

CG020 Genomics

Lesson 8

Structure and organization of genomes

Markéta Pernisová

Functional Genomics and Proteomics of Plants,
Mendel Centre for Plant Genomics and Proteomics,
CEITEC - Central European Institute of Technology, Masaryk University, Brno
marketa.pernisova@ceitec.muni.cz, www.ceitec.muni.cz



INVESTICE DO ROZVOJE VZDĚLÁVÁNÍ

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

Outline

1. Eukaryotic nuclear genomes
2. Genomes of prokaryotes and eukaryotic organelles
3. Virus genomes and mobile elements
4. Online sources
5. Literature

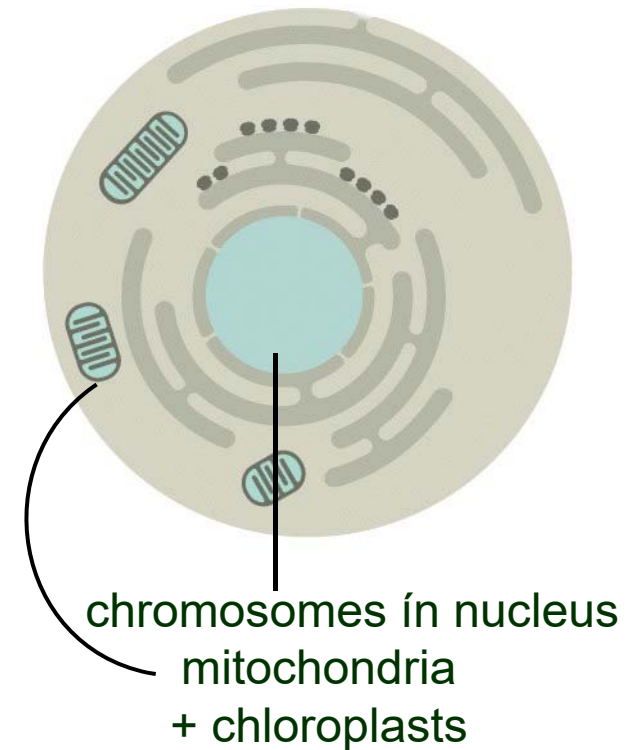
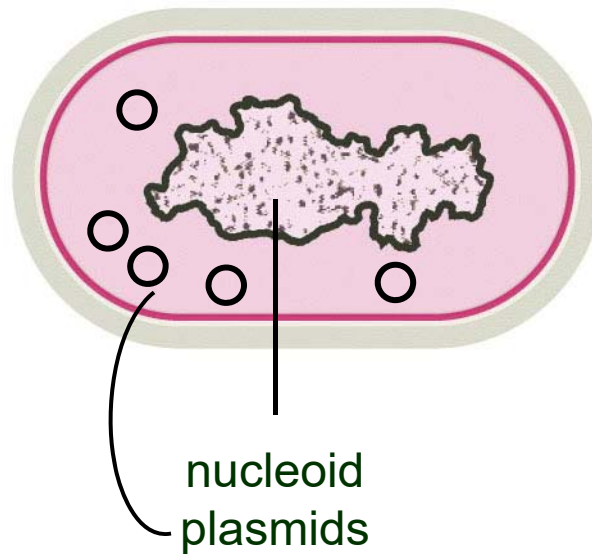
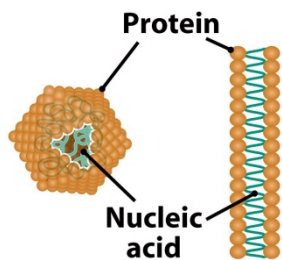


GENOME

Genome – set of genetic information of an organism

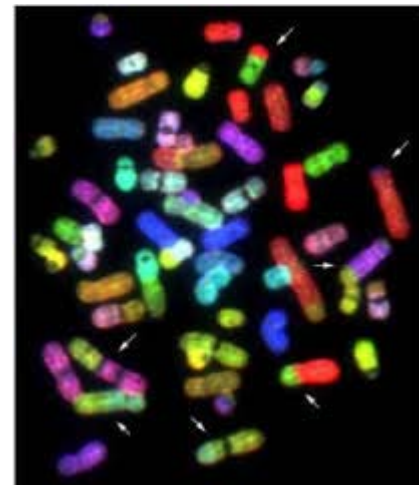
- complete biological information to construct, maintain and replicate/reproduce

- eukaryotic
- prokaryotic
- virus



EUKARYOTIC NUCLEAR GENOMES

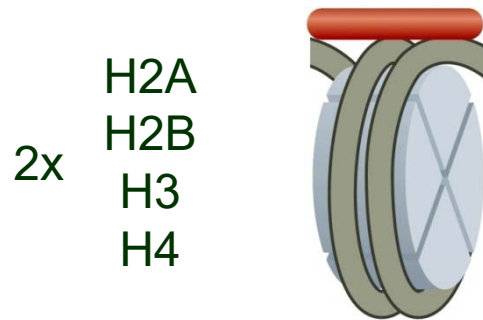
= set of linear DNA molecules, at least two chromosomes, without exceptions



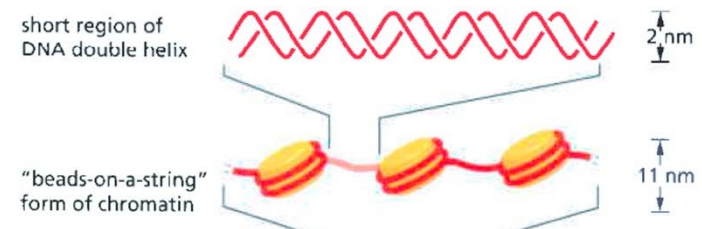
STRUCTURE OF CHROMOSOMES

DNA + histones = nucleosome

- 140-150 bp
- linker: 50-70 bp
- + linker histones



„beads-on-a-string“ form of
chromatin
– 11 nm



STRUCTURE OF CHROMOSOMES

30 nm chromatin fiber

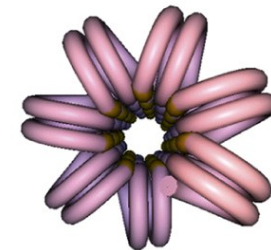
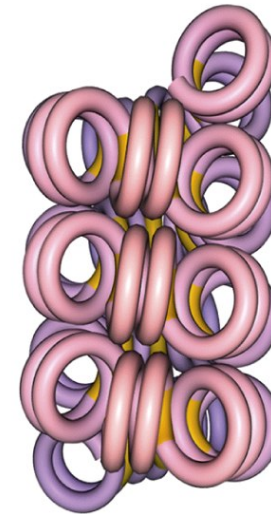
- interphase chromosomes

- several theories, 2 models:

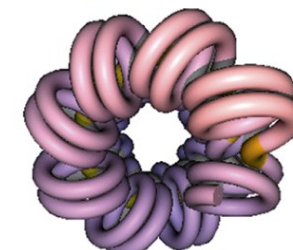
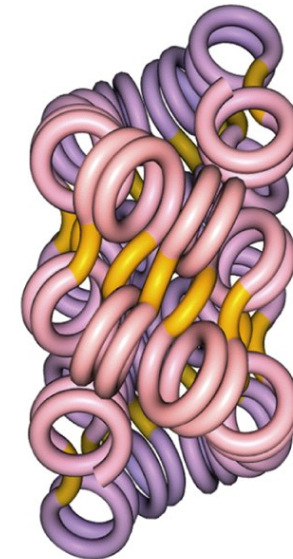
- solenoid model – only linker histones (e.g. H1)
- helical ribbon – linker histones + core histone „tails“ – chemical modifications of these tails open up the 30 nm fiber



(A) The solenoid model



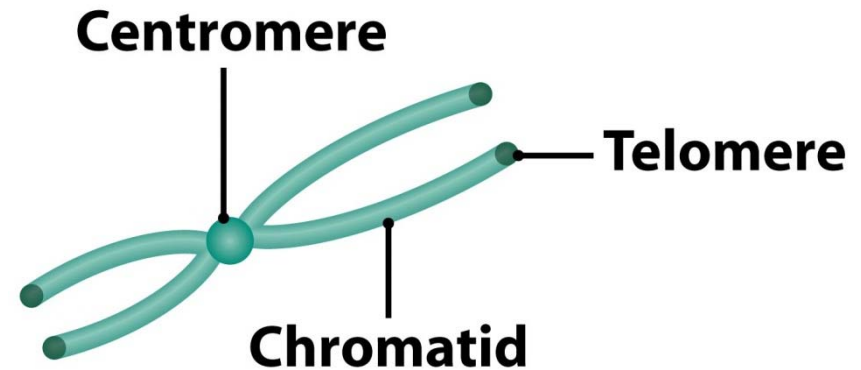
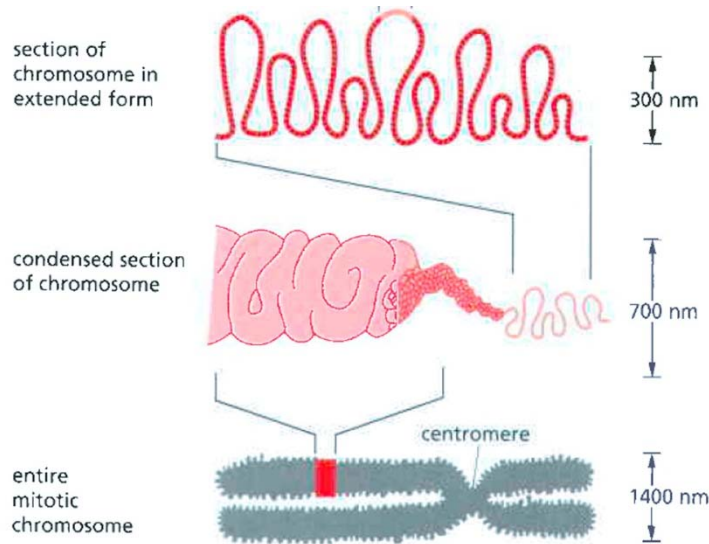
(B) The helical ribbon model



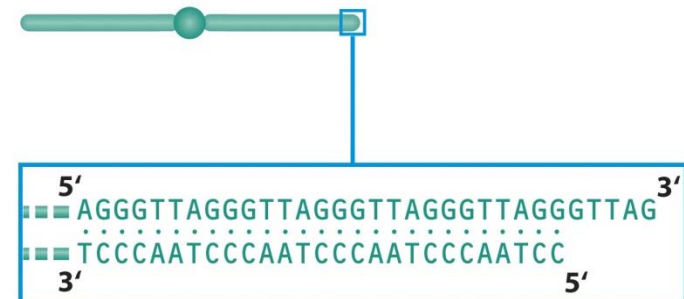
STRUCTURE OF CHROMOSOMES

Condensed metaphase chromosomes: 1400 nm

one chromatid: 700 nm



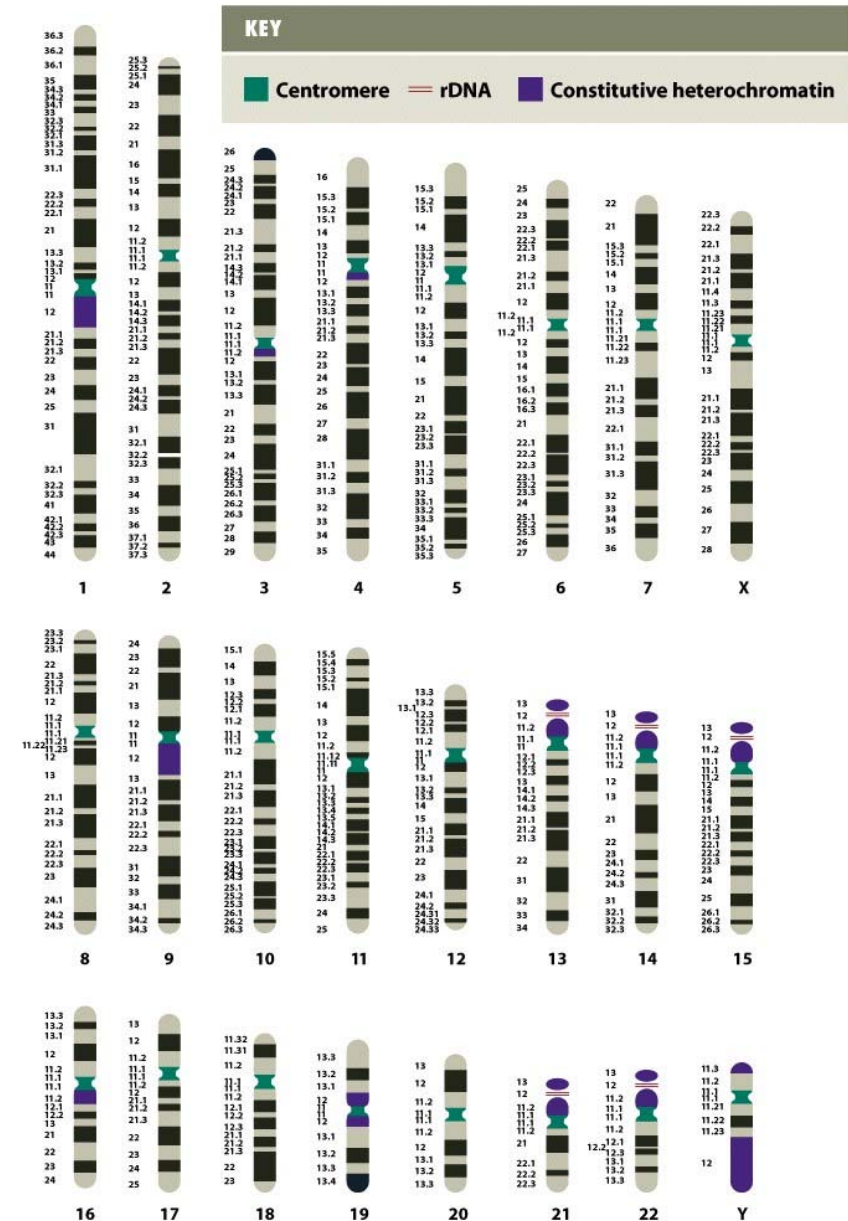
Centromeres, telomeres –
repetitive sequences



