

Molecular

Workshop on

Walter Salzburger

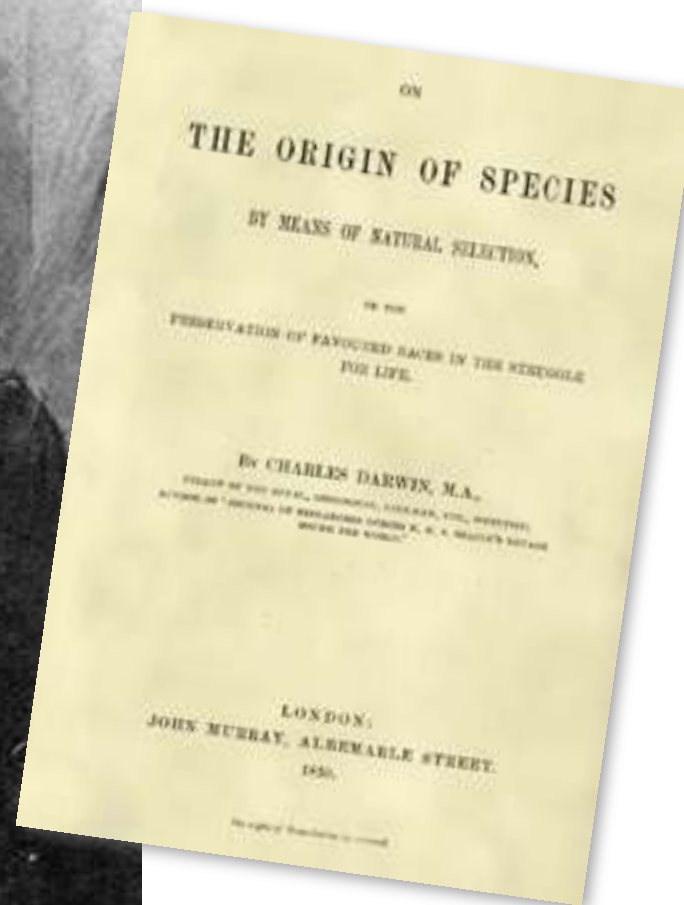
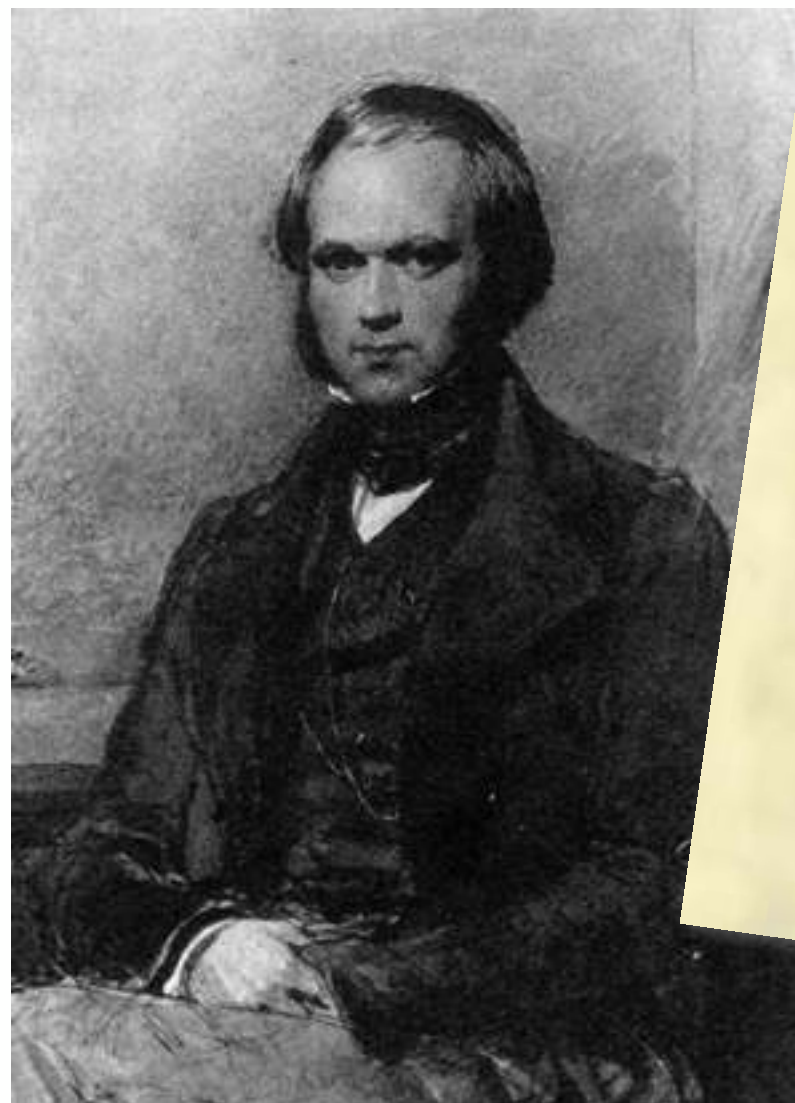
Zoological Institute
University of Basel, Switzerland



Molecular evolution deals with the process of evolution at the scale of DNA, RNA and proteins

1859

Charles R. Darwin publishes “*On the origin of species by means of natural selection*” and establishes the theory of evolution



Charles R. Darwin (1809-1882)

1866

Gregor Mendel publishes “*Experiments in plant hybridization*”. This paper established what eventually became formalized as the Mendelian laws of inheritance.



Gregor Mendel
(1822-1884)



1866

Gregor Mendel publishes “*Experiments in plant hybridization*”. This paper established what eventually became formalized as the Mendelian laws of inheritance.



1869

Johann Friedrich Miescher extracts what comes to be known as DNA from the nuclei of white blood cells.



Johann F. Miescher (1844-1895)

1900

Independently of one another, **Hugo de Vries** (1848-1935), **Erich von Tschermak-Seysenegg** (1871-1962) and **Carl Correns** (1864-1933) rediscover Mendel's published, but long neglected, paper outlining the basic laws of inheritance.



Hugo de Vries



Erich v. Tschermack



Carl Correns

1902

Theodor Boveri and **Walter Sutton** propose that chromosomes bear hereditary factors in accordance with Mendelian laws.



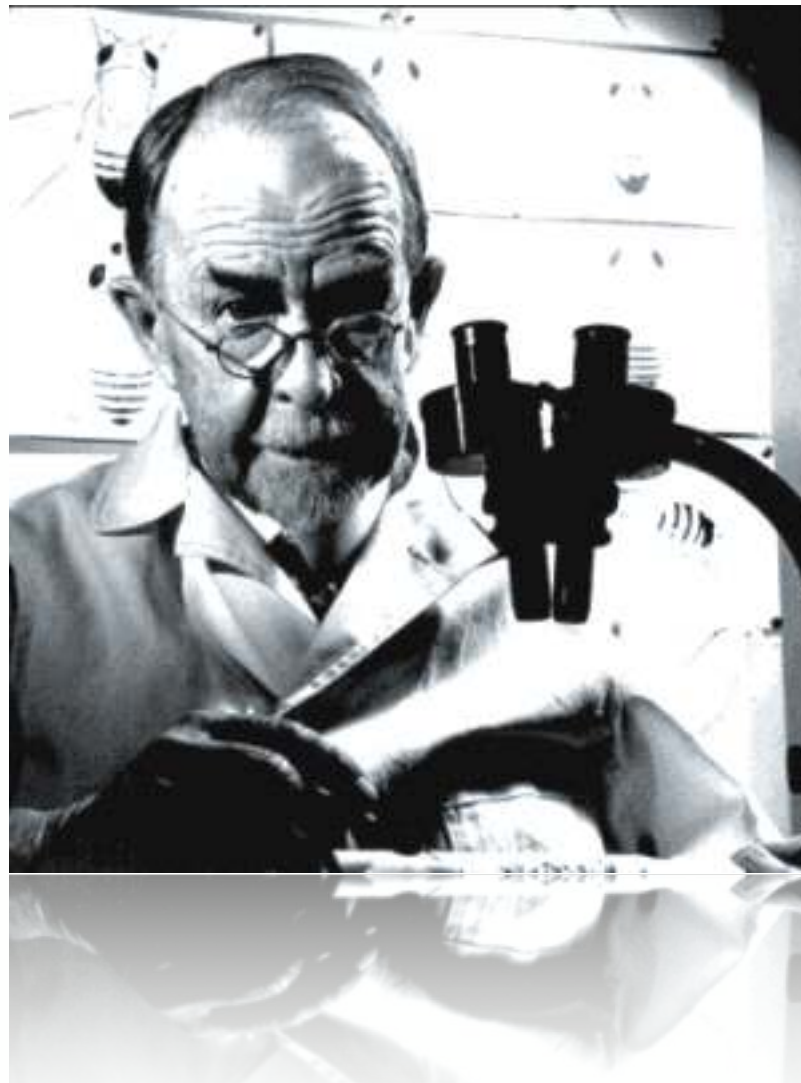
Walter Sutton
(1877-1916)



Theodor Boveri
(1862-1915)

1910

Thomas H. Morgan establishes the chromosomal theory of inheritance. He also discovered the recombination of homologous chromosomes during meiosis.



Thomas Hunt Morgan (1866-1945)

1944

Oswald T. Avery (1877-1955), **Maclyn McCarty** (1911-2005) and **Colin MacLeod** (1909-1972) identify deoxyribonucleic acid (DNA) as the “transforming principle”.



Oswald T. Avery
(1877-1955)

1950

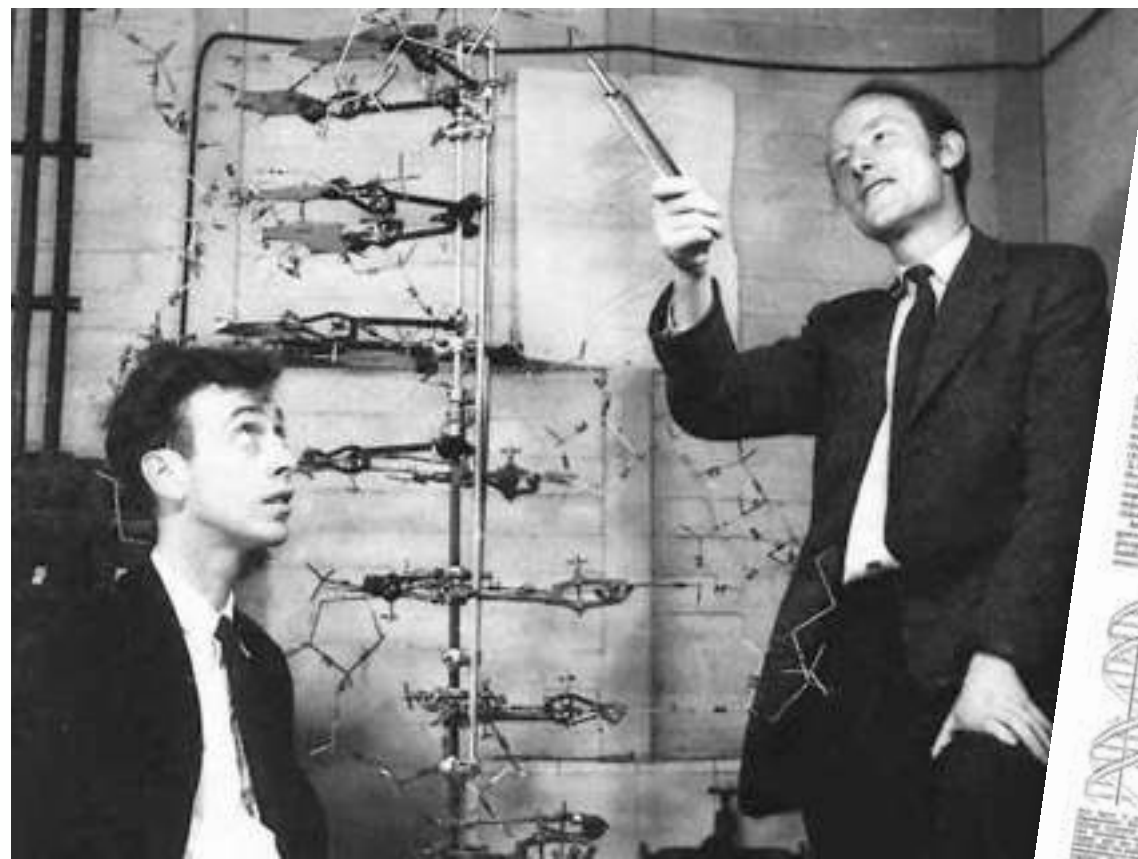
Erwin Chargaff discovers regularity in proportions of DNA bases. In all organisms he studied, the amount of adenine (A) equaled that of thymine (T), and guanine (G) equaled cytosine (C).



Erwin Chargaff (1905-2002)

1953

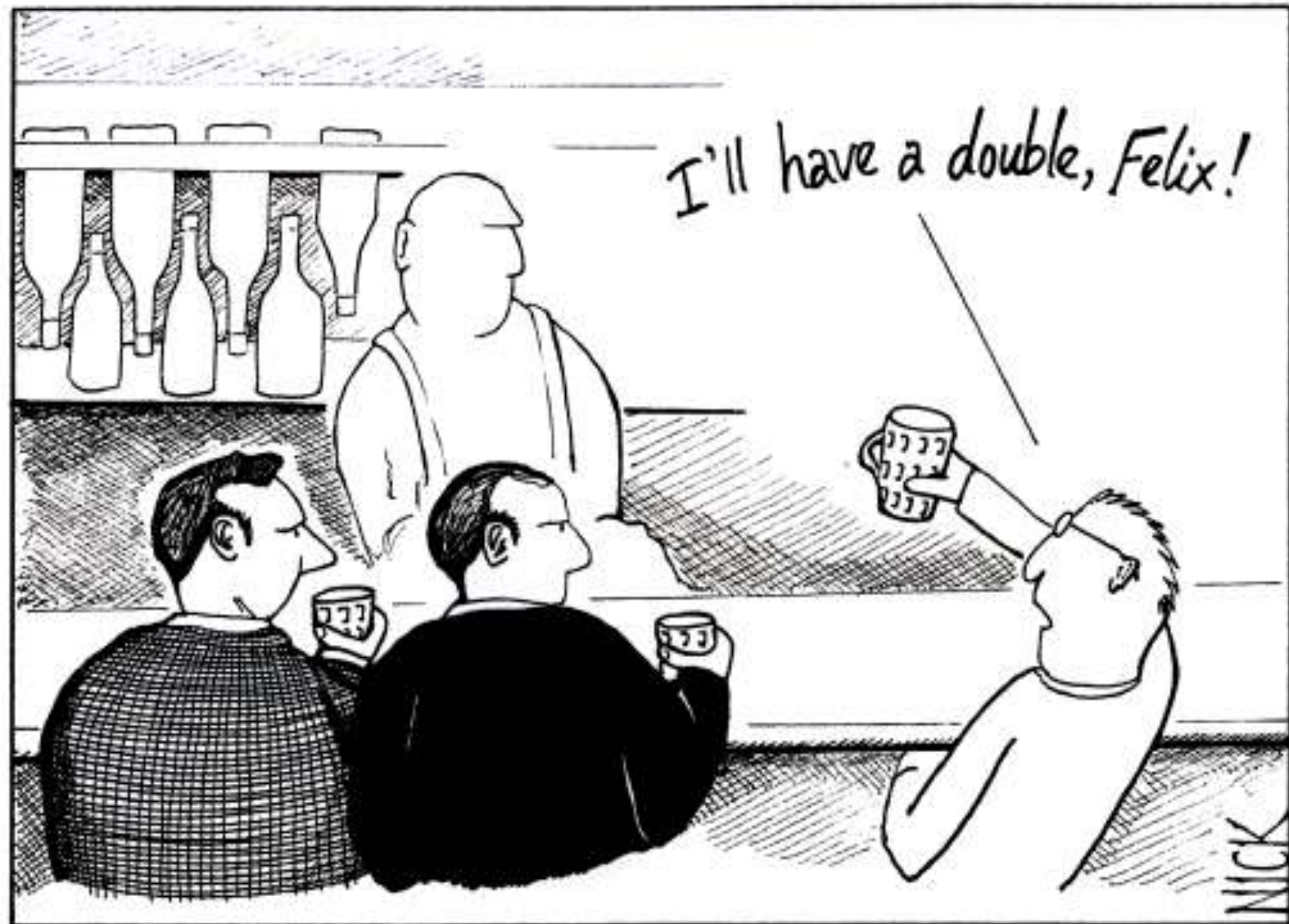
James Watson and Francis Crick discover the double helical structure of the DNA and that this structure meets the unique requirements for a substance that encodes genetic information.



James D. Watson (1928-)
Francis H. C. Crick (1916-2004)



1953



Cambridge, 1953. Shortly before discovering the structure of DNA, Watson and Crick, depressed by their lack of progress, visit the local pub.

