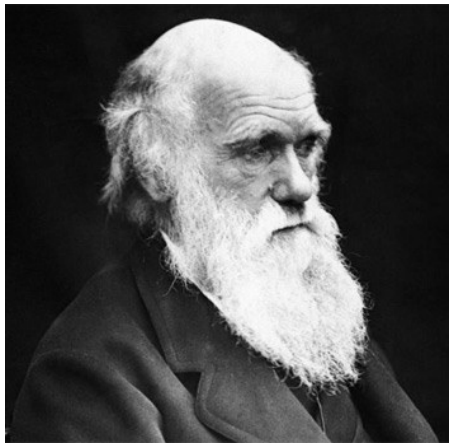
Evolutionary genetics
2/1 Zk/Z

Radka Reifová

Evolutionary genetics

- Combines Darwin's evolutionary theory (1859) and Mendel's theory of inheritance (1866).
- Aims to understand the diversity of life from a genetic perspective, using DNA sequences to study population history and evolution.



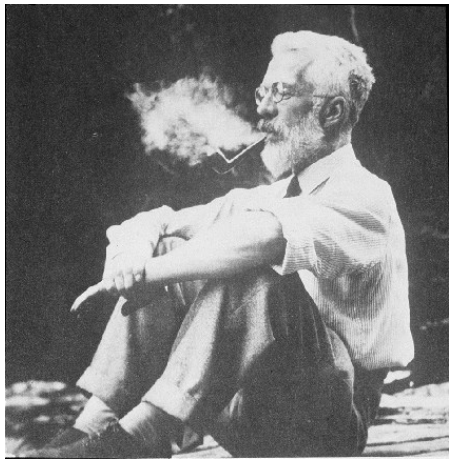
Charles Darwin

Johann Gregor Mendel

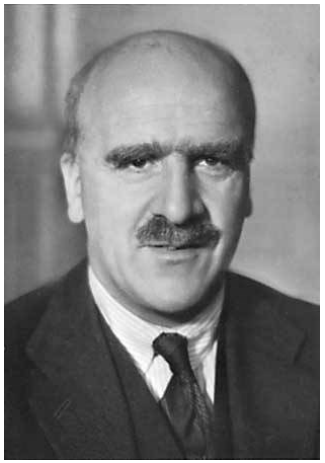
Milestones of evolutionary genetics

Modern synthesis
(1930s–1940s)

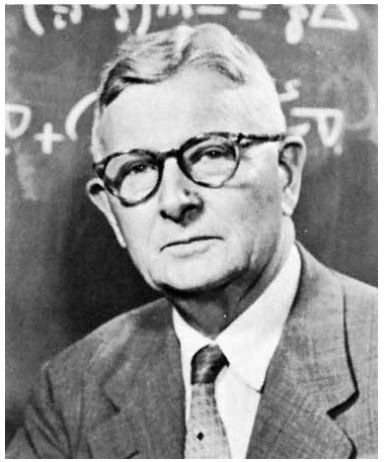
- synthesis between Mendelian genetics and Darwin's selection theory



Sir Ronald Fisher



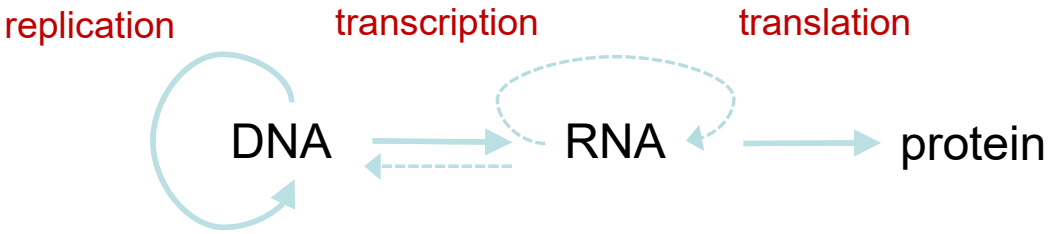
J. B. S. Haldane

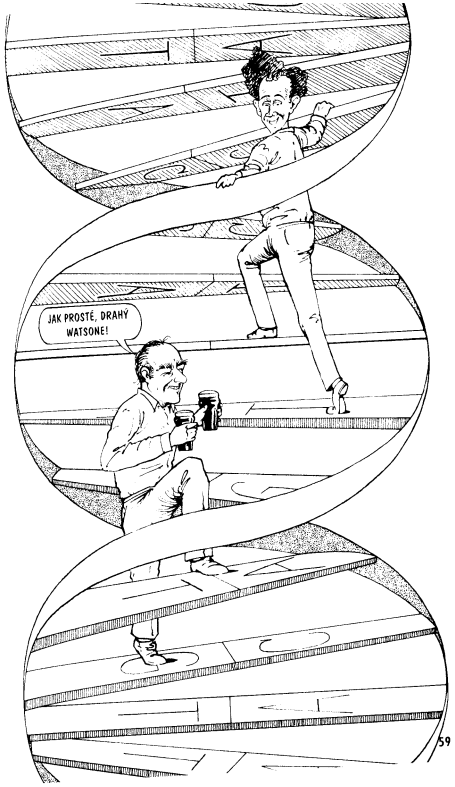


Sewall Wright

Milestones of evolutionary genetics

Discovery of DNA sturcture (1953)
Central dogma of molecular biology (1950s)





James Watson a Francis Crick

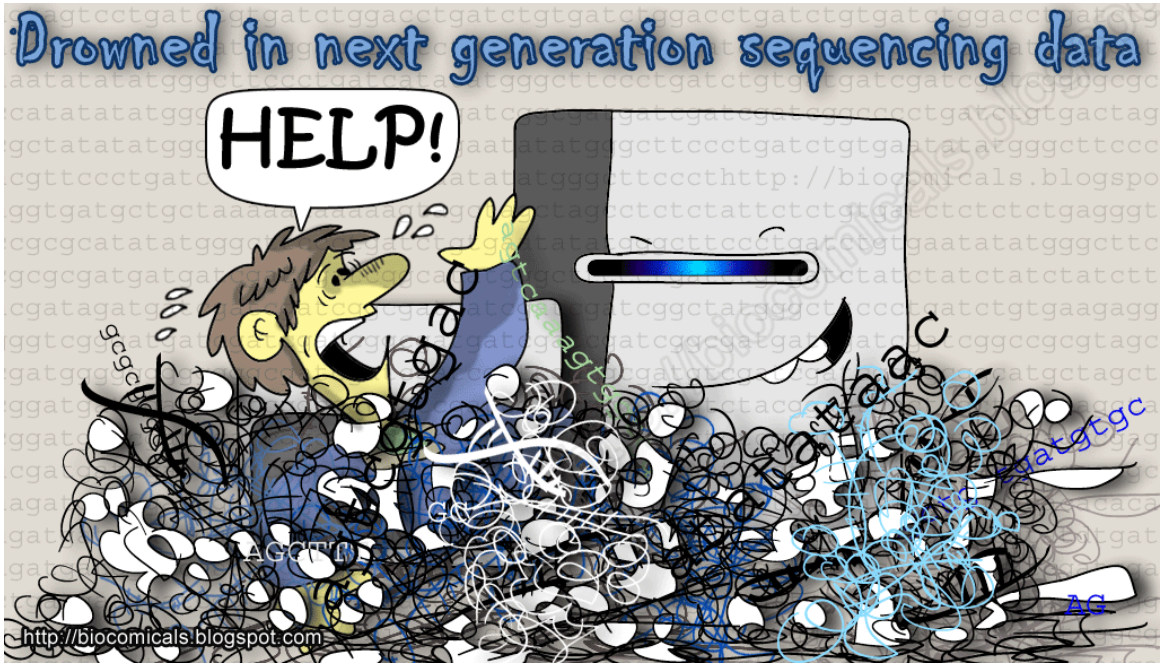
Theory of molecular evolution
(1970s)

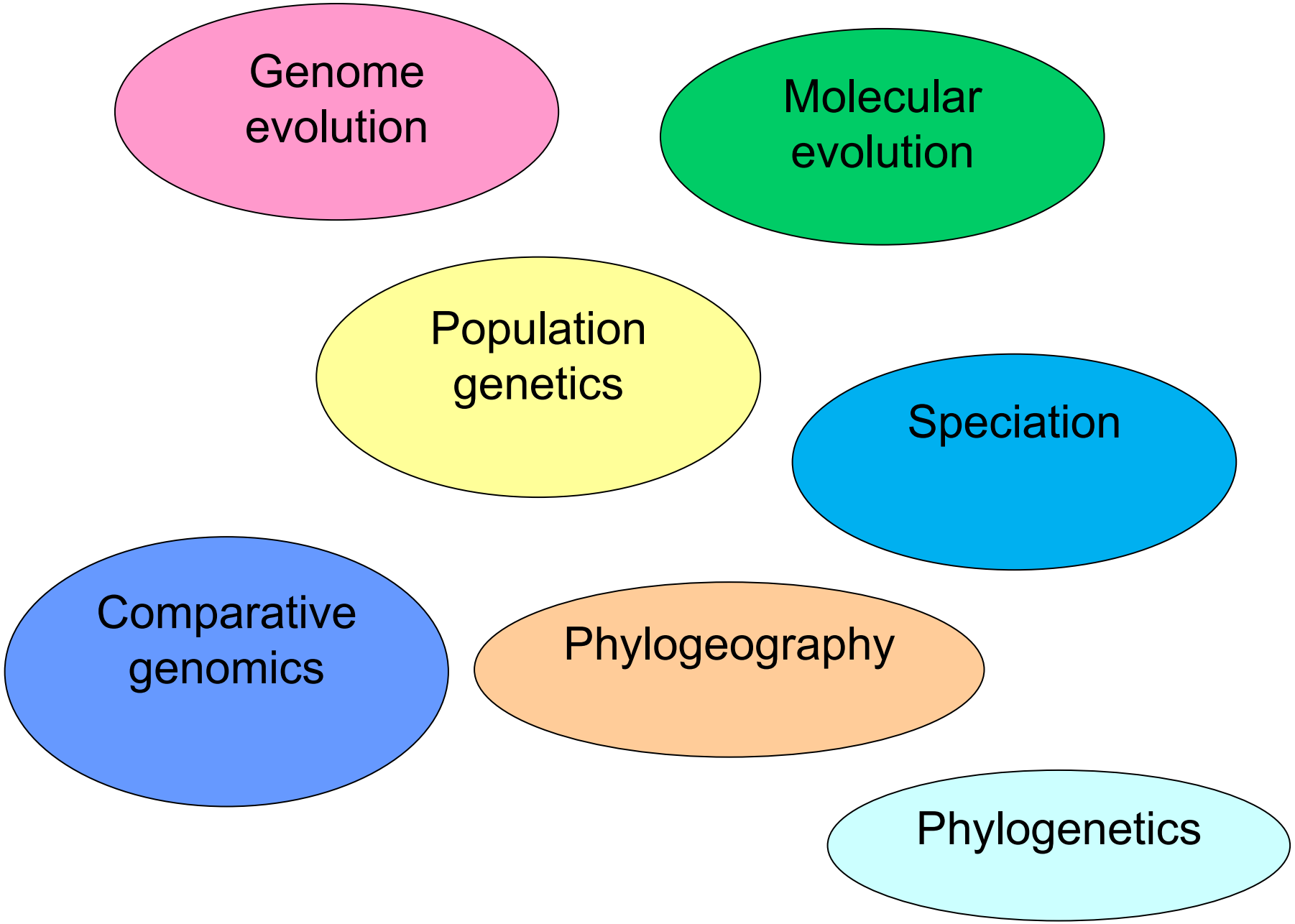
Motto Kimura

Milestones of evolutionary genetics

Discovery and development of sequencing technologies
(since 1970s)

- Sanger sequencing
- Next-generation sequencing methods

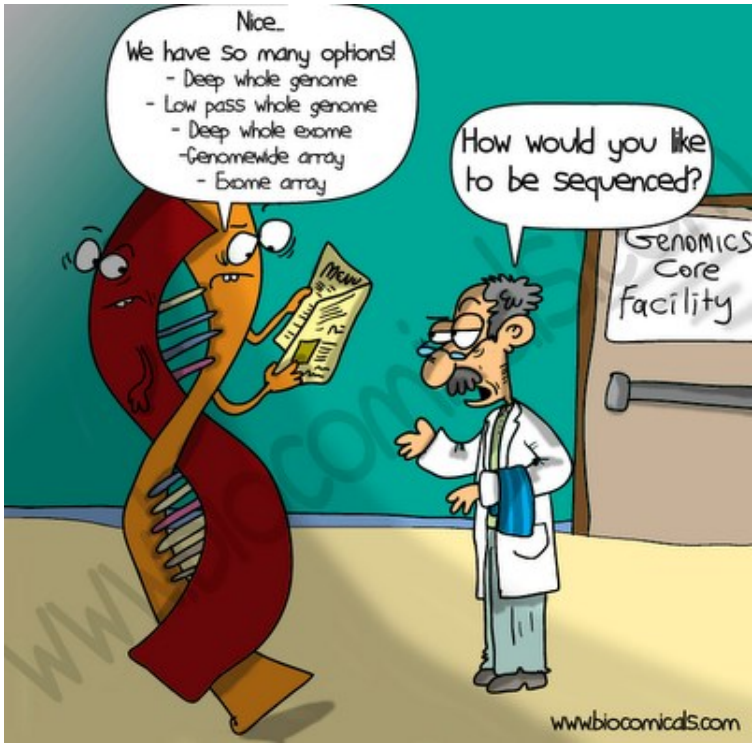


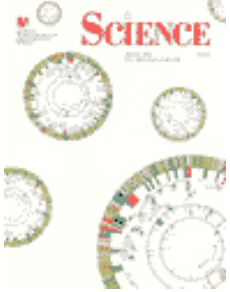

Genome evolution
Molecular evolution
Population genetics
Speciation
Comparative genomics
Phylogeography
Phylogenetics

DNA sequence data

- Sequences of individual genes
- Sequences of transcriptomes
- Restriction site associated DNA (RAD) sequences
- Whole genome sequences

  ……





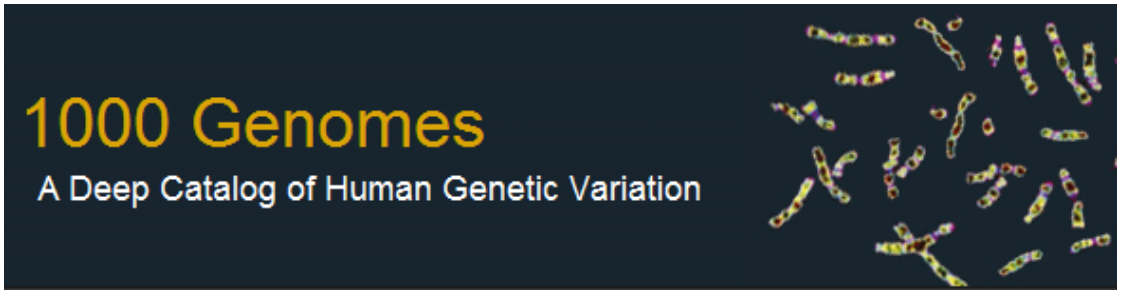
Genome sequences

- 1995 *Haemophilus influenze*
- 1996 *Saccharomyces cerevisiae*
- 1998 *Caenorhabditis elegans*,
- 2000 *Drosophila melanogaster*
- 2001 *Homo sapiens*
- 2002 *Mus musculus*
- 2005 *Pan troglodydes*
- 2010 *Homo neanderthalensis*

