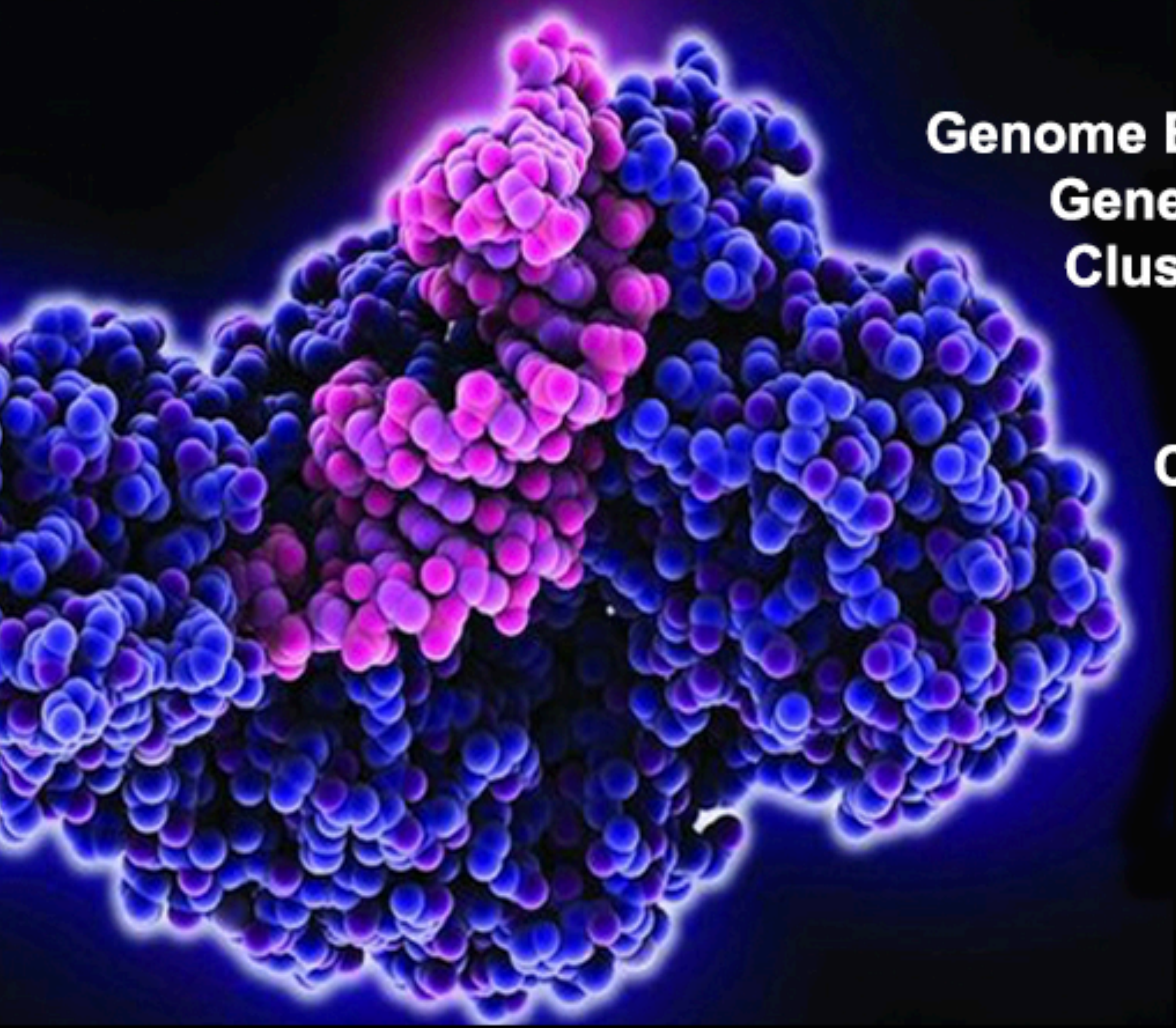
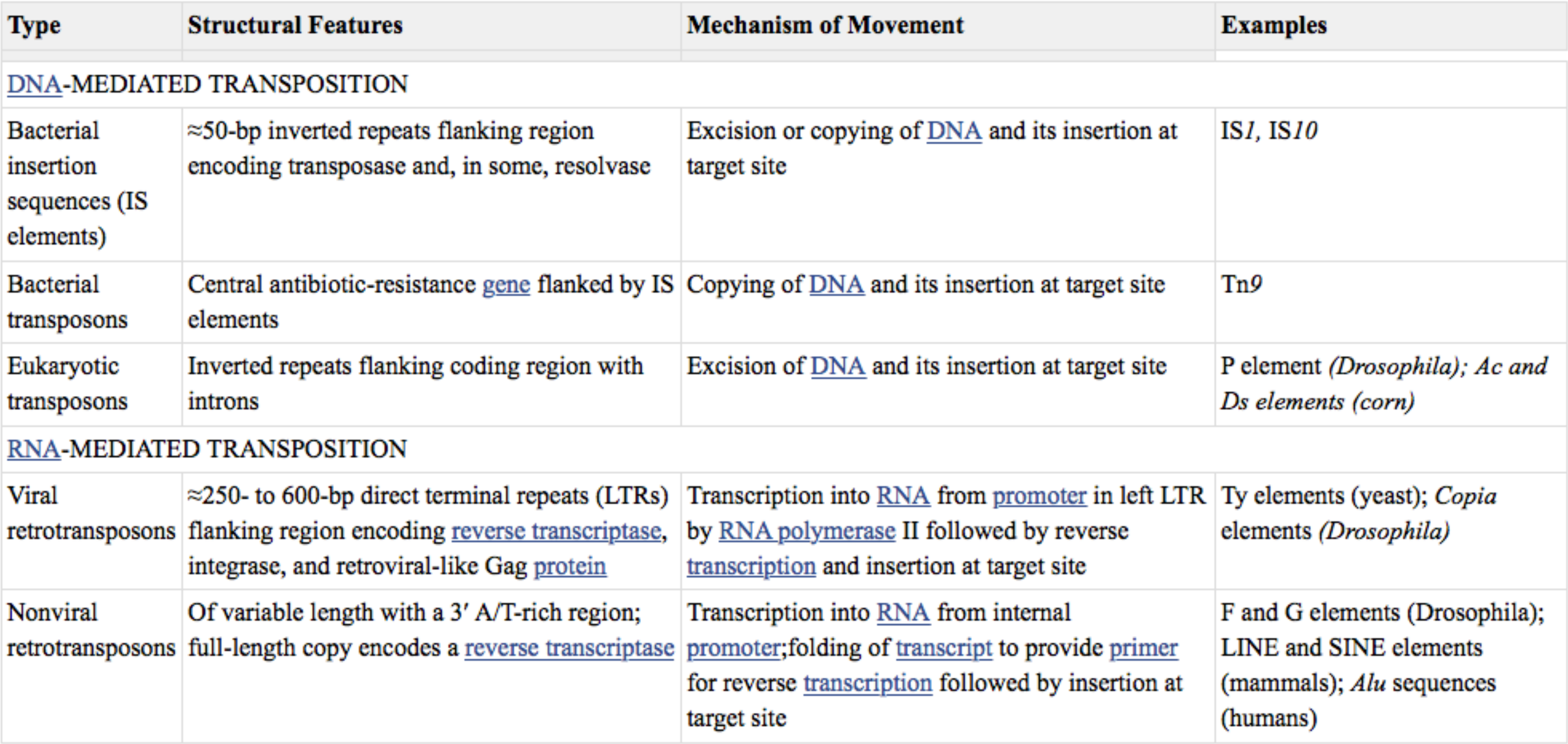
BIOL8620 Eukaryotic Genetics



Genome Evolution:
Gene numbers,
Clusters & Repeats

Chapter 5 & 6,

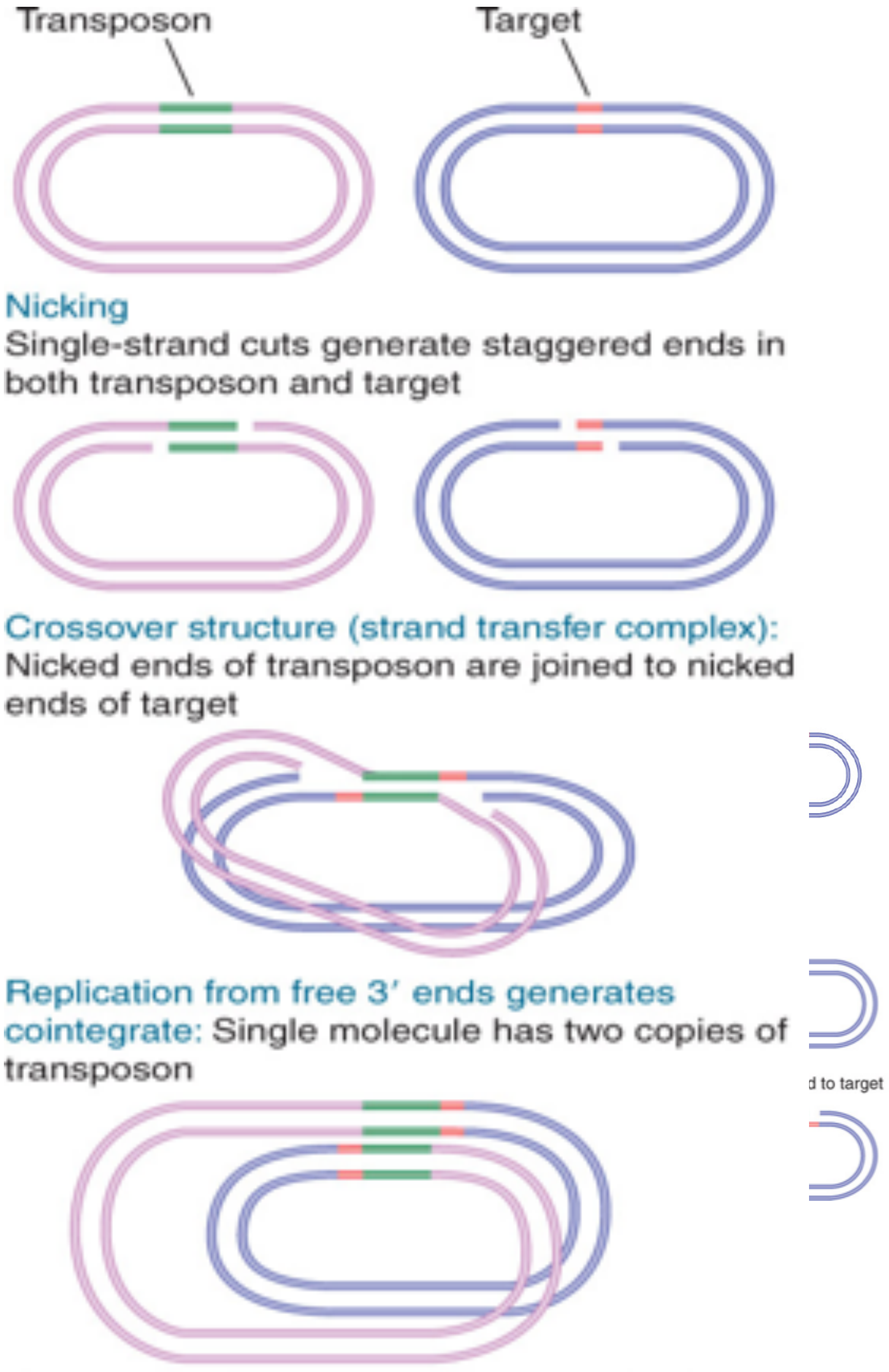
parts of 7 & 8



Type	Structural Features	Mechanism of Movement	Examples
DNA-MEDIATED TRANSPOSITION			
Bacterial insertion sequences (IS elements)	≈50-bp inverted repeats flanking region encoding transposase and, in some, resolvase	Excision or copying of DNA and its insertion at target site	IS*1*, IS*10*
Bacterial transposons	Central antibiotic-resistance gene flanked by IS elements	Copying of DNA and its insertion at target site	Tn*9*
Eukaryotic transposons	Inverted repeats flanking coding region with introns	Excision of DNA and its insertion at target site	P element *(Drosophila); Ac and Ds elements (corn)*
RNA-MEDIATED TRANSPOSITION			
Viral retrotransposons	≈250- to 600-bp direct terminal repeats (LTRs) flanking region encoding reverse transcriptase, integrase, and retroviral-like Gag protein	Transcription into RNA from promoter in left LTR by RNA polymerase II followed by reverse transcription and insertion at target site	Ty elements (yeast); *Copia* elements *(Drosophila)*
Nonviral retrotransposons	Of variable length with a 3′ A/T-rich region; full-length copy encodes a reverse transcriptase	Transcription into RNA from internal promoter;folding of transcript to provide primer for reverse transcription followed by insertion at target site	F and G elements (Drosophila); LINE and SINE elements (mammals); *Alu* sequences (humans)

NCBI Bookshelf. A service of the National Library of Medicine, National Institutes of Health.

DNA Transposition	Example
Replicative	Tn3 Tn7
Nonreplicative two-strand	Tn5 Tn7
Nonreplicative four-strand	Tn10 Ac autonomous (Ds) nonautonomous

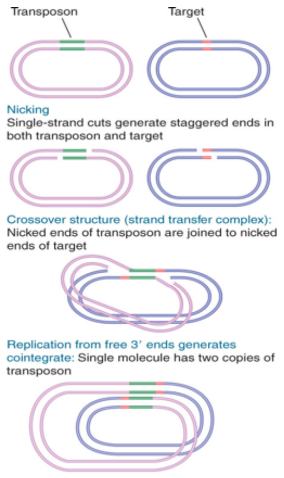


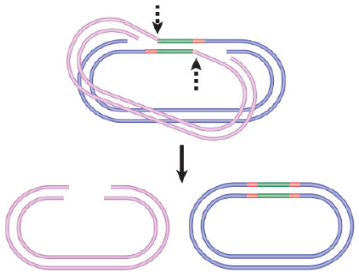
Nicking
Single-strand cuts generate staggered ends in both transposon and target

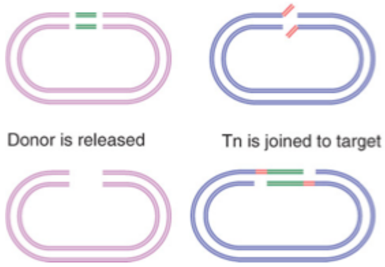
Crossover structure (strand transfer complex): Nicked ends of transposon are joined to nicked ends of target

Replication from free 3′ ends generates cointegrate: Single molecule has two copies of transposon

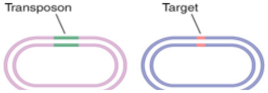
DNA Transposition	Example	Enzyme(s)
Replicative	Tn3	Transposase / Resolvase
	Tn7	Transposase (TnsB) - Endonuclease (TnsA)
Nonreplicative two-strand	Tn5	Transposase
	Tn7	Transposase (TnsB) + Endonuclease (TnsA)
Nonreplicative four-strand	Tn10	Transposase
	Ac autonomous (Ds) nonautonomous	Transposase (controlled by Methylation)

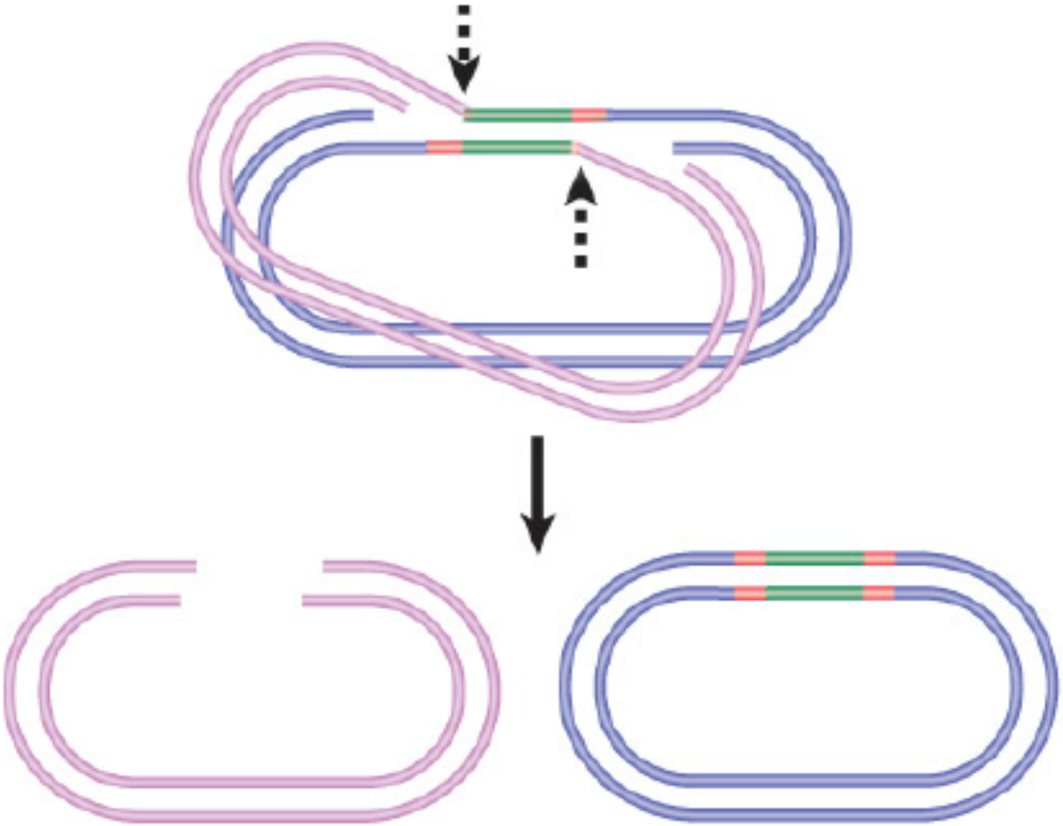

Transposon
Target
Nicking
Single-strand cuts generate staggered ends in both transposon and target
Crossover structure (strand transfer complex): Nicked ends of transposon are joined to nicked ends of target
Replication from free 3' ends generates cointegrate: Single molecule has two copies of transposon



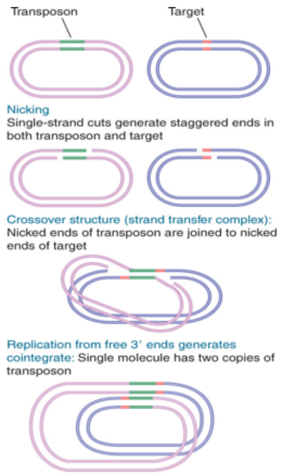


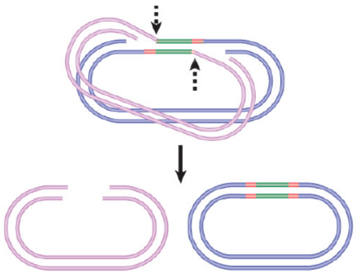
DNA Transposition	Example	Enzyme(s)
Replicative		
Nonreplicative two-strand		
Nonreplicative four-strand		

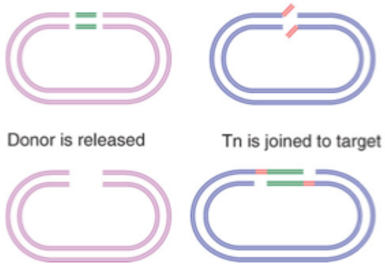




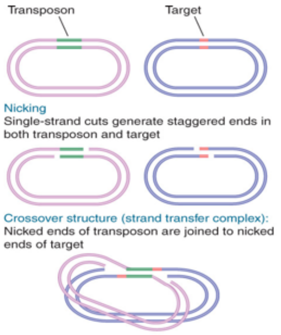
DNA Transposition	Example	Enzyme(s)
Replicative	Tn3	Transposase / Resolvase
	Tn7	Transposase (TnsB) - Endonuclease (TnsA)
Nonreplicative two-strand	Tn5	Transposase
	Tn7	Transposase (TnsB) + Endonuclease (TnsA)
Nonreplicative four-strand	Tn10	Transposase
	Ac autonomous (Ds) nonautonomous	Transposase (controlled by Methylation)


Transposon
Target
Nicking
Single-strand cuts generate staggered ends in both transposon and target
Crossover structure (strand transfer complex):
Nicked ends of transposon are joined to nicked ends of target
Replication from free 3' ends generates cointegrate: Single molecule has two copies of transposon



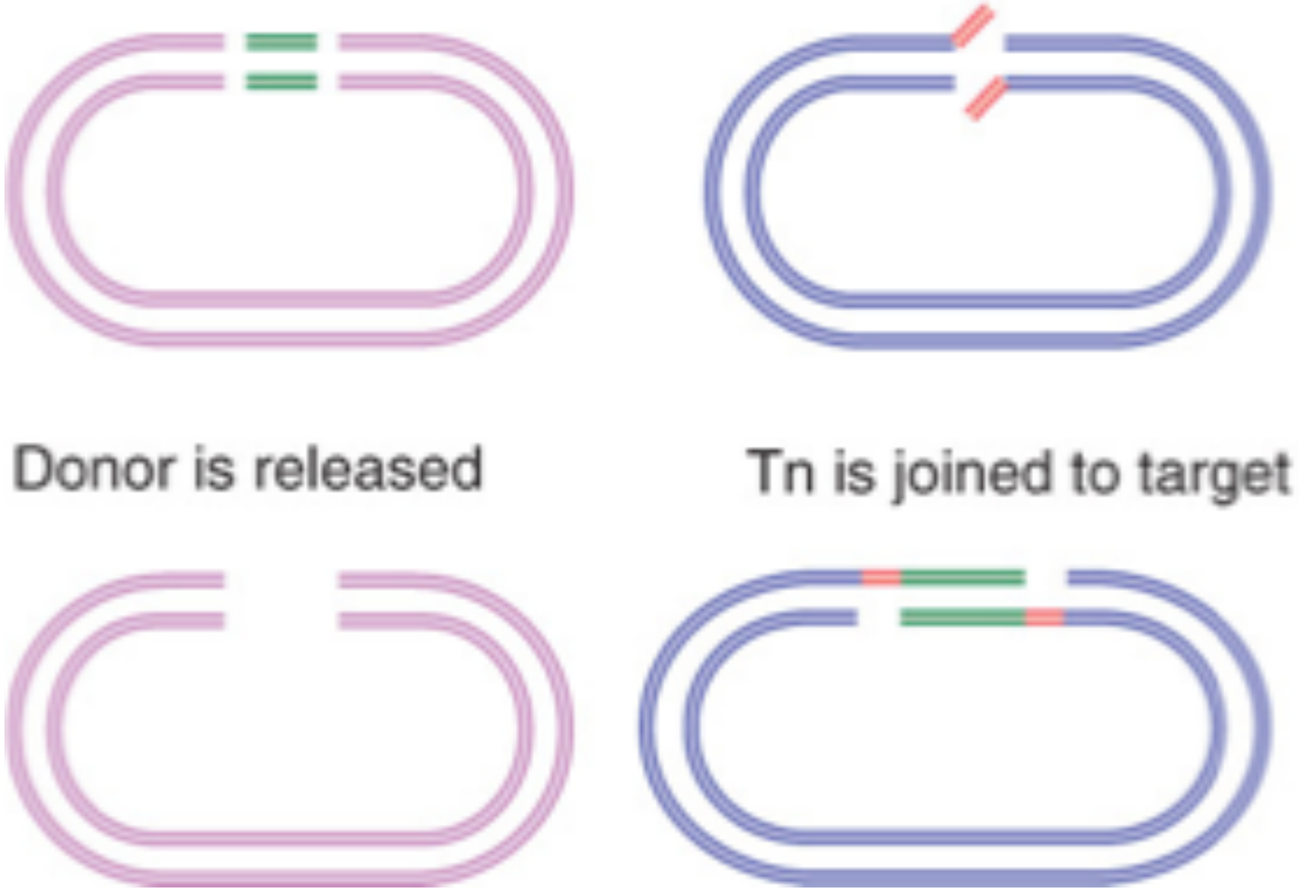


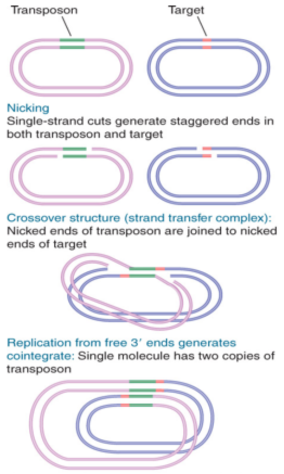
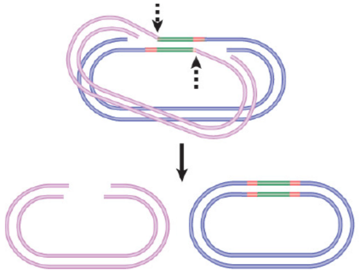
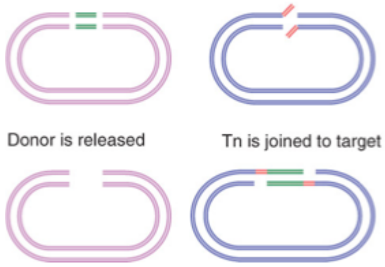
DNA Transposition	Example	Enzyme(s)
Replicative	Tn3	Transposase / Resolvase
	Tn7	Transposase (TnsB) - Endonuclease (TnsA)
Nonreplicative two-strand		
Nonreplicative four-strand		



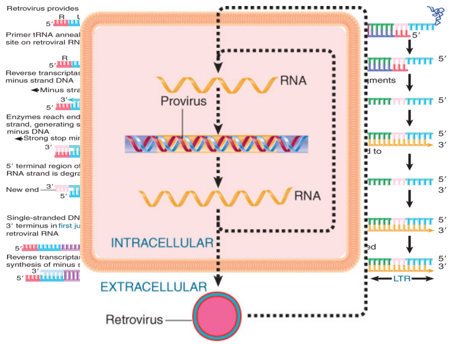
Nicking
Single-strand cuts generate staggered ends in both transposon and target

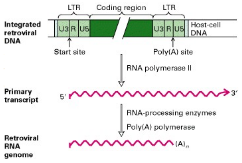
Crossover structure (strand transfer complex):
Nicked ends of transposon are joined to nicked ends of target

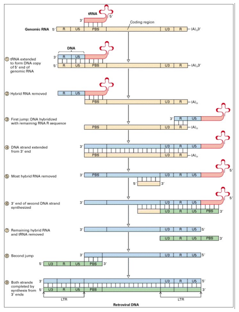


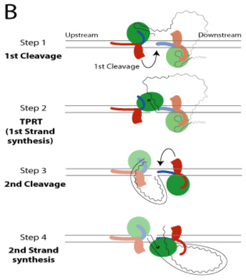
DNA Transposition	Example	Enzyme(s)	
Replicative	Tn3	Transposase / Resolvase	 Nicking Single-strand cuts generate staggered ends in both transposon and target Crossover structure (strand transfer complex): Nicked ends of transposon are joined to nicked ends of target Replication from free 3′ ends generates cointegrate: Single molecule has two copies of transposon
	Tn7	Transposase (TnsB) - Endonuclease (TnsA)	
Nonreplicative two-strand	Tn5	Transposase	
	Tn7	Transposase (TnsB) + Endonuclease (TnsA)	
Nonreplicative four-strand	Tn10	Transposase	
	Ac autonomous (Ds) nonautonomous	Transposase (controlled by Methylation)	

Retrotransposition	Example(s)	Enzyme(s)
Retroviruses	HIV, Feline Leukemia	Reverse Transcriptase + Integrase
LTR Retroelements	Ty elements, Copia-like elements ERV	Reverse Transcriptase + Integrase
TPRT Retroelements	LINES, L1 (humans) autonomous SINES, Alu1 nonautonomous	Reverse Transcriptase endonuclease Initial host transcription (controlled by Methylation)


Retrovirus provides
Primer tRNA anneal site on retroviral RN
Reverse transcriptas minus strand DNA
Enzymes reach end strand, generating s minus DNA
Strong stop mir
5′ terminal region of RNA strand is degra
New end
Single-stranded DN 3′ terminus in first ju retroviral RNA
Reverse transcripta synthesis of minus s
RNA
Provirus
RNA
INTRACELLULAR
EXTRACELLULAR
Retrovirus
LTR

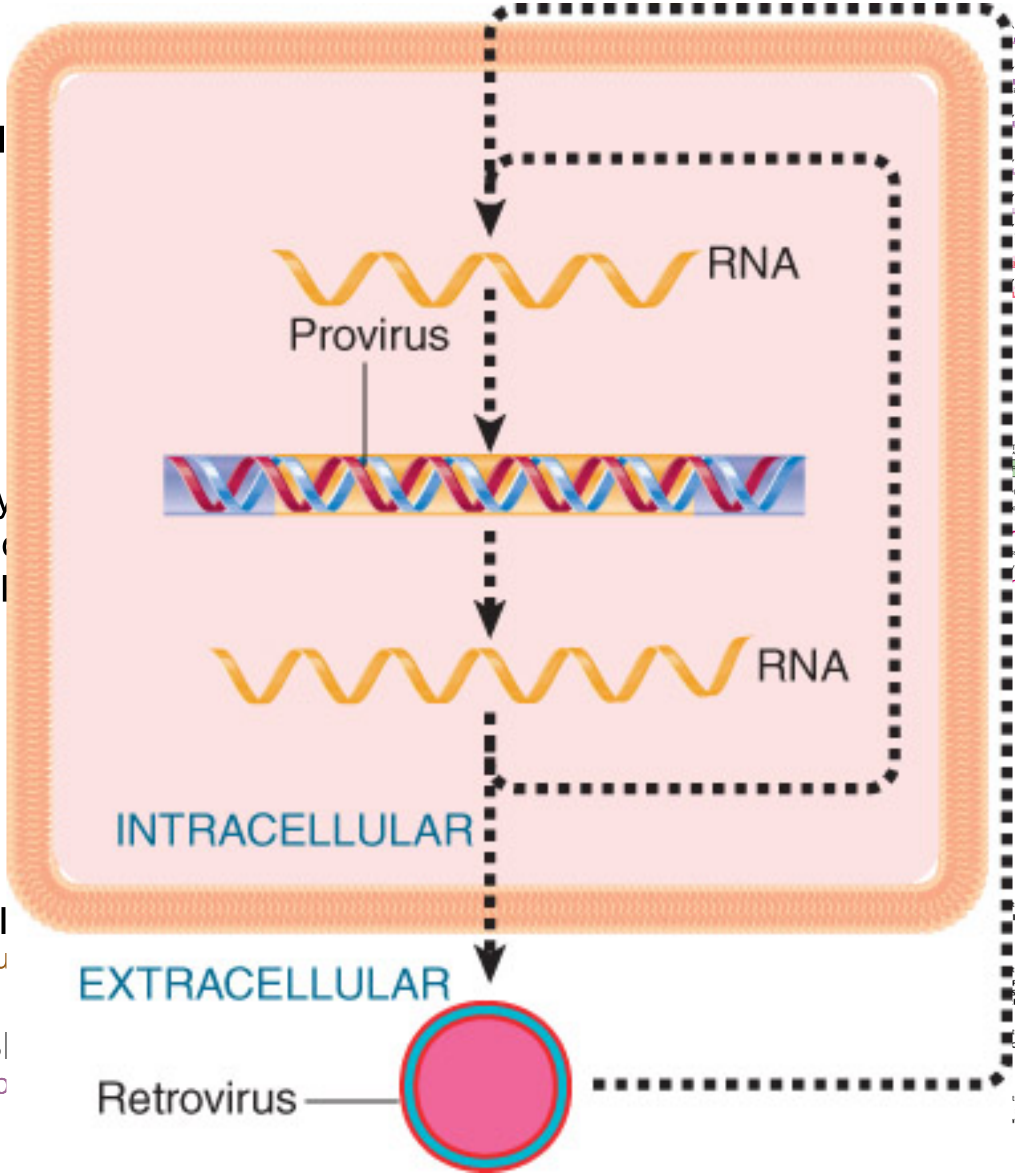


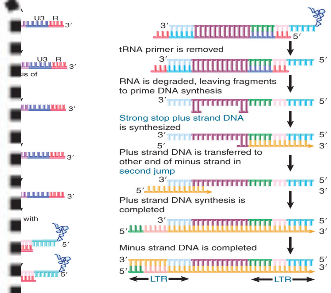




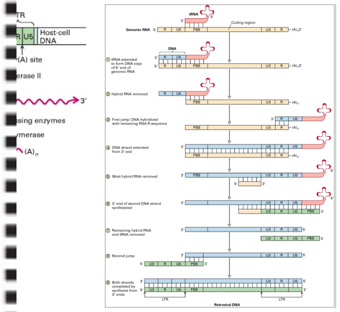
Retrotransposition Example(s) Enzyme(s)

