Genomes and Evolution

Shinichi Morishita

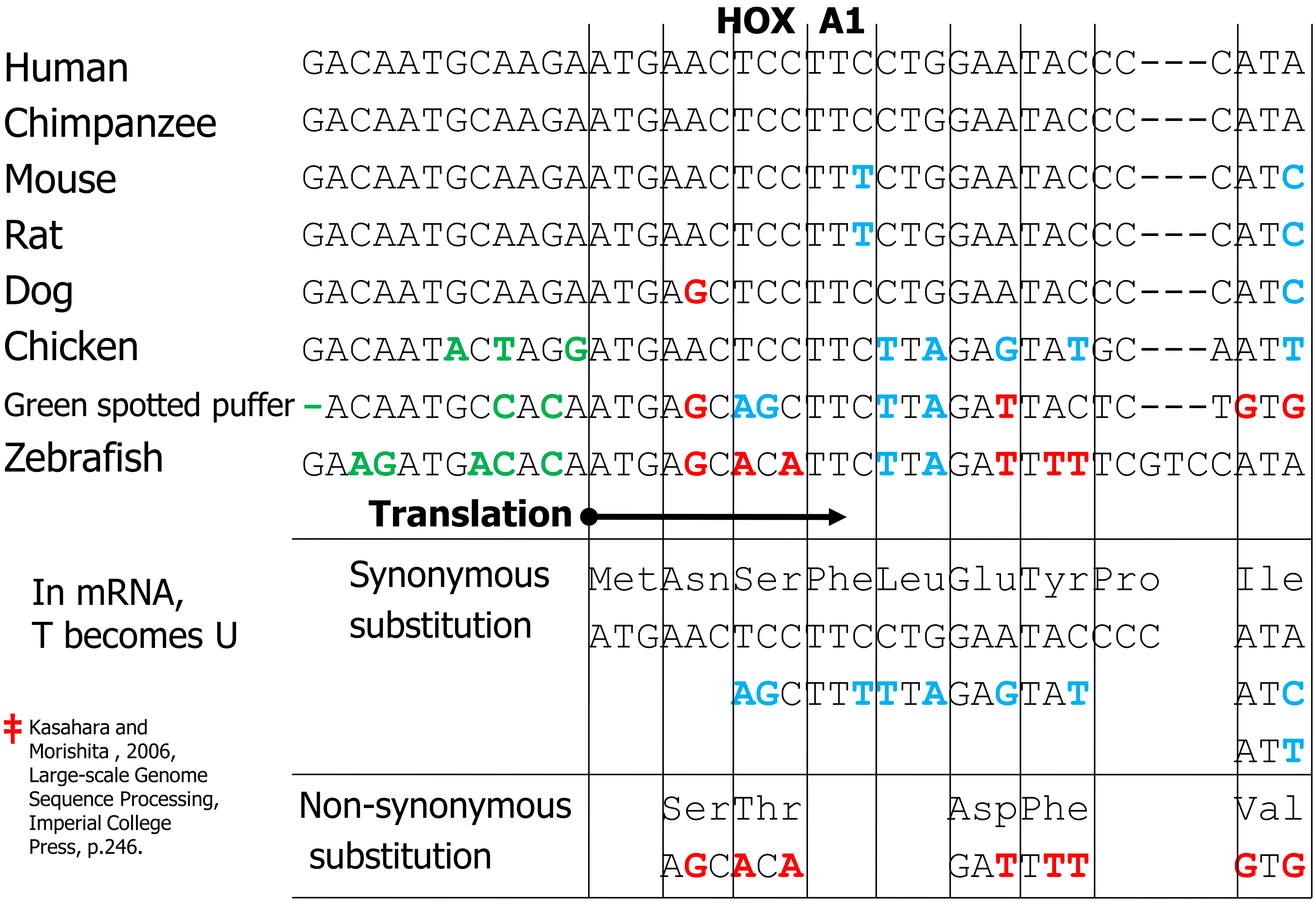
The University of Tokyo
Graduate School of Frontier Sciences,
Department of Computational Biology
Faculty of Science, Bioinformatics & Systems Biology

Evolution at Different Levels

- Nucleotide substitution, insertion & deletion
  Darwin's theory of natural selection
  Kimura's neutral theory
- Region insertion & deletion
  Transposons
- Whole-genome duplication
  Chromosome reorganization
- Ex. HOX genes
  Transcription factors that induce segmental structure
- Four HOX clusters present in mammals

Figure removed due to copyright restrictions

Molecular Biology of the Cell - Fifth Edition
Garland Science (2008)
Figure 22-46



HOX A1

Human	GACAATGCAAGAATGAACTCCTTCCTGGAATACCC---CATA
Chimpanzee	GACAATGCAAGAATGAACTCCTTCCTGGAATACCC---CATA
Mouse	GACAATGCAAGAATGAACTCCTTTCTGGAATACCC---CATC
Rat	GACAATGCAAGAATGAACTCCTTTCTGGAATACCC---CATC
Dog	GACAATGCAAGAATGAGCTCCTTCCTGGAATACCC---CATC
Chicken	GACAATACTAGGATGAACTCCTTCTTAGAGTATGC---AATT
Green spotted puffer	-ACAATGCCACAATGAGCAGCTTCTTAGATTACTC---TGTG
Zebrafish	GAAGATGACACAATGAGCACATTCTTAGATTTTTCGTCCATA

Translation →

In mRNA,
T becomes U

Synonymous substitution	Met	Asn	Ser	Phe	Leu	Glu	Tyr	Pro	Ile
	ATG	AAC	TCC	TTC	CTG	GAA	TAC	CCC	ATA
			AGC	TTT	TTA	GAG	TAT		ATC
									ATT
Non-synonymous substitution		Ser	Thr			Asp	Phe		Val
		AGC	ACA			GAT	TTT		GTG

ǂ Kasahara and Morishita , 2006, Large-scale Genome Sequence Processing, Imperial College Press, p.246.

- Harmful mutations are eliminated by natural selection. (*Darwin's theory of natural selection, 1859*)
- DNA retains not only mutations advantageous to survival. Rather, neutral mutations with no effect are predominant. (*Motoo Kimura's neutral theory, 1968*)

HOX A1

```
Human                 GACAATGCAAGAATGAACTCCTTCCTGGAATACCC---CATA
Chimpanzee            GACAATGCAAGAATGAACTCCTTCCTGGAATACCC---CATA
Mouse                 GACAATGCAAGAATGAACTCCTTTCTGGAATACCC---CATC
Rat                   GACAATGCAAGAATGAACTCCTTTCTGGAATACCC---CATC
Dog                   GACAATGCAAGAATGAGCTCCTTCCTGGAATACCC---CATC
Chicken               GACAATACTAGGATGAACTCCTTCTTAGAGTATGC---AATT
Green spotted puffer  -ACAATGCCACAATGAGCAGCTTCTTAGATTACTC---TGTG
Zebrafish             GAAGATGACACAATGAGCACATTCTTAGATTTTTCGTCCATA
```

```
Human       GACAATGCAAGAATGAACTCCTTCCTGGAATACCC---CATA
            ||xx|||xx|x|||||x|x|x|||x|x||x|xxx|   ||||
Zebrafish   GAAGATGACACAATGAGCACATTCTTAGATTTTTCGTCCATA

Blowfish    -ACAATGCCACAATGAGCAGCTTCTTAGATTACTC---TGTG
             |xx|||x|||||||||||xx||||||||||xx||   xx|x
Zebrafish   GAAGATGACACAATGAGCACATTCTTAGATTTTTCGTCCATA
```

ǂ *Kasahara and Morishita , 2006, Large-scale Genome Sequence Processing, Imperial College Press, p.246*

The Fossil Record and Molecular Clocks

Pauling and Zuckerkandl, 1962

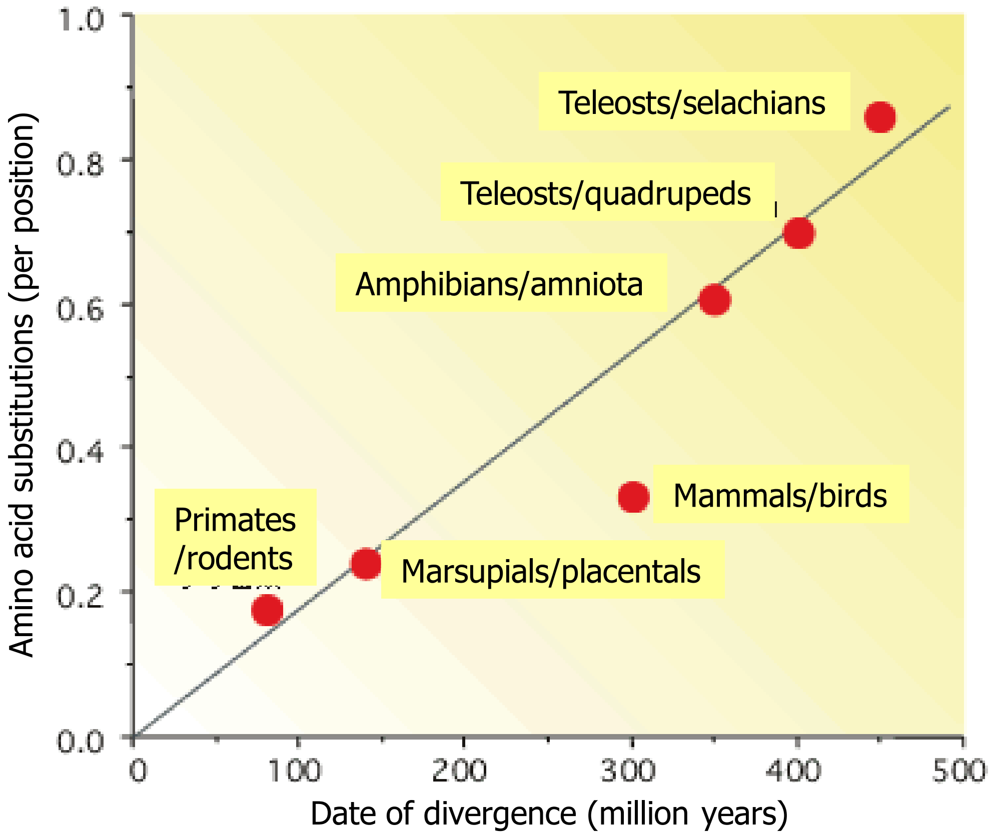
- Studied the number of substitutions in the 141 amino acids in hemoglobin α chains.
- Fossil record indicates strong correlation between time of divergence and rate of substitution.
- Suggested that the rate of substitution is more or less uniform

  ==> Calculation of date of divergence from rate of substitution

Our current understanding is that:

- The molecular clock is not necessarily accurate.
- Because the mutation rate varies with lineage and is not uniform.

	Gorilla	Horse	Dog	Newt	Carp
Human	1	18	23	62	68



‡ Takashi Miyata, *Shinka no Hanashi* ("On Evolution"), JT Biohistory Research Hall
http://www.brh.co.jp/katari/shinka/shinka09.html

Using Mitochondrial DNA

- Almost entirely matrilineal, making it easy to trace ancestry. (Some patrilineal.)
- c.16,500 base pairs in all
- Human and chimpanzee (diverged c.5 million years ago)

  Differences in mitochondrial DNA genes of about 9%

  (in nuclear DNA, about 1%)

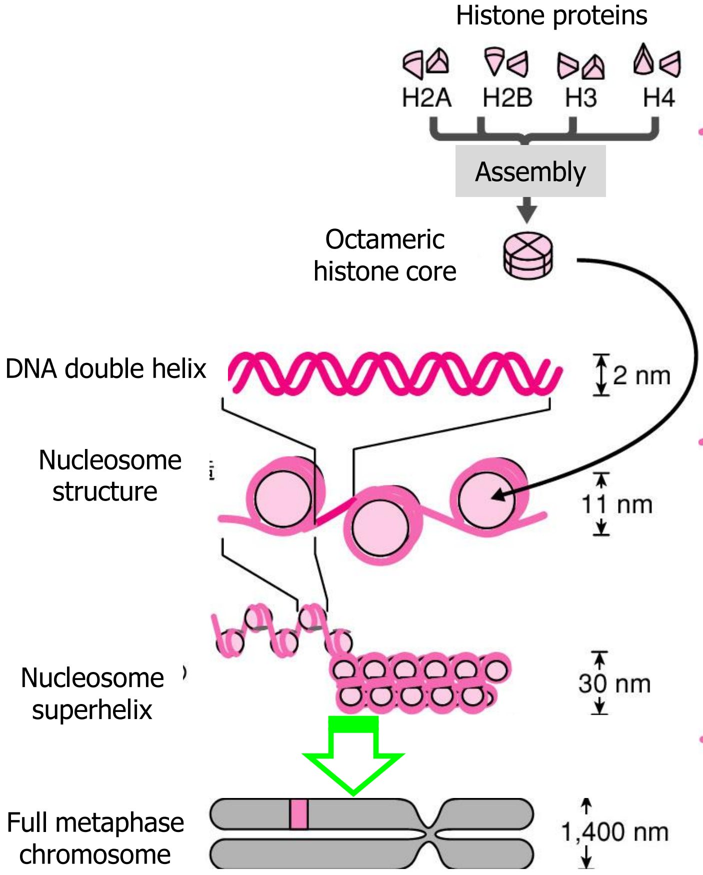
  Authoritative calculation of date of divergence in tens of thousands of years
- Alan Wilson, 1987

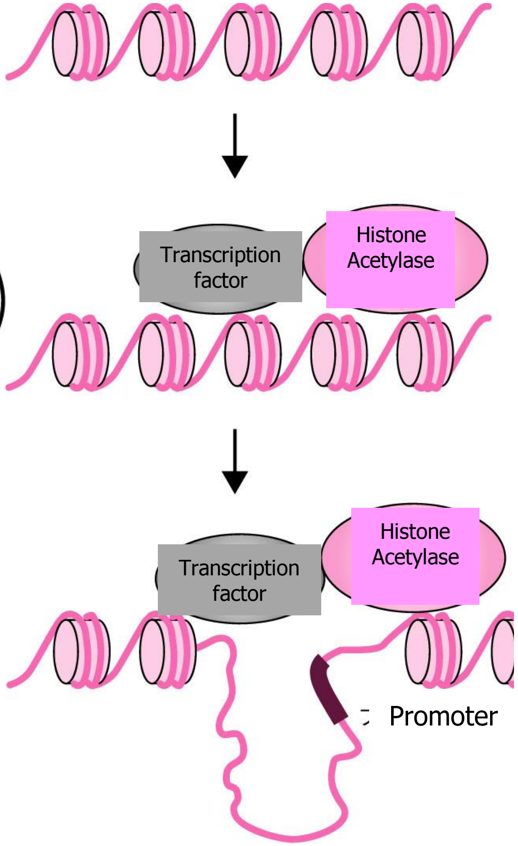
  Produced a genealogical tree from the mitochondria of 147 persons.

  Our ancestor is an African woman from 150,000 to 290,000 years ago (the Mitochondrial Eve hypothesis).
- 1991: 5,000-year-old mummy found in the Alps (the "Iceman").

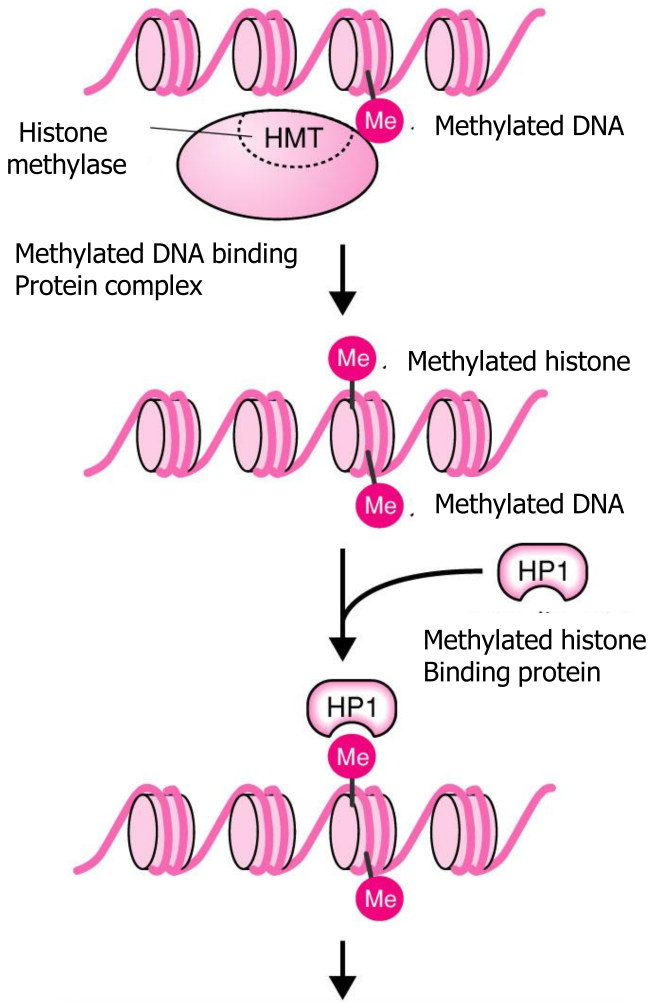
Chromatin Structure, Transcription and Evolution

‡ Figures from *Seimei Kagaku* ("Life Sciences"), 2nd rev. ed., Yodosha, 2008, Figures 4-6, 7, 8