Watson and Crick, depressed by their lack of progress, visit the local pub.
Cambridge, 1953. Shortly before discovering the structure of DNA, Watson

1977

Frederick Sanger (1918-) and **Walter Gilbert (1932)**
develop techniques for DNA sequencing

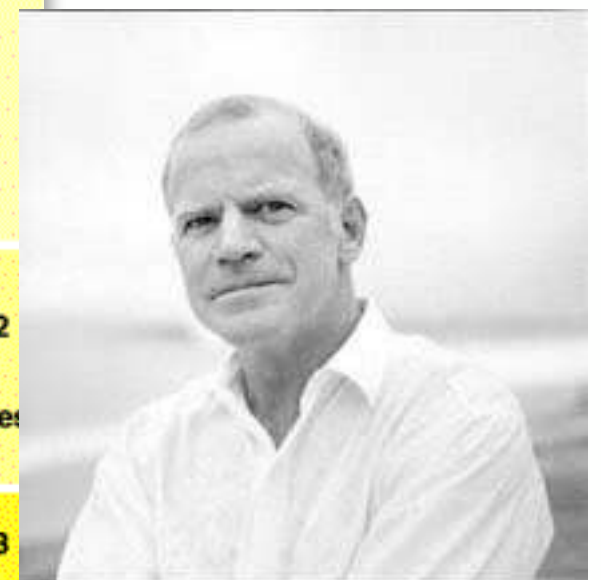
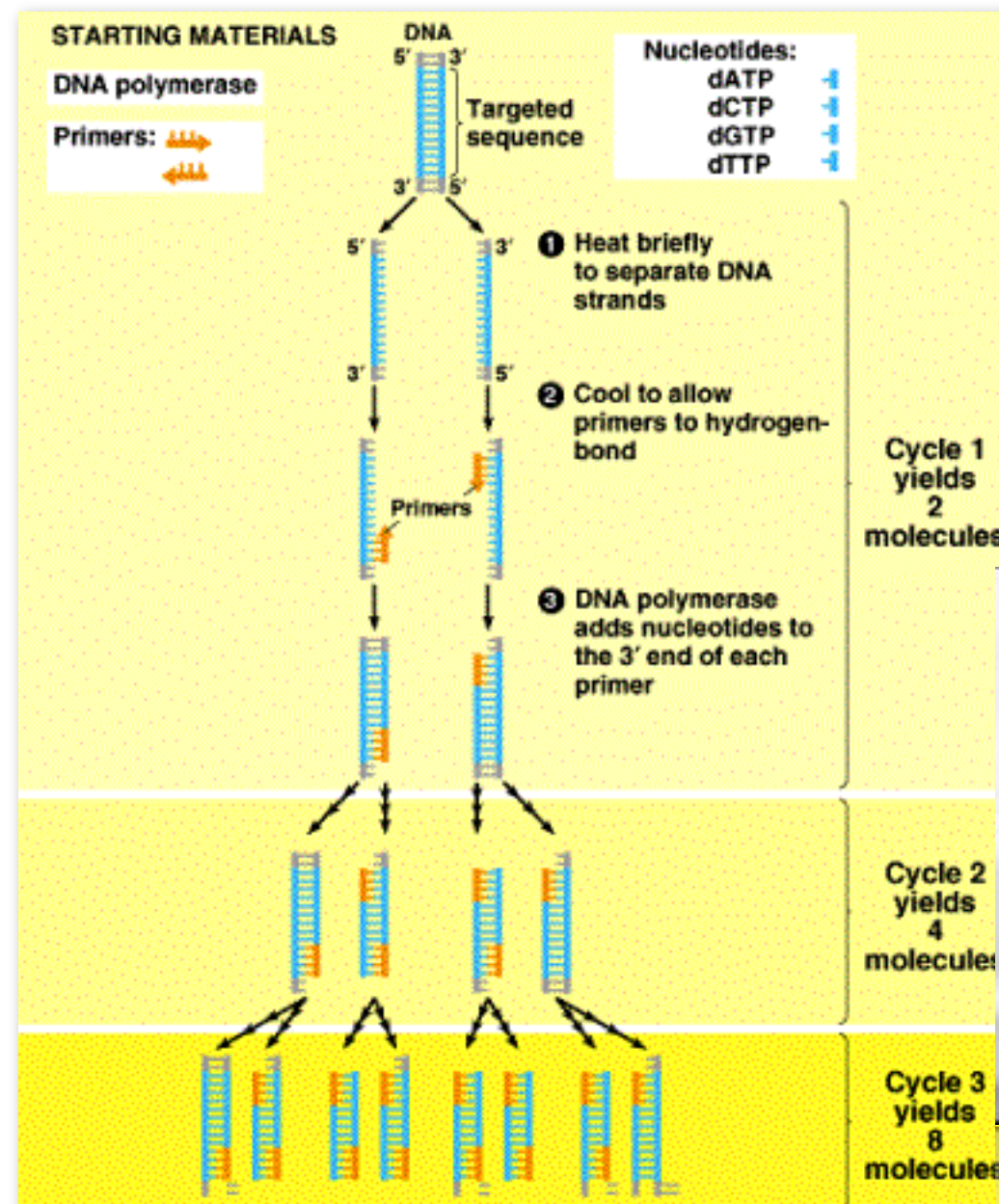


Walter Gilbert

Frederick Sanger

1983

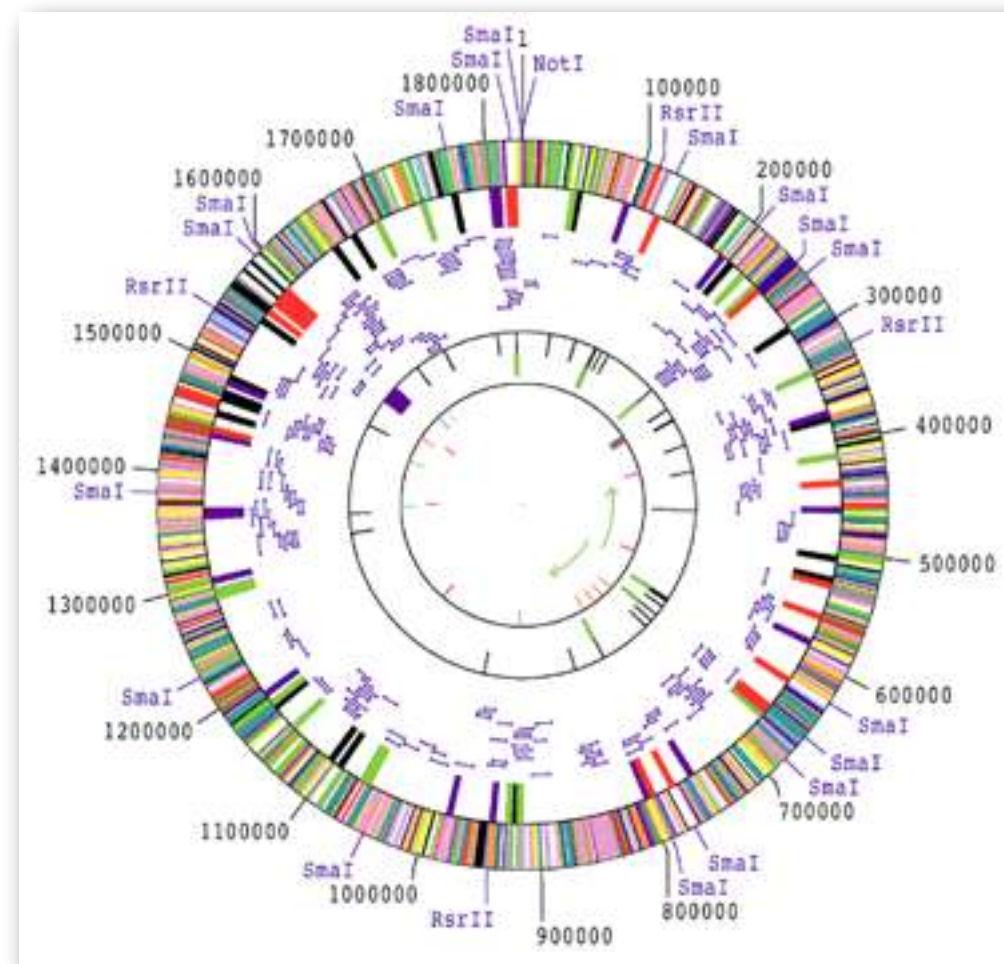
Kary B. Mullis (1944-) invents and helps to develop the polymerase chain reaction (PCR)



Kary B. Mullis

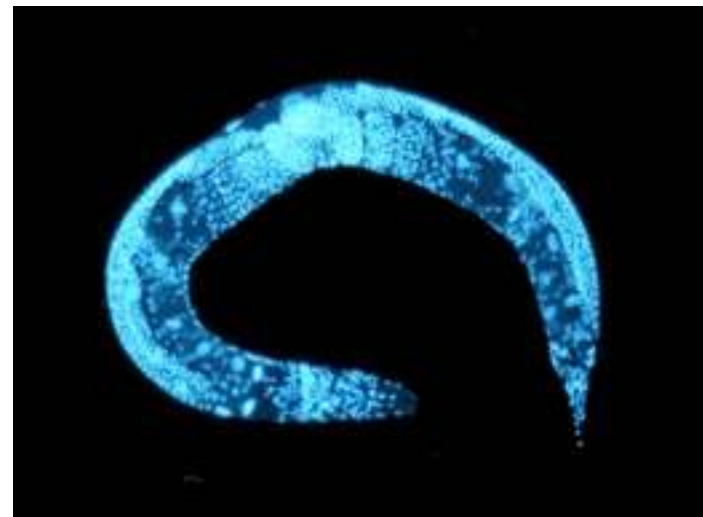
1995

1,830,137 bp of *Hamophilus influenzae* sequenced:
the first genome of a free living organisms
determined



1998

Caenorhabditis elegans sequenced



2000

Drosophila melanogaster sequenced



ncbi.nlm.nih.gov

NCBI Resources How To Sign in to NCBI

Genome Search

Genome Information by organism

Search by organism Clear

[Download Reports from FTP site](#)

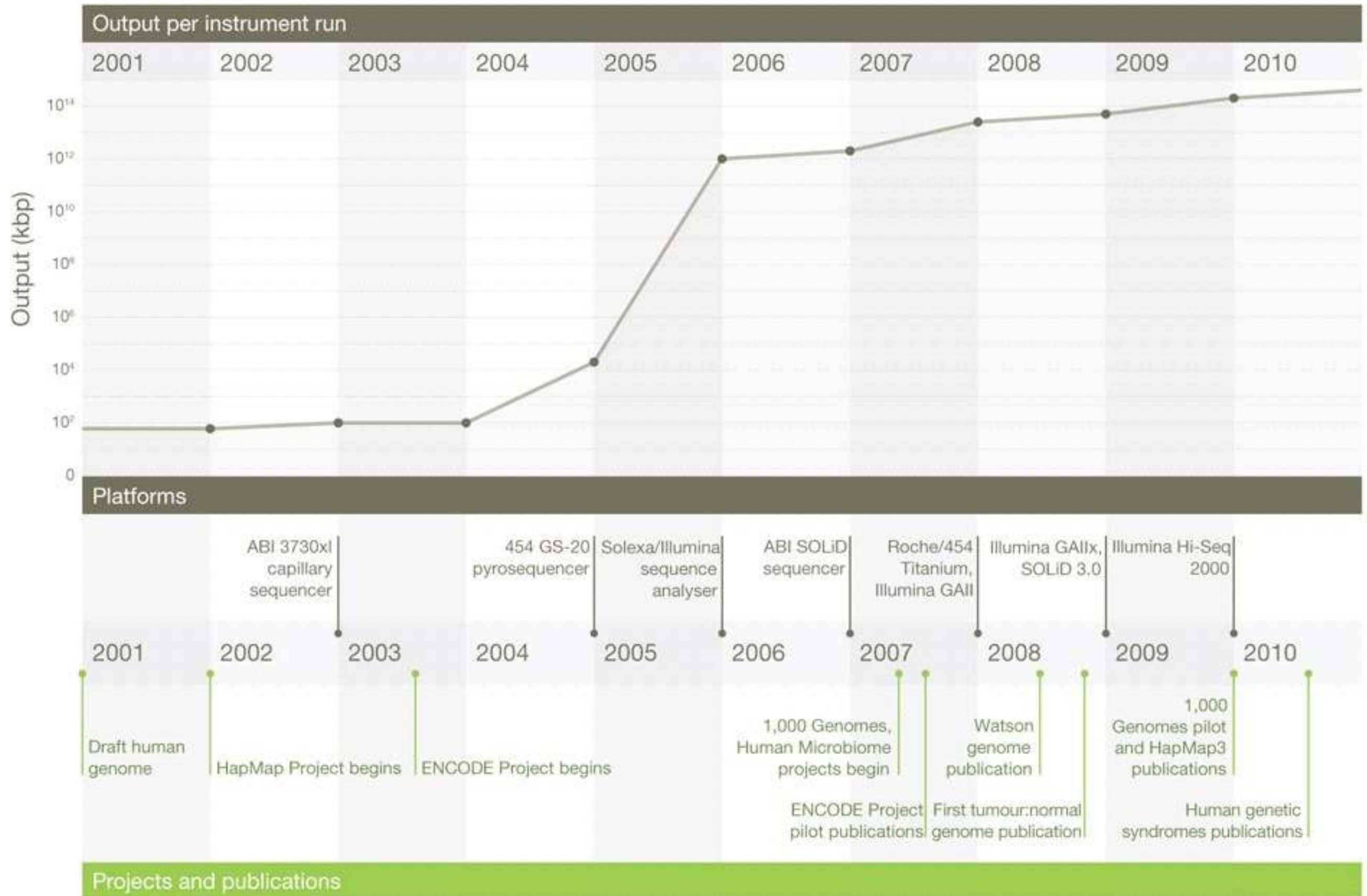
Overview [11223] Eukaryotes [1883] Prokaryotes [31111] Viruses [4483] Plasmids [4799] Organelles [6312]

[Download selected records](#)

Items 1 - 100 of 11223 << First < Prev Page 1 of 113 Next > Last >>

Organism/Name	Kingdom	Group	SubGroup	Size (Mb)	Chr	Organelles	Plasmids	Assemblies
	All	All	All					
'Chrysanthemum coronarium' phytoplasma	Bacteria	Tenericutes	Mollicutes	0.739592	-	-	-	1
Abaca bunchy top virus	Viruses	ssDNA viruses	Nanoviridae	0.008422	6	-	-	1
Abalone herpesvirus Victoria/AUS/2009	Viruses	dsDNA viruses, no RNA stage	unclassified	0.211518	1	-	-	1
Abalone shriveling syndrome-associated virus	Viruses	dsDNA viruses, no RNA stage	unclassified	0.034962	1	-	-	1
Abelson murine leukemia virus	Viruses	Retro-transcribing viruses	Retroviridae	0.005894	1	-	-	1
Abiotrophia defectiva	Bacteria	Firmicutes	Bacilli	2.04344	-	-	-	1
Abutilon Brazil virus	Viruses	ssDNA viruses	Geminiviridae	0.005271	2	-	-	1
Abutilon mosaic Bolivia virus	Viruses	ssDNA viruses	Geminiviridae	0.005399	2	-	-	1
Abutilon mosaic Brazil virus	Viruses	ssDNA viruses	Geminiviridae	0.005282	2	-	-	1
Abutilon mosaic virus	Viruses	ssDNA viruses	Geminiviridae	0.005217	2	-	-	1
Acanthamoeba astronyxis	Eukaryota	Protists	Other Protists	83.4326	-	-	-	1
Acanthamoeba castellanii	Eukaryota	Protists	Other Protists	116.06	-	1	-	3
Acanthamoeba culbertsoni	Eukaryota	Protists	Other Protists	56.5438	-	-	-	1
Acanthamoeba divionensis	Eukaryota	Protists	Other Protists	84.7694	-	-	-	1
Acanthamoeba healyi	Eukaryota	Protists	Other Protists	75.3182	-	-	-	1
Acanthamoeba lenticulata	Eukaryota	Protists	Other Protists	66.0257	-	-	-	1
Acanthamoeba lugdunensis	Eukaryota	Protists	Other Protists	99.4171	-	-	-	1
Acanthamoeba mauritaniensis	Eukaryota	Protists	Other Protists	106.836	-	-	-	1
Acanthamoeba palestinensis	Eukaryota	Protists	Other Protists	103.483	-	-	-	1
Acanthamoeba pearcei	Eukaryota	Protists	Other Protists	116.614	-	-	-	1
Acanthamoeba polyphaga	Eukaryota	Protists	Other Protists	120.416	-	-	-	1
Acanthamoeba polyphaga mimivirus	Viruses	dsDNA viruses, no RNA stage	Mimiviridae	1.18156	1	-	-	1
Acanthamoeba quina	Eukaryota	Protists	Other Protists	83.589	-	-	-	1

A brief history



Genetic Organization

Genetic Organization

chromosome

cell

gene

protein

desoxyribonucleic acid (DNA)