Retroviruses HI



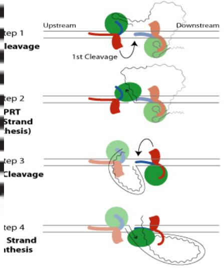

U3 R
tRNA primer is removed
RNA is degraded, leaving fragments to prime DNA synthesis
Strong stop plus strand DNA is synthesized
Plus strand DNA is transferred to other end of minus strand in second jump
Plus strand DNA synthesis is completed
Minus strand DNA is completed
LTR
LTR

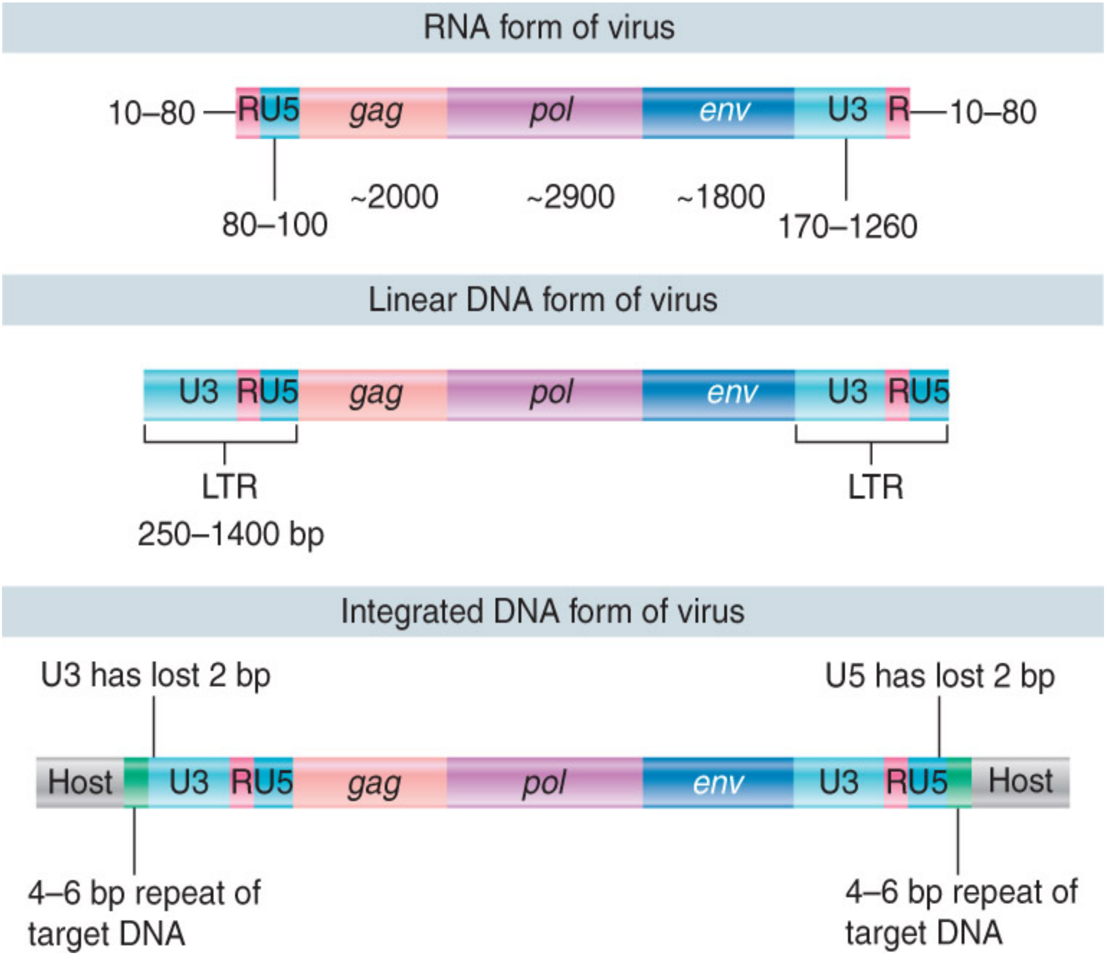
LTR Retroelements Ty
Co
El



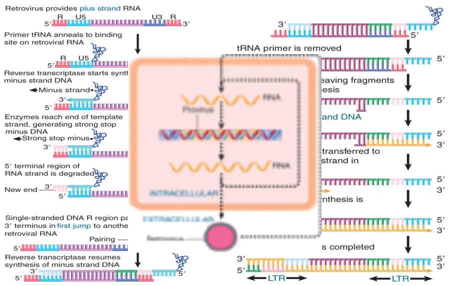
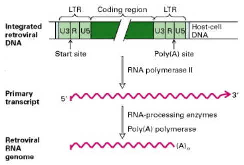
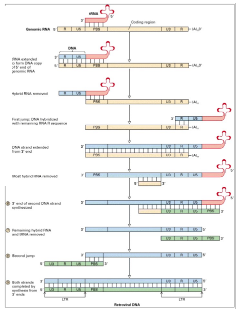
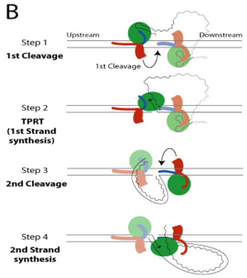
TPRT Retroelements LI
au

SI
no





- Retroviral genomes exist as RNA and DNA sequences
- A short sequence (R) is repeated at each end of the viral RNA.
  - The 5′ and 3′ ends are R-U5 and U3-R, respectively.

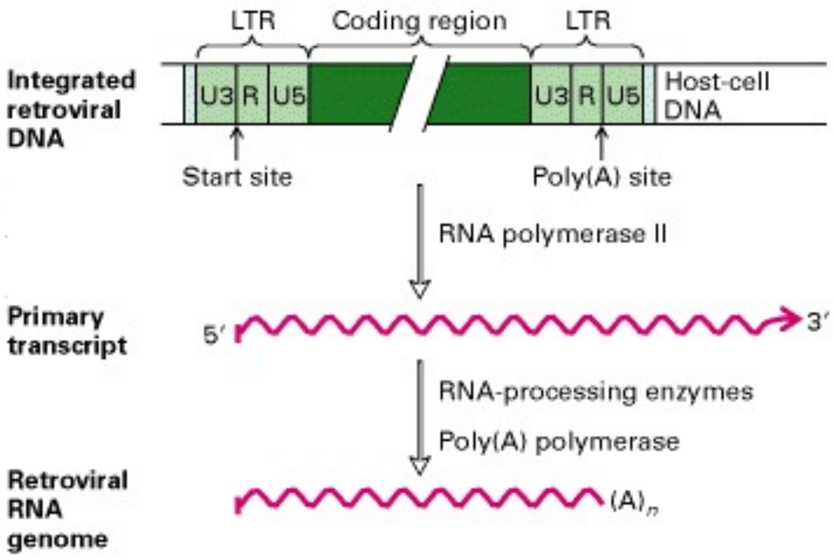
Retrotransposition	Example(s)	Enzyme(s)	
Retroviruses	HIV, Feline Leukemia	Reverse Transcriptase + Integrase	
LTR Retroelements	Ty elements, Copia-like elements ERV	Reverse Transcriptase + Integrase	 
TPRT Retroelements	LINES, L1 (humans) autonomous SINES, Alu1 nonautonomous	Reverse Transcriptase endonuclease Initial host transcription (controlled by Methylation)	

Retrovirus provides plus strand RNA
Primer tRNA anneals to binding site on retroviral RNA
Reverse transcriptase starts synth minus strand DNA
Minus strand
Enzymes reach end of template strand, generating strong stop minus DNA
Strong stop minus
5′ terminal region of RNA strand is degraded
New end
Single-stranded DNA R region pa 3′ terminus in first jump to anothe retroviral RNA
Pairing
Reverse transcriptase resumes synthesis of minus strand DNA
tRNA primer is removed
eaving fragments esis
and DNA
transferred to strand in
nthesis is
s completed
R
U5
U3
LTR
5′
3′

Retrotransposition

Example(s)

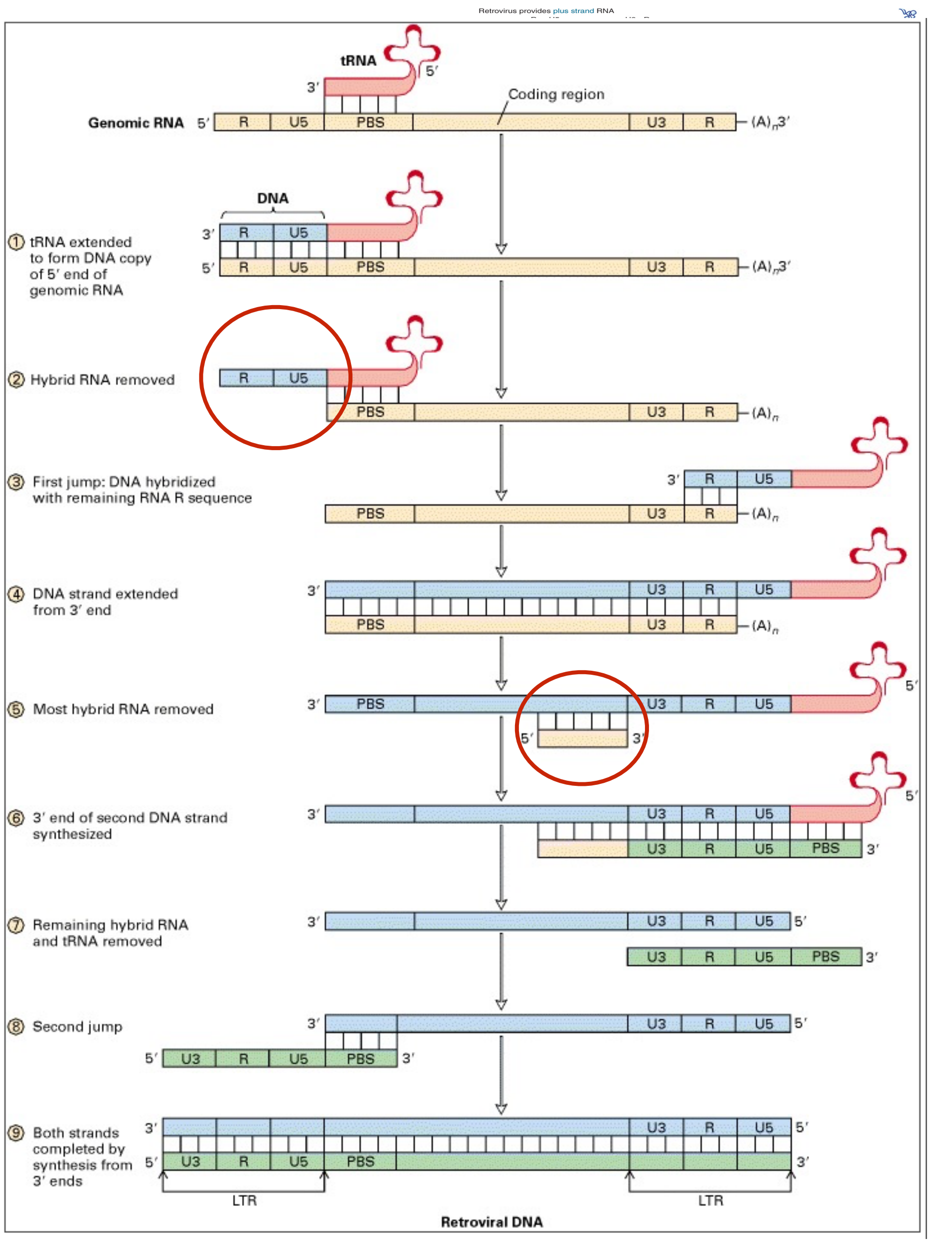
Enzyme(s)



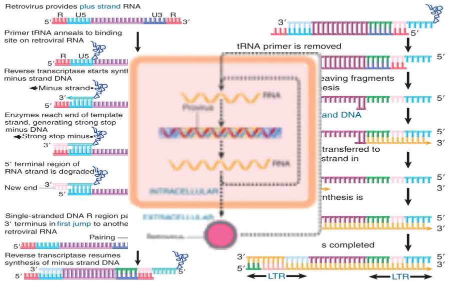
TPRT Retroelements

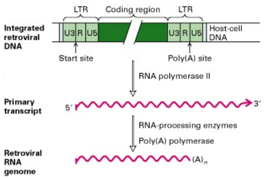
LINES, L1 (humans)
autonomous

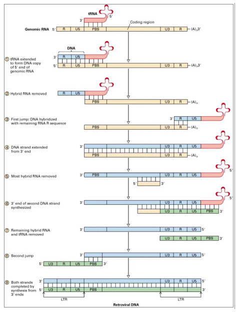
SINES, Alu1
nonautonomous

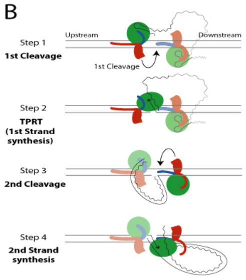

Retrovirus provides plus strand RNA
tRNA
3′
5′
Coding region
Genomic RNA
R
U5
PBS
U3
R
(A)$_n$3′
DNA
① tRNA extended to form DNA copy of 5′ end of genomic RNA
② Hybrid RNA removed
③ First jump: DNA hybridized with remaining RNA R sequence
④ DNA strand extended from 3′ end
⑤ Most hybrid RNA removed
⑥ 3′ end of second DNA strand synthesized
⑦ Remaining hybrid RNA and tRNA removed
⑧ Second jump
⑨ Both strands completed by synthesis from 3′ ends
LTR
LTR
Retroviral DNA

Retrotransposition	Example(s)	Enzyme(s)
Retroviruses	HIV, Feline Leukemia	Reverse Transcriptase + Integrase
LTR Retroelements	Ty elements, Copia-like elements ERV	Reverse Transcriptase + Integrase
TPRT Retroelements	LINES, L1 (humans) autonomous SINES, Alu1 nonautonomous	Reverse Transcriptase endonuclease Initial host transcription (controlled by Methylation)

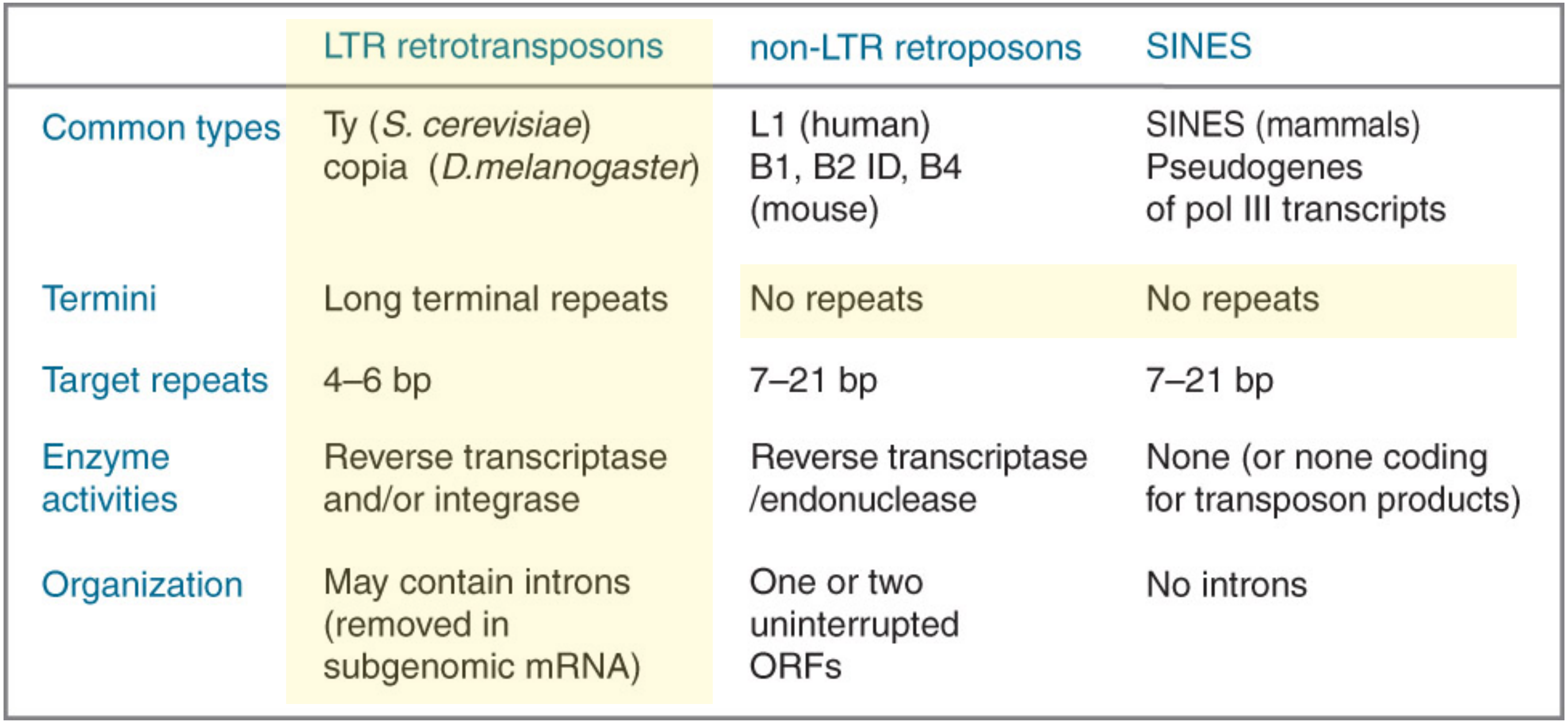

Retrovirus provides plus strand RNA
R
U5
U3
R
5′
3′
Primer tRNA anneals to binding site on retroviral RNA
Reverse transcriptase starts synth
minus strand DNA
Minus strand
Enzymes reach end of template strand, generating strong stop minus DNA
Strong stop minus
5′ terminal region of RNA strand is degraded
New end
Single-stranded DNA R region pa
3′ terminus in first jump to anothe
retroviral RNA
Pairing
Reverse transcriptase resumes synthesis of minus strand DNA
tRNA primer is removed
eaving fragments
esis
and DNA
transferred to
strand in
nthesis is
s completed
LTR
LTR




tRNA
Coding region
Genomic RNA
DNA
PBS
① tRNA extended to form DNA copy of 5′ end of genomic RNA
② Hybrid RNA removed
③ First jump: DNA hybridized with remaining RNA R sequence
④ DNA strand extended from 3′ end
⑤ Most hybrid RNA removed
⑥ 3′ end of second DNA strand synthesized
⑦ Remaining hybrid RNA and tRNA removed
⑧ Second jump
⑨ Both strands completed by synthesis from 3′ ends
LTR
LTR
Retroviral DNA

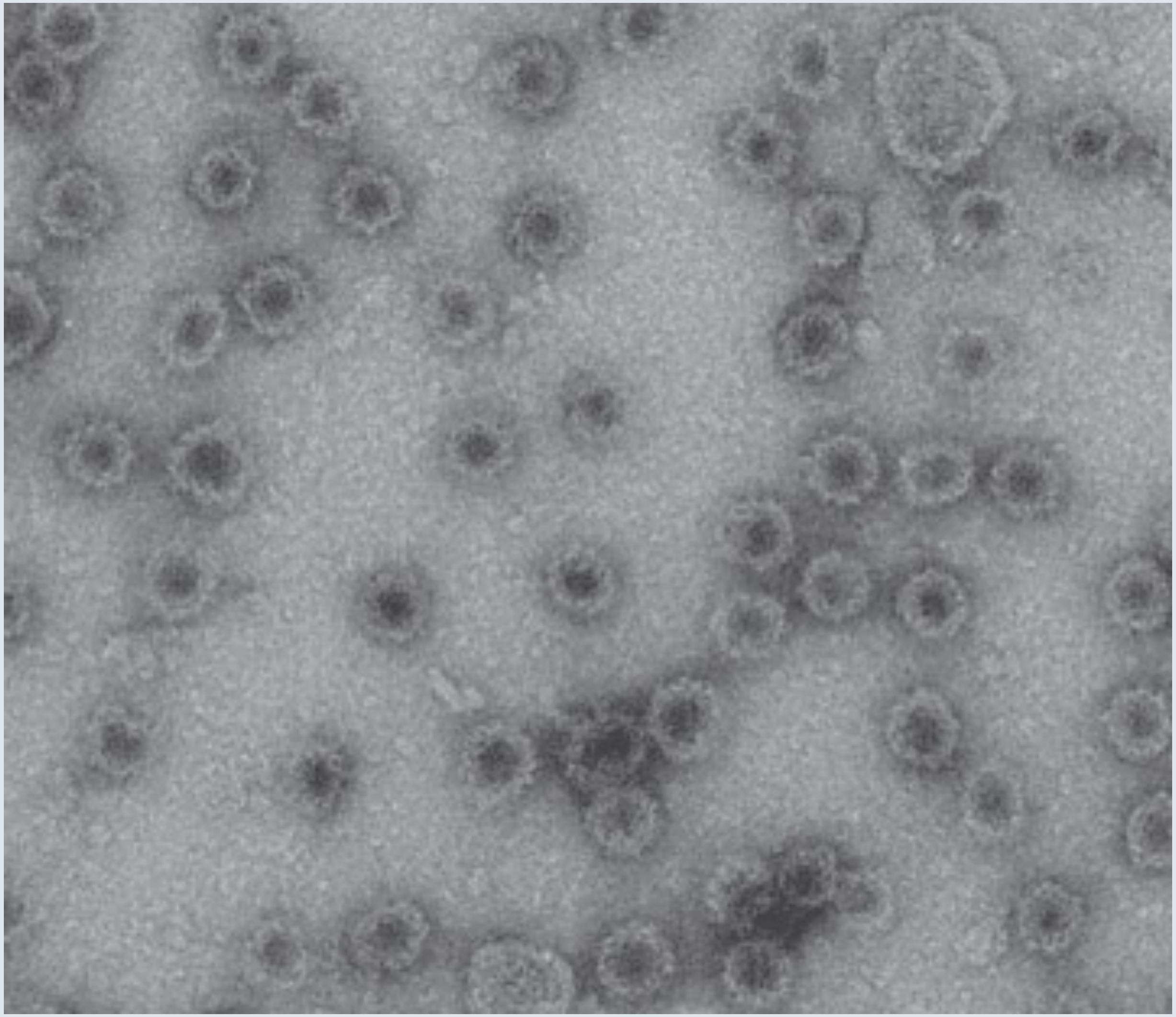


- **Retrotransposons** of the viral superfamily are transposons that mobilize via an RNA that does not form an infectious particle.
- Some **retrotransposons** directly resemble **retroviruses** in their use of LTRs. Others do not, and have no LTRs.



	LTR retrotransposons	non-LTR retroposons	SINES
Common types	Ty (*S. cerevisiae*) copia (*D.melanogaster*)	L1 (human) B1, B2 ID, B4 (mouse)	SINES (mammals) Pseudogenes of pol III transcripts
Termini	Long terminal repeats	No repeats	No repeats
Target repeats	4–6 bp	7–21 bp	7–21 bp
Enzyme activities	Reverse transcriptase and/or integrase	Reverse transcriptase /endonuclease	None (or none coding for transposon products)
Organization	May contain introns (removed in subgenomic mRNA)	One or two uninterrupted ORFs	No introns

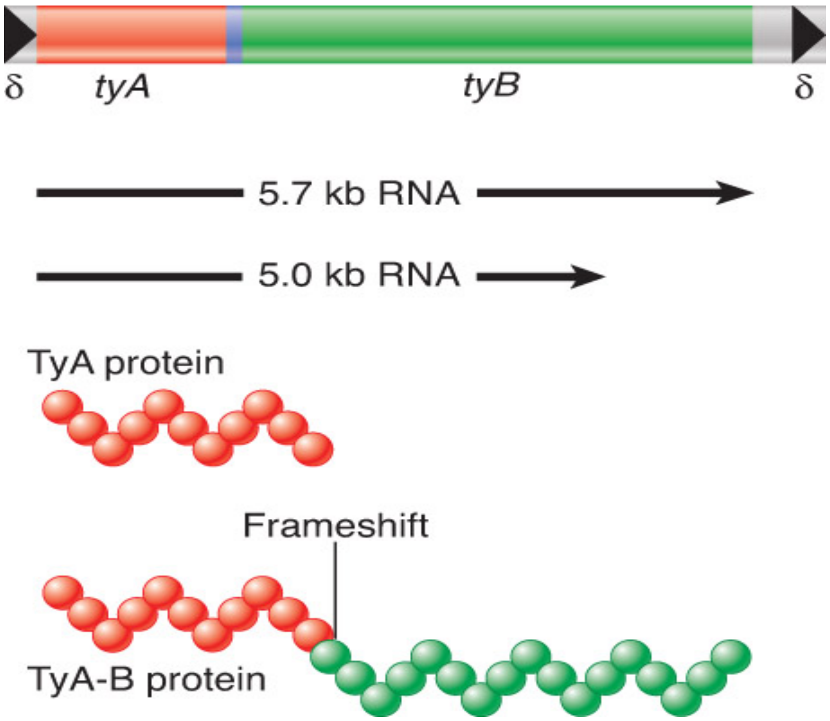
- Despite having an **RT activity**, LINES lack the **LTRs** of the viral superfamily and use a unique mechanism to prime the reverse transcription **rxn.**
- The non-viral superfamily may have originated from RNA sequences;
- **SINES** are derived from RNA polymerase III transcripts.



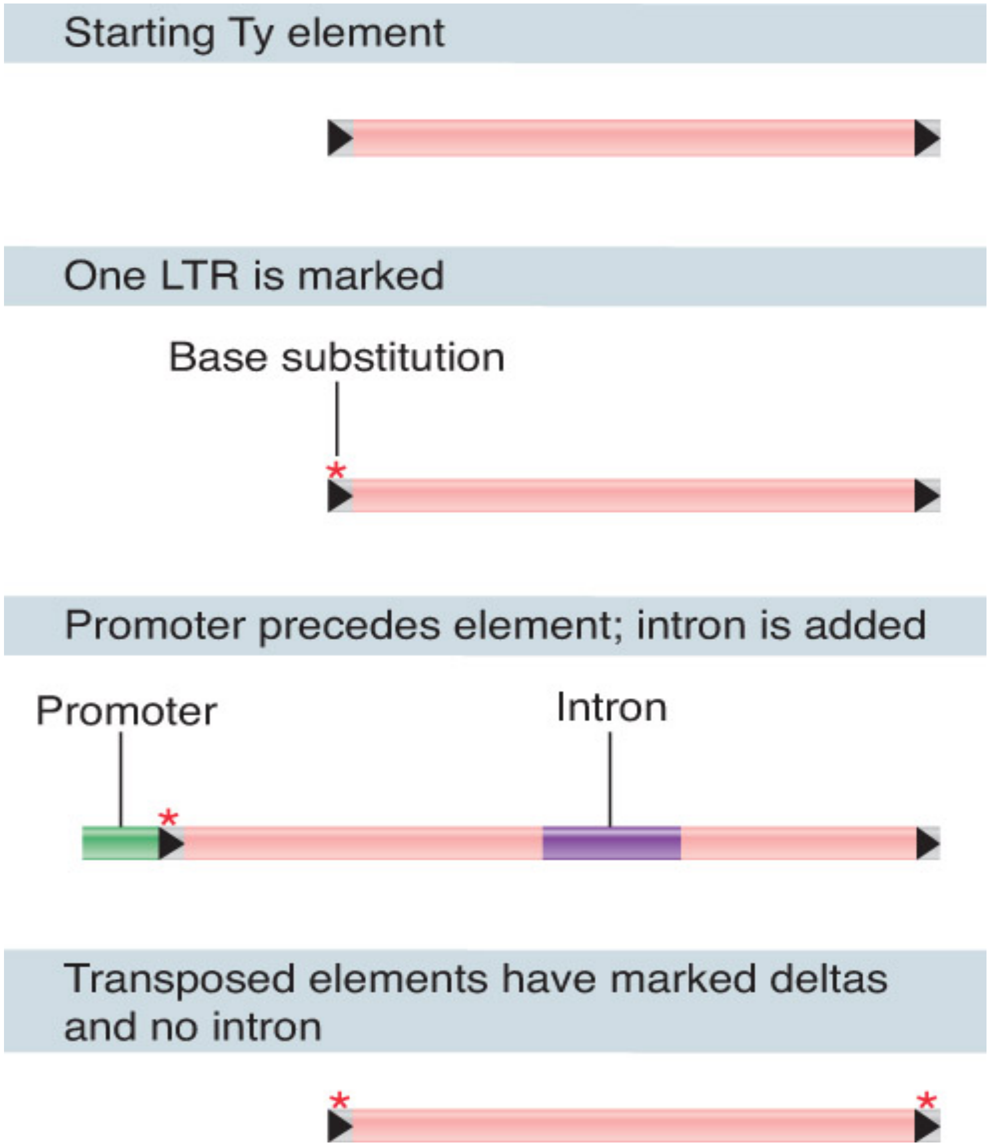
Ty elements in yeast generate virus-like particles.