DNA Methylation and Epigenetics

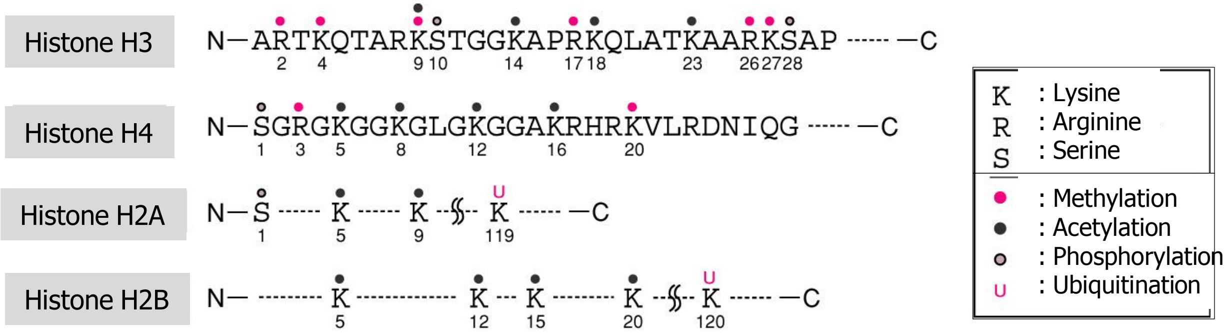


Heterochromatin formation

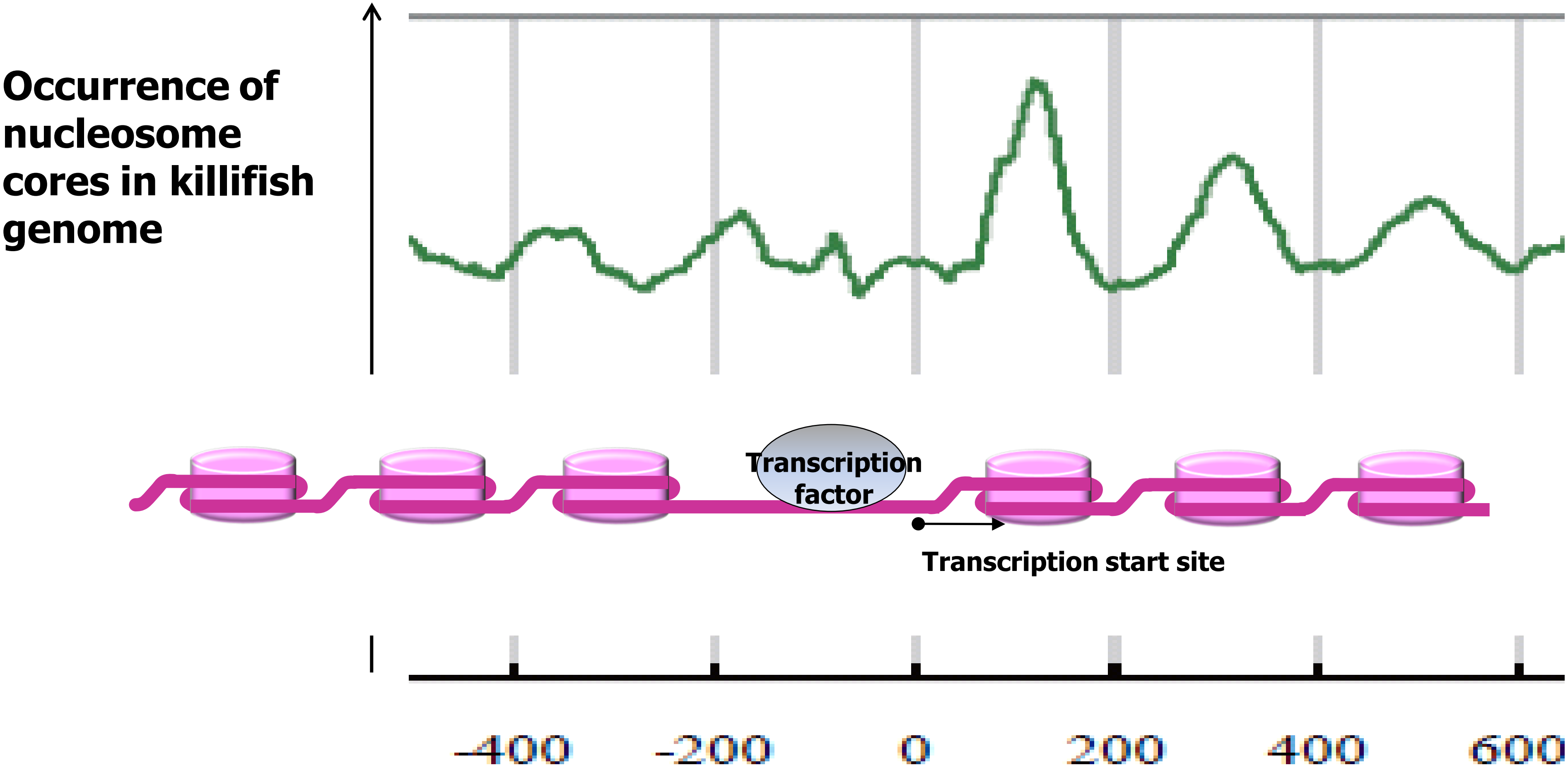
↓

Transcription inactivation

Epigenetic regulation of expression



Chromatin Structure, Transcription and Evolution

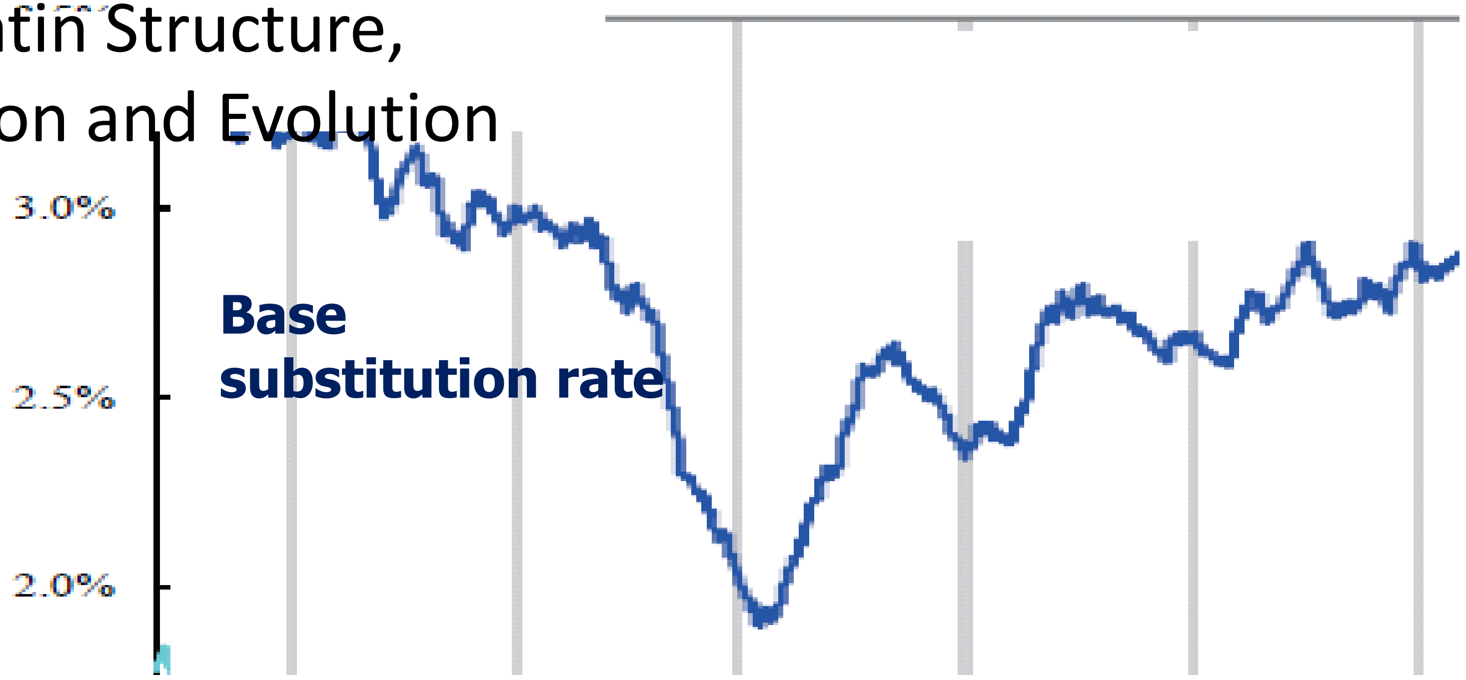


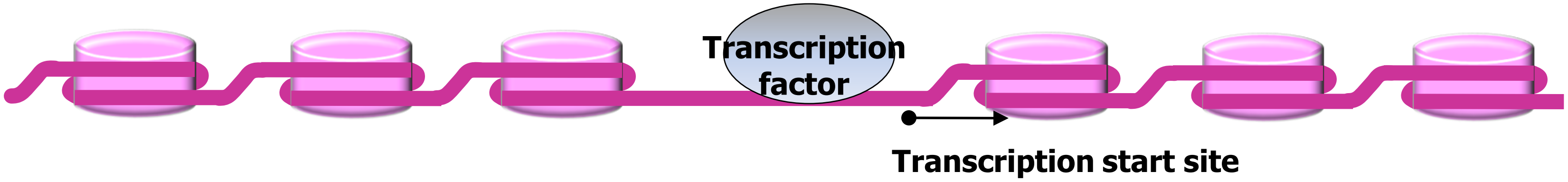
‡

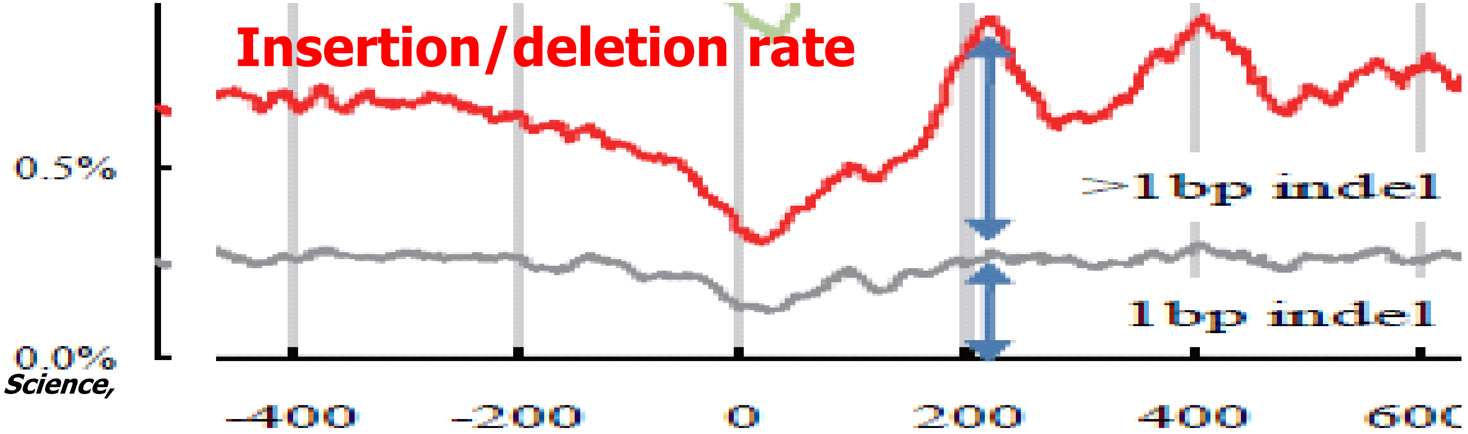
Chart from Sasaki et al., 2009, Science, 323(5912), 401-404

Chromatin Structure, Transcription and Evolution

Genomes of two killifish varieties compared





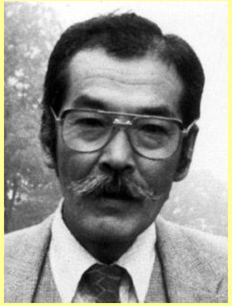


‡

Chart from Sasaki et al., 2009, Science, 323(5912), 401-404

Evolution by Duplication

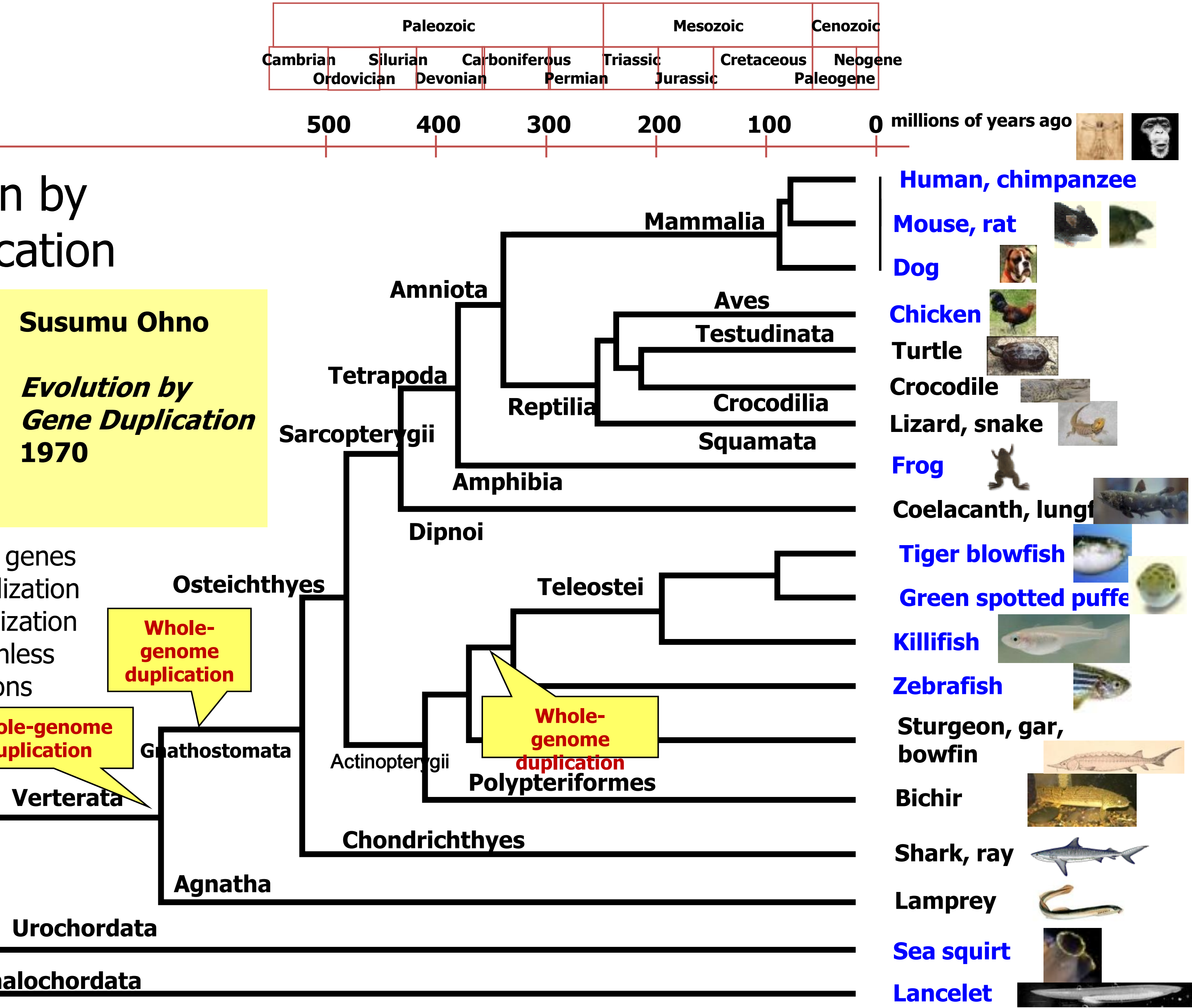
‡



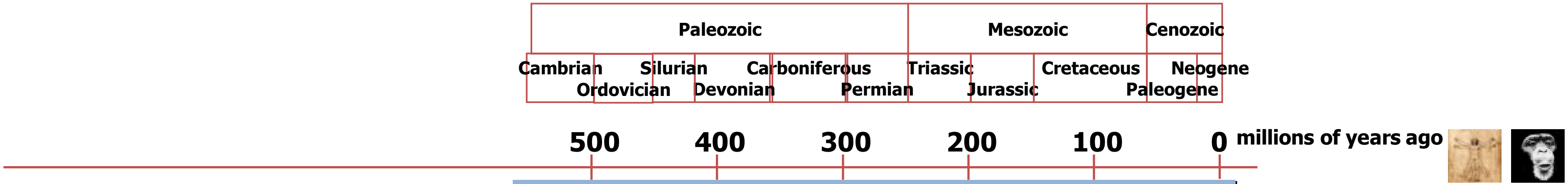
Susumu Ohno

Evolution by Gene Duplication
1970

Enrichment of genes
Neo-functionalization
Sub-functionalization
Mostly functionless due to mutations

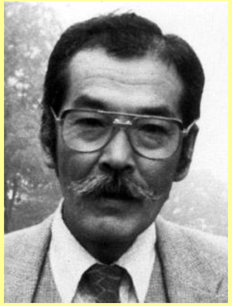


‡ *Nakatani et al., 2007 , Genome Res., 17, 1254-1265*



Evolution by Duplication

‡



Susumu Ohno

Evolution by Gene Duplication
1970

Enrichment of genes
Neo-functionalization
Sub-functionalization
Mostly functionless due to mutations

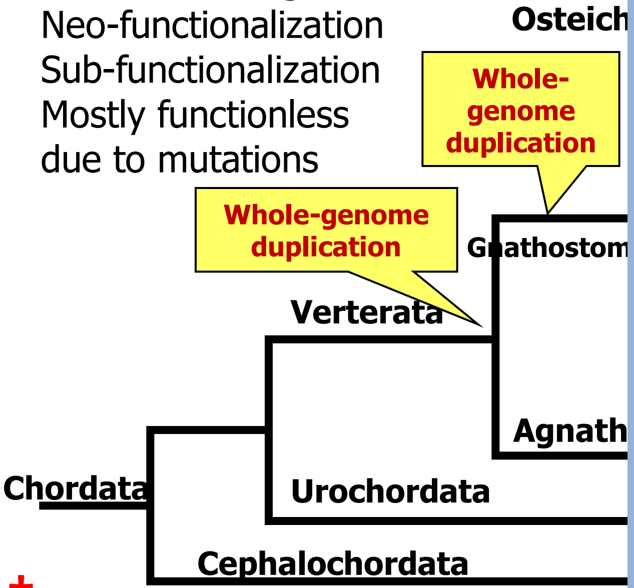


Figure removed due to copyright restrictions

Molecular Biology of the Cell - Fifth Edition
Garland Science (2008)
Figure 22-46

‡ *Nakatani et al., 2007 , Genome Res., 17, 1254-1265*

Inferences from Hox Genes (1997)

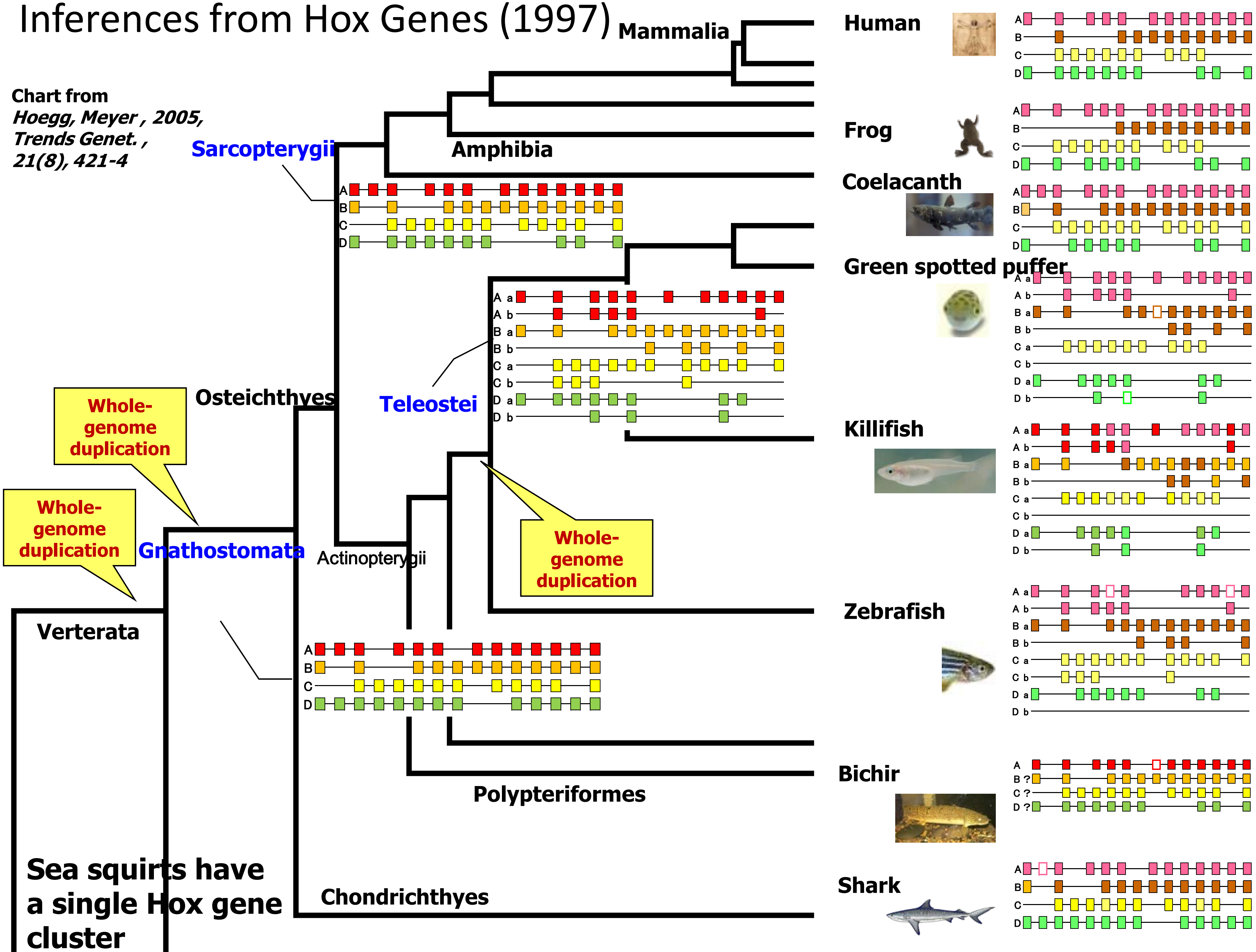


Chart from
Hoegg, Meyer , 2005,
Trends Genet. ,
21(8), 421-4

Whole-genome duplication

Whole-genome duplication

Whole-genome duplication

Sea squirts have a single Hox gene cluster

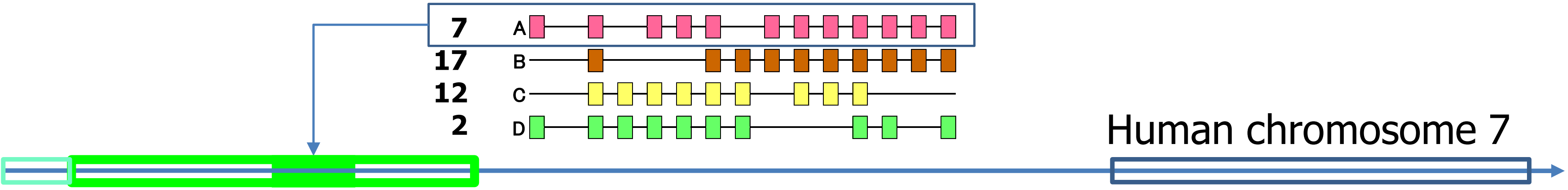
Can Whole-Genome Duplication Be Confirmed?

- The human genome has up to 300,000 base pairs where Hox genes have clustered.

- This represents a tiny fraction of the entire human genome (around three billion base pairs).

- Compare entire genomes.

Traces of Two Whole-Genome Duplications Detected in Vertebrate Lineage



The chart plots chromosomal paralogue (genes with similar sequences) found of genes in human chromosome 7.

Figure removed due to copyright restrictions

Nakatani et al., 2007 , Genome Res., 17, 1254-1265
supplemental research data

Human Chromosome

Figure removed due to copyright restrictions

Nakatani et al., 2007 , Genome Res., 17, 1254-1265
supplemental research data

Killifish Chromosome

Figure removed due to copyright restrictions

Nakatani et al., 2007 , Genome Res., 17, 1254-1265
supplemental research data

The chart plots killifish orthologues (genes with similar sequences) found of genes in human chromosome 7.

Figure removed due to copyright restrictions

Nakatani et al., 2007,
***Genome Res.*, 17, 1254-1265**
Figure4

Kasaahra M et al. Nature 447, 714-719 (2007)
Nakatani Y et al. Genome Res 17(9), 1254-1265 (2007).

Verterata (10)

Gnathostomata (40)

Osteichthyes (31)

Amniota (26)

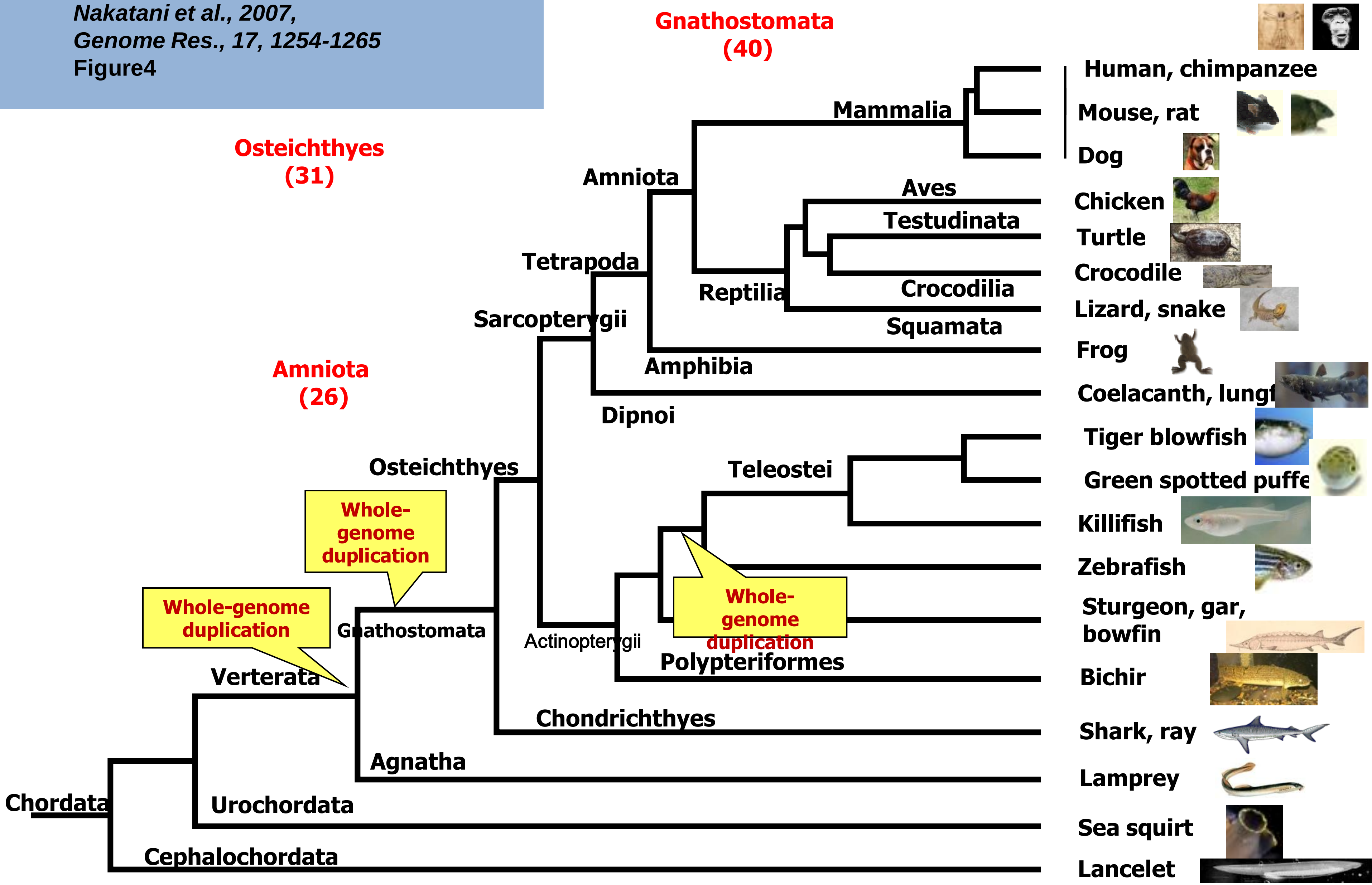


Figure removed due to copyright restrictions

Nakatani et al., 2007,
Genome Res., 17, 1254-1265
Figure4

Kasaahra M et al. *Nature* 447, 714-719 (2007)
Nakatani Y et al. *Genome Res* 17(9), 1254-1265 (2007).