GENOME 10K©

i5K

- 2012

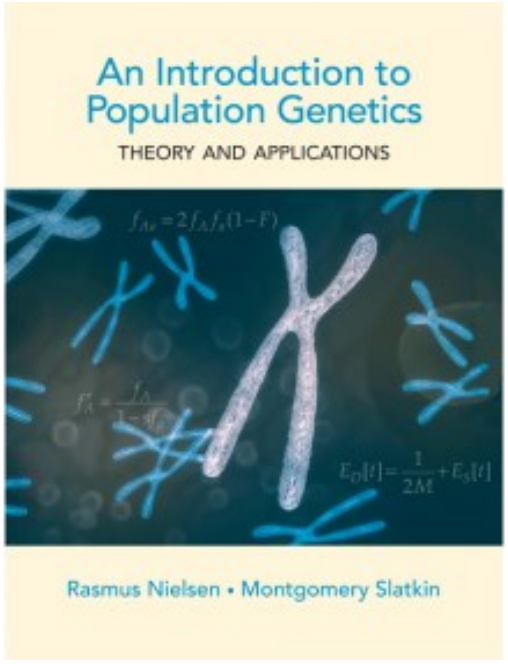


Schedule

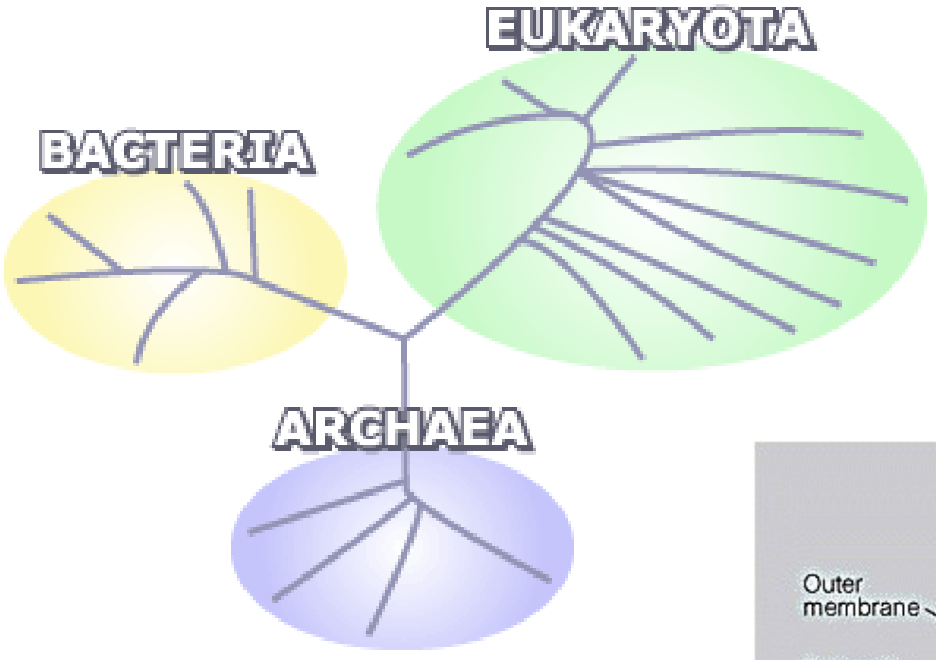
2.10.	Genes and genomes
9.10.	-
16.10.	-
23.10.	Inheritance
30.10.	**Students' presentations**
6.11.	Intoduction to population genetics
13.11.	Neutral evolution, molecular clocks
20.11.	Selection
27.11.	**Students' presentations**
4.12.	Functional genetics
11.12.	Speciation
18.12.	Gene genealogies and phylogenies
8.1.	**Students' presentations**

Materials for the exam

- **Presentations from lectures, student's presentations, recommended articles**
- An Introduction to Population Genetics (2013)
  Rasmus Nielsen and Montgomery Slatkin
- Other books about population and evolutionary genetics