Reprinted from J. Mol. Biol., vol. 292, H. A. AL-Khayat, et al., Yeast Ty retrotransposons..., pp. 65-73. Copyright 1999, with permission from Elsevier [http://www.sciencedirect.com/science/journal/00222836]. Photo courtesy of Dr. Hind A. AL-Khayat, Imperial College London, United Kingdom.

Ty elements in Yeast (~35 copies per genome) represent a third type of transposositional insertion....**Retrotransposition**.



Ty elements ~5.1kbp in size and encode for a 300 bp **tandem repeats** (δ's), which can be seen scattered around Yeast genomes. They cause **5 bp repeats** in target site and transpose through an **RNA intermediate**!!! How can we know this?



Starting Ty element

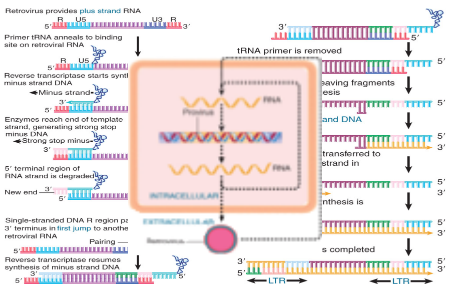
One LTR is marked

Promoter precedes element; intron is added

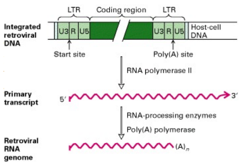
Transposed elements have marked deltas and no intron

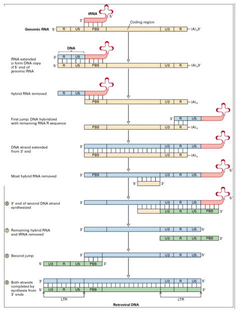
Ty transposes through a spliced RNA form

Retrotransposition	Example(s)	Enzyme(s)
Retroviruses	HIV, Feline Leukemia	Reverse Transcriptase + Integrase
LTR Retroelements	Ty elements, Copia-like elements ERV	Reverse Transcriptase + Integrase
TPRT Retroelements	LINES, L1 (humans) autonomous SINES, Alu1 nonautonomous	Reverse Transcriptase endonuclease Initial host transcription (controlled by Methylation)

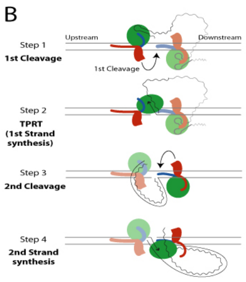


Retrovirus provides plus strand RNA
R
U5
U3
R
5′
3′
Primer tRNA anneals to binding site on retroviral RNA
Reverse transcriptase starts synth minus strand DNA
Minus strand
Enzymes reach end of template strand, generating strong stop minus DNA
Strong stop minus
5′ terminal region of RNA strand is degraded
New end
Single-stranded DNA R region pa 3′ terminus in first jump to anothe retroviral RNA
Pairing
Reverse transcriptase resumes synthesis of minus strand DNA
tRNA primer is removed
eaving fragments esis
and DNA
transferred to strand in
nthesis is
s completed
RNA
Provirus
INTRACELLULAR
EXTRACELLULAR
LTR
LTR



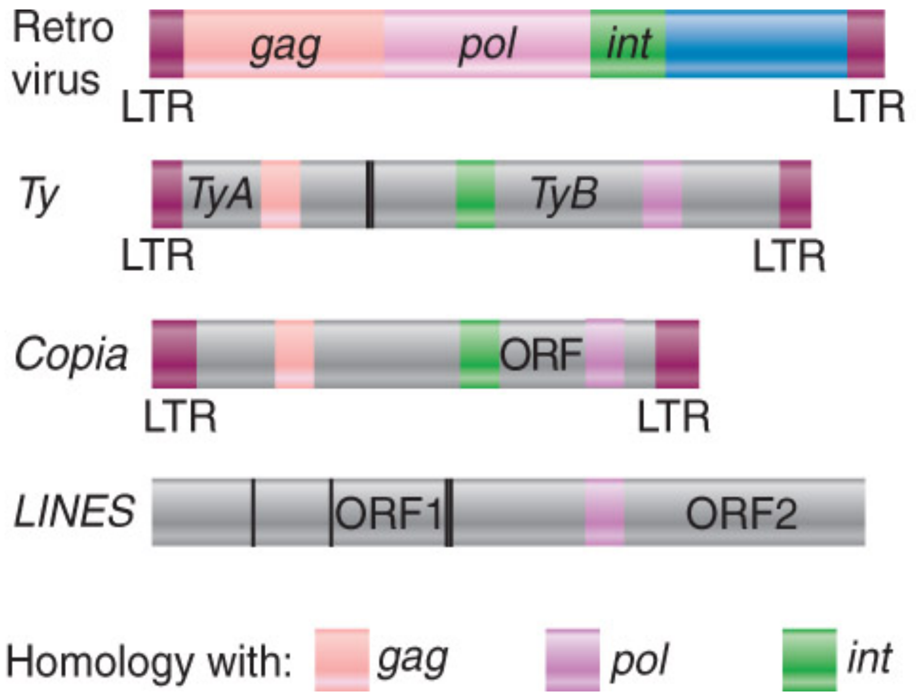


tRNA
Genomic RNA
Coding region
DNA
Hybrid RNA removed
First jump: DNA hybridized with remaining RNA R sequence
DNA strand extended from 3′ end
Most hybrid RNA removed
3′ end of second DNA strand synthesized
Remaining hybrid RNA and tRNA removed
Second jump
Both strands completed by synthesis from 3′ ends
LTR
LTR
Retroviral DNA

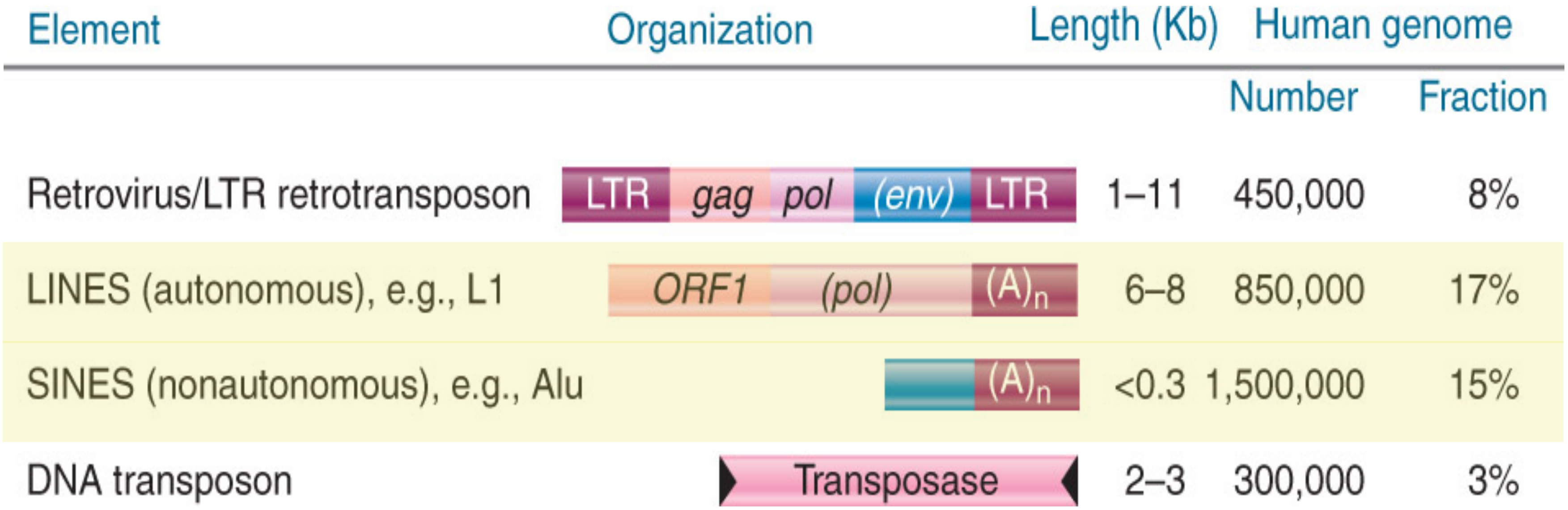


- Although **retroelements** that lack LTRs, also transpose *via* reverse transcriptase, they employ a distinct method of integration and are phylogenetically distinct from both **retroviruses** and **LTR retrotransposons**.

- Other elements can be found that were generated by an RNA-mediated transposition event, but they do not themselves code for enzymes that can catalyze transposition.



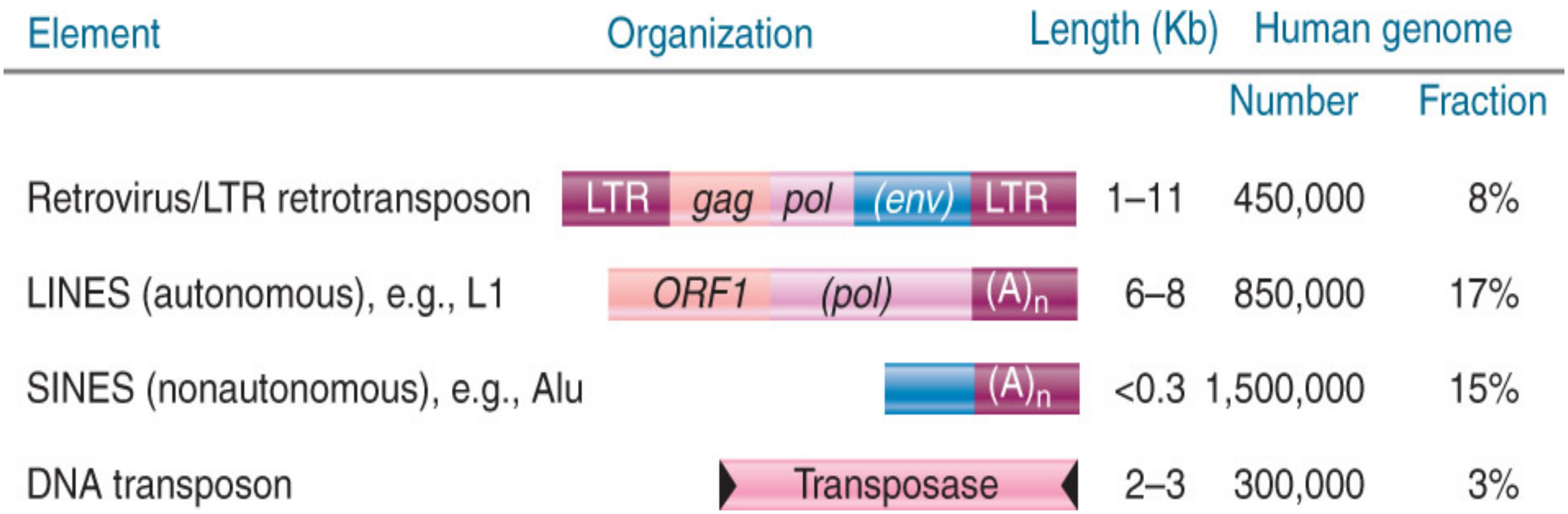
- **Retroelements** constitute almost half (48%) of the human genome.



Element	Organization	Length (Kb)	Human genome	
			Number	Fraction
Retrovirus/LTR retrotransposon	LTR *gag* *pol* *(env)* LTR	1–11	450,000	8%
LINES (autonomous), e.g., L1	*ORF1* *(pol)* $(A)_n$	6–8	850,000	17%
SINES (nonautonomous), e.g., Alu	$(A)_n$	<0.3	1,500,000	15%
DNA transposon	Transposase	2–3	300,000	3%

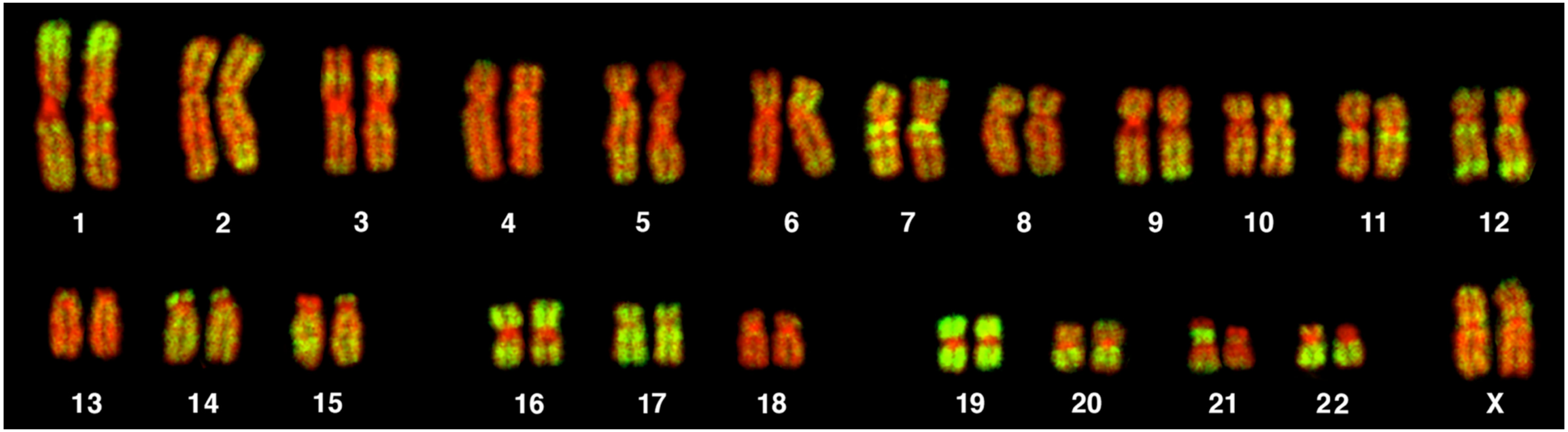
- LINES and SINES comprise a major part of the animal genome. They were originally defined by the existence of a large number of relatively short sequences that were related to one another.

- They are described as interspersed nuclear elements because of their common occurrence and widespread distribution. **L1** = active human LINES; **ALU** = active human SINES



Element	Organization	Length (Kb)	Human genome	
			Number	Fraction
Retrovirus/LTR retrotransposon	LTR *gag* *pol* *(env)* LTR	1–11	450,000	8%
LINES (autonomous), e.g., L1	*ORF1* *(pol)* $(A)_n$	6–8	850,000	17%
SINES (nonautonomous), e.g., Alu	$(A)_n$	<0.3	1,500,000	15%
DNA transposon	Transposase	2–3	300,000	3%

- **short-interspersed elements (SINEs)** – A major class of short (<500 bp) **nonautonomous** retrotransposons that occupy ~13 -15% of the human genome.
  - **Alu element** – One of a set of dispersed, related sequences, each ~300 bp long, in the human genome (members of the SINE family).



Karyotype from a female human lymphocyte (46, XX). Chromosomes were hybridized with a probe for Alu elements (green) and counterstained with TOPRO-3 (red). Alu elements were used as a marker for chromosomes and chromosome bands rich in genes.

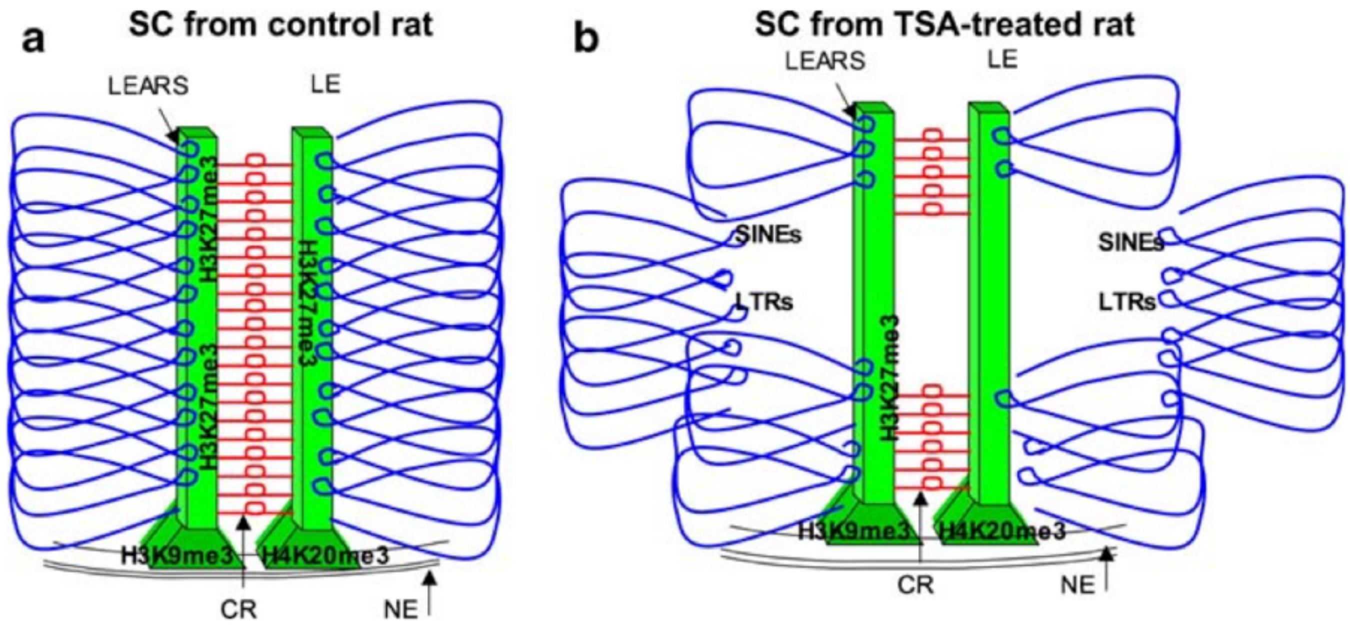
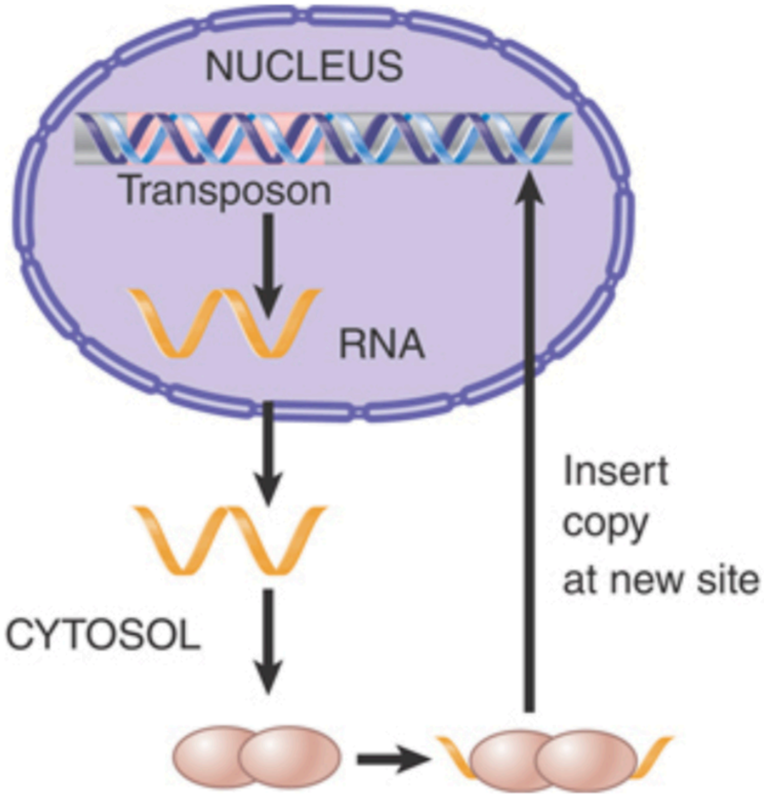
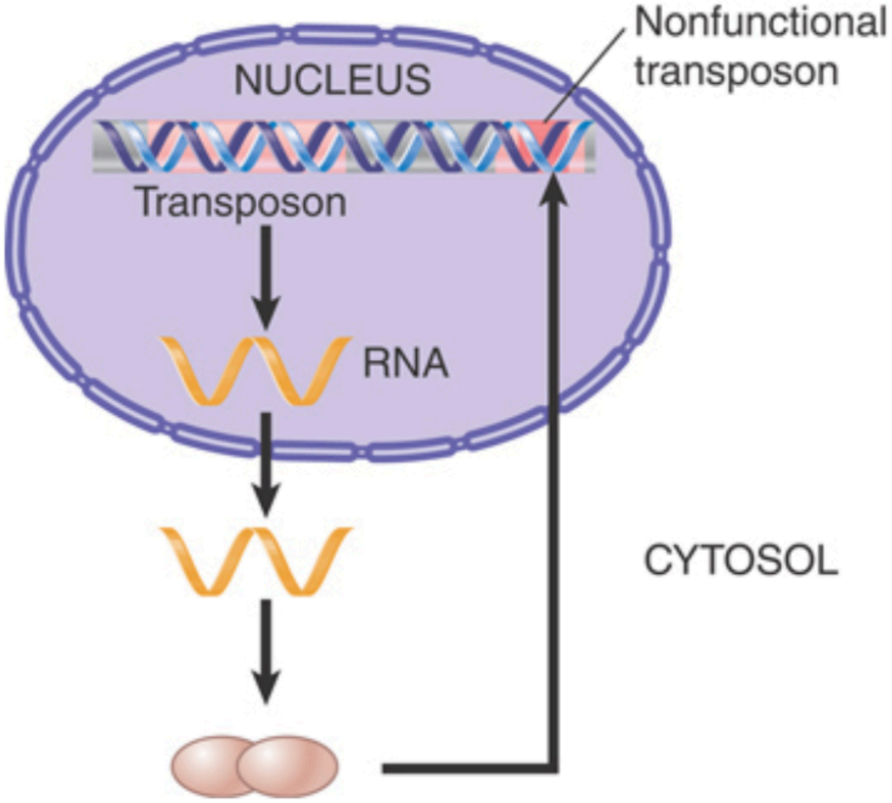


Fig. 14 Model of SC structure in control and TSA-treated rats. a SC of a control rat. The chromatin of homologous chromosomes is anchored to the lateral elements (*LE*) through lateral elements-associated repeat sequences (*LEARS*), for which chromatin structure is dictated by histone posttranslational modifications like H3K9me3, H3K27me3, and H4K20me3. **b** Upon inhibition of histone deacetylases, the presence of H3K27me3 in SINE and LTR sequences decreases dramatically, which could favor detachment of such sequences from the LEs. This is accompanied by alteration of the SC's central region (*CR*)



A LINE is transcribed into an RNA that is translated into proteins that assemble into a complex with the RNA. The complex translocates to the nucleus, where it inserts a DNA copy into the genome



A transposon is transcribed into an RNA that is translated into proteins that move independently to the nucleus, where they act on any pair of inverted repeats with the same sequence as the original transposon.

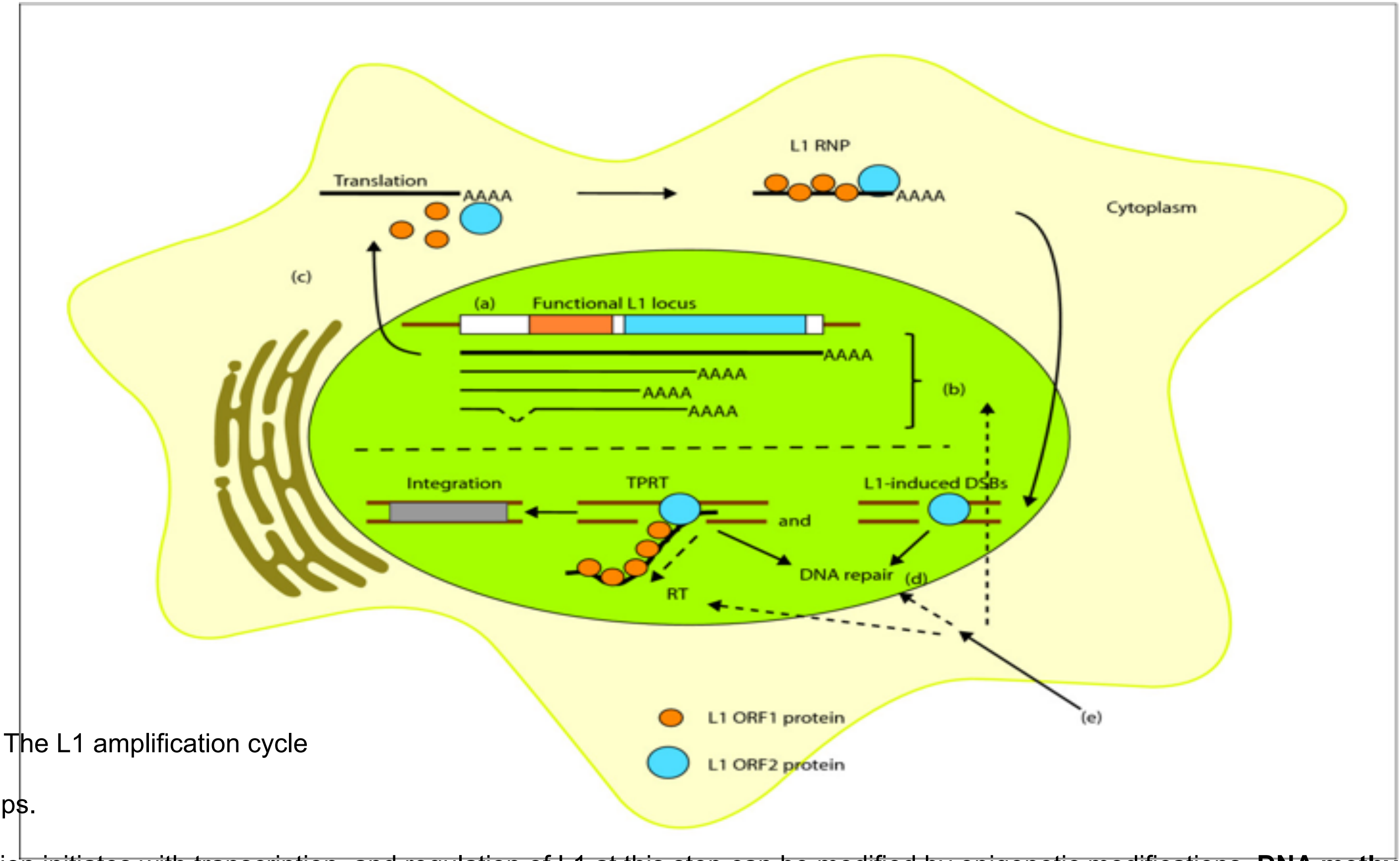


Figure 2

Modulators of the **L1** lifecycle. The L1 amplification cycle

can be divided into several steps.

(a) Transcription. L1 amplification initiates with transcription, and regulation of L1 at this step can be modified by epigenetic modifications, **DNA methylation, and recruitment of transcription factors**.

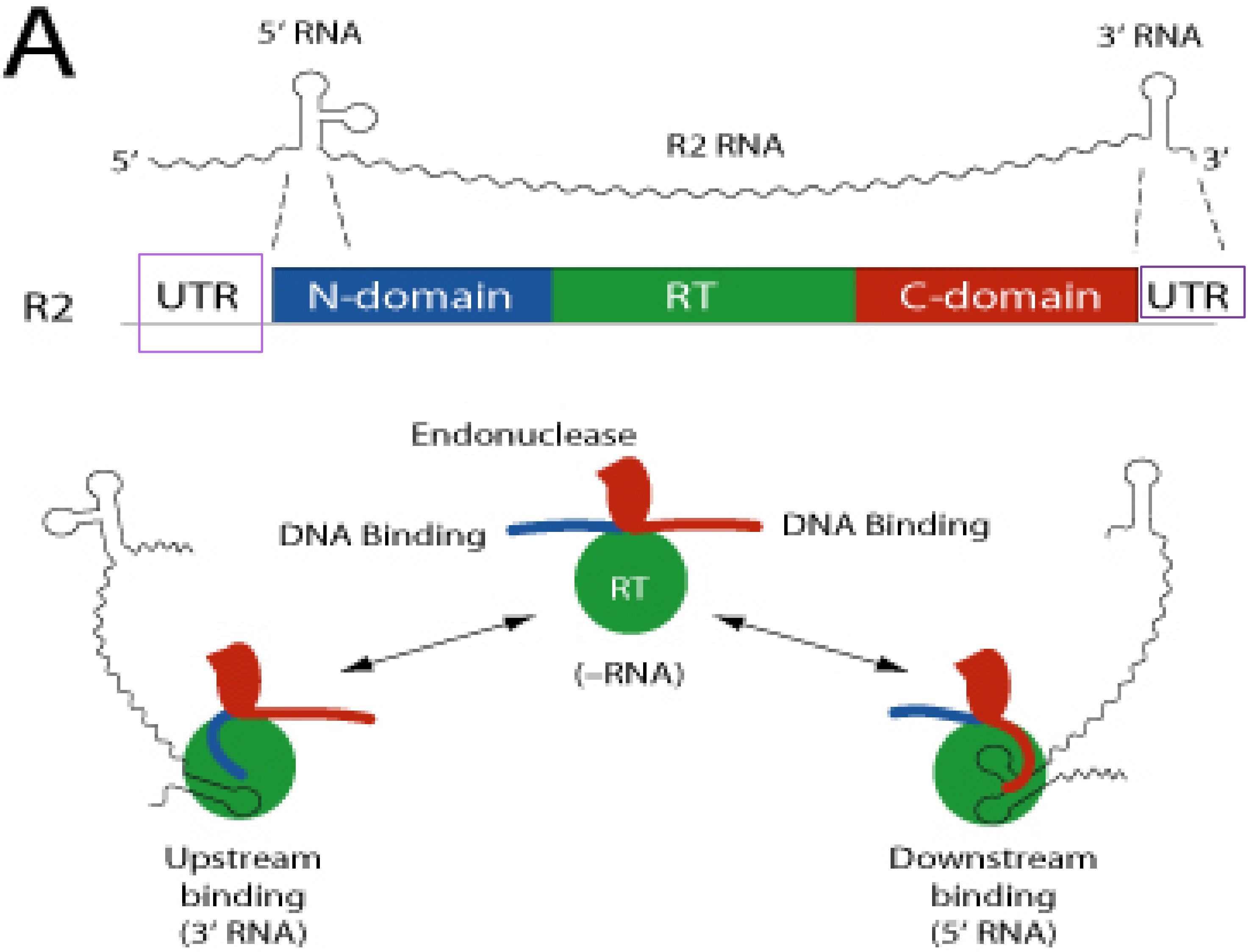
(b) Before leaving the nucleus, the number of retrocompetent full length **L1** transcripts can be reduced by RNA processing through premature polyadenylation and splicing.

