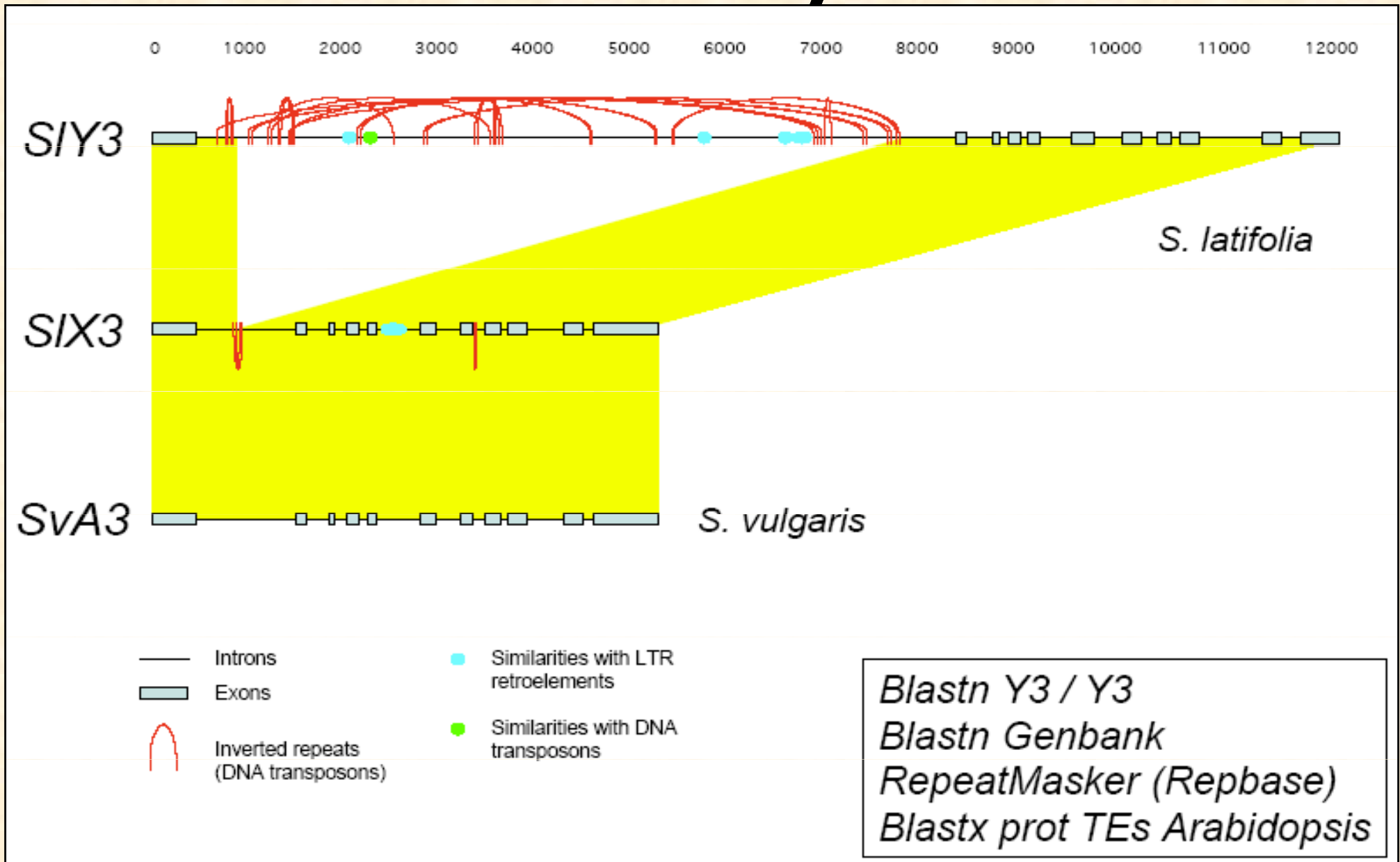
### Molecular evolution of *FOXP2*, a gene involved in speech and language

Wolfgang Enard\*, Molly Przeworski\*, Simon E. Fisher†, Cecilia S. L. Lai†, Victor Wiebe\*, Takashi Kitano\*, Anthony P. Monaco† & Svante Pääbo\*





# Geny na chromosomu Y mají delší introny



**INTRONY - STARÉ  
NEBO MLADÉ**

# Hypotézy původu intronů

## „Intron first“:

- původní organizmy obsahovaly introny
- prokaryota je ztratila

## „Intron late“:

- původní organizmy introny neobsahovaly
- eukaryota je získala

## Význam intronů:

1. Introny **užitečné nejsou**, ale organizmy se jich nedokáží zbavit
2. Introny **mají funkční** význam pro organizmy, jsou užitečné

# Introny byly v genech již na počátku ("intron first")

- studium vnitřní periodicity genů - stejné motivy v exonech i v sousedních intronech
- malá pravděpodobnost dlouhých úseků bez stop-kodonů,
- evoluční výhoda enzymatického aparátu, který vystřihne oblasti se stop-kodony a sestaví dlouhou mRNA

# Introny byly do genů vloženy až dodatečně ("intron late")

- Existuje řada různých intronů lišících se mechanismem vystřihování z RNA - vznikaly nezávisle
- Distribuce intronů v rámci fylogenetických stromů svědčí o dodatečném vložení spíše než o opakovaném nezávislém vymizení





# Introny jsou genomovými parazity

- Šíří se pouze v rámci genomu, vertikální přenos, aby nezabíjeli buňku, před translací se vystřihnou
- Samosestřih
- Splicesom - komplex kódovaný buňkou, původně parazitickými introny, kódují enzymy pro horizontální šíření v rámci genomu

# Introny jsou **užitečné** pro organizmy

## 1. Zvyšují evoluční potenciál organismu

- souvisí se vznikem eukaryot, v pozadí adaptivní radiace eukaryot,
- nenáhodná distribuce, odděluje funkční domény proteinů,
- stavební charakter genů urychluje evoluci nových proteinů,
- snižuje pravděpodobnost rekombinace v exonech (doménách)

## 2. Souvisí s existencí histonů

- oblasti v kontaktu s histony nepřístupné
- introny zpřístupňují regulační oblasti

## 3. Umožňují detekci, případně i reparaci mutací v exonech

- detekce chyb při přenosu informace, příklad „liché parity“
- introny jako kontrolní sekvence, sekundární struktura

## 4. Snižují riziko nelegitimní rekombinace

- paralogy a riziko nelegitimní rekombinace, nefunkční geny
- včlenění intronů do různých míst diferencuje geny