DNA double helix

pyrimidines

purines

DNA

mRNA

transcription

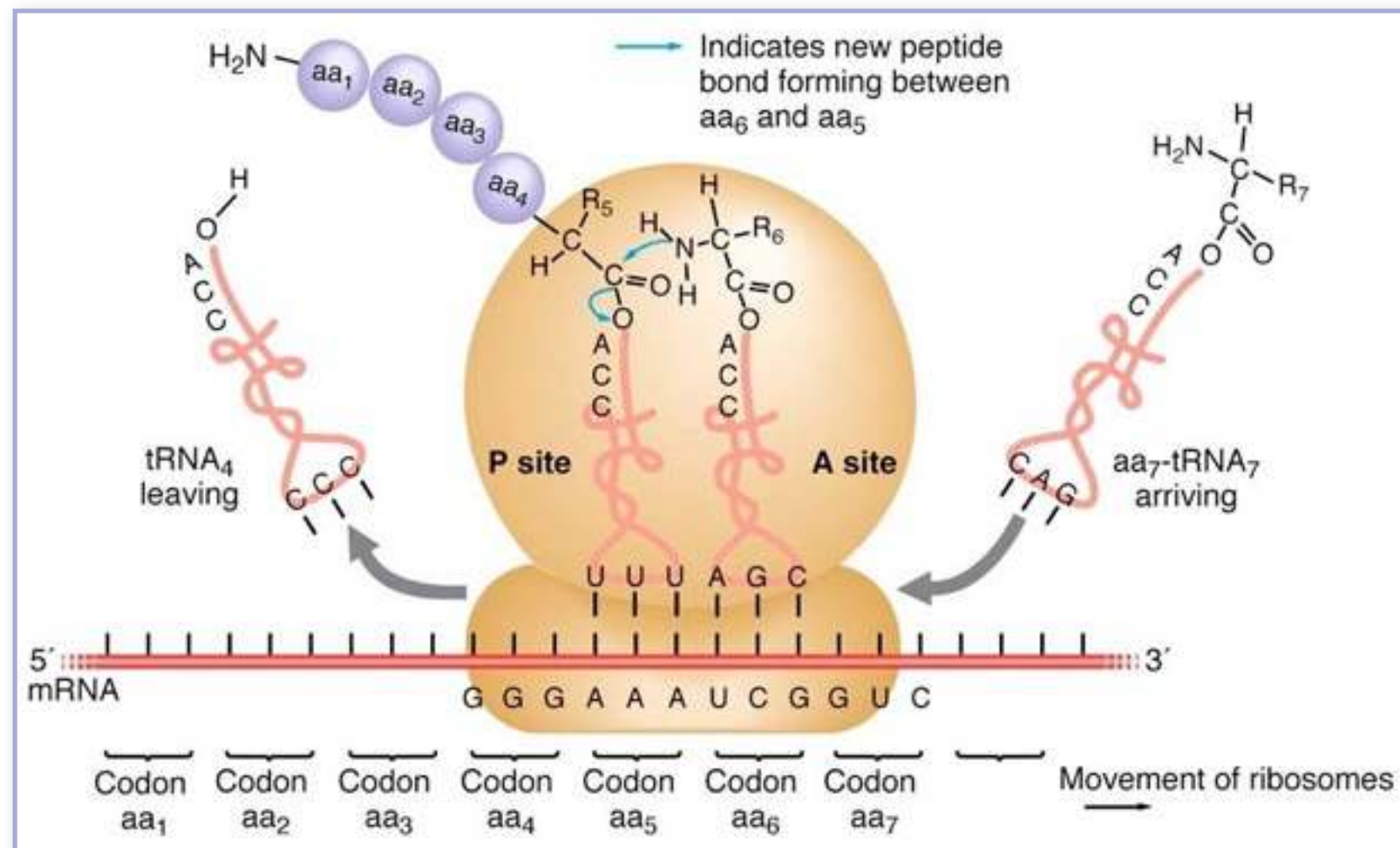
translation

protein



transcription

translation



**protein
structure**

The genetic code*

First Base in the Codon	Second Base in the Codon				Third Base in the Codon
	U	C	A	G	
U	Phenylalanine	Serine	Tyrosine	Cysteine	U
	Phenylalanine	Serine	Tyrosine	Cysteine	C
	Leucine	Serine	Stop	Stop	A
	Leucine	Serine	Stop	Tryptophan	G
C	Leucine	Proline	Histidine	Arginine	U
	Leucine	Proline	Histidine	Arginine	C
	Leucine	Proline	Glutamine	Arginine	A
	Leucine	Proline	Glutamine	Arginine	G
A	Isoleucine	Threonine	Asparagine	Serine	U
	Isoleucine	Threonine	Asparagine	Serine	C
	Isoleucine	Threonine	Lysine	Arginine	A
	Methionine	Threonine	Lysine	Arginine	G
G	Valine	Alanine	Aspartic acid	Glycine	U
	Valine	Alanine	Aspartic acid	Glycine	C
	Valine	Alanine	Glutamic acid	Glycine	A
	Valine	Alanine	Glutamic acid	Glycine	G

*Note that there is not just one 'universal' genetic code!

The degenerated genetic code

First Base in the Codon	Second Base in the Codon				Third Base in the Codon
	U	C	A	G	
U	Phenylalanine	Serine	Tyrosine	Cysteine	U
	Phenylalanine	Serine	Tyrosine	Cysteine	C
	Leucine	Serine	Stop	Stop	A
	Leucine	Serine	Stop	Tryptophan	G
C	Leucine	Proline	Histidine	Arginine	U
	Leucine	Proline	Histidine	Arginine	C
	Leucine	Proline	Glutamine	Arginine	A
	Leucine	Proline	Glutamine	Arginine	G
A	Isoleucine	Threonine	Asparagine	Serine	U
	Isoleucine	Threonine	Asparagine	Serine	C
	Isoleucine	Threonine	Lysine	Arginine	A
	Methionine	Threonine	Lysine	Arginine	G
G	Valine	Alanine	Aspartic acid	Glycine	U
	Valine	Alanine	Aspartic acid	Glycine	C
	Valine	Alanine	Glutamic acid	Glycine	A
	Valine	Alanine	Glutamic acid	Glycine	G

4-fold degenerated

2-fold degenerated

Molecular Evolution & Phenotypic Diversity

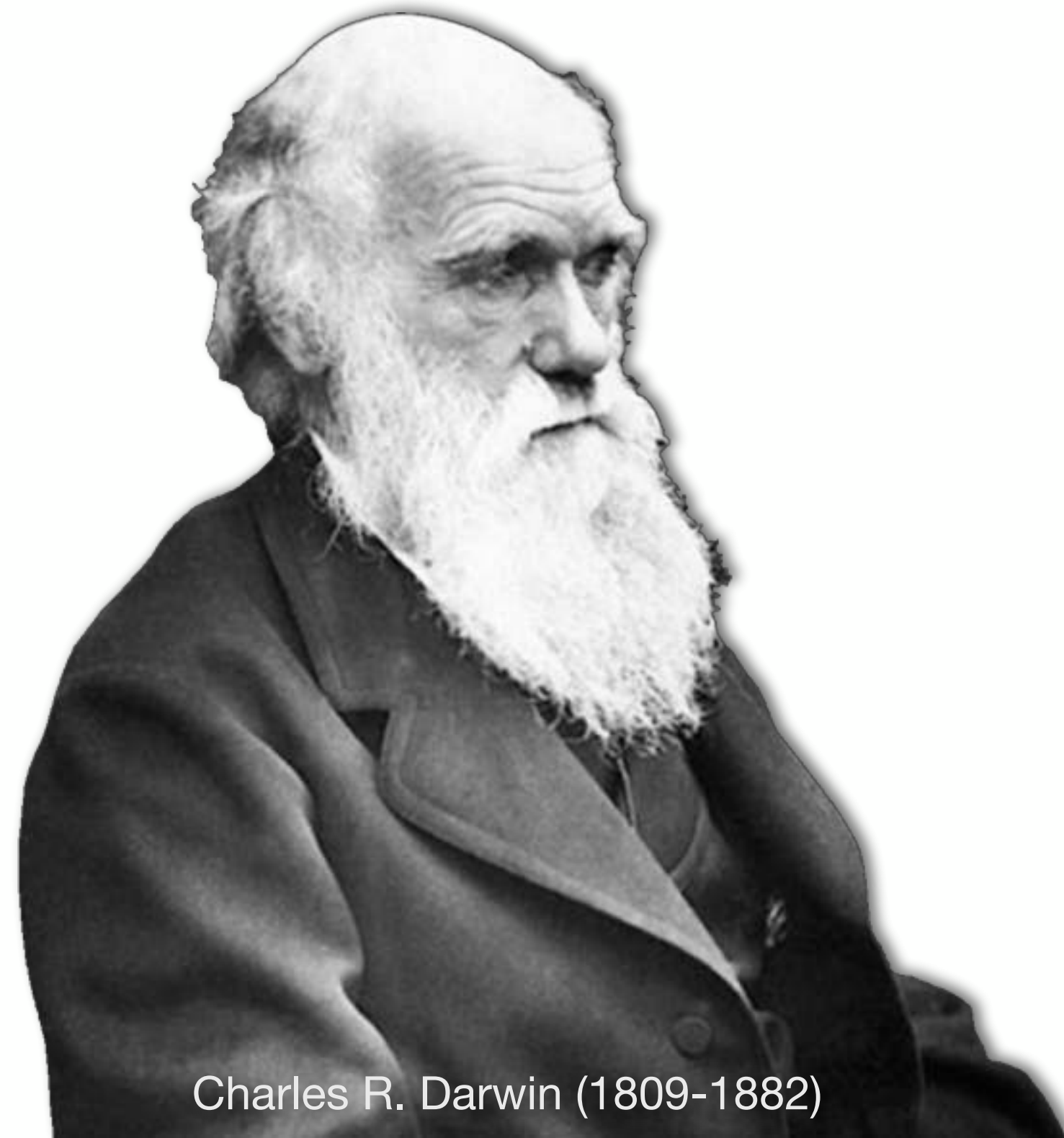


Walter Salzburger

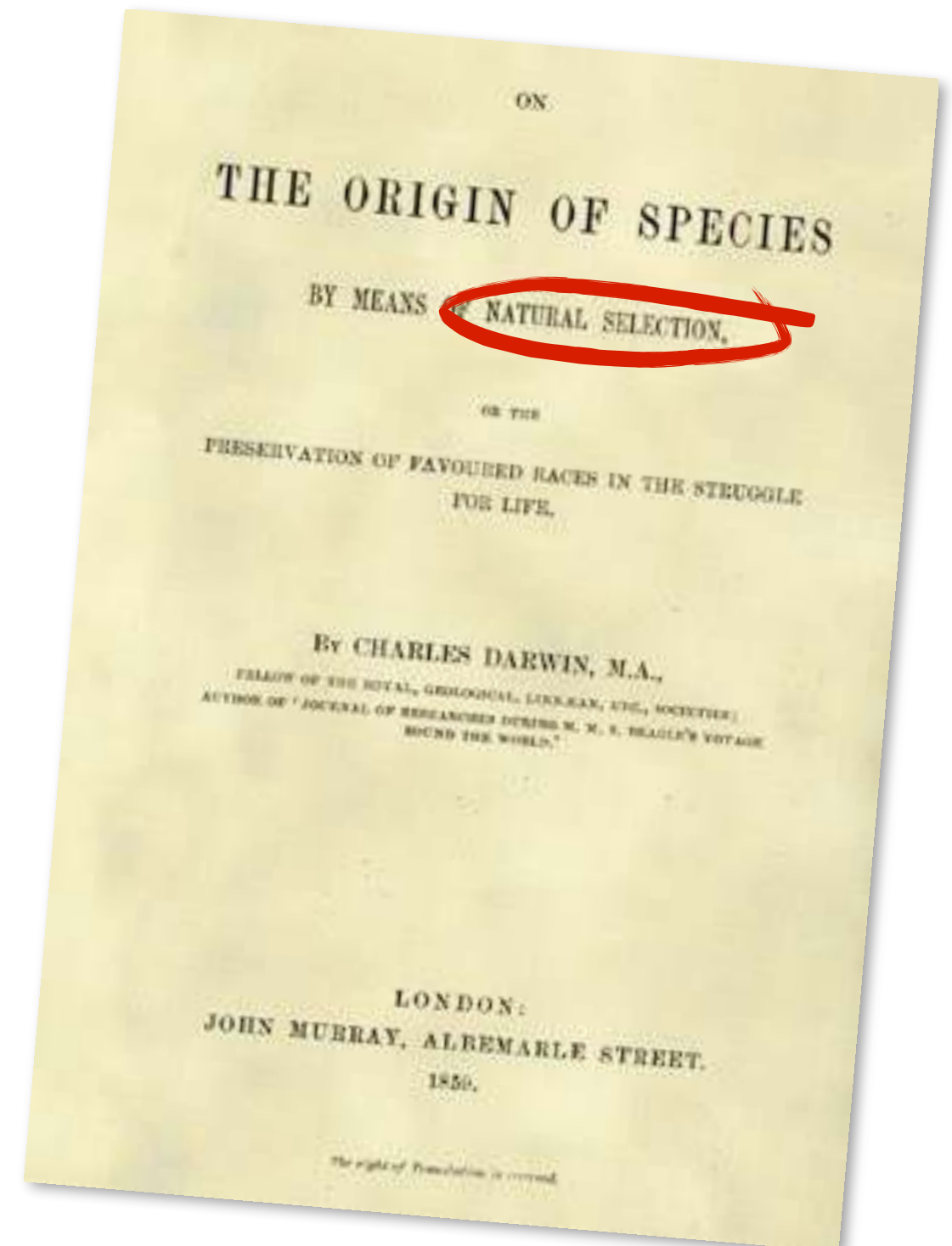
Zoological Institute
University of Basel, Switzerland



Natural Selection...



Charles R. Darwin (1809-1882)



Natural Selection...

- ▶ ... “is the process by which the forms of organisms in a population that are **best adapted** to the environment **increase in frequency** relative to less well-adapted forms over a **number of generations**” (Ridley 1996)



Sexual Selection...

- ...“is the selection on mating behavior, either through **competition among members of one sex** (usually males) for access to members of the other sex or through **choice by members of one sex** (usually females) for certain members of the other sex” (Ridley 1996)



images: www.crbs.umd.edu, www.smh.com.au

Natural Selection...

...Sexual Selection

fitness

competitors

sexual selection

individual fitness

other members of
the same sex

natural selection

fitness of the
genotype

other individuals in
the same
population

- **...operates** if the following **conditions** are met:

reproduction	organisms must reproduce to form new generations
heredity	offspring resemble parents (“like must produce like”)
trait variation	individuals in natural populations vary in (adaptive) traits
variation in fitness	individuals in natural populations vary in the number of their offspring that survive to reproduce (‘lifetime reproductive success’)

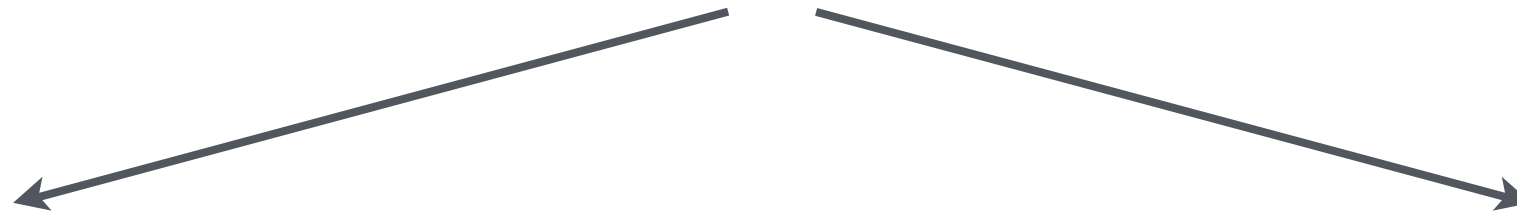
Natural Variation

- **Natural populations** show **variation** at all levels, from gross morphology to DNA sequences. Selection can only operate, if heritable variation exists.



