

Ecological and Evolutionary Genomics

BIO 594-0004

Spring 2019

Instructor: Carlos Prada
CBLS 181
401 874 5767
prada@edu

[GitHub](#)
<http://bio-594.slack.com/>

When and where do we meet:
Tuesdays 9 - 10:30 AM CBLS 435 Thursdays 9 -
10:30 AM CI 200

Learning Objective

The goal is to recognize the diverse possibilities that genomics allow to answer evolutionary and ecological questions. At the end of the course participants will be able to design experiments to answer their own questions and provide a detailed map of the methodological steps needed to carry out a robust study using genomic tools

What do we suppose to do in this class?

Discuss, discuss, discuss!

We are here to dissect to the best of our possibilities the assigned articles and then use that info to design a project that uses genomic tools to answer an ecological or an evolutionary question

We are meeting two days:

on Thursdays we will introduce a new subject (lecture)

on Tuesdays we will discuss two/three articles related to that subject

Why do we have a lecture/recitation before each discussion?

The presentation is key to have fruitful subsequent discussions because it will allow us to get a deeper understanding of the papers

The aim of the presentation is to:

- introduce the subject (theory, methods, controversies)
- discuss all the vocabulary in reviews and research papers
- an opportunity for participants to clarify concepts (**ASK!**)

Why do we have a lecture/recitation before each discussion?

The presentation is key to have fruitful subsequent discussions because it will allow us to get a deeper understanding of the papers.

The aim of the presentation is to:

- introduce the subject (theory, methods, controversies)
- discuss all the vocabulary in reviews and research papers
- an opportunity for participants to clarify concepts (**ASK!**)

At least one review paper with baseline info to develop the presentation. The review paper is **MANDATORY** for **non-presenters** too. Presenters should read also the upcoming discussion papers to make sure that they include all the necessary vocabulary in those papers in their presentations.

Presentation: 30-60 min (any strategy). Meet with me a week (seven days) before the presentation and as many times as you want.

What are we doing in this class?

There are four parts to this course:

Class participation (15%)

Student presentations (20%)
presentation (30-60 min)

Discussions (25%)
summary writeups
discussion writeups

Proposals (40%)
proposal (20%)
updated proposal (10%)
reviews from other proposals (10%)
panel discussions and panel summaries

What do we suppose to do in this class?

Discussions

summary writeups (each student)

(750-1000-word compare and contrast of the articles).

Each student has to submit at least 5 of them before each discussion (earlier than 9 am on Tuesdays).

Examples are provided [here](#)

discussion writeups (in pairs)

(1000-2000-word summary of the discussion, which needs to be uploaded to the class GitHub within 48 hours of the discussion and it is submitted by one of the two discussion leaders). Each student leads two discussions with a partner

What do we suppose to do in this class?

Proposals

proposal (5-page proposal, single spaced, 1", 2 aims)

updated proposal (addressing reviewers comments)

reviews from other proposals (each will review at least three proposals from their peers)

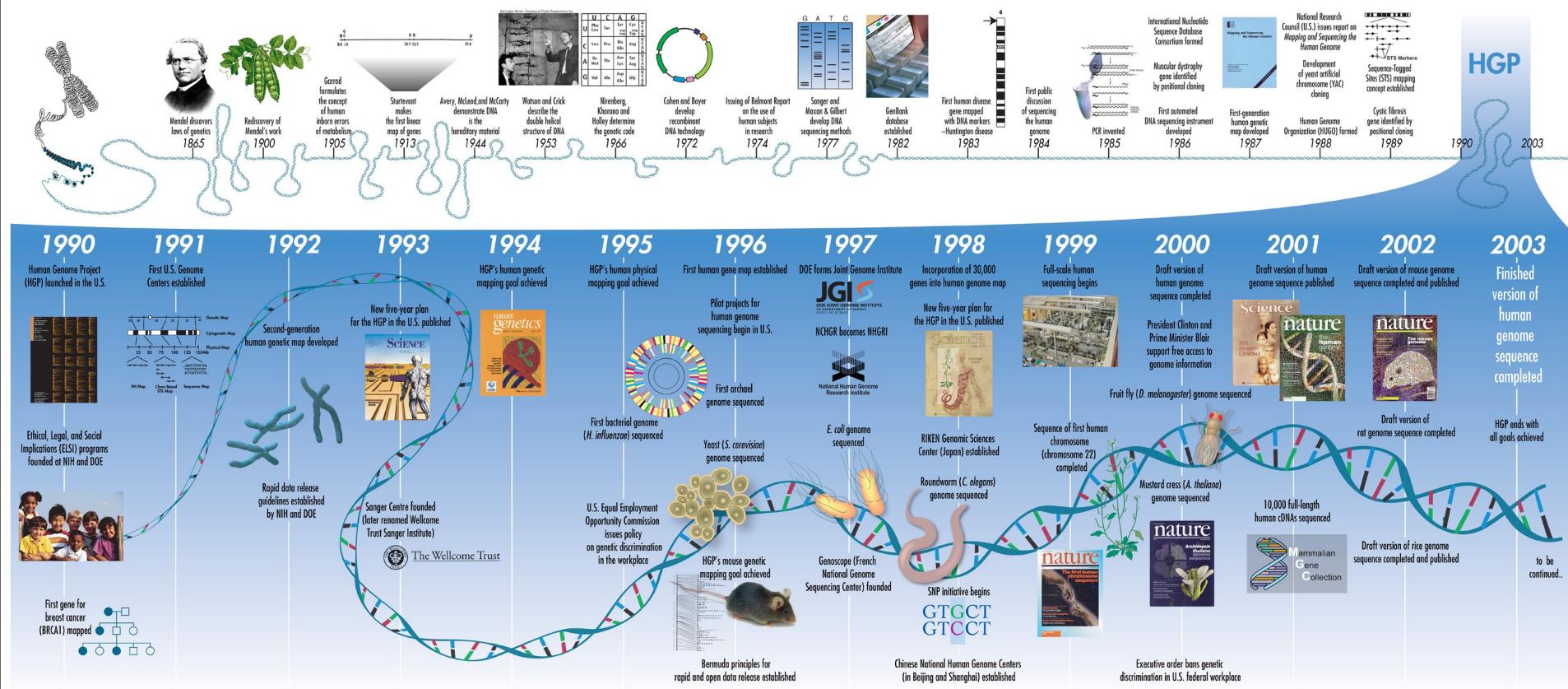
panel discussions and panel summaries (proposals will be evaluated and ranked as an NSF panel.
Panelists will have to generate a panel review for each proposal)

Discussion and Lecture Topics

Code	Date	Topic	Name
Intro	Jan 24 (Thu)	Introduction Holenhole et al. 2018	Carlos
No Class	Jan 29 (Tue)	No class	----
1	Jan 31 (Thu)	Population genomics, population structure and demography Schraiber_Akey 2015	Carlos
1a	Feb 5 (Tue)	DISCUSSION Rougemont et al. 2018 , Moreno-Mayar et al. 2018 & Suppl	Carlos
2	Feb 7 (Thu)	Seascape/Landscape Genomics Fu_Akey 2013.pdf & Rellstab et al. 2015.pdf	Amy
2a	Feb 12 (Thu)	DISCUSSION Brauer et al 2016.pdf , Hancock et al. 2012.pdf & Suppl	Amy/Jennifer
3	Feb 14 (Tue)	Correlation between Phenotype and Genotype Weigel_Nordborg_2015 & Mallarino_Abzhakov_2012	Maggie
3a	Feb 19 (Tue)	DISCUSSION Nadeau_et al_2016.pdf , Reid_et al_2016.pdf , Suppl & Bosse_et al_2016	Maggie/Matias
4	Feb 21 (Thu)	Physiology and Gene Expression Ritchie_et al_2015 & Conesa_et al_2016	Cassie
4a	Feb 26 (Tue)	DISCUSSION Bernal_et al_2018 , Lohman_et al_2018 & Walworth_2016	Cassie/Emma
5	Feb 28 (Thu)	Adaptive phenotypic plasticity and epigenetics Bossdorf_et al_2008 , Richards_et al_2017 & Donelson_et al_2017	Kevin
5a	Mar 5 (Tue)	DISCUSSION Ghalambor2015 , Ryu_et al_2018 & Lieu_et al_2018	Kevin/Elaine
6	Mar 7 (Thu)	Developmental biology Davidson_2006 & Mallarino_Abzhakov_2012	Jennifer
No Class	Mar 11-14	No class - Spring Break	----

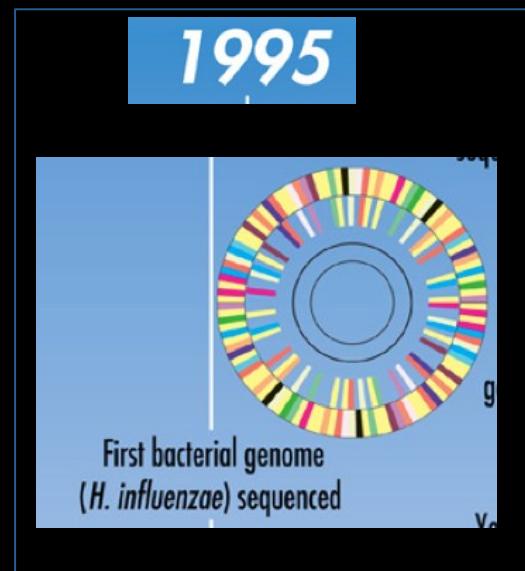
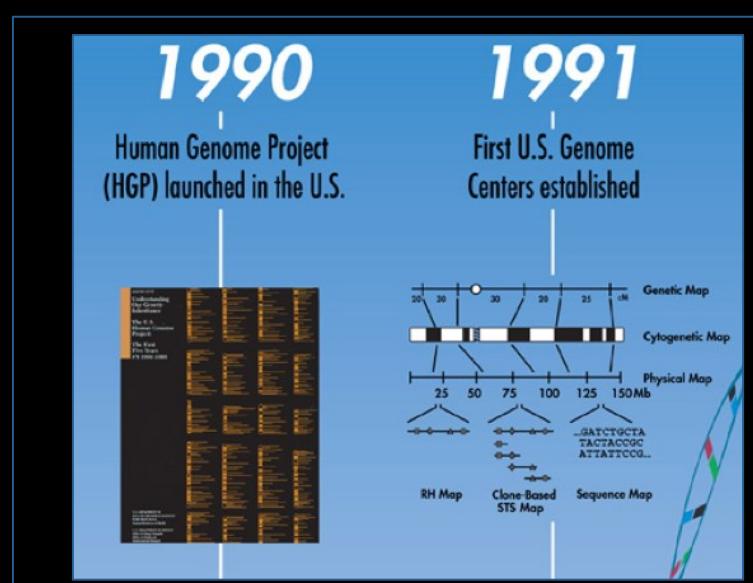
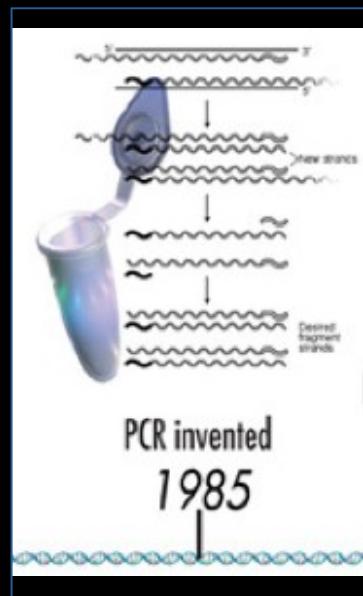
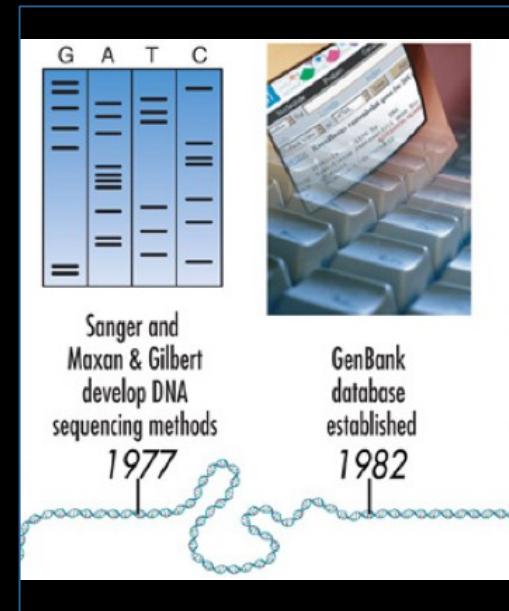
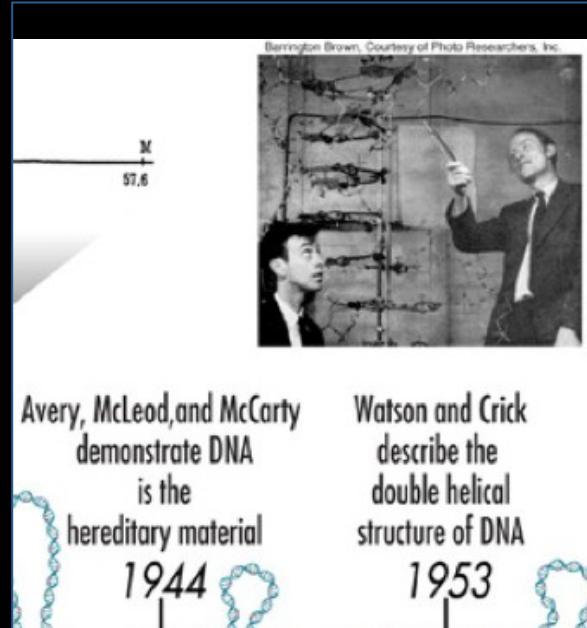
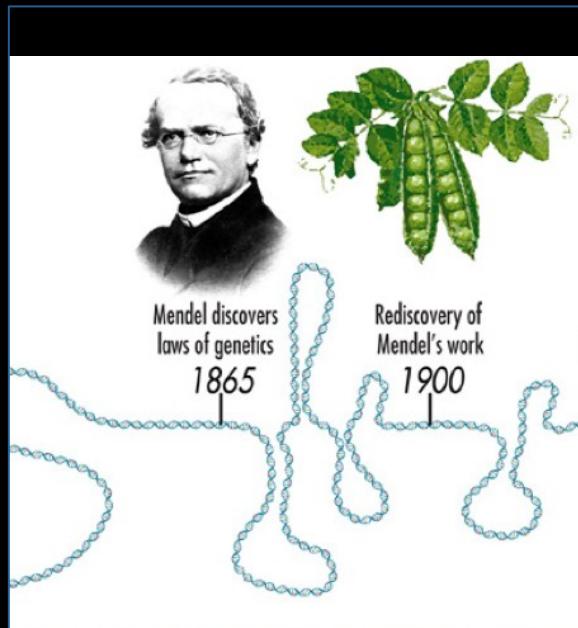
6a	Mar 19 (Tue)	DISCUSSION Israel_etal_2016 & Manceu_etal_2011.pdf	Jennifer/Erin
7	Mar 21 (Thu)	Symbiosis -The genomics of species interactions Gilbert_etal_2015	Ian
7a	Mar 26 (Tue)	DISCUSSION Belcaid_etal_2018 & Li_etal_2018	Ian/Maggie
8	March 28 (Thu)	Speciation and hybridization Feder_etal_2013 & Seehausen_etal_2014	Matias
No Class	Apr 1-7	No class - Carlos in Mexico	---
8a	Apr 9 (Tue)	DISCUSSION Edelman_et_al_2018 & Kautt_etal_2017	Matias/Amy
9	Apr 11 (Thu)	Climate change and conservation biology Hendricks_2018	Emma
9a	Apr 16 (Tue)	DISCUSSION and proposals due Bay_et_al_2018 & Barret_etal_2018	Emma/Erin
10	Apr 18 (Thu)	Community ecology eDNA and microbial genomics Deiner_et_al_2017 & knight_etal_2018	Elaine
No Class	April 23 (Tue)	No class -Carlos in Buffalo- but proposal reviews are due	----
10a	Apr 25 (Thu)	DISCUSSION Boussarie_etal_2018 & Sunagawa_etal_2015	Elaine/Cassie
Panel	Apr 30 (Tue)	Panel discussion -Summit summaries	Ian
Panel	May 2 (Thu)	Panel discussion -Summit summaries	Kevin
11	May 10 (Fri)	Lat day to submit revised proposals	