METAPHASE CHROMOSOMES

Human karyogram

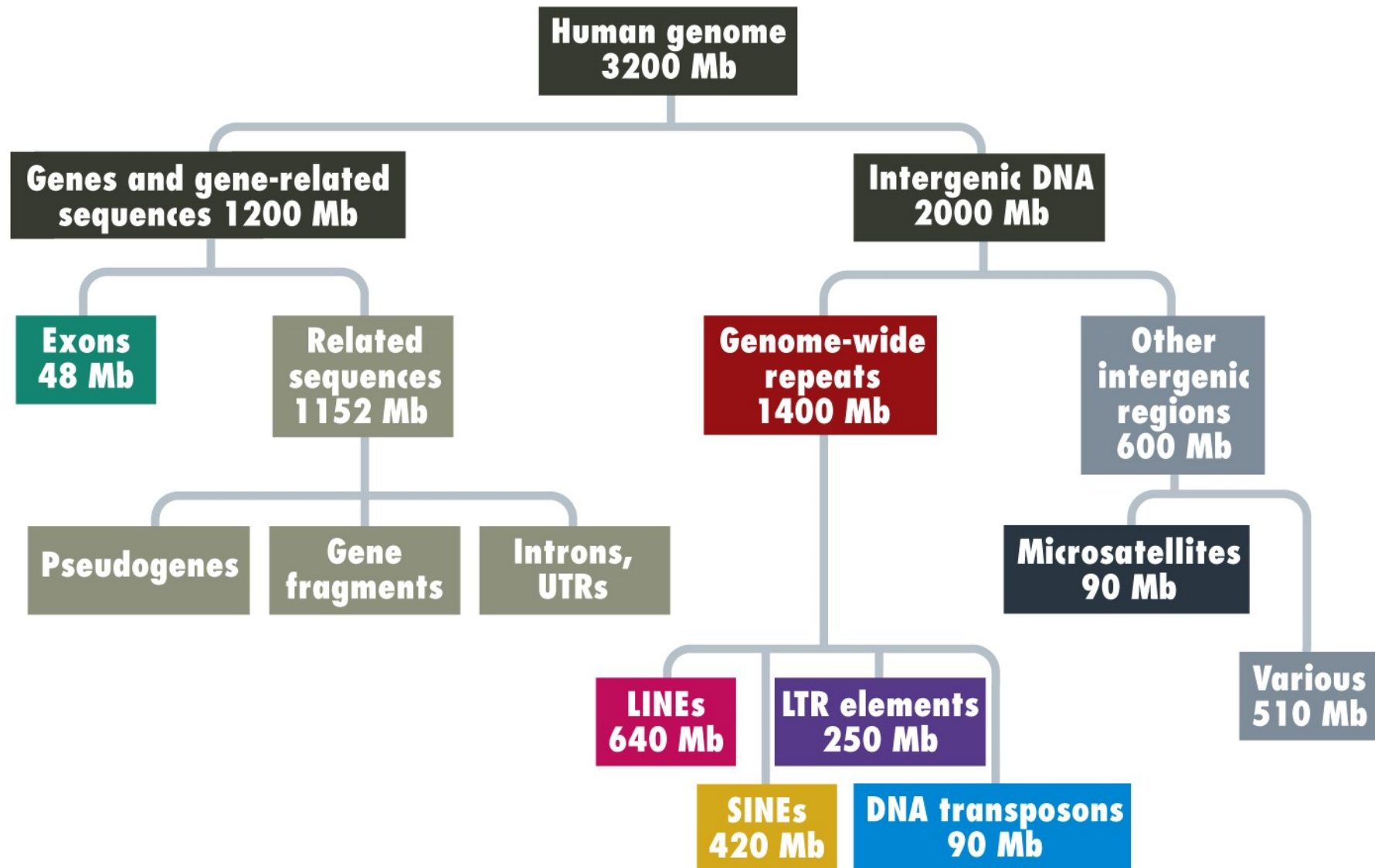
Technique	Banding pattern
G-banding	Dark bands are AT-rich Pale bands are GC-rich
R-banding	Dark bands are GC-rich Pale bands are AT-rich
Q-banding	Dark bands are AT-rich Pale bands are GC-rich
C-banding	Dark bands contain constitutive heterochromatin



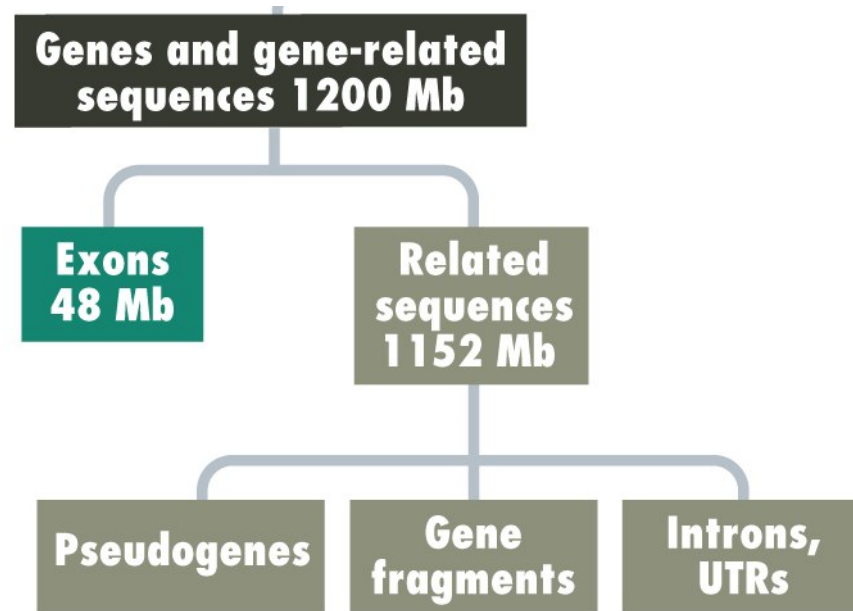
UNUSUAL CHROMOSOMES

- Minichromosomes
 - short, high gene density
 - e.g. chicken: 33 minichromosomes – 1/3 of genome ~ 75% of genes
- B chromosomes
 - individual, not in all populations
 - fragments of normal chromosomes, probably the result of unusual events during nuclear division
 - common in plants, associated with reduced viability
 - also in fungi, insects, animals
- Holocentric chromosomes
 - no single centromere, but multiple centromeric structures along chromosome
 - e.g. *Caenorhabditis elegans*

GENOME ORGANIZATION



GENES AND GENE-RELATED SEQUENCES



GENES

Genes

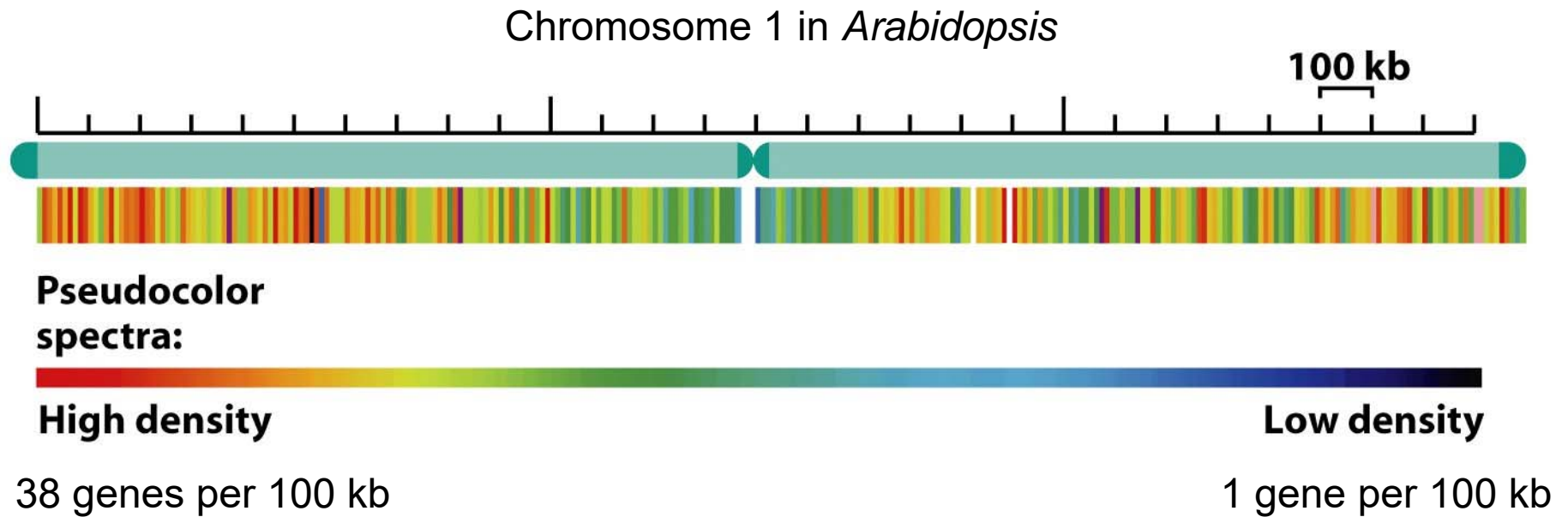
- UTR – untranslated region
- Introns – splicing
- Exons – functional products

Multigene families – groups of genes of identical or similar sequence

- simple (or classical)
 - gene duplication
 - e.g. genes coding ribosomal RNA in human genome:
 - 2000 genes for 5S rRNA – all in a single cluster on chromosome 1
 - 280 copies of a repeat unit for 28S, 5.8S and 18S rRNA – five clusters of 50-70 repeats on five chromosomes
- complex
 - similar sequences (but not the same), distinctive properties
 - e.g. mammalian globin genes – expressed at different developmental stages

GENE DISTRIBUTION ON CHROMOSOME

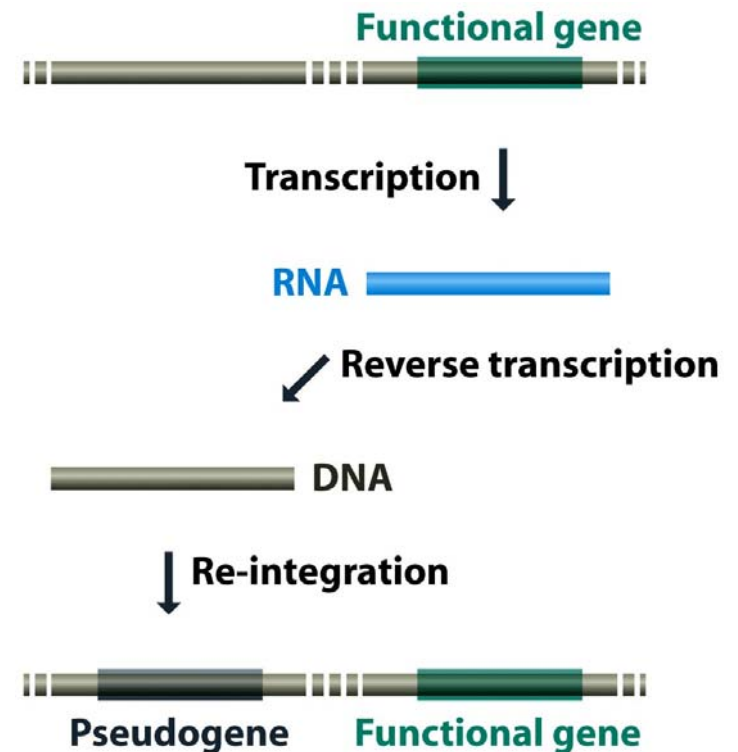
- uneven
- genes also in centromere, lower density