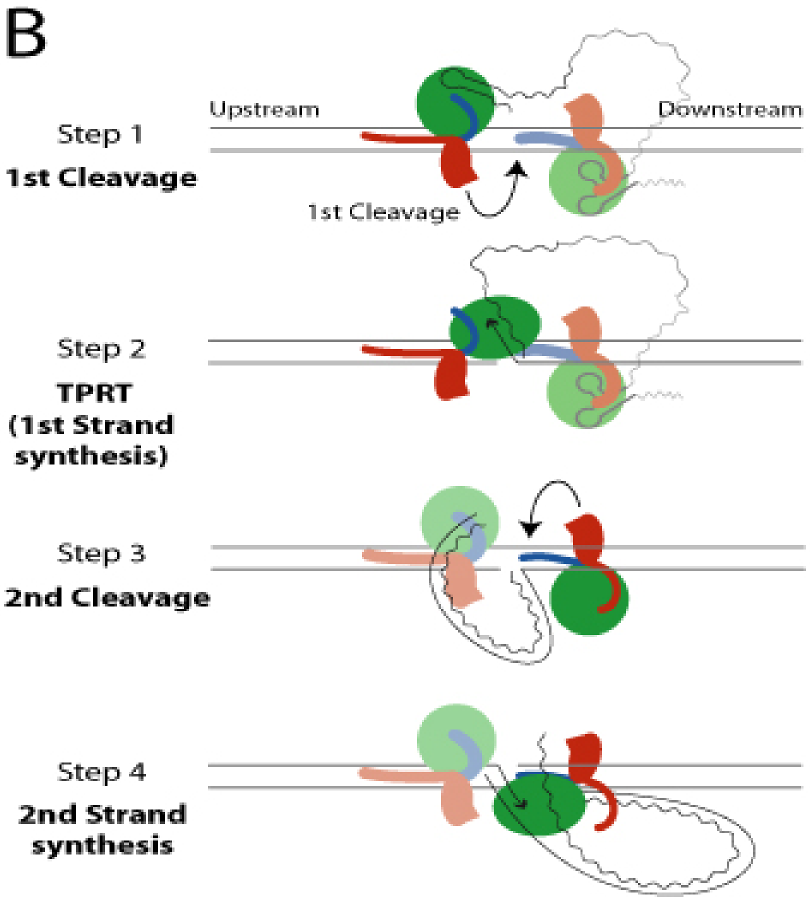
(c) Translation. Full length L1 enters the cytoplasm to be translated, producing **ORF1** and **ORF2** proteins for retrotransposition. The two proteins interact with the L1 transcript to form an **L1 ribonucleoprotein** particle (RNP). RNA interference can affect this step.

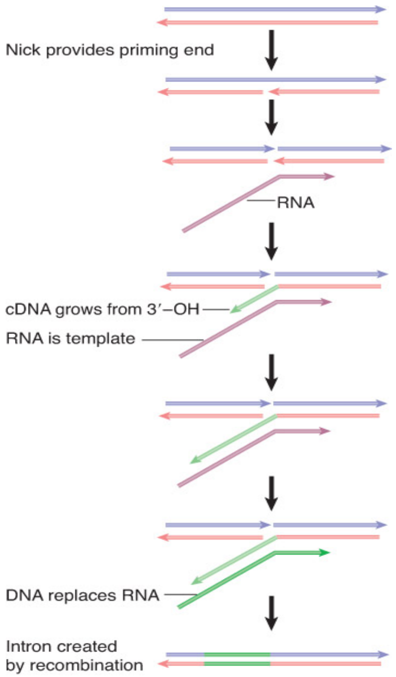
(d) Insertion of a new **L1** copy. The L1 RNP reaches the nucleus, where the DNA is cleaved by the L1 **ORF2** endonuclease activity. It is proposed that reverse transcription occurs through a process referred to as “target primed reverse transcription” (**TPRT**) [71]. The **L1 ORF2** reverse transcriptase activity generates the first strand of DNA. DNA repair proteins are likely to be involved in inhibiting the **L1** integration step.

(e) Effects of external stimuli. Ionizing radiation or heavy metals can affect L1 at multiple steps, such as transcriptional activation or altering DNA repair pathways.

Belancio *et al.: Genome Medicine* 2009, **1:97**


A
5′ RNA
3′ RNA
5′
R2 RNA
3′
R2
UTR
N-domain
RT
C-domain
UTR
Endonuclease
DNA Binding
DNA Binding
RT
(–RNA)
Upstream binding (3′ RNA)
Downstream binding (5′ RNA)





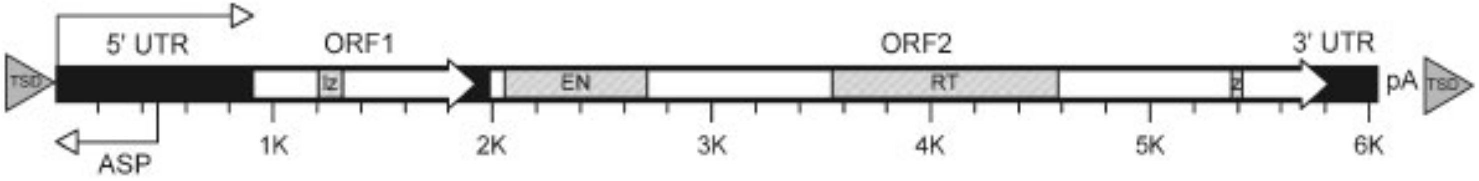
(Step 1) the **endonuclease** (red oval) from the upstream subunit is responsible for first strand cleavage.

(Step 2) The **RT** (green oval) of the upstream subunit catalyzes reverse transcription of the RNA template using the cleaved DNA target site as primer, a reaction we call **Target Primed Reverse Transcription**, **TPRT**.

(Step 3) The downstream subunit cleaves the second DNA strand.

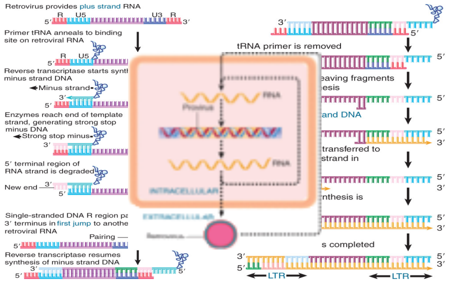
(Step 4) The downstream subunit provides the polymerase to perform **second strand TPRT** displacing the RNA template as it uses the first DNA strand as template.

- **DNA transposons** and **LTR elements** are believed to be “extinct” in the human genome, but the average human carries approximately **80 – 100 potentially active L1 elements** in a diploid genome.
- **LINES** and **SINES** are NOT extinct
- **L1** is active in BOTH the **germ line** and **somatic cells.**
- The full length human **L1 retrotransposon** is 6kb and contains
  - a 910 bp 5’-UTR with bidirectional promoter activity
  - An **ORF1** region which encodes an **RN binding protein** with a **leucine zipper domain**
  - An **ORF2** region which encodes a 150 kDa protein with **endonuclease and reverse transcriptase activities**
  - A 3’-UTR which contains a functional **polyadenylation sequence**
- The L1 element is flanked by 2 to 20 bp target site duplications.

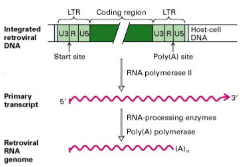


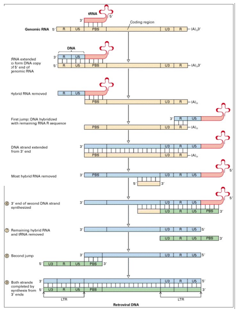
- The EN domain is thought to originate from a host endonuclease present in early eukaryotes. The element can move without a functional EN, but the endonuclease-independent integration is less efficient and occurs rarely.

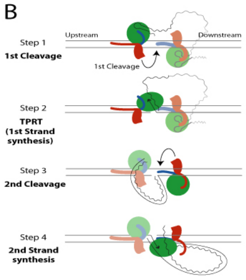
Retrotransposition	Example(s)	Enzyme(s)
Retroviruses	HIV, Feline Leukemia	Reverse Transcriptase + Integrase
LTR Retroelements	Ty elements, Copia-like elements ERV	Reverse Transcriptase + Integrase
TPRT Retroelements	LINES, L1 (humans) autonomous SINES, Alu1 nonautonomous	Reverse Transcriptase endonuclease Initial host transcription (controlled by Methylation)



Retrovirus provides plus strand RNA
R
U5
U3
R
5′
3′
Primer tRNA anneals to binding site on retroviral RNA
Reverse transcriptase starts synt
minus strand DNA
Minus strand
Enzymes reach end of template strand, generating strong stop minus DNA
Strong stop minus
5′ terminal region of RNA strand is degraded
New end
Single-stranded DNA R region pa
3′ terminus in first jump to anothe
retroviral RNA
Pairing
Reverse transcriptase resumes synthesis of minus strand DNA
tRNA primer is removed
eaving fragments
esis
and DNA
transferred to
strand in
nthesis is
s completed
LTR
LTR
RNA
Provirus
RNA
INTRACELLULAR
EXTRACELLULAR



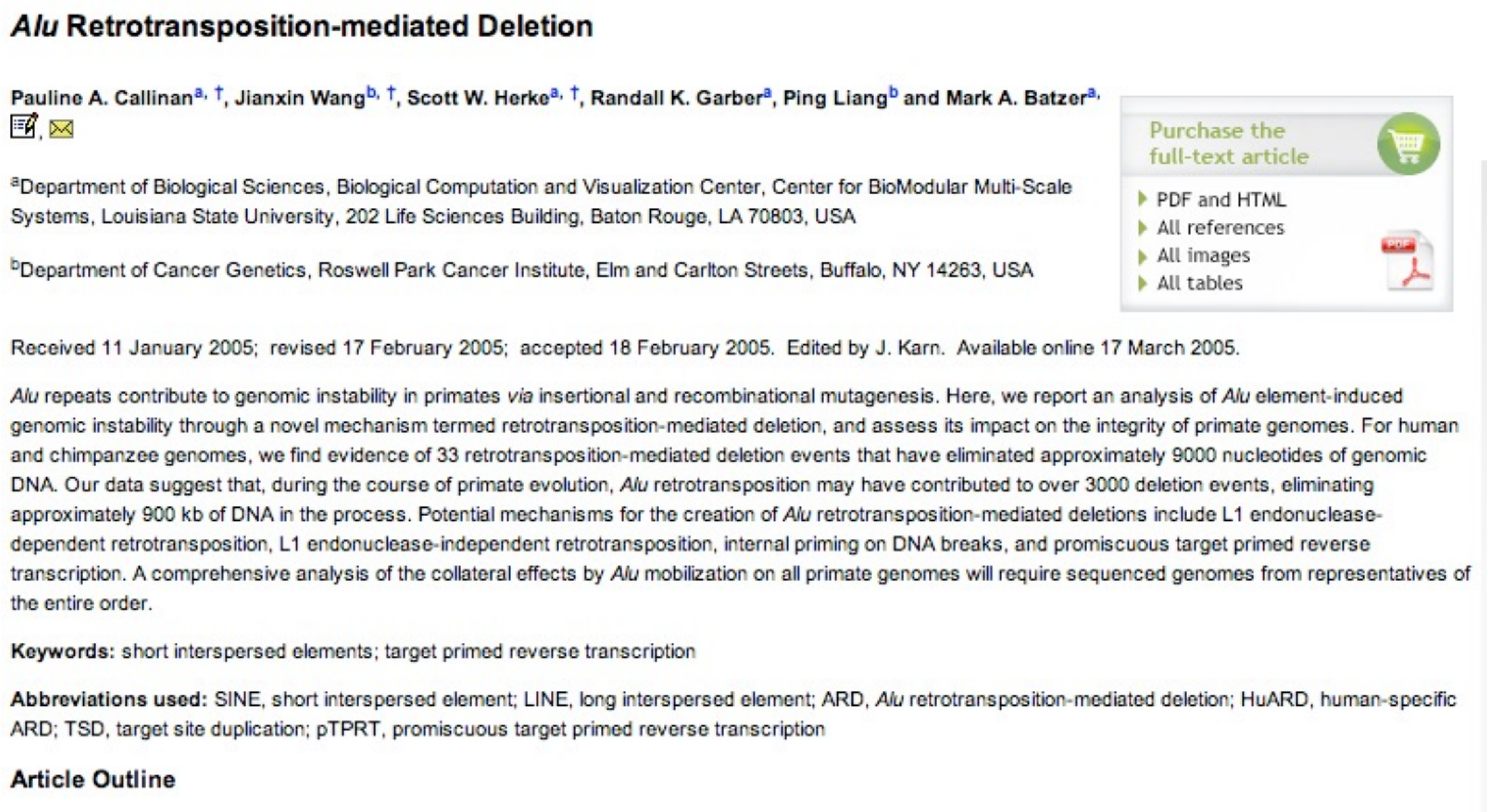




2005: Callinan Pauline A; Wang Jianxin; Herke Scott W; Garber Randall K; Liang Ping; Batzer Mark A

Alu retrotransposition-mediated deletion.

Journal of molecular biology 2005;348(4):791-800



Alu Retrotransposition-mediated Deletion

Pauline A. Callinan[a, †], Jianxin Wang[b, †], Scott W. Herke[a, †], Randall K. Garber[a], Ping Liang[b] and Mark A. Batzer[a,]

[a]Department of Biological Sciences, Biological Computation and Visualization Center, Center for BioModular Multi-Scale Systems, Louisiana State University, 202 Life Sciences Building, Baton Rouge, LA 70803, USA

[b]Department of Cancer Genetics, Roswell Park Cancer Institute, Elm and Carlton Streets, Buffalo, NY 14263, USA

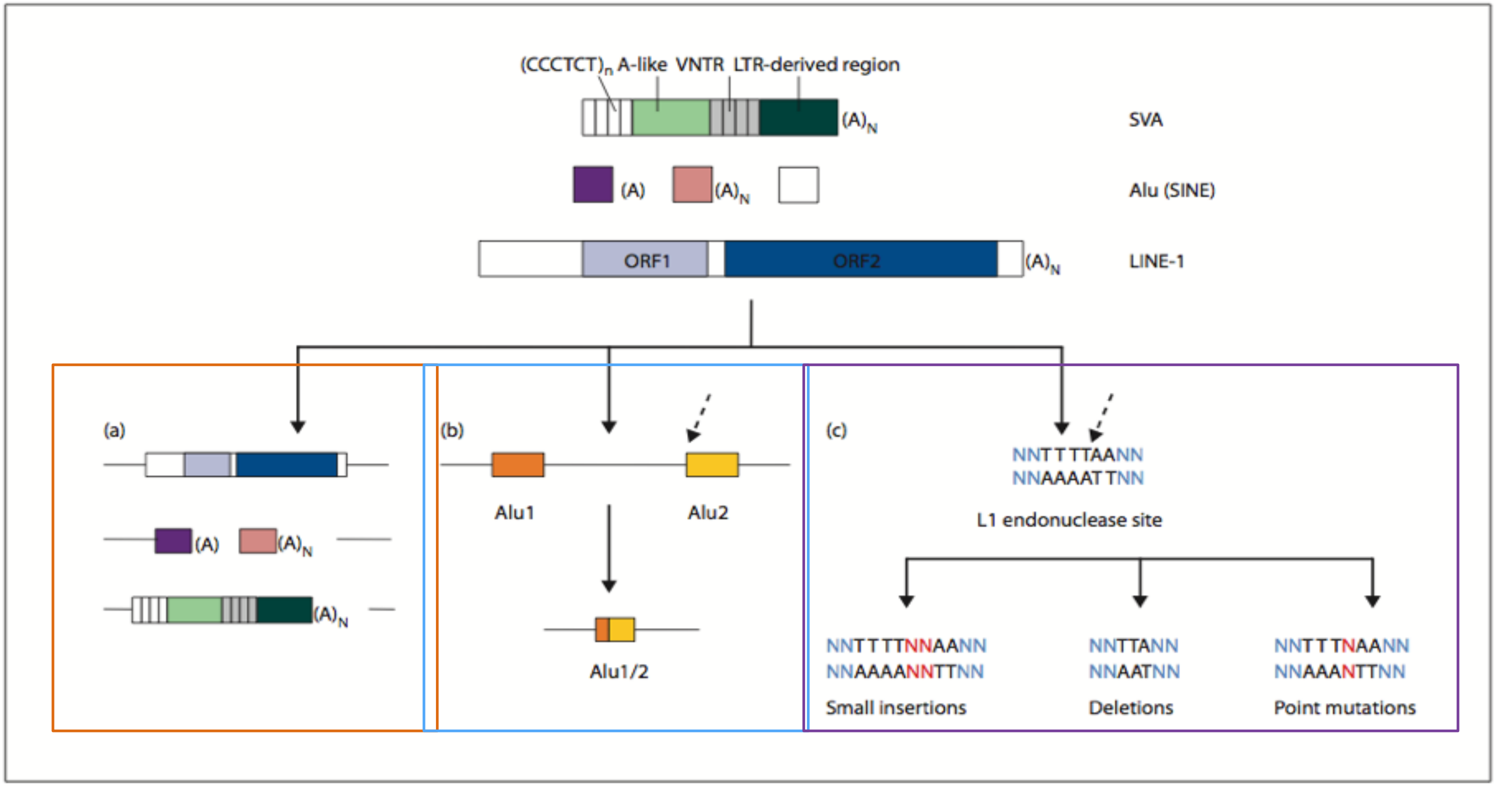
Received 11 January 2005; revised 17 February 2005; accepted 18 February 2005. Edited by J. Karn. Available online 17 March 2005.

Alu repeats contribute to genomic instability in primates *via* insertional and recombinational mutagenesis. Here, we report an analysis of *Alu* element-induced genomic instability through a novel mechanism termed retrotransposition-mediated deletion, and assess its impact on the integrity of primate genomes. For human and chimpanzee genomes, we find evidence of 33 retrotransposition-mediated deletion events that have eliminated approximately 9000 nucleotides of genomic DNA. Our data suggest that, during the course of primate evolution, *Alu* retrotransposition may have contributed to over 3000 deletion events, eliminating approximately 900 kb of DNA in the process. Potential mechanisms for the creation of *Alu* retrotransposition-mediated deletions include L1 endonuclease-dependent retrotransposition, L1 endonuclease-independent retrotransposition, internal priming on DNA breaks, and promiscuous target primed reverse transcription. A comprehensive analysis of the collateral effects by *Alu* mobilization on all primate genomes will require sequenced genomes from representatives of the entire order.

Keywords: short interspersed elements; target primed reverse transcription

Abbreviations used: SINE, short interspersed element; LINE, long interspersed element; ARD, *Alu* retrotransposition-mediated deletion; HuARD, human-specific ARD; TSD, target site duplication; pTPRT, promiscuous target primed reverse transcription

Article Outline



L1 expression leads to different types of DNA damage.

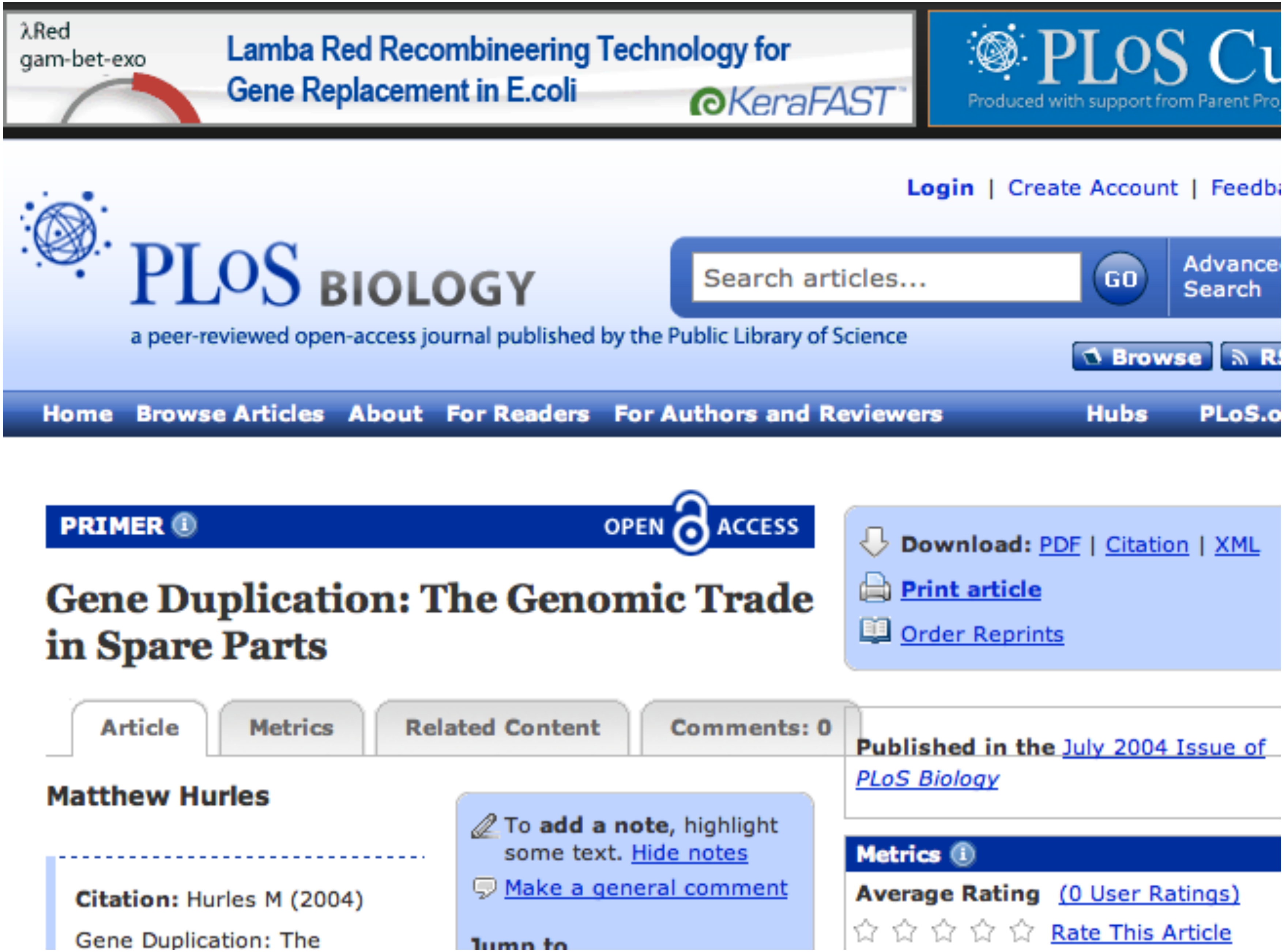
Schematic structures of an **SVA element** (labeled SVA), showing the CCCTCT repeat, the **Alu derived (A-like) region**, the variable number tandem repeat (VNTR) region, and the long terminal repeat (LTR)derived region; an Alu element (labeled Alu (SINE)), showing left (purple) and right (pink) halves separated by the Arich region (A) and the variable length Atail ((A)n) followed by the 3' region (white), which has a variable length and sequence; and an L1 element (labeled LINE1), showing open reading frame (ORF)1 (light blue) and ORF2 (dark blue) and the 5' untranslated region, interORF region and 3' untranslated region (white).

(a) The typical insertion of these elements into the genome, which can lead to insertional mutagenesis. In breast cancer **BRCA1** and **BRCA2** are also known to be disrupted by **TE insertion**

(b) Dispersed repetitive elements such as **Alu elements** can undergo **non-allelic homologous recombination**, which can cause a deletion (shown) or duplication (not shown). The dashed arrow indicates the potential site of DNA damage by an L1 endonuclease that may help initiate these recombination events.

(c) Potential outcomes of the repair of the **L1 induced double strand breaks** (**DSBs**). The **L1** recognition site is in black; surrounding sequence is in blue; inserted nucleotides are in red. The associated changes are typical of what might be seen with repair of the DSB by **non homologous end joining (NHEJ) mechanisms**. It is also possible that the sites are simply re-ligated with no mutation occurring, or alternatively, these sites may cause recombination, as shown in **(b).**

Belancio *et al.: Genome Medicine* 2009, **1:97**



PLoS BIOLOGY

a peer-reviewed open-access journal published by the Public Library of Science

PRIMER ⓘ OPEN ACCESS

Gene Duplication: The Genomic Trade in Spare Parts

Matthew Hurles

Citation: Hurles M (2004) Gene Duplication: The

Download: PDF | Citation | XML

Print article

Order Reprints

Published in the July 2004 Issue of *PLoS Biology*

To **add a note,** highlight some text. Hide notes

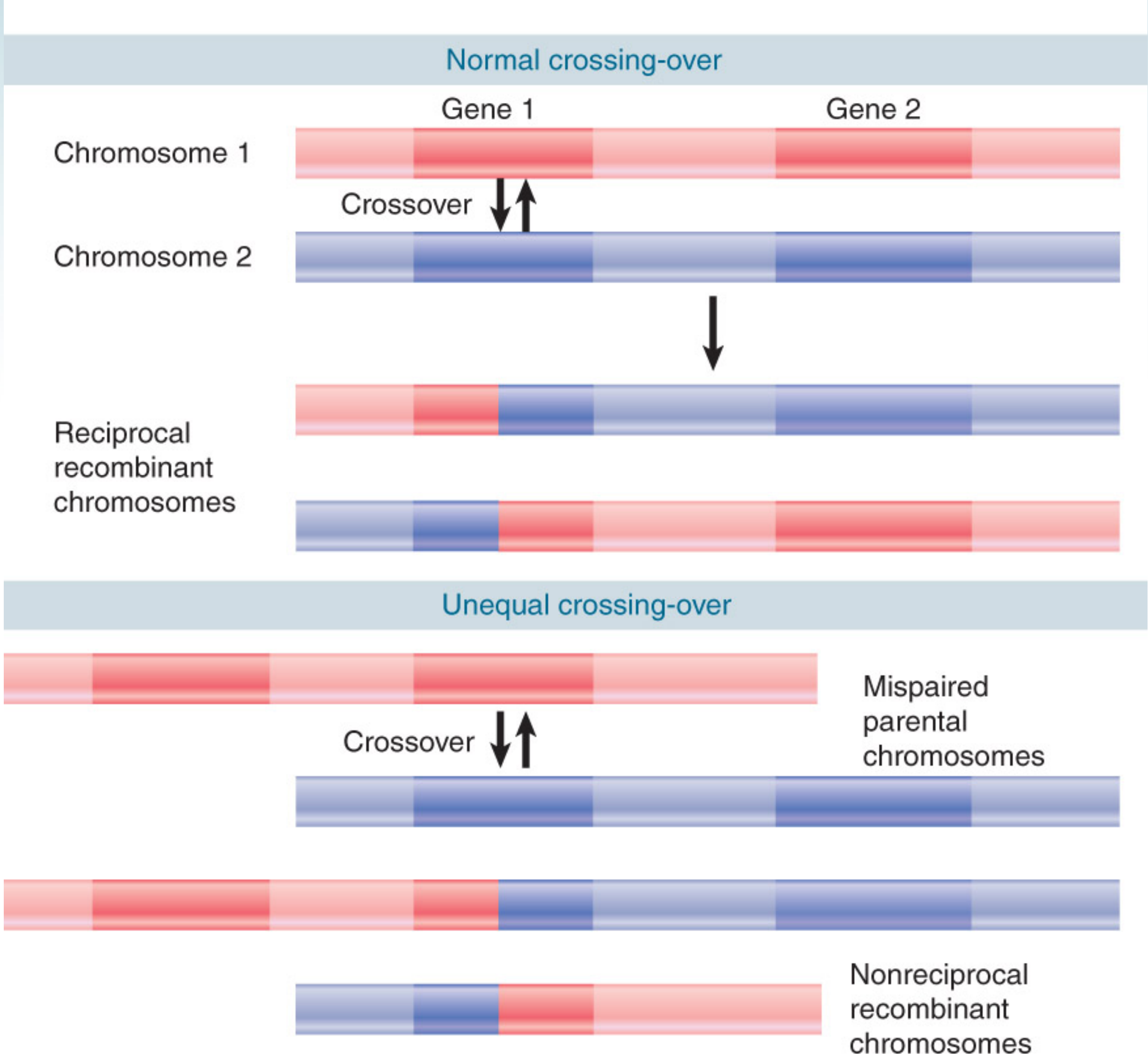
Make a general comment

Jump to

Metrics ⓘ

Average Rating (0 User Ratings)

☆ ☆ ☆ ☆ ☆ Rate This Article


Normal crossing-over
Gene 1
Gene 2
Chromosome 1
Crossover
Chromosome 2
Reciprocal recombinant chromosomes
Unequal crossing-over
Mispaired parental chromosomes
Crossover
Nonreciprocal recombinant chromosomes

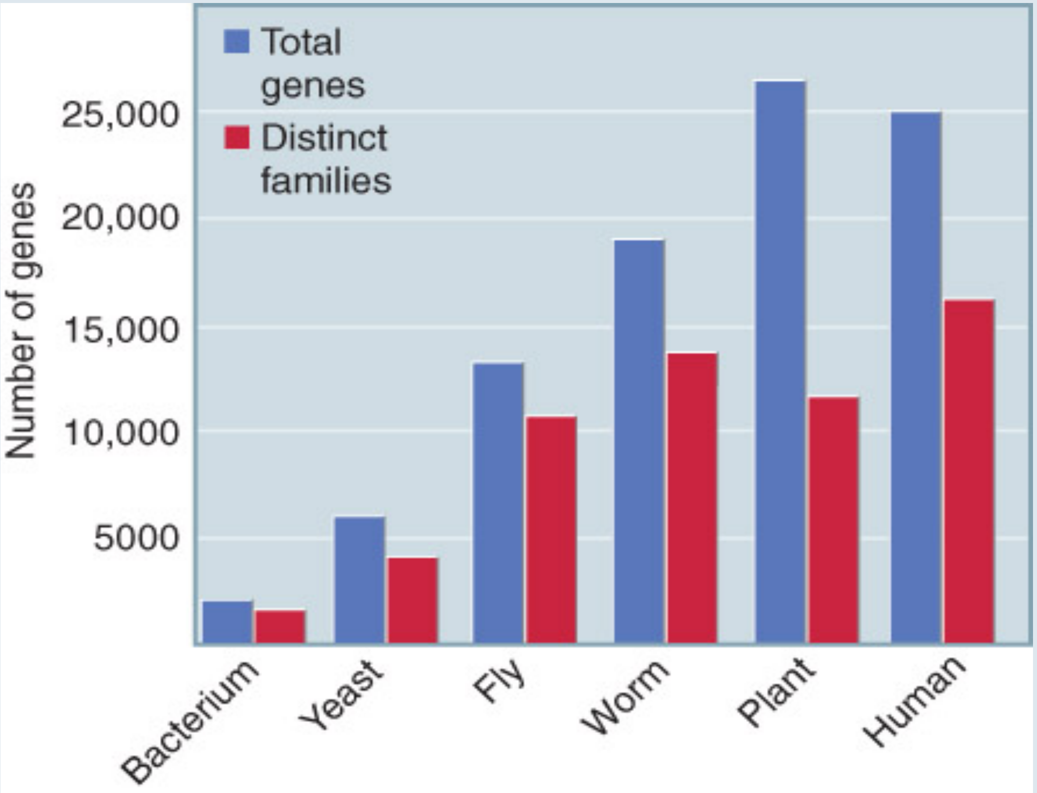
Gene NUMBERS and FUNCTION can be changed by unequal crossing-over

- **gene families –** sets of genes within a genome that code for related or identical proteins or RNAs.
  - The members were derived by duplication of an ancestral gene followed by accumulation of changes in sequence between the copies.
  - Most often the members are related but not identical.