A timeline to genomics



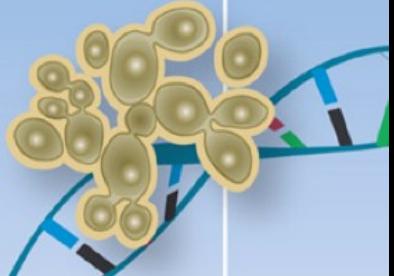
NHGRI – National Human Genome Research Institute

<http://www.genome.gov/dmd/img.cfm?node=Photos/Graphics&id=85325>



1996

Yeast (*S. cerevisiae*) genome sequenced



2000

Fruit fly (*D. melanogaster*) genome sequenced



1998

Roundworm (*C. elegans*) genome sequenced



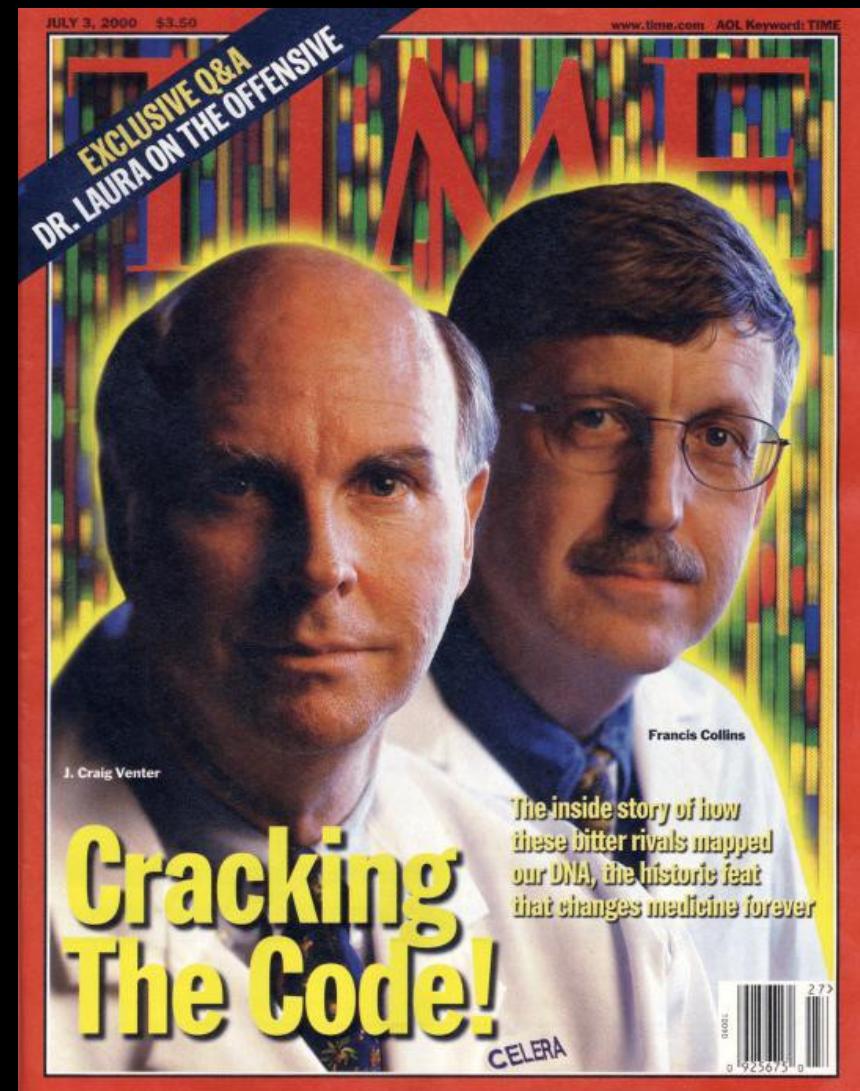
1999

Full-scale human sequencing begins



The Human Genome

Copyright © The McGraw-Hill Companies, Inc. Permission required for reproduction or display.



February 2001

Growth in genomics thanks to advances in...

MICROTECHNOLOGY

DNA/RNA sequencing...

COMPUTATION

human genome assembly: 500 million trillion sequence comparisons

COMMUNICATION

online sequence and expression databases

Genomes and Databases

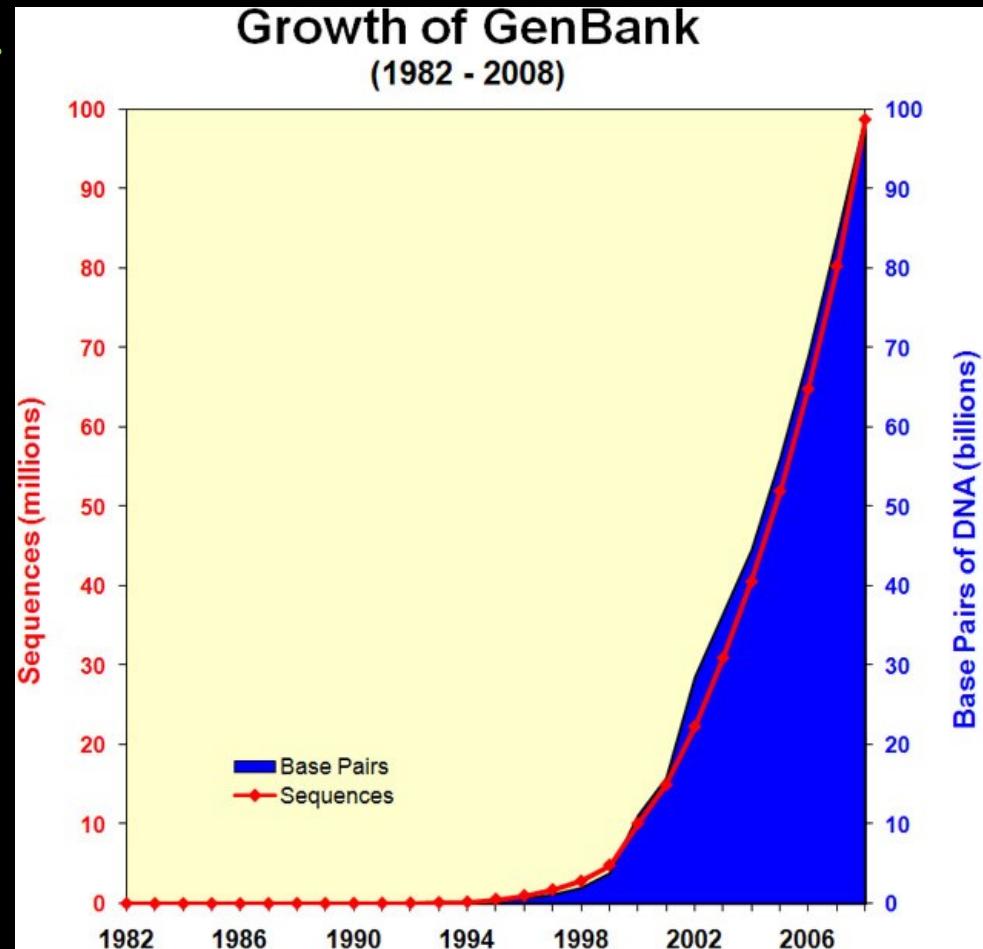
Sequence Accession numbers are IDs that allow sequences to be easily deposited in and retrieved from online databases

CNGB, DDBJ, EMBL, GenBank

The four main sequence databases. They are redundant and work in concert

GenBank is the NIH genetic sequence database, an annotated collection of all publicly available DNA sequences

GenBank is maintained by the NCBI (National Center for Biotechnology Information)



GENOMICS

Seeks to understand the structure and function of all genes in an organism based on knowing the organism's entire DNA sequence and extensive reliance on powerful computer technologies

Rachel Schwartz

<http://rachelss.github.io>

Jon Puritz

https://github.com/jpuritz/BIO_594_2019

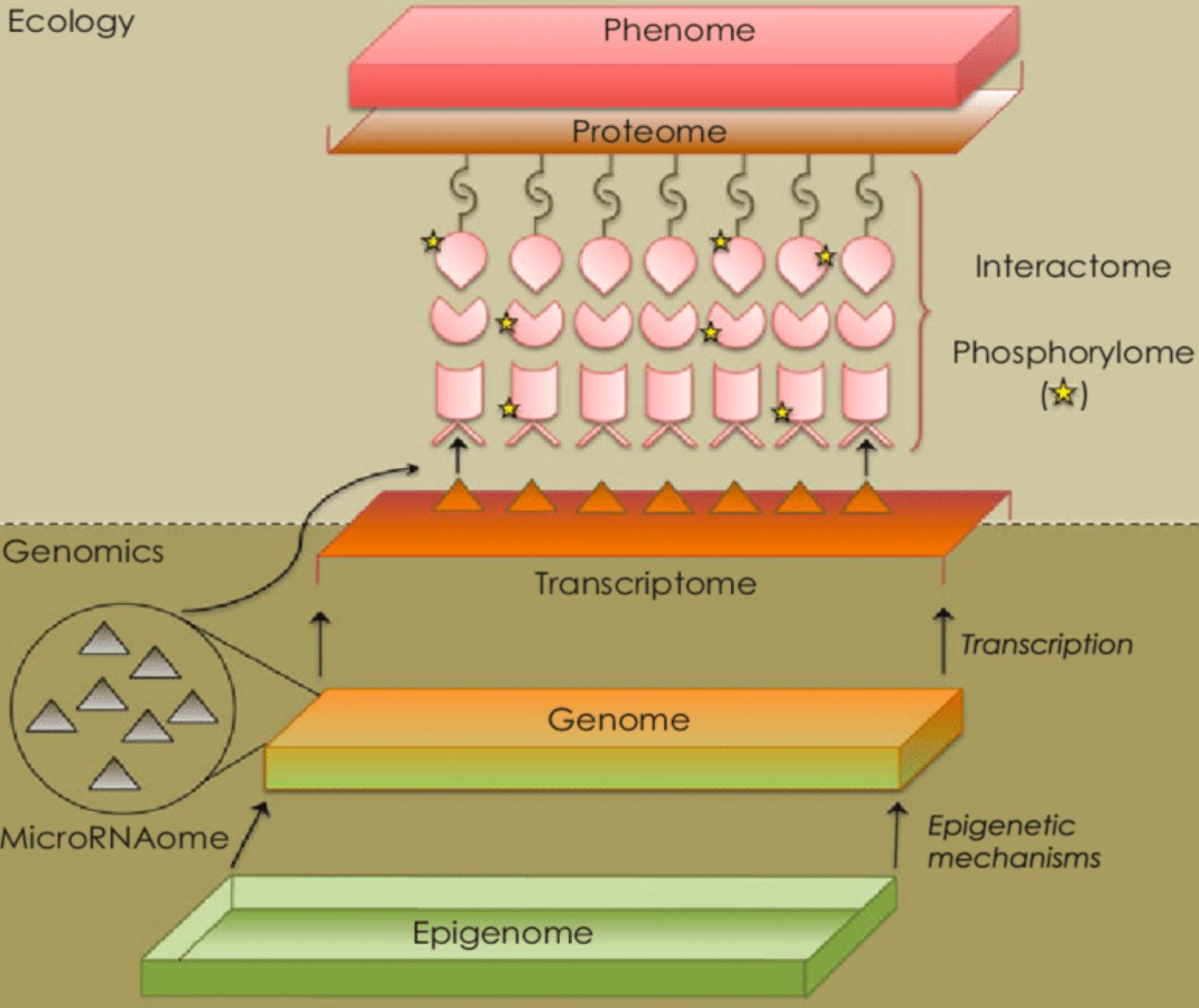
GENOMICS

Seeks to understand the structure and function of all genes in an organism based on knowing the organism's entire DNA sequence and extensive reliance on powerful computer technologies

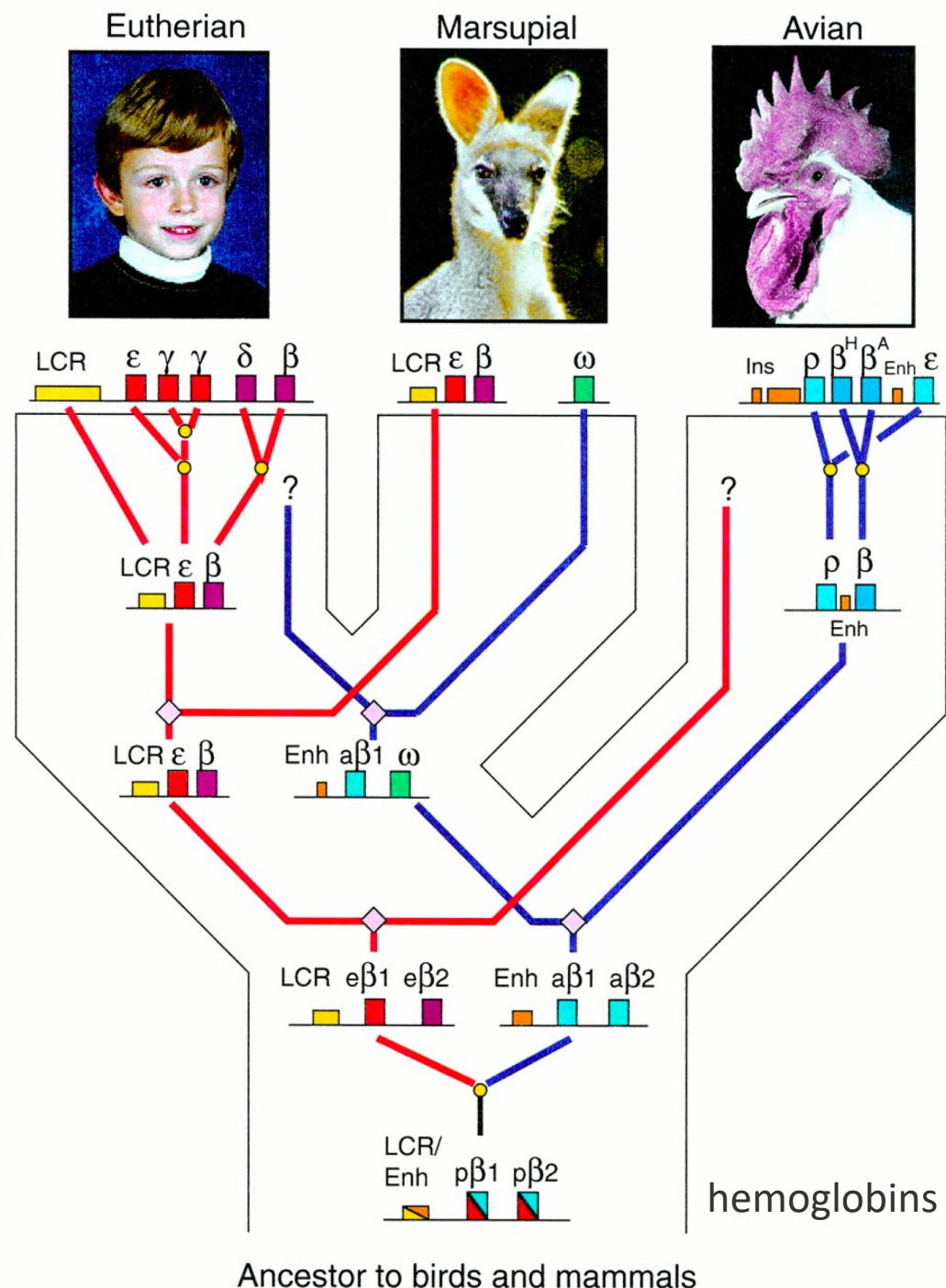
ECOLOGICAL GENOMICS

Seeks to understand how genomes interact with, integrate cues from, and are shaped by the environment