Verterata (10)

Gnathostomata (40)

Osteichthyes (31)

Amniota (26)

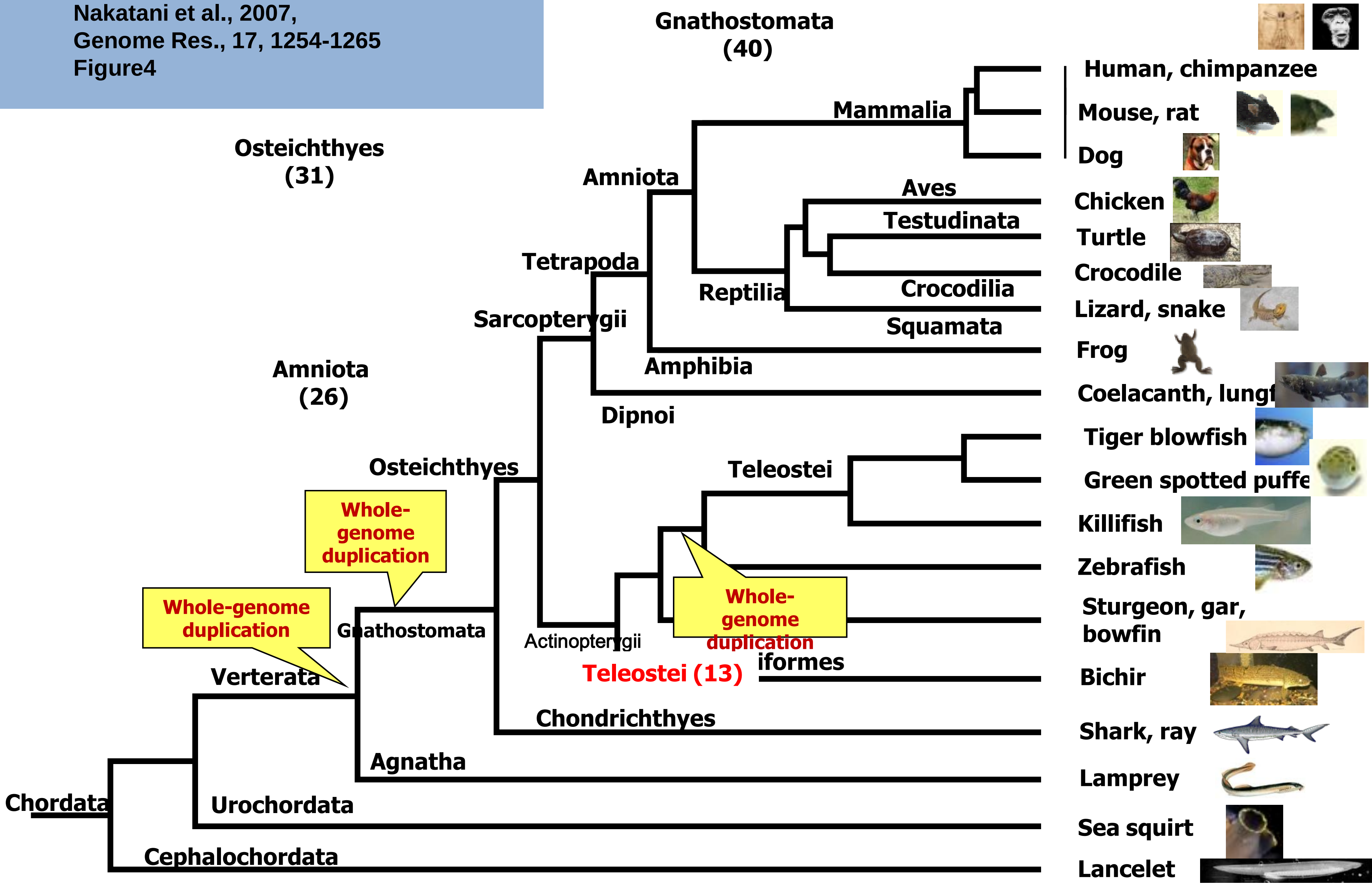


Figure removed due to copyright restrictions

Molecular Biology of the Cell - Fifth Edition
Garland Science (2008)
Figure 4-14

Global Focus on Knowledge Lecture Series
Puzzling Out the System of Life

Genomes and Evolution:

The Path to Humanity

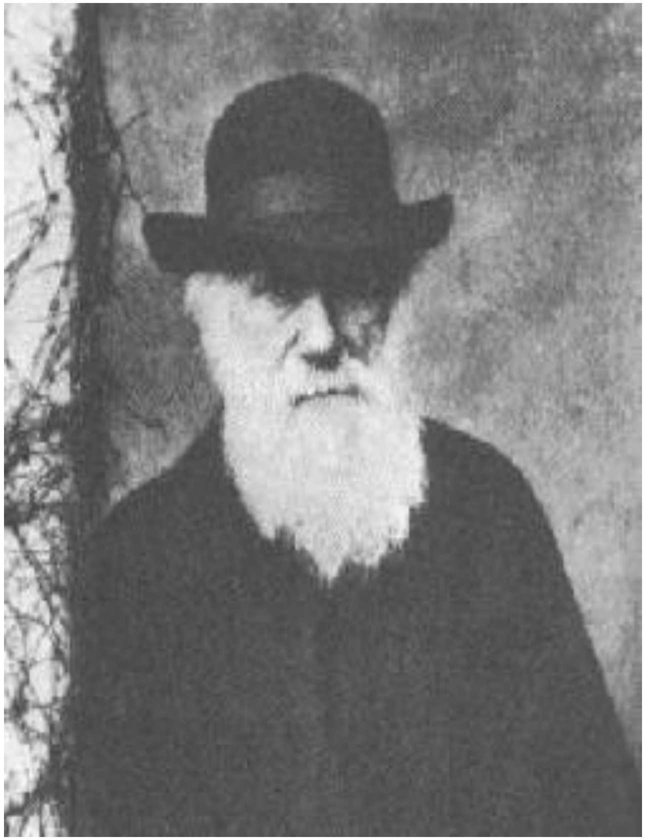
Naruya Saitou

National Institute of Genetics

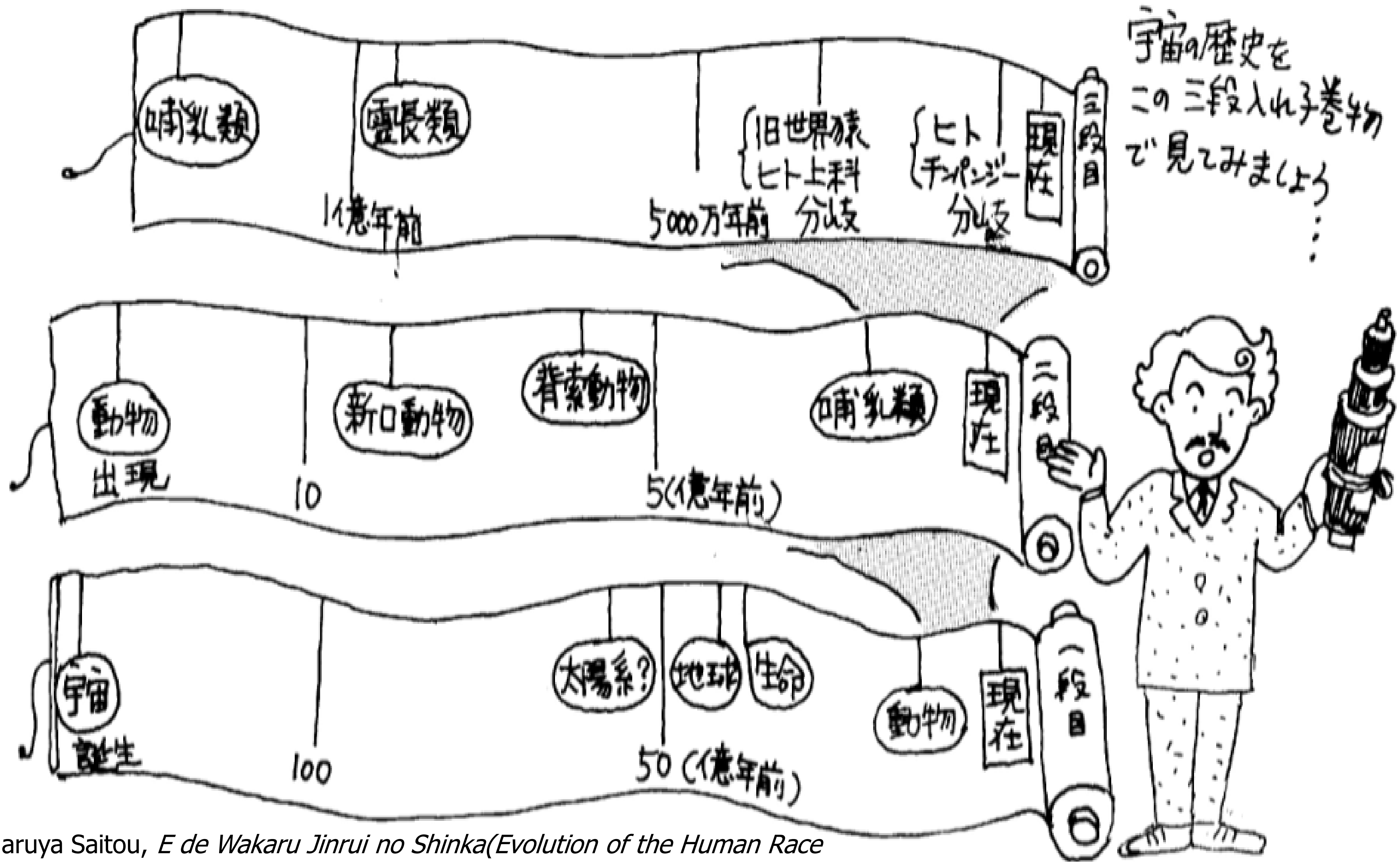
SAITOU
LABORATORY
NIG SOKENDAI DDBJ / MISHIMA

April 22, Year 9 of the Scientific Era
Bldg 18 Lecture Hall, Komaba Campus, The University of Tokyo

Charles Darwin (1809-1882)



A History of Life and the Universe, in Three Volumes



Naruya Saitou, *E de Wakaru Jinrui no Shinka(Evolution of the Human Race which is Understood with Pictures),* Koudan Sha, 2009.

Illustrations: Saori Yasutomi

"It took at most 100 million years for the Earth to cool down."

Lord Kelvin

Alpha rays

Beta rays

Uranium 238 Thorium 234 Protactinium 234

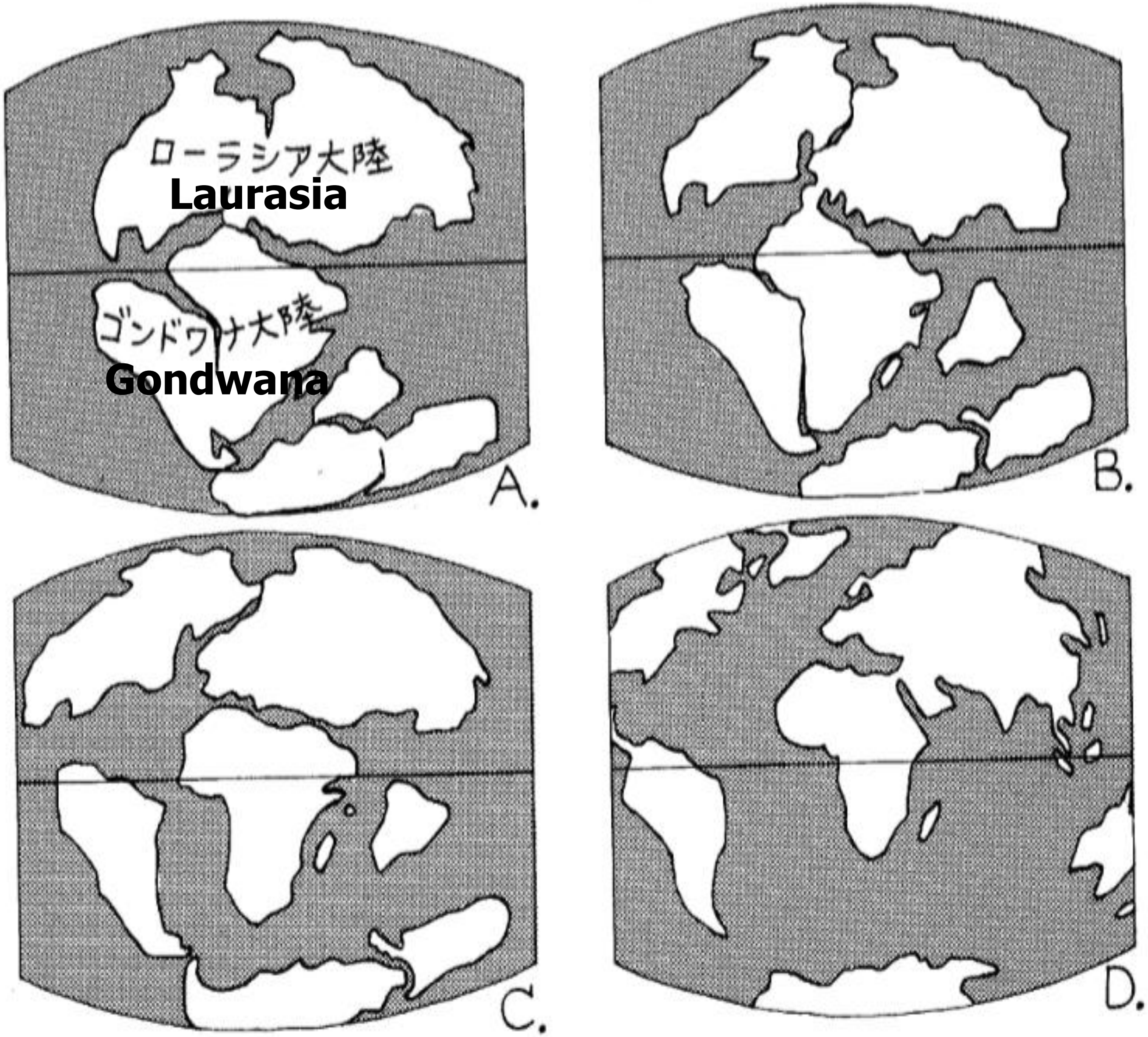
Pierre and Marie Curie

"Given the energy emitted from radioisotopes, it's taken longer than that . . ."

‡ Naruya Saitou, *E de Wakaru Jinrui no Shinka(Evolution of the Human Race which is Understood with Pictures)*, Koudan Sha, 2009.

Illustrations: Saori Yasutomi

200 Million Years of Continental Drift

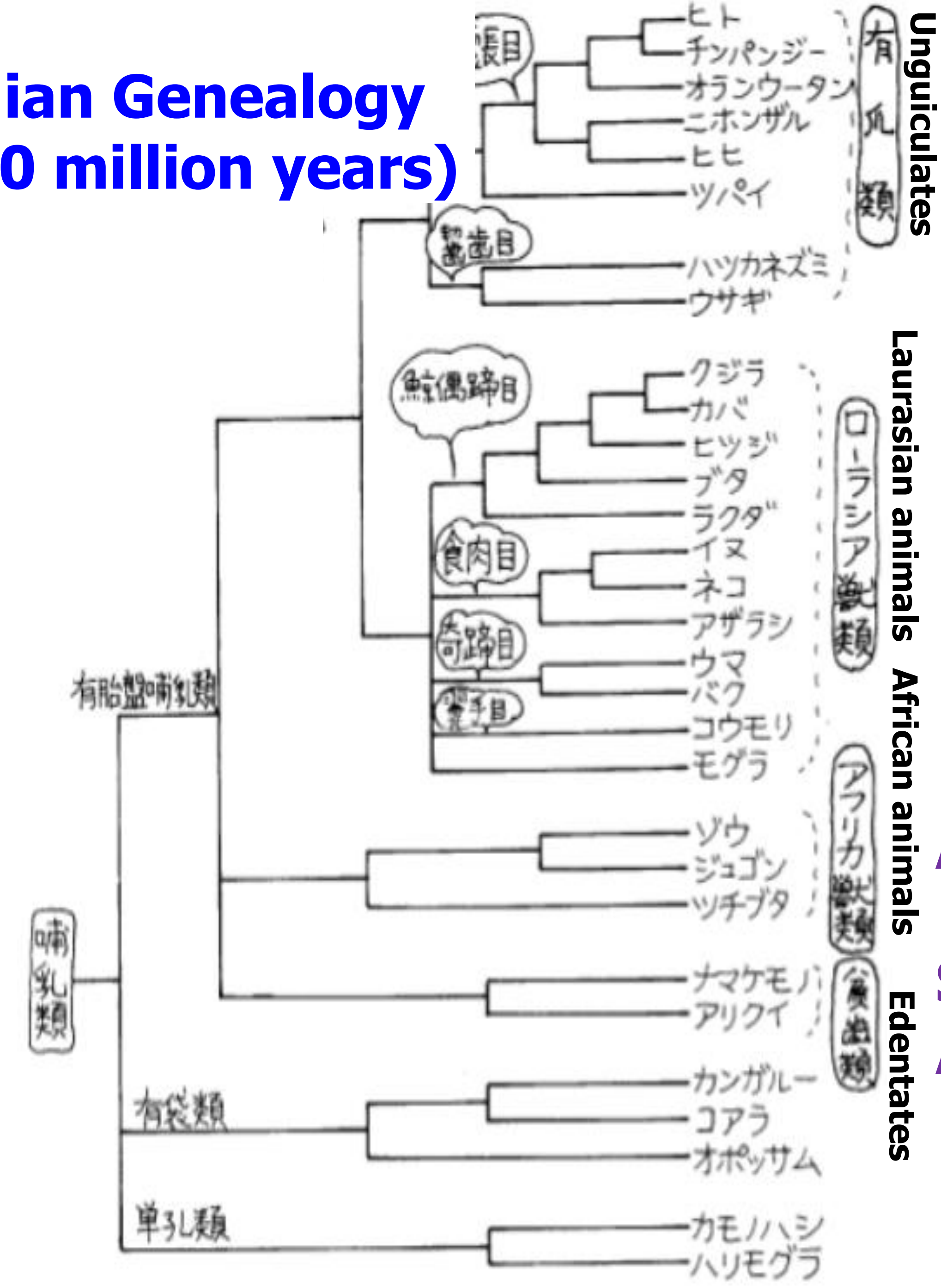


‡

Naruya Saitou, *E de Wakaru Jinrui no Shinka(Evolution of the Human Race which is Understood with Pictures),* Koudan Sha, 2009.

Illustrations: Saori Yasutomi

Mammalian Genealogy (Past 100 million years)



Unguiculates

Laurasian animals

African animals

Edentates

Laurasia

Laurasia

Africa

Gondwana

South America

Australia

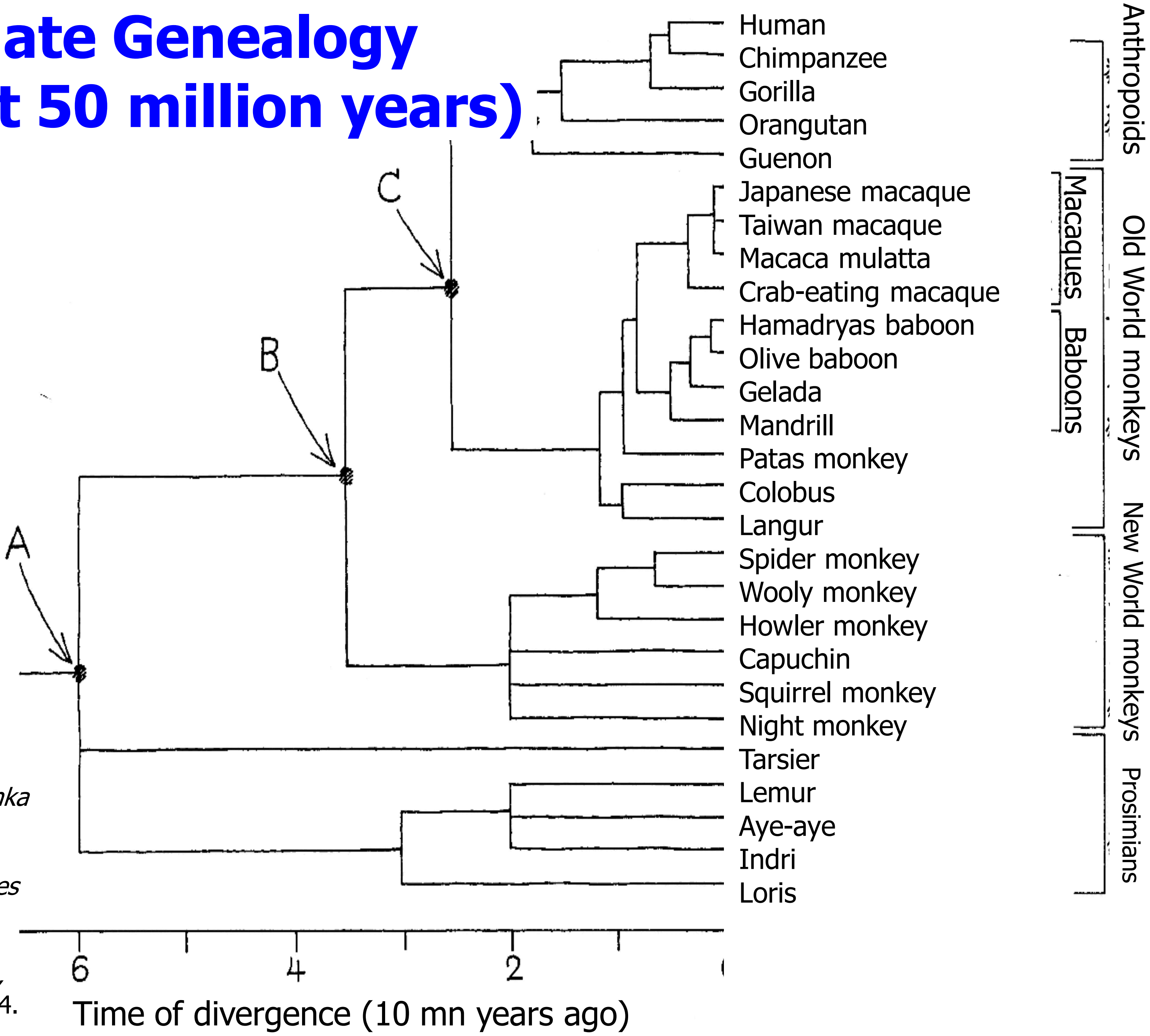
Australia

イラスト:安富佐織

‡

Naruya Saitou, *E de Wakaru Jinrui no Shinka(Evolution of the Human Race which is Understood with Pictures)*, Koudan Sha, 2009.