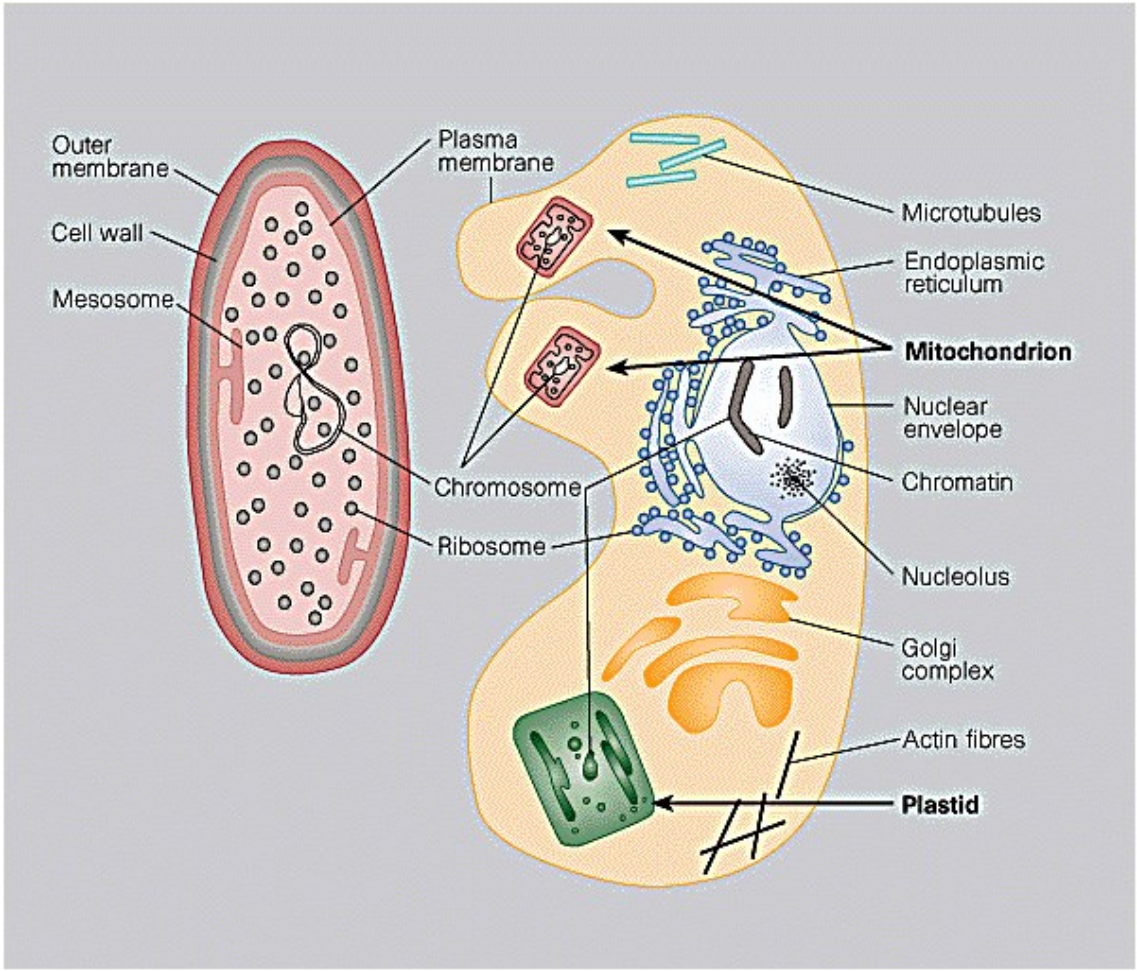


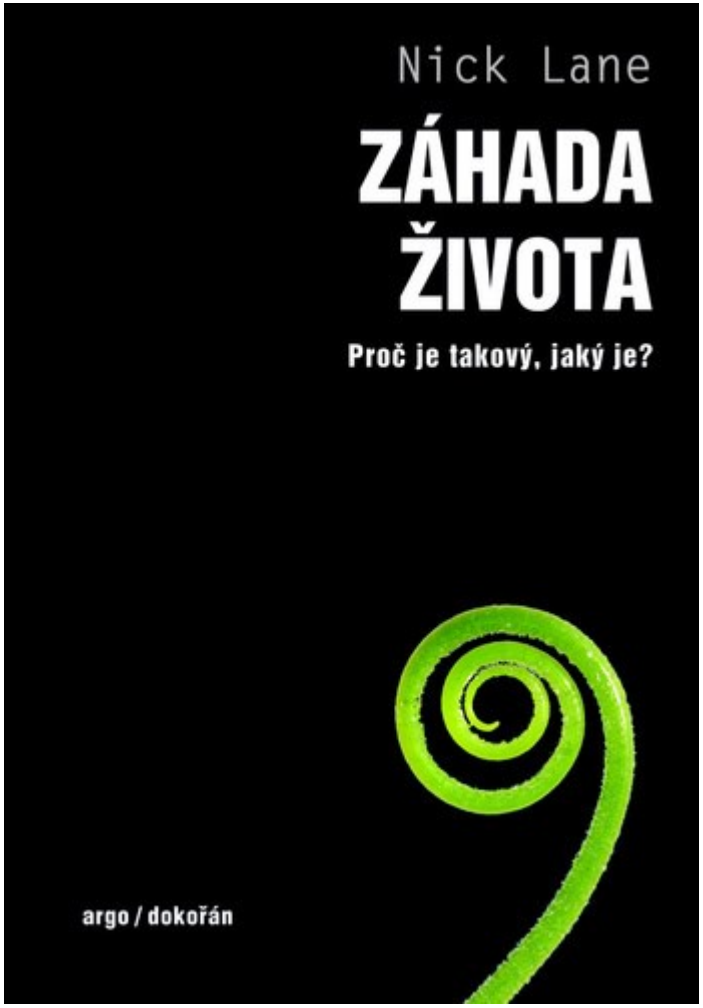
Genes a genomes

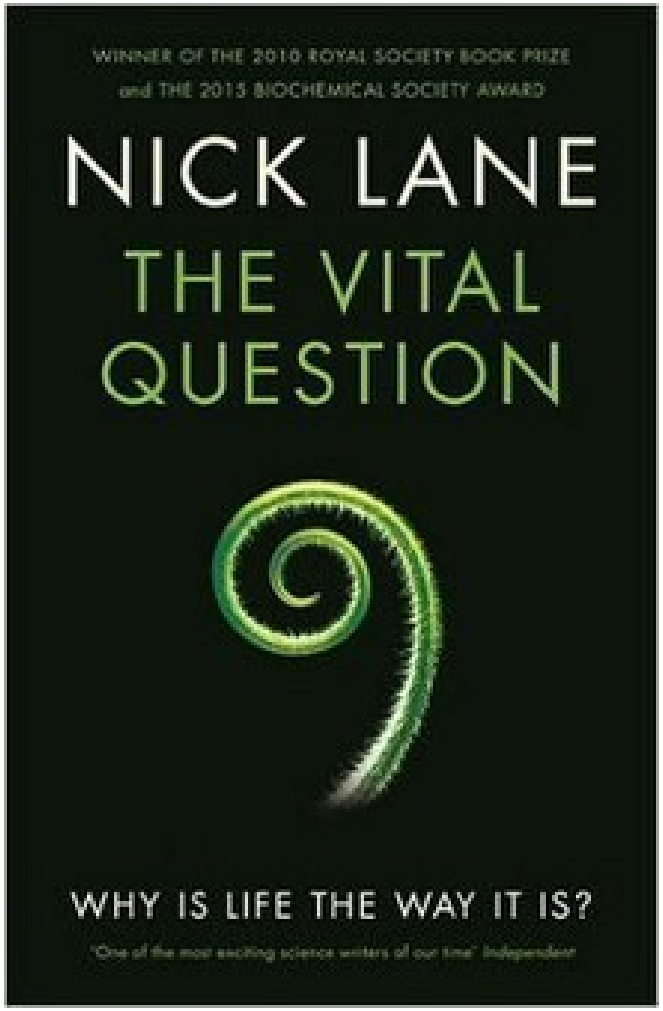


Genome of eukaryota, bakteria and archea

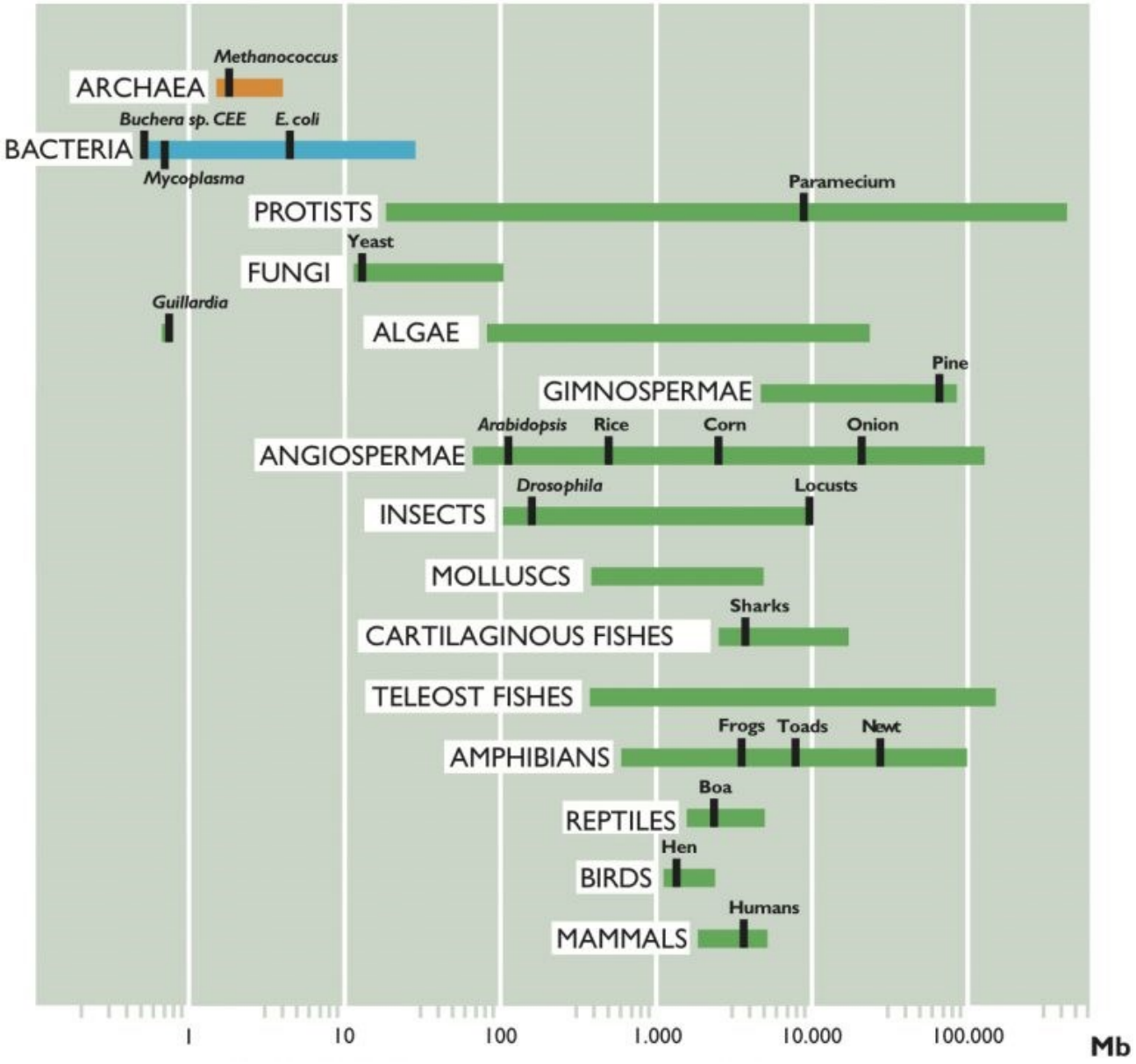
Endosymbiosis theory



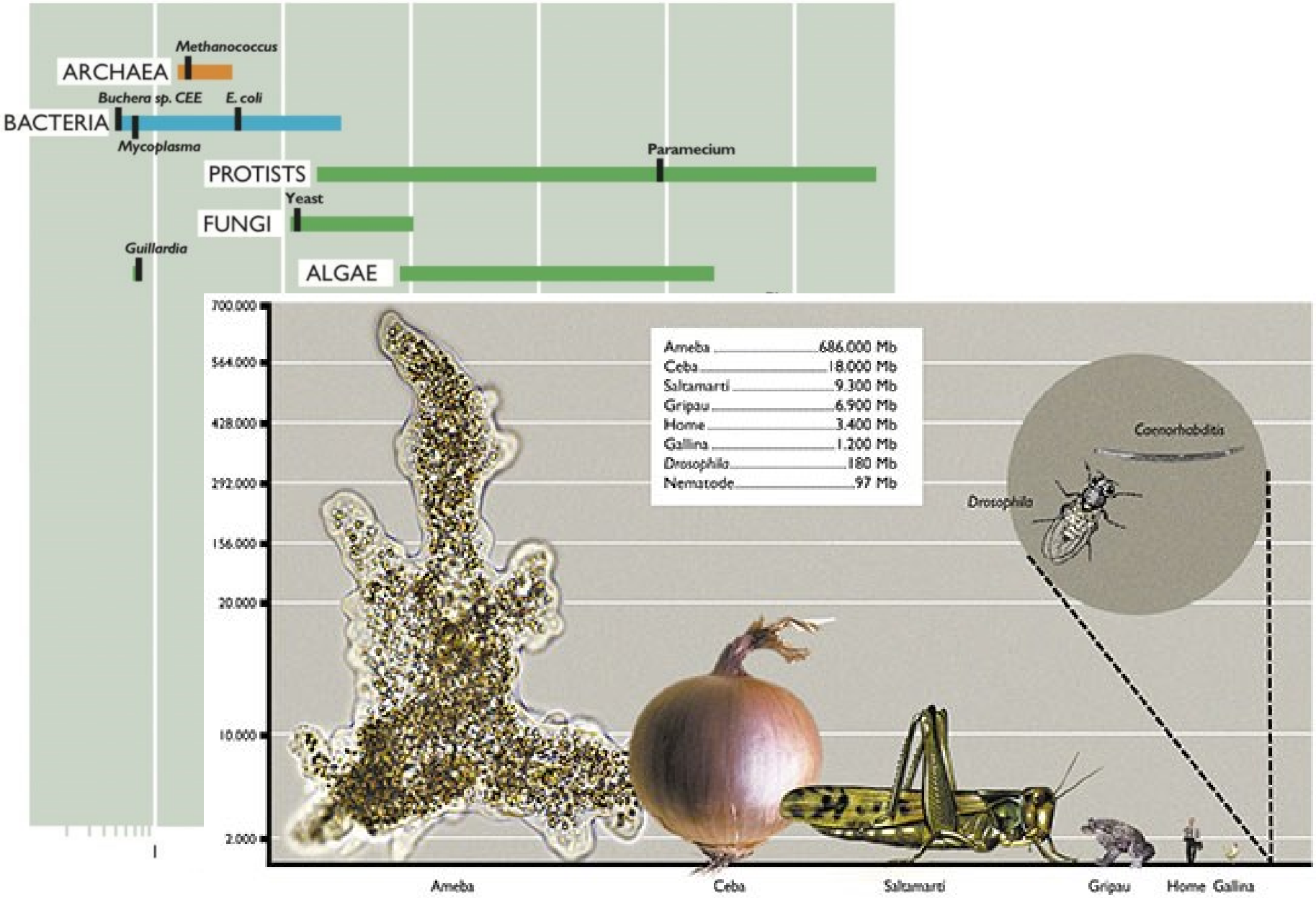

Nick Lane
ZÁHADA ŽIVOTA
Proč je takový, jaký je?
argo / dokořán


WINNER OF THE 2010 ROYAL SOCIETY BOOK PRIZE
and THE 2015 BIOCHEMICAL SOCIETY AWARD
NICK LANE
THE VITAL QUESTION
WHY IS LIFE THE WAY IT IS?

Genome size and C value paradox

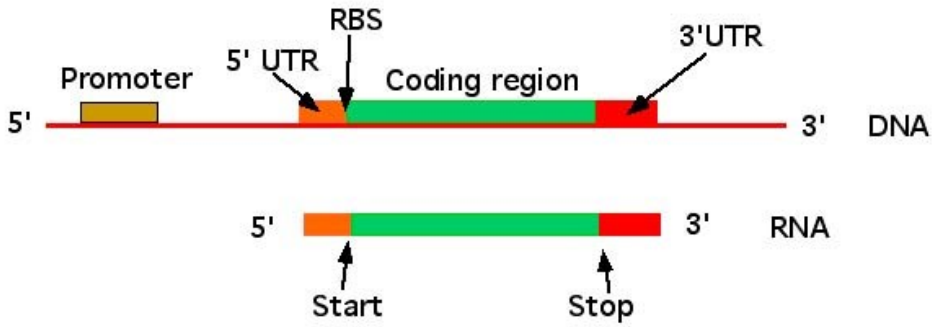


Genome size and C value paradox

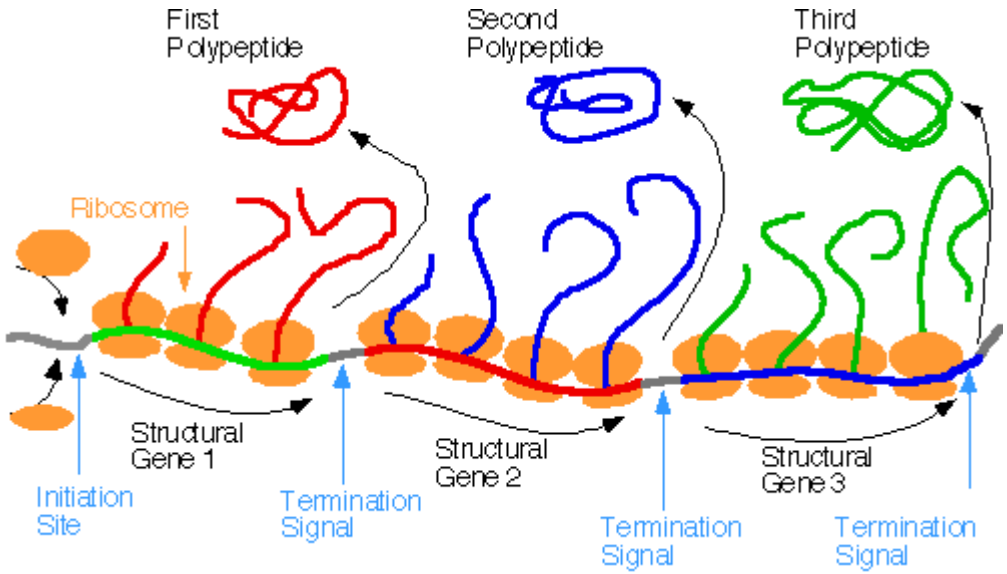


Protein coding genes – bacteria and archea

- Simple genes (without introns).

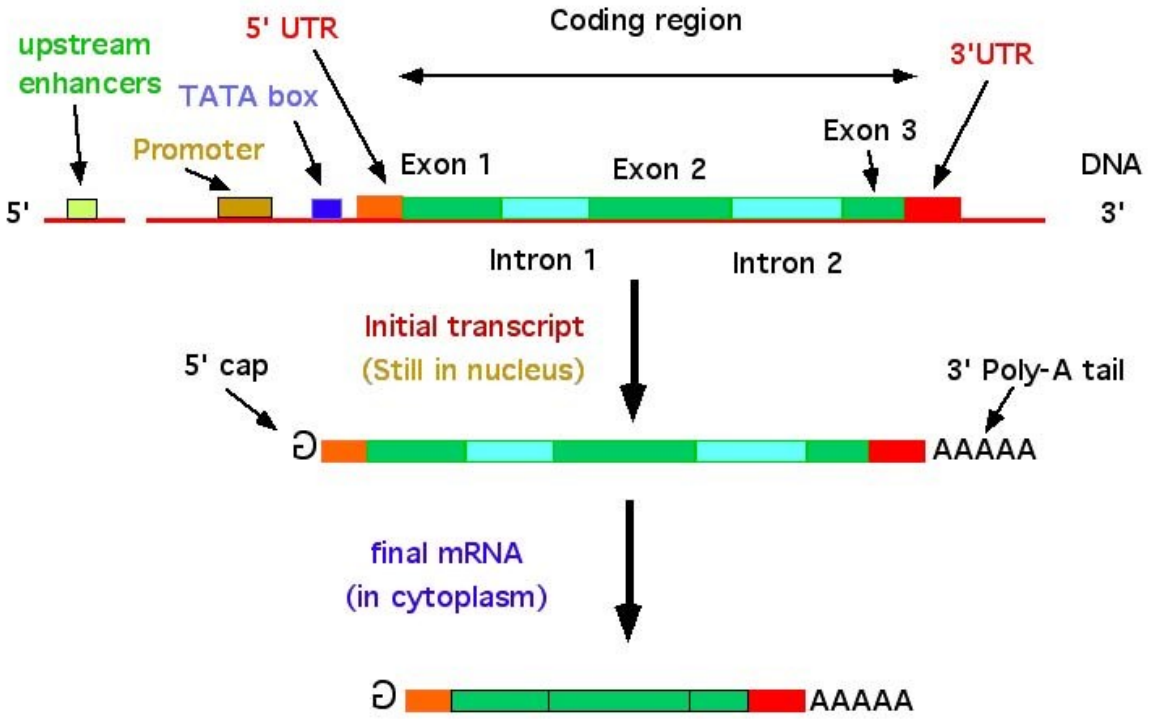


- Operons



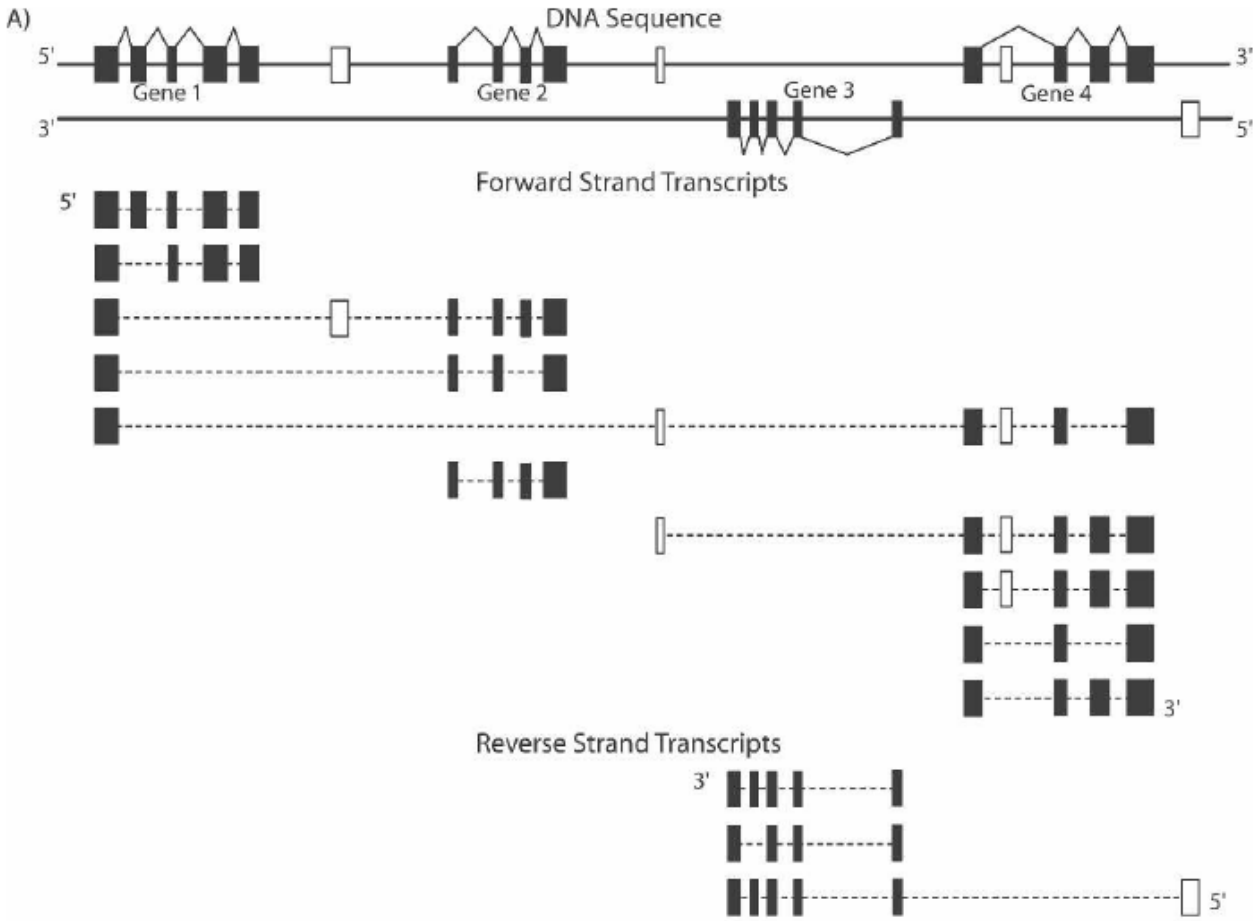
Protein coding genes – eukaryotes

- Contains exons and introns
- Splicing by spliceosome
- Alternative splicing, more proteins from a single gene
- Regulatory regions (promoters, enhancers)



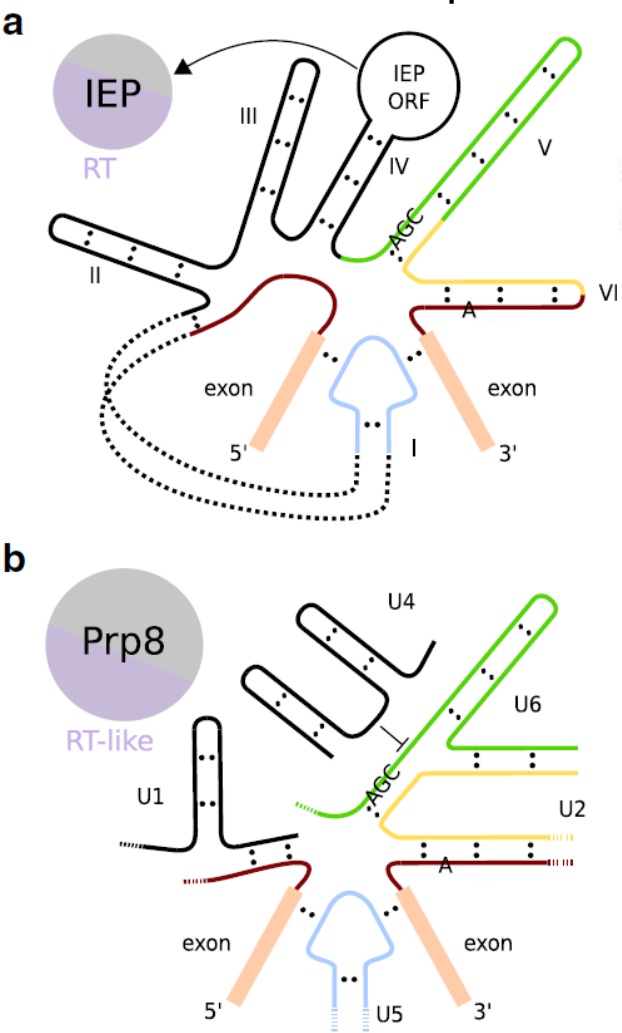
Alternative splicing
Alternative promoters and terminators of transcription