Human: 1-64 genes per 100kb

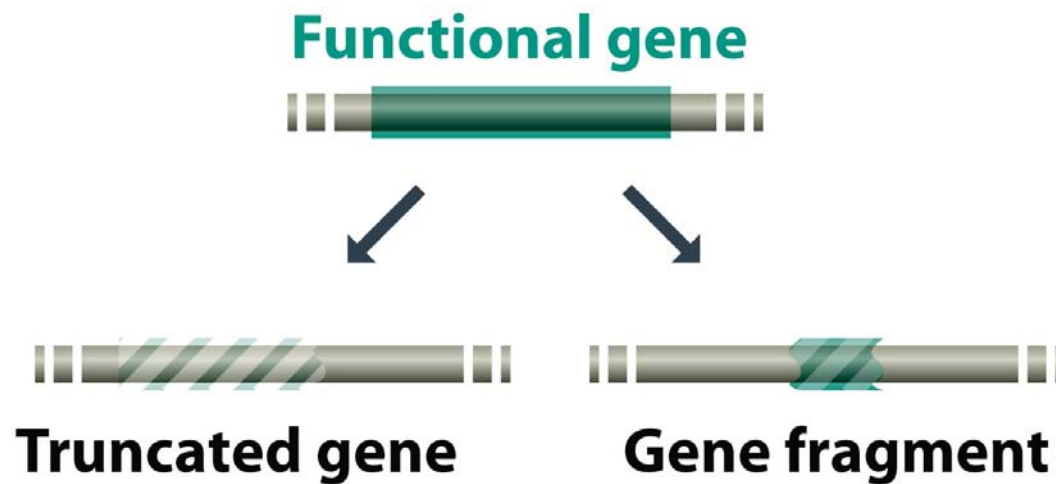
PSEUDOGENES

- evolutionary relics
- 2 groups
 - conventional – arise due to mutation
 - partially functional or nonfunctional
 - processed – derived from mRNA copy by RT
 - no introns
 - lacks regulatory sequence upstream of gene
 - nonfunctional



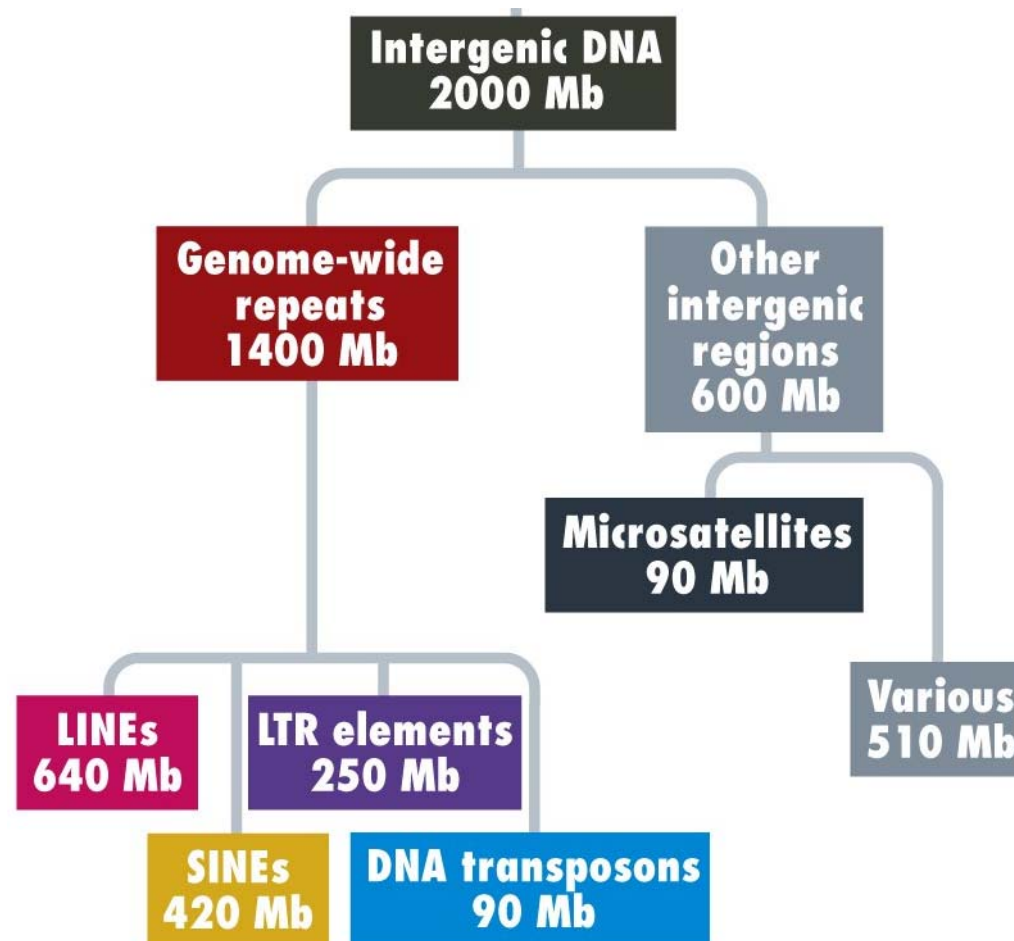
GENE FRAGMENTS

- Truncated genes
- Gene fragments



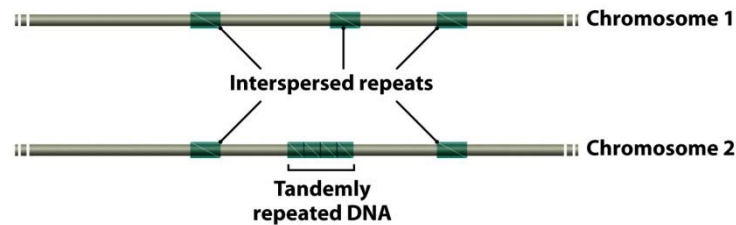
INTERGENIC DNA

- „junk“ DNA – not true



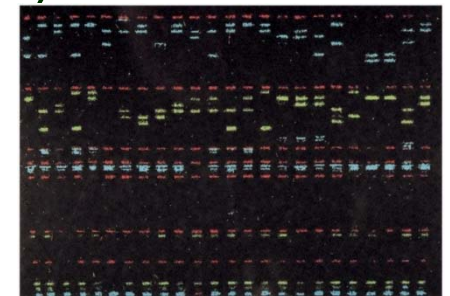
REPETITIVE DNA

- Tandemly repeated DNA
 - repeat units are placed next to each other in an array
- Interspersed repeats
 - genome-wide
 - repeat units are distributed randomly around the genome



TANDEMLY REPEATED DNA

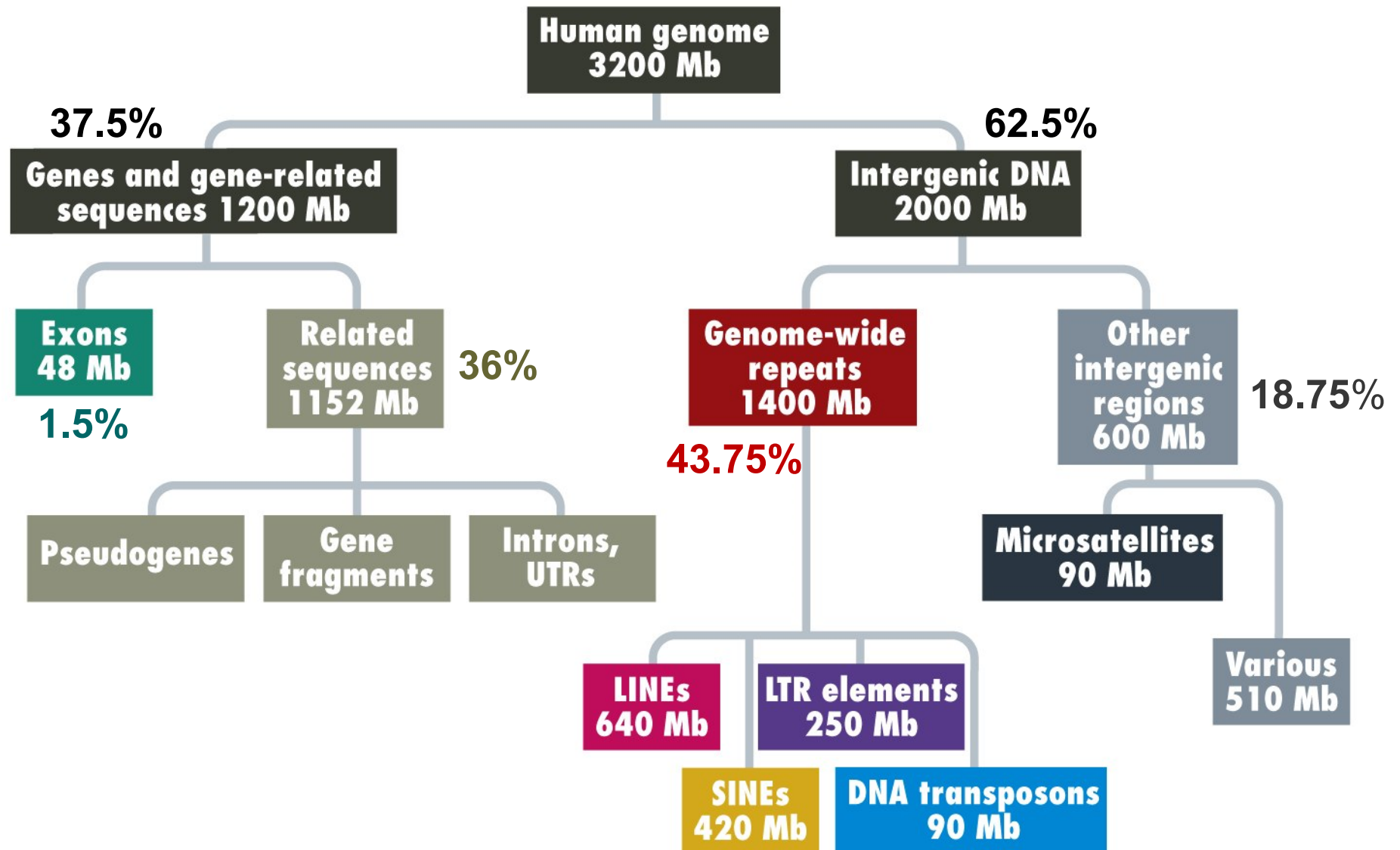
- ~ satellite DNA
- arised due to errors during genome replication
- long series of tandem repeats, hundreds of kilobases in length
- single genome: several different types of satellite DNA, each with a different repeat unit (5-200 bp)
- in centromeres or LTR
- Minisatellites („variable number of tandem repeats“ - VNTRs)
 - repeat unit up to 25 bp, clusters up to 20 kb
 - telomeres
- Microsatellites („simple tandem repeats“ - STRs)
 - repeat unit up to 13 bp, clusters up to 150 bp
 - function not clear
 - use: genetic profiling



INTERSPERSED REPEATS

- random
- arise due to transposition
- some of them descended from transposable viruses
- LINEs (long interspersed nuclear elements)
 - over 300 bp
- SINEs (short interspersed nuclear elements)
 - up to 300 bp

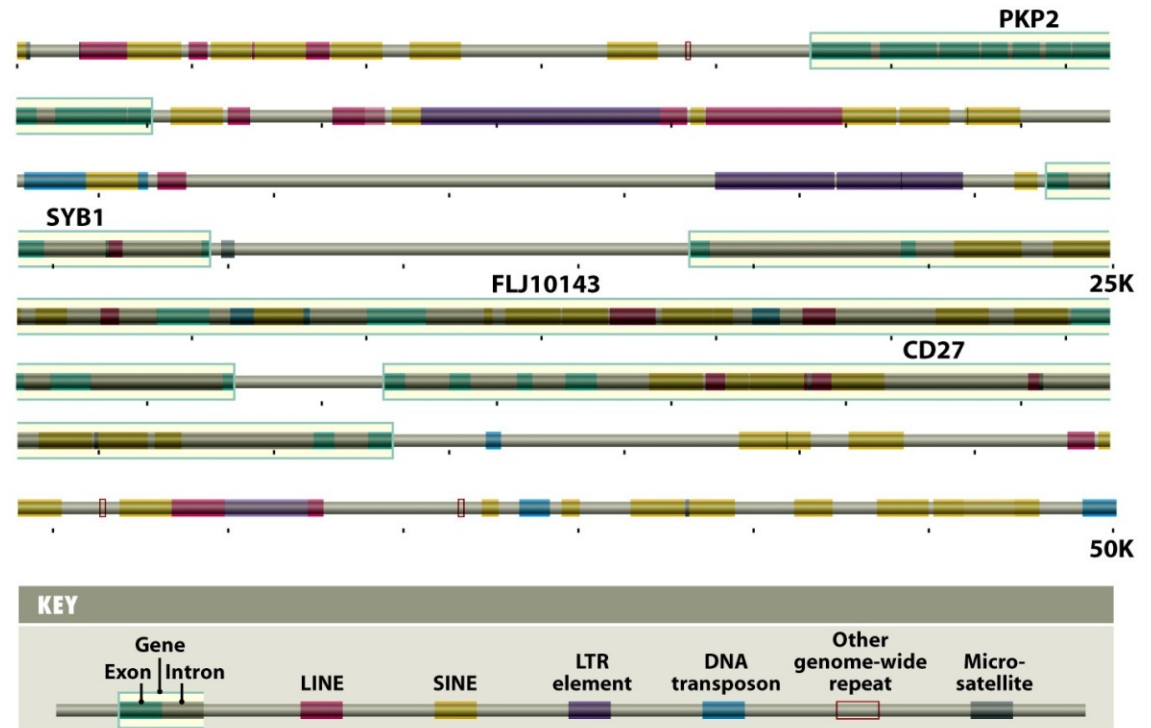
HUMAN GENOME



NUCLEAR GENOME ORGANIZATION

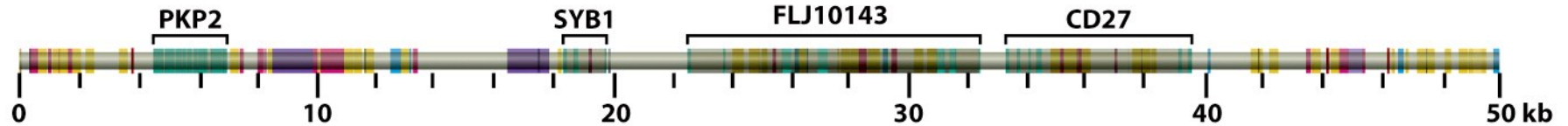
Human genome – 50 kb section

- 4 genes
- 88 repeats
 - LINEs
 - SINEs
 - LTRs
 - DNA transposones
- 7 microsatellites (4 in introns)
- 30% noncoding, nonrepetitive, single-copy DNA of no known function