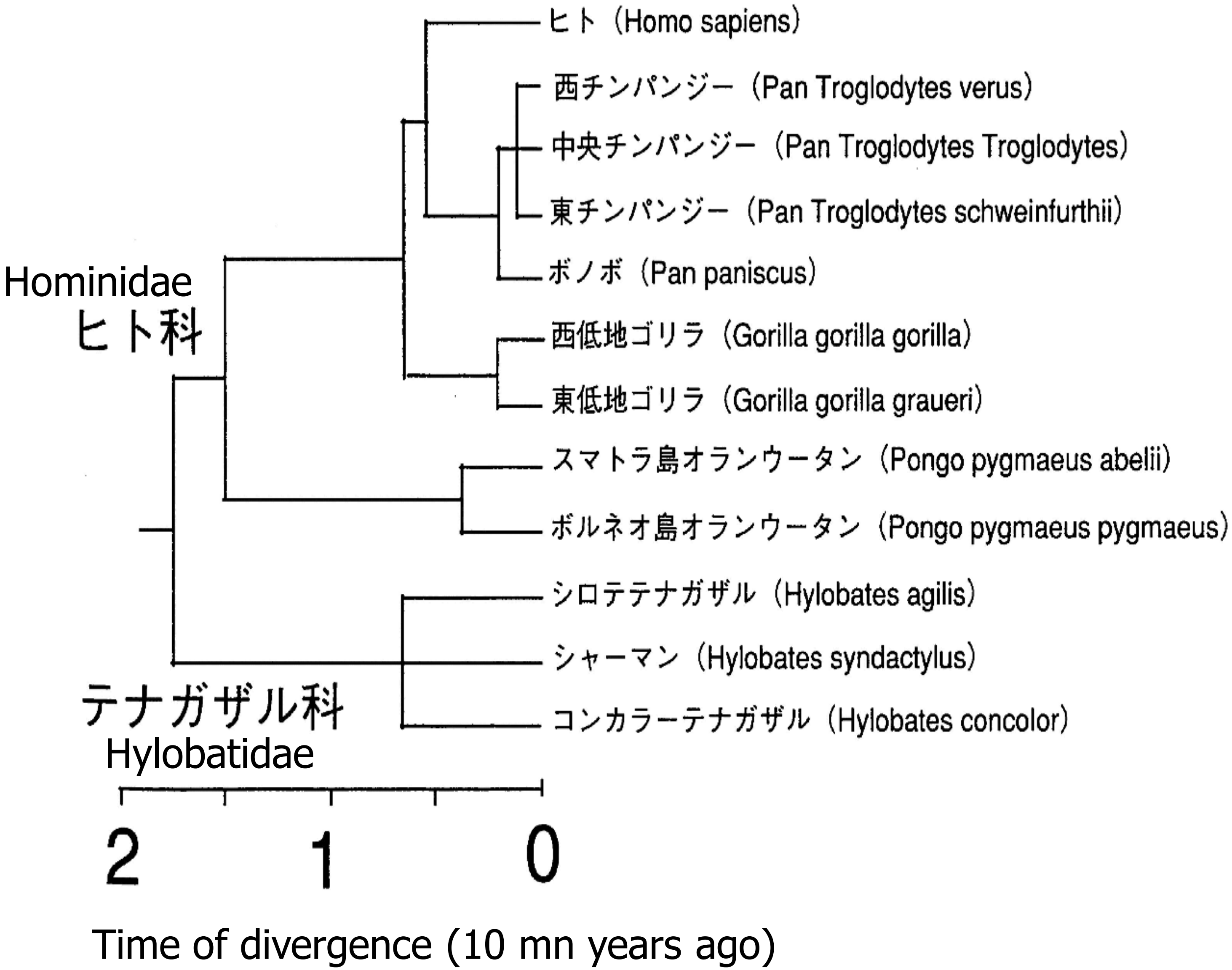
Primate Genealogy (Past 50 million years)



ǂ

Naruya Saitou, *Genomu to Shinka – Genomu kara Tachinoboru Seimei (Genomes and Evolution – Life Ascendent from Genomes),* Shin'yosha, 2004.

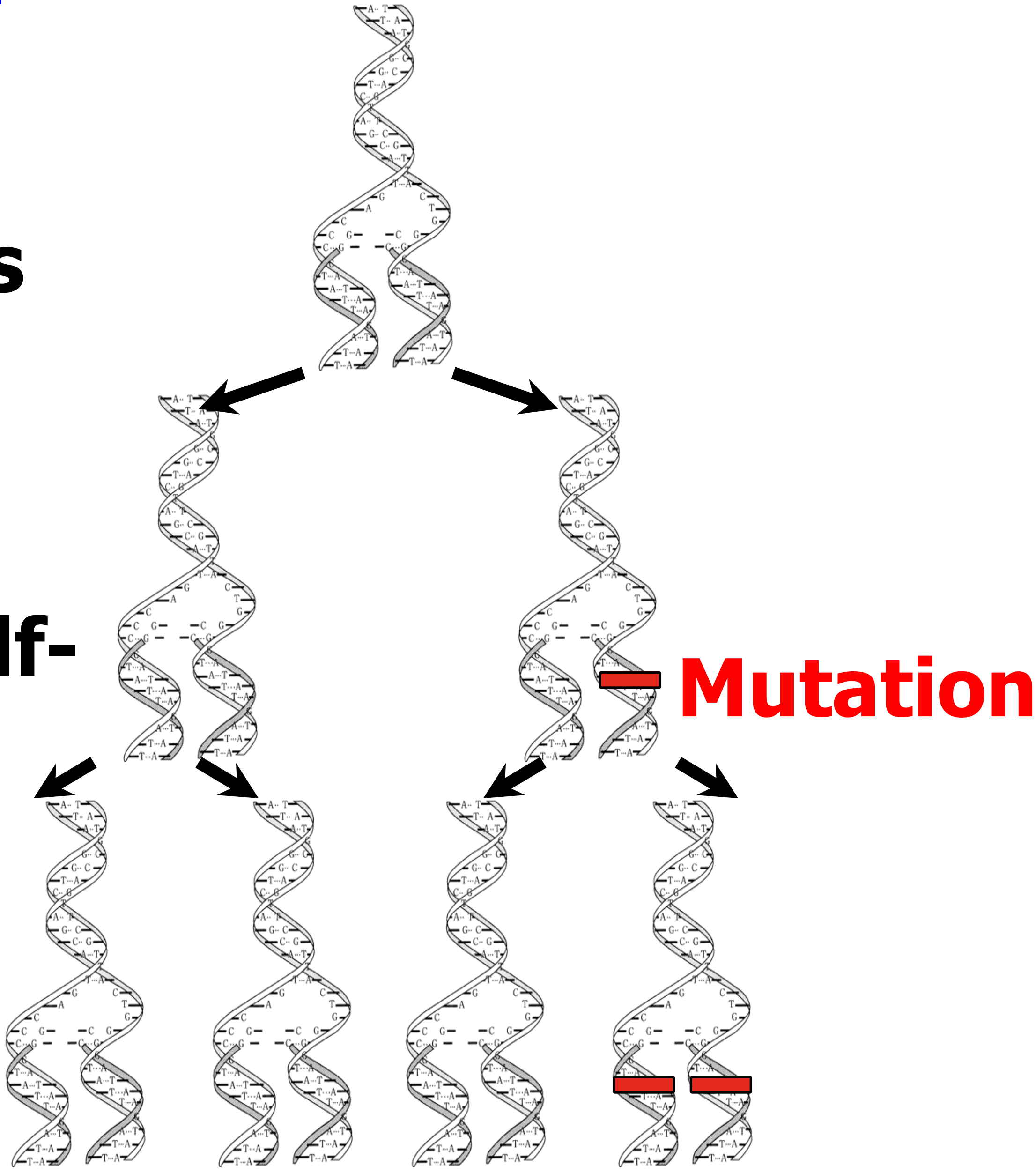
Hominoid Genealogy
(Past 20 million years)



Naruya Saitou, *Genomu to Shinka – Genomu kara Tachinoboru Seimei (Genomes and Evolution – Life Ascendent from Genomes),* Shin'yosha, 2004.

The Heart of Evolution:

DNA accumulates mutations while undergoing repeated self-replication.



Different Micromutations

Nucleotide replacement

Original sequence: ACCTATTTTGCTG

New sequence: ACCTGTTTTGCTG

Nucleotide insertion

Original sequence: ACCTATTTTGCTG

New sequence: ACCAGTTATTTTGCTG

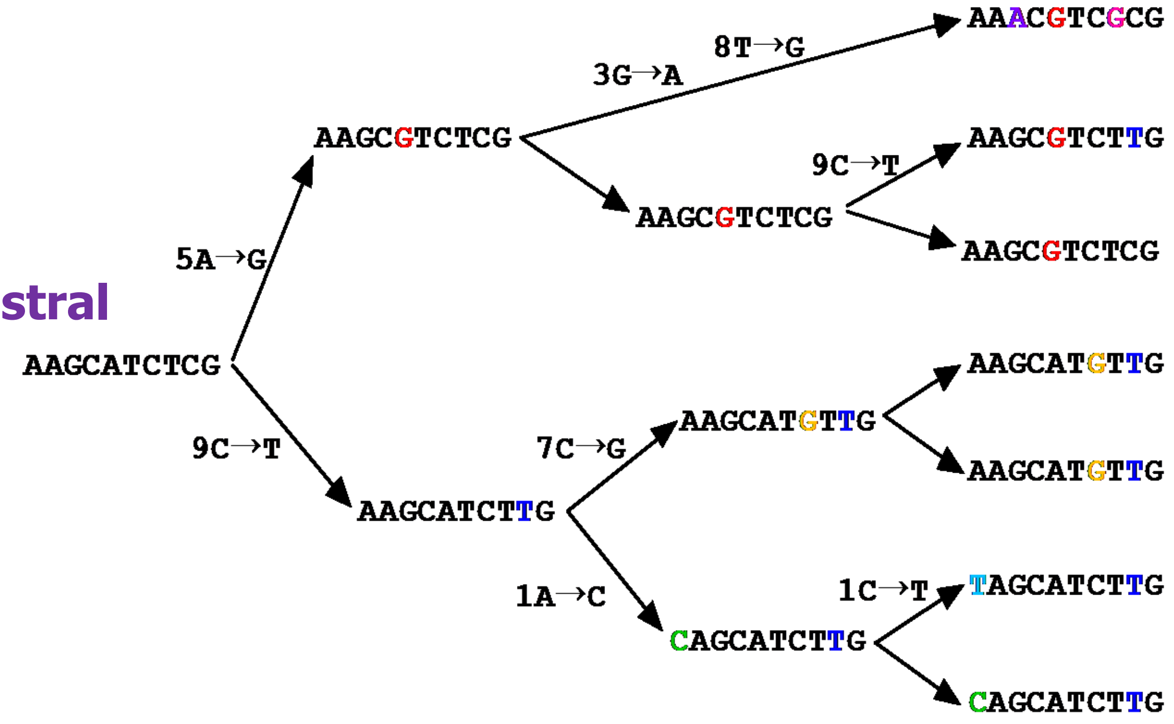
Nucleotide deletion

Original sequence: ACCTATTTTGCTG

New sequence: ACCTATTGCTG

Mutations Accumulating in DNA

DNA of ancestral life form



AAGCATCTCG

5A→G

AAGCGTCTCG

3G→A

8T→G

AAACGTCGCG

AAGCGTCTCG

9C→T

AAGCGTCTTG

AAGCGTCTCG

9C→T

AAGCATCTTG

7C→G

AAGCATGTTG

AAGCATGTTG

AAGCATGTTG

1A→C

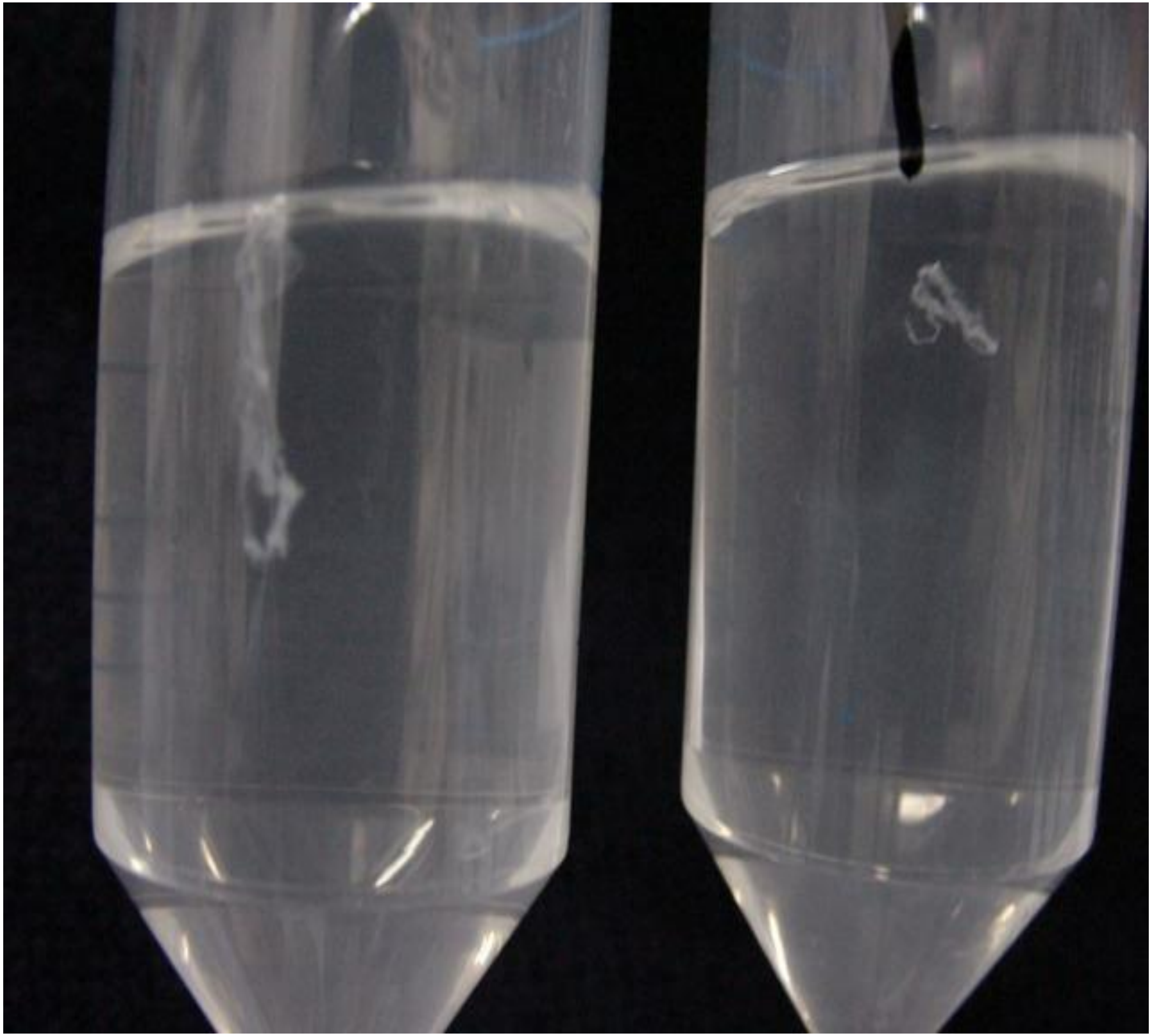
CAGCATCTTG

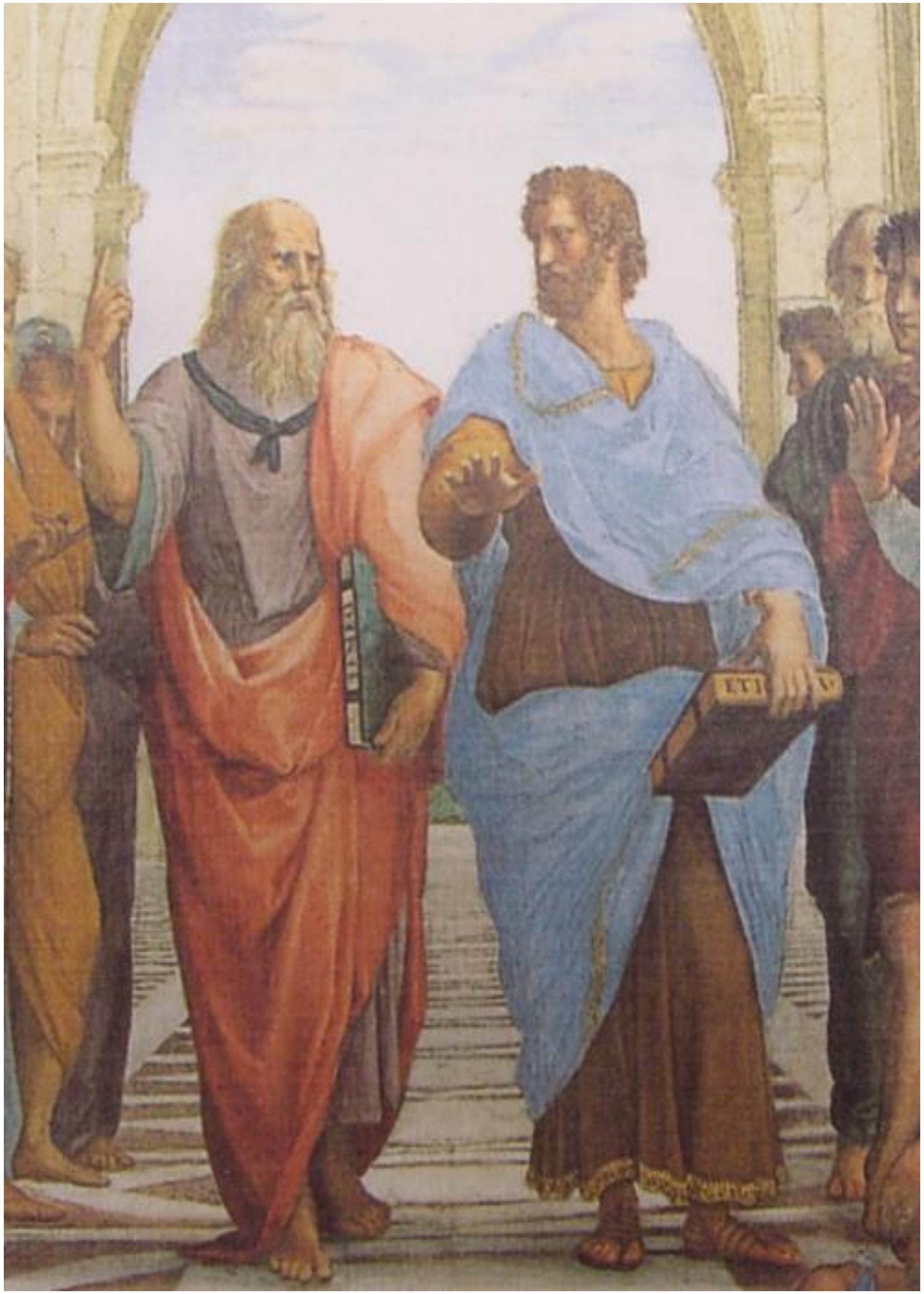
1C→T

TAGCATCTTG

CAGCATCTTG

DNA of current life forms





Matter and Essence

Raphael's *Scuola di Atene,* detail

Left: Plato
Right: Aristotle

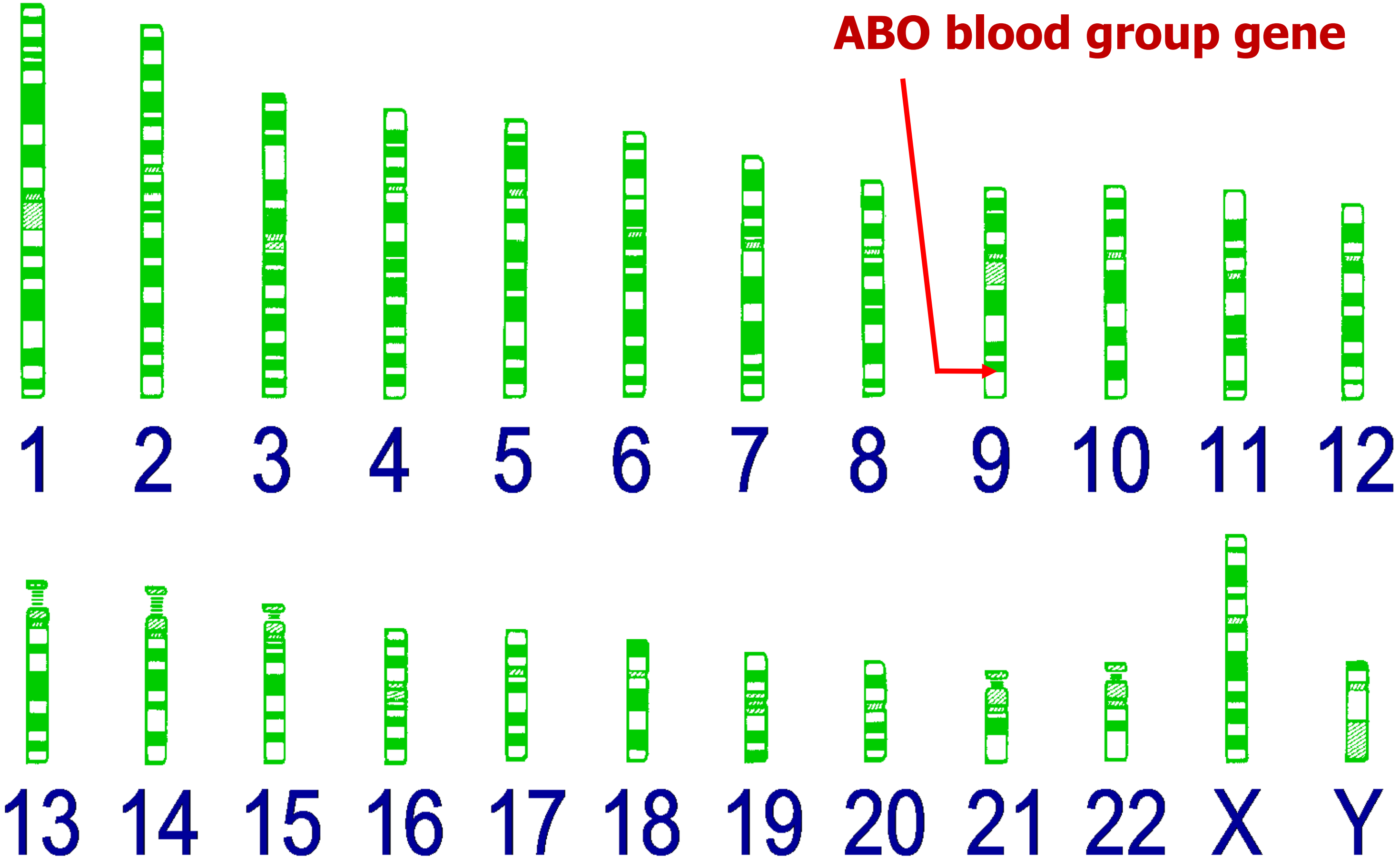
What Is a Genome?