- **gene clusters –** Groups of adjacent genes that are identical or related.

- **pseudogenes –** Inactive but stable components of the genome derived by mutation of an ancestral active gene.
  - Usually they are inactive because of mutations that block transcription or translation or both.

- The sum of the number of unique genes and the number of gene families is an estimate of the number of types of genes.

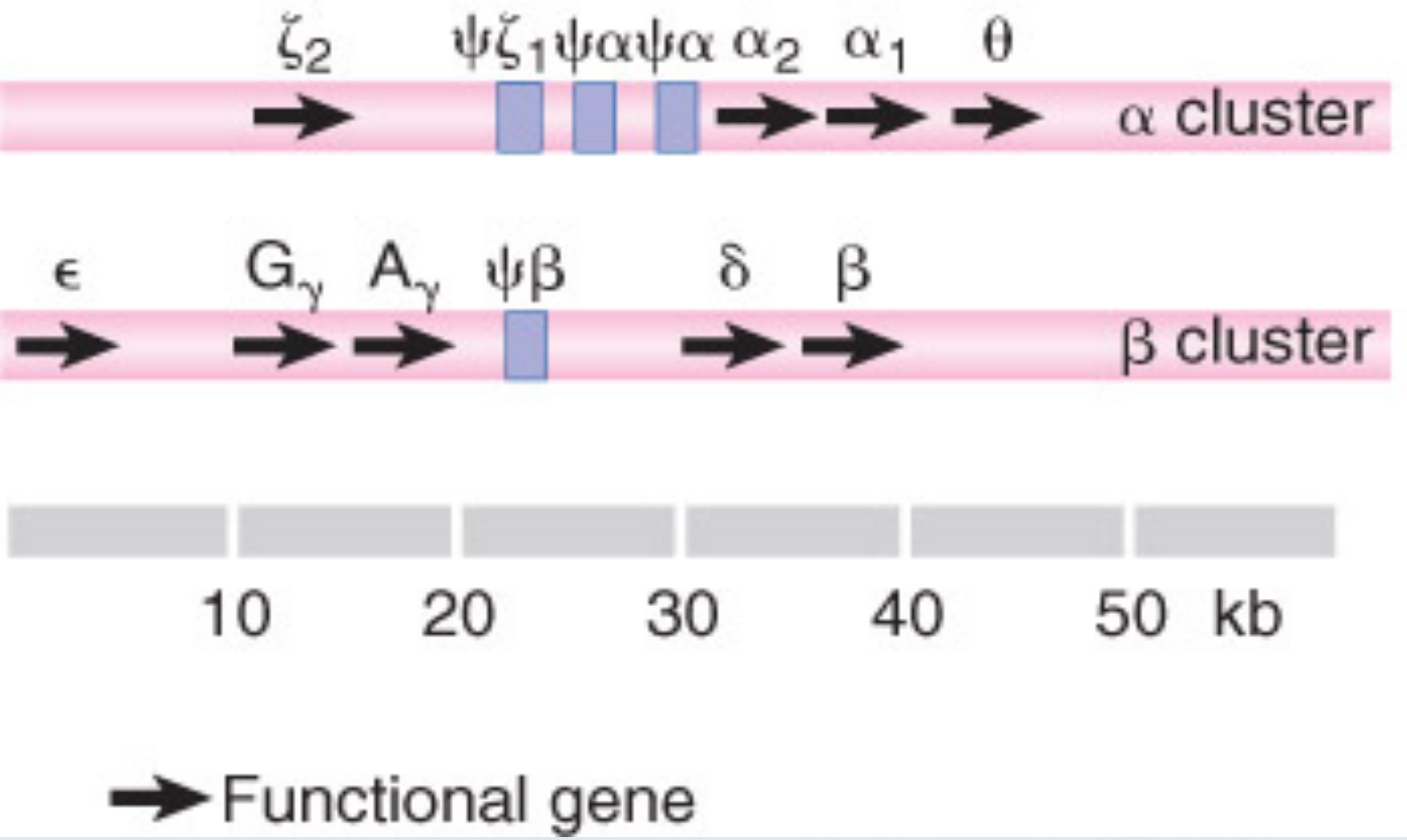


Many genes in a genome are duplicated and, as a result, form a number of different gene families - more so in higher eukaryotes

- **gene families –** sets of genes within a genome that code for related or identical proteins or RNAs.
  - The members were derived by duplication of an ancestral gene followed by accumulation of changes in sequence between the copies.
  - Most often the members are related but not identical.

- **gene clusters –** Groups of adjacent genes that are identical or related.

- **pseudogenes –** Inactive but stable components of the genome derived by mutation of an ancestral active gene.
  - Usually they are inactive because of mutations that block transcription or translation or both.

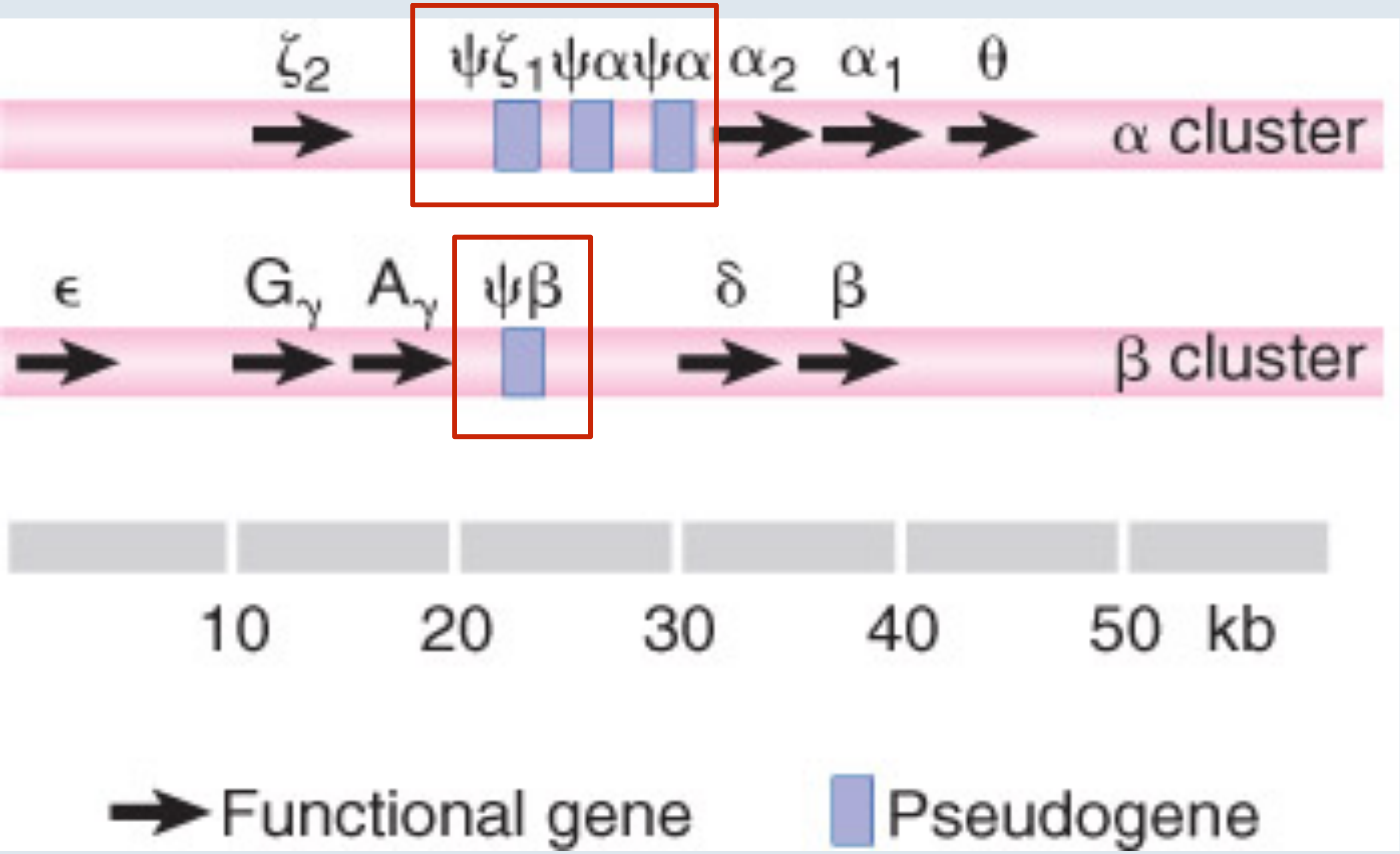


Each of the α-like and β-like globin gene families is organized into “clusters”, which includes functional genes and pseudogenes.

- **gene families –** sets of genes within a genome that code for related or identical proteins or RNAs.
  - The members were derived by duplication of an ancestral gene followed by accumulation of changes in sequence between the copies.
  - Most often the members are related but not identical.

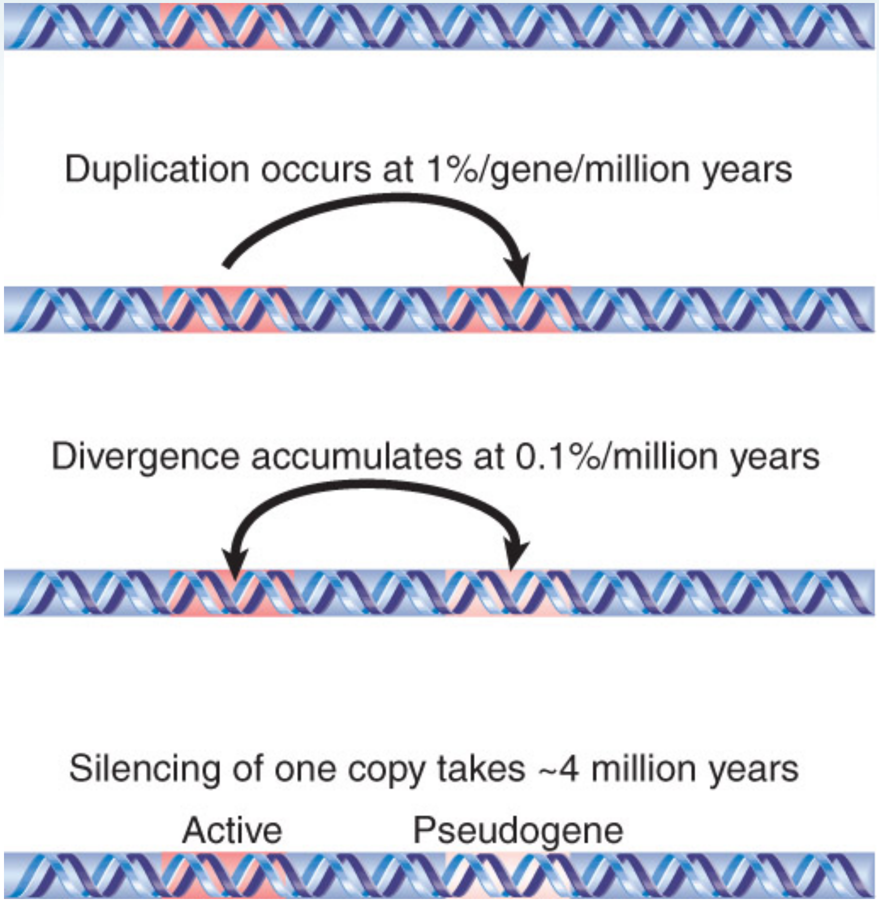
- **gene cluster –** A group of adjacent genes that are identical or related.

- **pseudogenes –** Inactive but stable components of the genome derived by mutation of an ancestral active gene.
  - Usually they are inactive because of mutations that block transcription / translation or both.



Each of the α-like and β-like globin gene families is organized into clusters, which includes functional genes and pseudogenes.

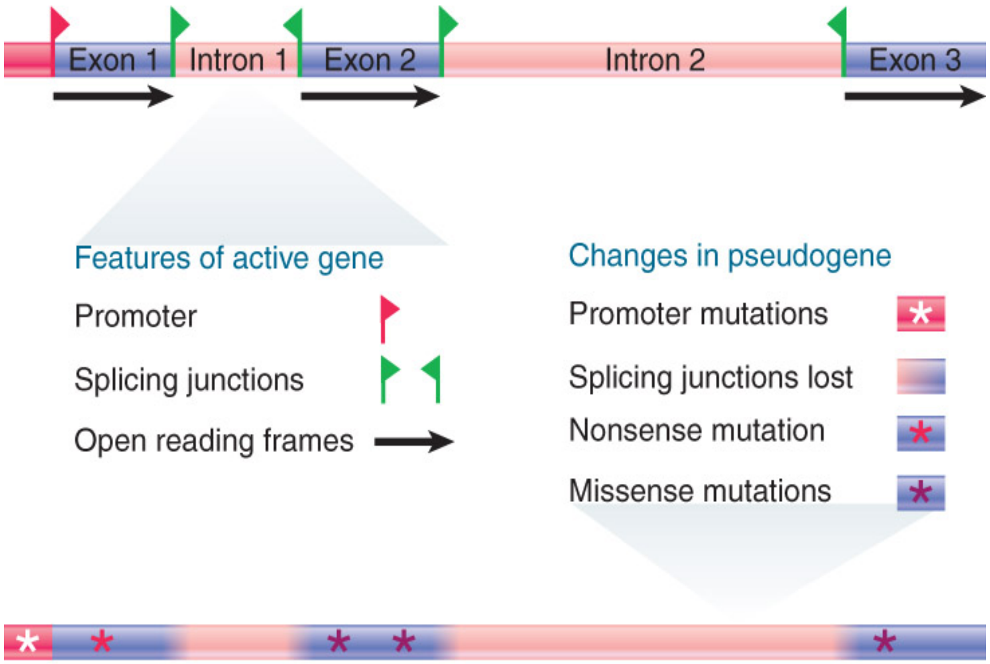
Gene Duplication is a Major Force in Genome Evolution


Duplication occurs at 1%/gene/million years
Divergence accumulates at 0.1%/million years
Silencing of one copy takes ~4 million years
Active
Pseudogene

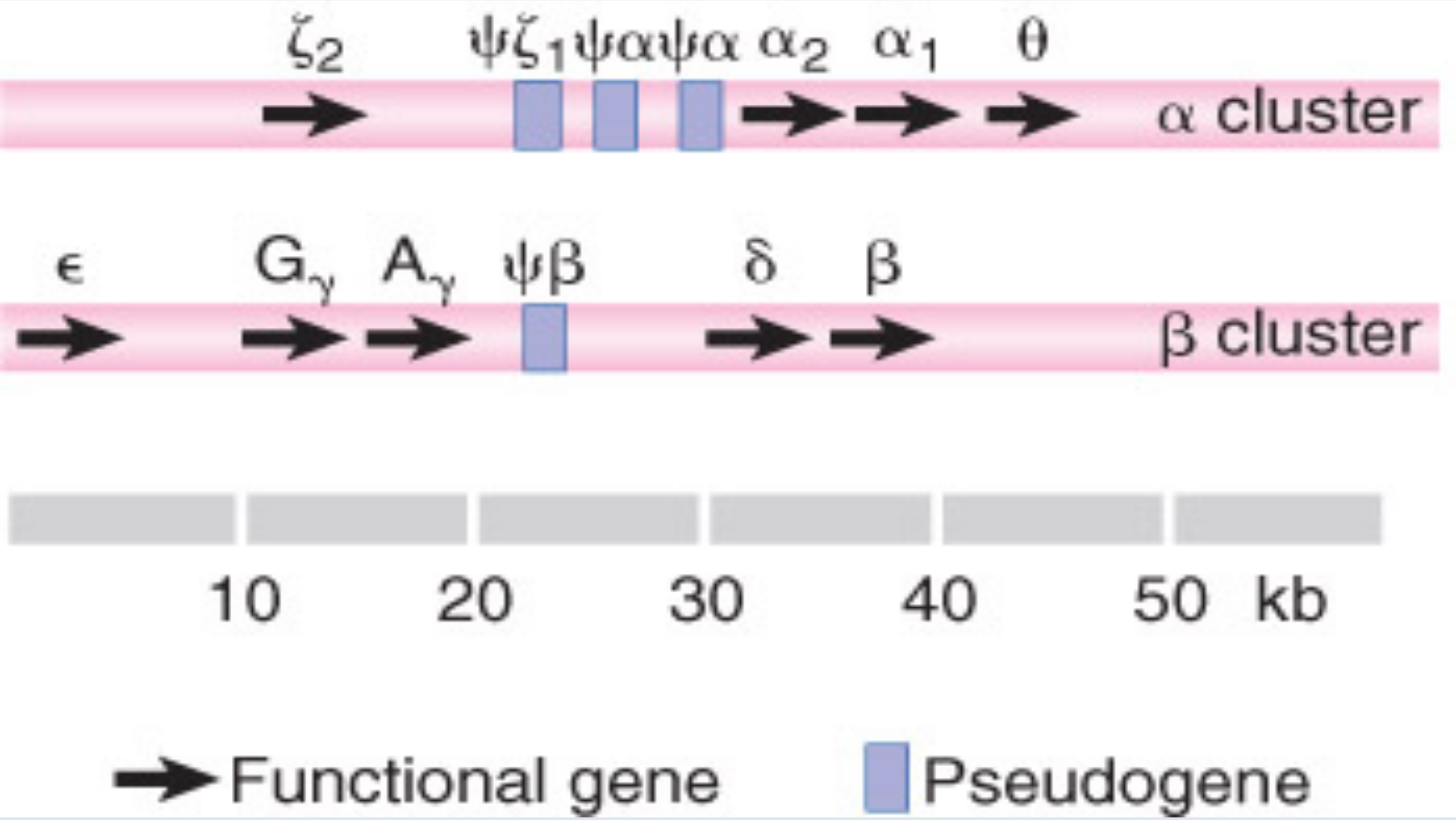
After a globin gene has been duplicated, differences may accumulate between the copies

Pseudogenes Are Nonfunctional Gene Copies

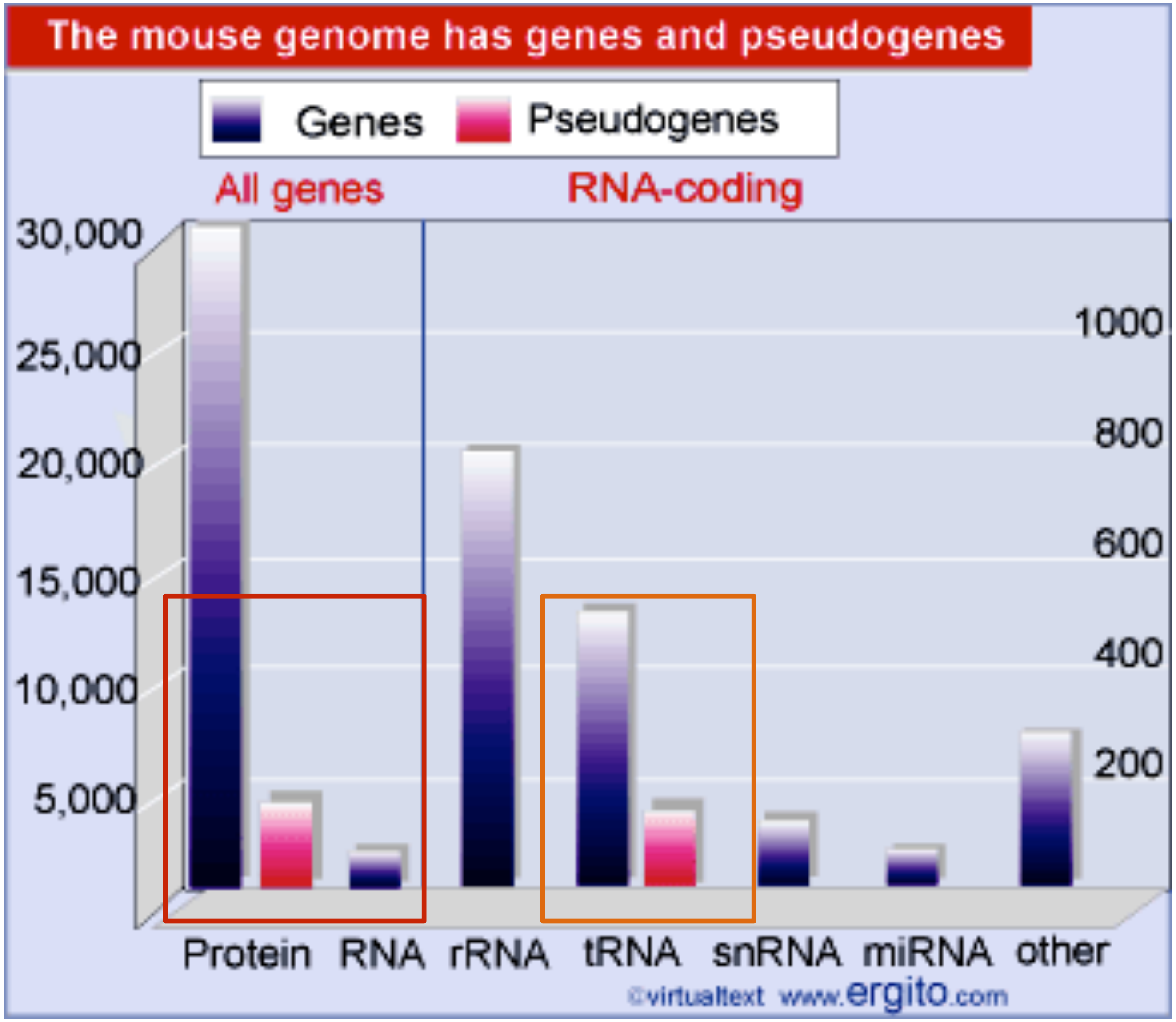
- **Processed pseudogenes** result from reverse transcription and integration of mRNA transcripts.
- **Nonprocessed pseudogenes** result from incomplete duplication or second-copy mutation of functional genes.
- Some pseudogenes may gain functions different from those of their parent genes, such as regulation of gene expression, and take on different names.

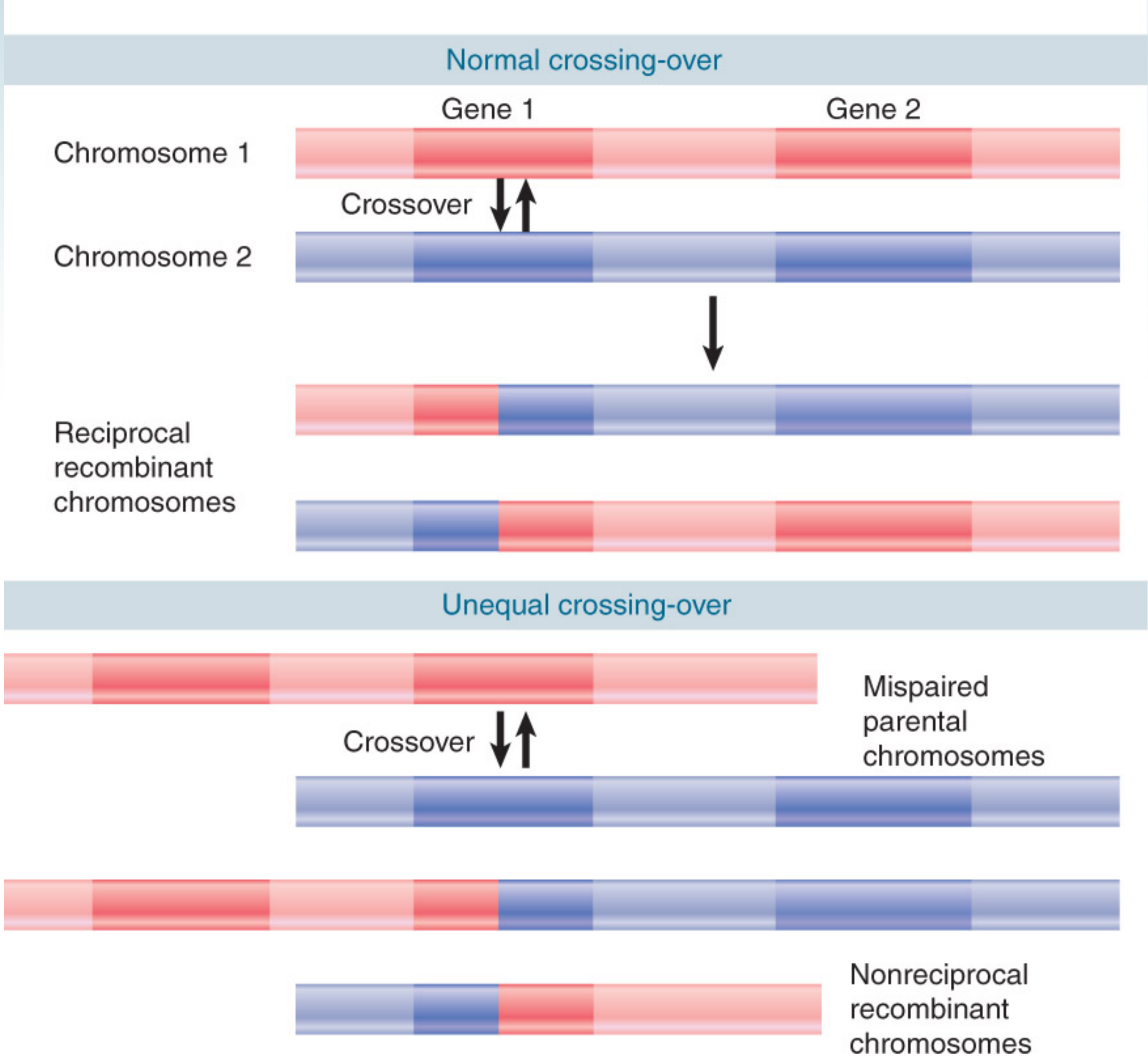


Many changes have occurred in a beta-globin gene since it became a pseudogene



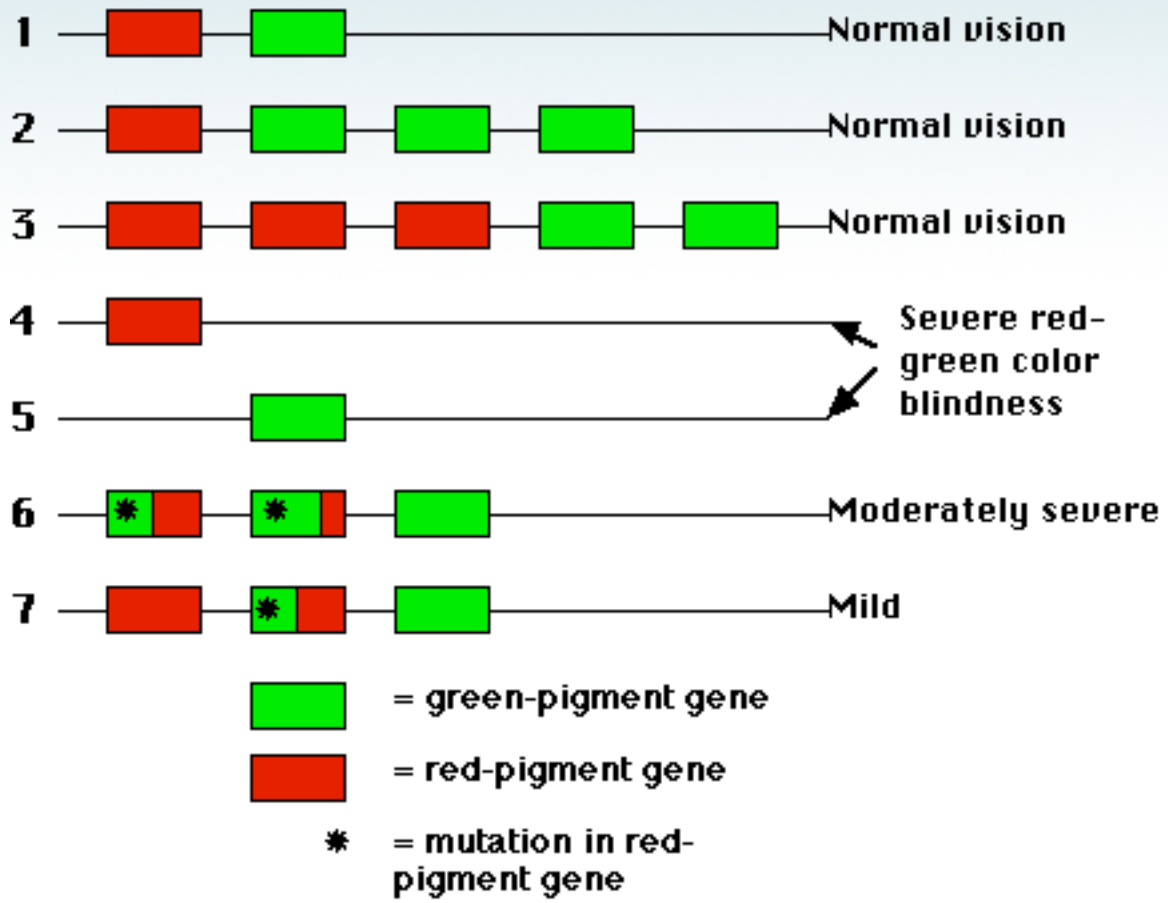
Each of the α-like and β-like globin gene families is organized into a single cluster, which includes functional genes and pseudogenes.