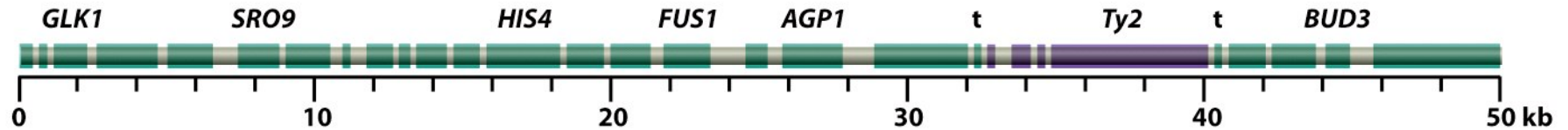


GENOME ORGANIZATION COMPARED

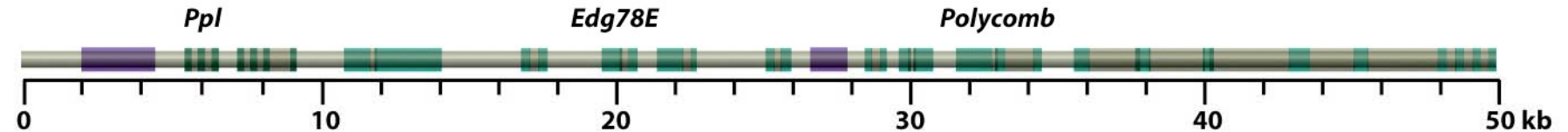
(A) Human



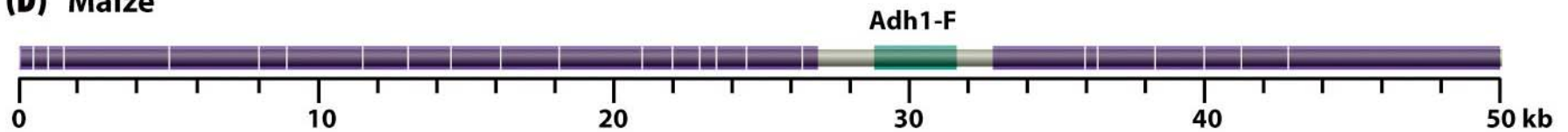
(B) *Saccharomyces cerevisiae*



(C) *Drosophila melanogaster*



(D) Maize

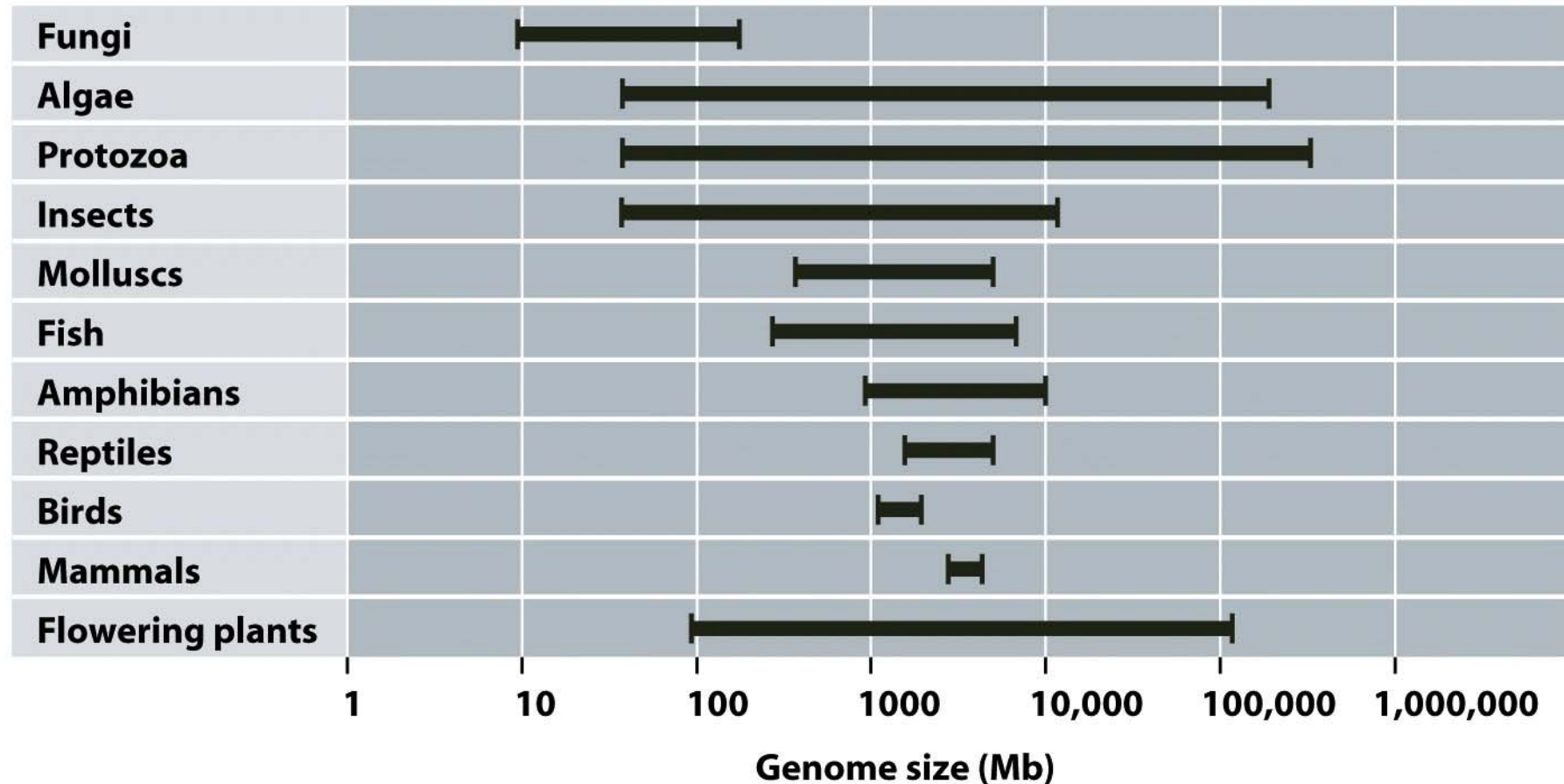


KEY



GENOME ORGANIZATION

- C-value paradox (C-value enigma) – size of the genome does NOT correlate with organism complexity



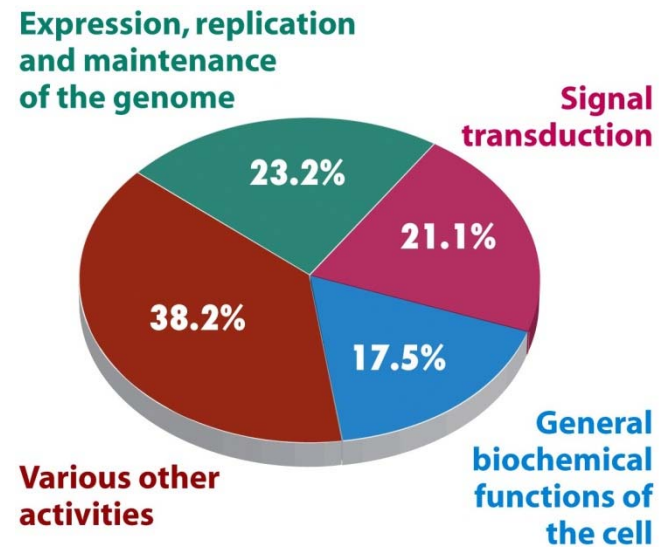
GENOME ORGANIZATION

- size of the genome does not correlate with the number of genes

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

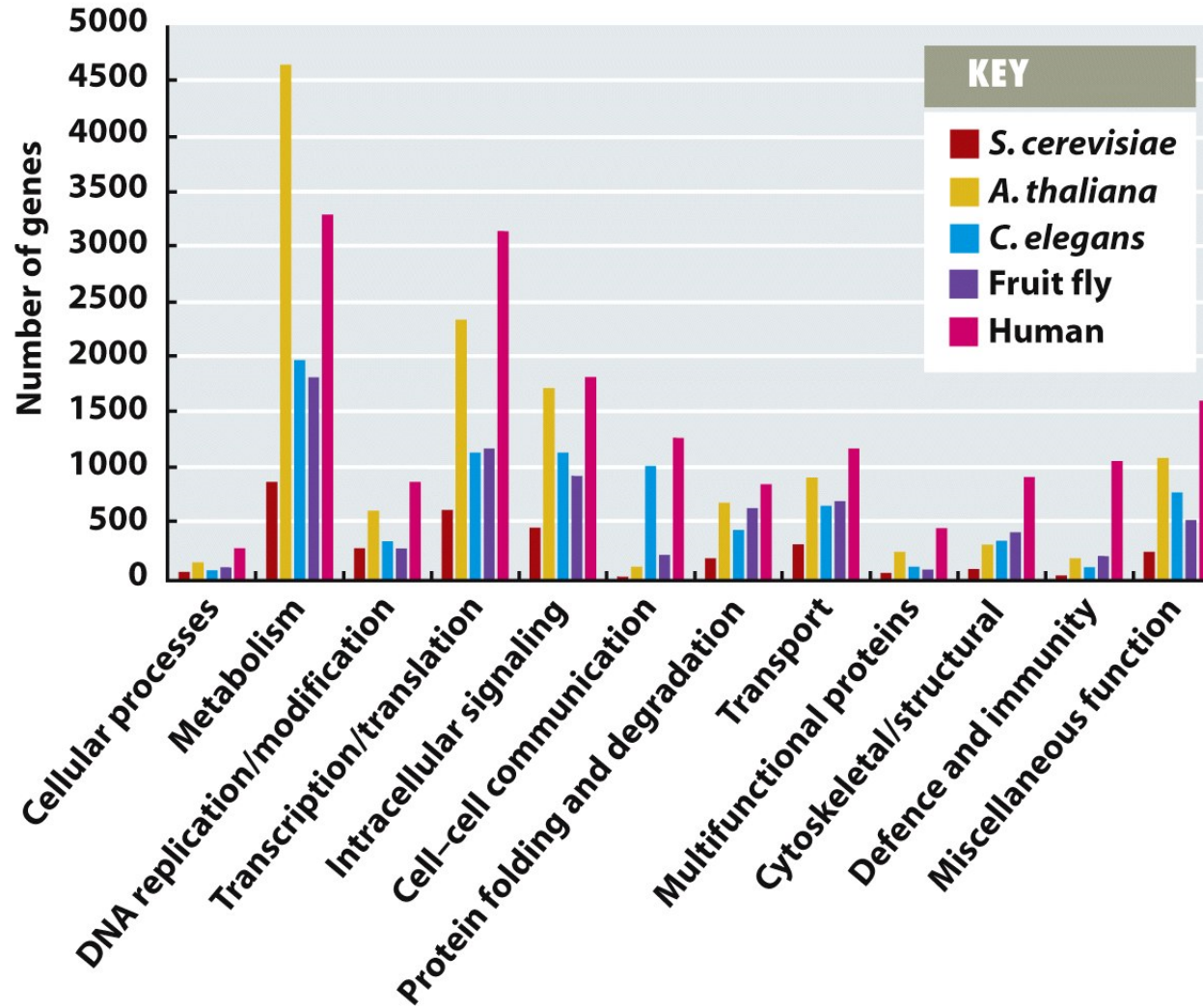
GENE CATALOG

- Organisms with sequenced genome
- Human gene catalog:



- cannot tell us what makes a human being ...