- A genome is the sum of its genes and its chromosomes. (Hans Winkler, 1920)
- Functional definition: The minimal set of genes required for the life of an organism. (Hitoshi Kihara, 1930s)
- Structural definition: The largest self-replicating unit of nucleotide sequences. (Naruya Saitou, 2004)

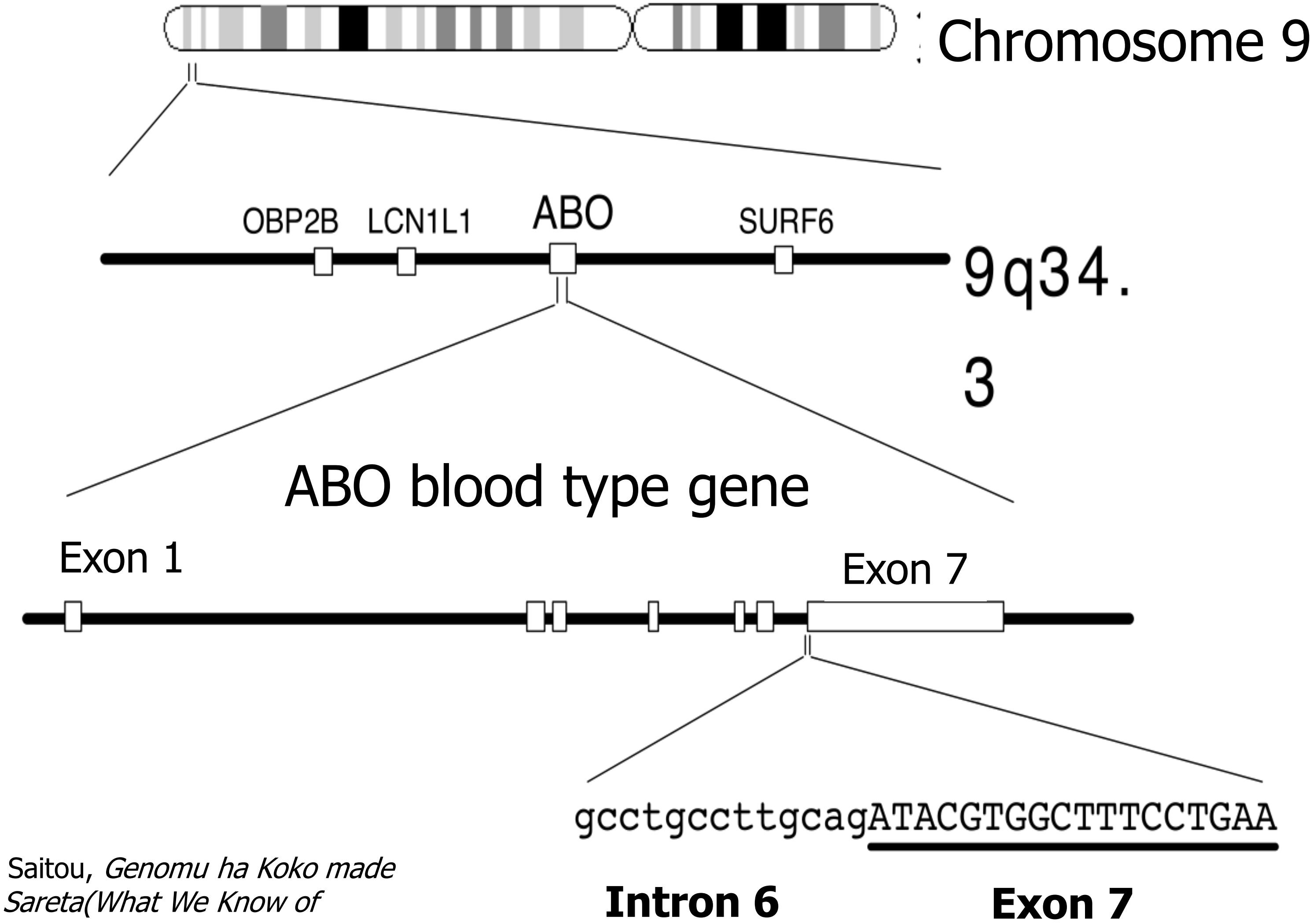
Human Chromosomes

Human genome

: 3 billion nucleotides
22 autosomal chromosomes and
2 sex chromosomes
22,000 to 50,000 genes



From Chromosome to Nucleotide Sequence: The ABO Blood Type Gene



Chromosome 9

OBP2B LCN1L1 ABO SURF6

9q34.3

ABO blood type gene

Exon 1

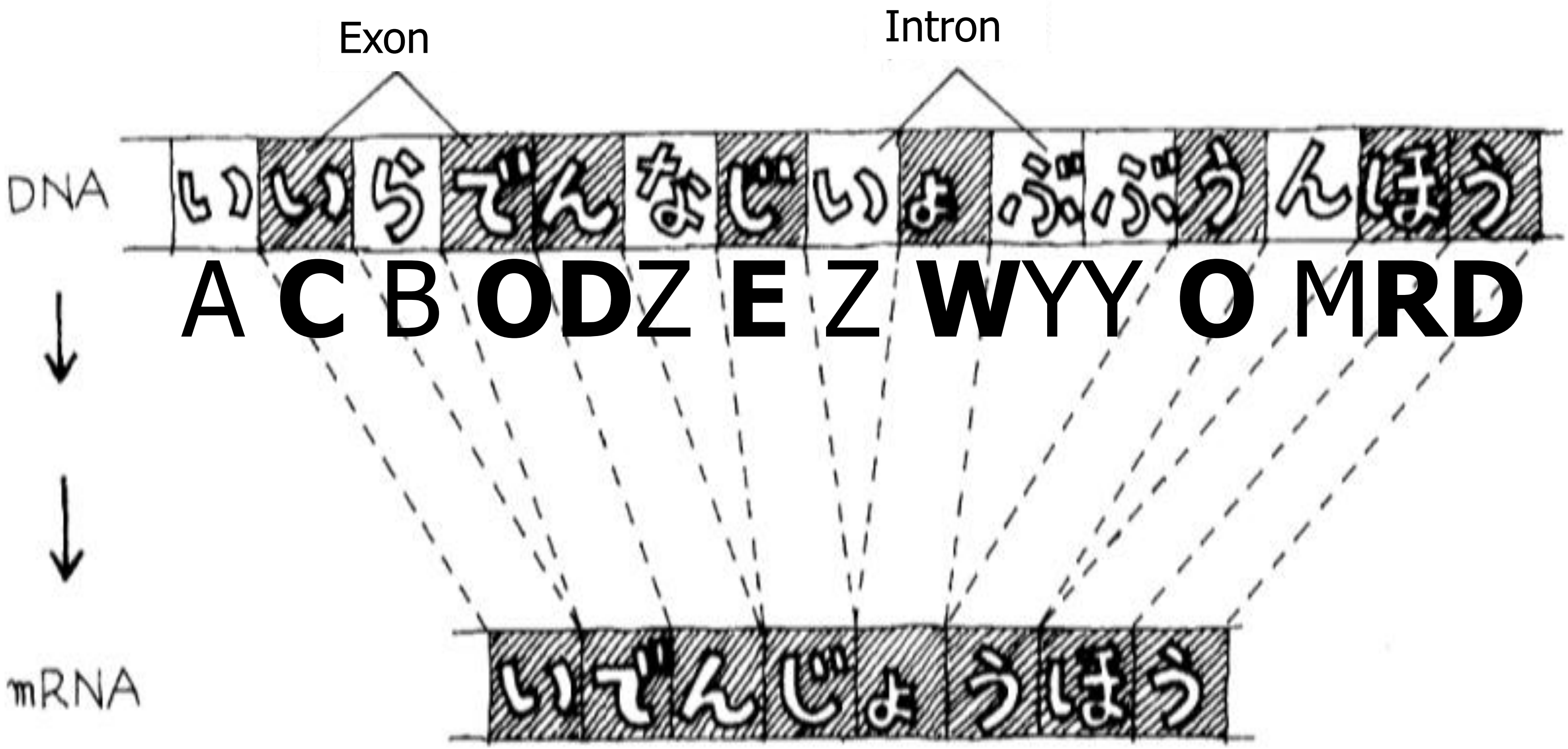
Exon 7

gcctgccttgcagATACGTGGCTTTCCTGAA

Intron 6 **Exon 7**

‡ Naruya Saitou, *Genomu ha Koko made Kaimei Sareta(What We Know of Genomes),* Wedge, 2007.

Splicing (Exon Patching)



A **C** B **OD**Z **E** Z **W**YY **O** M**RD**

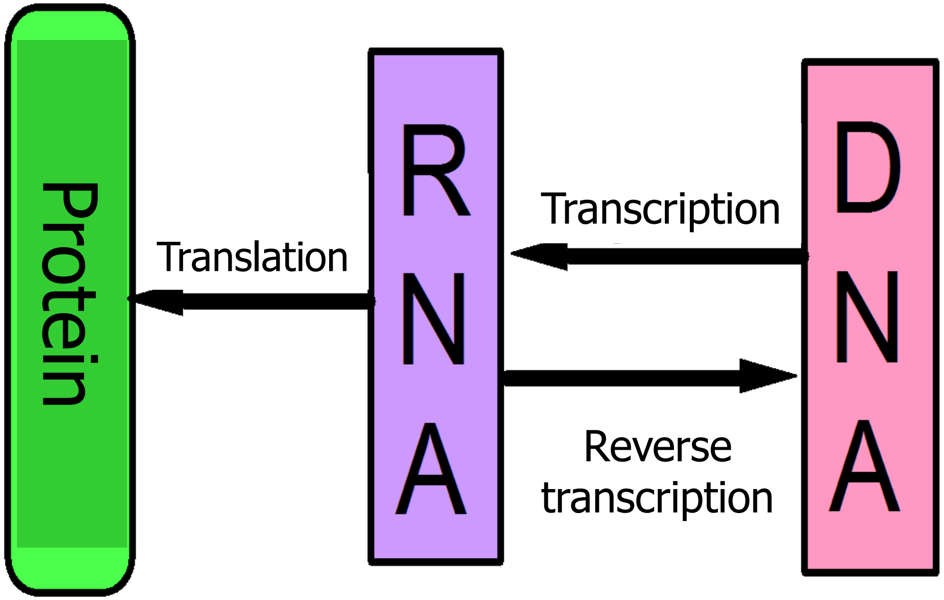
COD E WORD

‡ Naruya Saitou, *E de Wakaru Jinrui no Shinka(Evolution of the Human Race which is Understood with Pictures),* Koudan Sha, 2009.

Illustrations: Saori Yasutomi

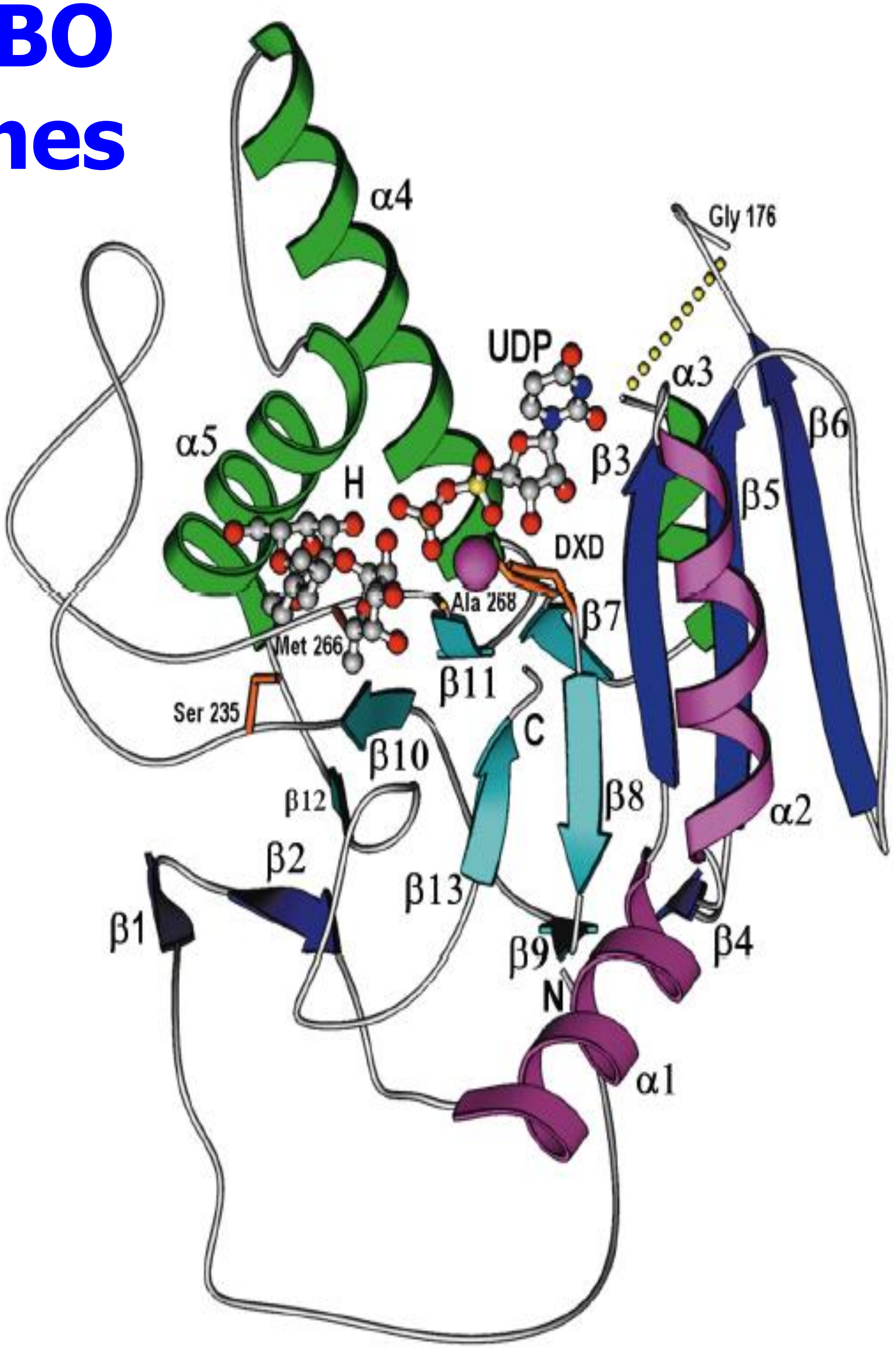
Direction of Flow of Genetic Information

(The central drama of molecular biology)



3D Structure of ABO Blood Type Enzymes

(Patenaude et al, 2002)

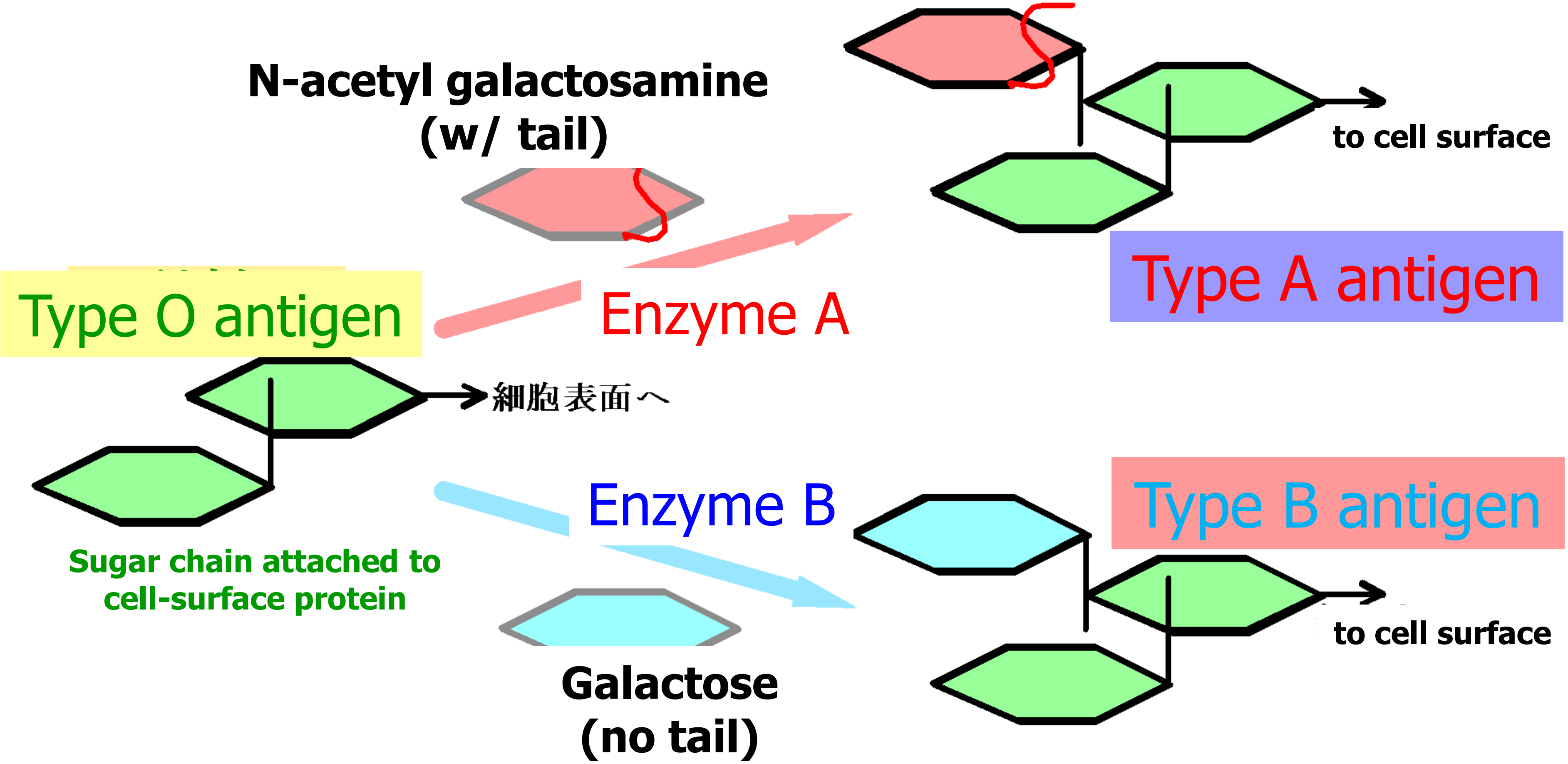


‡

Reprinted by permission from Macmillan Publishers Ltd:
Sonia I. Patenaude1,et al.,
Nature Structural Biology 9, 685 – 690, copyright (2002)

Structure of ABO Blood Type Antigens

The ABO gene produces glycosyltransferase



N-acetyl galactosamine
(w/ tail)

to cell surface

Type A antigen

Type O antigen

Enzyme A

→細胞表面へ

Enzyme B

Type B antigen

Sugar chain attached to
cell-surface protein

to cell surface

Galactose
(no tail)

Frame-shift mutation caused by the deletion of one nucleotide (X) in the O allele

Nucleotide sequences:

```
A   AAGGATGTCCTCGTGGTGACCCCTTGGCTGGCTCCCATTGTCTGGGAGGCACATTCAAC
B                                                       G
O                     X                                   G
```

Protein amino acid sequences:

A/B	**K**	**D**	**V**	**L**	**V**	**V**	**T**	**P**	**W**	**L**	**A**	**P**	**I**	**V**	**W**	**E**	**G**	**T**	**F**	**N**	
O	**K**	**D**	**V**	**L**	**V**	**V**		**P**	**L**	**G**	**W**	**L**	**P**	**L**	**S**	**G**	**R**	**A**	**H**	**S**	**T**

The differences (+) in two amino acids conferring a difference in enzyme activity between transferase A and transferase B, and the differences(!) between two nucleotides in gene A and gene B causing that to happen

Genetic nucleotide sequences:

```
A GATTTCTACTACCTGGGGGGTTCTTCGGGGGTCGGTGCAAGAG
B            A      C
                ！       ！
```