Natural Variation

- ▶ **Natural populations** show **variation** at all levels, from gross morphology to DNA sequences. Selection can only operate, if heritable variation exists.
- ▶ Natural variation is generated by two processes:



recombination

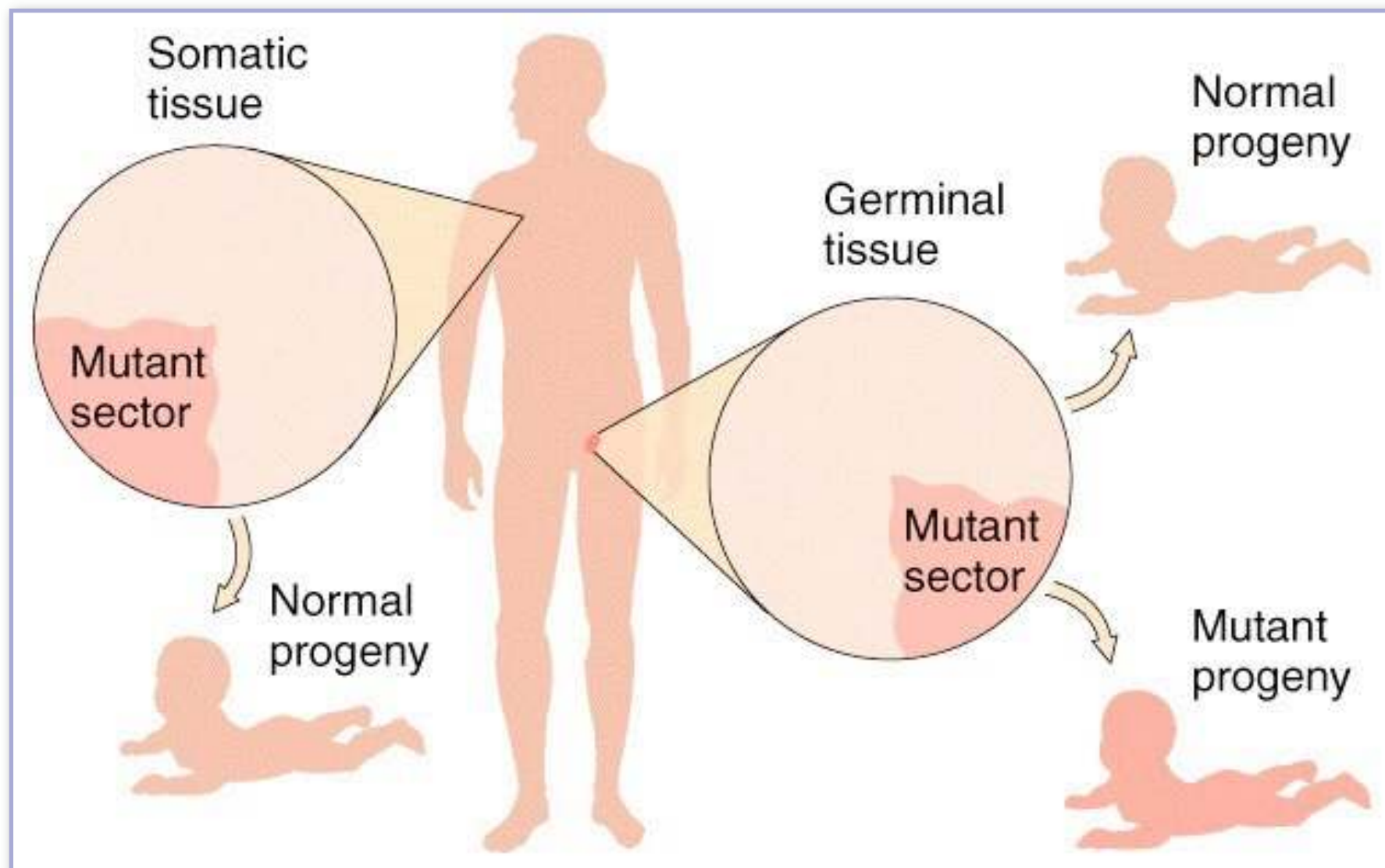
“reshuffling” of genetic material
by introducing or breaking up
physical linkage

mutation

generation of new genetic
variation by “mistakes” during
the copying of a DNA strand

Natural Variation

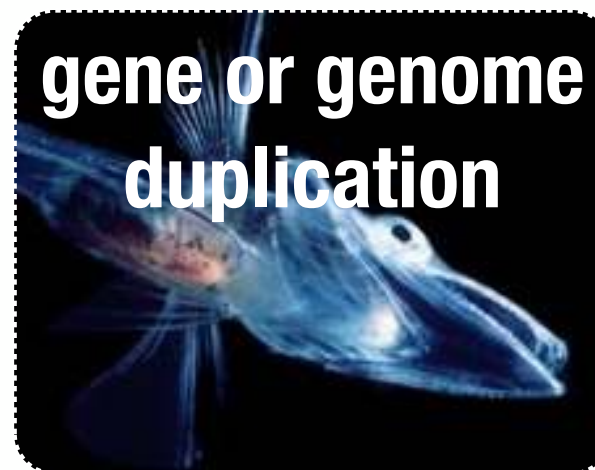
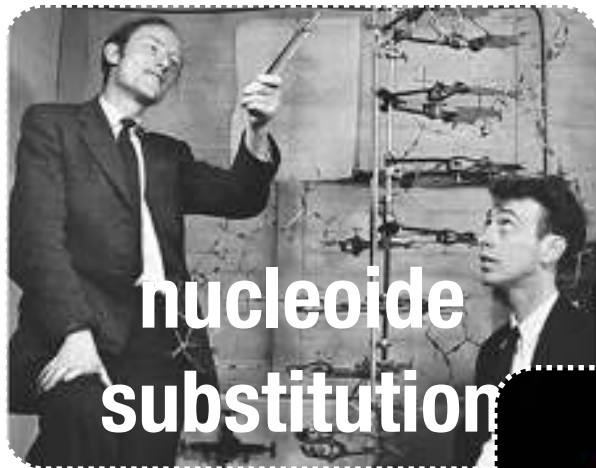
- ▶ New mutations are only transmitted to the next generation, if they occur in **germinal tissue**!



Phenotypic Evolution

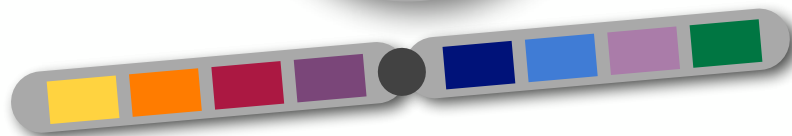
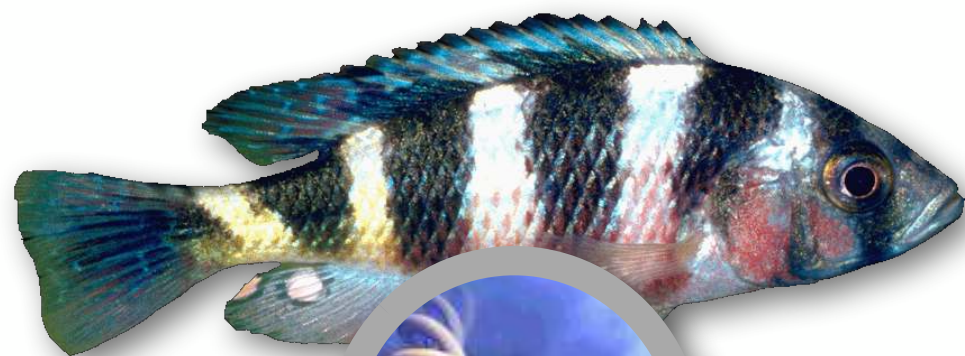
“The primary cause of evolution is the mutational change of genes”

Nei and Kumar (2000)



Phenotypic Evolution

phenome



morphology
ecology
behavior



genome
structure

and/or

coding
sequence

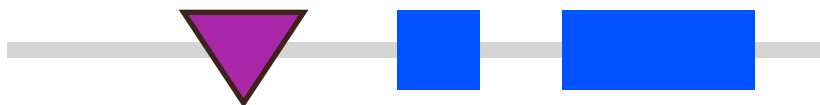
and/or

regulatory
regions

ATG AAC GCA TGG AGG...
Met Asn Ala Trp Arg

genome

ATG AAC **GTA** TGG AGG...
Met Asn **Val** Trp Arg



Nucleotide substitutions

original DNA sequence:

C	C	G	C	T	C	G	T	C	A	A	C	T	A	G
GLY			GLU			GLN			LEU			ILE		

synonymous mutation:

C > T

C	C	G	C	T	T	G	T	C	A	A	C	T	A	G
GLY			GLU			GLN			LEU			ILE		

non-synonymous mutation:

A > C

C	C	G	C	T	C	G	T	C	C	A	C	T	A	G
GLY			GLU			GLN			ILE			ILE		

frameshift mutation:

insert C

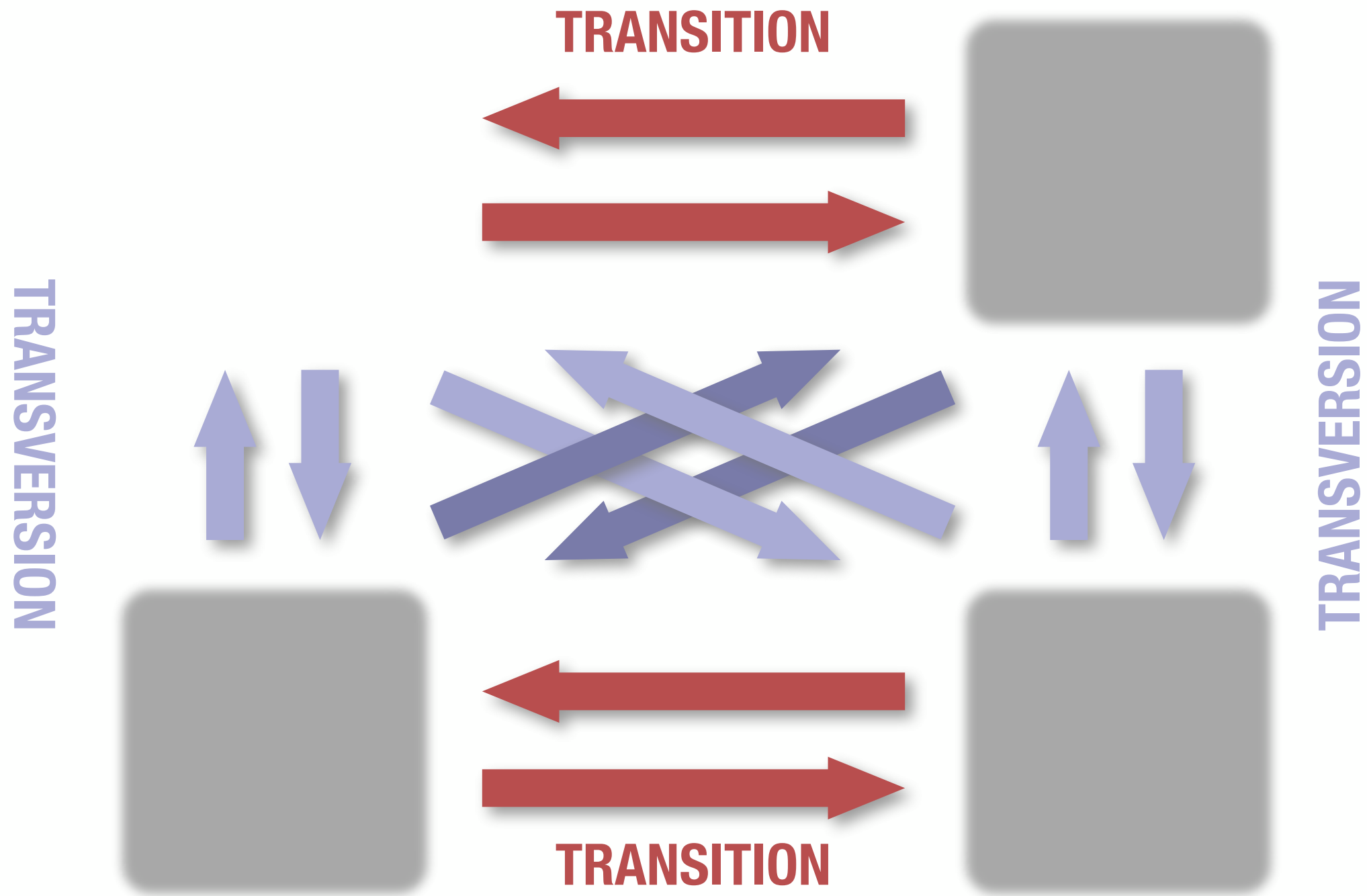
C	C	G	C	C	T	C	G	T	C	A	A	C	T	A
GLY			GLY			ALA			VAL			ASP		

stop mutation:

G > A

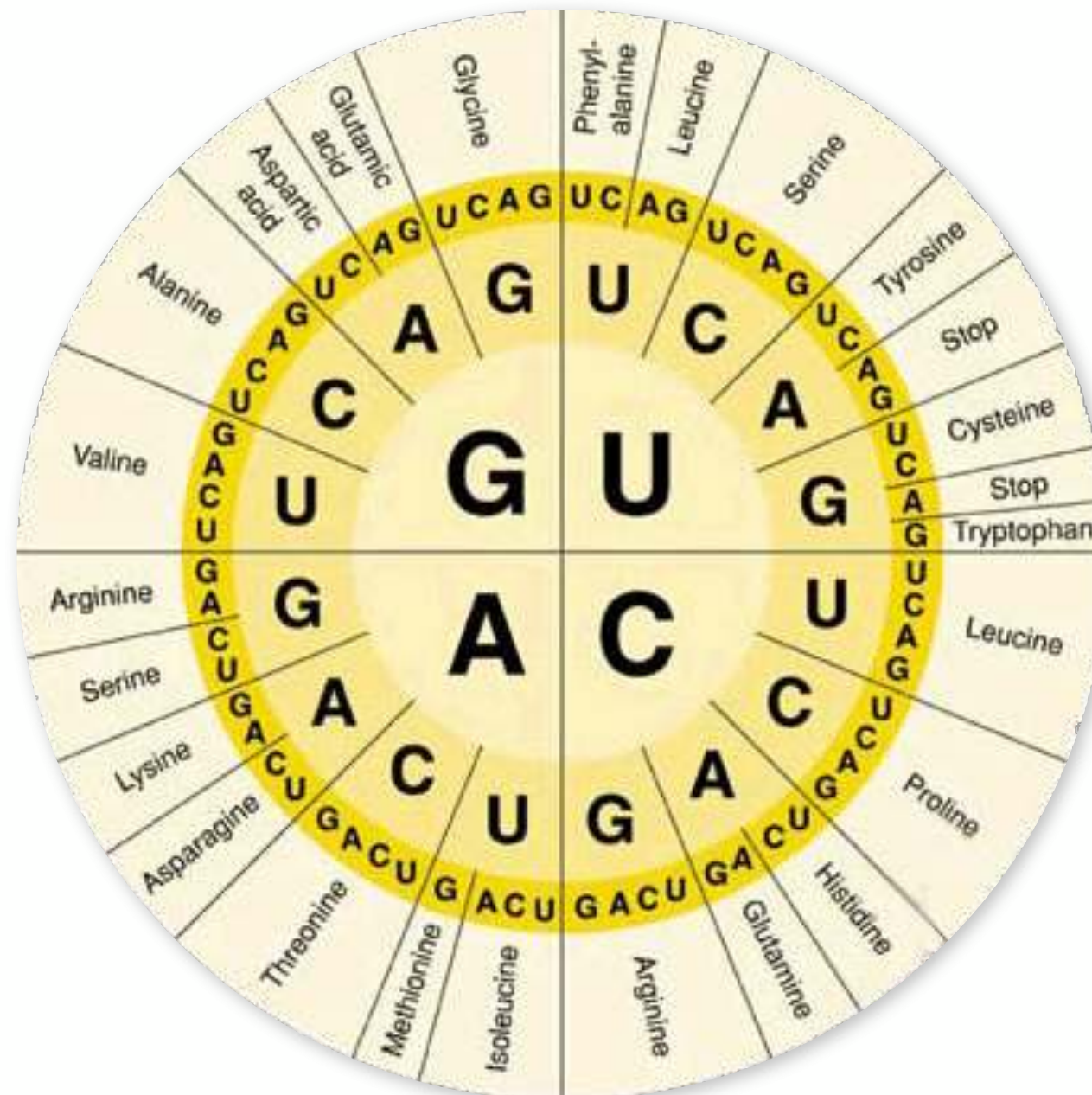
C	C	G	C	T	C	A	T	C	A	A	C	T	A	G
GLY			GLU			STOP!								

Nucleotide substitutions



Transition mutations outnumber transversions!