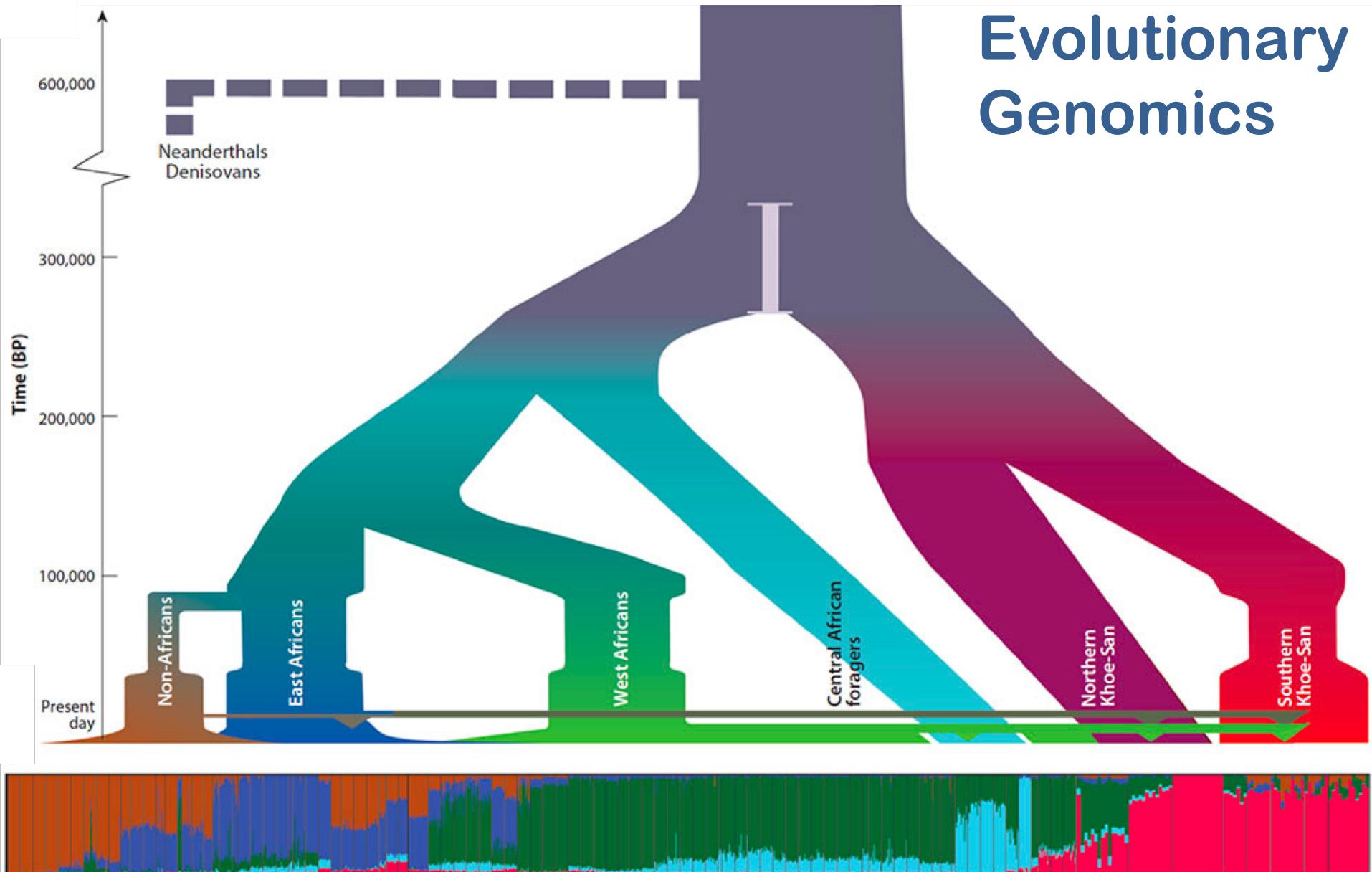
Ecological Genomics



Evolutionary Genomics



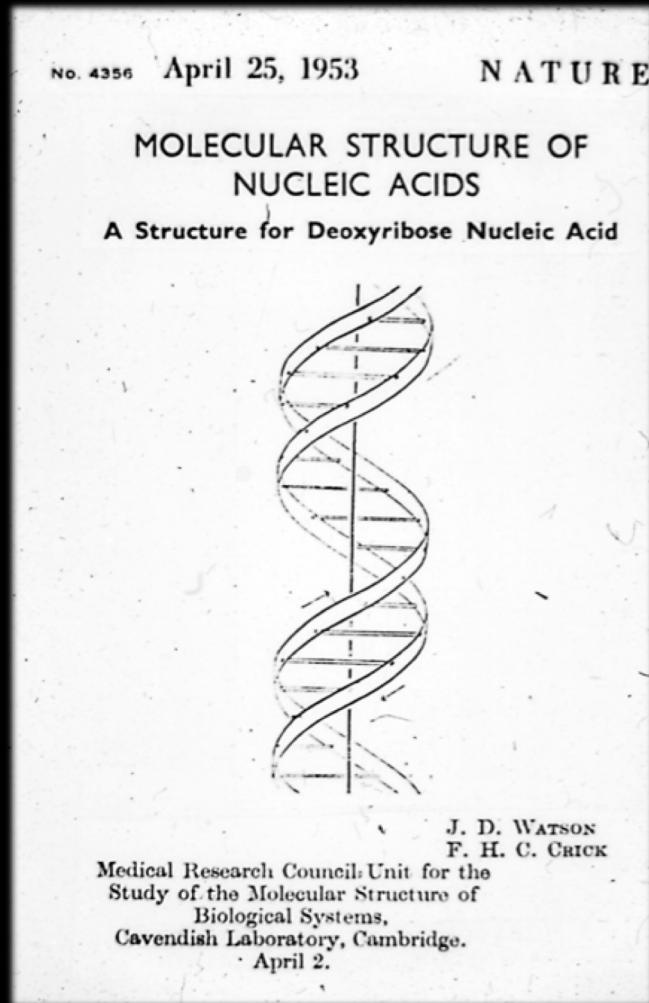
Evolutionary Genomics



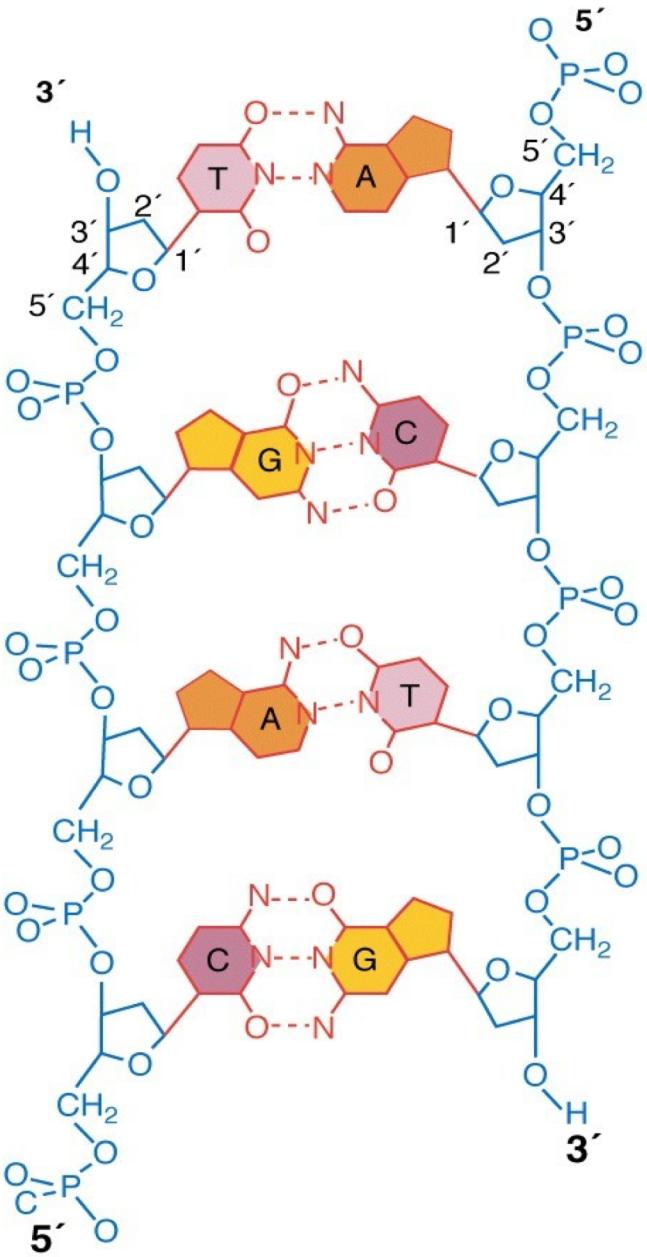
Issue in ecology and evolution	Analytical methods and metrics
<i>Broad-sense genomics</i>	Mostly evolutionary genomics
Estimation of genetic diversity	Heterozygosity, allelic diversity, nucleotide diversity
Effective population size	Linkage disequilibrium (LD), two-sample methods
Population structure, admixture	Bayesian clustering, principal component analysis (PCA)
Source population assignment	Clustering methods
Inbreeding	Identity-by-descent methods
<i>Narrow-sense genomics</i>	Mostly ecological genomics
Mapping phenotypic traits	Genome-wide association studies (GWAS)
Fine-scale demographic history	Coalescent, diffusion approximation methods
Fine-scale estimates of current historic hybridization	Phylogenetic, haplotype-based methods
Loci for local adaptation	Outlier methods, genotype-environment association (GEA), multilocus covariance
Loci for inbreeding depression	GWAS
Loci for adaptive introgression	Outlier, cline analysis
Defining population units on local adaptation	Outlier, GEA

The Structure of DNA

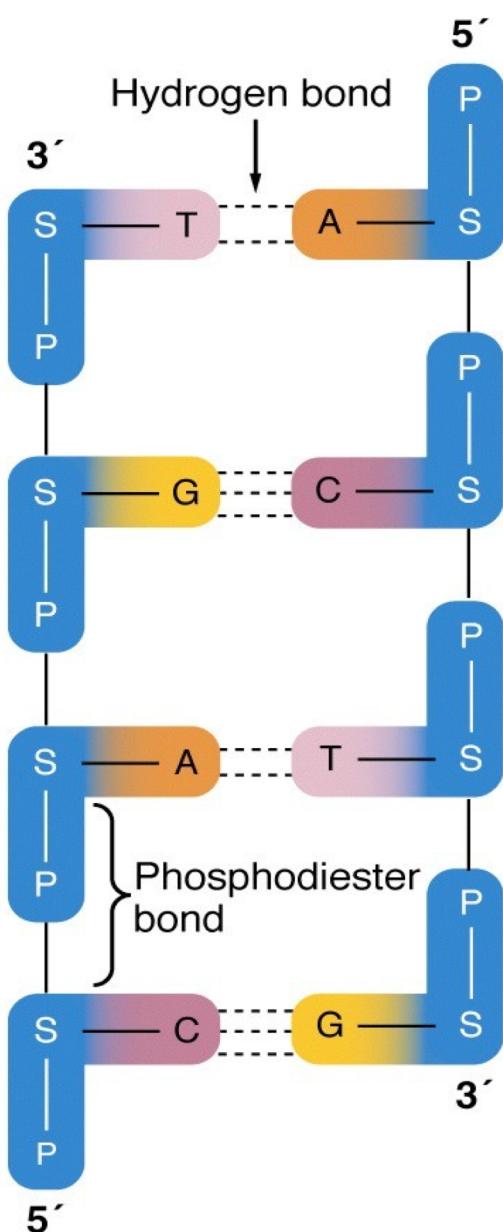
April, 1953



The double helix



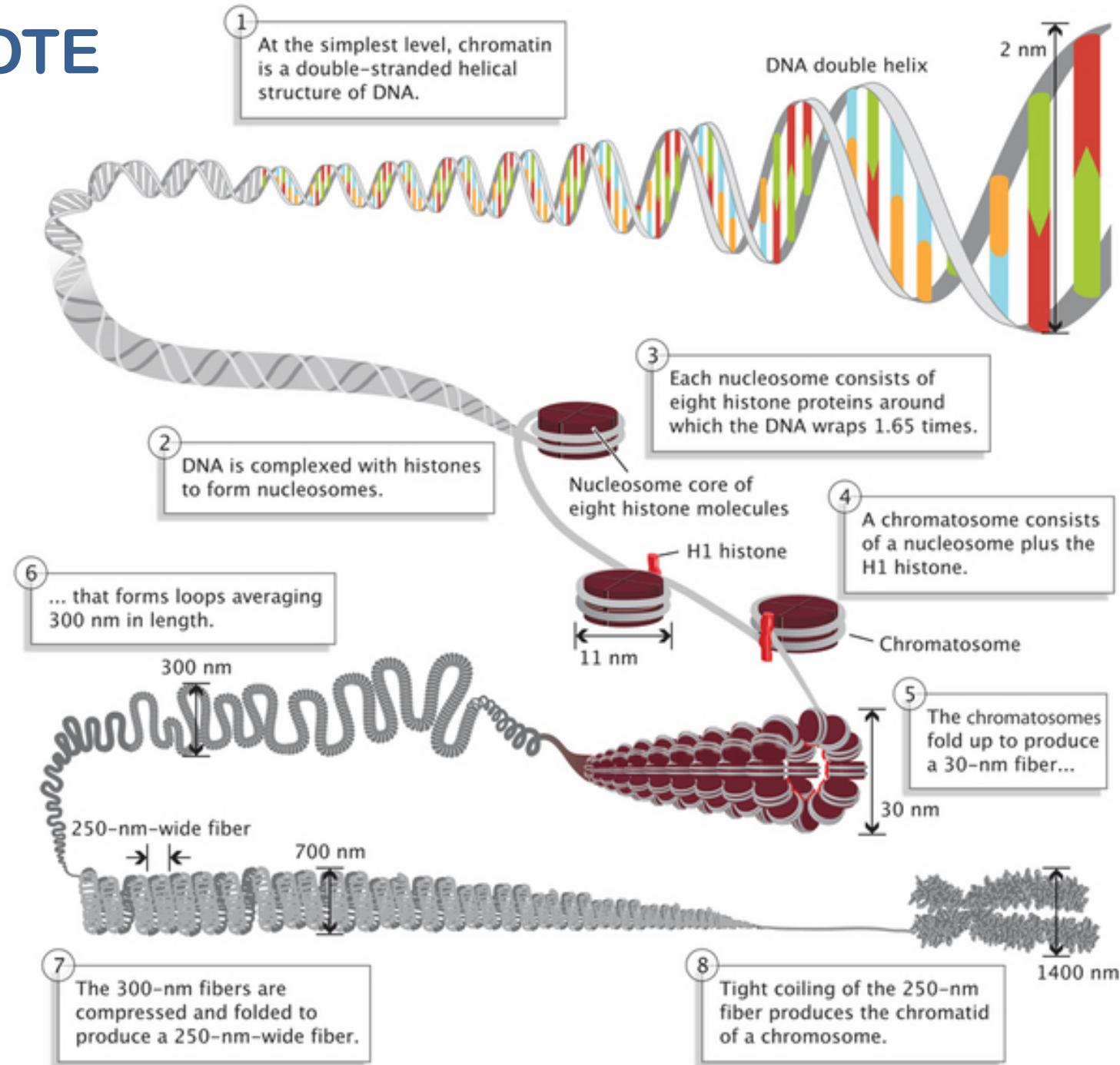
(b)