Enzyme protein amino acid sequences:

A	D	F	Y	Y	L	G	G	F	F	G	G	S	V	Q	R
B	D	F	Y	Y	M	G	A	F	F	G	G	S	V	Q	R
						+		+							

(Fumiichiro Yamamoto et al., *Nature*, 1990)

Synonymous and Non-synonymous Substitution in Protein Translation Regions

Synonymous substitution:

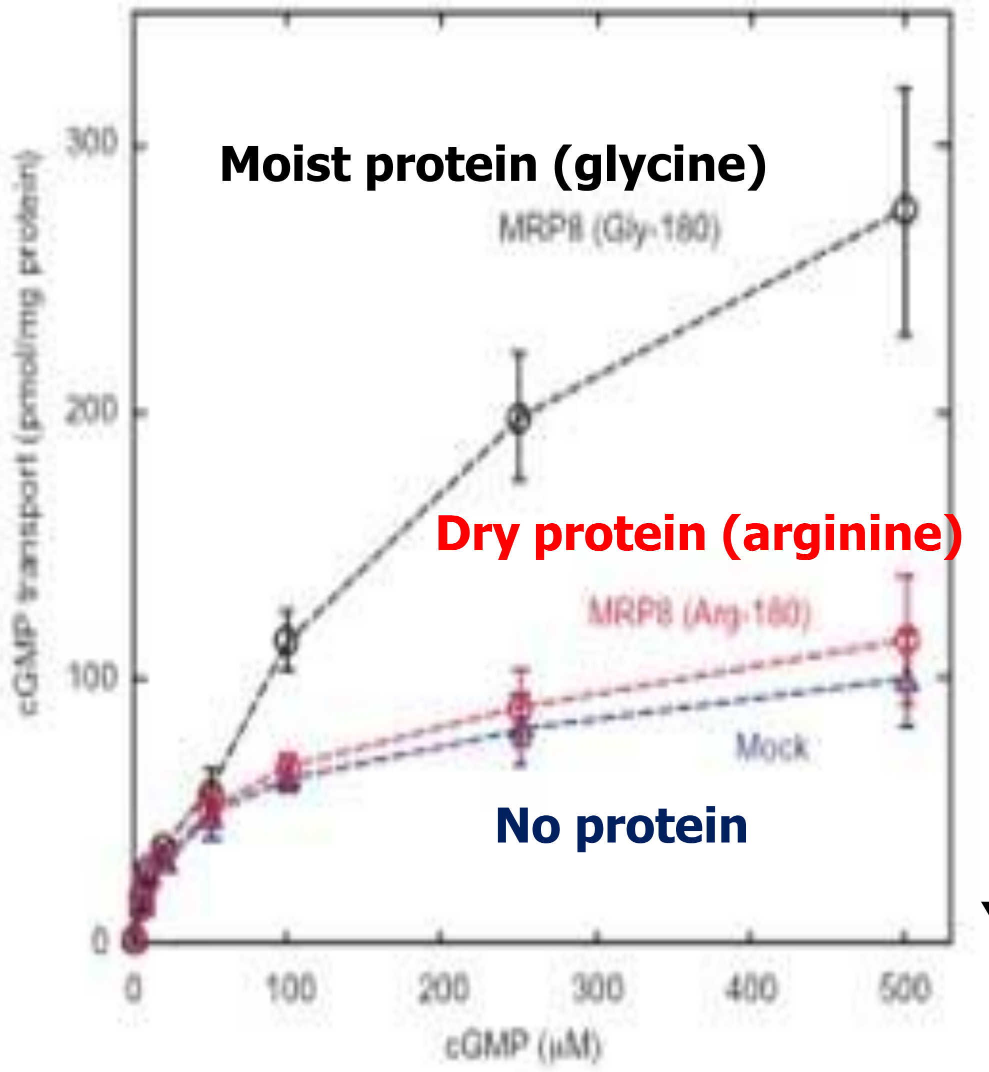
Nucleotide substitution (neutral only)
that does not modify amino acids

Non-synonymous substitution:

Nucleotide substitution (neutral and selective)
that modifies amino acids

Differences in the Function of ABCC11 Gene Product Are the Cause of the Differences in Earwax

Yoshiura et al (2006)



Moist protein (glycine)

Dry protein (arginine)

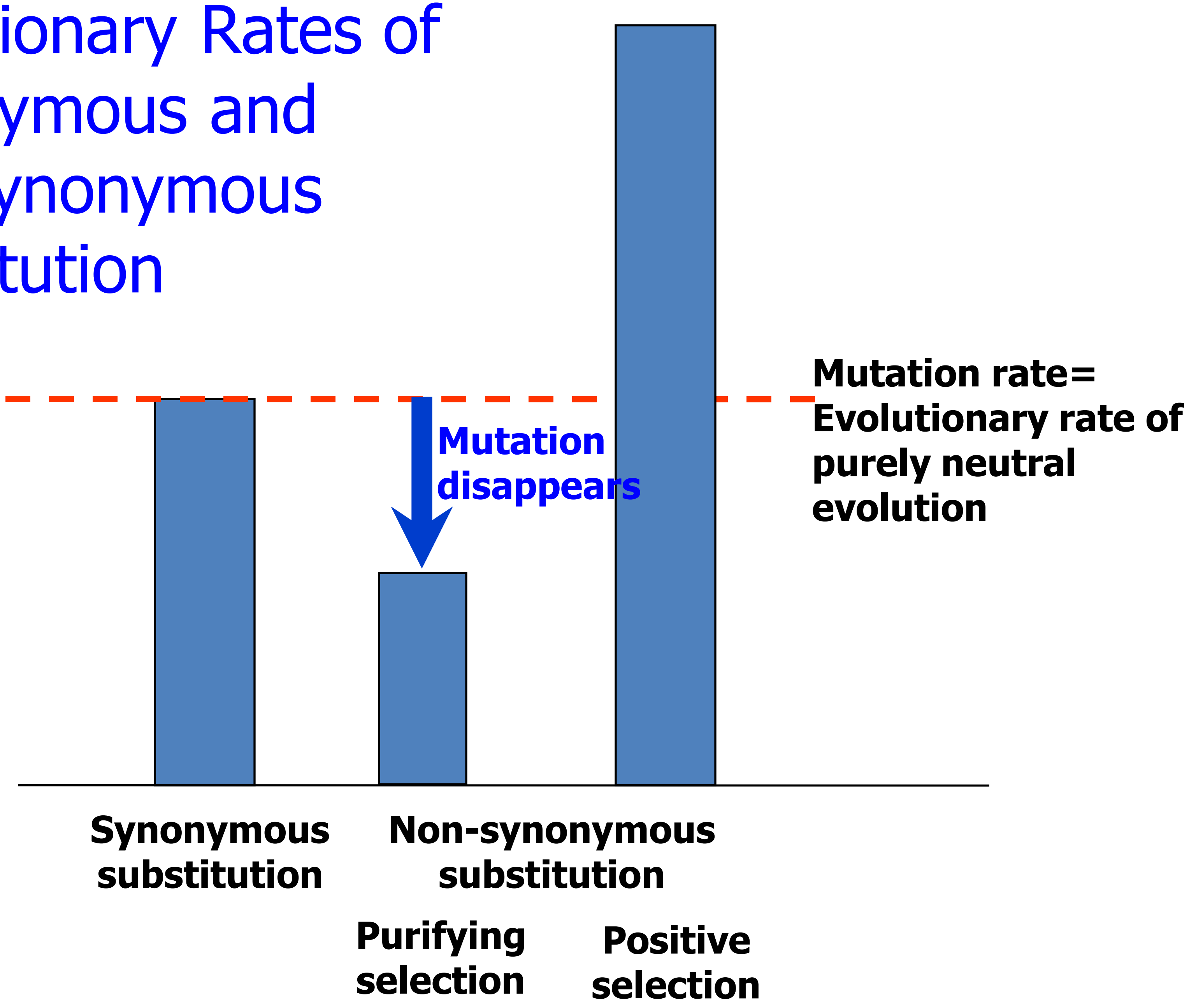
No protein

Dry (function-less) resulted from moist (functional) with modification of a single amino acid

Yoshiura et al (2006)

‡ Reprinted by permission from Macmillan Publishers Ltd, *Nature Genetics 38, 324 – 330, copyright (2006)*

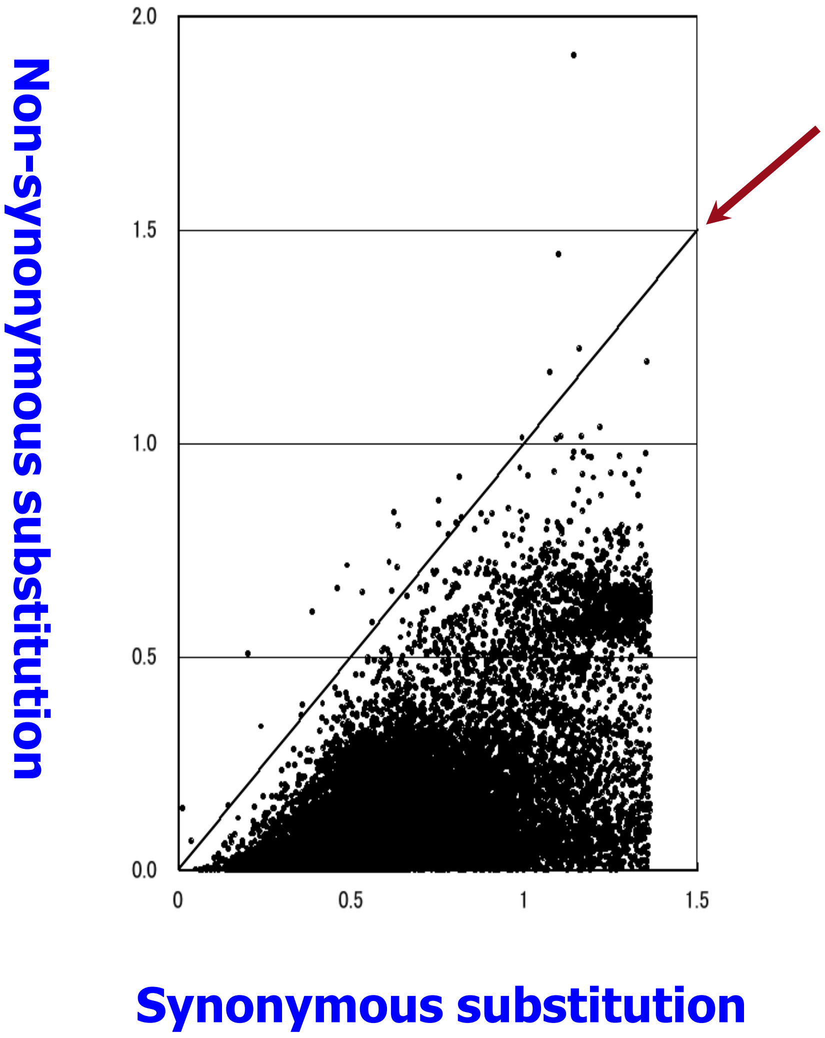
Evolutionary Rates of Synonymous and Non-synonymous Substitution



Mutation rate= Evolutionary rate of purely neutral evolution

Synonymous/Non-synonymous Substitution Ratio in Comparison of Human and Mouse Genes

Saitou, *Genomu Shinka-gaku Nyuumon*(Introduction to Genome Evolution), 2007



Non-synonymous substitution = Synonymous substitution

‡ Naruya Saitou, *Genomu Shinka Gaku Nyuumon(Introduction to Genome Evolution),* Kyouritsu Shuppan, 2007.

- **Entire human genome (3.2 billion nucleotides)**
  - **Genes and gene-related regions (1.2 billion nucleotides)**
    - **Genes (48 million nucleotides) <-- Just 1.5% of the whole**
    - **Gene-related regions (1,152 million nucleotides)**
  - **Regions between genes (2.0 billion nucleotides)**
    - **Dispersed repetitive sequences (1.4 billion nucleotides)**
      - **Long dispersed repetitive sequences LINE (640 million nucleotides)**
      - **Short dispersed repetitive sequences SINE (420 million nucleotides)**
      - **Other repetitive sequences SINE (340 million nucleotides)**
    - **Other intergenic regions (600 million nucleotides)**
      - **Microsatellites (90 million nucleotides)**
      - **Non-microsatellites (510 million nucleotides)**

International Human Genome Sequencing Consortium (2004)

- Entire human genome (3.2 billion nucleotides) Almost entirely junk DNA
  - Genes and gene-related regions (1.2 billion nucleotides)
Almost entirely junk DNA
    - Genes (48 million nucleotides) <-- Just 1.5% of the whole
    - Gene-related regions (1,152 million nucleotides)
Almost entirely junk DNA
  - Regions between genes (2.0 billion nucleotides)
Almost entirely junk DNA
    - Dispersed repetitive sequences (1.4 billion nucleotides)
Almost entirely junk DNA
      - Long dispersed repetitive sequences LINE (640 million nucleotides)
Almost entirely junk DNA
      - Short dispersed repetitive sequences SINE (420 million nucleotides)
Almost entirely junk DNA
      - Other repetitive sequences SINE (340 million nucleotides)
Almost entirely junk DNA
    - Other intergenic regions (600 million nucleotides)
Almost entirely junk DNA
      - Microsatellites (90 million nucleotides)
Almost entirely junk DNA
      - Non-microsatellites (510 million nucleotides)
Almost entirely junk DNA

International Human Genome Sequencing Consortium (2004)

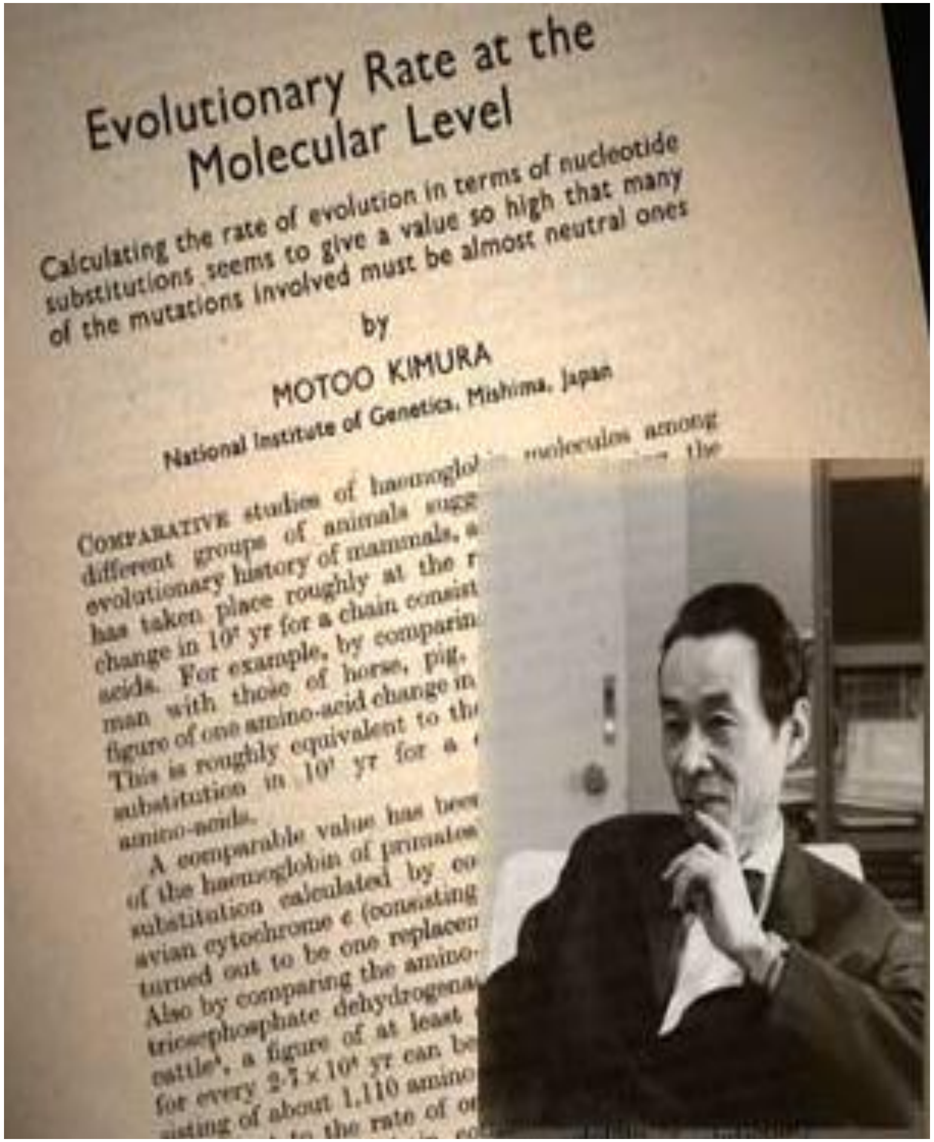
Mouse–Rat Differences in Intergenic Regions and in Synonymous Sites

Regions Compared	Sites compared	Nucleotide Substitutions
Intergenic regions	27, 931, 288	0. 15 ～ 0. 16
Synonymous sites	955, 105	0. 16 ～ 0. 17

Naruya Saitou, *Genomu Shinka Gaku Nyuumon(Introduction to Genome Evolution)*, 2007.
Source data from Abe et al (2004)

The bulk of our genome is the product of neutral evolution!

Prof. Motoo Kimura and
his paper outlining neutral evolution theory (1968)



Evolutionary Rate at the Molecular Level

Calculating the rate of evolution in terms of nucleotide substitutions seems to give a value so high that many of the mutations involved must be almost neutral ones

by

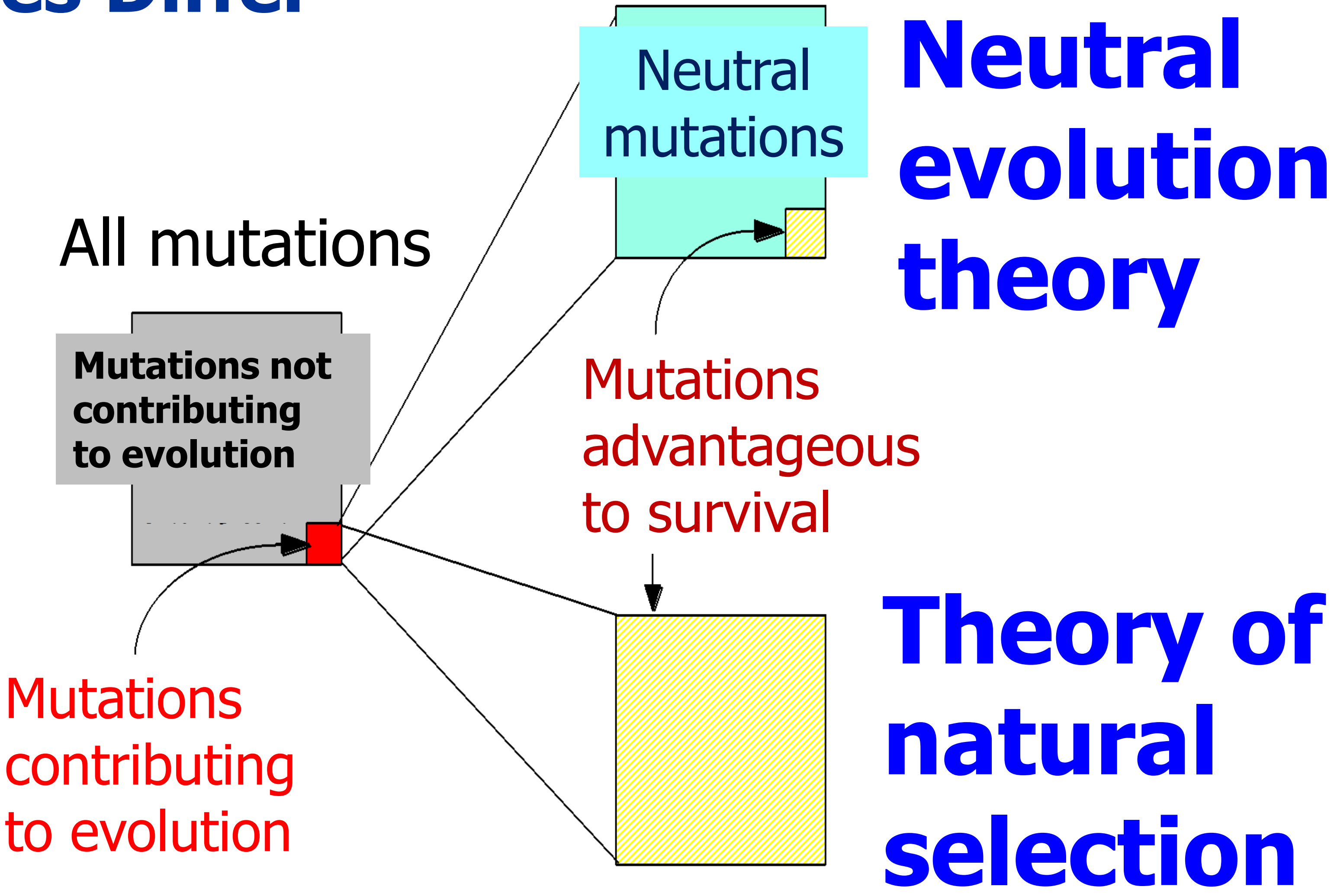
MOTOO KIMURA

National Institute of Genetics, Mishima, Japan

COMPARATIVE studies of haemoglo molecules among
different groups of animals sugg the
evolutionary history of mammals, a
has taken place roughly at the r
change in 10^7 yr for a chain consist
acids. For example, by comparin
man with those of horse, pig,
figure of one amino-acid change in
This is roughly equivalent to th
substitution in 10^9 yr for a
amino-acids.

A comparable value has bee
of the haemoglobin of primates
substitution calculated by co
avian cytochrome c (consisting
turned out to be one replacem
Also by comparing the amino-
triosephosphate dehydrogena
cattle[4], a figure of at least
for every $2\cdot 7 \times 10^6$ yr can be
sisting of about 1,110 amino
the rate of or

How the Neutral and Natural Selection Theories Differ



All mutations

Mutations not contributing to evolution

Mutations contributing to evolution

Neutral mutations

Neutral evolution theory

Mutations advantageous to survival

Theory of natural selection

‡ Naruya Saitou, *E de Wakaru Jinrui no Shinka(Evolution of the Human Race which is Understood with Pictures)*, Koudan Sha, 2009.

"Of all these mutations, just this one box has ones that contribute to evolution."