The mouse genome has genes and pseudogenes
Genes
Pseudogenes
All genes
RNA-coding
30,000
25,000
20,000
15,000
10,000
5,000
1000
800
600
400
200
Protein
RNA
rRNA
tRNA
snRNA
miRNA
other
©virtualtext www.ergito.com


Normal crossing-over
Gene 1
Gene 2
Chromosome 1
Crossover
Chromosome 2
Reciprocal recombinant chromosomes
Unequal crossing-over
Mispaired parental chromosomes
Crossover
Nonreciprocal recombinant chromosomes

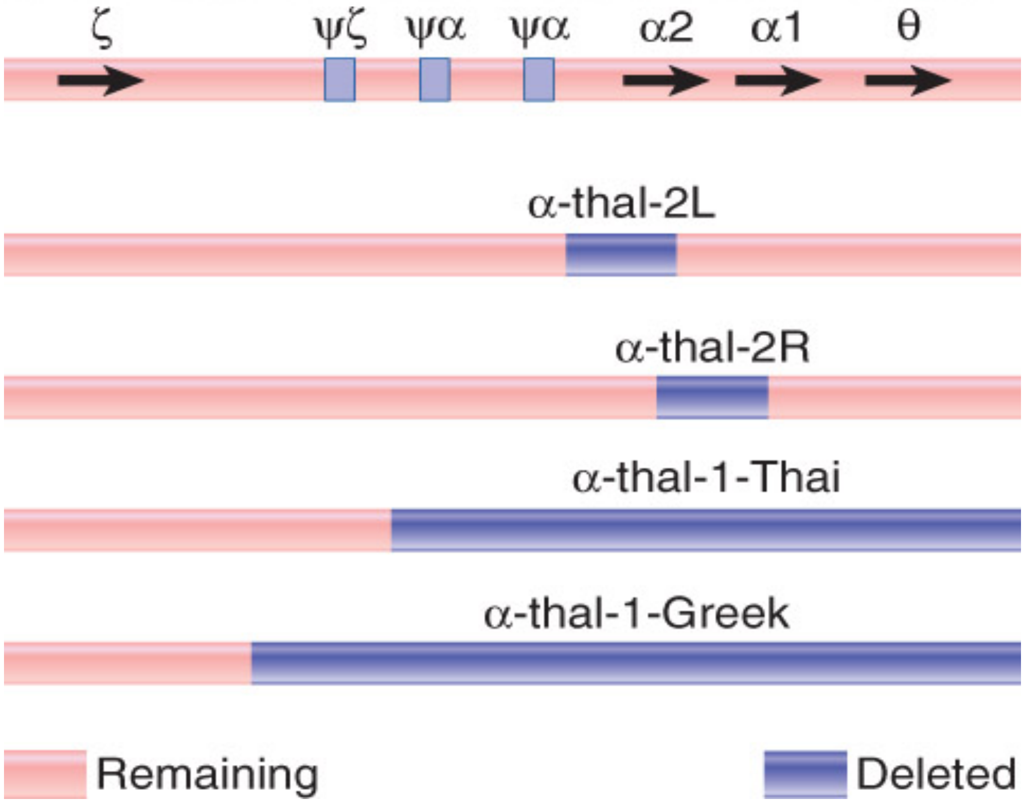
Gene NUMBERS and FUNCTION can be changed by unequal crossing-over

Representative X chromosomes
(each male has only one)



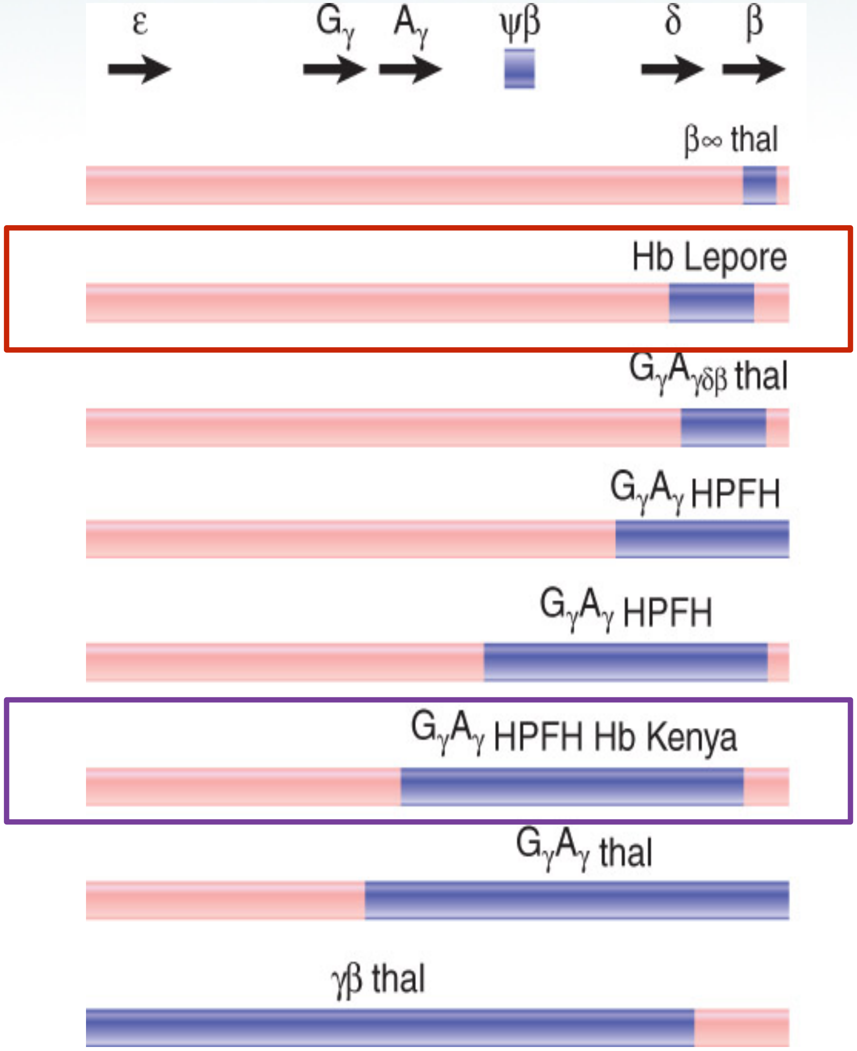
Some males have as many as 9 copies of genes encoding the red and green opsin genes, when two are enough. The sequences of the red and green genes are effectively very similar at 98% of their nucleotides. This high degree of similarity creates the risk of mismatches in synapsis during meiosis with **unequal crossing over.**

- Different **thalassaemias** are caused by various deletions that eliminate α- or β-globin genes.
  - The severity of the disease depends on the individual deletion.



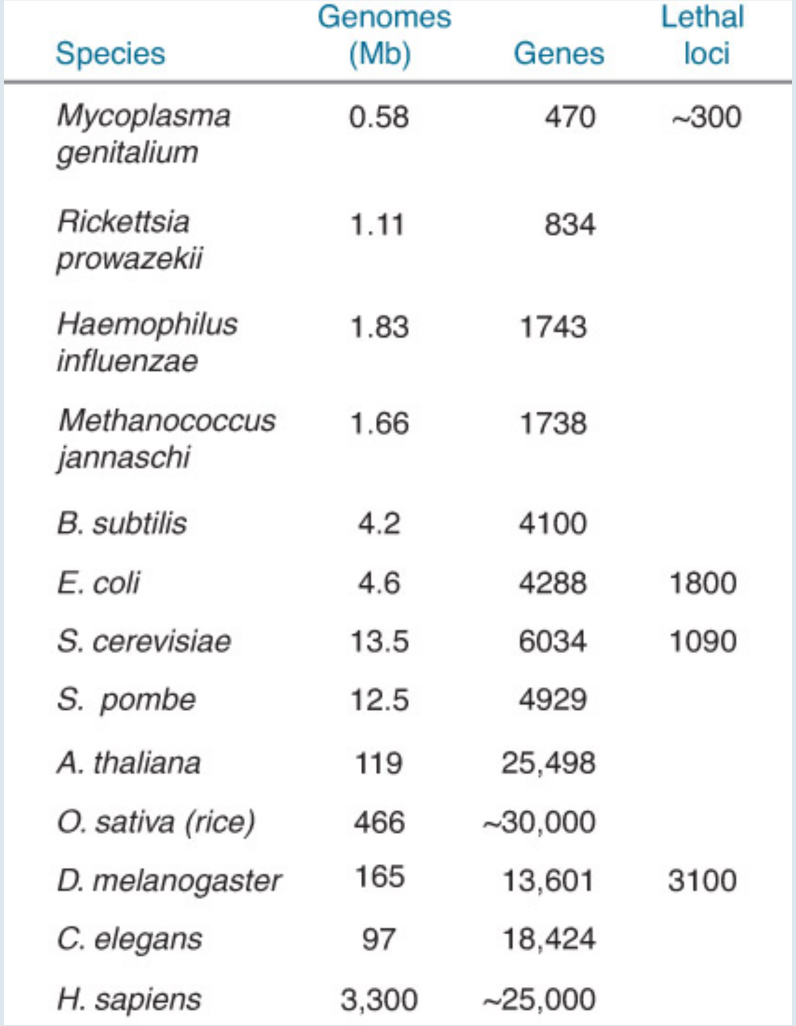
α-thalassaemias result from various deletions in the α- globin gene cluster

Unequal Crossing-over Rearranges Gene Clusters



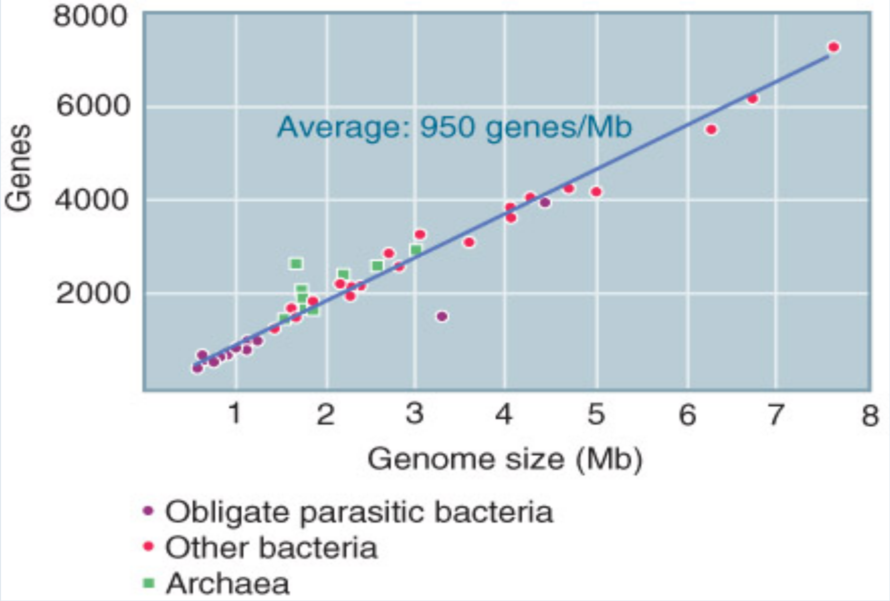
- **Hb Lepore –** An unusual globin protein that results from unequal crossing-over between the δ and β genes.

- Hb Kenya – A fusion gene produced by unequal crossing-over between the Aγ- and β-globin genes

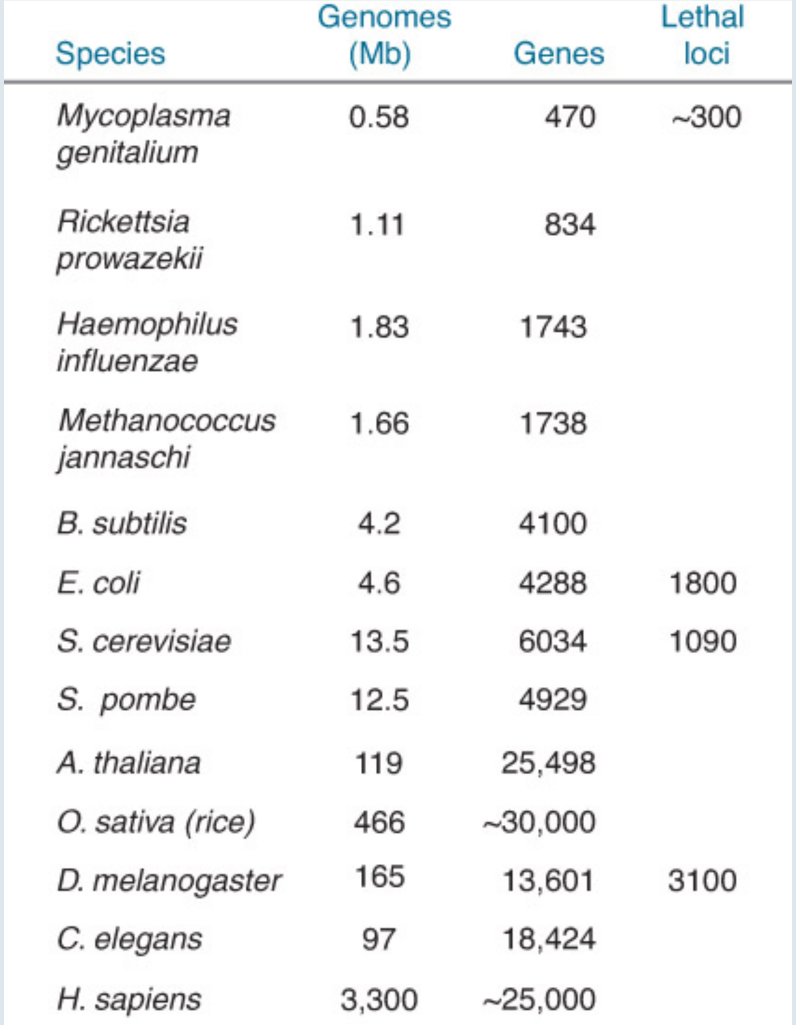


Species	Genomes (Mb)	Genes	Lethal loci
Mycoplasma genitalium	0.58	470	~300
Rickettsia prowazekii	1.11	834	
Haemophilus influenzae	1.83	1743	
Methanococcus jannaschi	1.66	1738	
B. subtilis	4.2	4100	
E. coli	4.6	4288	1800
S. cerevisiae	13.5	6034	1090
S. pombe	12.5	4929	
A. thaliana	119	25,498	
O. sativa (rice)	466	~30,000	
D. melanogaster	165	13,601	3100
C. elegans	97	18,424	
H. sapiens	3,300	~25,000	

Genome sizes and gene numbers are known from complete sequences for several organisms.

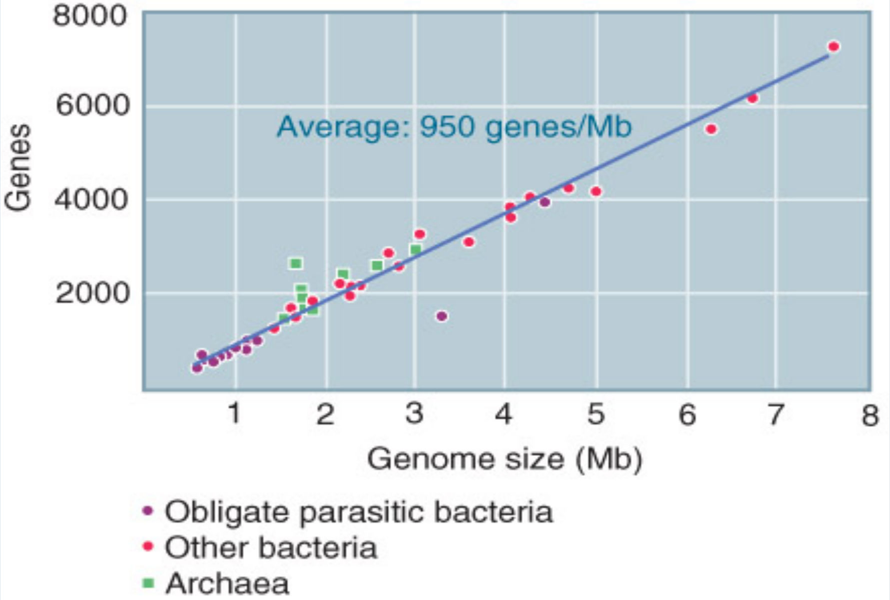


- The minimum number of genes for a parasitic prokaryote is about 500; for a free-living non-parasitic prokaryote it is about 1500.

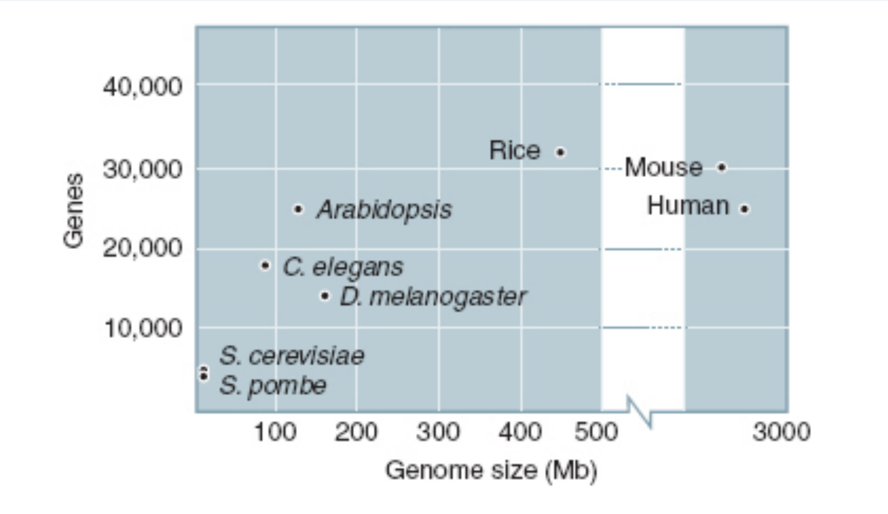


Species	Genomes (Mb)	Genes	Lethal loci
Mycoplasma genitalium	0.58	470	~300
Rickettsia prowazekii	1.11	834	
Haemophilus influenzae	1.83	1743	
Methanococcus jannaschi	1.66	1738	
B. subtilis	4.2	4100	
E. coli	4.6	4288	1800
S. cerevisiae	13.5	6034	1090
S. pombe	12.5	4929	
A. thaliana	119	25,498	
O. sativa (rice)	466	~30,000	
D. melanogaster	165	13,601	3100
C. elegans	97	18,424	
H. sapiens	3,300	~25,000	

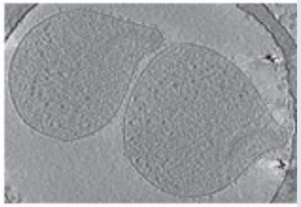
Genome sizes and gene numbers are known from complete sequences for several organisms.



- The minimum number of genes for a parasitic prokaryote is about 500; for a free-living non-parasitic prokaryote it is about 1500.



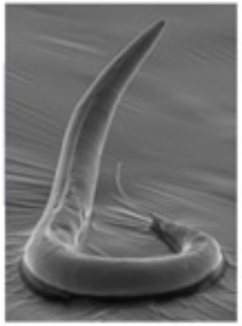
500 genes
Intracellular (parasitic) bacterium



1500 genes
Free-living bacterium

5000 genes
Unicellular eukaryote

13,000 genes
Multicellular eukaryote



25,000 genes
Higher plants

25,000 genes
Mammals

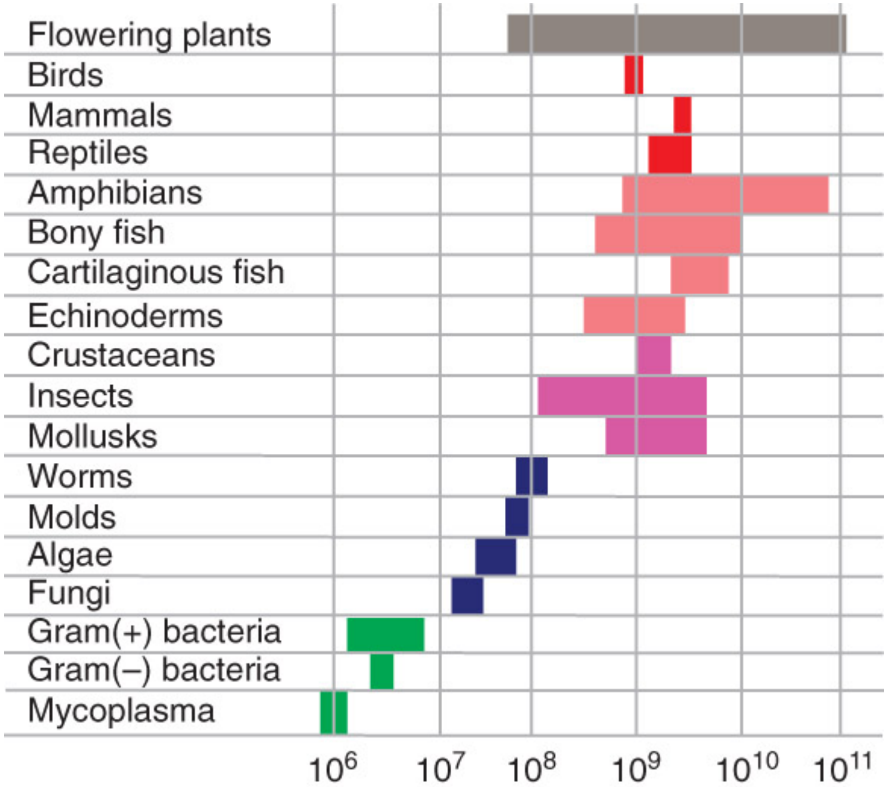
The “minimum” gene number required for any type of organism increases with its complexity……..

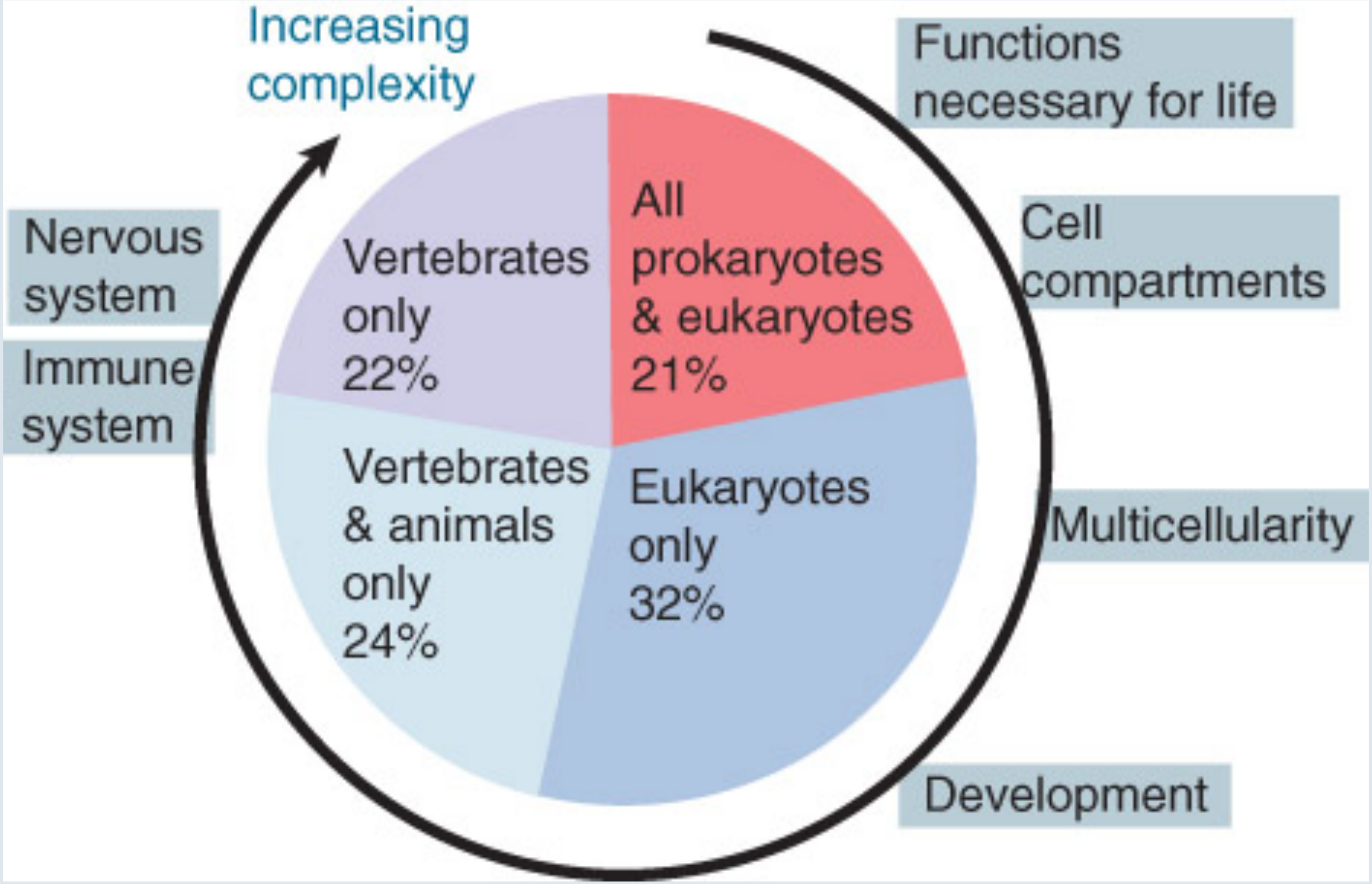
Courtesy of Eishi Noguchi, Drexel University College of Medicine.

- There is no definitive correlation between genome size and genetic complexity.

- **C-value –** The total amount of DNA in the genome (per haploid set of chromosomes)

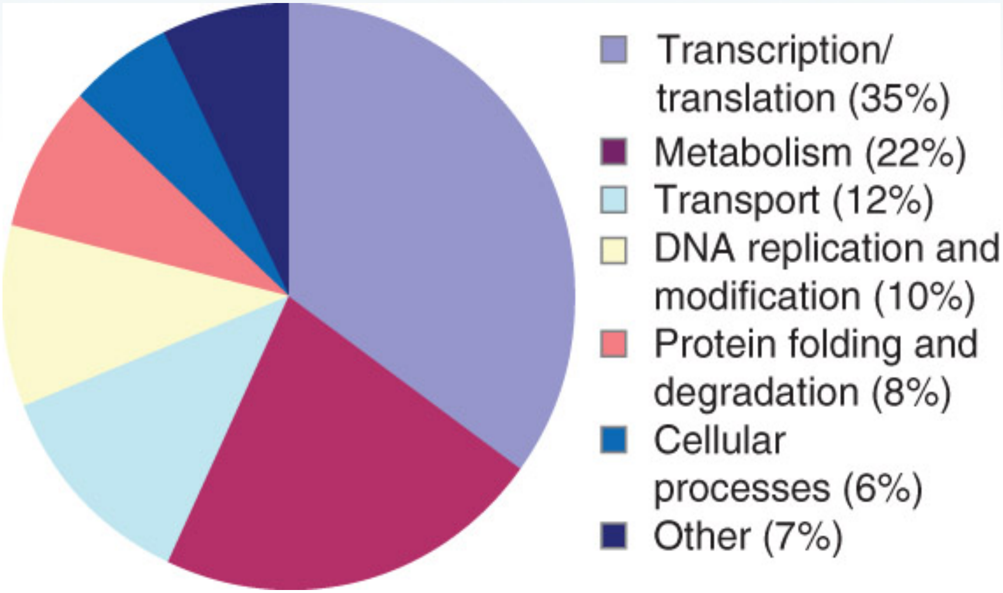
- **C-value paradox –** The lack of relationship between the DNA content (C-value) of an organism and its coding potential.

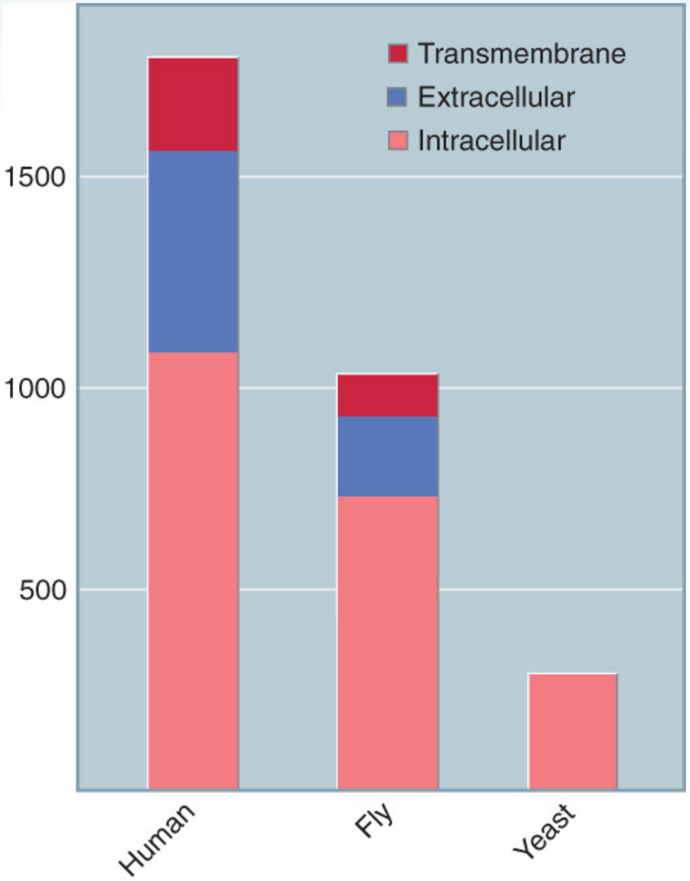




Human genes can be classified according to how widely their homologues are distributed in other species.

Morphological Complexity Evolves by Adding New Gene Functions



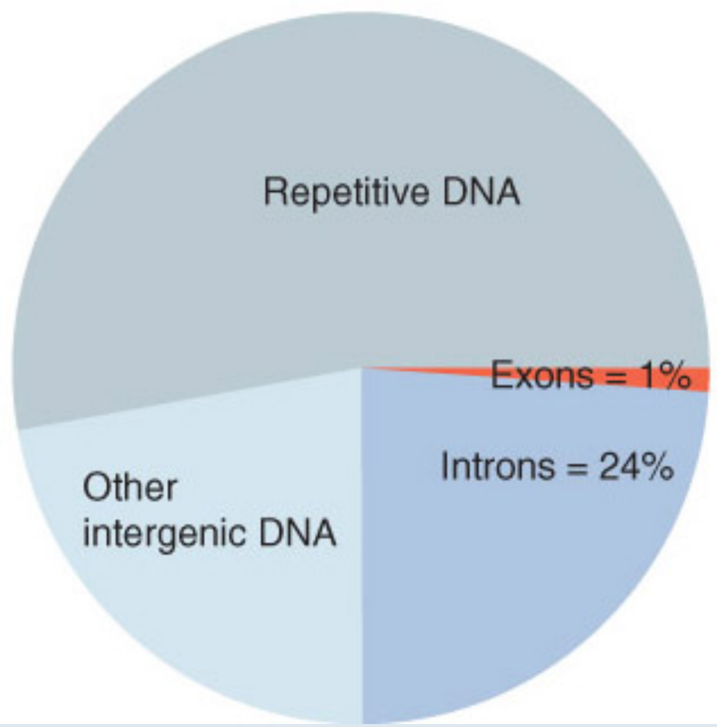


Common eukaryotic proteins are concerned with essential cellular functions

Increasing complexity in eukaryotes is accompanied by accumulation of new proteins for **transmembrane** and **extracellular** functions

The Human Genome Has Fewer Genes Than Originally Expected

- Only 1% of the human genome consists of exons.
- Exons comprise ~5% of each gene, so genes (exons plus introns) comprise ~25% of the genome.
- The human genome has between 20,000 to 25,000 genes.



Genes occupy 25% of the human genome, but protein-coding sequences are only a small part of this fraction, ~1%.