DNA consists of
two antiparallel
poly**nucleotide**
chains

-Sugar
-Phosphate
backbone

EUKARYOTE GENOME



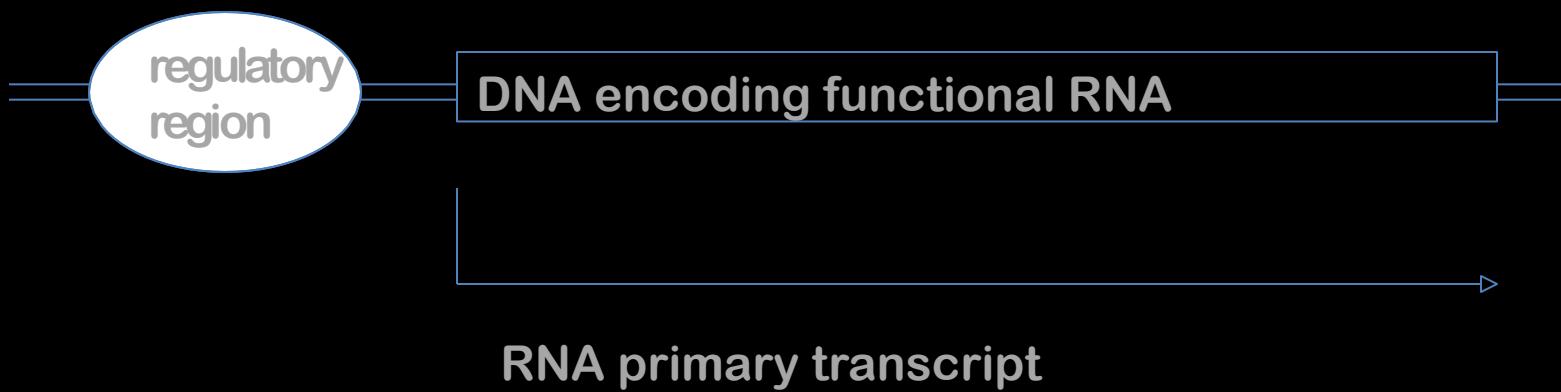
DNA features

- **Replication:** each strand serves as template for synthesis of complement, using rules of base pairing
- **Information:** specified by sequence of nucleotides; may be transcribed/copied into RNA
- **Mutation:** replacement, insertion, deletion of nucleotide results in altered sequence

The Structure of Genes

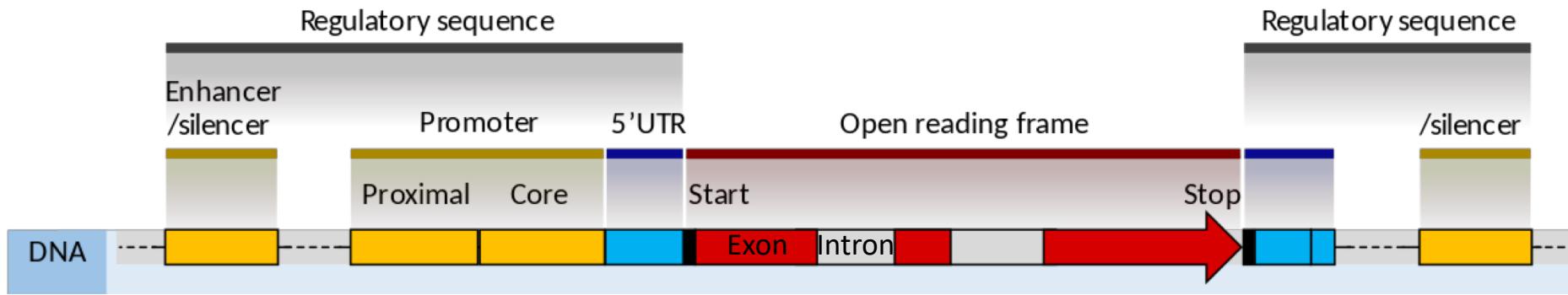
Structure of genes

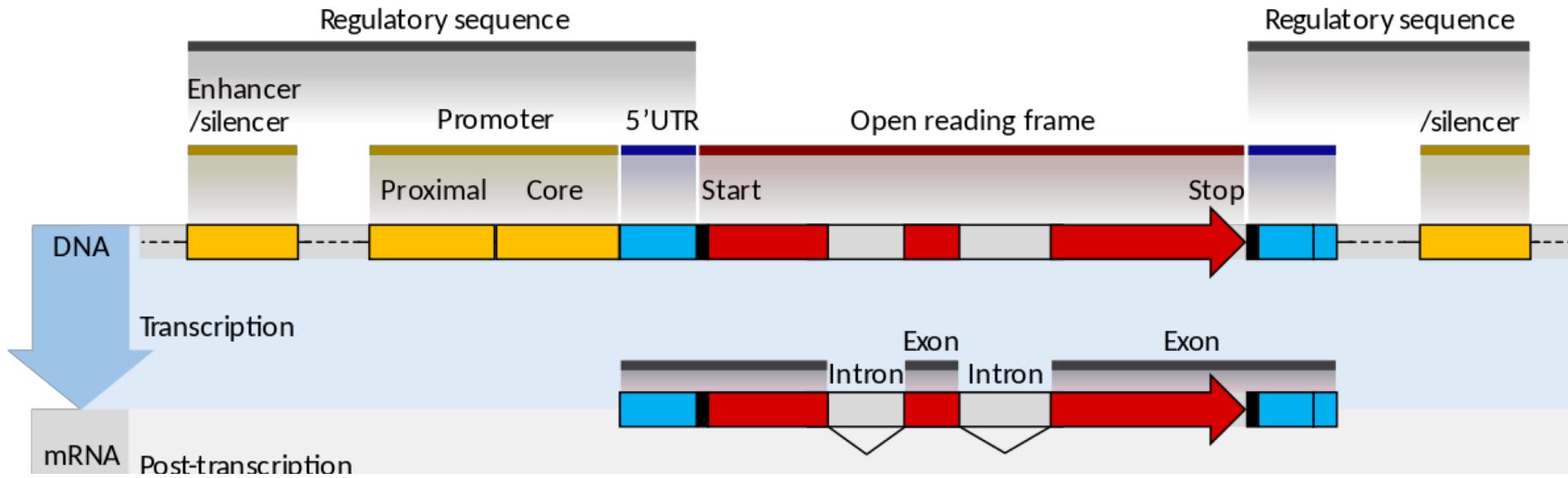
- Gene = transcriptional unit – functional element
- Gene encodes **coding RNA (mRNA)** or
non-coding RNA (tRNA, rRNA, miRNA)

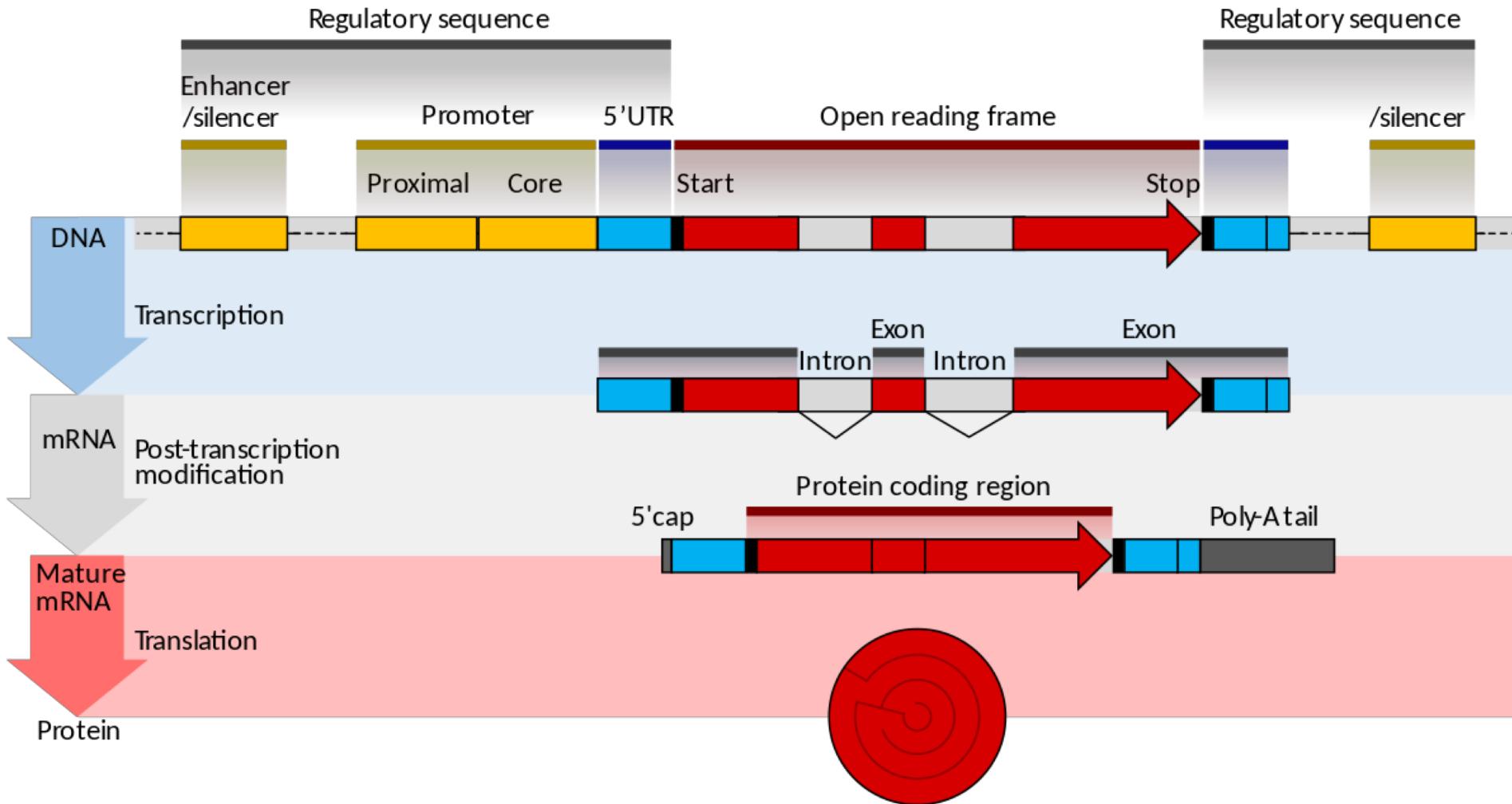


- To some researchers, gene actually includes its adjacent regulatory region(s)

Genes have exons and introns







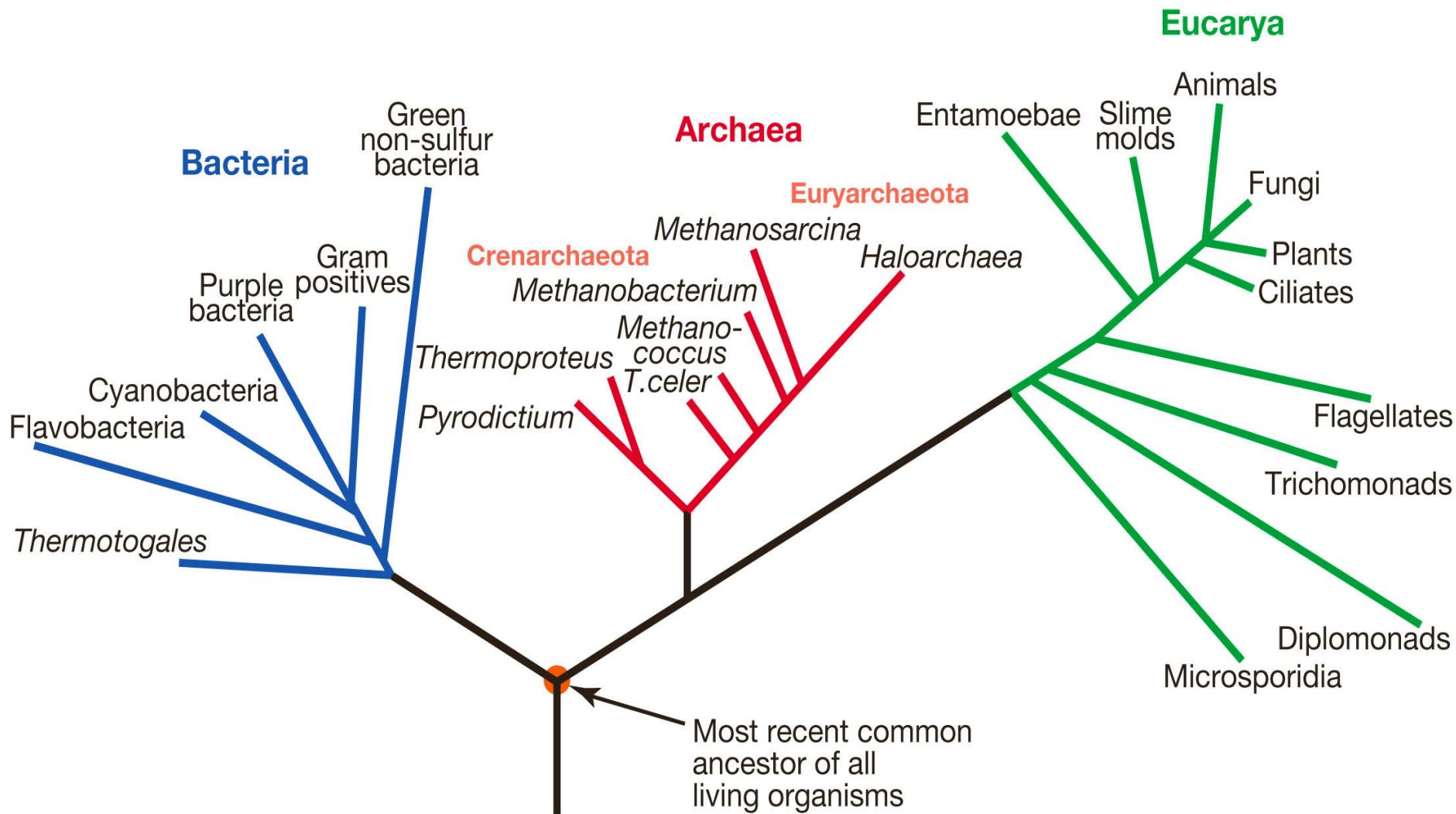
The Structure of Genomes

What is a genome?