Mutation Hold

What does the box hold?

"Just a few are advantageous to survival. The others are neutral."

"No, no! They all contribute to survival."

Evolutionary contribution

Evolutionary contribution

Neutral Evolution Theory

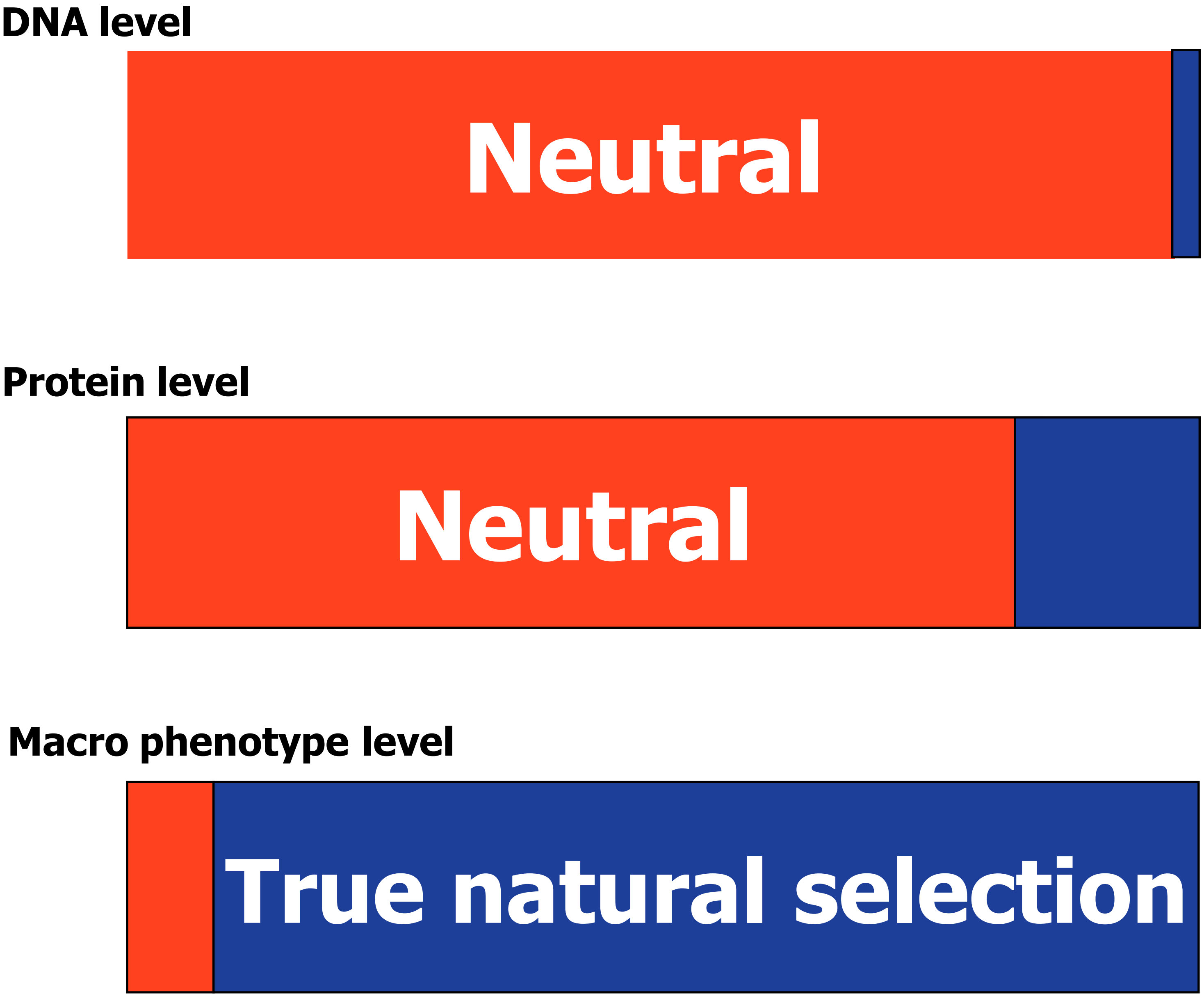
Theory of Natural Selection

Illustrations: Saori Yasutomi

Genes That Evolved Non-neutrally

- mmunological system genes
- Viral genes that attack host immunological systems
- Some cell-surface molecules (e.g. blood type)
- Genes for plant systems that prevent self-fertilization
- Surface proteins in marine animal sperm (differentiation from other species)
- Causative protein in snake venom

Proportion of Neutral Mutations at Three Levels as Reckoned by Many Researchers



DNA level

Neutral

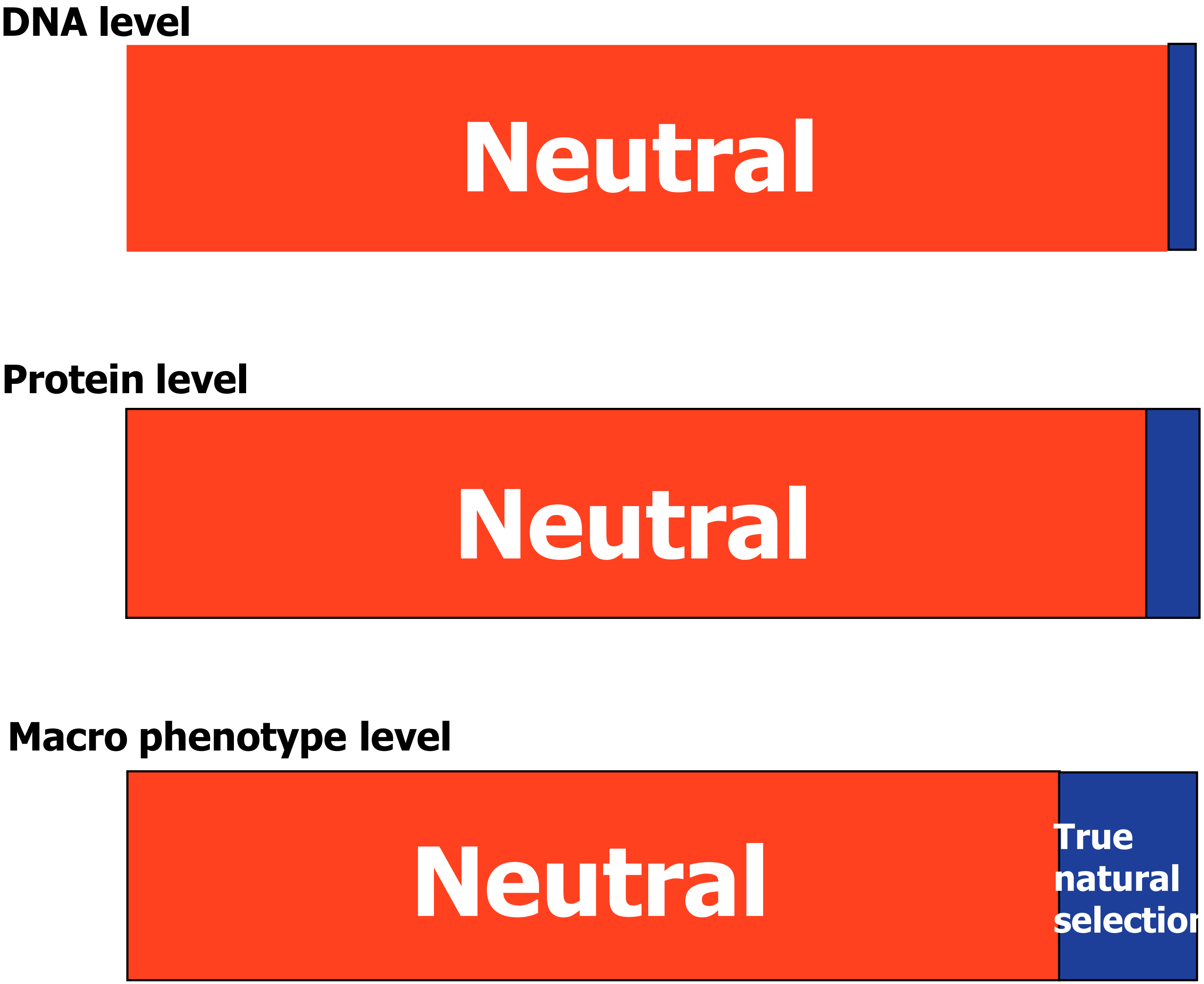
Protein level

Neutral

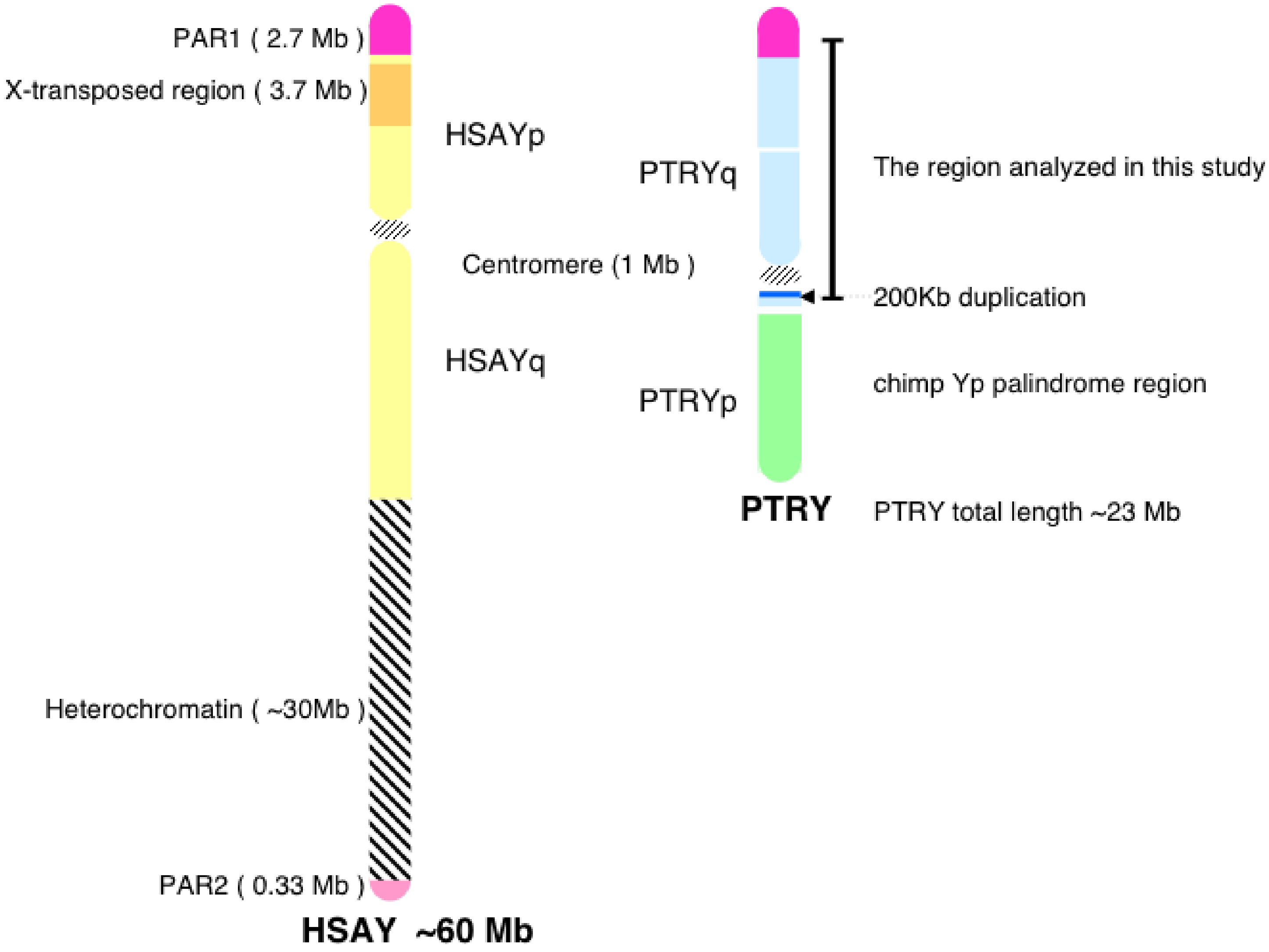
Macro phenotype level

True natural selection

Proportion of Neutral Mutations at Three Levels as Reckoned by Me



Comparative Analysis of Human and Anthropoid Genomes



Reprinted by permission from Macmillan Publishers Ltd:
Nature Genetics 38, 158 – 167, copyright (2006)

Kuroki et al (2006)

‡

‡ Reprinted by permission from Macmillan Publishers Ltd: *Nature, 413, 519–523, copyright (2001)*

Pedigree of a family exhibiting language disorder (Lai et al., 2001)

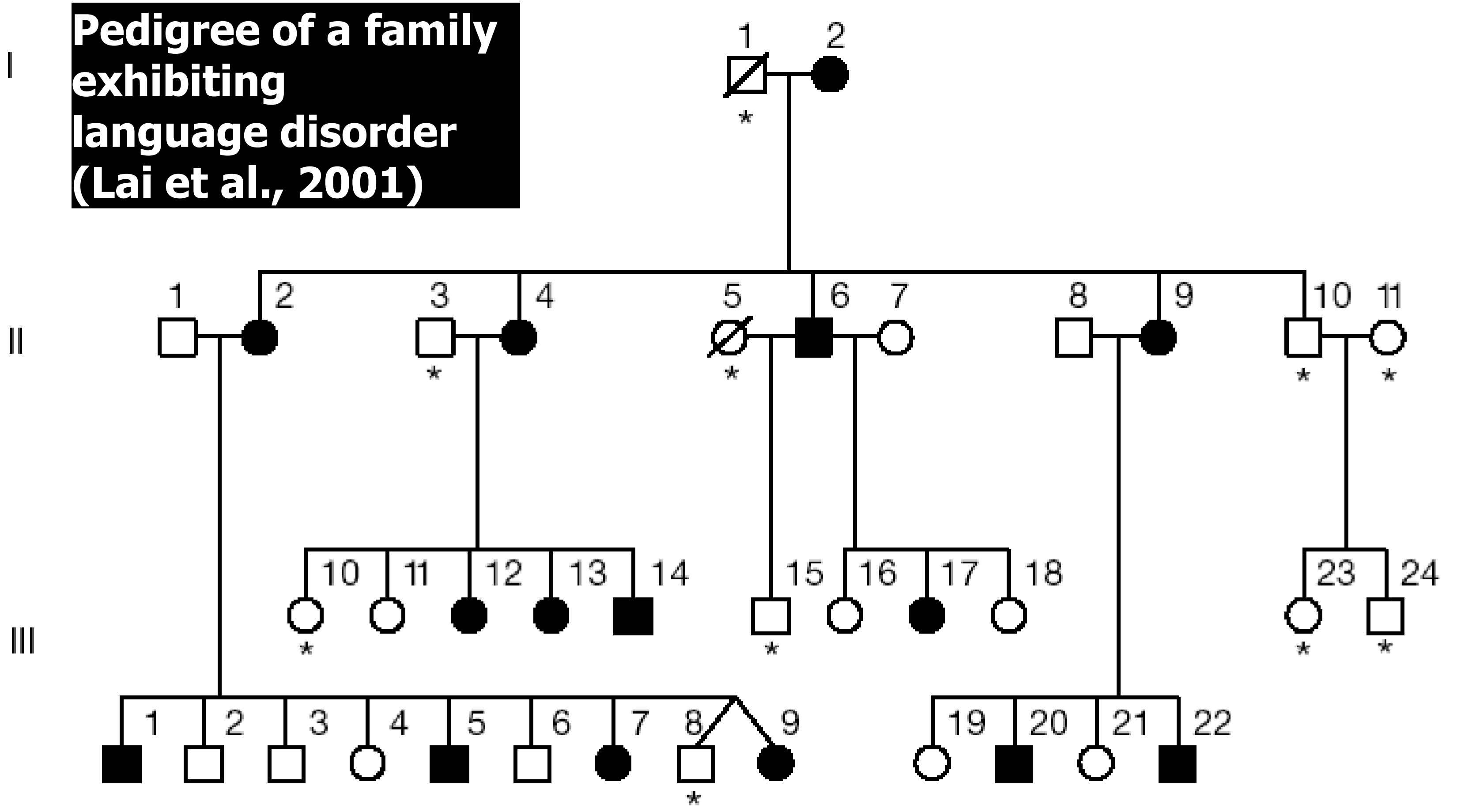
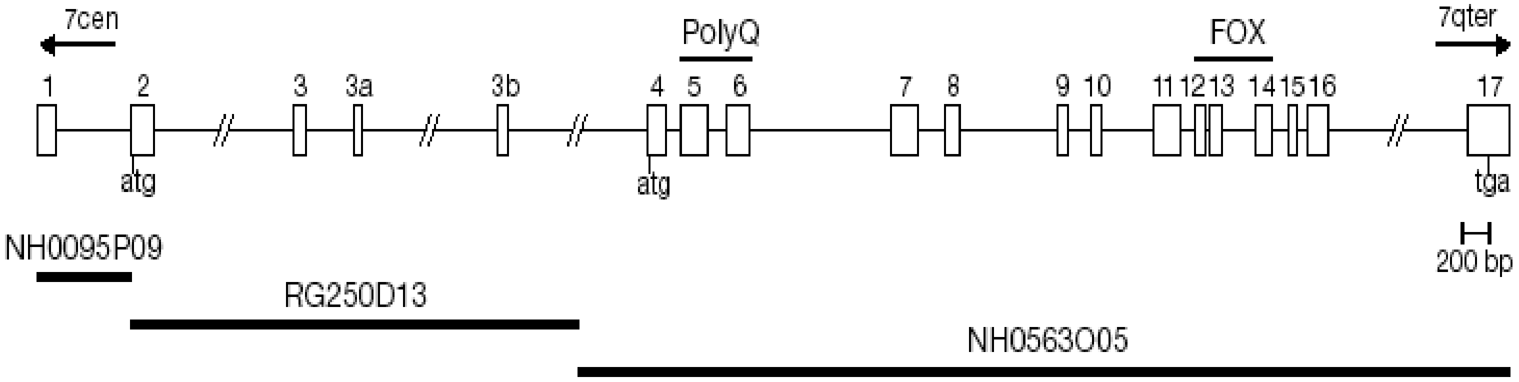


Figure 1 Pedigree of the KE family. Affected individuals are indicated by filled symbols. Asterisks indicate those individuals who were unavailable for genetic analyses. Squares are males, circles are females, and a line through a symbol indicates that the person is deceased.

A Language Gene?
The Structure of the FOXP2 Gene

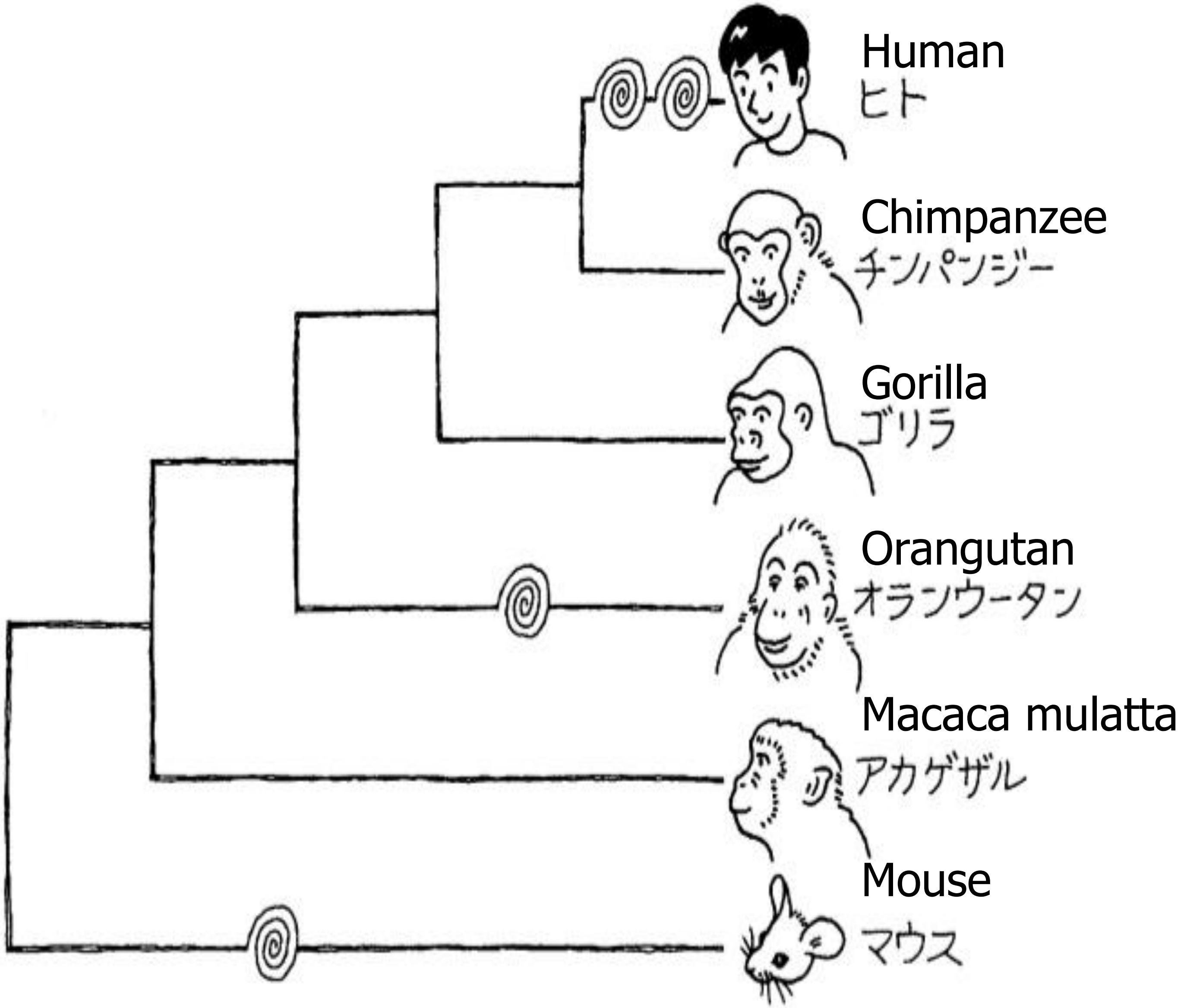


Approximate location in human genome:
Human chromosome 7, on the long arm (q)

Reprinted by permission from Macmillan Publishers Ltd:
Nature, 413, 519–523, copyright (2001)

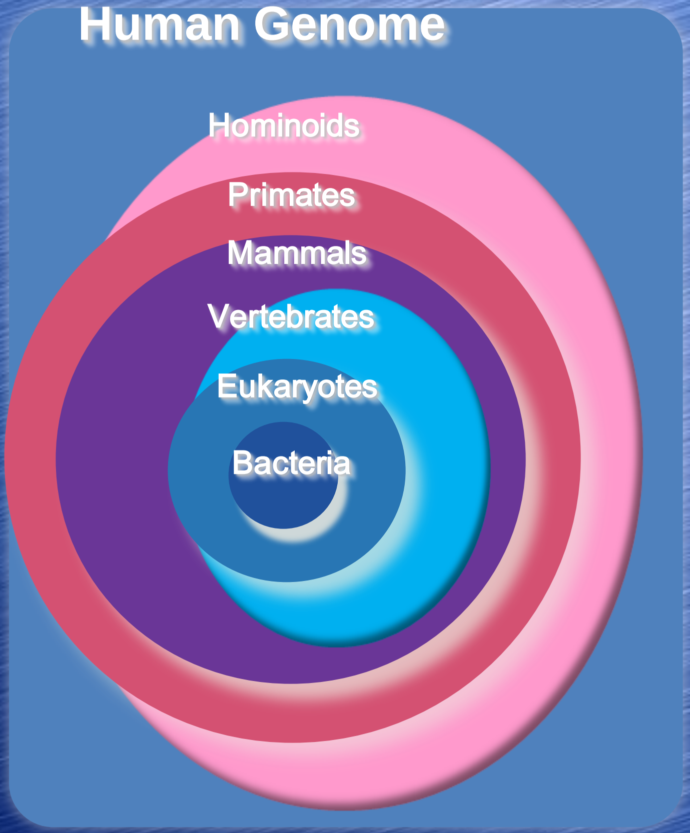
Lai et al,. (2001)

Did Changes in Amino Acids of the FOXP2 Protein Contribute to Capacity for Language?