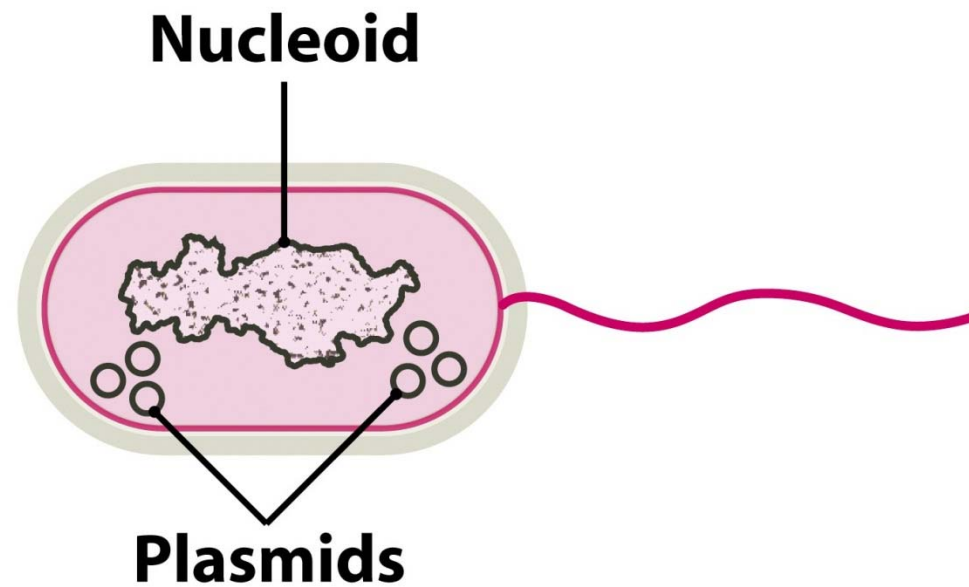
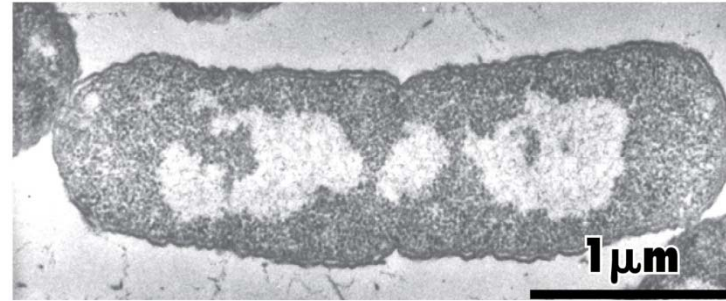
GENE CATALOG



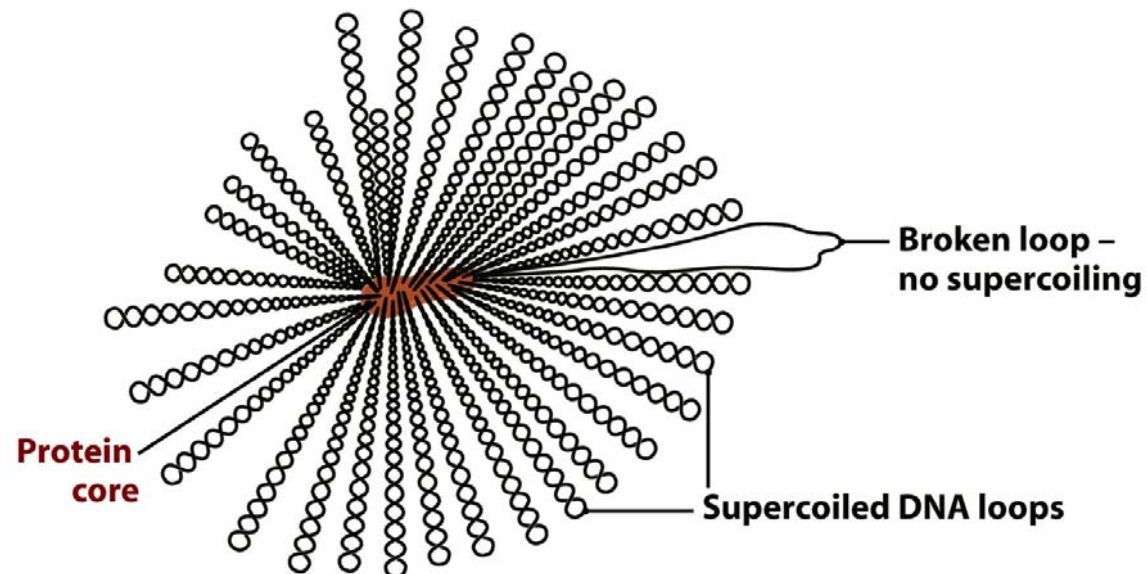
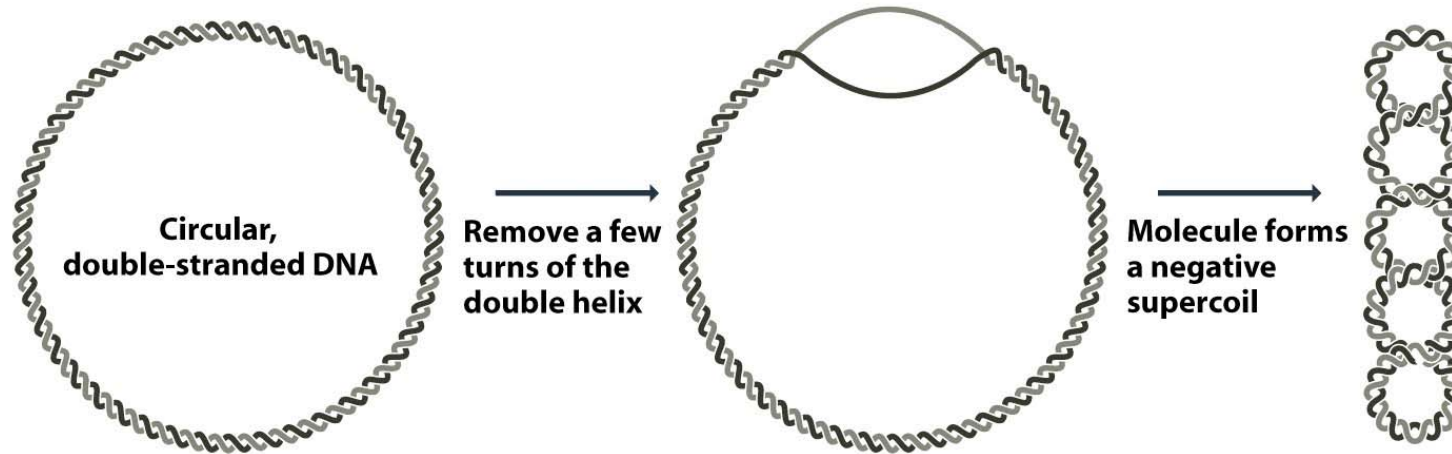
PROKARYOTIC GENOME

PROKARYOTIC GENOME

- Prokaryota
 - bacteria
 - archaea

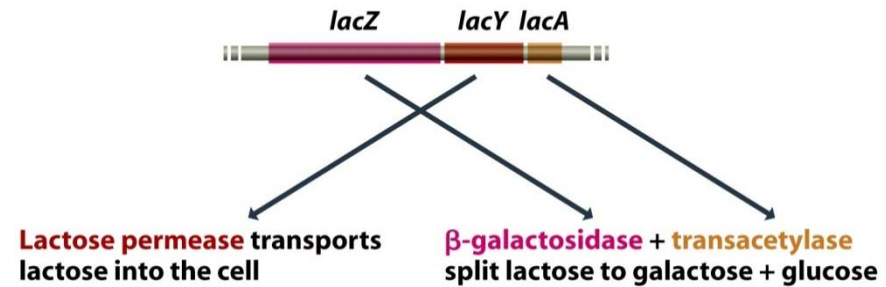


PROKARYOTIC GENOME

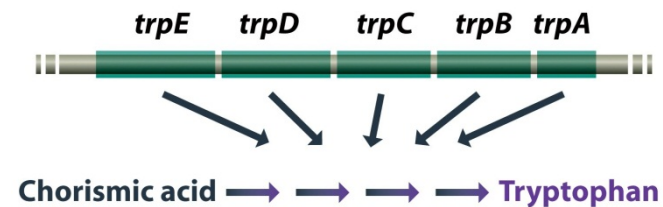


OPERON

- lactose operon
 - utilization of lactose



- tryptophan operon
 - same biochemical pathway



- *Methanococcus jannaschii* (archaea) and *Aquifex aeolicus* (bacteria)
 - different functions



PLASMIDS

- additional genetic information
- adaptation to environmental conditions
- advantage for the host
- in some cases inserted into the main genome

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

PROKARYOTIC GENOMES

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	

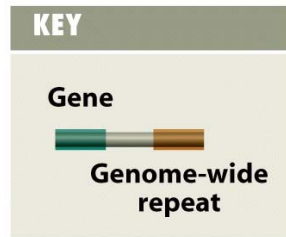
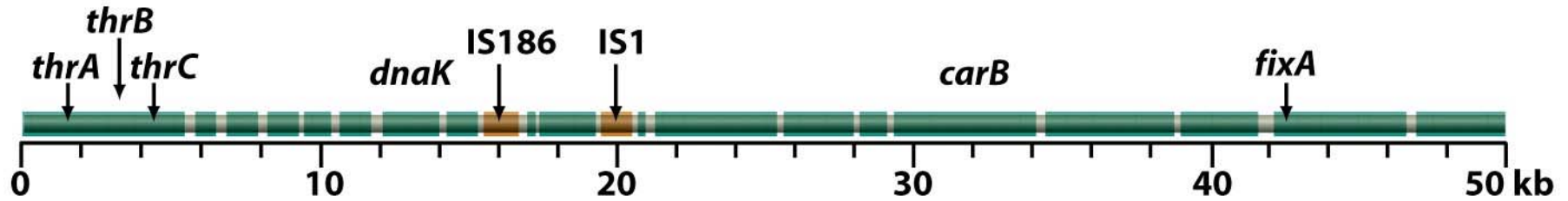
SIZE OF PROKARYOTIC GENOMES

- in most cases size of the genome correlates with the number of genes
- average: 950 genes per 1Mb

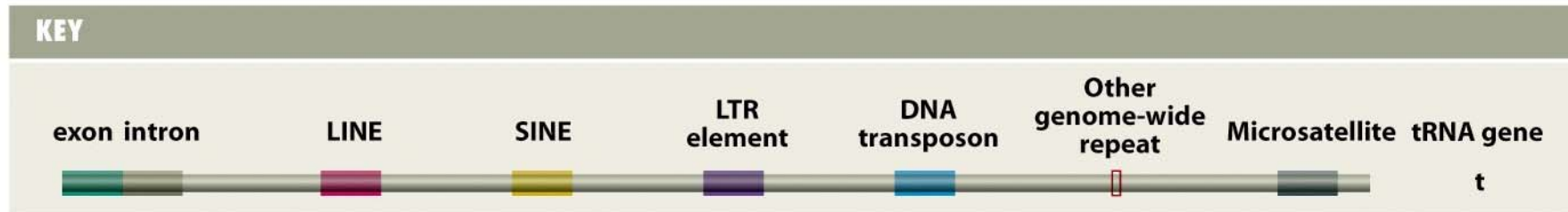
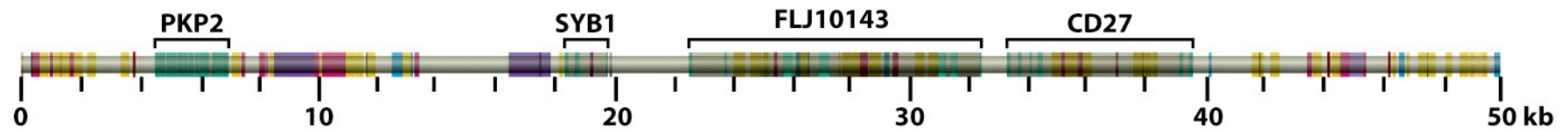
Species	Size of genome (Mb)	Approximate number of genes
Bacteria		
<i>Mycoplasma genitalium</i>	0.58	500
<i>Streptococcus pneumoniae</i>	2.16	2300
<i>Vibrio cholerae</i> El Tor N16961	4.03	4000
<i>Mycobacterium tuberculosis</i> H37Rv	4.41	4000
<i>Escherichia coli</i> K12	4.64	4400
<i>Yersinia pestis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

PROKARYOTES vs EUKARYOTES

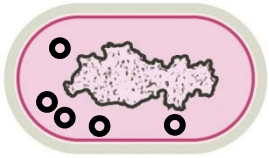
E. coli



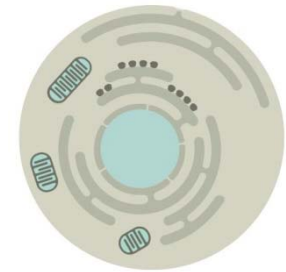
human



PROKARYOTES vs. EUKARYOTES



- nucleoid
- free in cytoplasm
- -
- plasmids
- compact
- majority of genome: coding sequences
- **operons**
- few repetitive sequences
- -



- chromosomes
- in nucleus
- **introns**
- -
- **C-value paradox**
- majority of genome: noncoding sequences
- -
- large number of repetitive sequences
- mitochondrial and chloroplast genomes

EUKARYOTIC ORGANELLE GENOMES

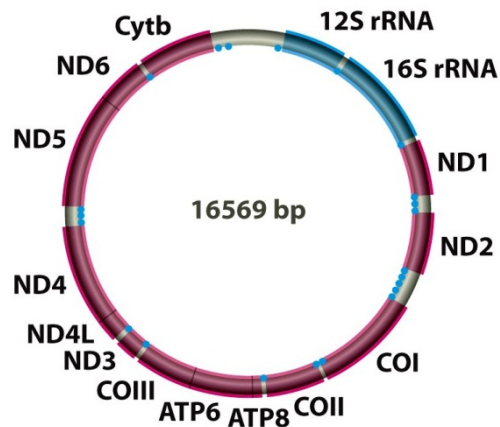
ORGANELLE GENOMES

- endosymbiont theory of organelle origin
 - relics of free-living bacteria
 - symbiotic association with the precursor of the eucaryotic cell
 - endosymbiosis
- from 1 up to a 100 copies in one mitochondrion
- heritability – as one copy ???
- transfer DNA from organelles into nucleus and between organelles
 - *Arabidopsis*
 - mitochondrial genome contains nuclear and chloroplast DNA
 - nuclear genome contains sequences of chloroplast and mitochondrial DNA
 - vertebrates
 - mitochondrial DNA in nuclear genome

MITOCHONDRIAL GENOME

- circular or linear
- 1 mitochondrion – 10 identical molecules = approximately 8000 in one cell (human)
- rRNA, tRNA, respiratory chain components, ribosomal proteins, transcription, translation, transport proteins ...