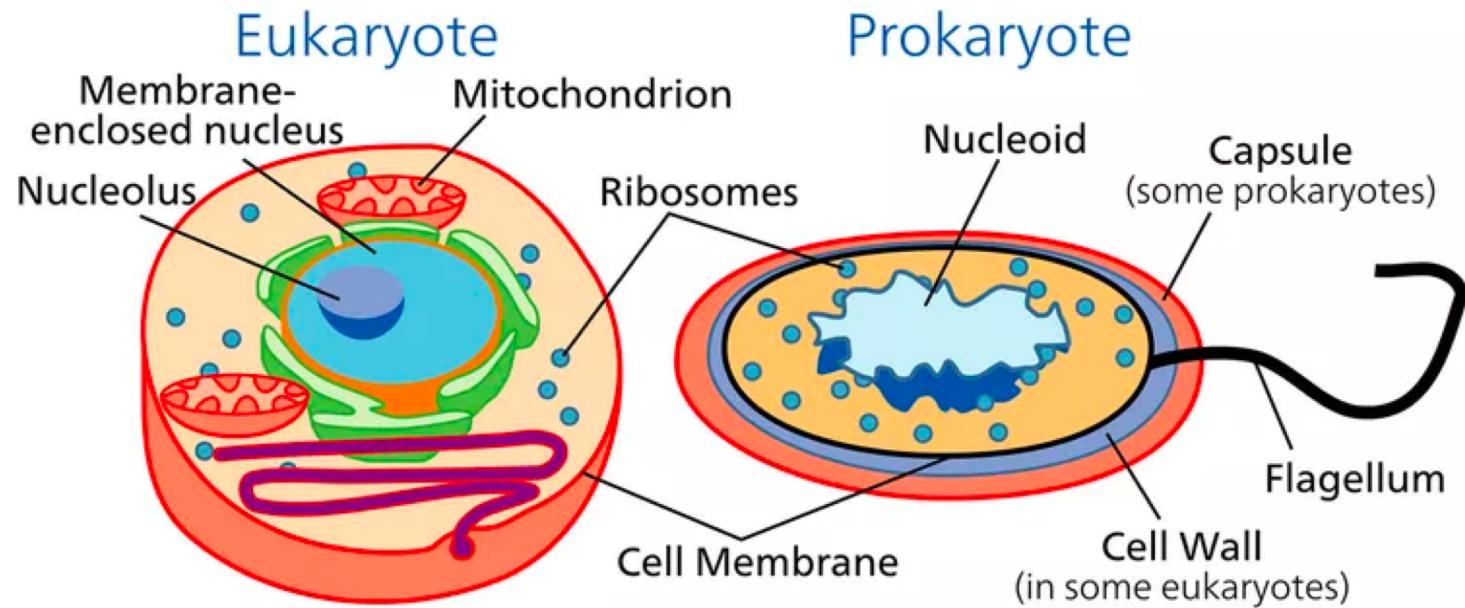
The genome represents the whole genetic information of an organism

It is physically formed by DNA, except in some viruses where it is constituted by RNA

The 3 major domains of life



EUKARYOTES vs PROKARYOTES



NUCLEAR MATERIAL

→ Euk – within membrane-bound nucleus

ORGANELLES

→ Euk have other membrane-bound organelles (mitochondria, chloroplasts)

CELL MEMBRANE

→ Euk have semi-permeable membrane with internal cytoskeleton

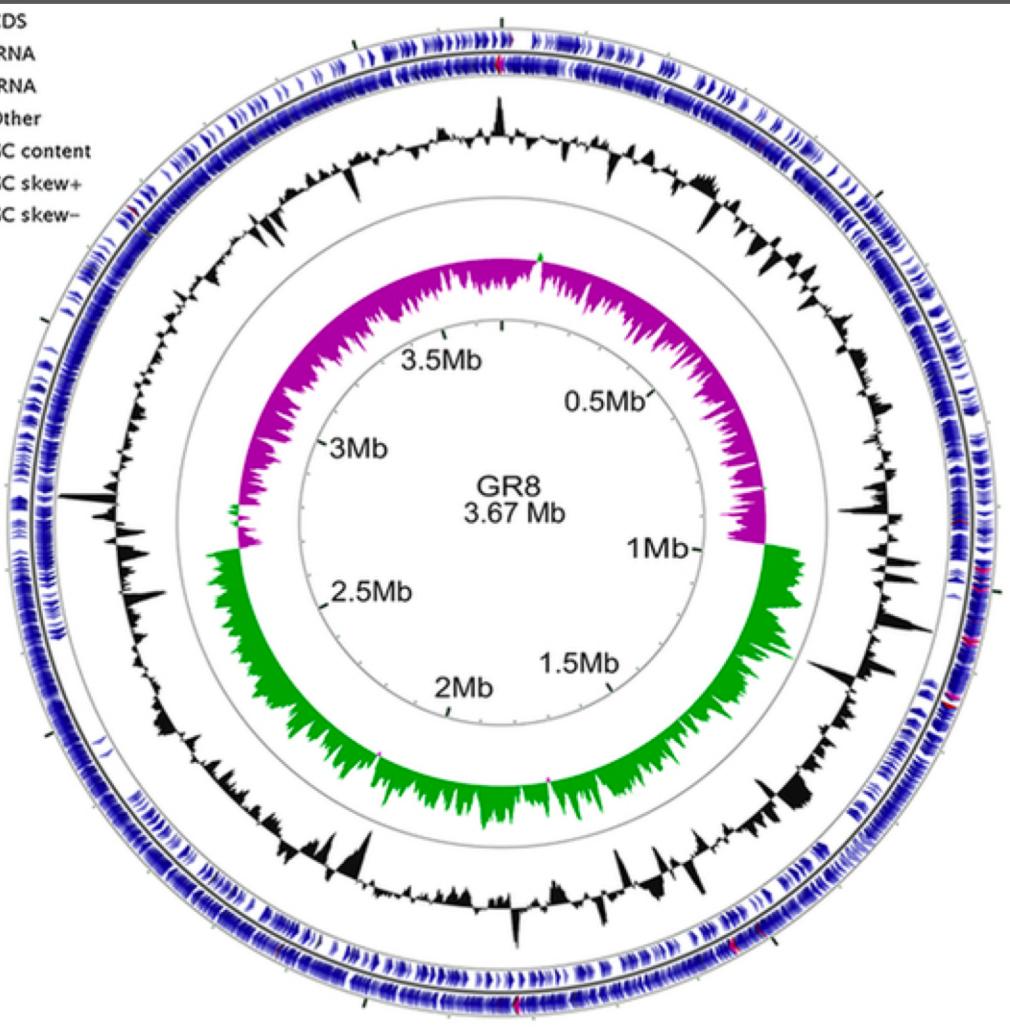
→ Prok have rigid plasma membrane

Very different genome landscape in prokaryotes and eukaryotes

- In **prokaryotes**, genes are compactly arranged, with little or no spacer sequences in between (short intergenic regions) = most of the genome is *coding* DNA
- In **eukaryotes**, there is considerable spacer DNA between genes (large intergenic regions) and within genes (introns) = most of the genome is '*non-coding*' DNA

Prokaryotic genome

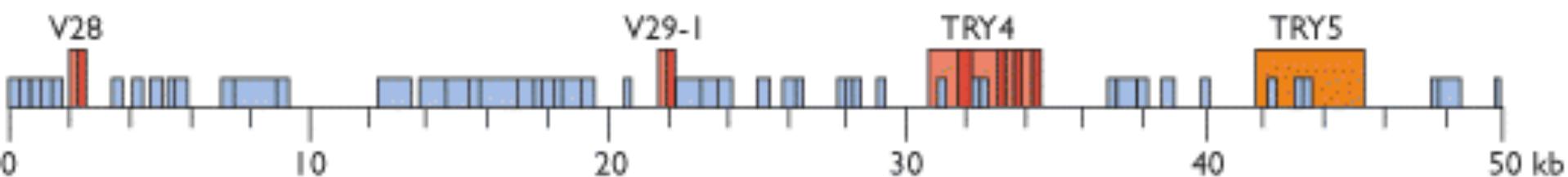
- Usually **circular**
- Gene--dense: genes are close together with little intergenic spacer (compact genome)
- **No introns**
- Shorter genes with low frequency of gene families
- Genes are organized in **operons**
 - tandem cluster of coordinately regulated genes
 - Several genes transcribed as single mRNA