- On average 6.3 alternative transcripts per gene

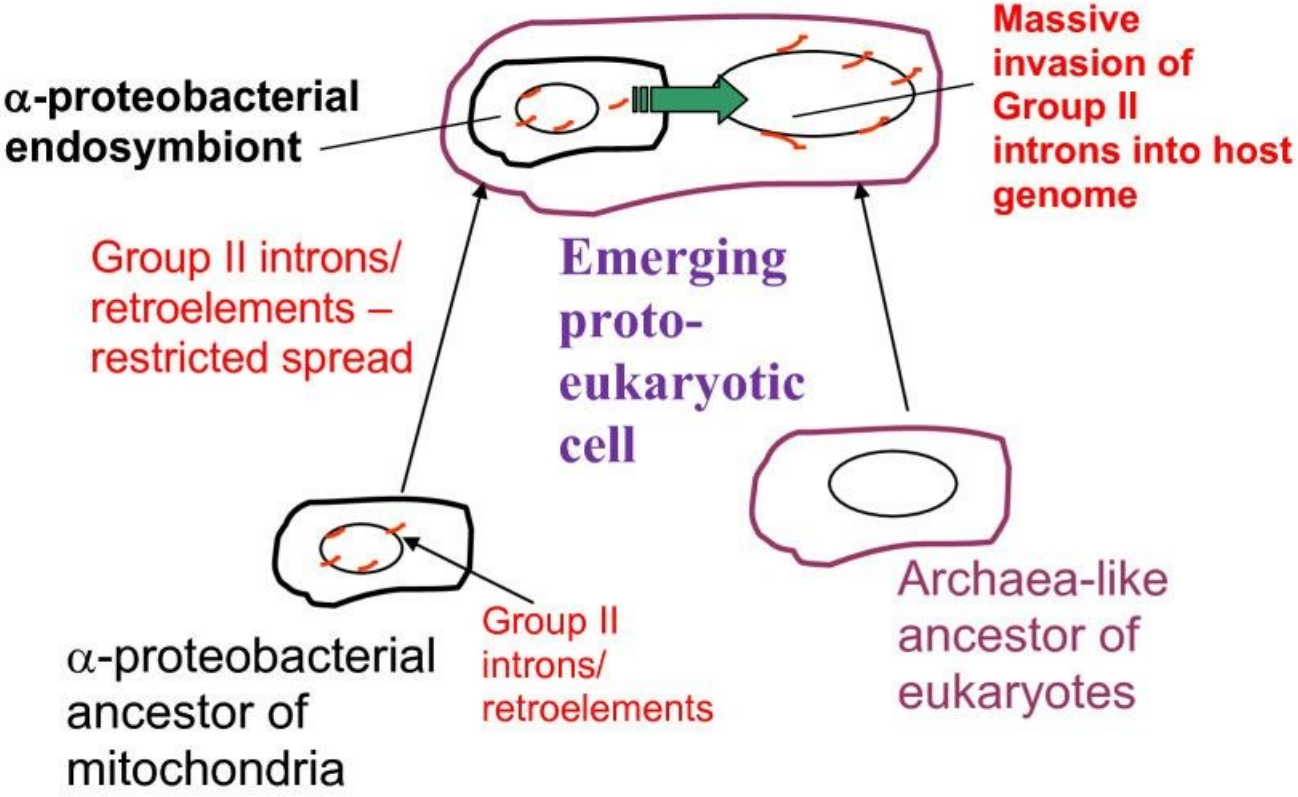


Origin of spliceosome from bacterial selfsplicing introns

Intron encoded protein

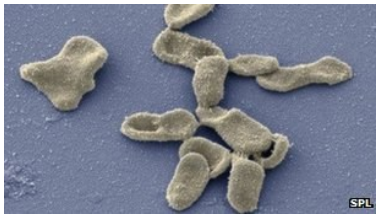


Part of spliceosome



Numbers of protein coding genes in genomes

Mycoplasma genitalium

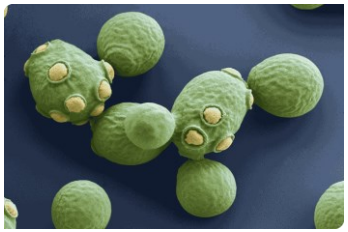


470

Haemophilus influenzae

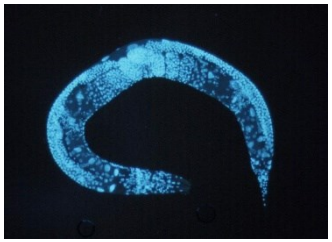
2 000

Saccharomyces cerevisiae



7 000

Caenorhabditis elegans



20 000

Drosophila melanogaster

14 000

Homo sapiens

21 000

Mus musculus

23 000

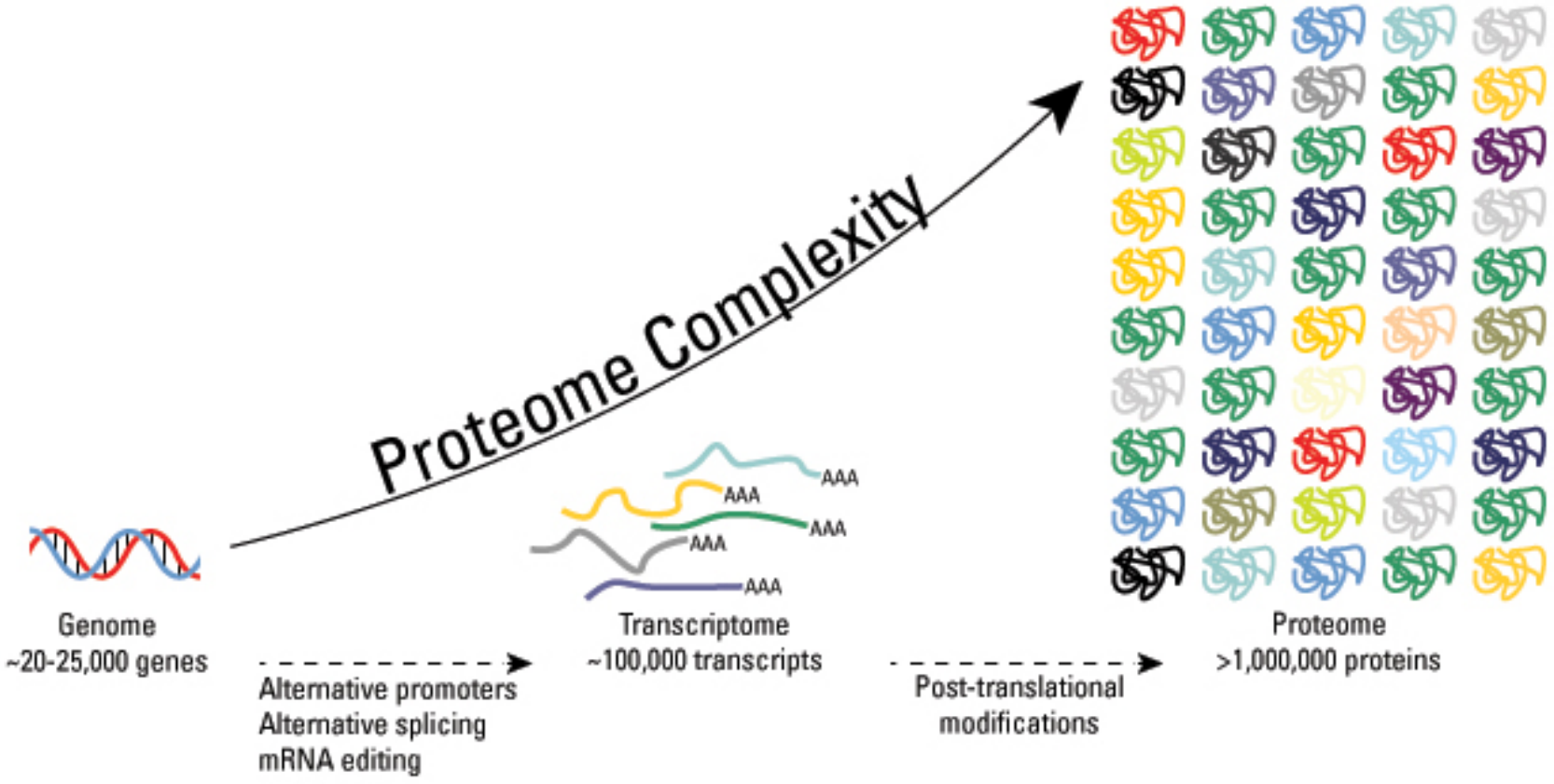
Danio rerio

26 000

Arabidopsis thaliana

25 000

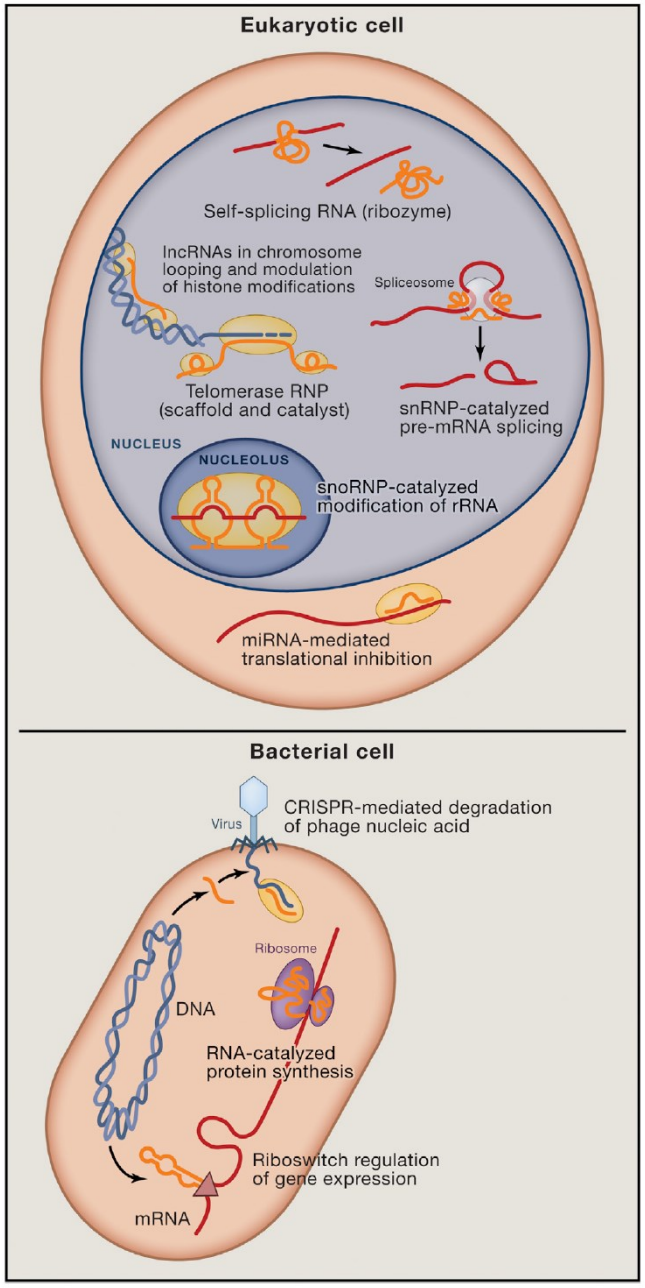
Alternative splicing, alternative promoters and post-translational modifications increase proteome complexity



RNA genes

- **genes for functional non-coding RNAs**

- Regulation of gene expressions
- Epigenetic inheritance
- Ribozymes (catalyze biochemical reactions)
- Possible remnants of RNA world



Small non-coding RNAs

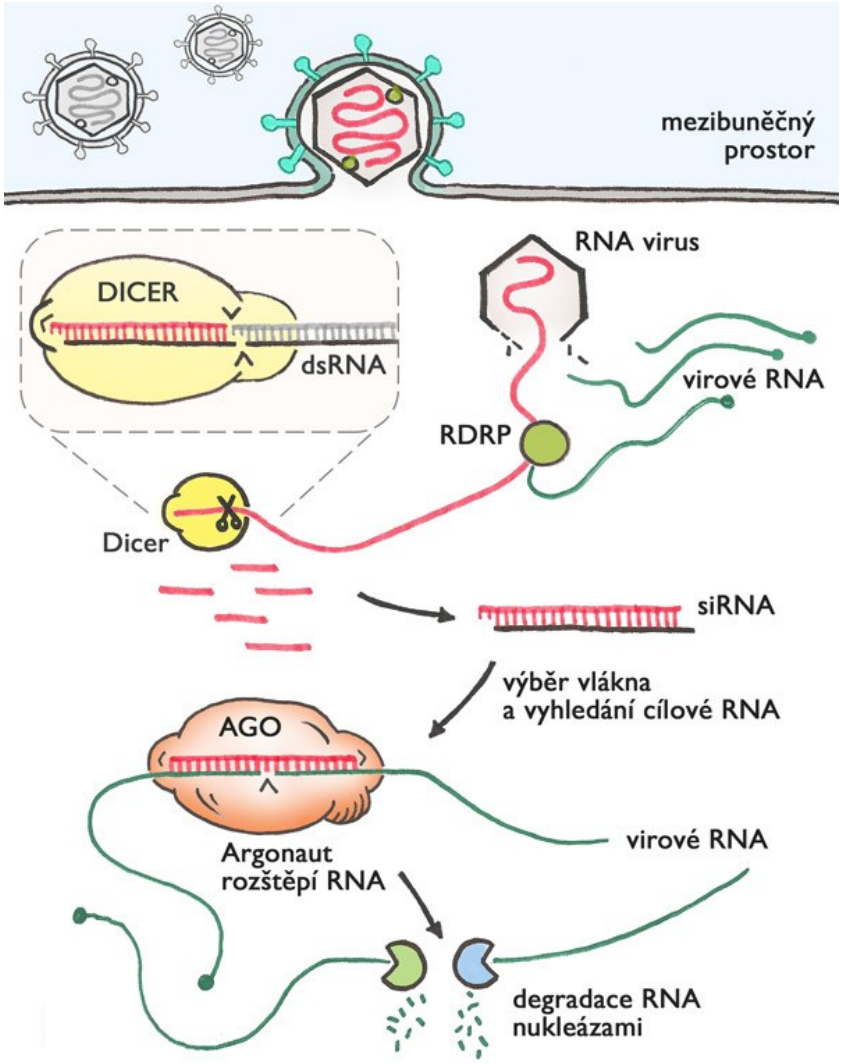
- Usually shorter than 100 nt, regulatory ncRNA ~20-30 nt long