Adaptive Sequence Evolution



Adaptive Sequence Evolution

dN/dS ratio

- ▶ dN/dS (Ka/Ks) provides information about the form of sequence evolution in a gene*:
 - ▶ **$dN/dS < 1$: purifying selection** (non-synonymous sites are evolving slower than synonymous sites)
 - ▶ **$dN/dS = 1$: neutrality** (non-synonymous and synonymous sites are evolving at equal rates)
 - ▶ **$dN/dS > 1$: positive selection** (non-synonymous sites are evolving faster than synonymous sites; natural selection has acted to change the amino acid sequence)

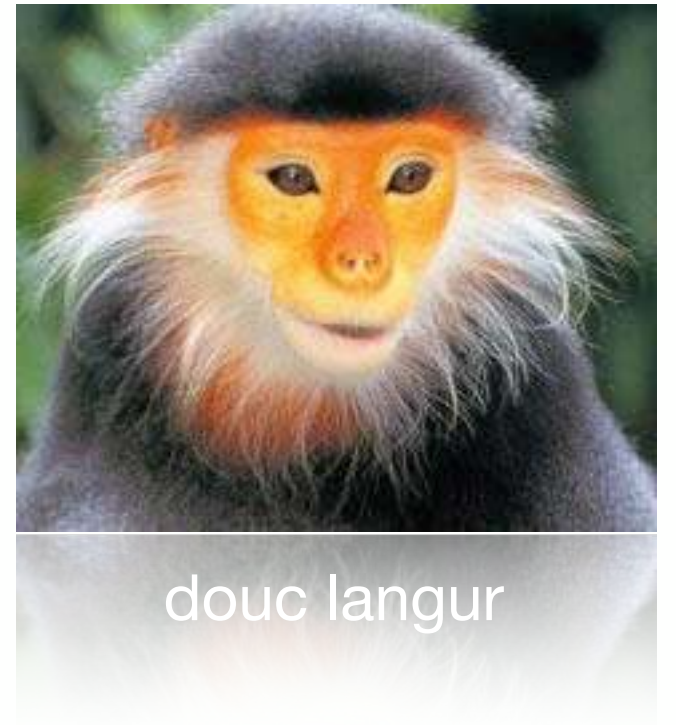
* dS is used as proxy for the neutral rate of molecular evolution

Adaptive Sequence Evolution

Pancreatic ribonuclease in colobine monkeys



Zhang et al. (2002)



Adaptive Sequence Evolution

Pancreatic ribonuclease in colobine monkeys

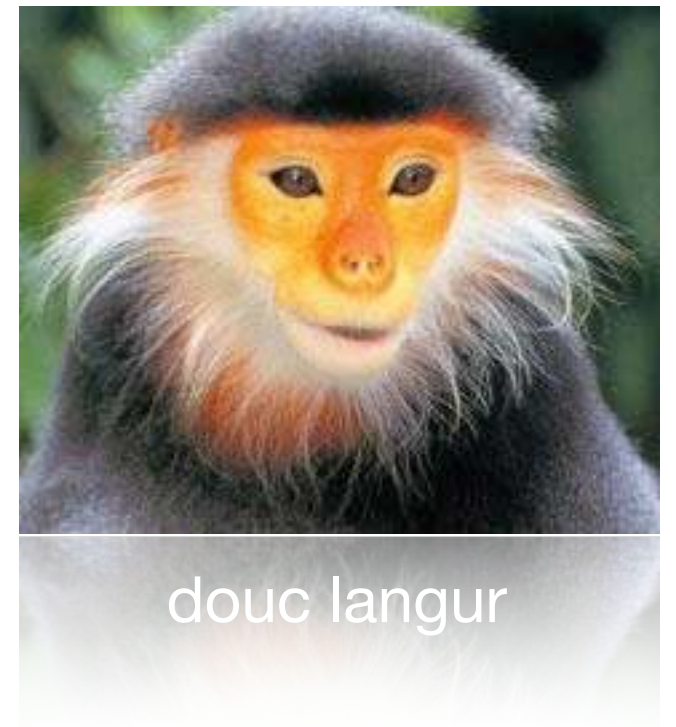
$$K_a = 0.0310$$

substitutions per non-synonymous sites

$$K_s = 0.0077$$

substitutions per synonymous sites
(and non-coding sites)

$$K_a/K_s = 4.026$$



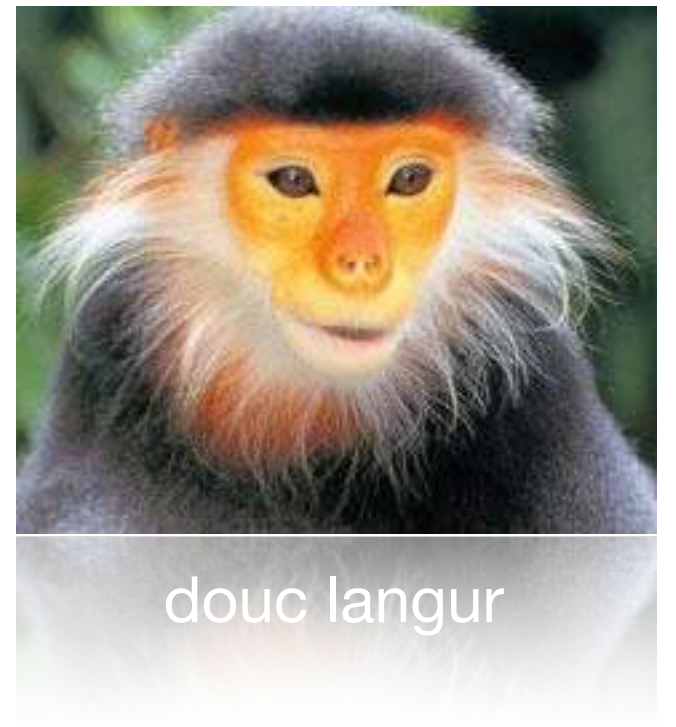
douc langur

Adaptive Sequence Evolution

Pancreatic ribonuclease in colobine monkeys



Zhang et al. (2002)

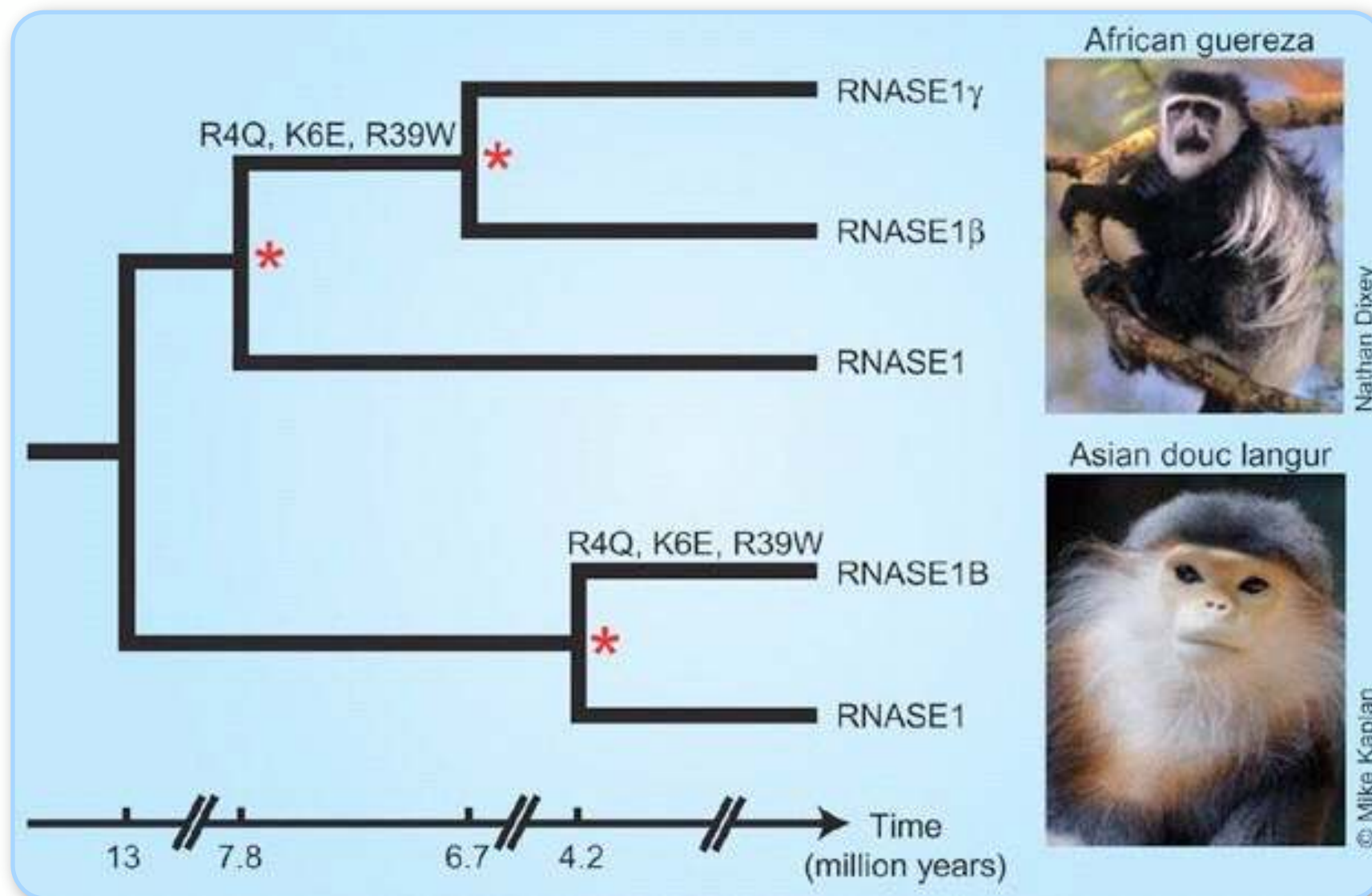


douc langur

adaptation to different pH in the colobine small intestine

Adaptive Sequence Evolution

Pancreatic ribonuclease in leaf-eating monkeys



Prud'homme & Carroll (2006)



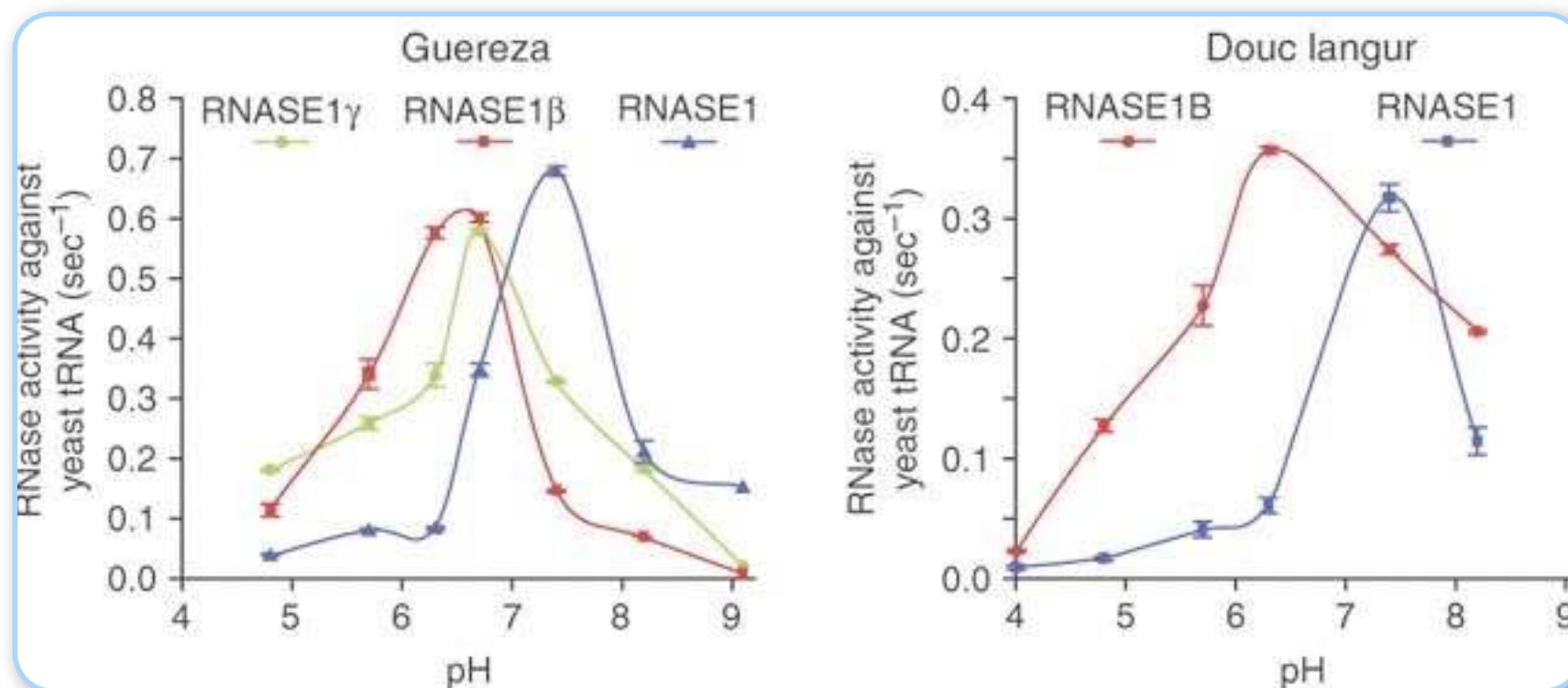
douc langur



guereza

Adaptive Sequence Evolution

Pancreatic ribonuclease in leaf-eating monkeys



Zhang (2006)



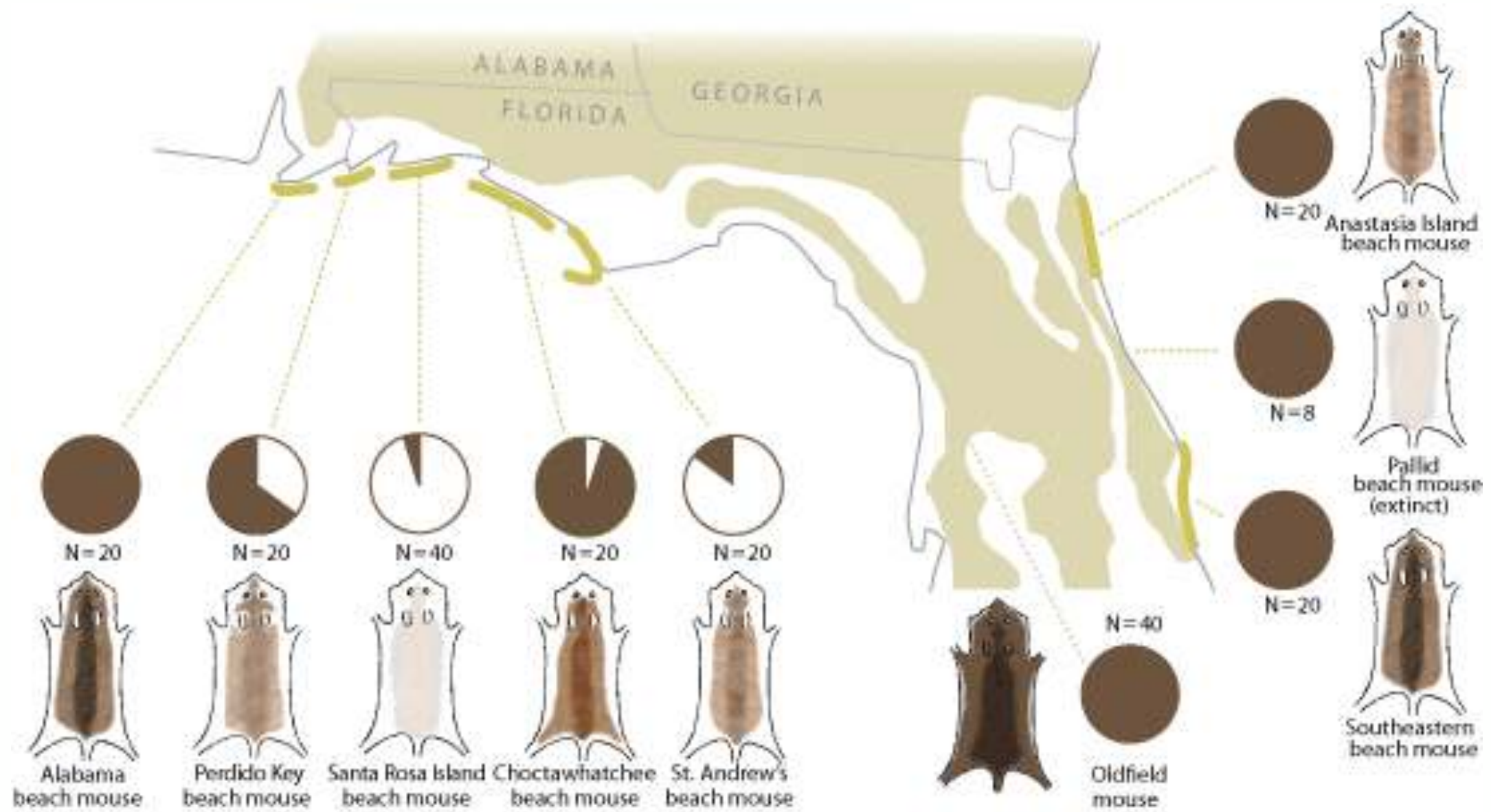
douc langur



guereza

Coding Sequence Evolution

beach mice



Hoekstra et al. (2006)

Coding Sequence Evolution

beach mice



Nachman et al. (2003)

Coding Sequence Evolution

phenotype	<i>mc1r</i> alignment position								
	52	205	313	325	478	633	699	720	819
dark (1371)	T	C	G	T	T	C	C	T	C
	.	.	.	.	.	.	.	.	.
dark (1373)	.	.	T	.	.	.	.	.	.
	.	.	.	.	.	.	.	.	.
dark (1384)	.	.	T	.	.	.	.	.	.
	C	.	.	C	C	T	A	.	.
dark (1412)	.	.	.	.	.	.	.	.	.
	C	T	.	C	C	T	A	.	.
light (1387)	C	.	.	C	C	T	A	G	T
	C	.	.	C	C	T	A	G	T
light (1403)	C	.	.	C	C	T	A	.	.
	C	.	.	C	C	T	A	.	.