PROKARIOTIC GENOME

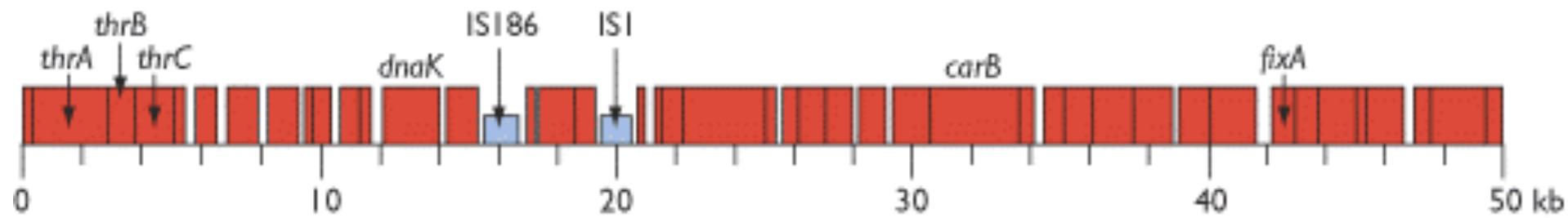


(A) Human

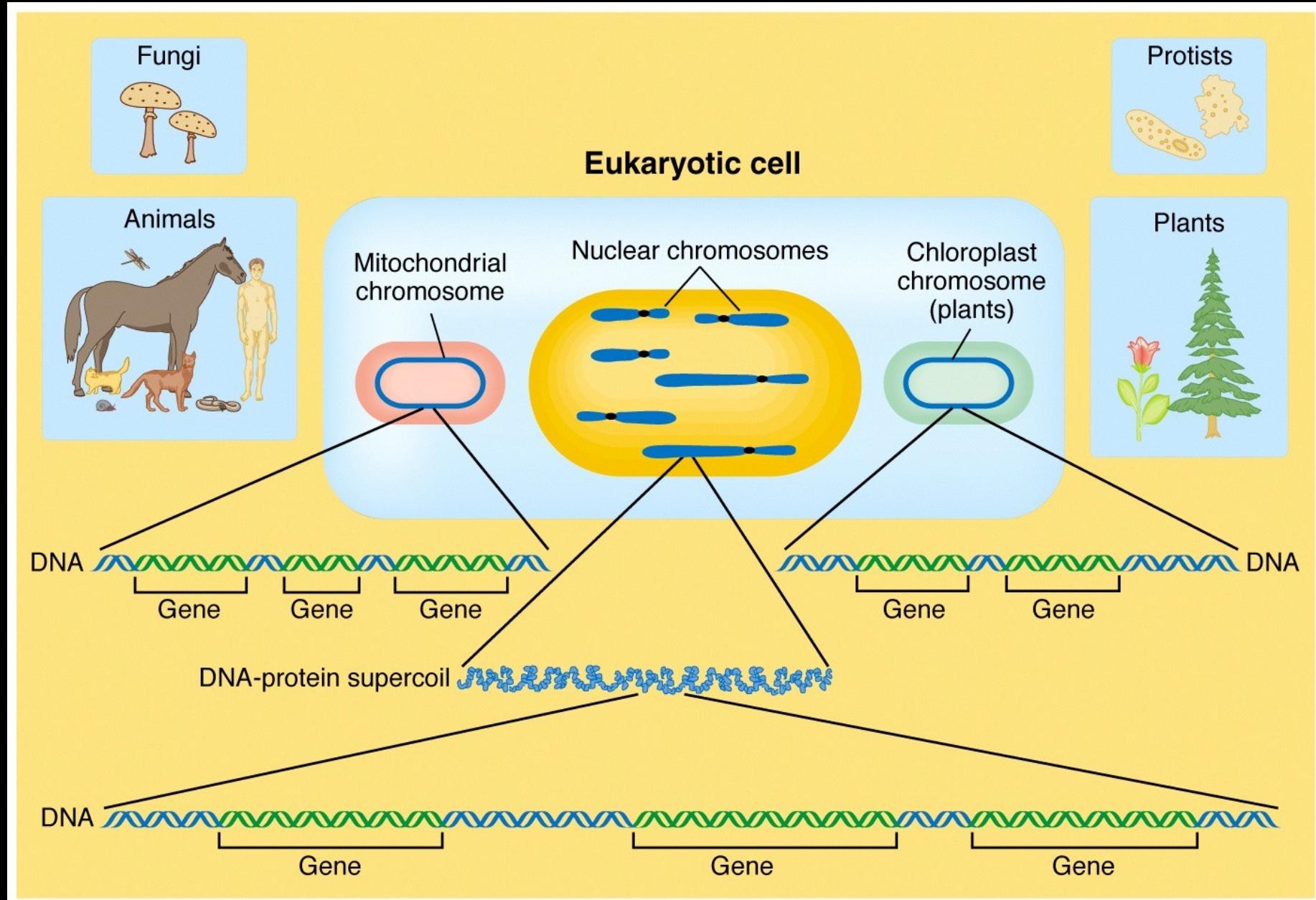


Escherichia coli – non-coding DNA only 11% of total genome

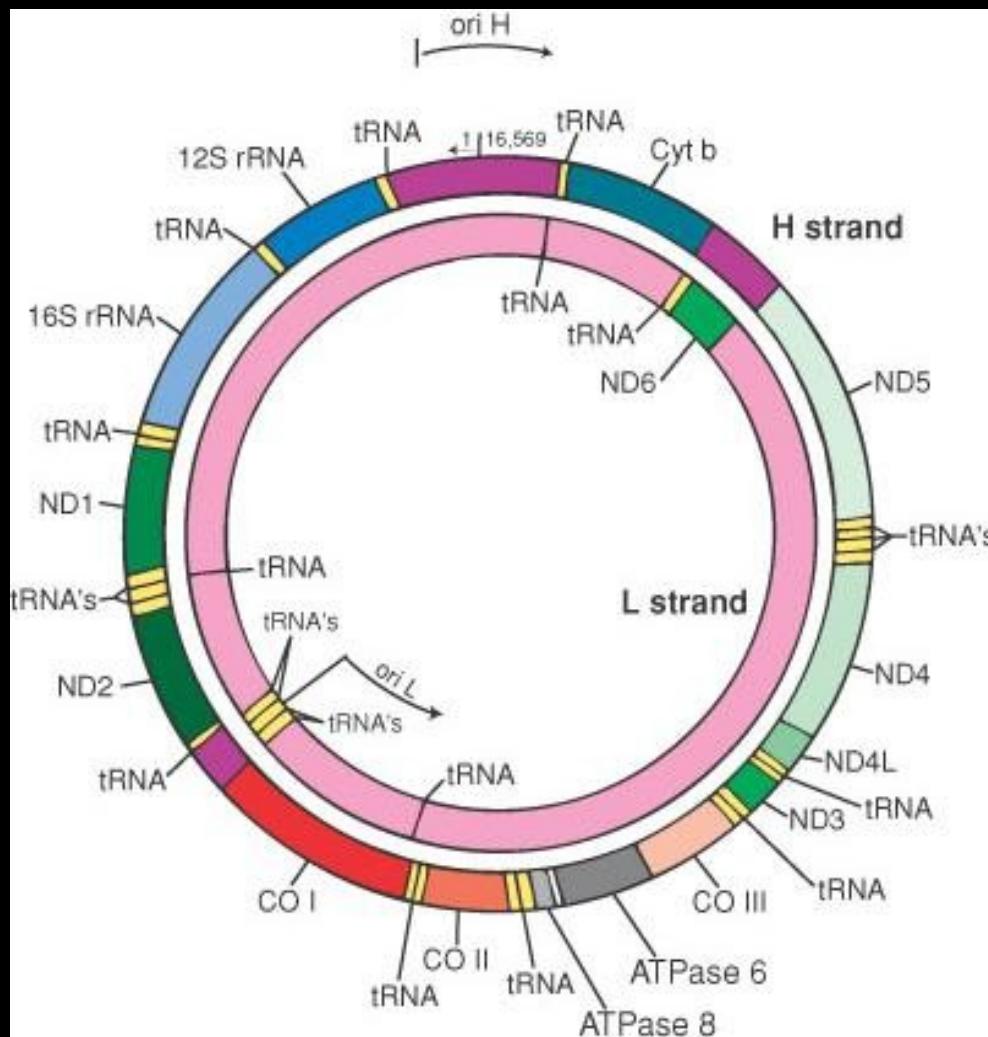
(E) Escherichia coli



Eukaryotes: 2 or 3 different genomes per cell

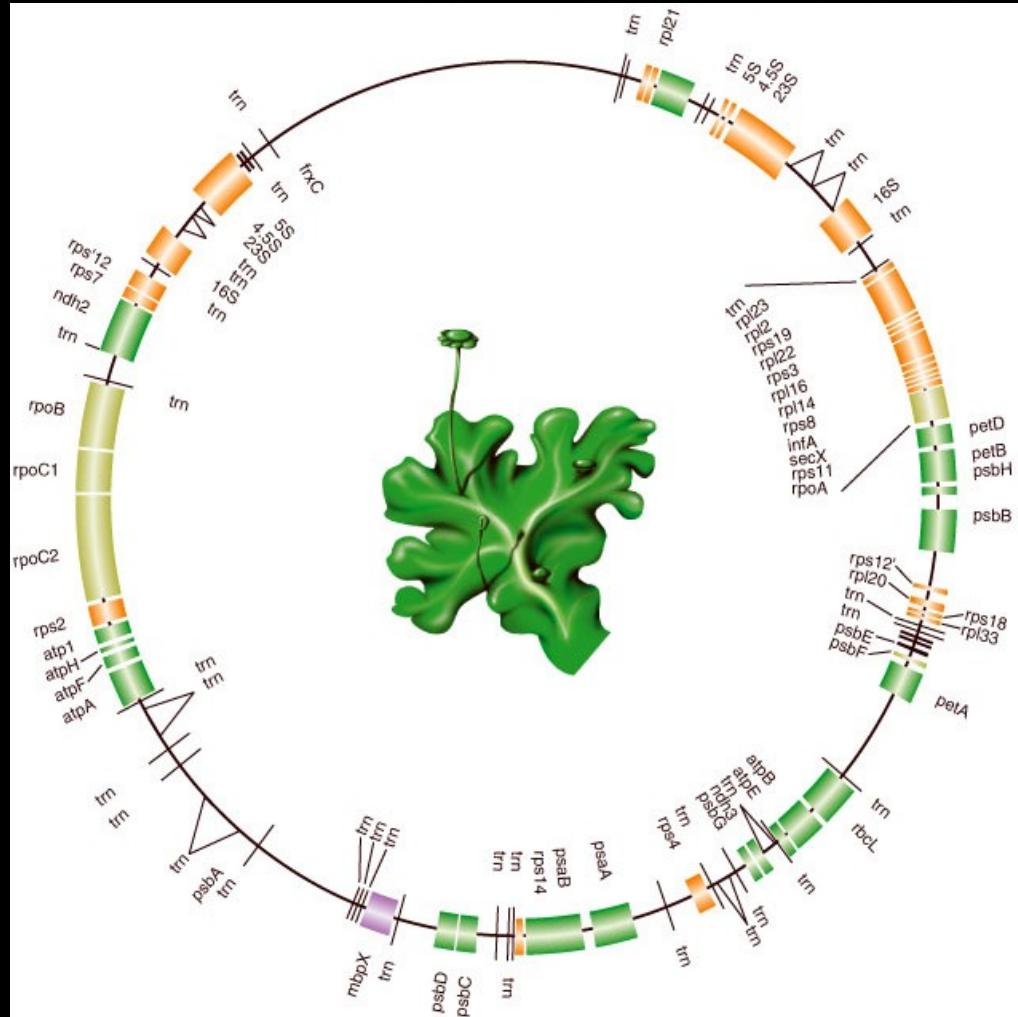


The Mitochondrial Genome



- Circular, 2--10 copies per mitochondrion
- Resembles a reduced prokaryotic genome in terms of organization and gene numbers
- The human mitochondrion genome has 37 genes: 13 protein-coding genes involved in the production of energy and 24 non-coding genes involved in mRNA translation (tRNAs, rRNA)
- Maternally inherited

The Chloroplast Genome (plants)



-- Circular

-- Resembles a reduced prokaryotic genome in terms of organization and gene numbers

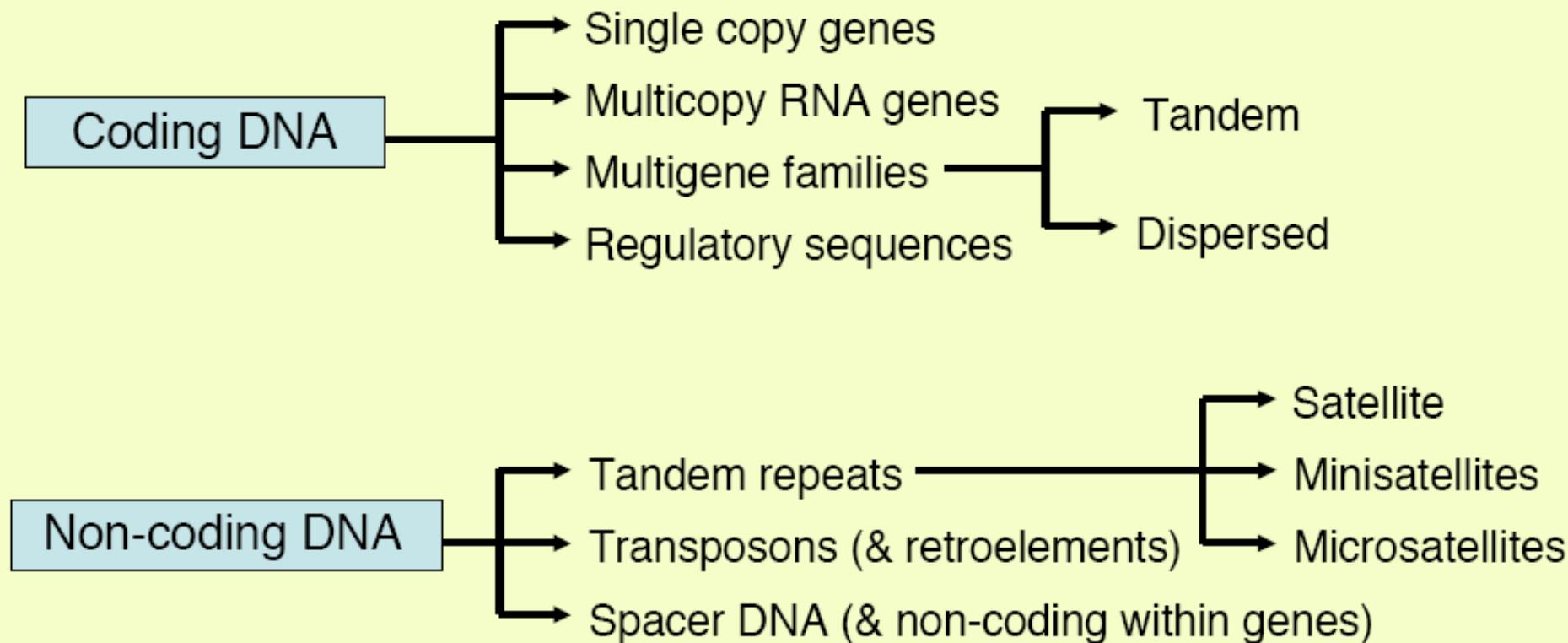
-- Encodes genes mostly involved in photosynthesis and electron transport

-- 'Maternally' inherited

Marchantia (moss) CpDNA 121 kb

Contents of a eukaryotic genome

(adapted from Page & Holmes, *Molecular Evolution*, 1998)



The non-coding DNA of eukaryotic nuclear genomes

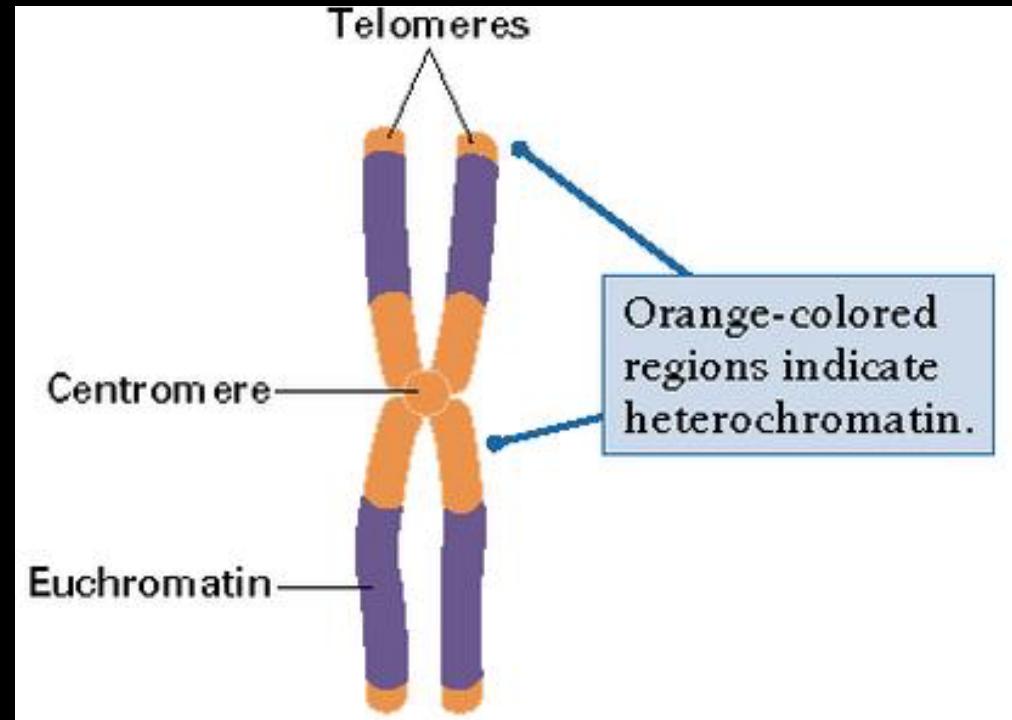
-Non-coding DNA:

introns

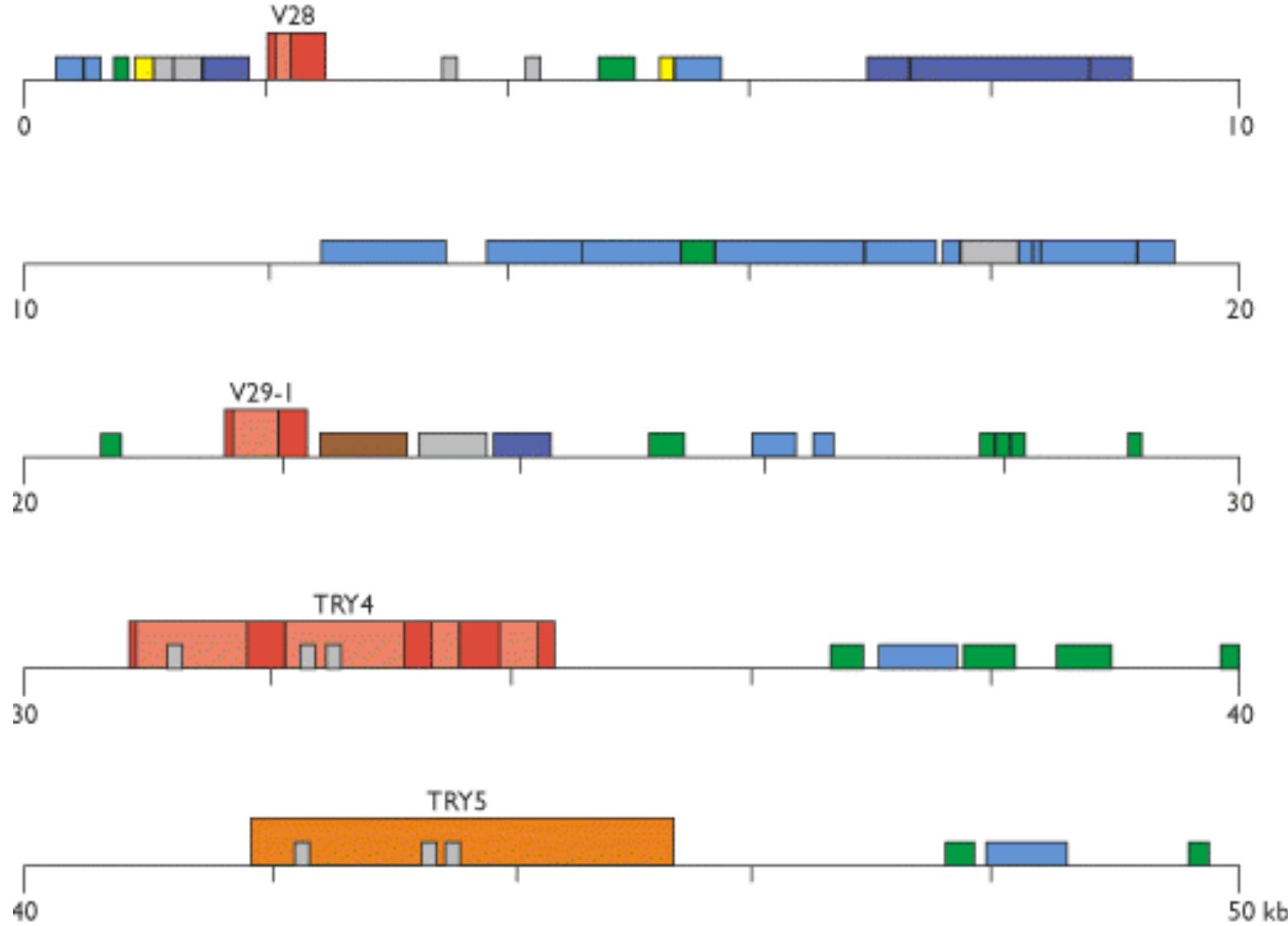
intergenic regions

centromeric regions

telomeric regions



- Most non-coding DNA is formed by identical or nearly identical repeated units (repetitive DNA)
- Two types of repetitive DNA:
 - Tandem repeats (e.g. satellite DNA at centromeres and telomeric repeats at telomeres)
 - Interspersed repeats (mostly transposable elements)