- Ribosomal RNA (rRNA)
- Transfer RNA (tRNA)
- Small nuclear RNA (snRNA)
- Small nucleolar RNA (snoRNA)

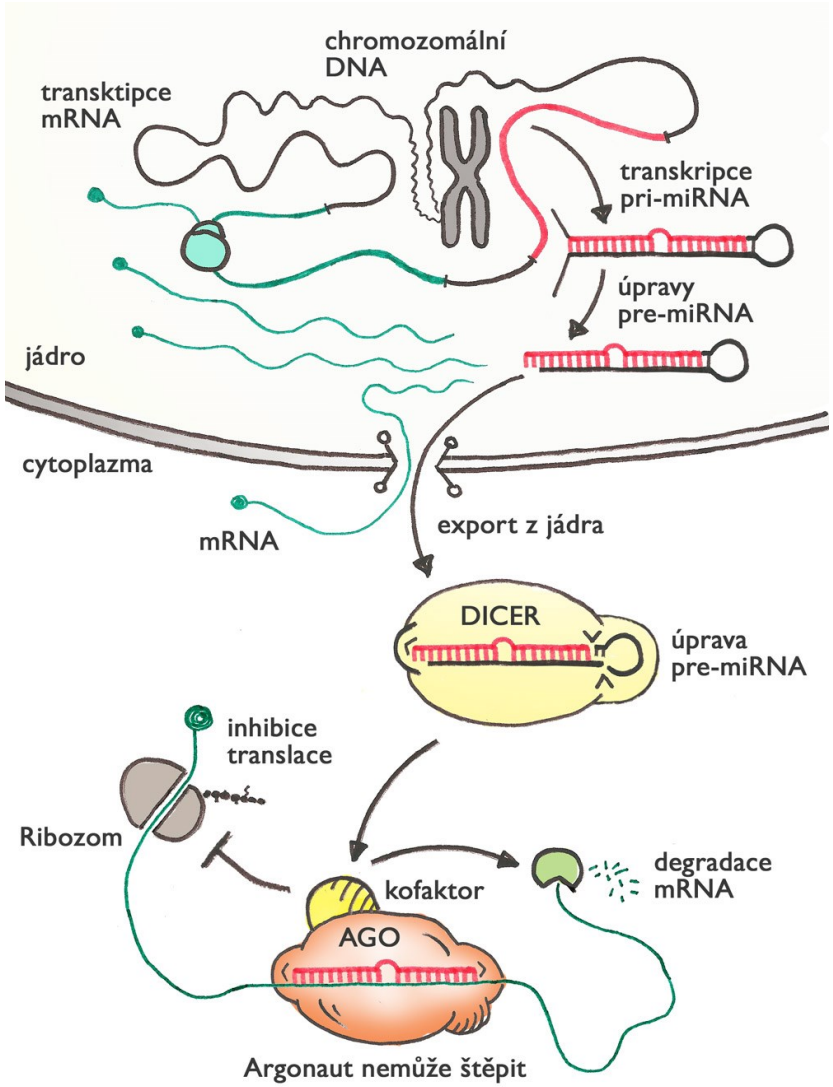
Regulatory RNA

- Small interfering RNA (siRNA) – defence against RNA viruses
- Micro RNA (miRNA) – regulation of gene expression
- Piwi interacting RNA (piRNA) – defence against transposones

siRNA



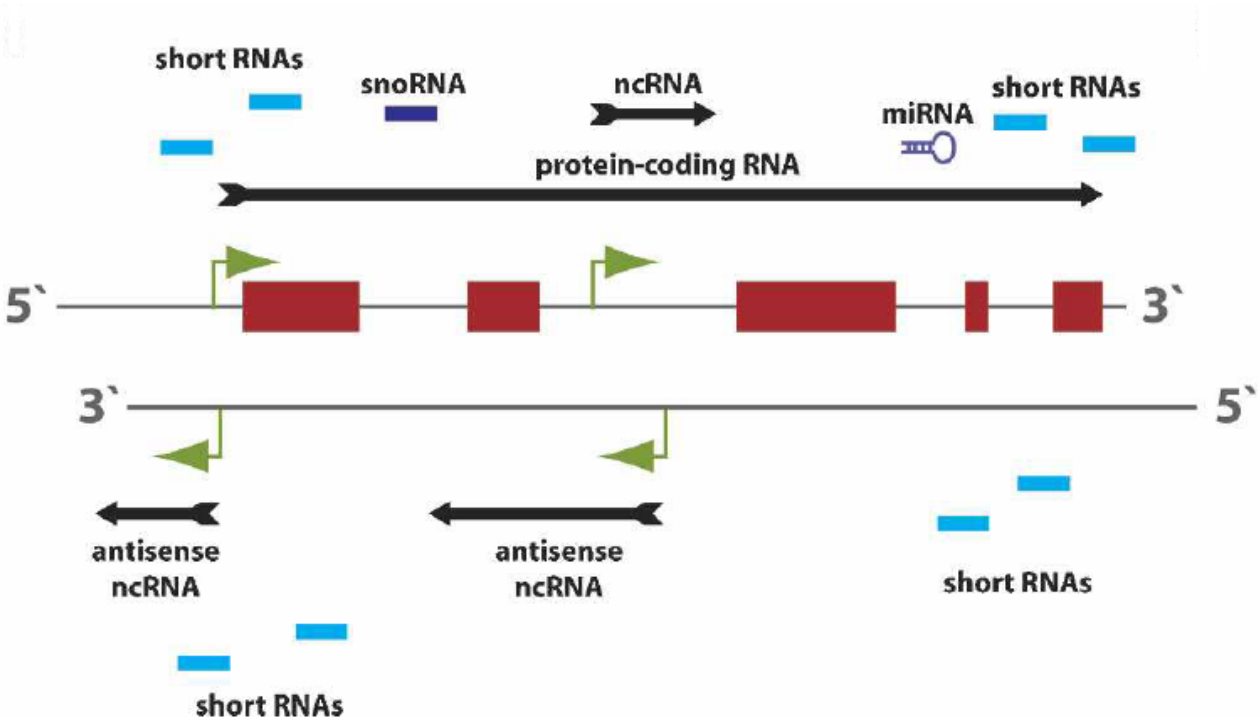
miRNA



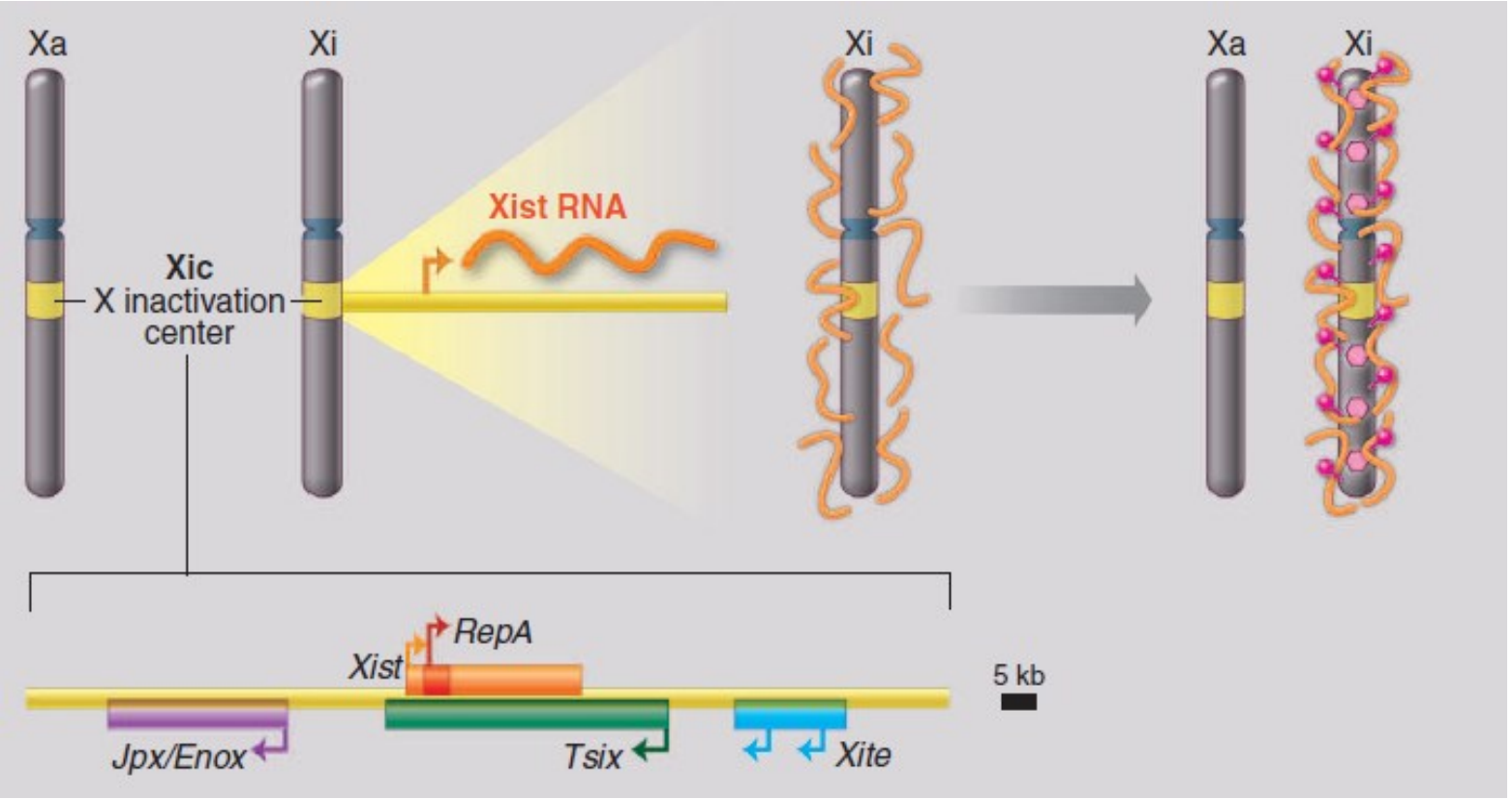
Petr Svoboda a Radek Jankele, Vesmír, 2015

Long non-coding RNAs

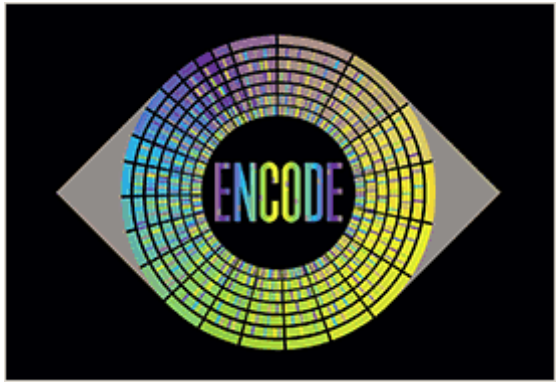
- Usually longer than100 nt
- Function of many of them unknown
- Often overlap with coding genes
- Regulation of gene expression



Xist non-coding RNA



Lee, Science (2012).



ENCODE (Encyclopedia of DNA Elements)

a project which aims to identify functional elements in the human genome (protein coding genes, RNA genes, regulatory regions etc.)

ARTICLE

doi:10.1038/nature11247

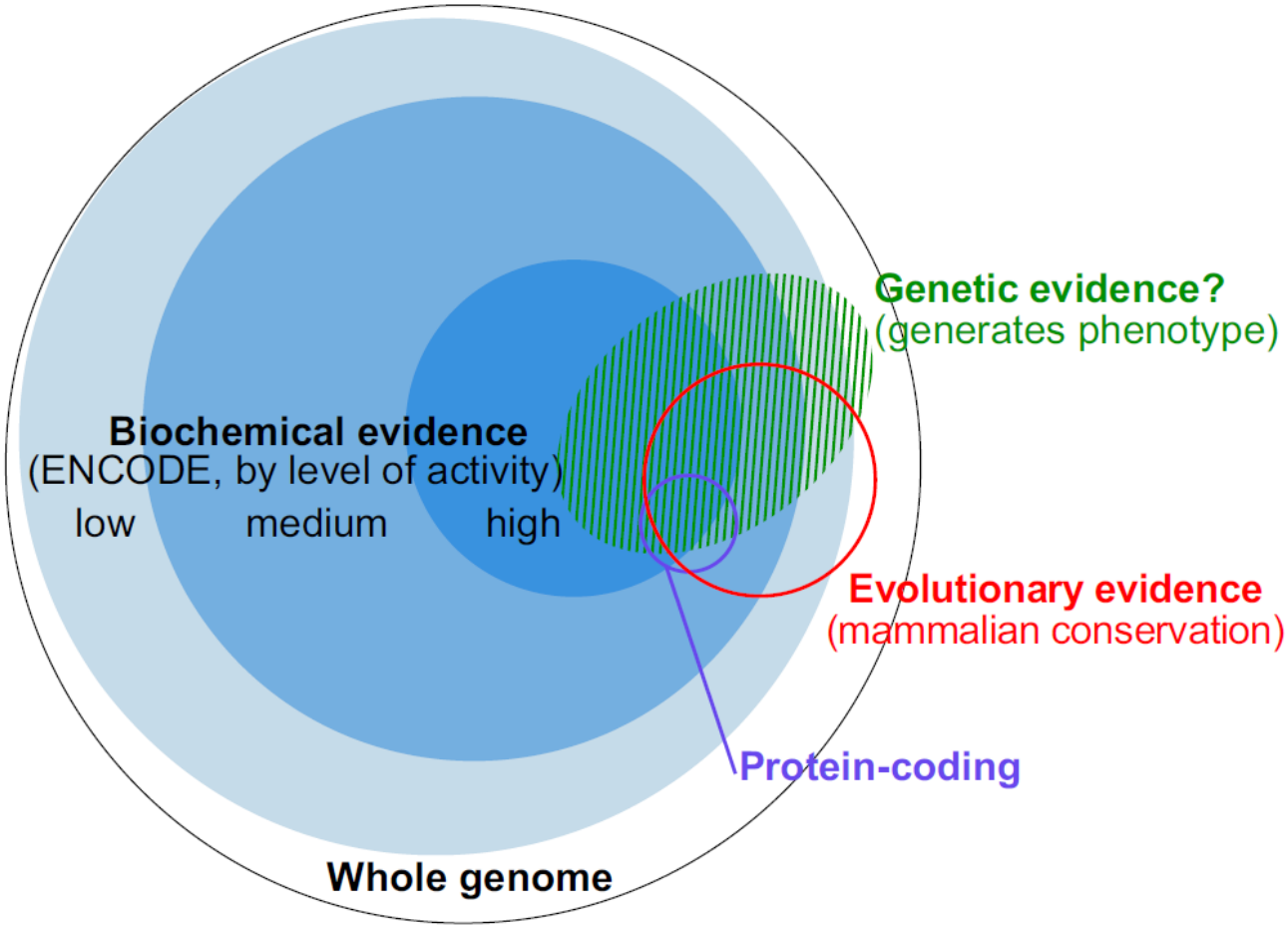
An integrated encyclopedia of DNA elements in the human genome

The ENCODE Project Consortium*

6 SEPTEMBER 2012 | VOL 489 | NATURE | 57

- **about 80% genome is transcribed**

What fraction of human genome is functional?