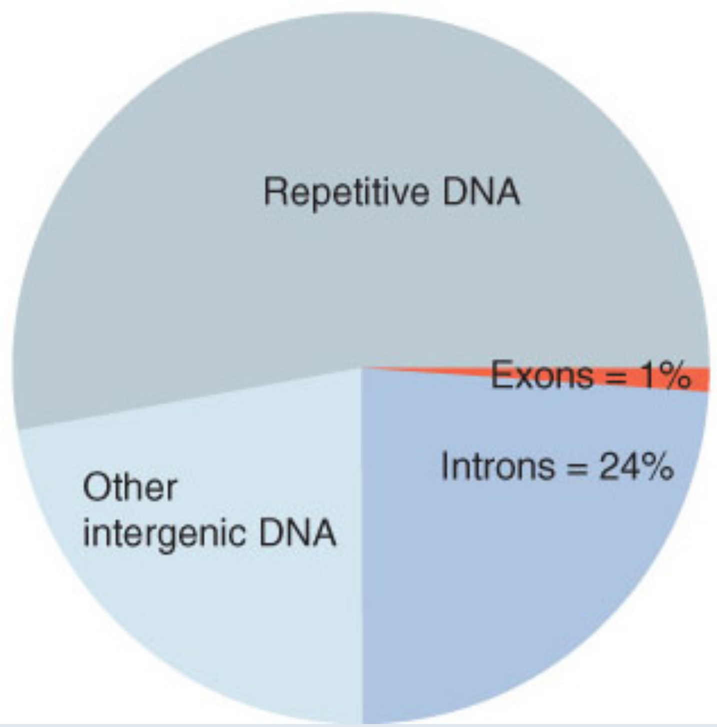
- Repeated sequences (present in more than one copy) account for >50% of the human genome.

- The great bulk of repeated sequences consists of copies of nonfunctional transposons.

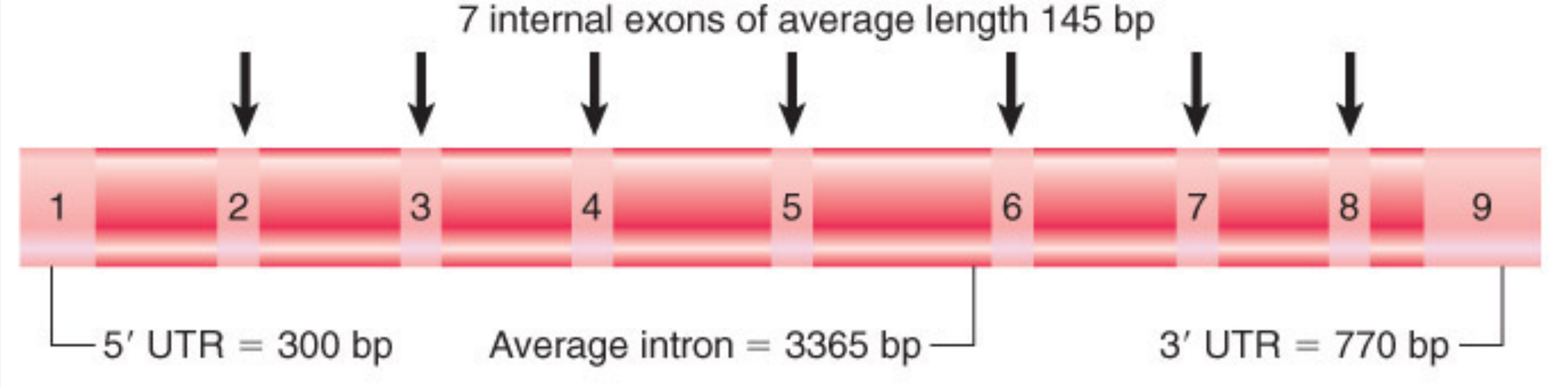
- There are many duplications of large chromosome regions.

- There are many duplications of large chromosome regions.

Indeed, the largest component of the human genome consists of transposons.



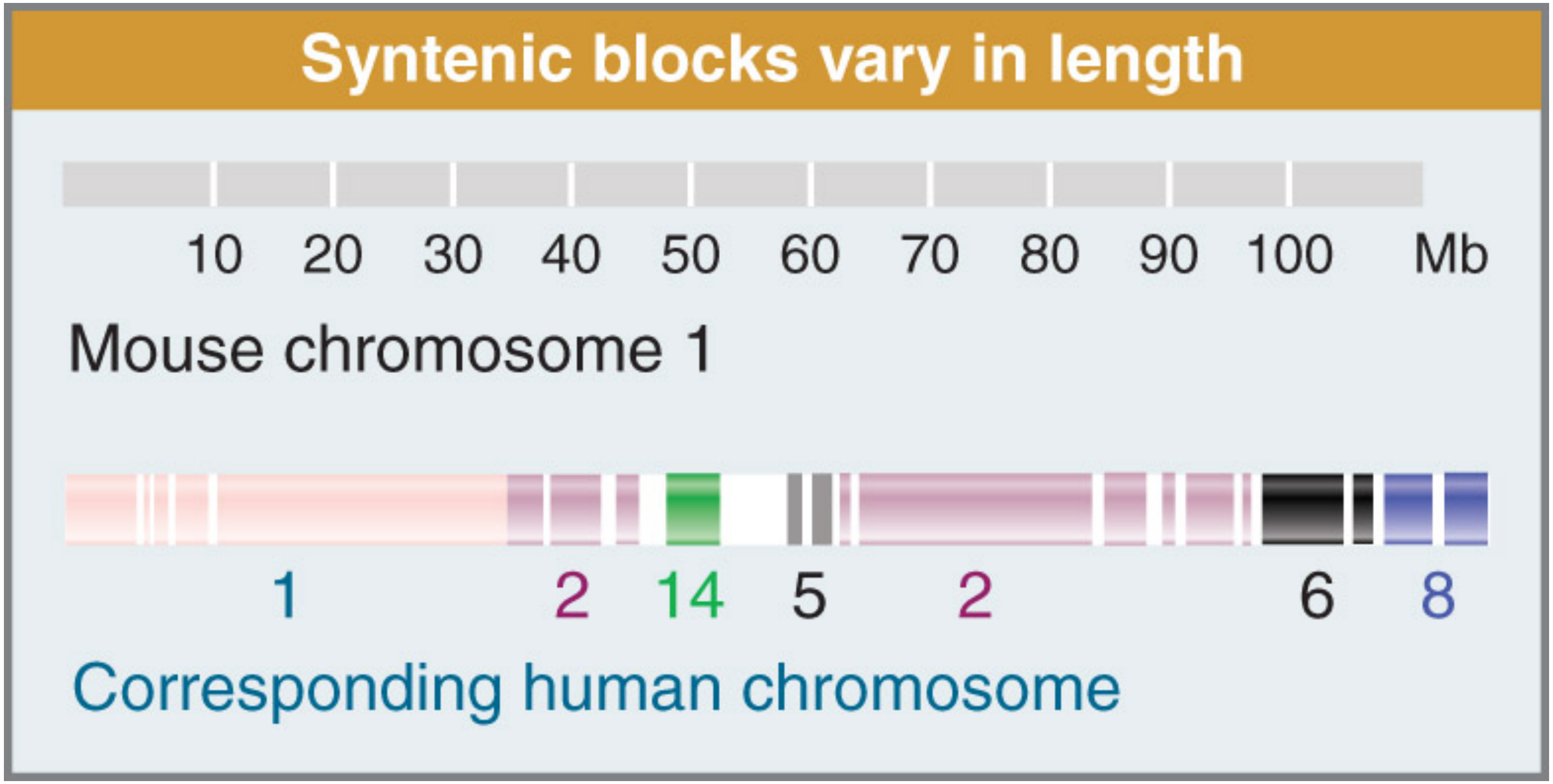
The Human Genome Has Fewer Genes Than Originally Expected

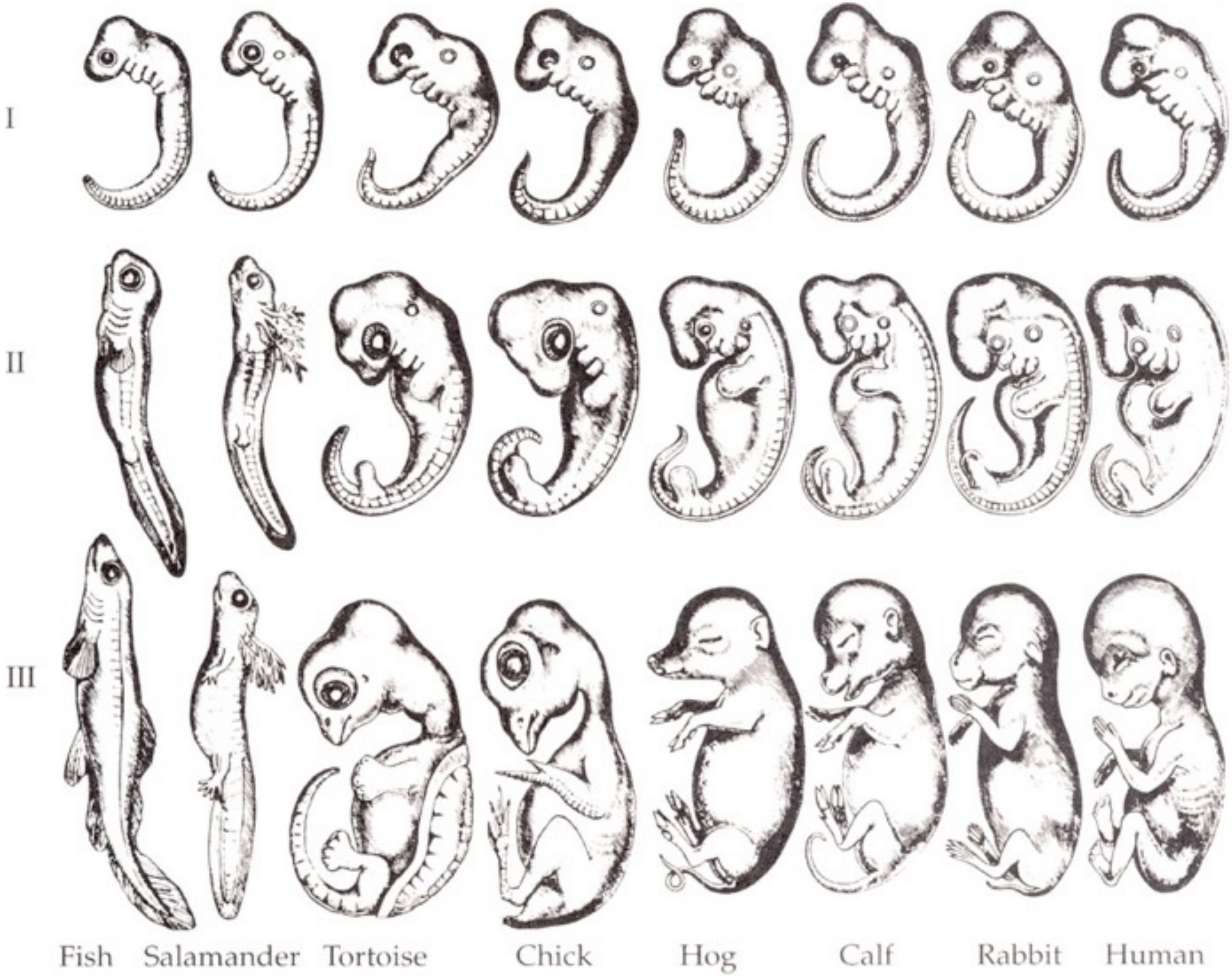


The average human gene is ~27 kb long and has 9 exons, usually comprising two longer exons at each end and seven smaller, internal exons.

- **~60%** of human genes are **alternatively** spliced.
- Up to **80%** of the alternative splices change protein sequence, so the **proteome** has upward of **50,000 to 60,000 members**.

- **Syntenic** relationships can be extensive, as seen between mouse and human genomes, where most of the similar active genes are in a syntenic region.


Syntenic blocks vary in length
10
20
30
40
50
60
70
80
90
100
Mb
Mouse chromosome 1
1
2
14
5
2
6
8
Corresponding human chromosome

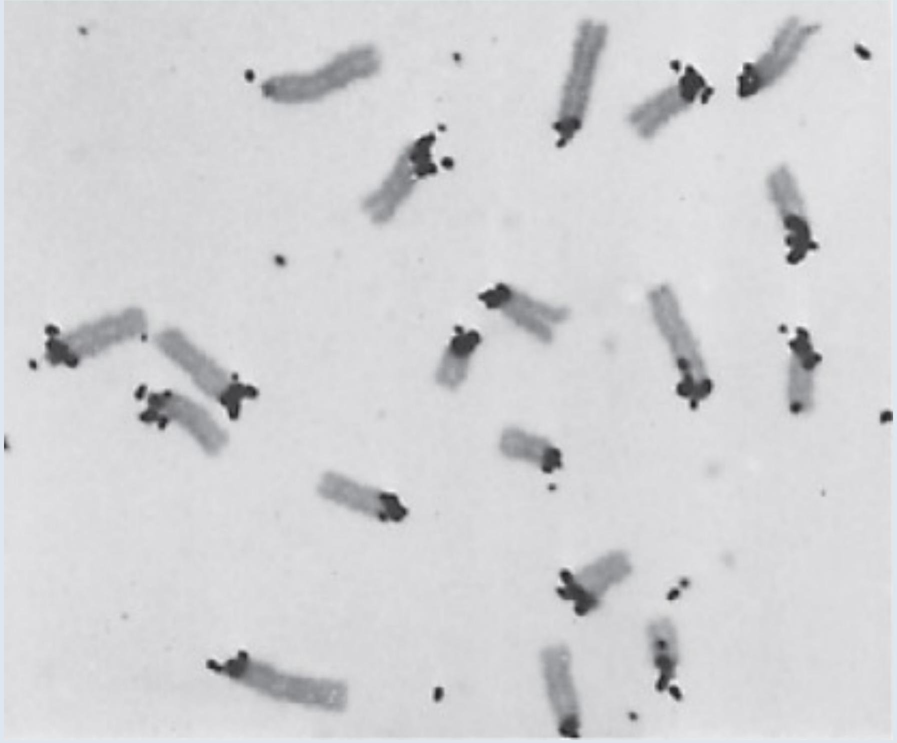


“Ontogeny recapitulates phylogeny ??” Haeckel, 1870’s

- **satellite DNA –** DNA that consists of many tandem repeats (identical or related) of a short basic repeating unit

- **minisatellite –** DNAs consisting of tandemly repeated copies of a short repeating sequences, with more repeat copies than a **micro**satellite but fewer than a **satellite**.
  - The length of the repeating unit is measured in tens of base pairs.
  - The number of repeats varies between individual genomes.

- Satellite DNA is often the major constituent of centromeric **heterochromatin**.
- As opposed to **euchromatin** – Regions that comprise most of the genome in the interphase nucleus which are less tightly coiled than heterochromatin, and contain most of the active or potentially active single-copy genes.

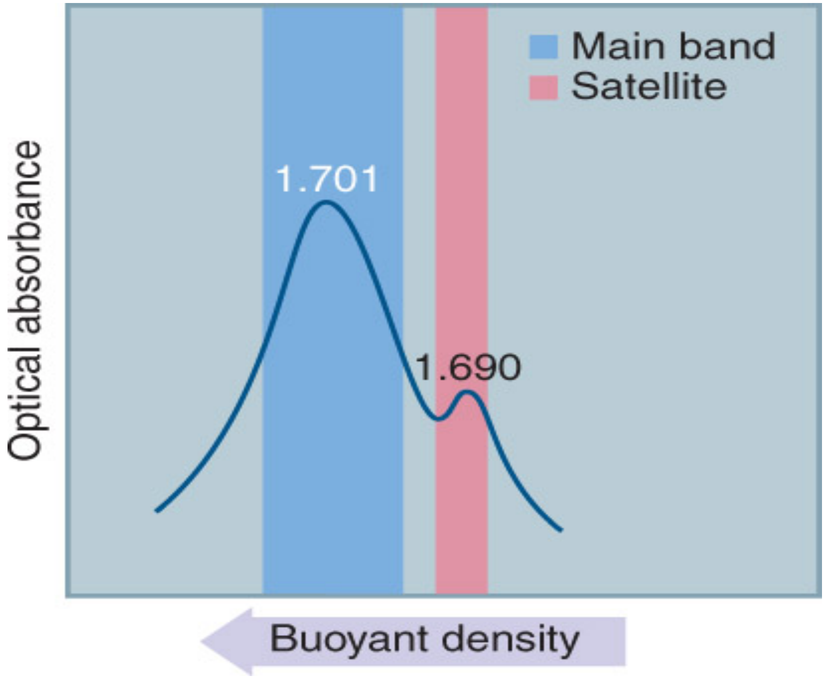


Cytological hybridization shows that mouse satellite DNA is located at the centromeres.

Photo courtesy of Mary Lou Pardue and Joseph G. Gall, Carnegie Institution.

- Highly repetitive DNA (or satellite DNA) has a very short repeating sequence and no coding function….

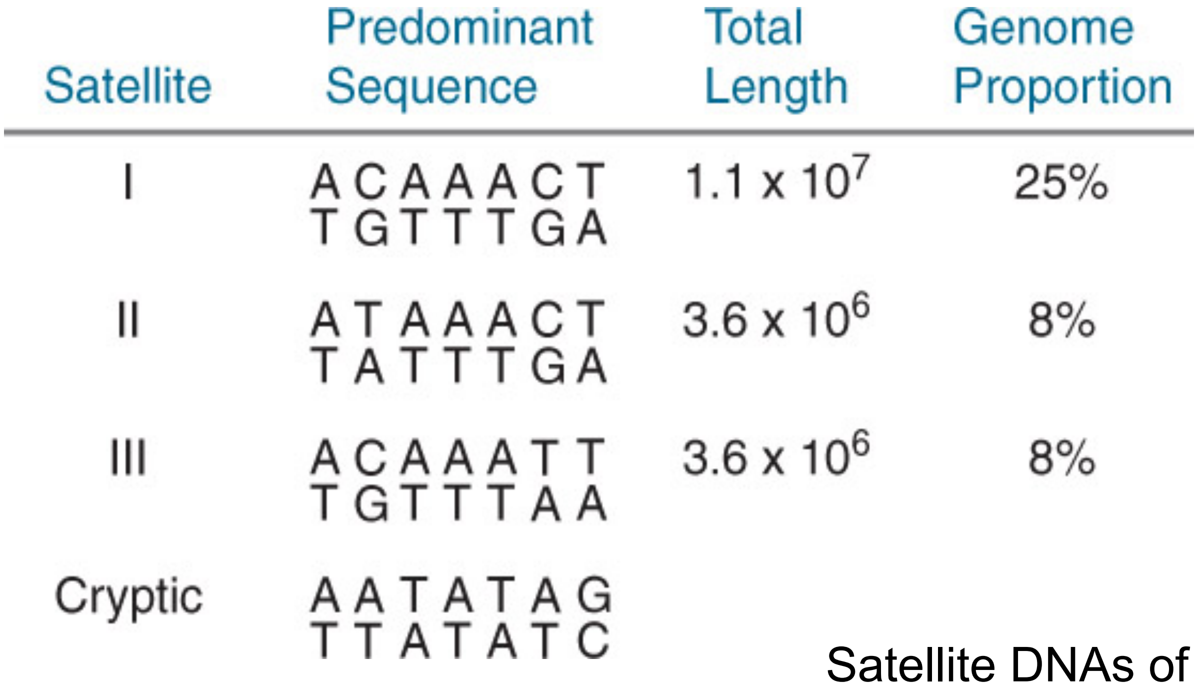
- **simple sequence DNA –** Short repeating units of DNA sequence.
- **Satellite DNA** occurs in large blocks that can have distinct physical properties.



Mouse DNA is separated into a main band and a satellite by centrifugation through a density gradient of CsCl

Arthropod Satellite DNA Have Very Short Identical Repeats

- The repeating units of arthropod satellite DNAs are only a few nucleotides long.
  - Most of the copies of the sequence are identical.



Satellite	Predominant Sequence	Total Length	Genome Proportion
I	ACAAACT TGTTTGA	1.1×10^7	25%
II	ATAAACT TATTTGA	3.6×10^6	8%
III	ACAAATT TGTTTAA	3.6×10^6	8%
Cryptic	AATATAG TTATATC		

Satellite DNAs of *D. virilis* are related

Mammalian Satellites Consist of Hierarchical Repeats

- Mouse satellite DNA appear to have evolved through duplication and mutation of a short repeating unit to give a basic repeating unit of 234 bp in which the original half-, quarter-, and eighth-repeats can be recognized.

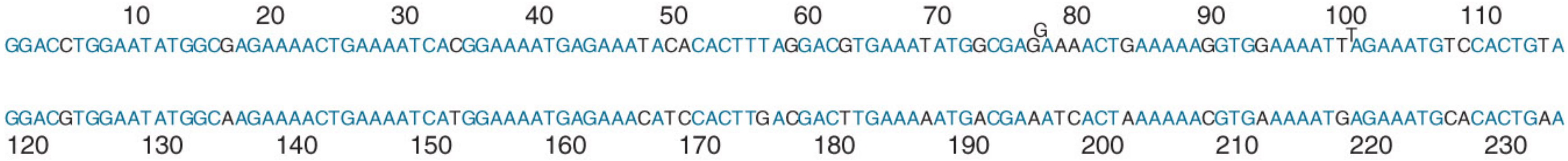
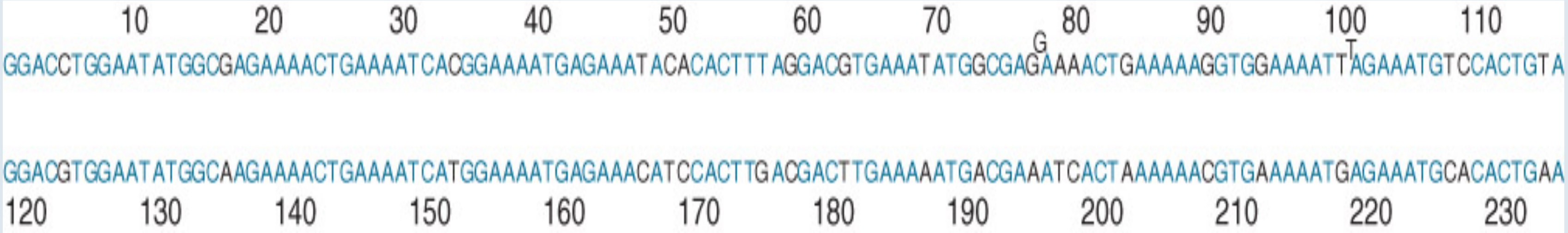
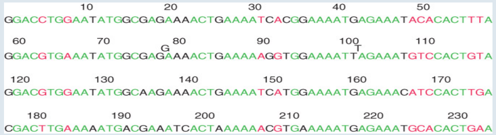
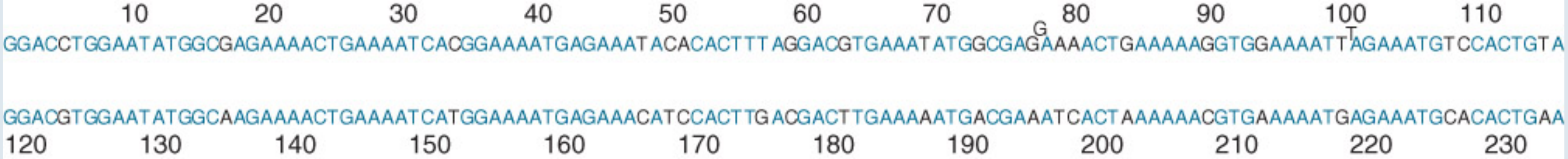

10 20 30 40 50 60 70 80 90 100 110
G T
GGACCTGGAATATGGCGAGAAAACTGAAAATCACGGAAAATGAGAAATACACACTTTAGGACGTGAAATATGGCGAGAAAACTGAAAAAGGTGGAAAATTAGAAATGTCCACTGTA
GGACGTGGAATATGGCAAGAAAACTGAAAATCATGGAAAATGAGAAACATCCACTTGACGACTTGAAAAATGACGAAATCACTAAAAAACGTGAAAAATGAGAAATGCACACTGAA
120 130 140 150 160 170 180 190 200 210 220 230

FIGURE 15: The repeating unit of mouse satellite DNA contains two half-repeats, which are aligned to show the identities (in blue)


10 20 30 40 50 60 70 80 90 100 110
G
T
GGACCTGGAATATGGCGAGAAAACTGAAAATCACGGAAAATGAGAAATACACACTTTAGGACGTGAAATATGGCGAGAAAACTGAAAAAGGTGGAAAATTAGAAATGTCCACTGTA
GGACGTGGAATATGGCAAGAAAACTGAAAATCATGGAAAATGAGAAACATCCACTTGACGACTTGAAAAATGACGAAATCACTAAAAAACGTGAAAAATGAGAAATGCACACTGAA
120 130 140 150 160 170 180 190 200 210 220 230

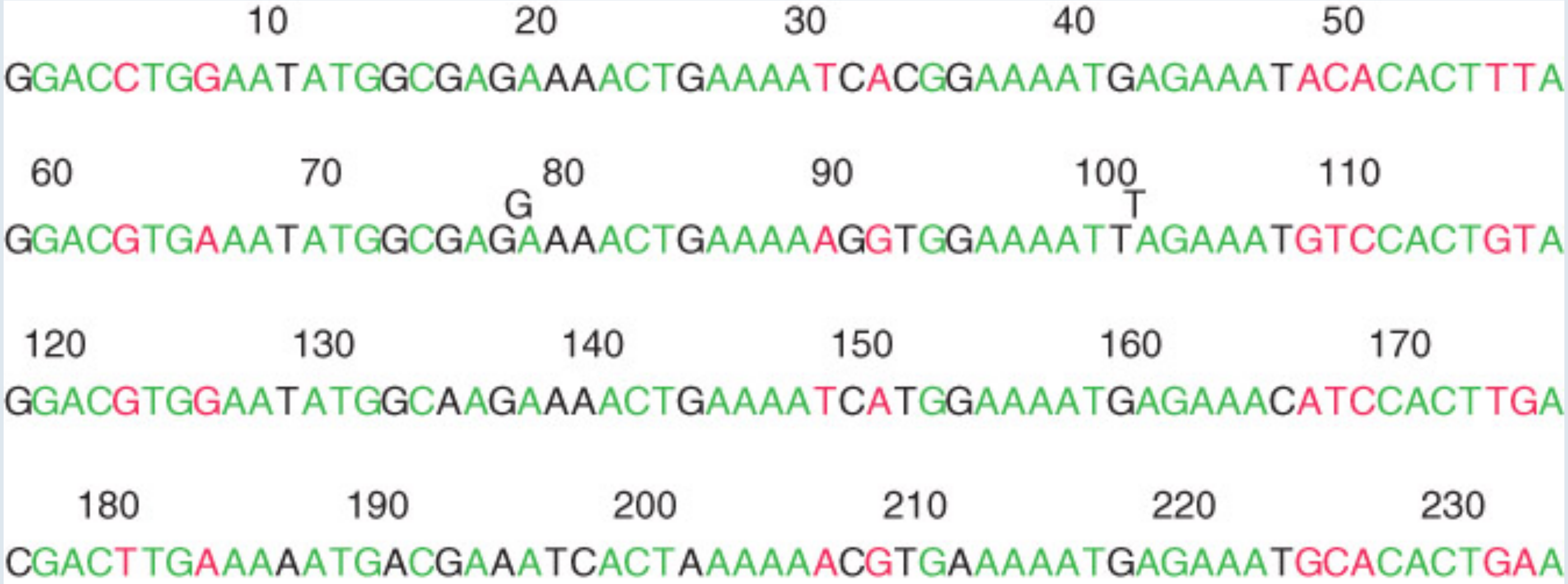

10 20 30 40 50
GGACCTGGAATATGGCGAGAAAACTGAAAATCACGGAAAATGAGAAATACACACTTTA
60 70 80 90 100 110
G
T
GGACGTGAAATATGGCGAGAAAACTGAAAAAGGTGGAAAATTAGAAATGTCCACTGTA
120 130 140 150 160 170
GGACGTGGAATATGGCAAGAAAACTGAAAATCATGGAAAATGAGAAACATCCACTTGA
180 190 200 210 220 230
CGACTTGAAAAATGACGAAATCACTAAAAAACGTGAAAAATGAGAAATGCACACTGAA

The repeating unit of mouse satellite DNA contains two half-repeats, which are aligned to show the identities (in blue), along with significant additional homologies between the first and second half of each half-repeat.



```
          10        20        30        40        50        60        70       G80        90       T100       110
GGACCTGGAATATGGCGAGAAAACTGAAAATCACGGAAAATGAGAAATACACACTTTAGGACGTGAAATATGGCGAGAAAACTGAAAAAGGTGGAAAATTAGAAATGTCCACTGTA

GGACGTGGAATATGGCAAGAAAACTGAAAATCATGGAAAATGAGAAACATCCACTTGACGACTTGAAAAATGACGAAATCACTAAAAAACGTGAAAAATGAGAAATGCACACTGAA
120       130       140       150       160       170       180       190       200       210       220       230
```