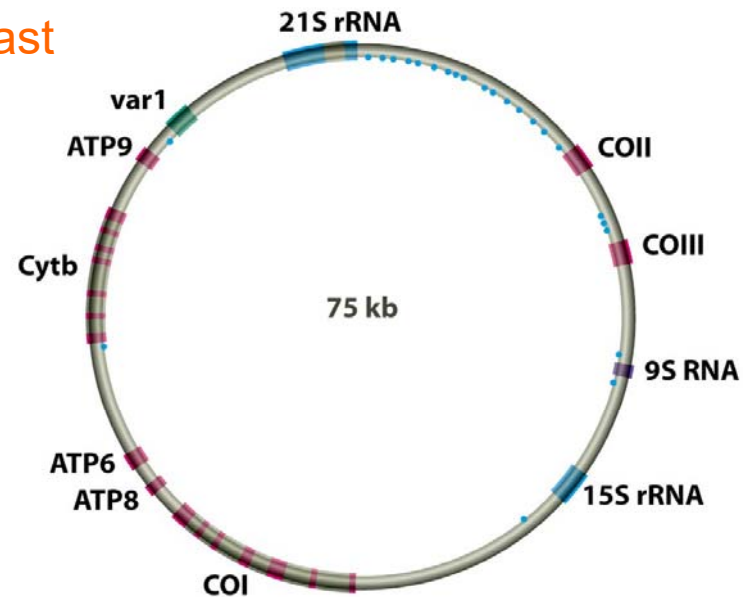
human



KEY

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

yeast



KEY

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

SIZE OF ORGANELLE GENOMES

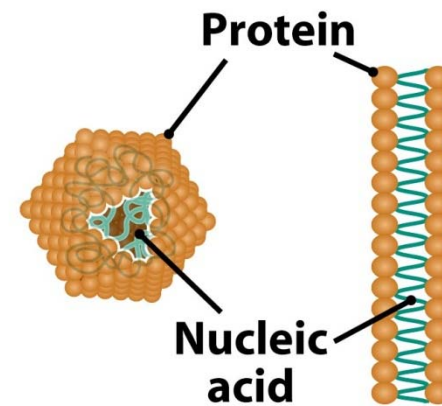
Species	Type of organism	Genome size (kb)
Mitochondrial genomes		
<i>Plasmodium falciparum</i>	Protozoan (malaria parasite)	6
<i>Chlamydomonas reinhardtii</i>	Green alga	16
<i>Mus musculus</i>	Vertebrate (mouse)	16
<i>Homo sapiens</i>	Vertebrate (human)	17
<i>Metridium senile</i>	Invertebrate (sea anemone)	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
Chloroplast genomes		
<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

VIRUS GENOMES AND MOBILE ELEMENTS



VIRUS GENOMES

- virus – nucleoprotein particle
- dependent on the host = obligate parasites – they need ribosomes and translational apparatus to synthesise the protein coat
 - bacteriophages (~ phages)
 - eukaryotic
- virus genome
 - DNA or RNA
 - circular or linear
 - ss or ds
 - segmented or nonsegmented

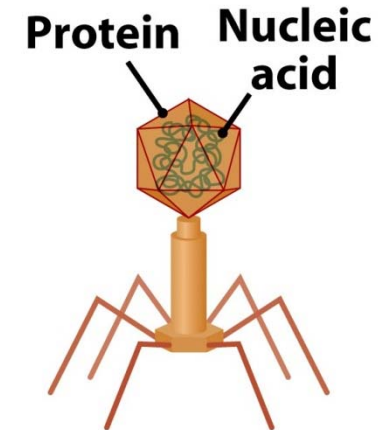


Icosahedral

MS2

Filamentous

M13



Head-and-tail

T4, λ NK2

Slide 42

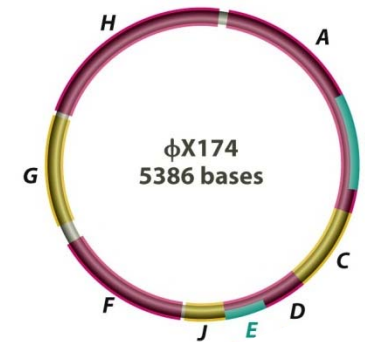
NK2

Pozor, jsou to všechny bakteriofágy... Nebylo by od věci třeba na ten Icosahedral dát adenovirus (common cold)

Nagavalli Kiran; 05.08.2016

BACTERIOPHAGE GENOMES

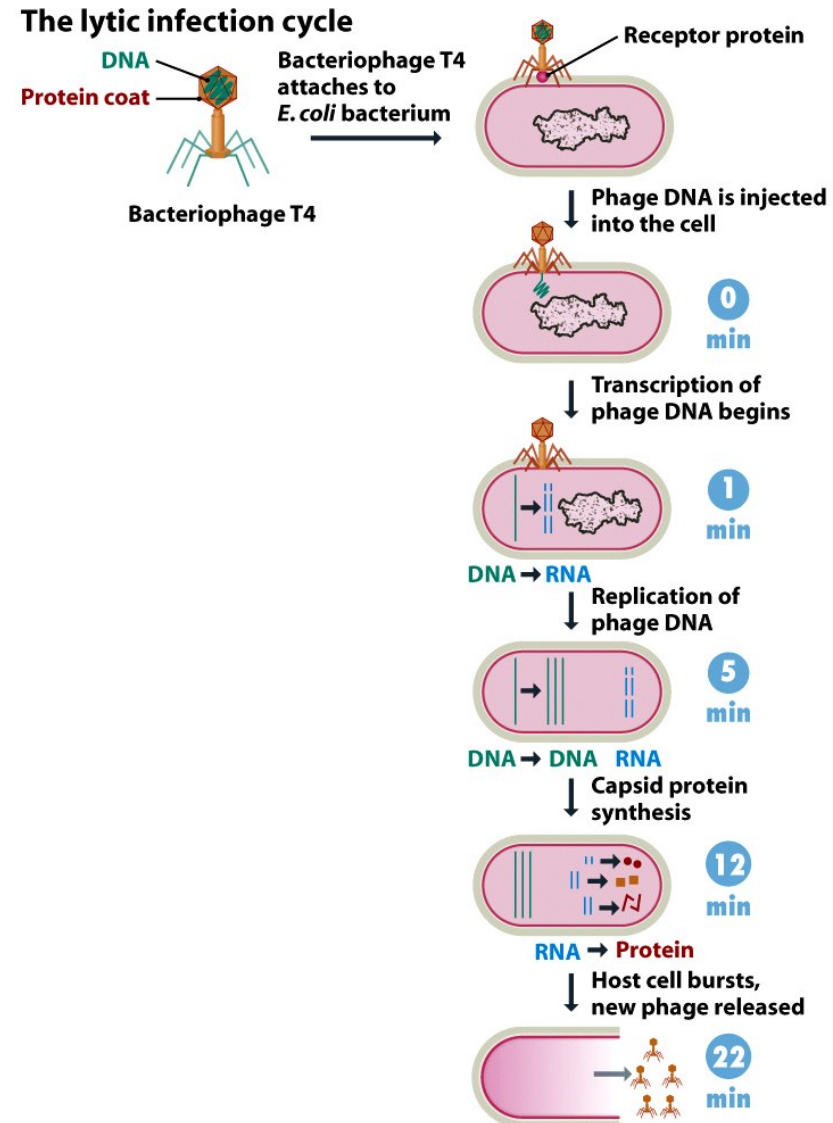
- number of genes: 3-200
- overlapping genes
- phages
 - lytic (virulent), e.g. T4
 - lysogenic (temperate), e.g. phage λ



Phage	Host	Capsid structure	Genome structure	Genome size (kb)	Number of genes
λ	<i>Escherichia coli</i>	Head-and-tail	Double-stranded linear DNA	49.5	48
ϕ 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+

LYTIC INFECTION

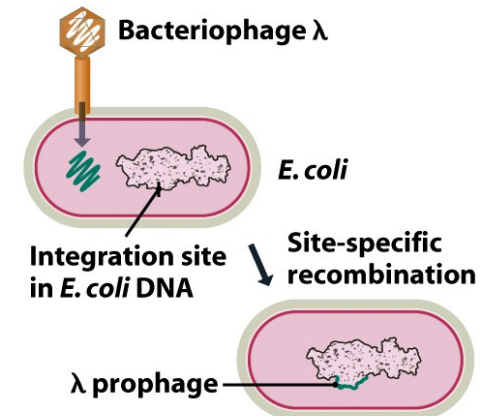
- = virulent, productive
- e.g. phage T4
- fast cell lysis and death
- latent period – time needed for the phage reproduction in host
 - 22 minutes



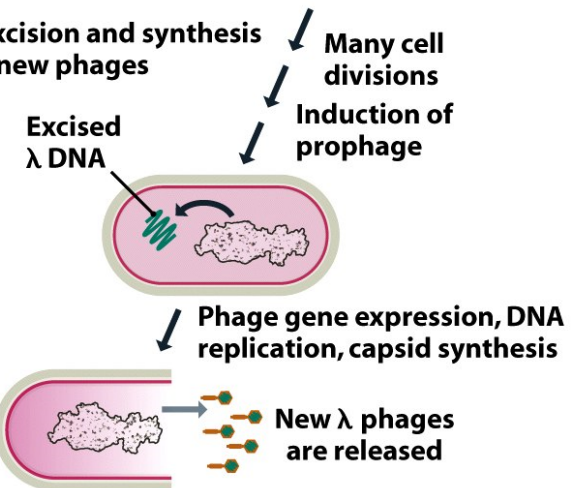
LYSOGENIC INFECTION

- = temperate, quiescent
- e.g. phage λ
- immediately after phage DNA entry – virus genome integration into the host genome by site-specific recombination – prophage
- induction of the prophage excision – chemical or physical factors – probably connected with DNA damage

(A) Integration into the host DNA



(B) Excision and synthesis of new phages



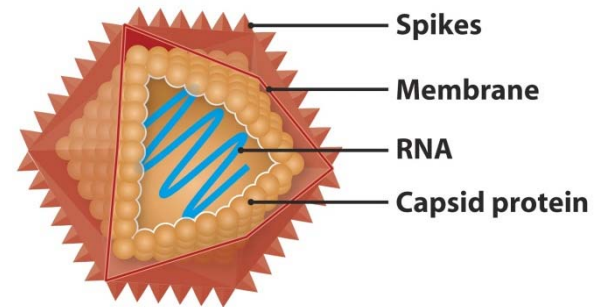
EUKARYOTIC VIRUSES

- variable
 - DNA or RNA; ds or ss; circular or linear; segmented or nonsegmented
- size: 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

EUKARYOTIC VIRUSES

- capsid – icosahedral or filamentous
- lipid membrane – derived from host
- plant viruses – usually RNA
- lytic and lysogenic infection
- e.g. viral retroelements
 - retroviruses – RNA genome
 - pararetroviruses – DNA genome