Kellis et al. 2014

Pseudogenes

- Nonfunctional gene copies (shortened, premature stop-codons, usually not transcribed)
- In the human genome ~ 14 000 pseudogenes (13% transcriptionaly active).

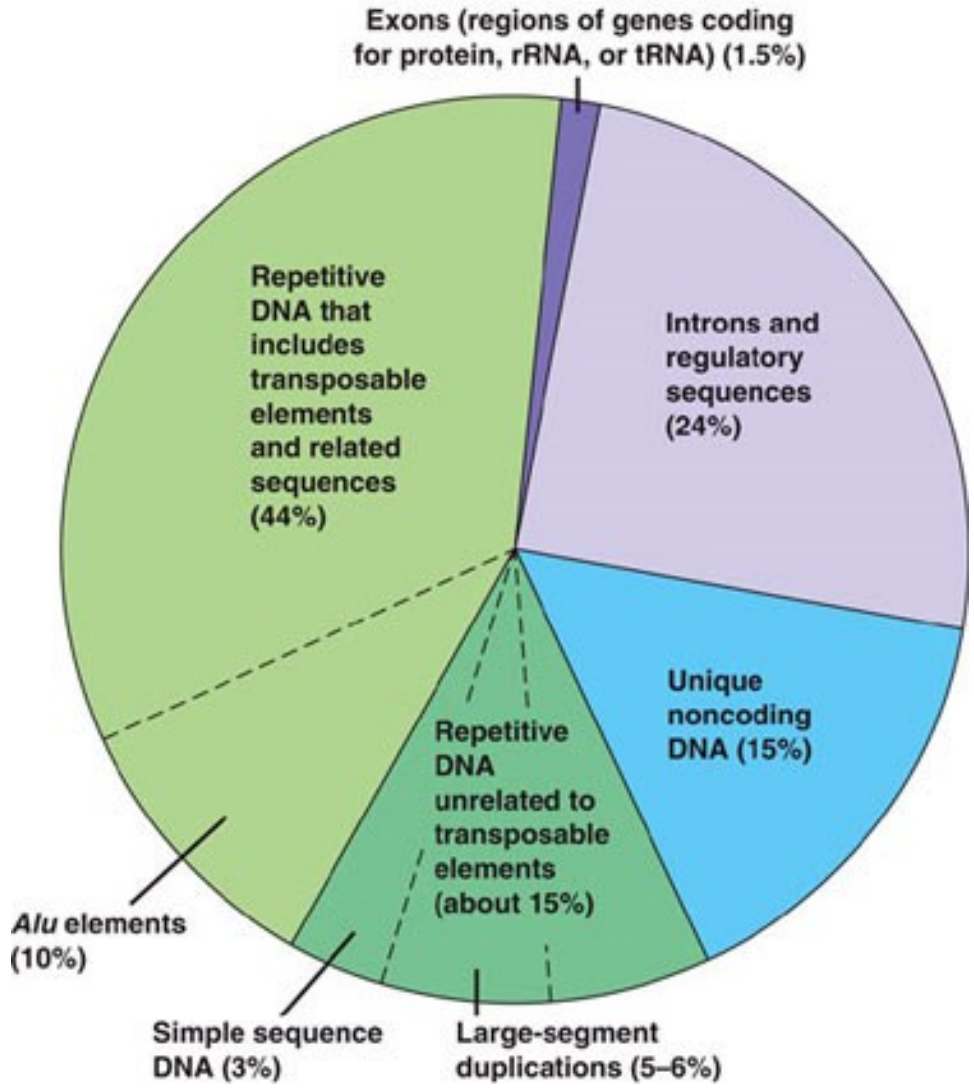
Nonprocessed pseudogenes

Originate by genome duplication. Contain exons and introns, often also surrounding sequences including promoters.

Processed pseudogenes

Originate by reverse transcription of mRNA. No introns, polyA.

Composition of human genome



„junk“ DNA

Just because of you don't understand, you can't call us "Junk"!
http://www.biocomicals.com

Repetitive sequences

Tandem repeats

microsatellites – repeat motif 1-6 bp long (e.g. CACACACA……)

minisatellites – repeat motif up to 25 bp, e.g. telomeric sequences

satellites – repeat motif several hundreds bp long or even longer,
e.g. centromeric sequences

Repetitive sequences

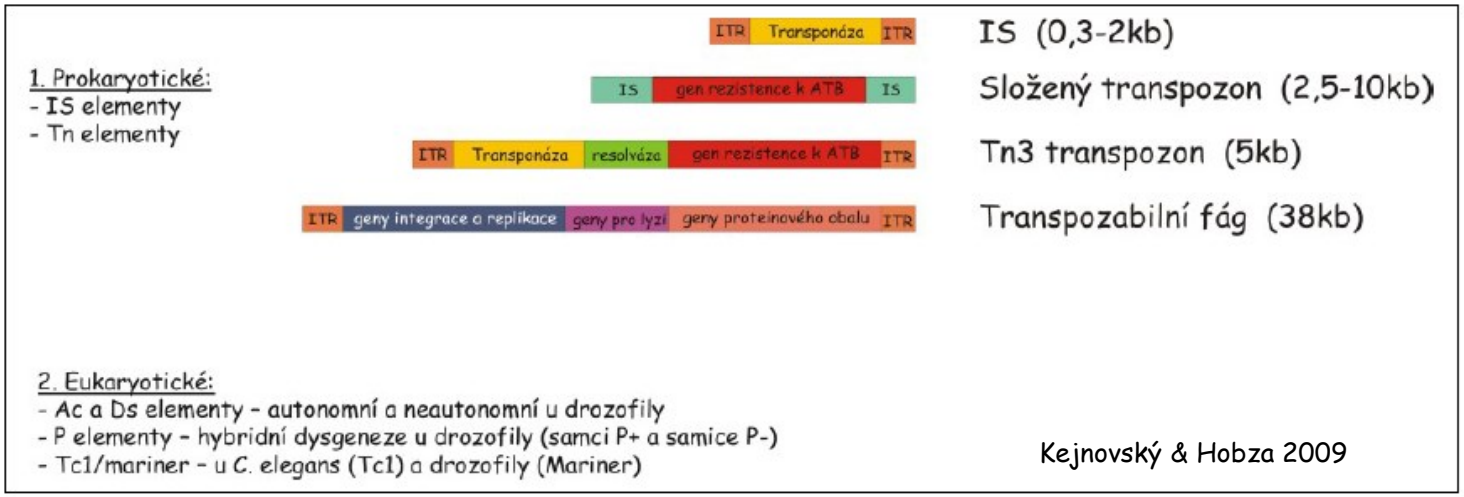
Mobile DNA elements (transposable elements)

DNA transposons – move by transposition through DNA molecule

retrotransposons – move by retrotransposition through RNA molecule

DNA transposons

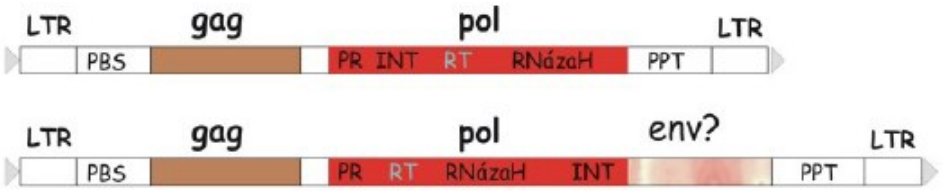
- Code an enzym **transposase**.
- Sometimes additional genes (e.g. genes for resistance to antibiotics in bacteria)
- IS elements and Tn elements in prokaryotes.
  P elements in Drosophila.
  Ac a Ds elements in corn (Barbara McClintock).

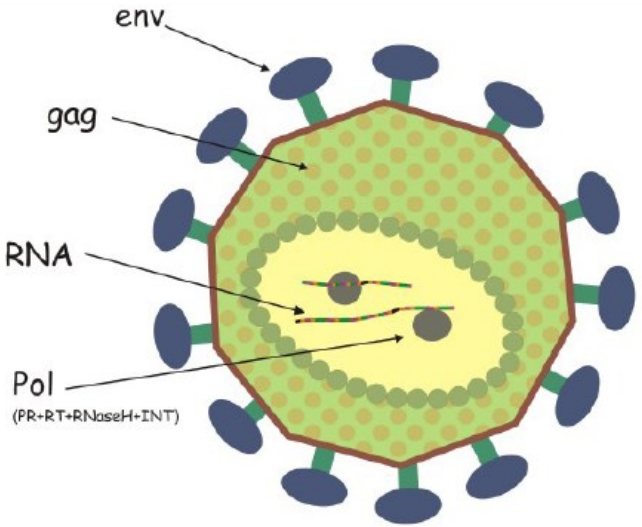


LTR retrotransposons

- Originated from retroviruses.
- Long terminal repeats (LTR) contain regulátory sequences, promoters.
- Code enzyme **reverse transkriptase.**
- In *Vicia faba* cca 1 milion copies of element Ty1-copia (half of the genome).

Ty1-copia, Ty3-copia





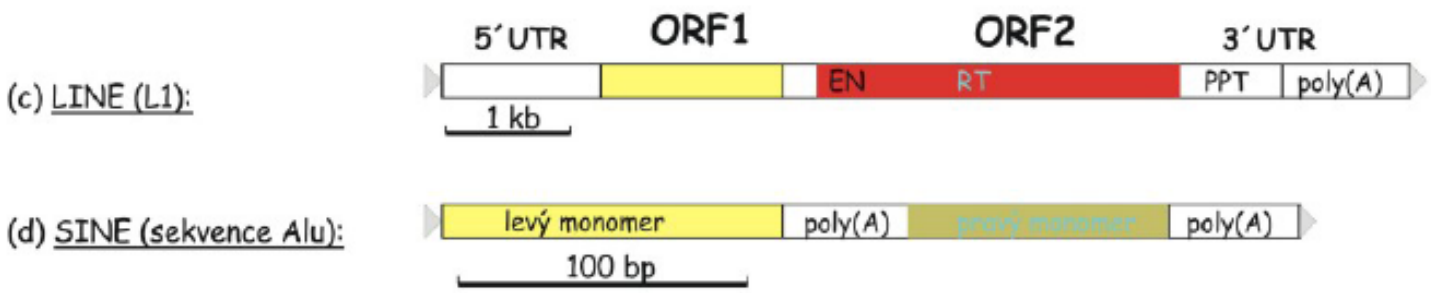
- Similar structure have **endogenous retroviruses**. Remnants of ancient once infectious exogenous retroviruses.

gag, pol, env ..... geny
LTR ... dlouhé koncové repetice
PBS .... místo vazby primeru
PR ..... proteáza
INT .... integráza
RT ..... reverzní transkriptáza
PPT .... polypurinový úsek
..... přímé repetice

Kejnovský & Hobza 2009

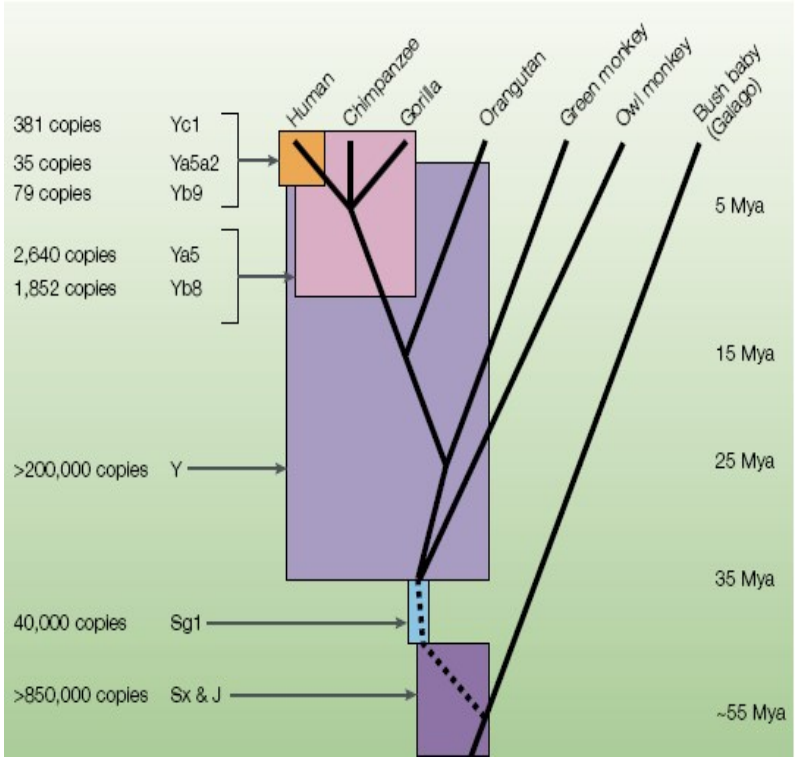
nonLTR retrotrasposons

- Do not have LTR. Poly(A) on 3' end.
- Most common transposons in mammals, including human.
- **LINE (Long interspersed nuclear elements)**
  Code its own reverse transkriptase.
- **SINE (Short interspersed nuclear elements)**
  Do not have its own transcriptase.
  Their movement dependent on the presence of LINE elements.
- **Alu seuence** (SINE): the most common transposon in humans (1,5 milionů copies, 13% of the genome). Originated from a gene coding tRNA 7SLRNA



Evolution of transposable elements

- Periods of TE activity and silence.
- Most TE in the genome are shortened and non-functional. But can be copied if there are at least some functional TEs providing enzymes.
- Relatively frequent horizontal gene transfer.



The expansion of Alu subfamilies (Nat Rev Genet)

Negative effects of TEs

Increase in the genome size

- 3-45% of the genome in animals.
  Up to 80% of the genome in plants (corn)
- Higher energetical demands for the cell.
- Enlargement of the cell.

Giant genomes of salamanders

- Genome size ~120 Gb.
  Large amount of LTR retrotransposons.
- Increased cell size.
- Simplification of nervous and a optical system.
- Some species have red blood cells without the nucleus.
- Some species have only four fingers on the legs.

Increase in the mutation rate

- **Insertion mutagenesis.**
- **Alteration of gene expression of nearby genes.**
- **Ectopic recombination**

- To prevent movement of TEs in the genome, organisms heterochromatinize the genomic regions containg TEs.
- Important role of piwi RNAs.

Positive effects of TEs

Regulation of mutation rate in response to stress

- Stress conditions (UV or gamma radiation, infection by patogen, injury, change of temperature…) can lead to TE activation.
- Increased mutation rate can lead to new adaptations.