AS change
AS position

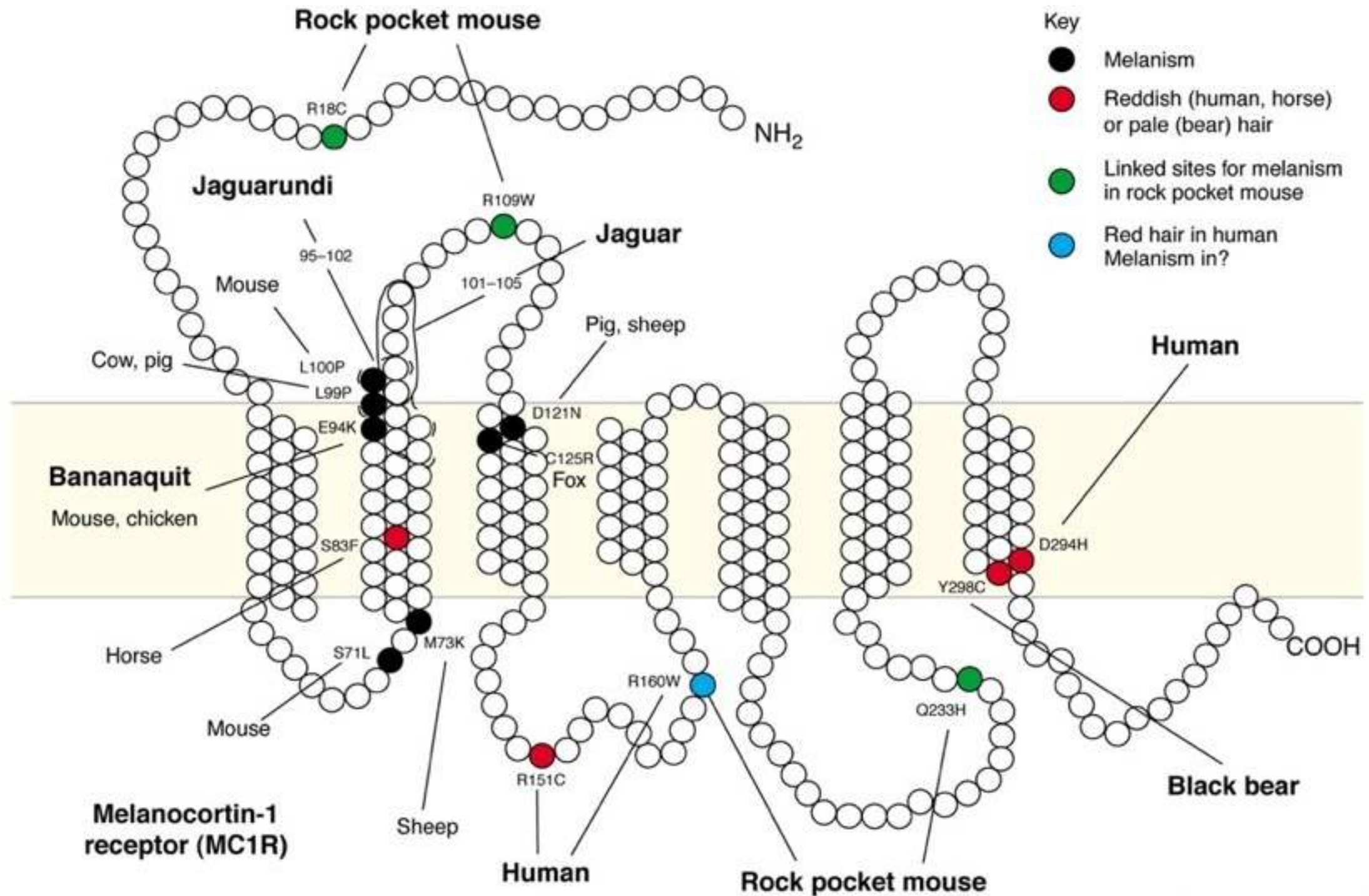
R<>C
18

R<>W R<>W
109 160

Q<>H
233

Nachman et al. (2003)

Coding Sequence Evolution



TRENDS in Genetics

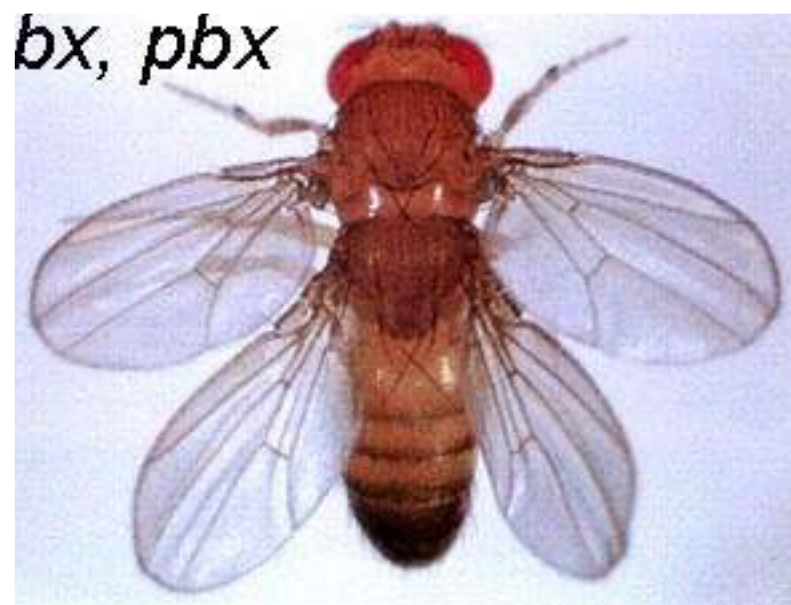
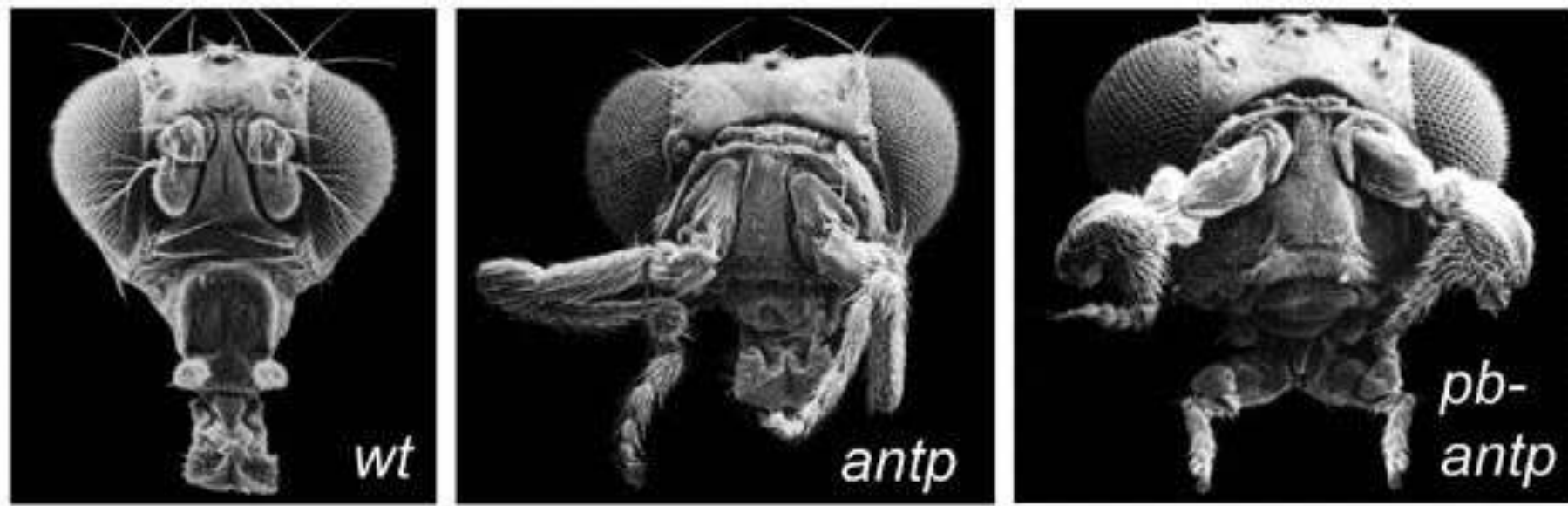
Regulatory Evolution

Hox genes

- ▶ **=homeotic genes:** genes that regulate the identity of body regions
- ▶ Hox genes contain a **homeobox**, a 180 bp region of DNA encoding a particular class of DNA-binding domains. Approximately 20 families of homeobox-containing genes exist.
- ▶ The homeobox encodes the **homeodomain**
- ▶ Mutations in homeotic genes cause the transformation of one body region or part into the likeness of another

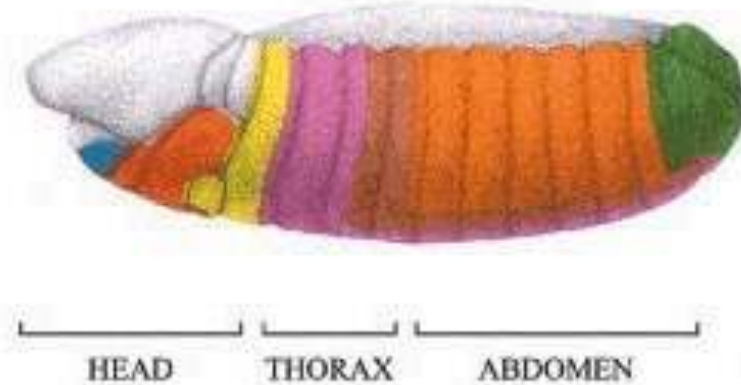
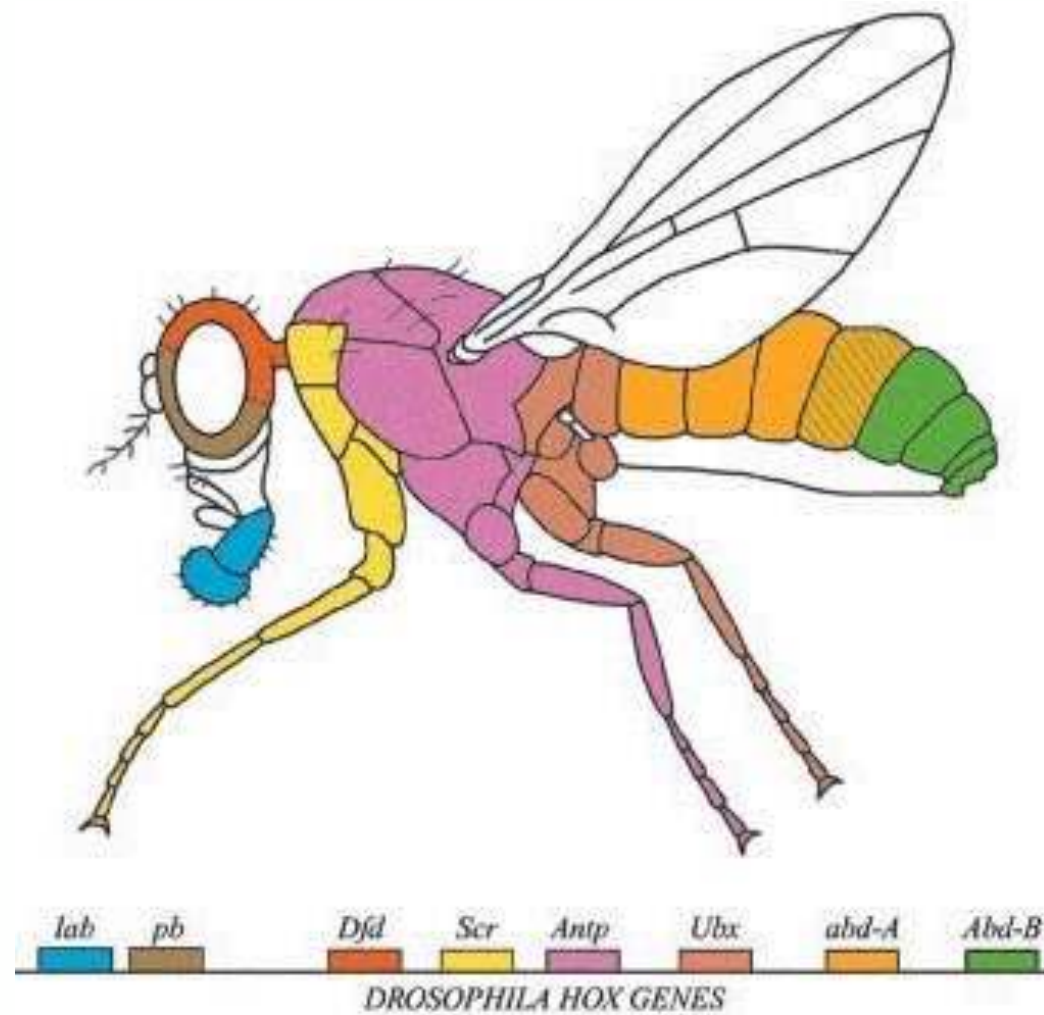
Regulatory Evolution

Hox genes, Drosophila



Regulatory Evolution

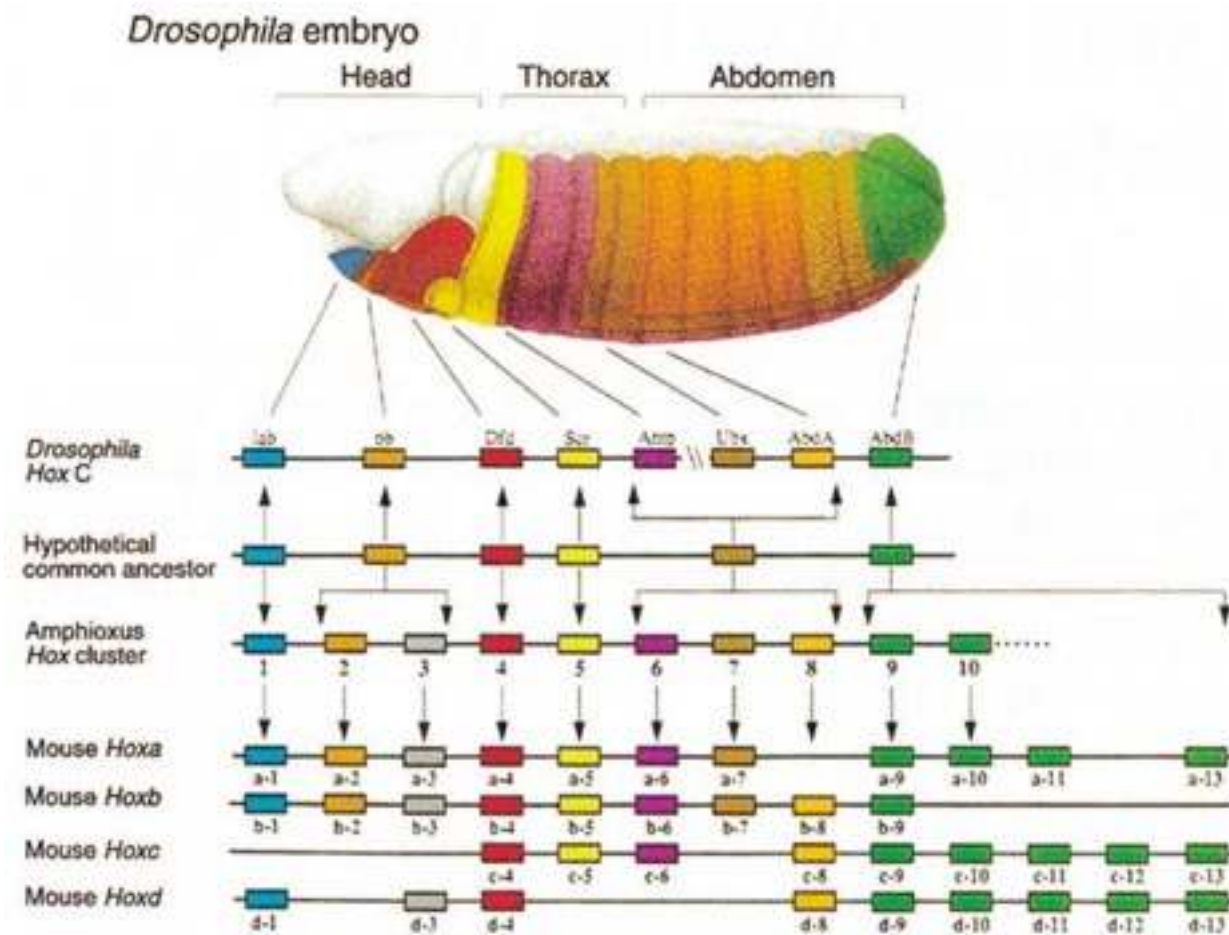
Hox genes and animal body plan



Carroll et al. (2001)

Regulatory Evolution

Hox genes and animal body plan



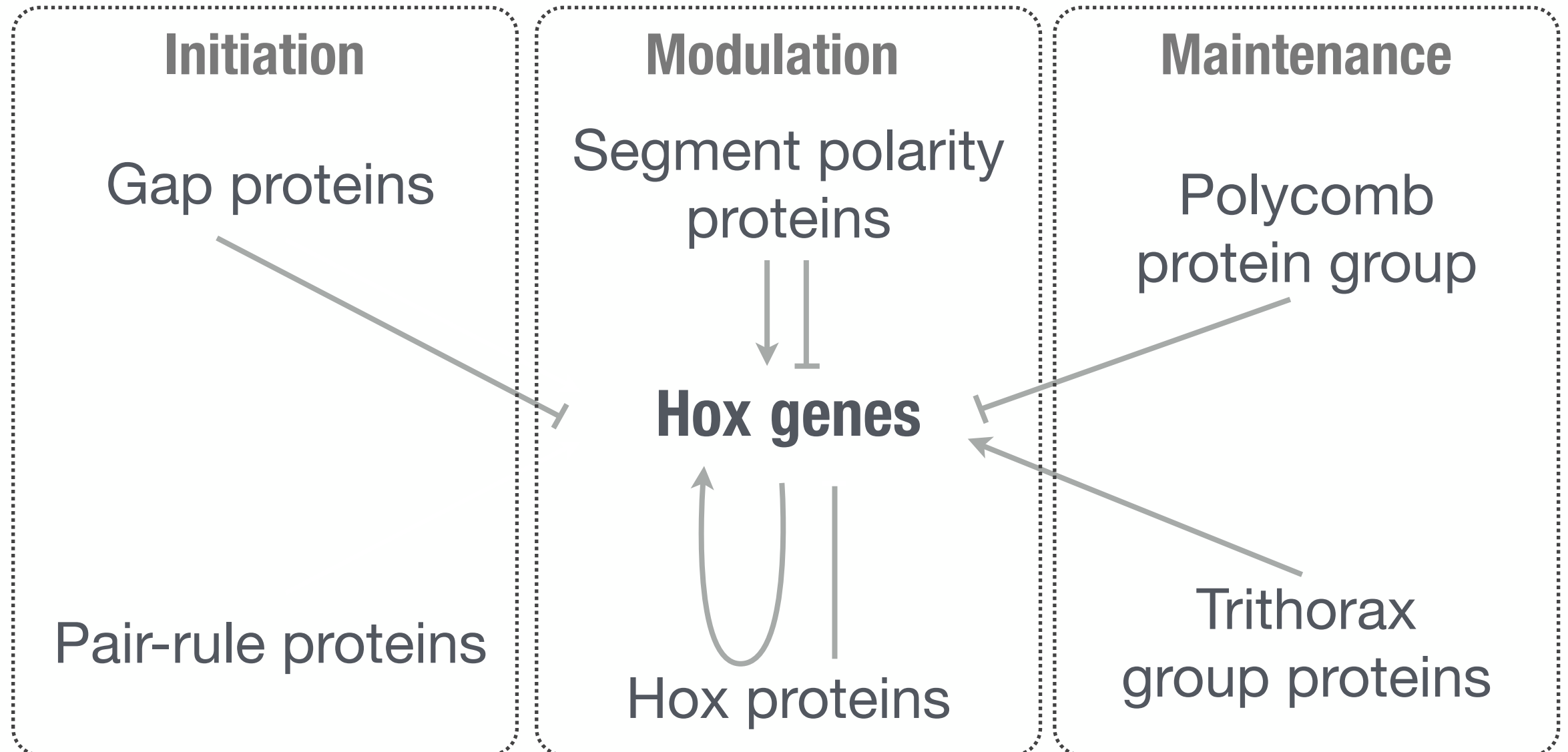
Mouse embryo



Carroll et al. (2001)