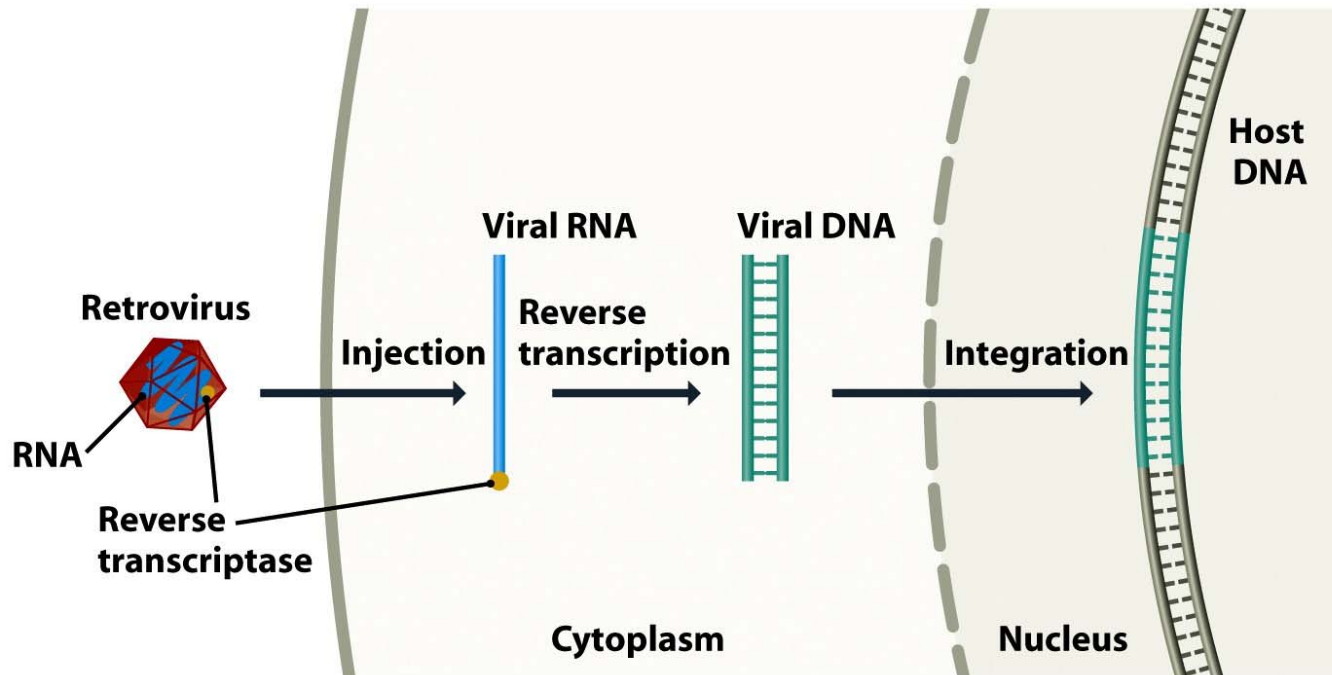
RETROVIRUSES

- genome – each of three genes encode polyproteins which are cleaved, after translation, into two or more functional gene products
 - *gag* – viral core structure = group antigens
 - *pol* – reverse transcriptase, integrase, protease functions
 - *env* – viral capsid proteins = envelope
 - LTR – important regulatory regions for transcription and replication



RETROVIRUSES

- integration of retroviral genome into the host genome

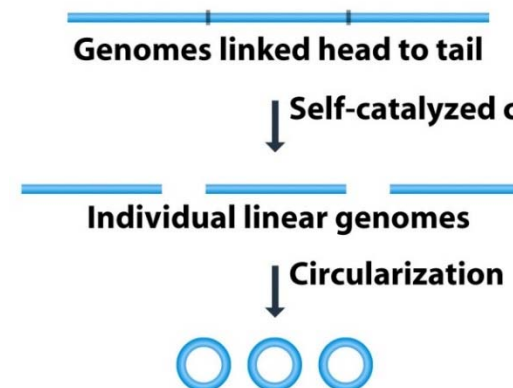


VIRUSOIDS AND VIROIDS

- satellite viruses or virusoids – especially in plants
 - RNA molecules, 320-400 bases
 - satellite virus – shares the capsid with the genome of the helper virus
 - virusoid – encapsulated on its own
- viroid
 - RNA molecule, 240-375 bases, no genes, never become encapsidated = naked RNA

- circular single-stranded molecules
- replicated by enzymes coded by the host or helper virus
- self-catalyzed cleavage
- probably related to evolution of RNA splicing

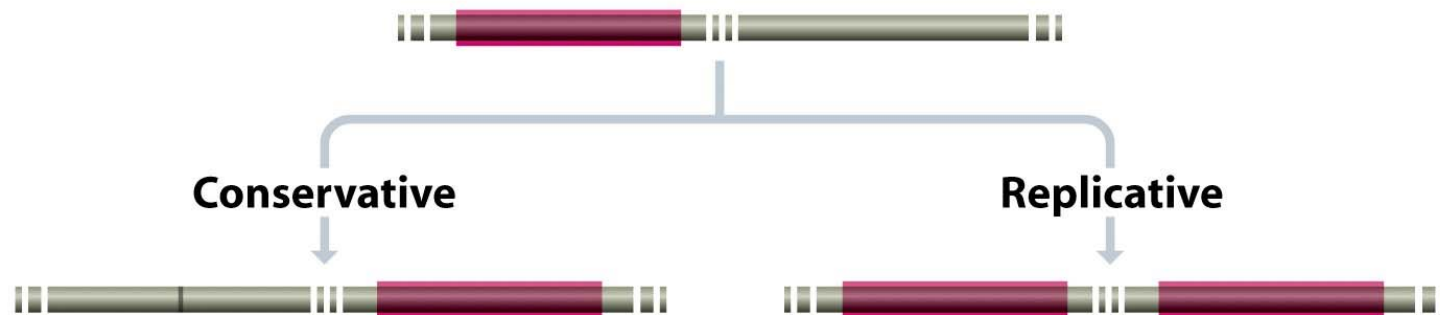
Self-catalyzed cleavage of viroid and virusoid RNAs



MOBILE ELEMENTS

MOBILE ELEMENTS

- = transposons
- transposition – segment of DNA can move from one position to another in a genome
 - conservative
 - replicative
- involves recombination

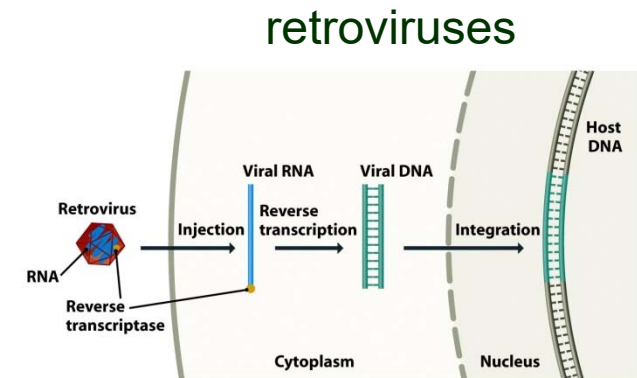
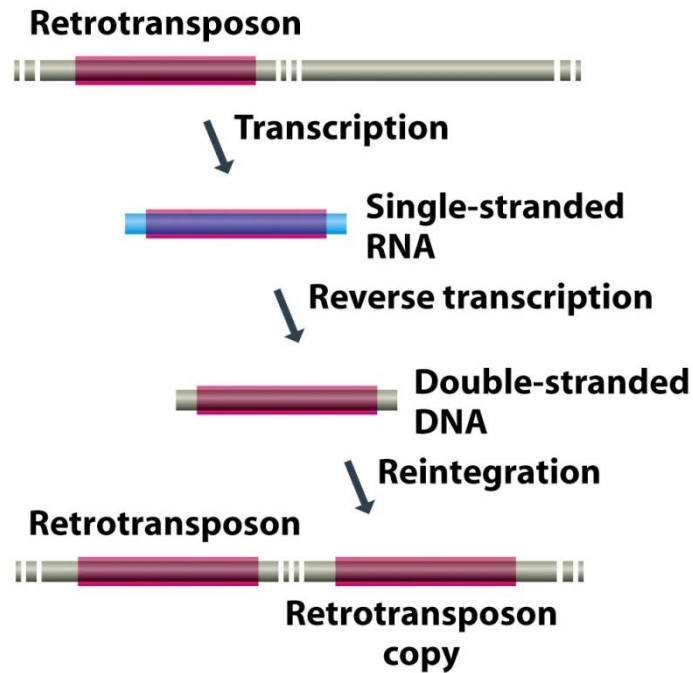


MOBILE ELEMENTS

- RNA transposons
 - retrotransposons with LTR
 - retrotransposons without LTR
- DNA transposons
 - in prokaryotic genomes
 - Insertion sequence (IS)
 - Composite transposon
 - Tn3-type transposon
 - Transposable phage
 - ...
 - in eukaryotic genomes
 - Ac/Ds
 - Spm
 - ...

RETROTRANSPOSONS

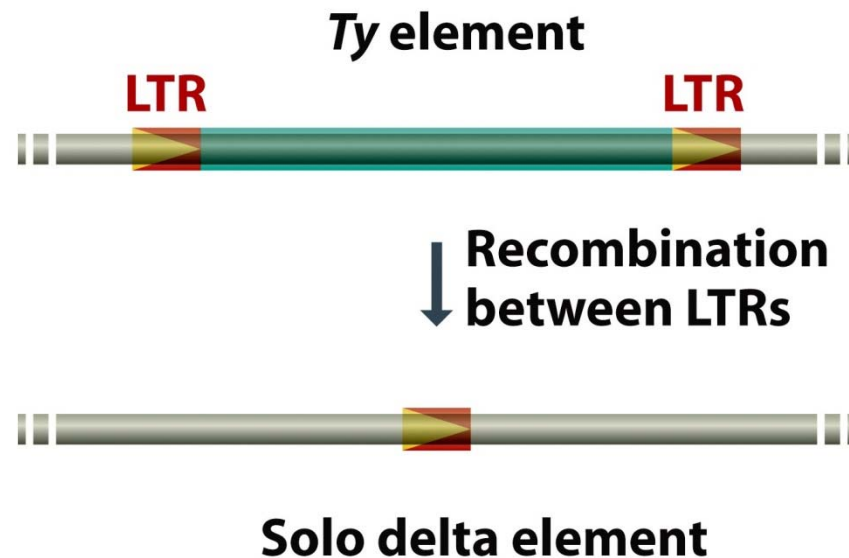
- transposition via RNA intermediate



- retrotransposons
 - with LTR sequence
 - without LTR sequence

RETROTRANSPOSONS WITH LTR

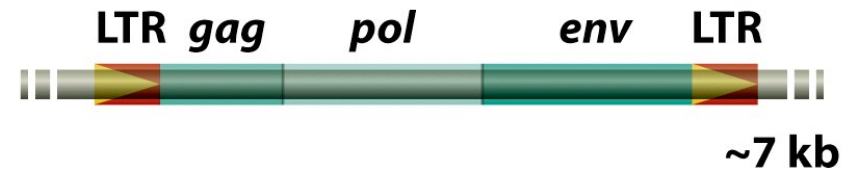
- *Ty* element
 - first discovered in yeast
 - 6.3 kb, 25-35 copies
- „delta“ element
 - LTR sequences
 - 330 bp
 - around 100 copies



RETROTRANSPOSONS WITH LTR

- *Ty1/copia*
 - most abundant
 - *env* gene is missing
 - cannot form infectious virus particles - cannot escape from the host cell
 - form virus like particles (VPL)
- *Ty3/gypsy*
 - *env* equivalent
 - some of them form infectious viruses
- endogenous retroviruses (ERV)
 - humans, mammals

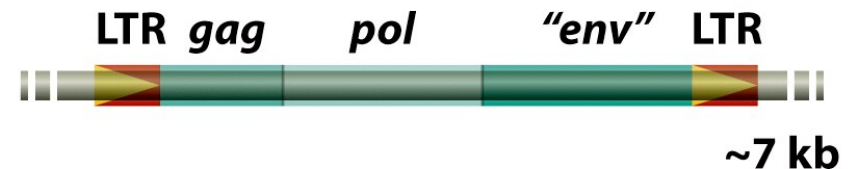
(A) Viral retroelement



(B) *Ty1/copia* retroelement



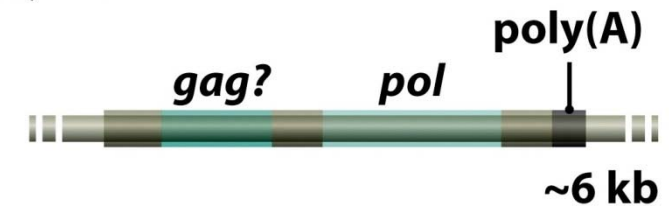
(C) *Ty3/gypsy* retroelement



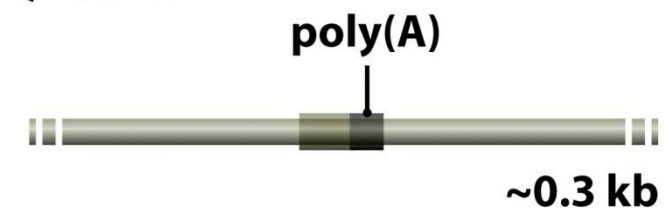
RETROTRANSPOSONS WITHOUT LTR

- retroposons
 - LINEs (long interspersed nuclear elements)
 - *pol* gene
 - functional reverse transcriptase
 - SINEs (short interspersed nuclear elements)
 - 100-400 bp
 - no gene
 - „borrow“ reverse transcriptase from LINE
 - e.g. Alu element