Tam1 , Tam3 u hledíku
(1000x větší aktivita při 15°C)
Reme1 u melounu (aktivace UV)

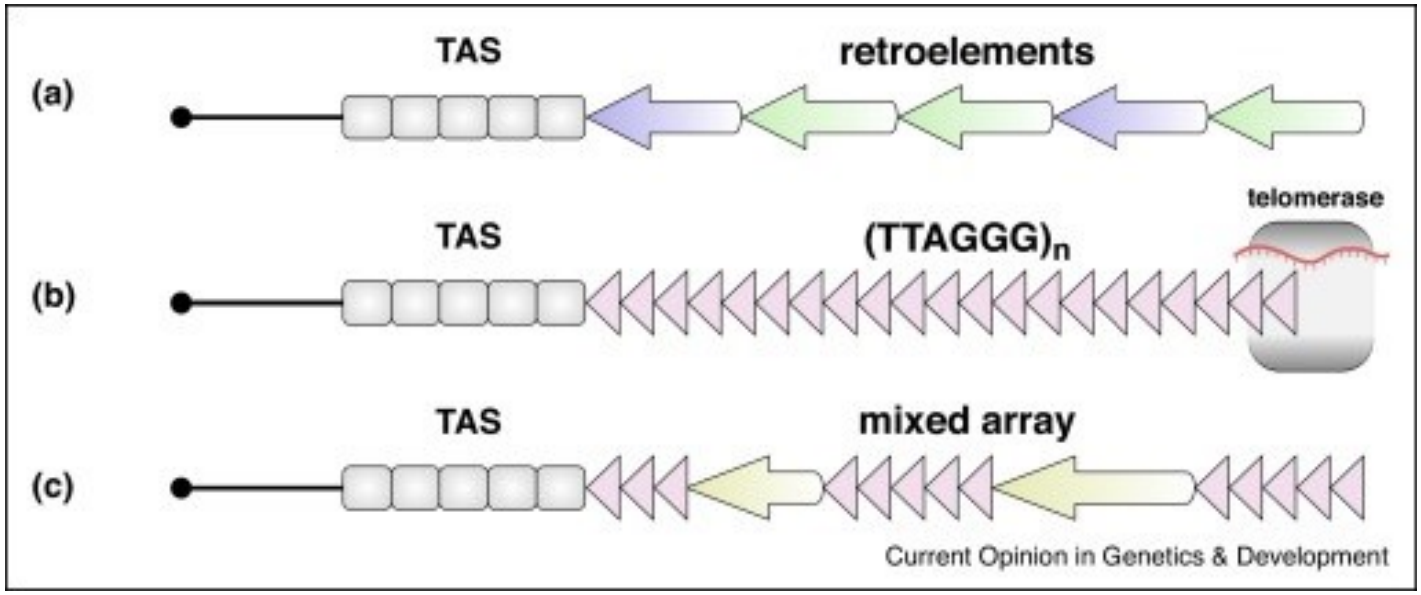
TEs make genome more dynamic

- Gene duplications.
- Structural changes (translocations, inversions)
- „Exon shuffling“ (Helitrons).

Helitrons = DNA transposons, transpose by a rolling circle replication mechanism. Can move parts of genes. Creation of new chimeric genes.

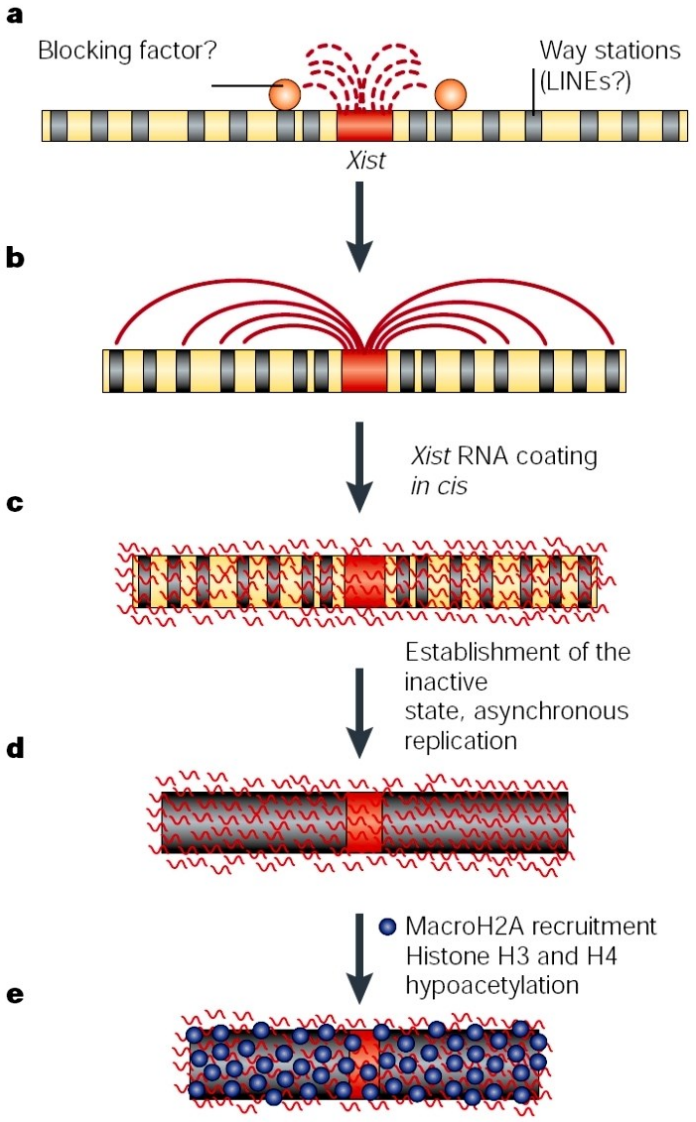
Molecular domestication of TEs

- Retrotrasposons make up **telomeres** in Drosophila melanogaster (non-LTR retrotransposons Het-A, TART, TAHRE).
  Telomere elongation is achieved by TE retrotransposition.

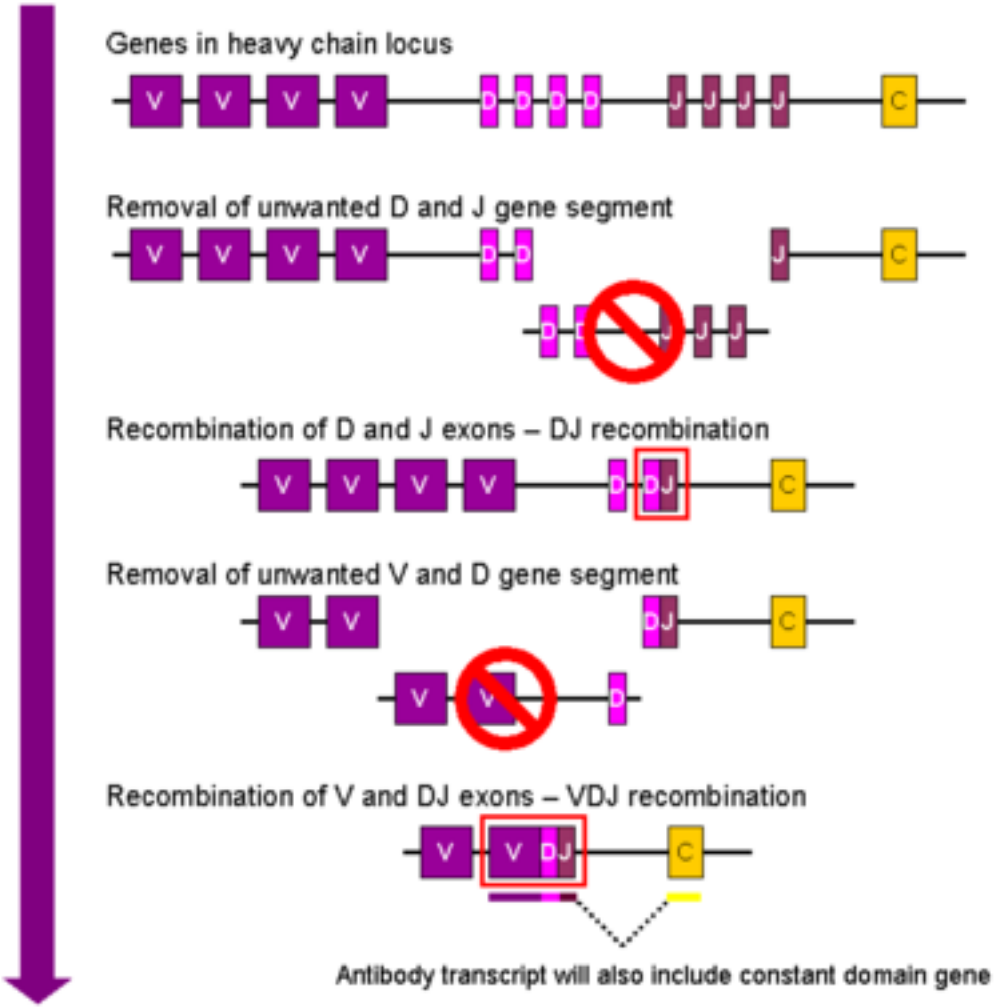


Formation of heterochromatin

- TEs often silenced by heterochromatinization.
- They sometimes make an important part of centromeric heterochromatin.
- LINE elements play a role in inactivation of the X chromosome in female somatic cells in mammals.



V(D)J recombination in imunoglobulins



Important role of RAG proteins. Derived from transposase.

Industrial melanism in the Peppered moth

The dark form (carbonaria) originated in the 1. half of the 19. century (industrial revolution) by insertion of TE to the vicinity of cortex gene, which incresed its expression.

van't Hof, A. et al. 2016.

Syncytin – derived from env from endogenous retrovirus HERV –W.
Important role in placenta development in mammals and viviparous lizards.

PNAS

Check for updates

PNAS PLUS

SEE COMMENTARY

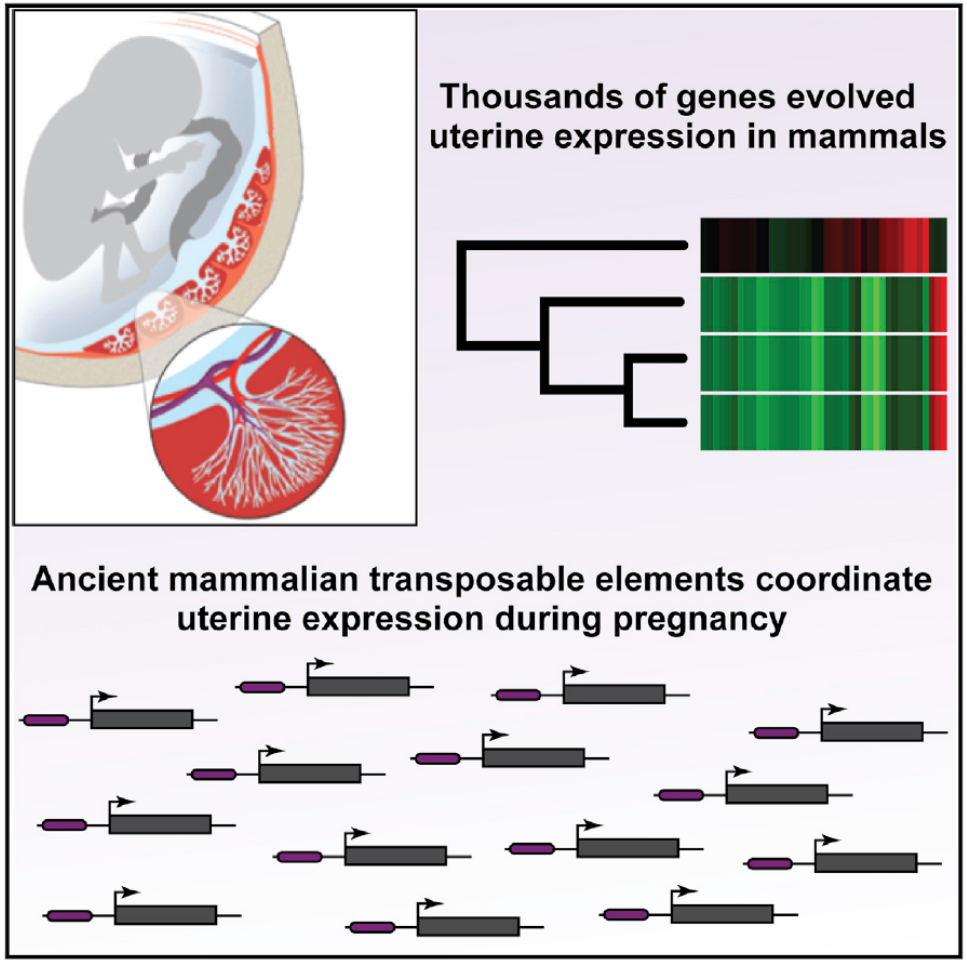
An endogenous retroviral envelope syncytin and its cognate receptor identified in the viviparous placental *Mabuya* lizard

Guillaume Cornelis[a,b,1,2], Mathis Funk[a,b,1], Cécile Vernochet[a,b], Francisca Leal[c,3], Oscar Alejandro Tarazona[c,4], Guillaume Meurice[d], Odile Heidmann[a,b], Anne Dupressoir[a,b], Aurélien Miralles[e], Martha Patricia Ramirez-Pinilla[c], and Thierry Heidmann[a,b,5]

[a]Unité Physiologie et Pathologie Moléculaires des Rétrovirus Endogènes et Infectieux, CNRS UMR 9196, Gustave Roussy, Villejuif, F-94805, France; [b]UMR 9196, Université Paris-Sud, Orsay, F-91405, France; [c]Laboratorio de Biologia Reproductiva de Vertebrados, Escuela de Biologia, Universidad Industrial de Santander, 680002 Bucaramanga, Colombia; [d]Plateforme de Bioinformatique, INSERM US23/CNRS UMS3655, Gustave Roussy, Villejuif, F-94805, France; and [e]Institut de Systématique, Evolution, Biodiversité, Muséum National d'Histoire Naturelle, CNRS UPMC EPHE, Sorbonne Universités, Paris, F-75005, France

Origin of large evolutionary innovations

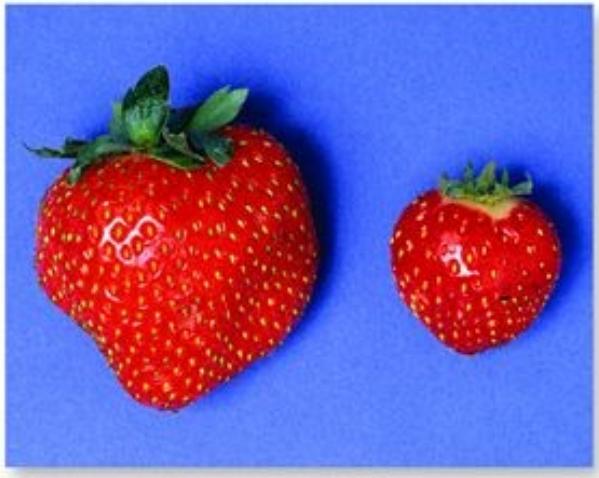
- Simultaneous change of gene expression of many genes.



Cell Reports, 2015

Polyploidization

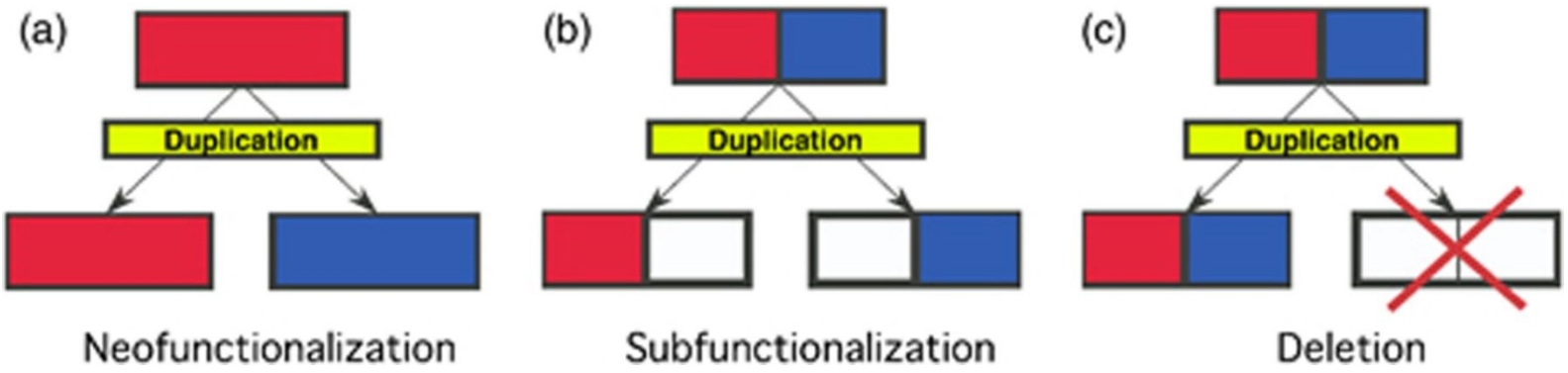
- Common especially in plants (60–70% angiosperms are polyploids).
- Problems with meiosis. Often transition to asexual reproduction.
- Functional redundancy, can enhance adaptive evolution.