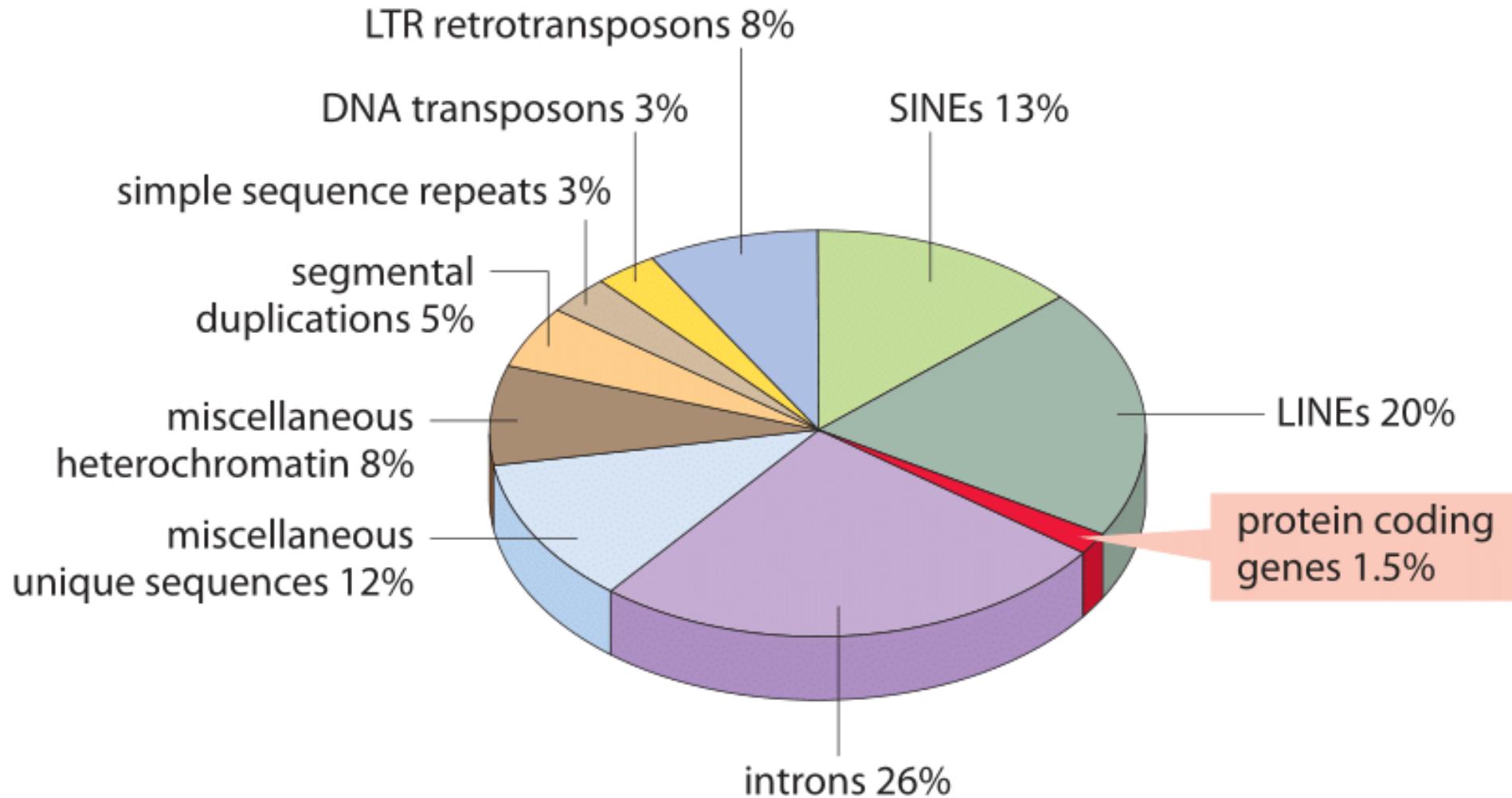


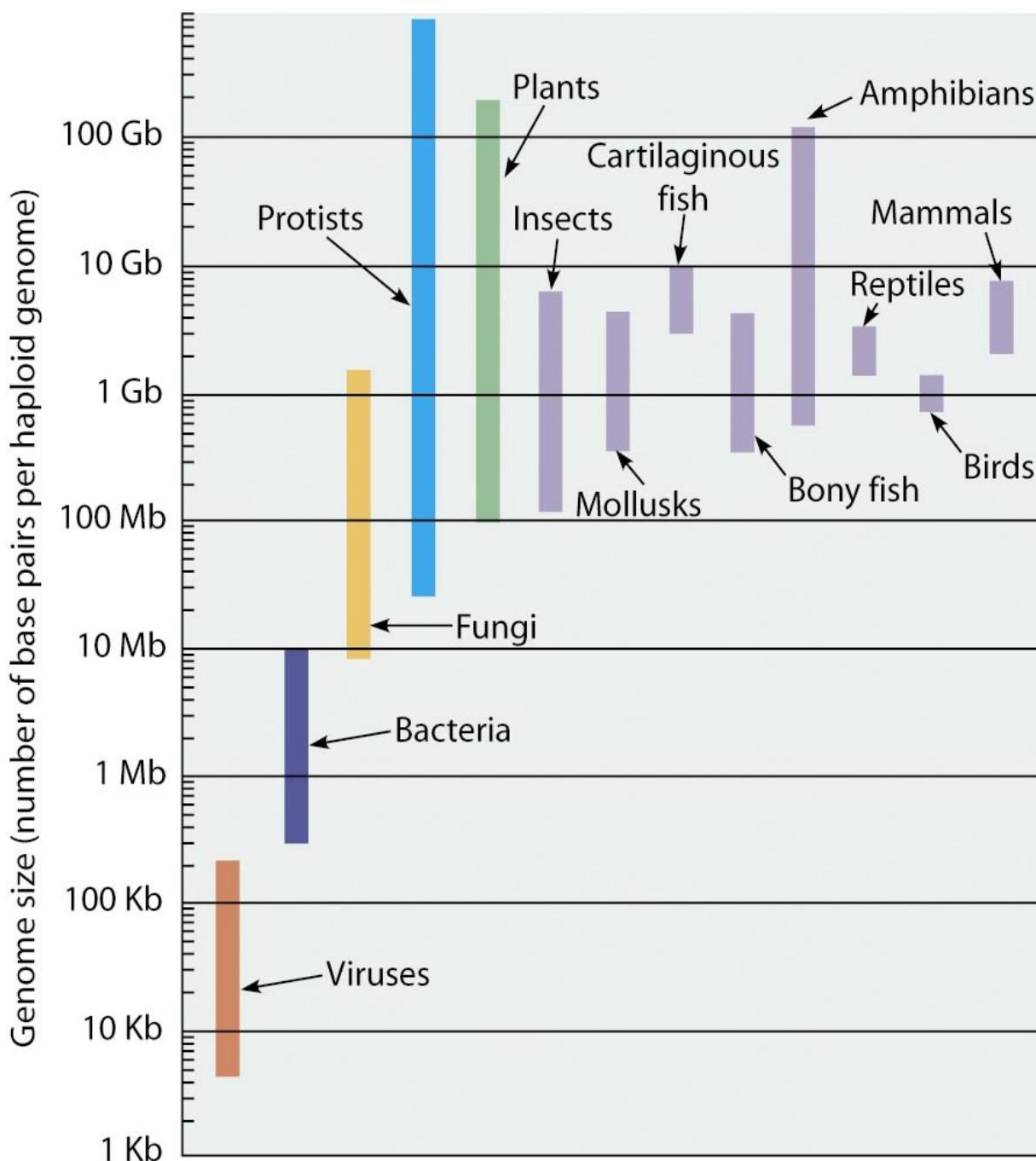
50-kb segment of the human b T-cell receptor locus on chromosome 7

from T.A. Brown, *Genomes 2nd Ed.*

KEY	
Exon	LINE
Intron	SINE
Pseudogene	DNA transposon
	Other genome-wide repeat
	LTR element
	Microsatellite

In Eukaryotic genomes VERY small part contains genes





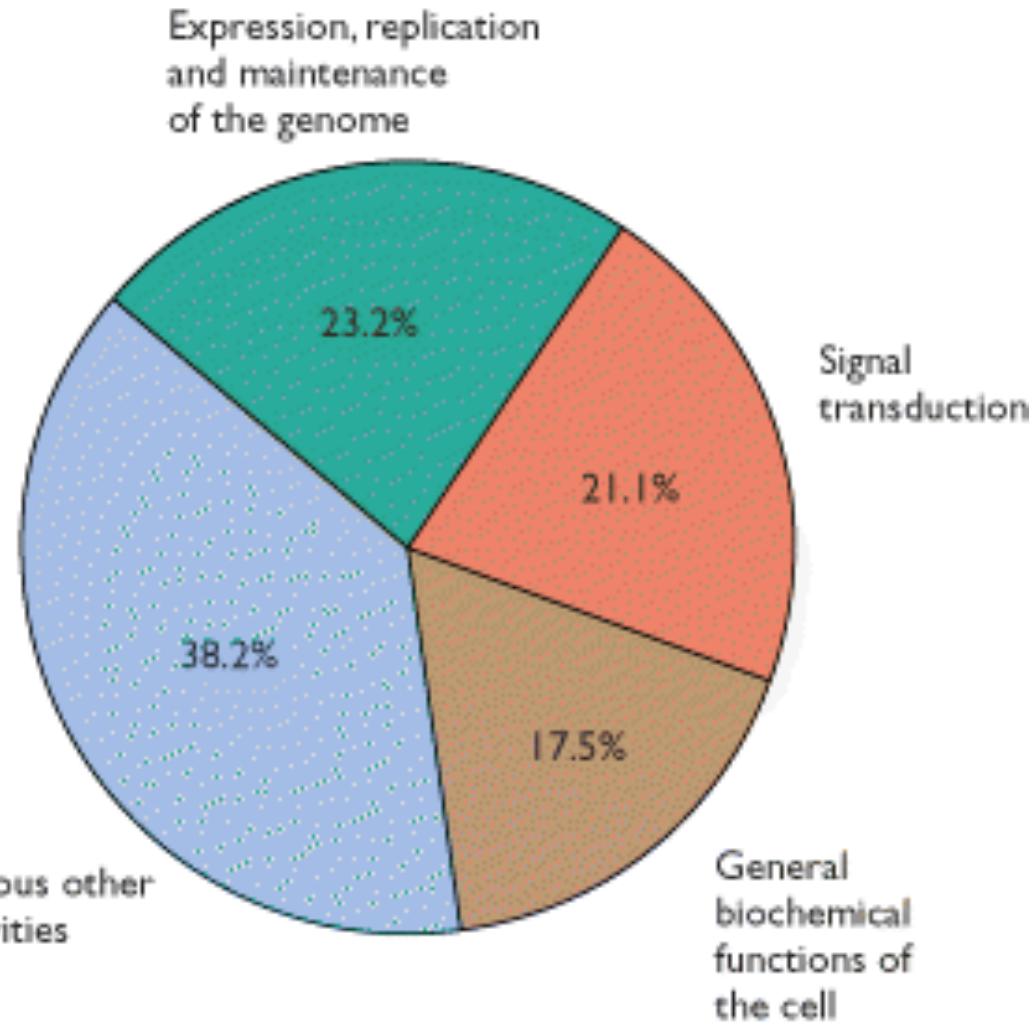
Gene number and genome size

Genome size comparison

Species	Chromosomes	Genes	Base pairs
 Human <i>(Homo sapiens)</i>	46 (23 pairs)	21,000	3.1 billion
 Mouse <i>(Mus musculus)</i>	40	21,000	2.7 billion
 Puffer fish <i>(Fugu rubripes)</i>	44	28,000	365 million
 Malaria mosquito <i>(Anopheles gambiae)</i>	6	14,000	289 million
 Fruit fly <i>(Drosophila melanogaster)</i>	8	14,000	137 million
 Roundworm <i>(C. elegans)</i>	12	19,000	97 million
 Bacterium * <i>(E. coli)</i>	1	5,000	4.1 million