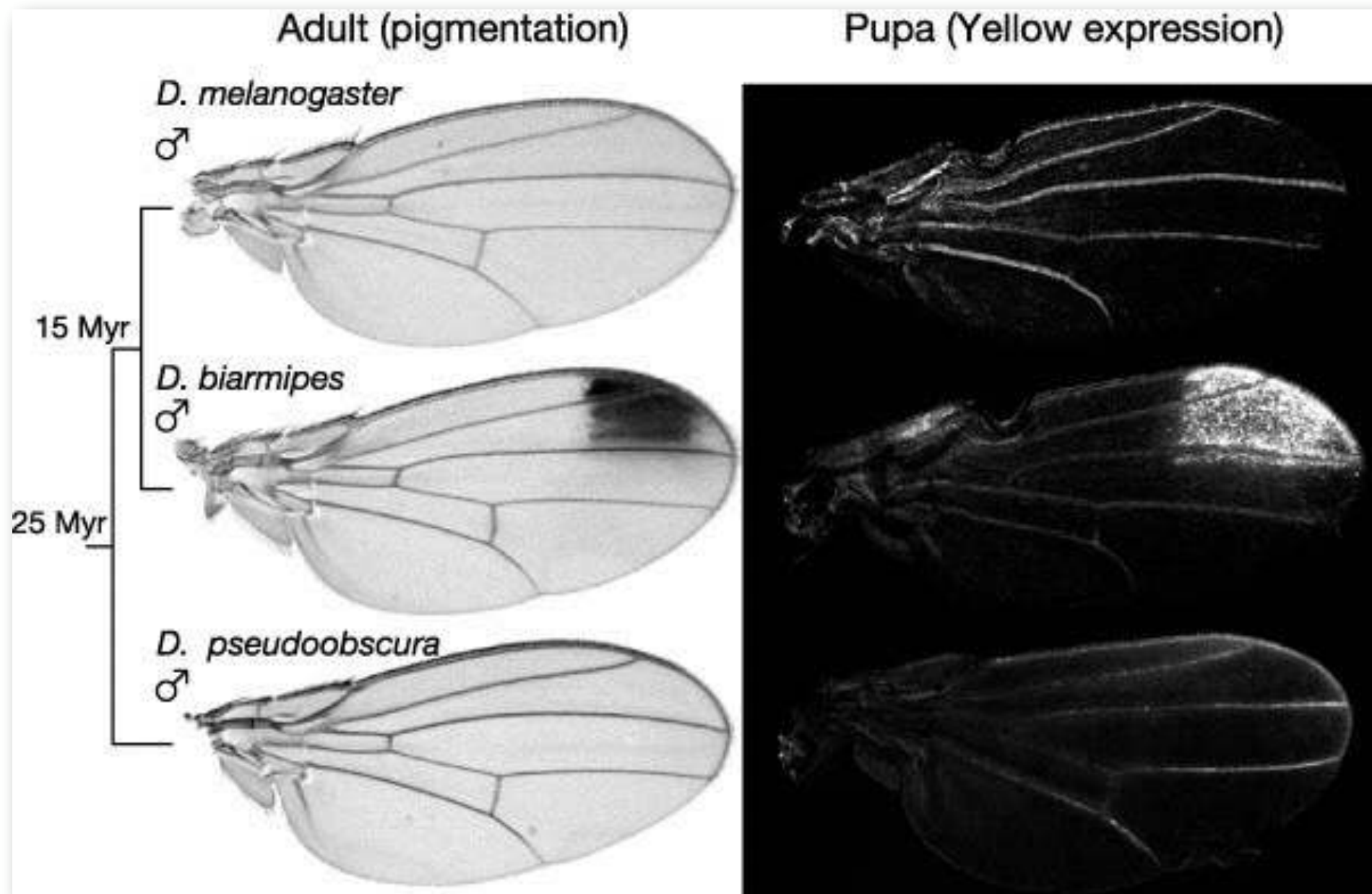
Regulatory Evolution

Hox gene regulation



Regulatory Evolution

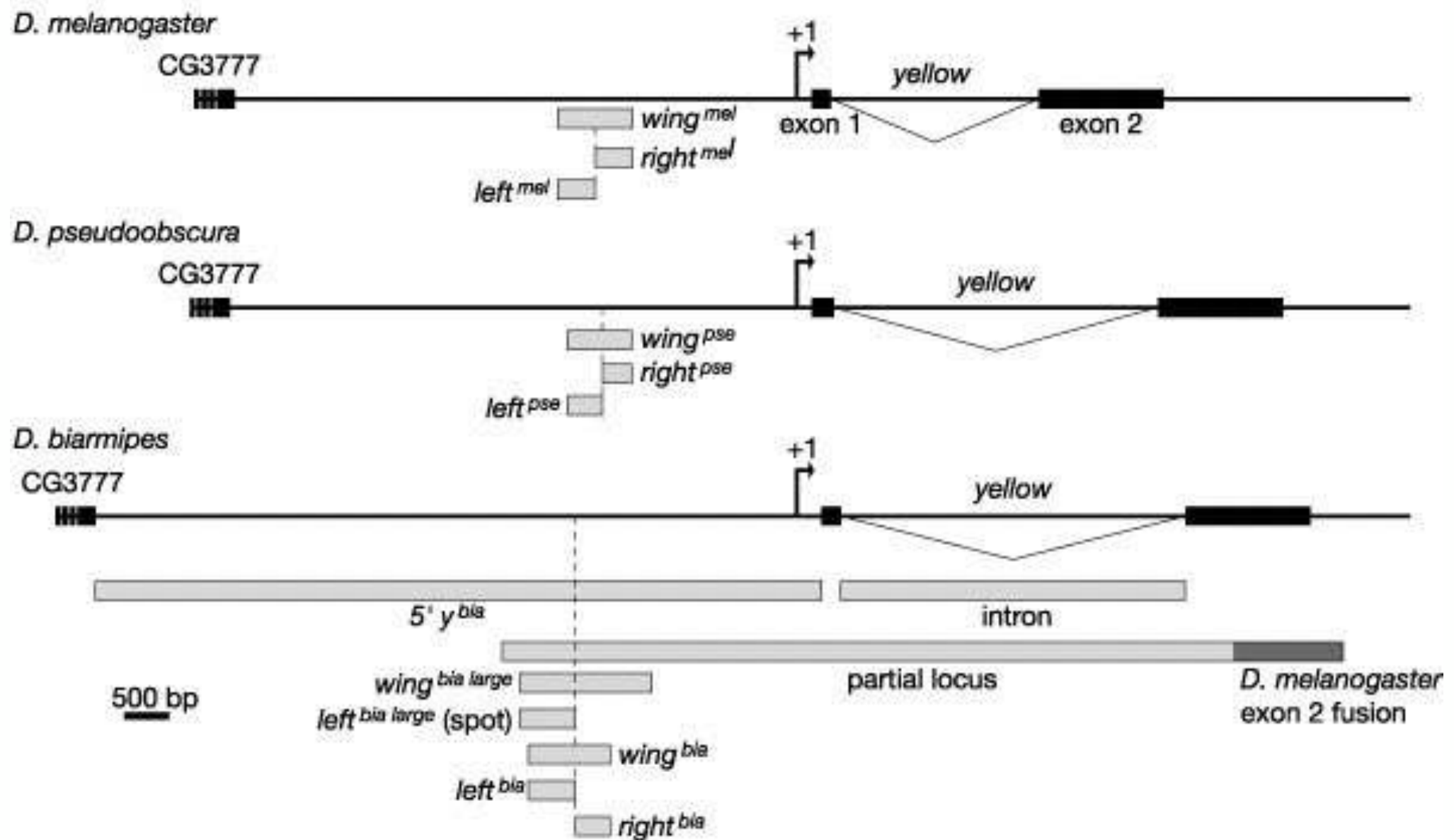
cis-regulatory evolution, *Drosophila*



Gompel et al. 2005

Regulatory Evolution

cis-regulatory evolution, Drosophila

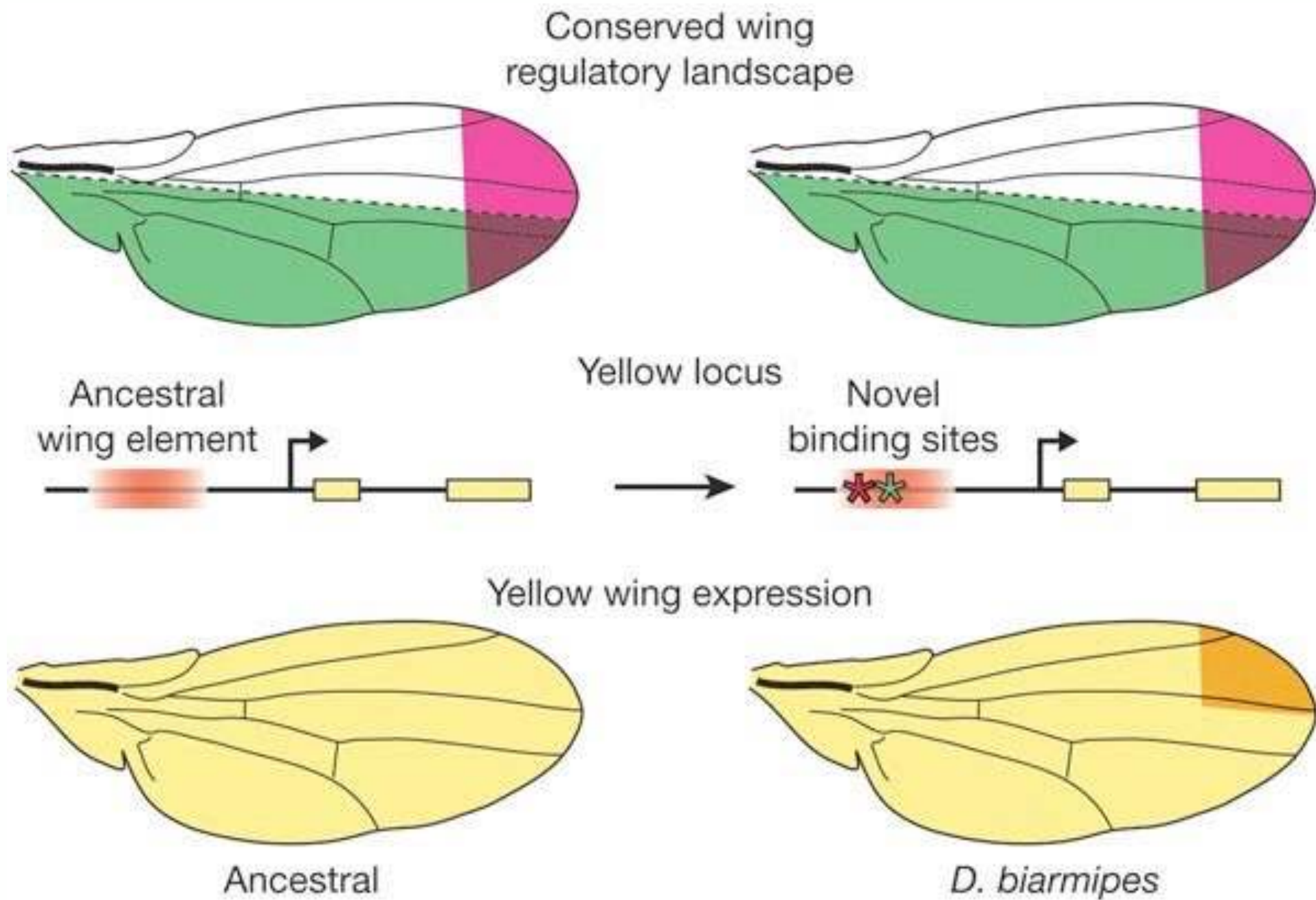


Gompel et al. 2005

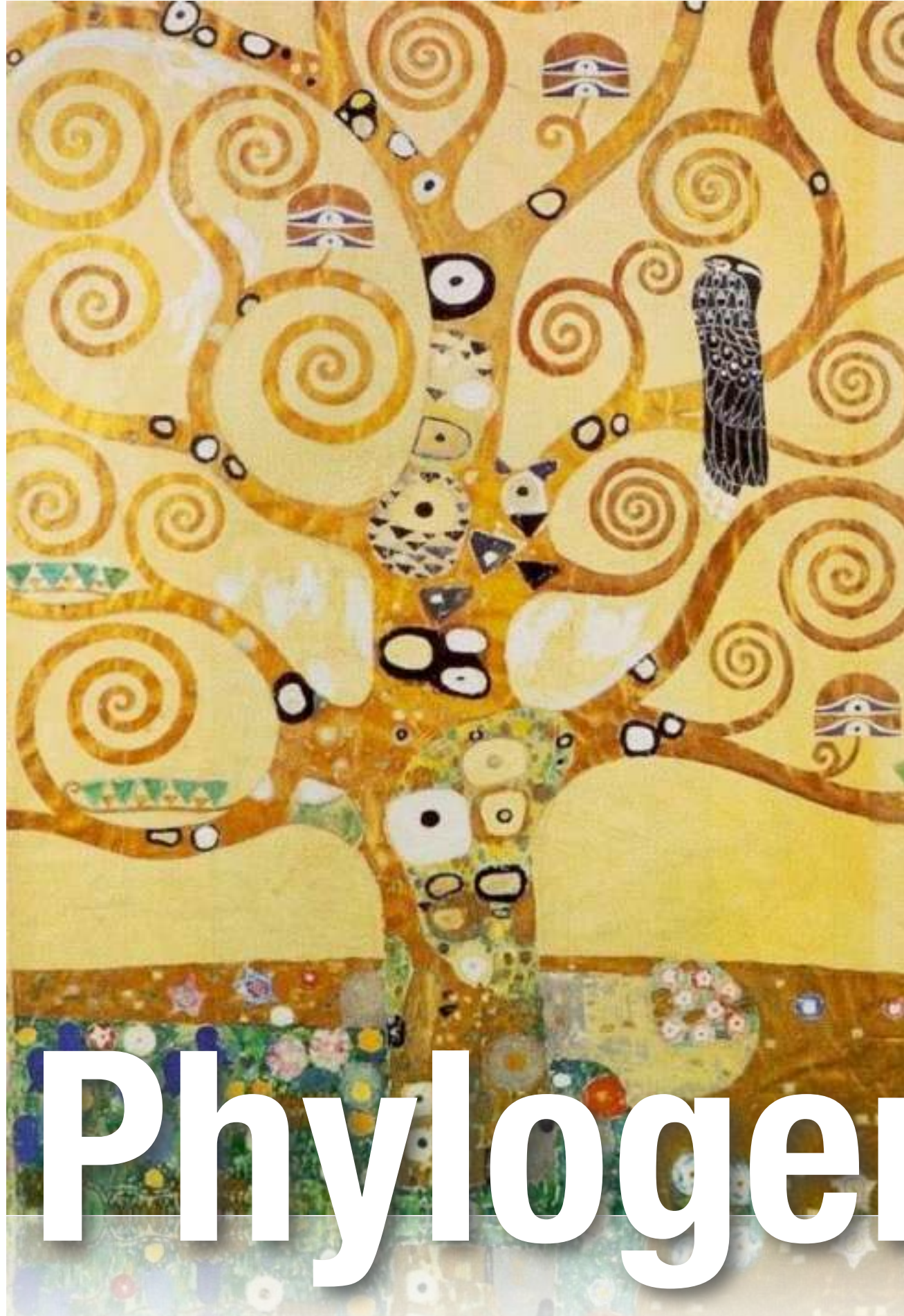
yellow locus

Regulatory Evolution

cis-regulatory evolution, *Drosophila*



“The tree of life”, Gustav Klimt (1907)