(A) LINE



(B) SINE



DNA TRANSPOSONS IN PROKARYOTES

- do not require RNA intermediate
- less common than retrotransposons
- IS – insertion sequence
 - conservative and replicative transposition
- composite transposons
- Tn3-type
 - lacks IS
 - replicative transposition
- Transposable phages
 - replicative transposition

(A) Insertion sequence



(B) Composite transposon



(C) Tn3-type transposon



(D) Transposable phage

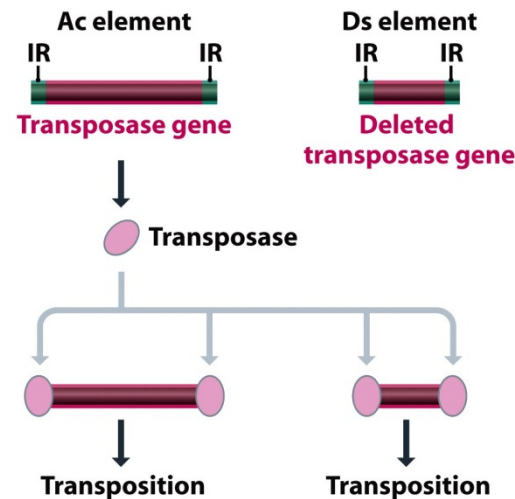


DNA TRANSPOSONS IN EUKARYOTES

“JUMPING GENES”

NK3

- Human genome
 - 350 000 transposons
 - inverted tandem repeats (ITR)
 - gene for transposase
 - usually nonfunctional
- Maize
 - Ac/Ds elements
 - Spm element
- *Drosophila*
 - P element

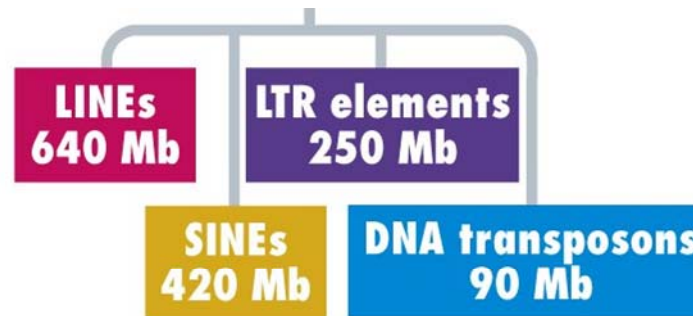


NK3

Předpokládám, že budeš mluvit o B. McClintockové, proto jsem přidal její pojmenování kukříčích transposonů

Nagavalli Kiran; 05.08.2016

MOBILE ELEMENTS IN THE HUMAN GENOME



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

SUMMARY

- Eukaryotic nuclear genome
 - chromosomes
 - genes
 - intergenic DNA
 - gene catalog
- Prokaryotic genome
 - nukleoid
 - plasmids
- Mitochondrial and chloroplast genomes
- Virus genomes
 - bacterial viruses – phages
 - eukaryotic viruses
- Mobile elements
 - RNA transposons
 - DNA transposons

ONLINE AND LITERATURE SOURCES

ONLINE SOURCES

The screenshot shows the GOLD website interface. At the top left is the GOLD logo with the text "Genomes Online Database". To the right is a "JGI/IMG Gold User Log in" button. Below the logo is a navigation menu with links: Home, Search, Distribution Graphs, Biogeographical Metadata, Statistics, References, Team, Help, News. The main content area is titled "Welcome to the Genomes OnLine Database" and "GOLD Release v.5". It includes a summary of the database's purpose: "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." On the left, there is a table showing counts for Studies (20354), Biosamples (59036), Sequencing Projects (59171), and Analysis Projects (44919), with a "Download Excel Data file" link below. The main content is divided into four sections: 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), and Organisms (Organisms 53794, Archaea 935, Bacteria 39183, Eukarya 9250). Below these sections is a three-step process: 1. Register (with GOLD logo and "Register" button), 2. Annotate (with a database icon and "Annotate" button), and 3. Publish (with SIGS logo and "Publish" button).

Studies	20354
Biosamples	59036
Sequencing Projects	59171
Analysis Projects	44919

[Download Excel Data file](#)

Welcome to the Genomes OnLine Database

GOLD Release v.5
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.

Studies	Biosamples	Projects	Organisms
<ul style="list-style-type: none">Metagenomic 545Non-Metagenomic 19809	<ul style="list-style-type: none">Classification<ul style="list-style-type: none">EcosystemsHost-associated 11813Engineered 1655Environmental 6740	<ul style="list-style-type: none">Complete Projects 6649Permanent Drafts 23552Incomplete Projects 26572Targeted Projects 1404	<ul style="list-style-type: none">Organisms 53794Archaea 935Bacteria 39183Eukarya 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](#)

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

ONLINE SOURCES



Genomes Online Database

[Home](#)
[Search](#)
[Distribution Graphs](#)
[Biogeographical Metadata](#)
[Statistics](#)
[References](#)
[Team](#)
[Help](#)
[News](#)

Studies ⁱ	20354
Biosamples ⁱ	59038
Sequencing Projects ⁱ	59171
Analysis Projects ⁱ	44919

Select Fields

GOLD Project ID +	Project Name +	Project Status Complete and Published [remove]
Go0000002	Dictyoglomus turgidum DSM 6724	Complete and Published
Go0000003	Methanospaerula palustris E1-9c, DSM 19958	Complete and Published
Go0000005	Paenibacillus sp. Y412MC10	Complete and Published
Go0000007	Erwinia pyrifoliae Ep1/98	Complete and Published
Go0000008	Thermoanaerobacterium thermosaccharolyticum DSM 571	Complete and Published
Go0000009	Starkeya novella DSM 508	Complete and Published
Go0000013	Corynebacterium kroppenstedtii DSM 44385	Complete and Published
Go0000014	Petrogla mobilis SJ95	Complete and Published
Go0000015	Delftia acidovorans SPH-1	Complete and Published
Go0000016	Clostridium phytofermentans ISDg	Complete and Published
Go0000017	Shewanella baltica OS195	Complete and Published
Go0000018	Hemiselmis andersenii CCMP 644 (nucleomorph)	Complete and Published
Go0000019	Prochlorococcus marinus MIT 9211	Complete and Published
Go0000020	Methanococcus maripaludis C8	Complete and Published
Go0000021	Herpetosiphon aurantiacus DSM 785	Complete and Published
Go0000023	Sorangium cellulosum So ce 58	Complete and Published
Go0000024	Candida albicans SC5314	Complete and Published
Go0000025	Caldivirga maquilingensis IC-187	Complete and Published
Go0000026	Salinispora arenicola CNS-205	Complete and Published
Go0000027	Desulfococcus oleovorans Hxd3	Complete and Published
Go0000030	Azorhizobium caulinodans ORS 571	Complete and Published
Go0000033	Acaryochloris marina MBIC11017	Complete and Published
Go0000034	Alkaliphilus oremlandii OhLAs	Complete and Published
Go0000037	Shewanella pealeana ANG-SQ1, ATCC 700345	Complete and Published
Go0000038	Rickettsia massiliae MTU5	Complete and Published

RESET

[1 - 25] of 6649 [NEXT](#) Show results.

ONLINE SOURCES

Genomes Pages - At the EBI

- Complete genomes
- Archaea
- Archaeal virus
- Bacteria
- Eukaryota
- Organelle
- Phage
- Plasmid
- Viroid
- Virus
- Links
 - WGS info
 - EnsemblGenomes
 - Genome Reviews
 - Integr8 (proteomes)
 - Fasta33 Server
 - Ensembl

Databases > Nucleotide > The European Nucleotide Archive > Complete Genomes

Access to Completed Genomes

The first completed genomes from [viruses](#), [phages](#) and [organelles](#) were deposited into the EMBL Database in the early 1980's. Since then, molecular biology's shift to obtain the complete sequences of as many genomes as possible combined with major developments in sequencing technology resulted in hundreds of complete genome sequences being added to the database, including [Archaea](#), [Bacteria](#) and [Eukaryota](#). These web pages give access to a large number of complete genomes, [help](#) is available to describe the layout.



Whole Genome Shotgun Sequences (WGS)

Methods using whole genome shotgun data are used to gain a large amount of genome coverage for an organism. WGS data for a growing number of organisms are being submitted to DDBJ/EMBL/GenBank.

[More information about WGS projects...](#)

Last 40 Genome Entries

Date	Accession	Description
08-NOV-2014	KM390013.1	Corallorhiza bulbosa plastid
08-NOV-2014	KM390014.1	Corallorhiza maculata var. maculata voucher Freudenstein 2919 (OS) plastid
08-NOV-2014	KM390015.1	Corallorhiza maculata var. mexicana voucher Barrett 232b (OS) plastid
08-NOV-2014	KM390016.1	Corallorhiza maculata var. occidentalis voucher Freudenstein 2095 (OS) plastid
08-NOV-2014	KM390017.1	Corallorhiza macrantha voucher Salazar A (OS) plastid
08-NOV-2014	KM390018.1	Corallorhiza mertensiana voucher Freudenstein 1999 (OS) plastid
08-NOV-2014	KM390019.1	Corallorhiza trifida voucher Freudenstein 2787a MI plastid
08-NOV-2014	KM390020.1	Corallorhiza wisteriana voucher Freudenstein 2462 (OS) plastid
08-NOV-2014	KM390021.1	Corallorhiza odontorhiza plastid
08-NOV-2014	LC009435.1	Andinoacara rivulatus mitochondrion
08-NOV-2014	AB859014.1	Equus caballus mitochondrion, strain: trotter
08-NOV-2014	CP008741.1	Bifidobacterium longum subsp. longum GT15
08-NOV-2014	CP007241.1	Streptococcus pyogenes strain 1E1
08-NOV-2014	CP009811.1	Bacillus subtilis strain Bs-916
08-NOV-2014	CP009812.1	Streptococcus pyogenes strain HKU360
08-NOV-2014	CP009748.1	Bacillus subtilis strain ATCC 13952
08-NOV-2014	CP009749.1	Bacillus subtilis strain ATCC 19217
08-NOV-2014	CP009754.1	Streptomyces sp. CCM_MD2014

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

ONLINE SOURCES

GNN Genome News Network

Home | About | Topics Search GNN

FEATURE



Genomes of the World

A pufferfish, a boxer, and a parasite that frequents swimming pools are among the organisms whose genomes have recently been sequenced. For more sequenced genomes visit GNN's [Quick Guide](#).

NEWS STORIES

Bioterrorism, Cloning, and Stem Cells

Diseases and Personalized Medicine
Alzheimer's; Cancer; Cystic Fibrosis; Depression and Bipolar Disorder; Diabetes; HIV/AIDS; Heart Disease; Huntington's; Multiple Sclerosis; Obesity; Parkinson's; SARS; Schizophrenia; more...

Drugs and Gene Tests
Antibiotics; Drug Development; Vaccines

Energy and the Environment
Extremophiles; Oceans; Toxic Cleanup

Genes and Genomes
Biobanks; Evolution; Human Genome; Sex and Reproduction

Microbes

Technology
Microarrays; Movies and Imaging; Nanotechnology

Weird Science

A Quick Guide to Sequenced Genomes



What's a Genome?



Art Gallery



Glossary

Bioethics Central

Stem Cells: Policies and Players

Book Reviews

Timeline

NEWS BY TOPIC | ARCHIVE | SEARCH

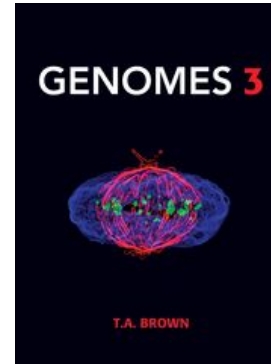
[GNN gratefully acknowledges its collaboration with the Albert and Mary Lasker Foundation.](#)

Genome News Network is an editorially independent online publication of the J. Craig Venter Institute.
© 2000 - 2004 J. Craig Venter Institute.
All rights reserved. This material may not be published, broadcast, rewritten or redistributed.

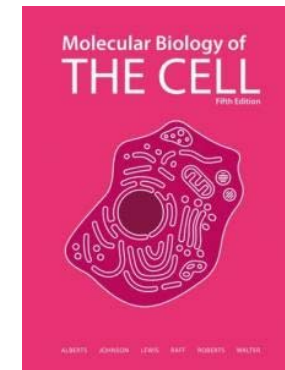
<http://www.genomenetwork.org/>

LITERATURE

- T. A. Brown: Genomes



- Alberts et al.:
Molecular Biology of the Cell



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



+ internet, scientific papers ...