‡ Illustrations: Saori Yasutomi

Reverse-Nested Structure of the Human Genome



Human Genome

Hominoids

Primates

Mammals

Vertebrates

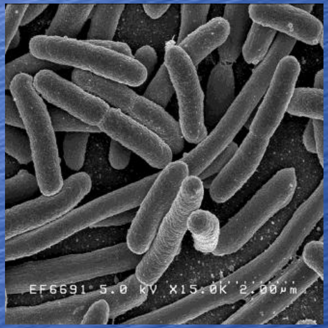
Eukaryotes

Bacteria

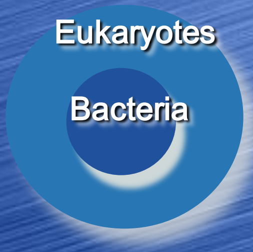
Reverse-Nested Structure of the Human Genome

Bacteria

Bacteria



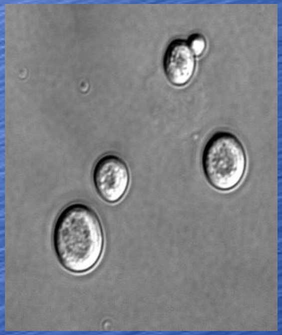
Reverse-Nested Structure of the Human Genome



Eukaryotes

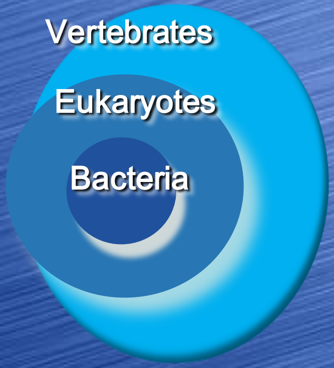
Bacteria

Eukaryotes

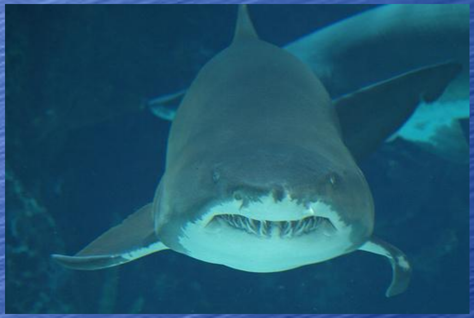


‡

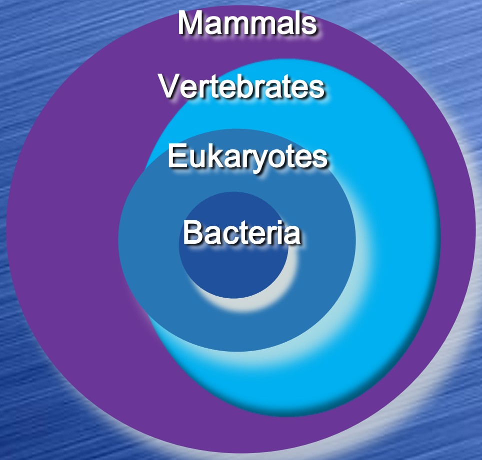
Reverse-Nested Structure of the Human Genome



Vertebrates

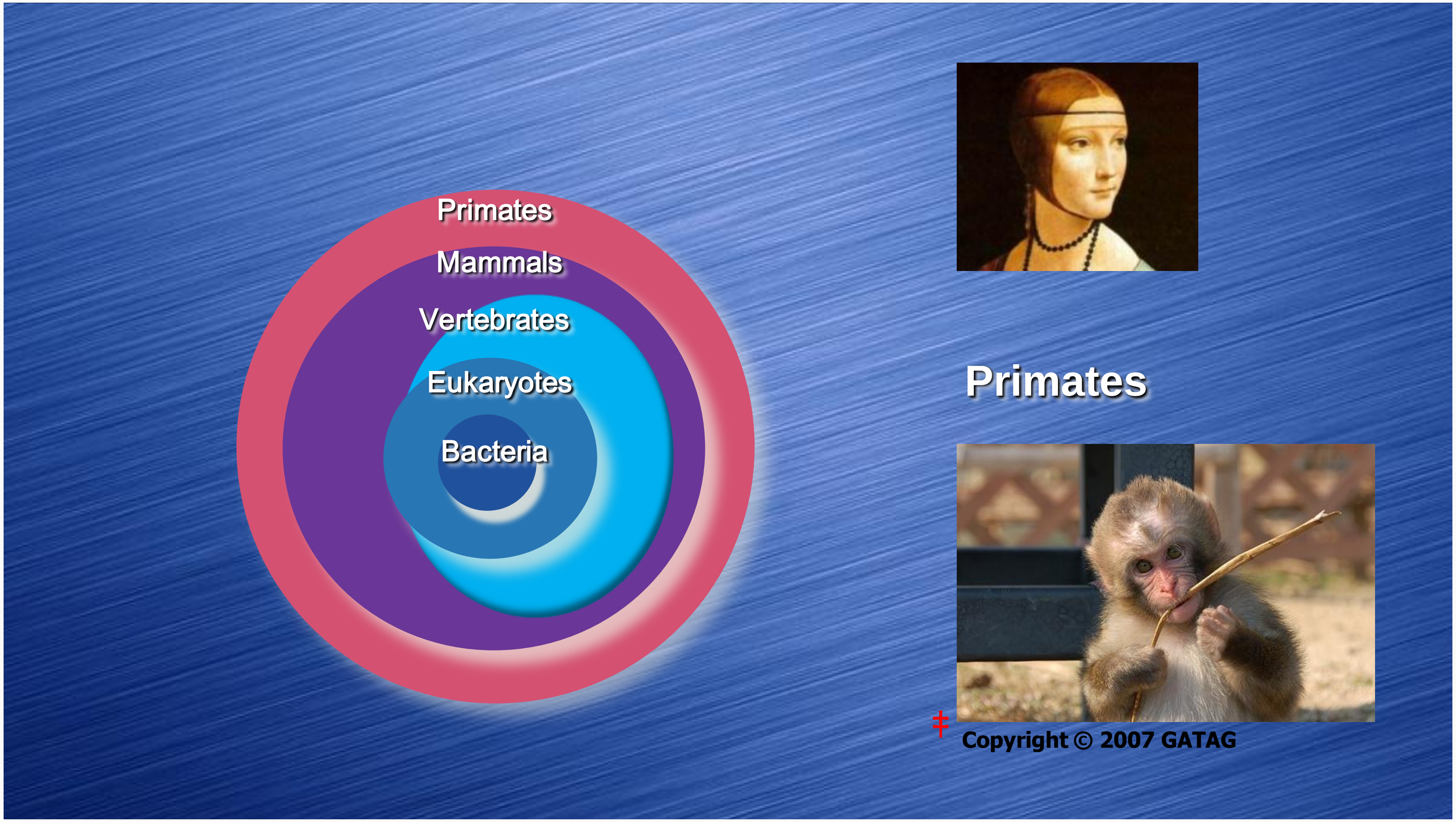


Reverse-Nested Structure of the Human Genome



Mammals

Reverse-Nested Structure of the Human Genome

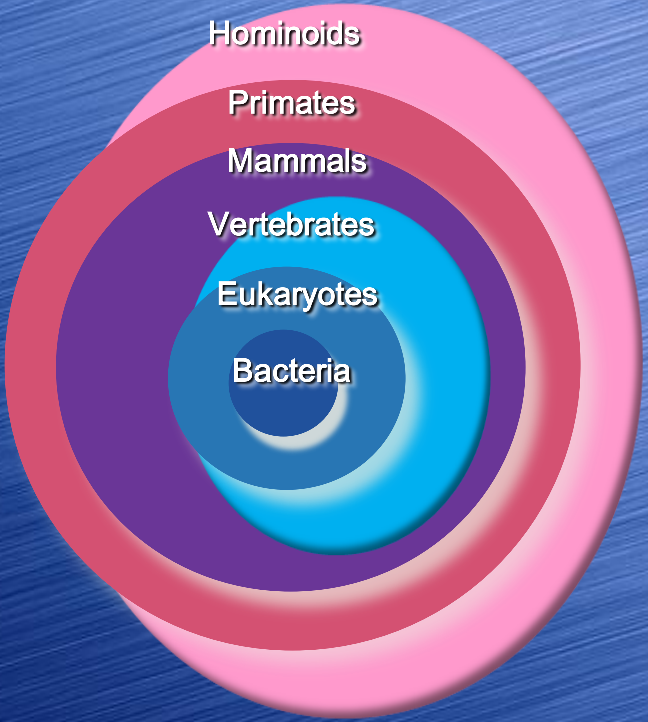


Primates
Mammals
Vertebrates
Eukaryotes
Bacteria

Primates

‡

Reverse-Nested Structure of the Human Genome



Hominoids

Primates

Mammals

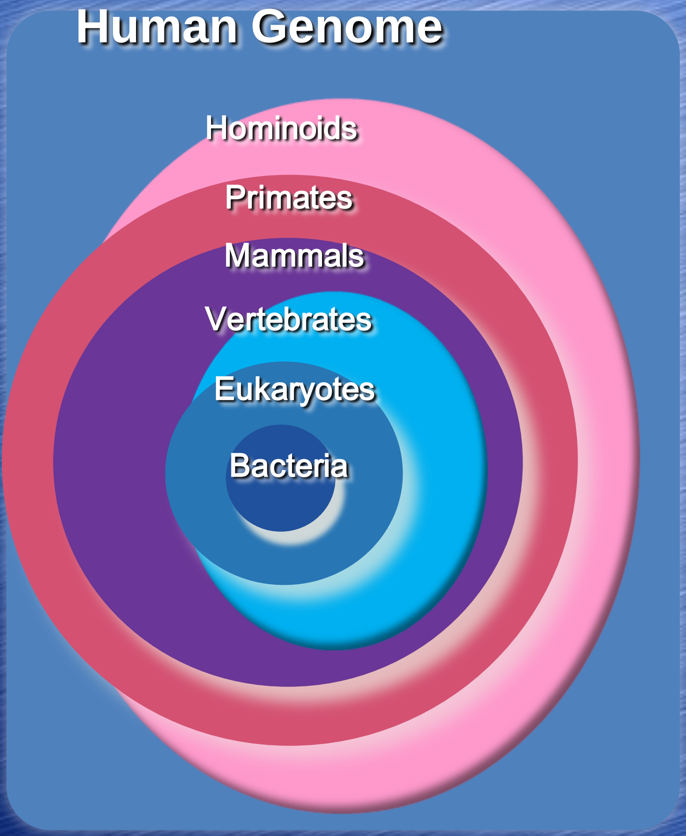
Vertebrates

Eukaryotes

Bacteria

Hominoids

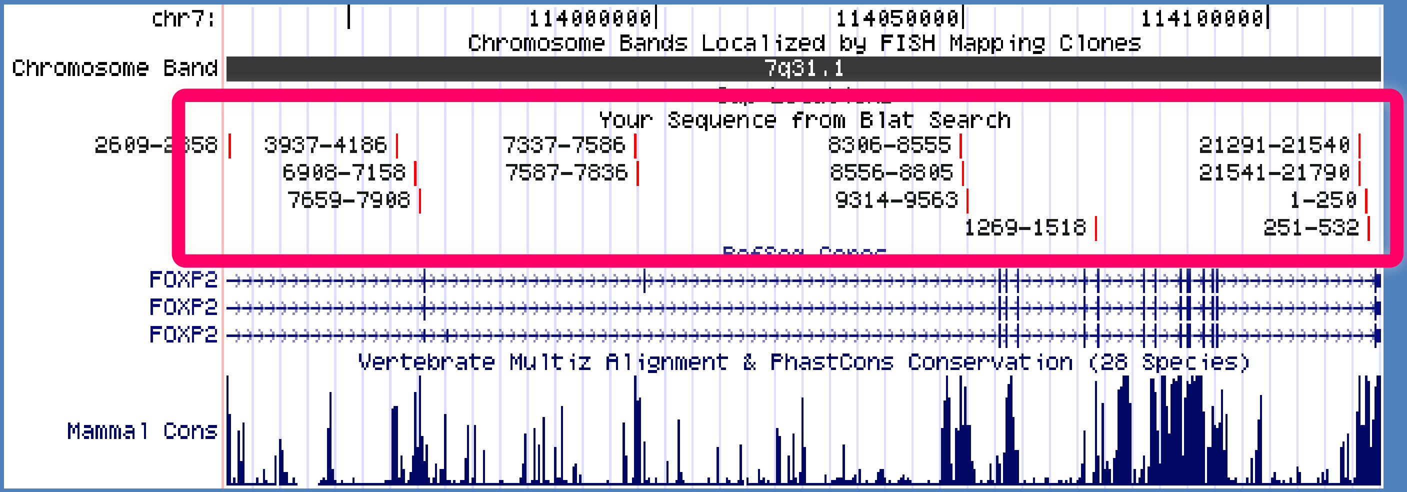
Reverse-Nested Structure of the Human Genome



Hominoids

DNA Regions Considered Evolutionarily Significant in All Mammals

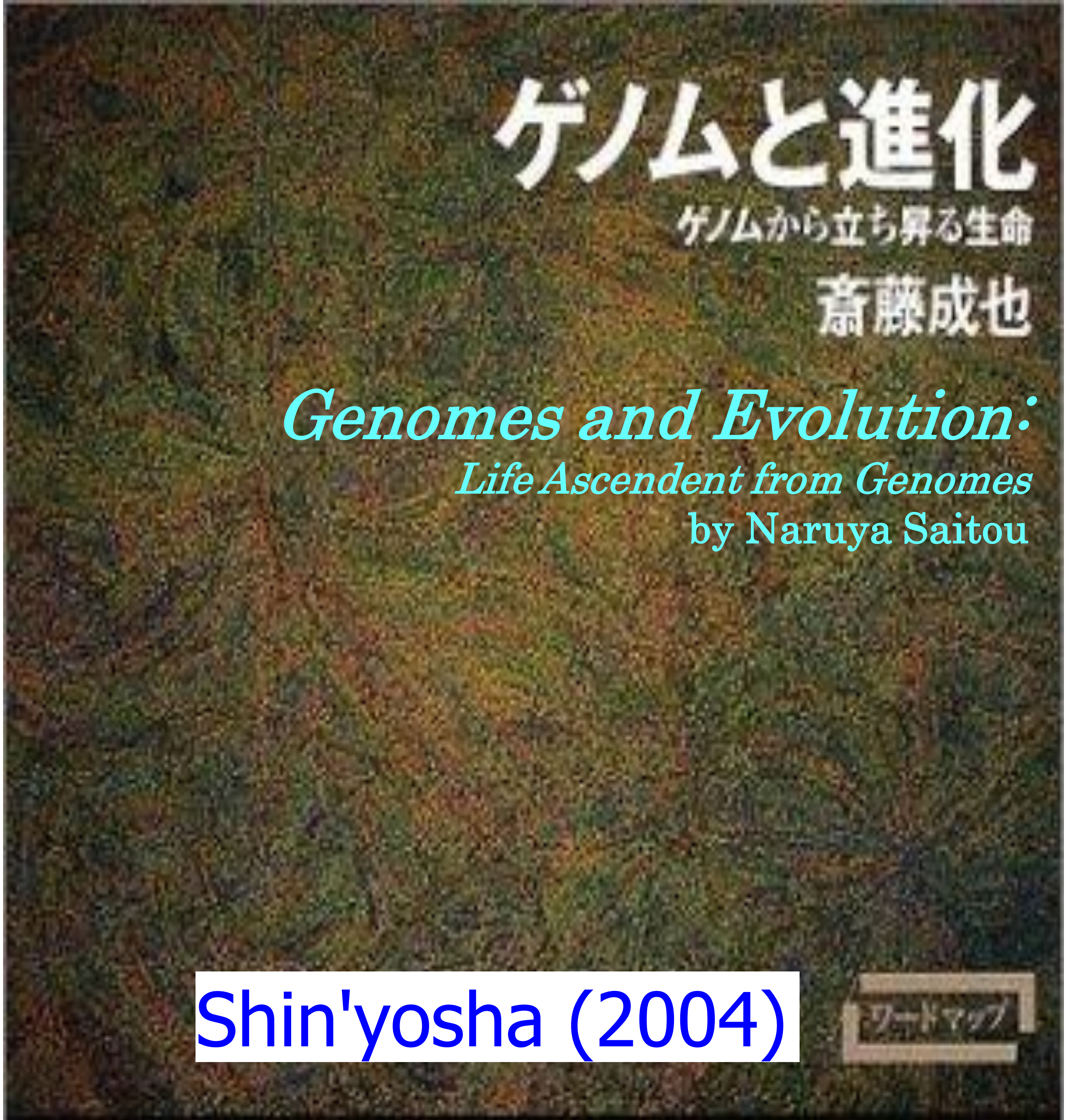
FOXP2



Humans and Other Life Forms

Humans and Chimpanzees, and the Rest?

Humans and Large Anthropoids, and the Rest?



Genomes and Evolution:

Life Ascendent from Genomes

by Naruya Saitou

Shin'yosha (2004)

‡

Naruya Saitou, *Genomu to Shinka – Genomu kara Tachinoboru Seimei (Genomes and Evolution – Life Ascendent from Genomes),* Shin'yosha, 2004.

Published March 2006

The Iwanami Shoten Evolution Series Vol. 2

Idenshi to Genomu no Shinka

(Evolution of Genes and Genomes)

Hajime Ishikawa, Noriyuki Sato,
Mariko Hasegawa and Naruya Saitou, eds.

Published March 2007
Naruya Saitou, ed.
Genomu ha Koko made Kaimei Sareta
(What We Know of Genomes)
(Wedge Sensho)

- “An Integrated Understanding of Matter and Life as Disclosed by Genomes,” by Akiyoshi Wada
- “Decoding the Human Genome, and Metagenomes,” by Yoshiyuki Sakaki
- “Genomes and Animal Body Formation, and Gene Networks,” by Noriyuki Sato
- “Tracing to Humans the History of Life Written in Genome Sequences,” by Naruya Saitou
- A Genomics Roundtable: "From microorganisms to Neanderthals" (Asao Fujiyama, Shohei Hattori, Shin'ichi Morishita, Naruya Saitou)

Introduction to Genome Evolution,
by Naruya Saitou
Kyouritsu Shuppan
(2007)

ゲノム進化学入門
CD-ROM付
斎藤成也著
共立出版

ǂ

Up Next on Friday, May 1:

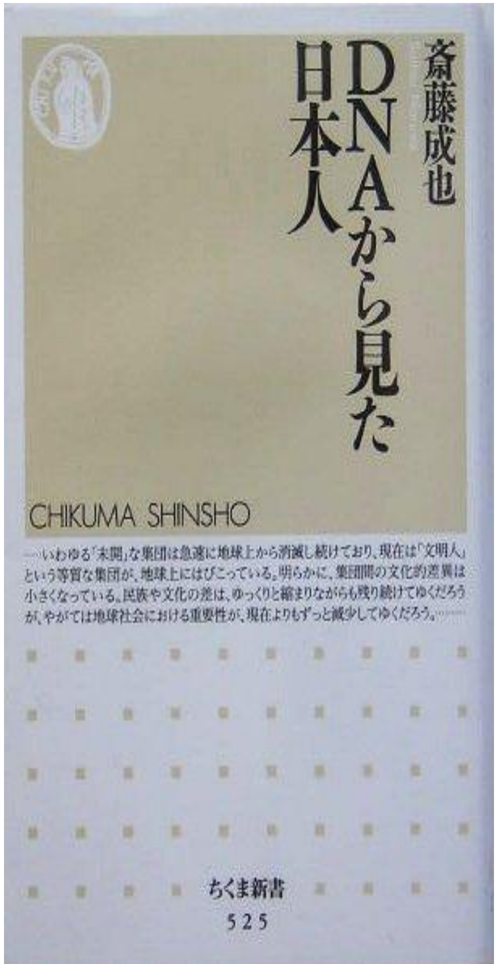
"What Our Genome Says about the Origins of the Japanese"

Naruya Saitou

DNA kara Mita Nihonjin

(The Japanese in Terms of DNA)

Chikuma Shinsho, 2005



斎藤成也

DNAから見た
日本人

CHIKUMA SHINSHO

……いわゆる「未開」な集団は急速に地球上から消滅し続けており、現在は「文明人」という等質な集団が、地球上にはびこっている。明らかに、集団間の文化的差異は小さくなっている。民族や文化の差は、ゆっくりと縮まりながらも残り続けてゆくだろうが、やがては地球社会における重要性が、現在よりもずっと減少してゆくだろう。……

ちくま新書
525

‡