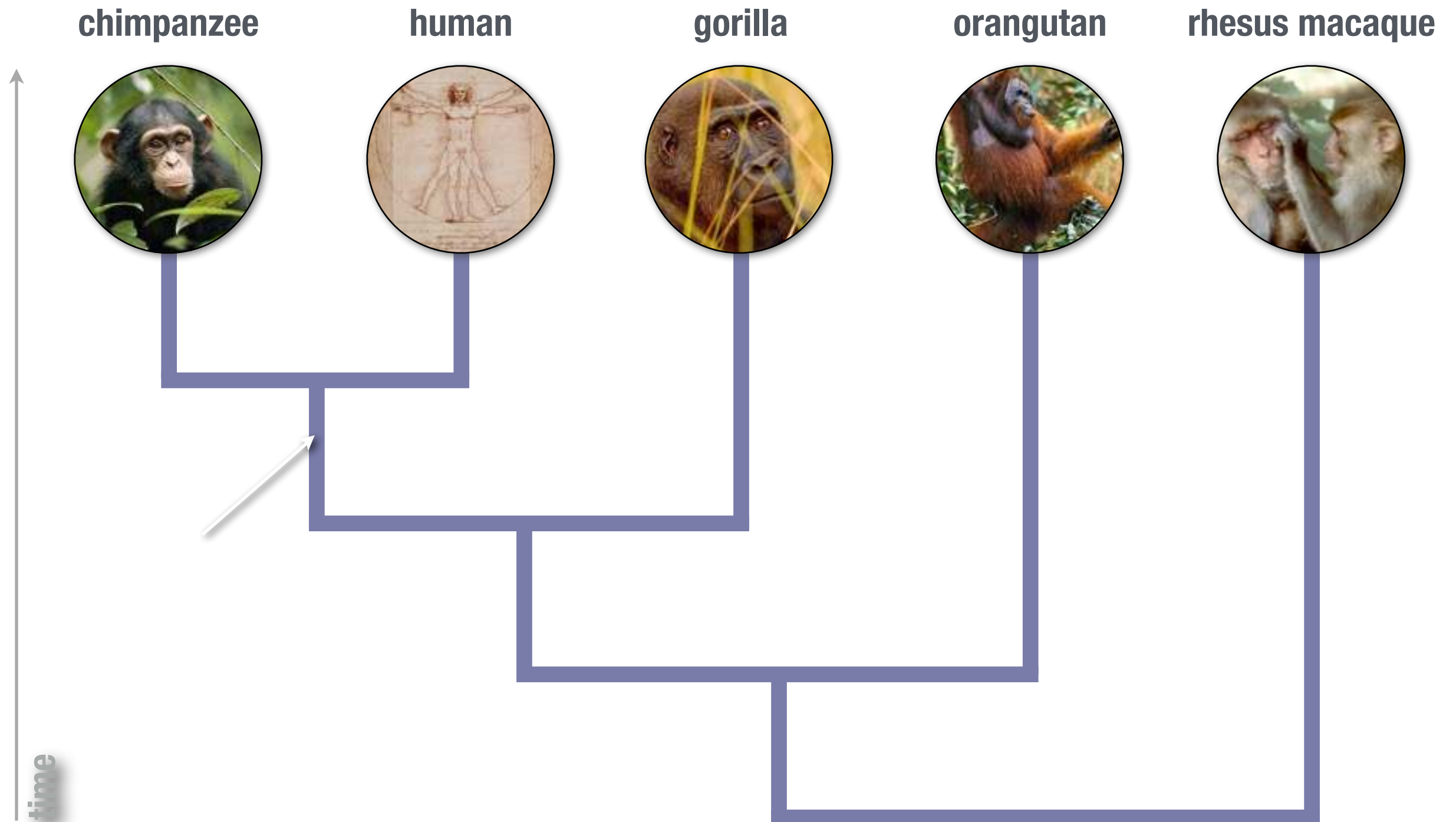
Phylogeny

What is a phylogeny?

- ▶ A **phylogney=phylogenetic tree=evolutionary tree=tree** is a **branching diagram...**
- ▶ ...that shows the evolutionary relationships between taxa.
- ▶ A **taxon** is a named taxonomic group, *e.g.*:
 - ▶ a family (*Felidae*)
 - ▶ a genus (*Felis*)
 - ▶ a species (*Felis catus*)
 - ▶ a single individual ('Nufta')



What is a phylogeny?

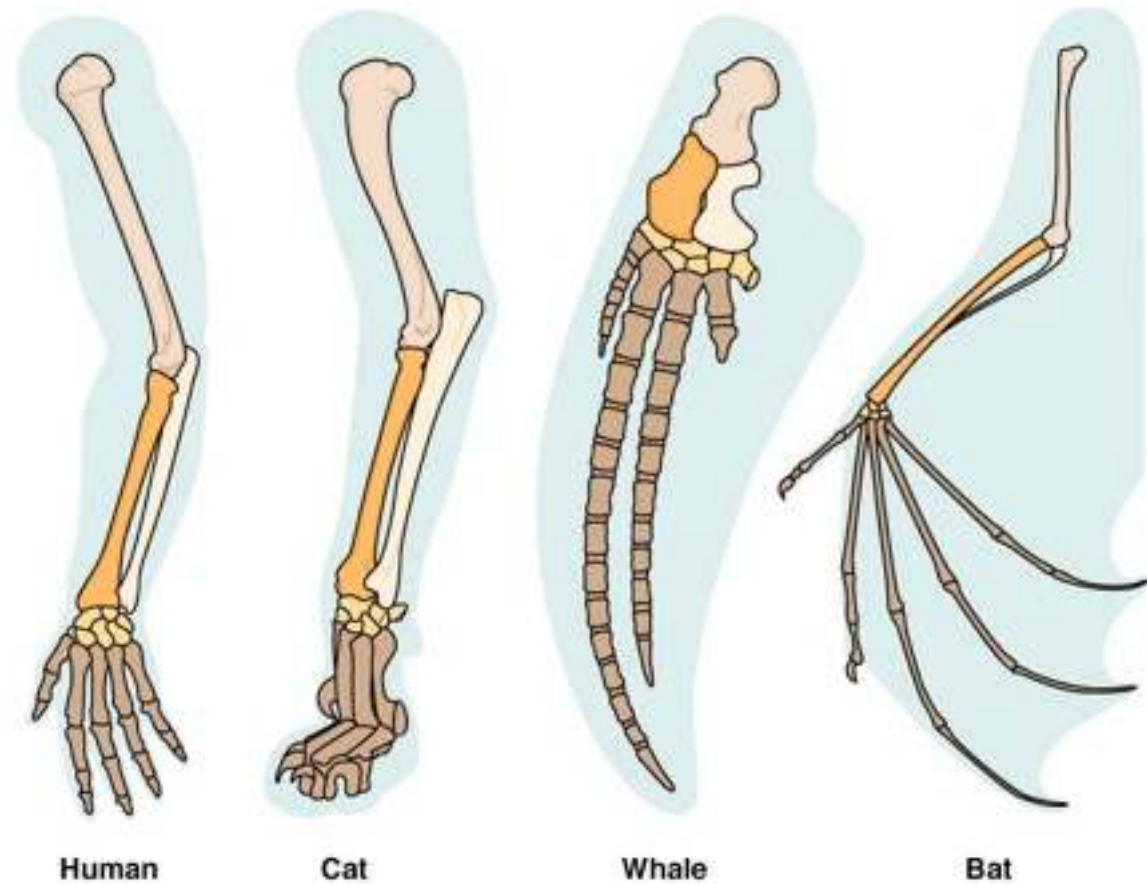


Phylogeny reconstruction

- **Phylogenies** are inferred from **homologous characters**

Homology:

A homology is a character that is shared between two or more species and that was already present in their common ancestor. Hence, (morphological) similarity between homologous characters is due to common ancestry.



©1999 Addison-Wesley Longman, Inc.

Phylogeny reconstruction

- **Phylogenies cannot be inferred from homoplasious characters**

Homoplasy:

A homoplasy is a character that is shared between two or more species and that was **not** present in their common ancestor. Similarity between homoplasious characters is **not** due to shared ancestry (but due to e.g., chance or similar selection regimes).

placental



flying squirrel



gray wolf

marsupial



flying possum



Tasmanian tiger

images: Encyclopaedia Britannica

Phylogeny reconstruction

► Homology can be inferred by several **criteria**

structural similarity	Homologous characters are likely to have the same fundamental structure .
relations between parts	Homologous characters usually have the same relations to surrounding characters (e.g., homologous bones are connected in a similar way with their neighboring bones).
embryonic development	Homologous characters are likely to have the same embryonic development . A character that looks similar in adult forms but has a different developmental history, is unlikely to be homologous.
intermediate forms	Characters are likely to be homologous, if they can be linked through intermediate forms (e.g., homology between swim bladder of fishes and tetrapod lungs).

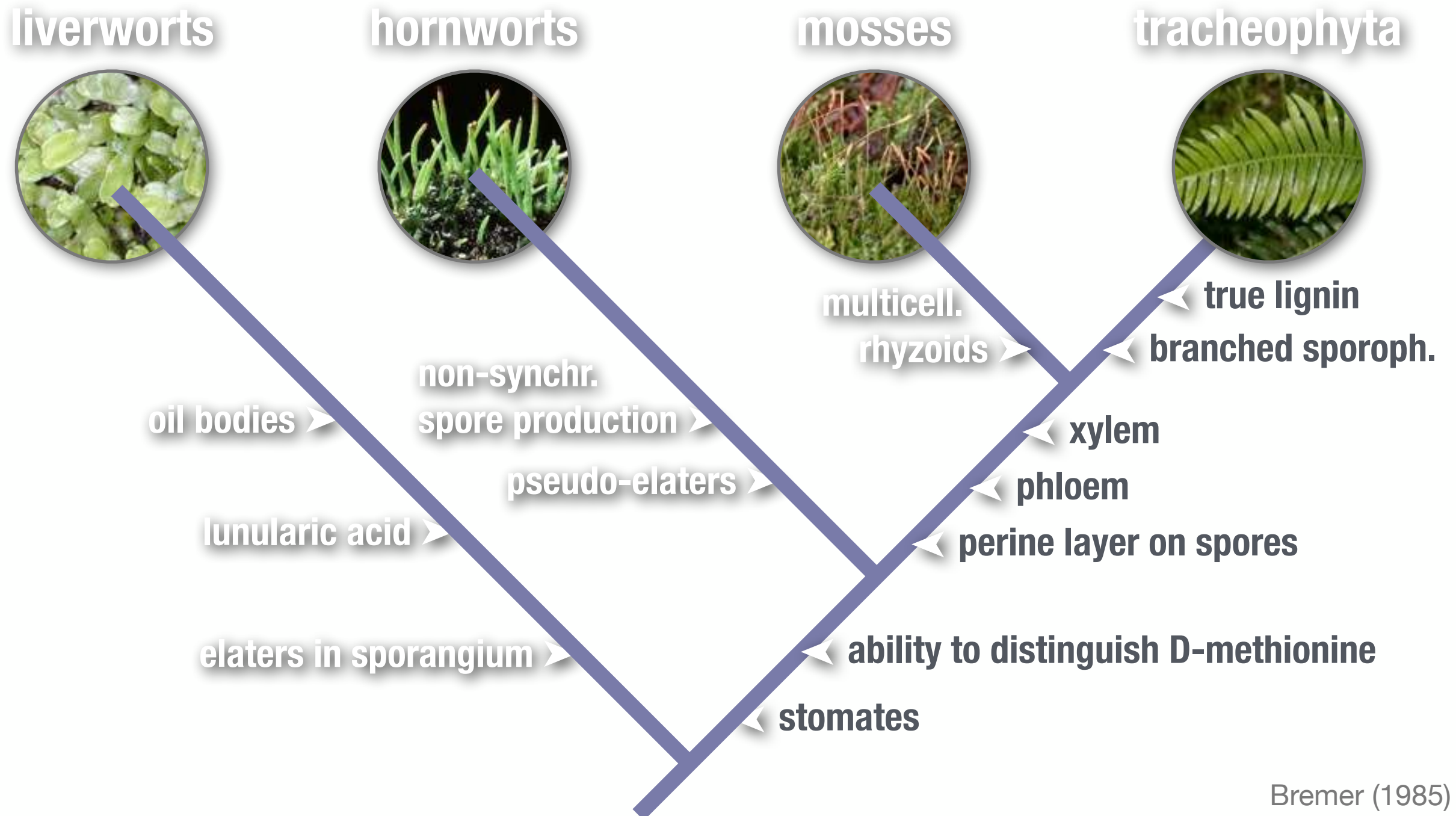
Phylogeny reconstruction

- ▶ Phylogenies can be inferred from **morphological characters** using **cladistic techniques***:
 - ▶ “morphological” refers to all observable characters in an organism: meristic and morphometric measurements; physiological, reproductive, behavioral characteristics, etc.
 - ▶ cladistic techniques are (usually) the only way to analyze fossil data
- ▶ Nowadays, phylogenies are inferred from **genetic data** using **phylogenetic techniques***:
 - ▶ genetic data are homologous by nature (“the ultimate homologous data”)
 - ▶ “Molecular phylogenetics” uses sophisticated statistical techniques based on discrete models of DNA sequence evolution

*Some of the analyses used are the same for morphological and genetic data

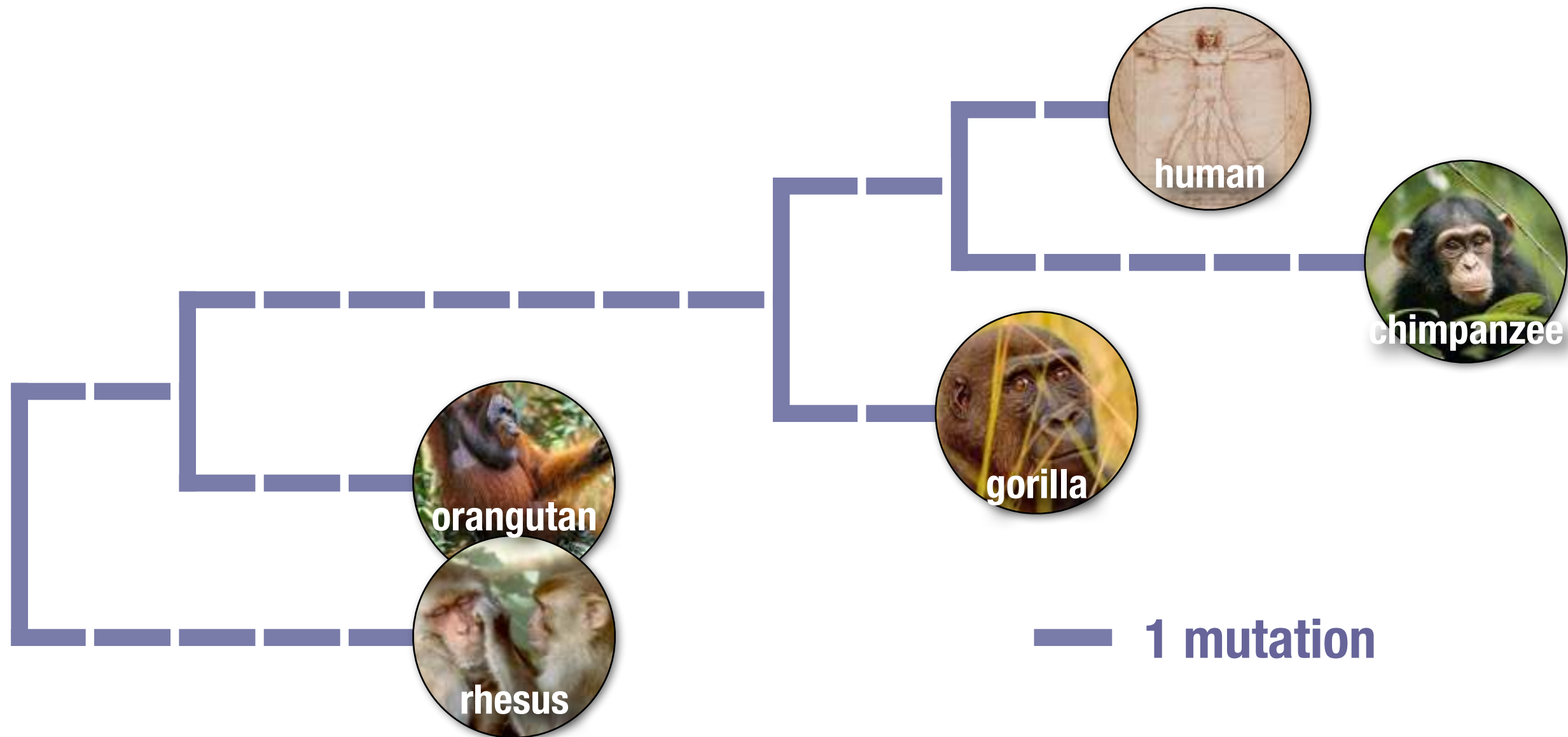
Phylogeny reconstruction

- Phylogeny of plants based on **morphological characters**



Phylogeny reconstruction

- Phylogeny of higher primates based on **genetic characters**



phylogeny based on DNA sequences of the *foxp2* gene (Enard et al. 2002)