*Bacterial chromosomes are chromonemes, not true chromosomes

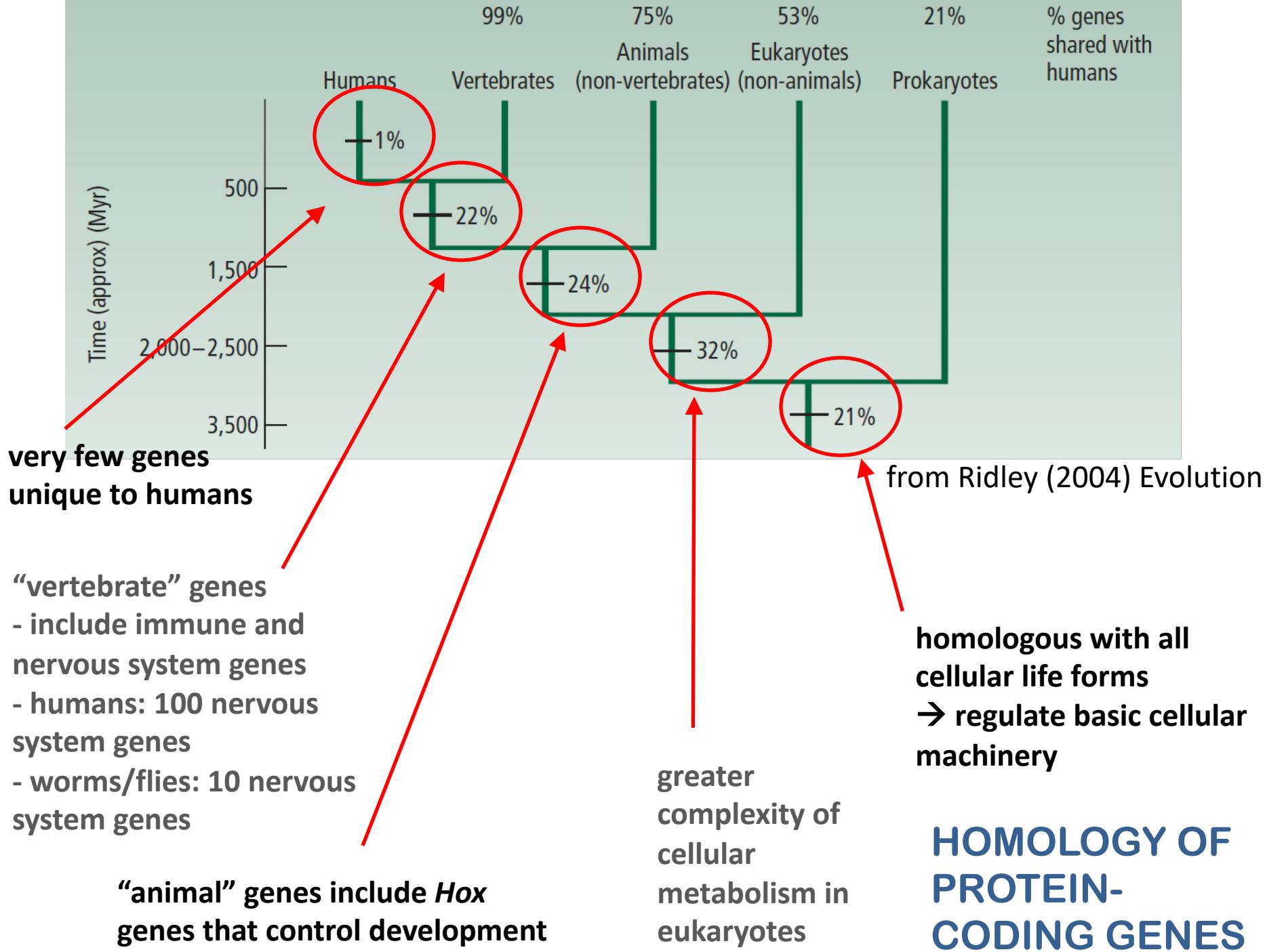
GENE CONTENT EUkARYOTIC GENOME



Categorization of the identified human protein-coding genes.

- omits ~ 10 000 genes whose functions are not yet known.
- segment labeled 'various other activities' includes, among others, proteins involved in biochemical transport processes and protein folding, immunological proteins, and structural proteins.

Based on *Figure 15* of Venter *et al.* (2001).



GENOME EVOLUTION

TWO ways in which NEW GENES could be acquired by a genome

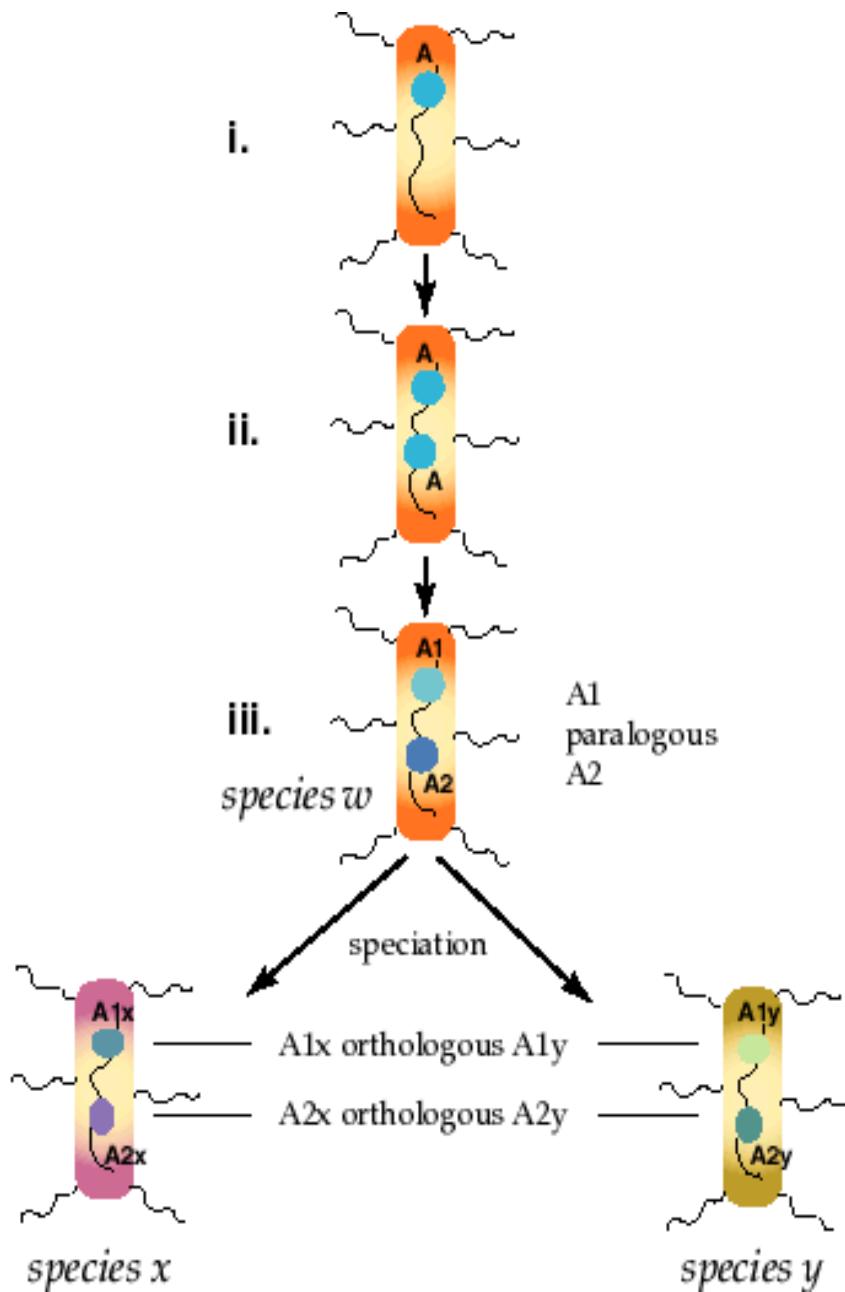
1. By duplicating existing genes in the genome.

- whole genome duplication
- duplication of single or part of chromosome
- duplication of single gene or group of genes

2. By acquiring genes from other species

- lateral gene transfer
- retroviruses and transposable elements

Gene Duplication: Orthologs and Paralogs...

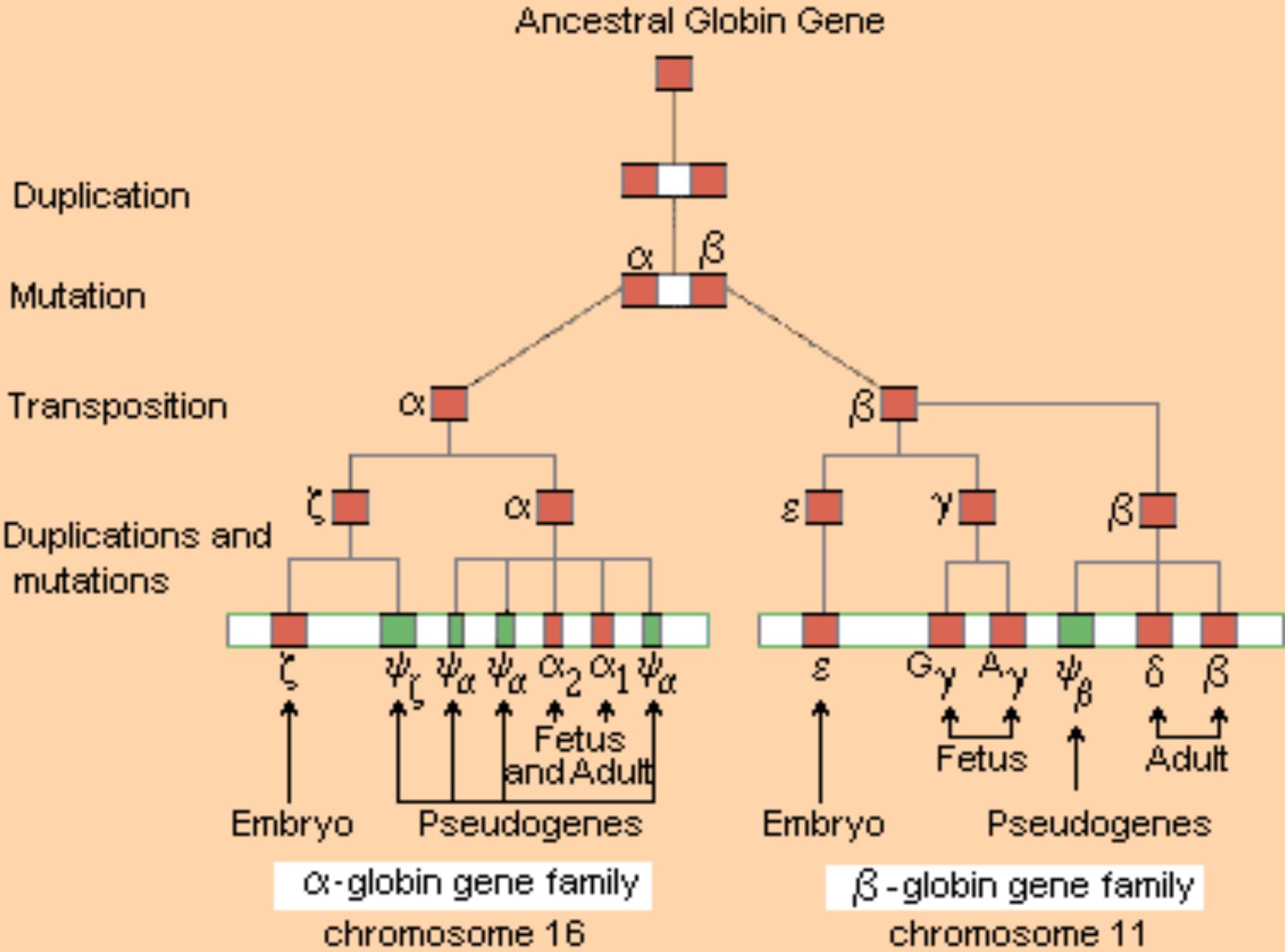


Two genes are **paralogous** if derived from a **duplication** event, **orthologous** if derived from a **speciation** event.

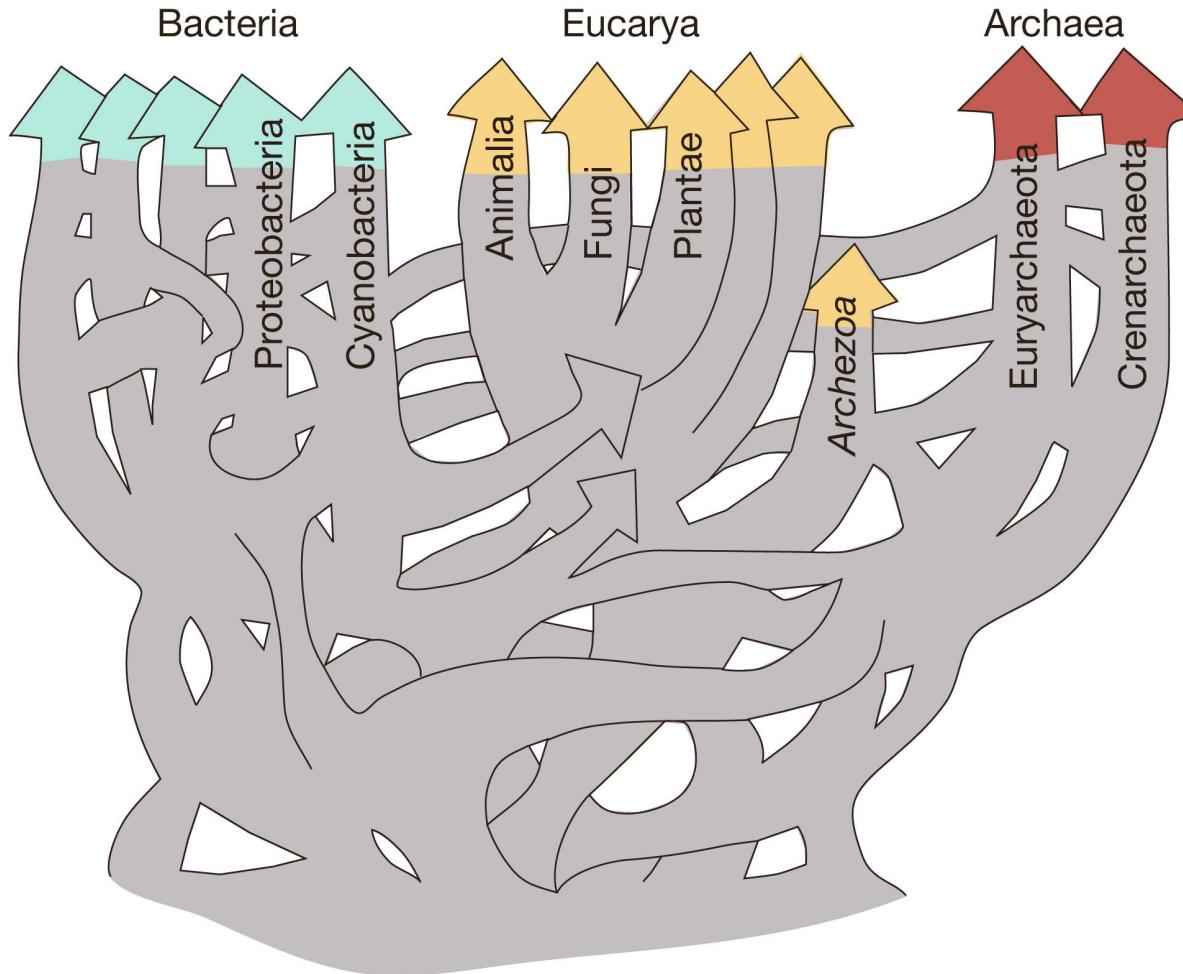
- Gene "A" in species "w"
- is duplicated producing initially two copies of A.
- With time the 2 copies diverge forming related genes A1 and A2. These two genes are **paralogous** to one another. **Paralogy** involves comparisons **within a species**.
- 2 species, x and y evolve from species w (common ancestor). Descendants of the A1 and A2 genes are now called A1x, A1y, and A2x, A2y to reflect which species they now occupy. A1x is **orthologous** to A1y and A2x is **orthologous** to A2y. **The comparison is between two species**. A1x and A2y are **paralogously** related as are A2x and A1y.

DUPLICATION FOLLOWED BY SUB-FUNCTION PARTITIONING...

Evolutionary Time ↓



Horizontal gene transfer rampant



common ancestor not a single species, but community of species swapping genes

**“TREE OF LIFE” concept is overly simplistic
“WEB OF LIFE” more appropriate?**

Next class

Population genomics, population
structure and demography

Schraiber_Akey 2015