(a) Octaploid strawberries (left) and diploid strawberries (right)

(b) Diploid daylilies (left) and tetraploid relative (right)

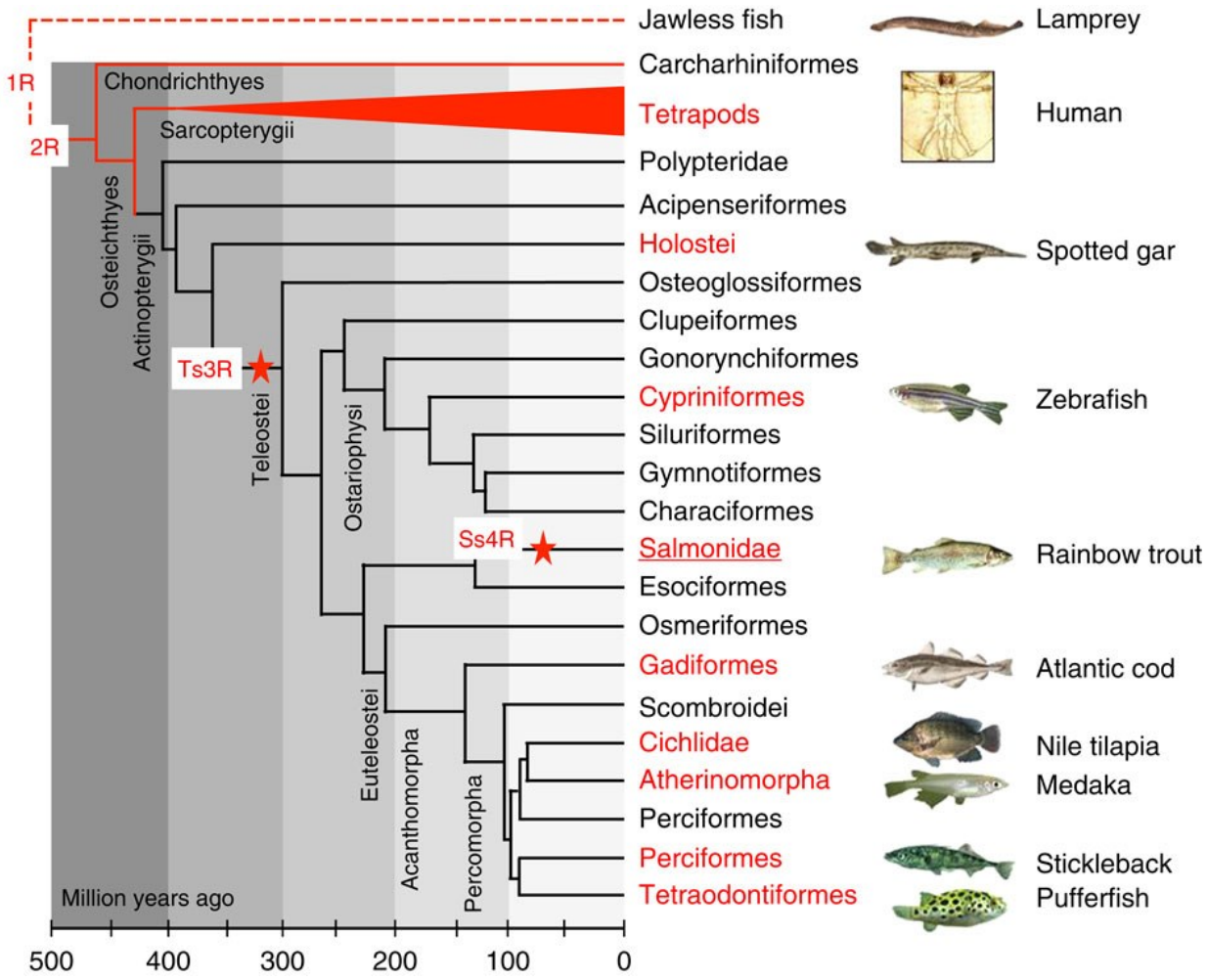
Duplications



- Faster adaptive evolution.
- Evolution of reproductive isolation (loss of different copies in different populations can result in sterility or inviability of hybrids). Can speed up speciation.

Paleopolyploidy in vertebrates

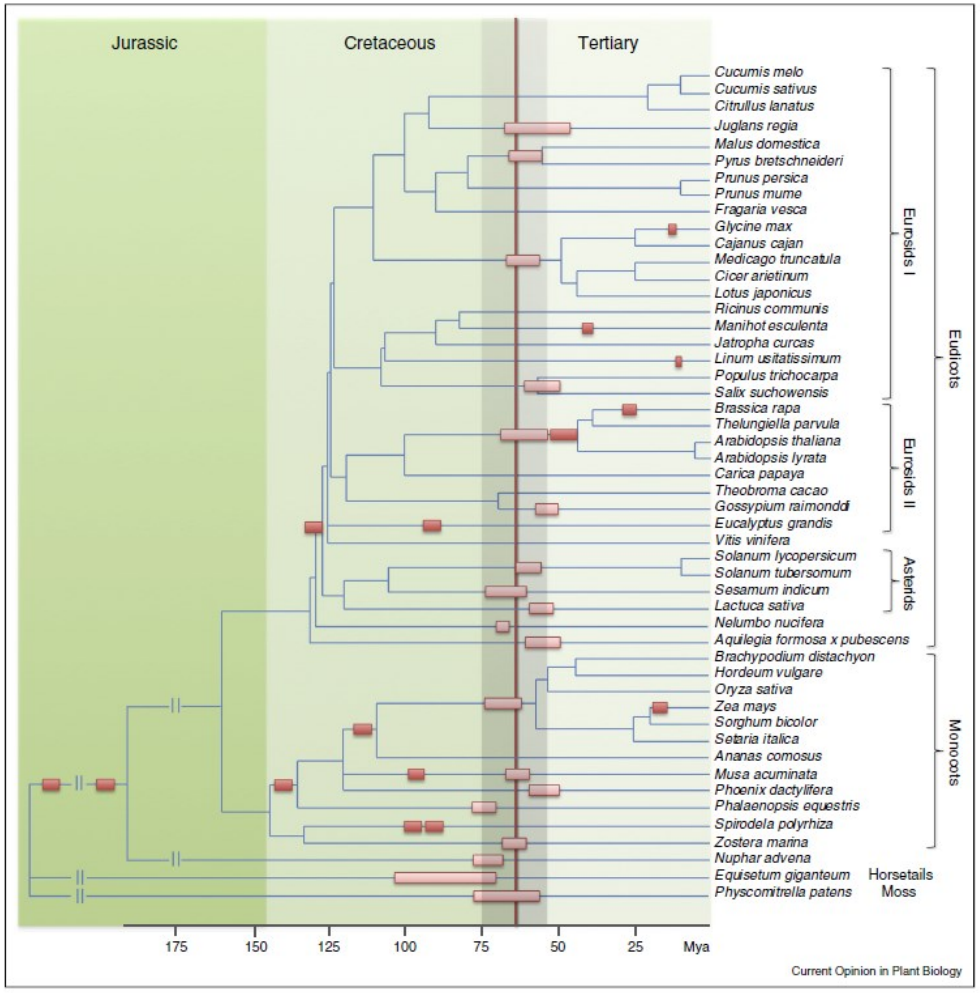
- 2 in the ancestor of all vertebrates.
- 1 in the ancestor of Teleostei.
- 1 in Salmonidae fish.



Berthelot et al. 2014

Paleopolyploidy in angiosperms

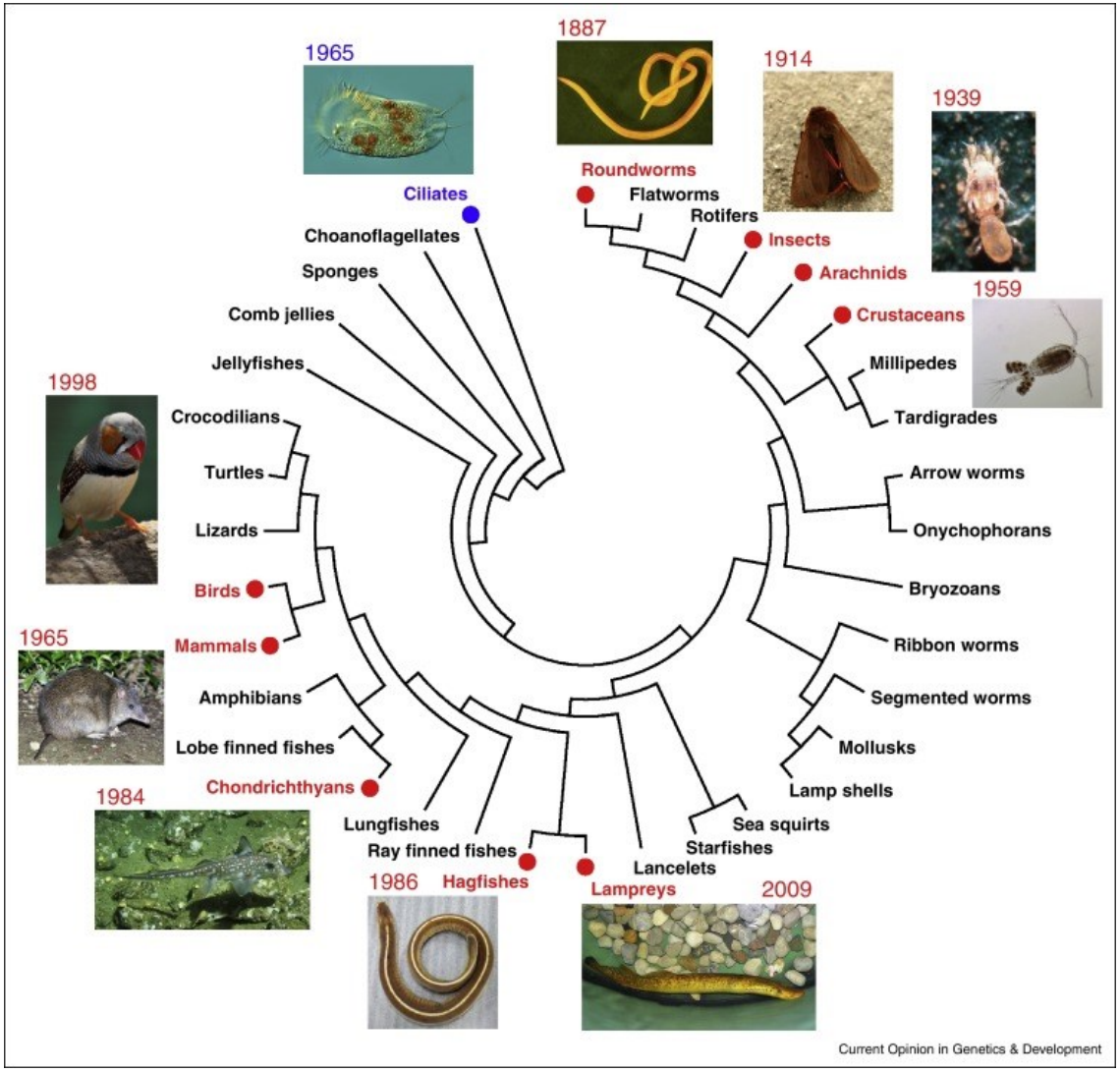
- Relatively more whole genome duplications close to K-T boundary. Polyploid plants might better survived in markedly changed environments compared to diploids.

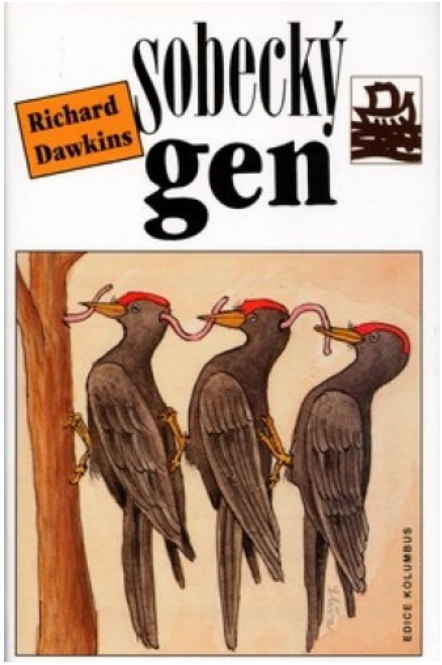
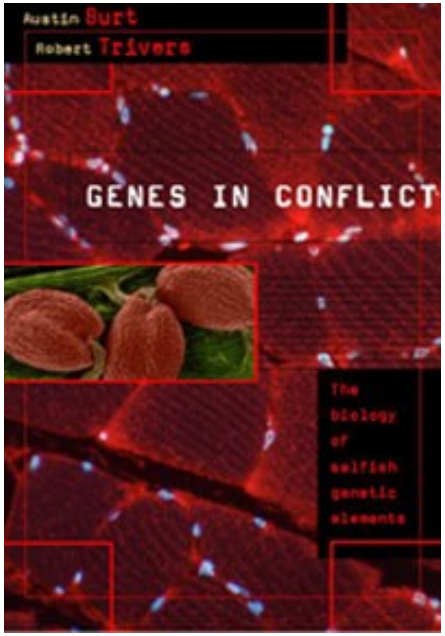
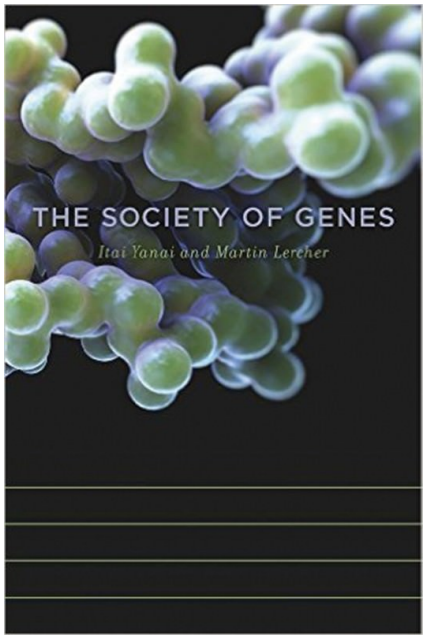


haus and Van de Peer 2016

Programmed DNA elimination

Elimination of parts of the genome from some or all somatic cells.



Genome as ecosystems

- Genes and other functional elements as species.
- Aleles of genes as individuals of particular species (compete over reproduction, „selfish genes“)
- Interactions among genes (parasitims, mutualism). Intergenomic conflicts.