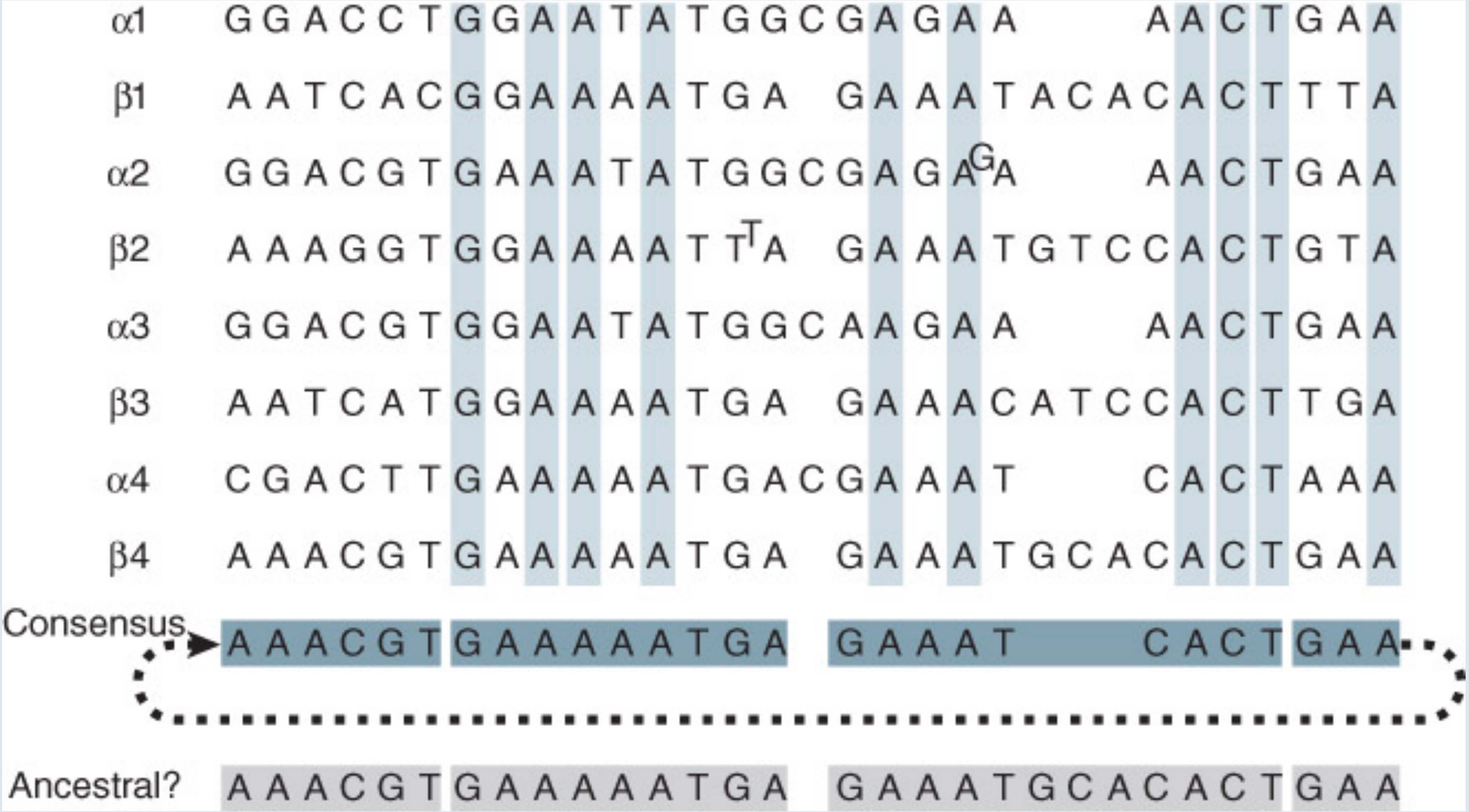
```
         10        20        30        40        50
GGACCTGGAATATGGCGAGAAAACTGAAAATCACGGAAAATGAGAAATACACACTTTA

 60        70       G80        90       T100      110
GGACGTGAAATATGGCGAGAAAACTGAAAAAGGTGGAAAATTAGAAATGTCCACTGTA

 120       130       140       150       160       170
GGACGTGGAATATGGCAAGAAAACTGAAAATCATGGAAAATGAGAAACATCCACTTGA

   180       190       200       210       220       230
CGACTTGAAAAATGACGAAATCACTAAAAAACGTGAAAAATGAGAAATGCACACTGAA
```

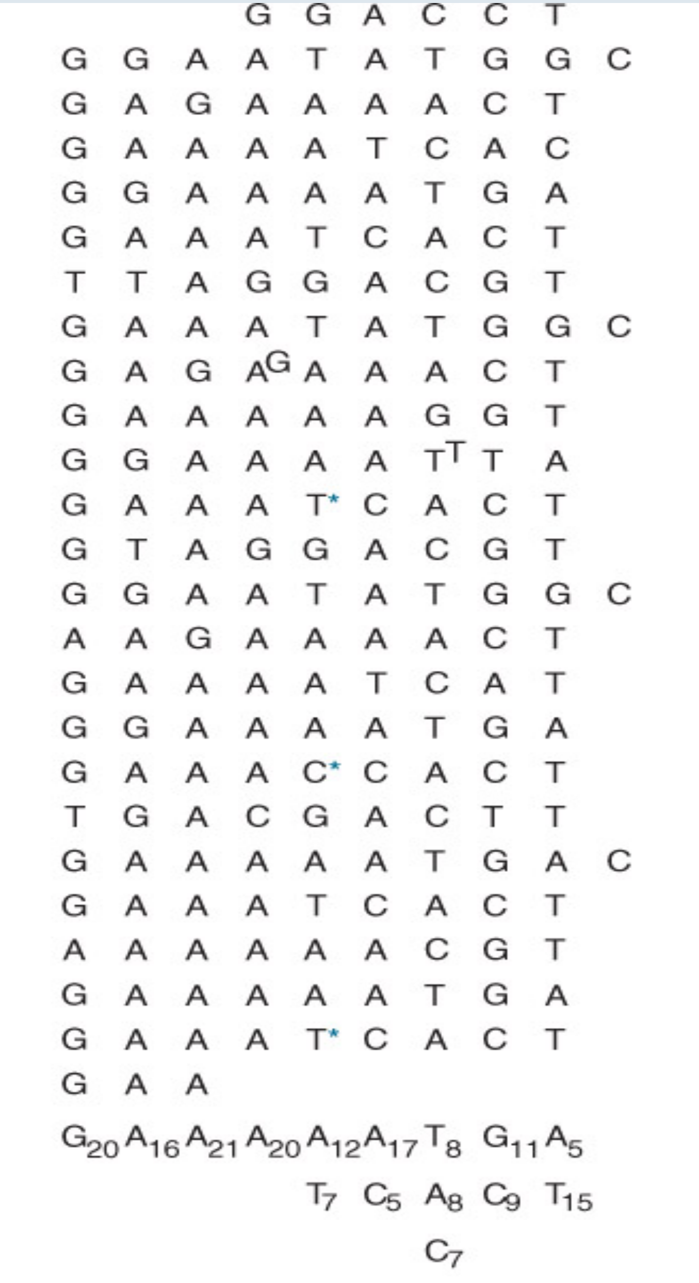
The repeating unit of mouse satellite DNA contains two half-repeats, which are aligned to show the identities (in blue), along with significant additional homologies between the first and second half of each half-repeat.



```
α1          G G A C C T G G A A T A T G G C G A G A A       A A C T G A A
β1          A A T C A C G G A A A A T G A   G A A A T A C A C A C T T T A
                                                   G
α2          G G A C G T G A A A T A T G G C G A G A A       A A C T G A A
                                       T
β2          A A A G G T G G A A A A T T A   G A A A T G T C C A C T G T A
α3          G G A C G T G G A A T A T G G C A A G A A       A A C T G A A
β3          A A T C A T G G A A A A T G A   G A A A C A T C C A C T T G A
α4          C G A C T T G A A A A A T G A C G A A A T       C A C T A A A
β4          A A A C G T G A A A A A T G A   G A A A T G C A C A C T G A A
Consensus   A A A C G T G A A A A A T G A   G A A A T       C A C T G A A

Ancestral?  A A A C G T G A A A A A T G A   G A A A T G C A C A C T G A A
```

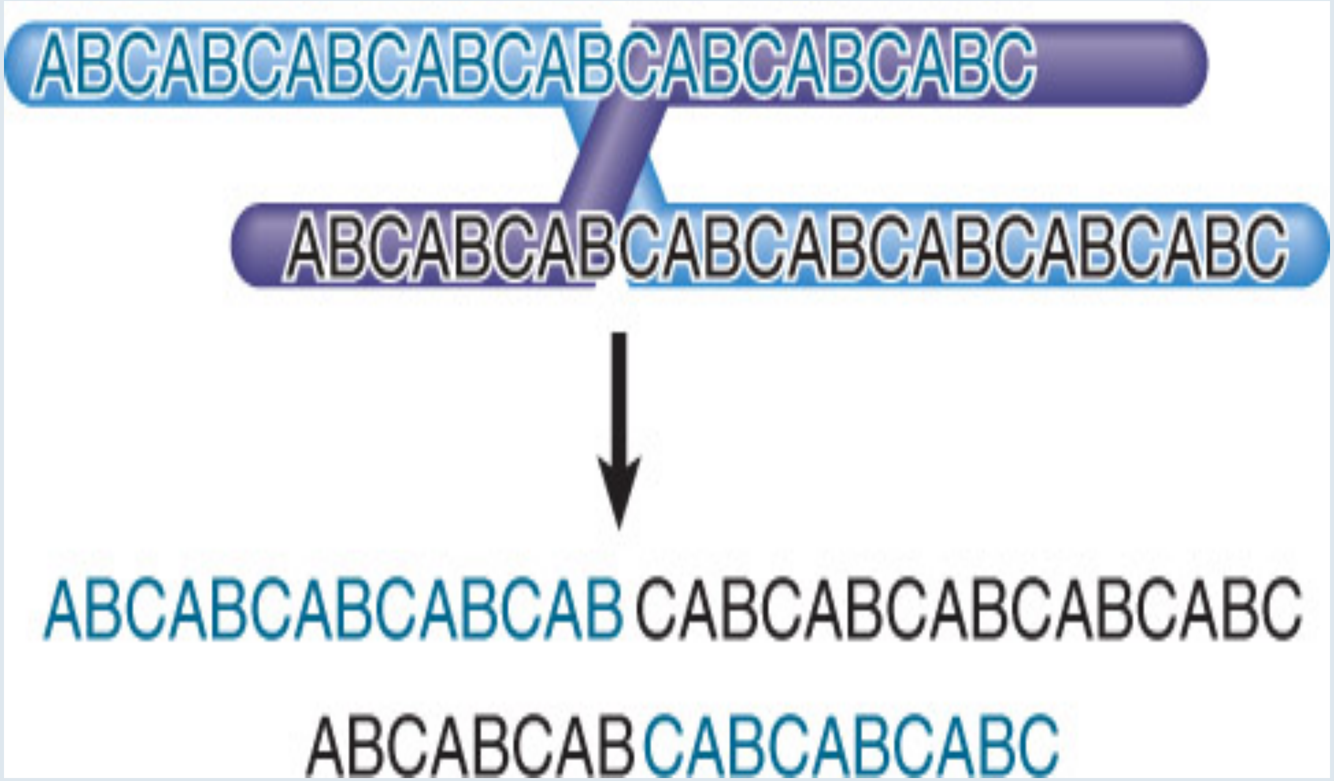
The alignment of eighth-repeats shows that each quarter-repeat consists of an α and a β half.

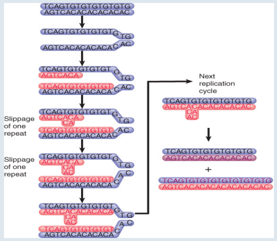


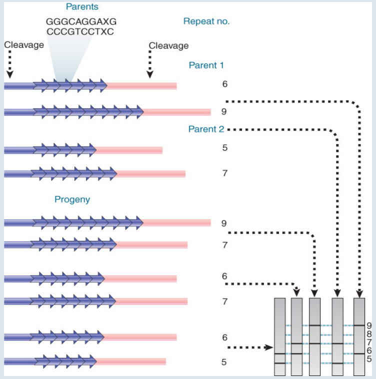
			G	G	A	C	C	T	
G	G	A	A	T	A	T	G	G	C
G	A	G	A	A	A	A	C	T	
G	A	A	A	A	T	C	A	C	
G	G	A	A	A	A	T	G	A	
G	A	A	A	T	C	A	C	T	
T	T	A	G	G	A	C	G	T	
G	A	A	A	T	A	T	G	G	C
G	A	G	A^G	A	A	A	C	T	
G	A	A	A	A	A	G	G	T	
G	G	A	A	A	A	T^T	T	A	
G	A	A	A	T*	C	A	C	T	
G	T	A	G	G	A	C	G	T	
G	G	A	A	T	A	T	G	G	C
A	A	G	A	A	A	A	C	T	
G	A	A	A	A	T	C	A	T	
G	G	A	A	A	A	T	G	A	
G	A	A	A	C*	C	A	C	T	
T	G	A	C	G	A	C	T	T	
G	A	A	A	A	A	T	G	A	C
G	A	A	A	T	C	A	C	T	
A	A	A	A	A	A	C	G	T	
G	A	A	A	A	A	T	G	A	
G	A	A	A	T*	C	A	C	T	
G	A	A							
G_{20}	A_{16}	A_{21}	A_{20}	A_{12}	A_{17}	T_8	G_{11}	A_5	
				T_7	C_5	A_8	C_9	T_{15}	
						C_7			

* indicates inserted triplet in β sequence
C in position 10 is extra base in α sequence

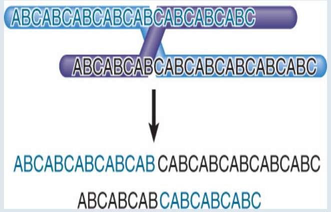
Eventually giving rise to the existence of an overall consensus sequence is shown by effectively writing the satellite sequence as a 9 bp repeat.

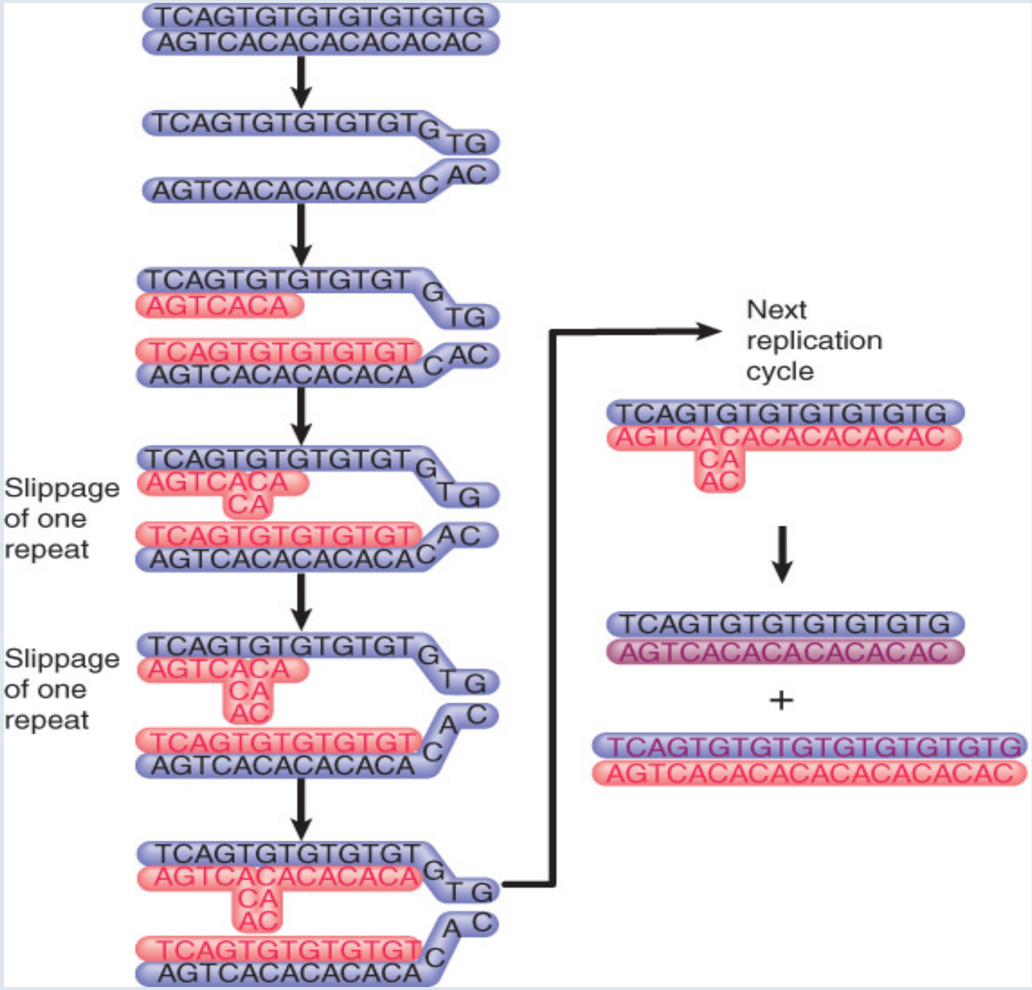


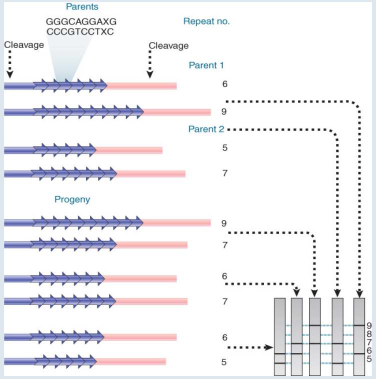




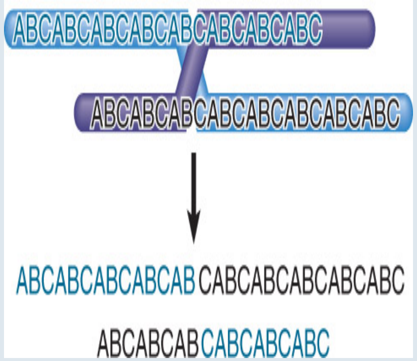
Alleles may differ by number of repeats at a minisatellite locus, so digestion generates restriction fragments that differ in length.

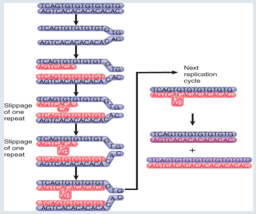


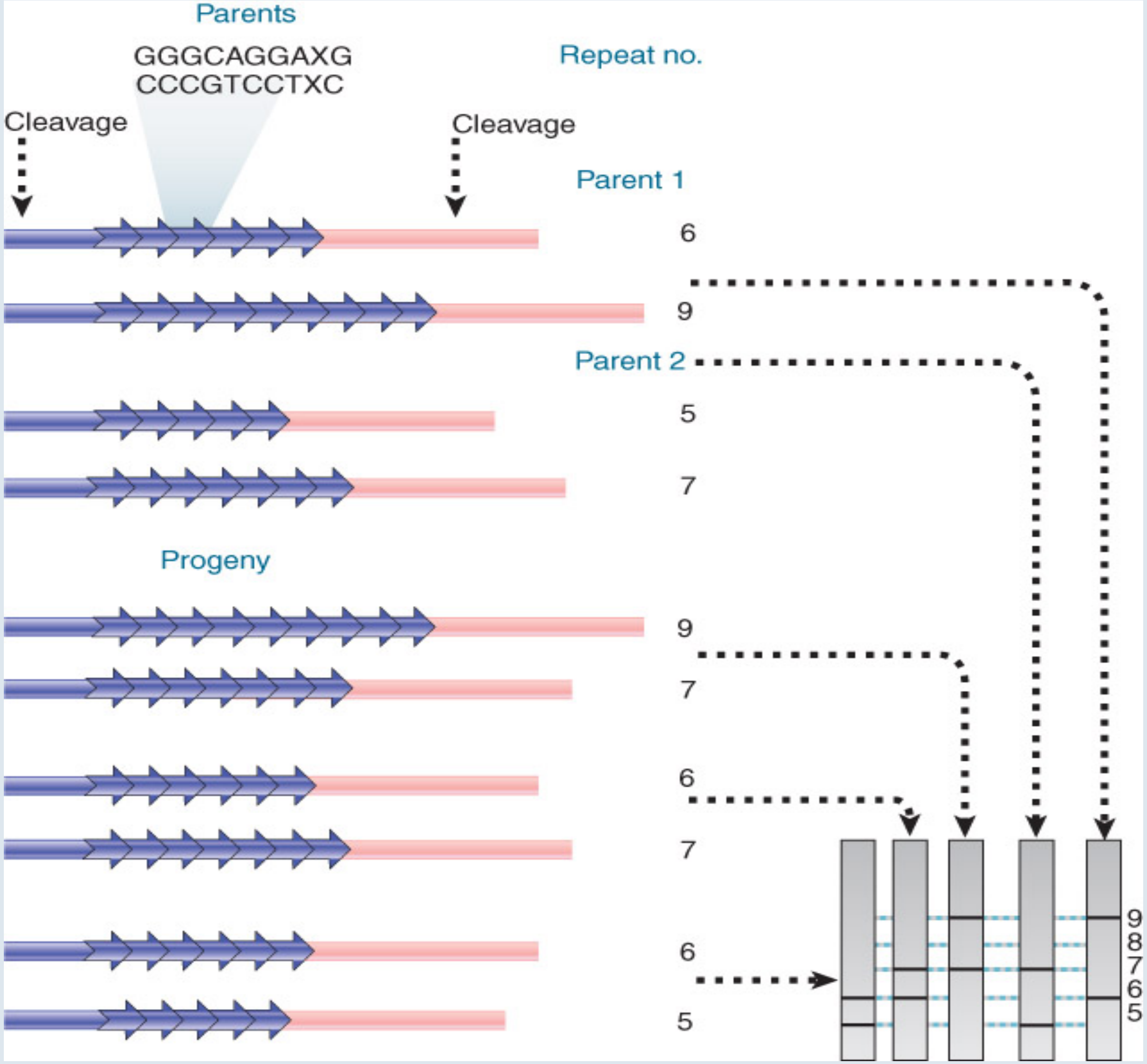




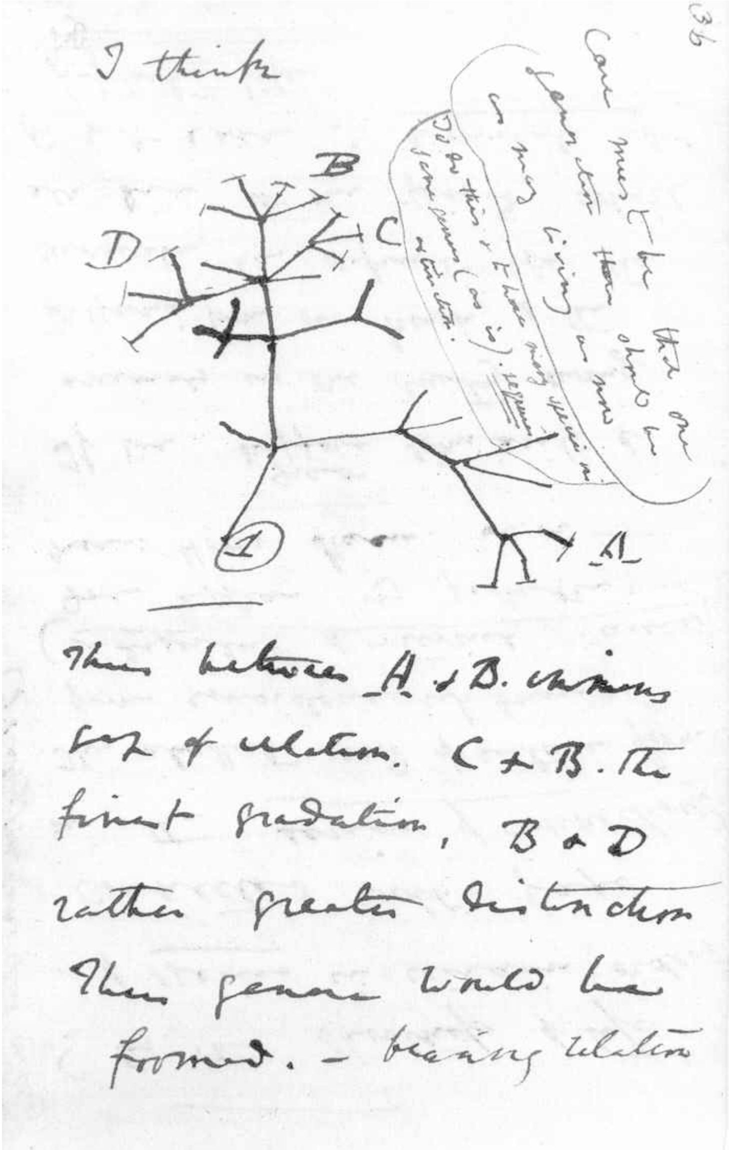
Alleles may differ by number of repeats at a minisatellite loci, so digestion of endonuclease sites flanking these repeats…generates restriction fragments that differ in length.







Alleles may differ by number of repeats at a minisatellite locus, so digestion generates restriction fragments that differ in length. VNTR's…

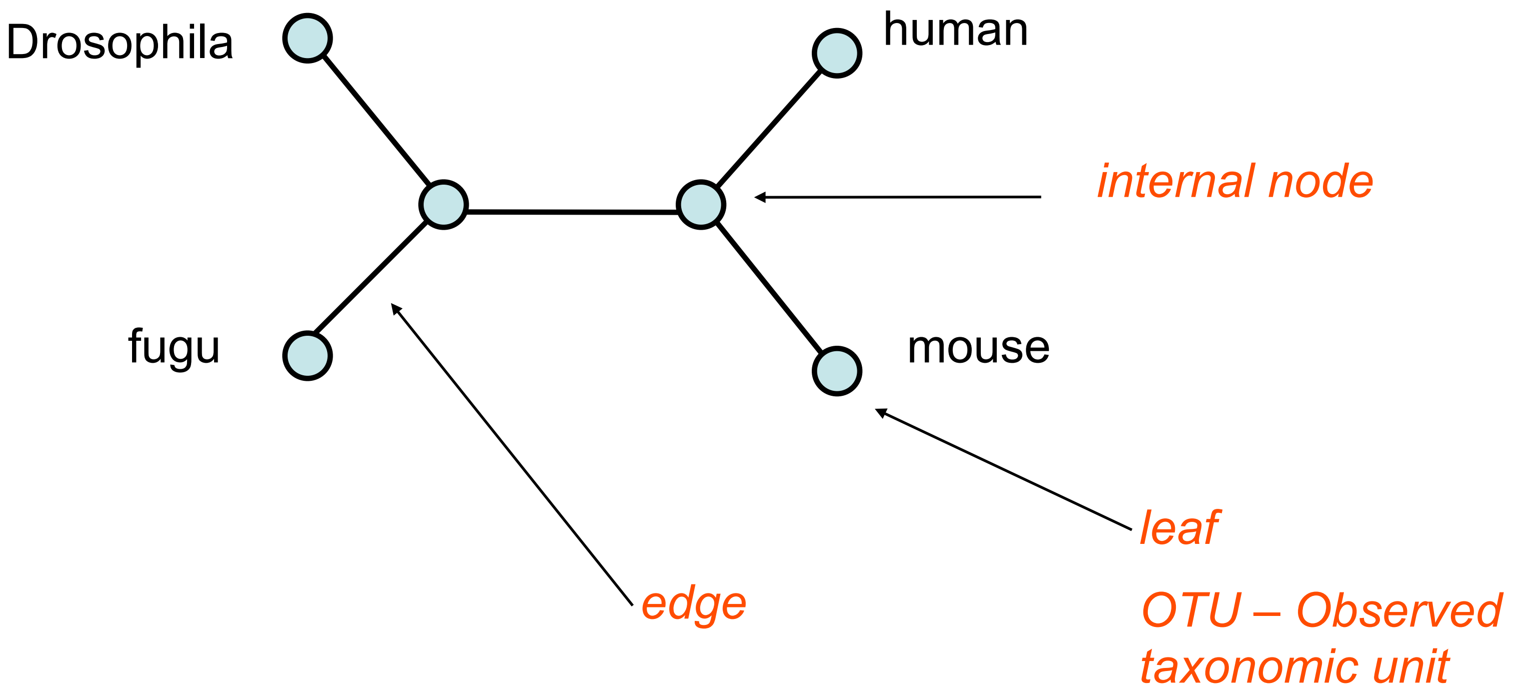

36
I think
B
C
D
1
A

Case must be that one generation then should be as many living as now. To do this & to have many species in same genus (as is). requires extinction.

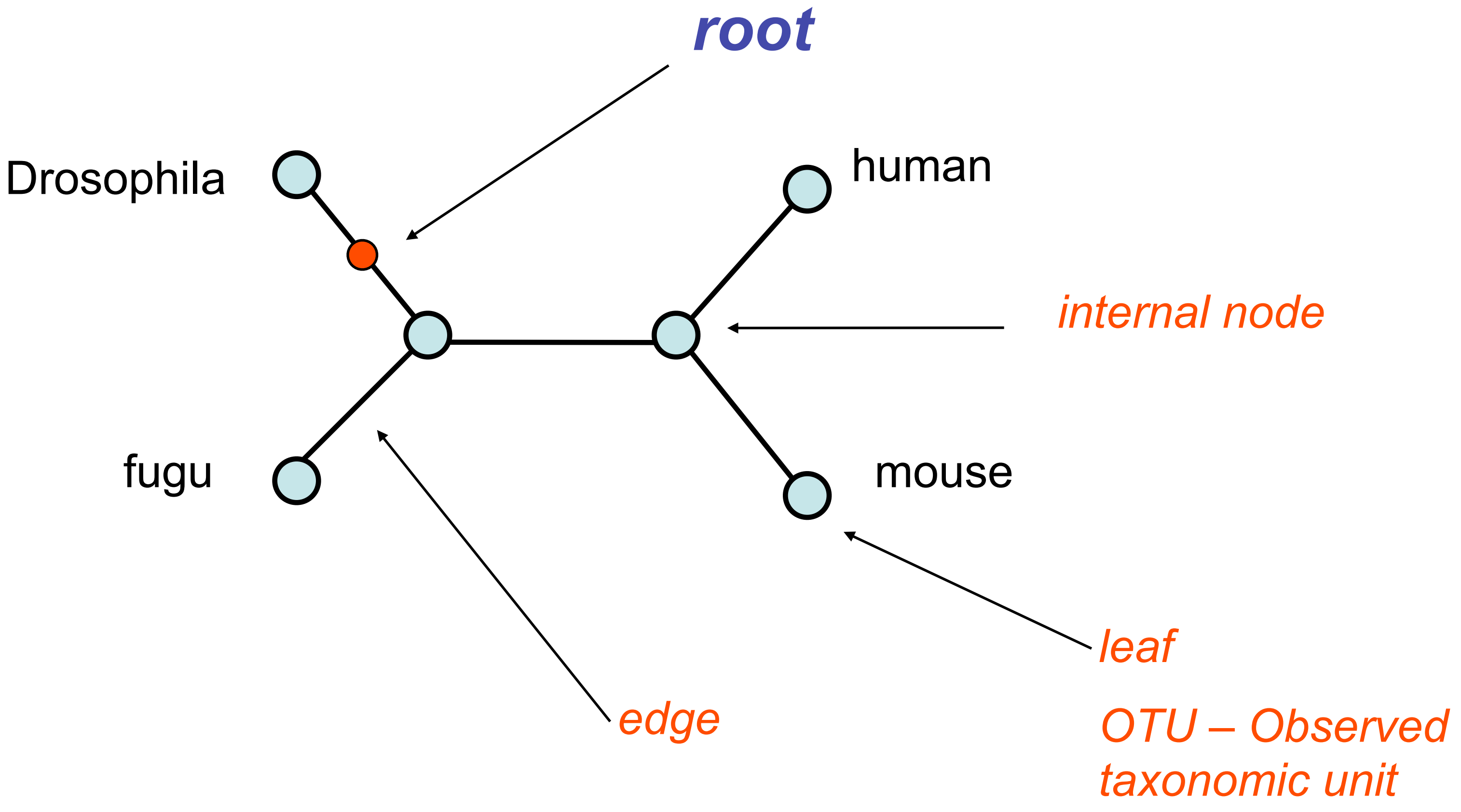
Thus between A & B. immense gap of relation. C & B. the finest gradation, B & D rather greater distinction Thus genera would be formed. – bearing relation

Exert from Darwin’s diary

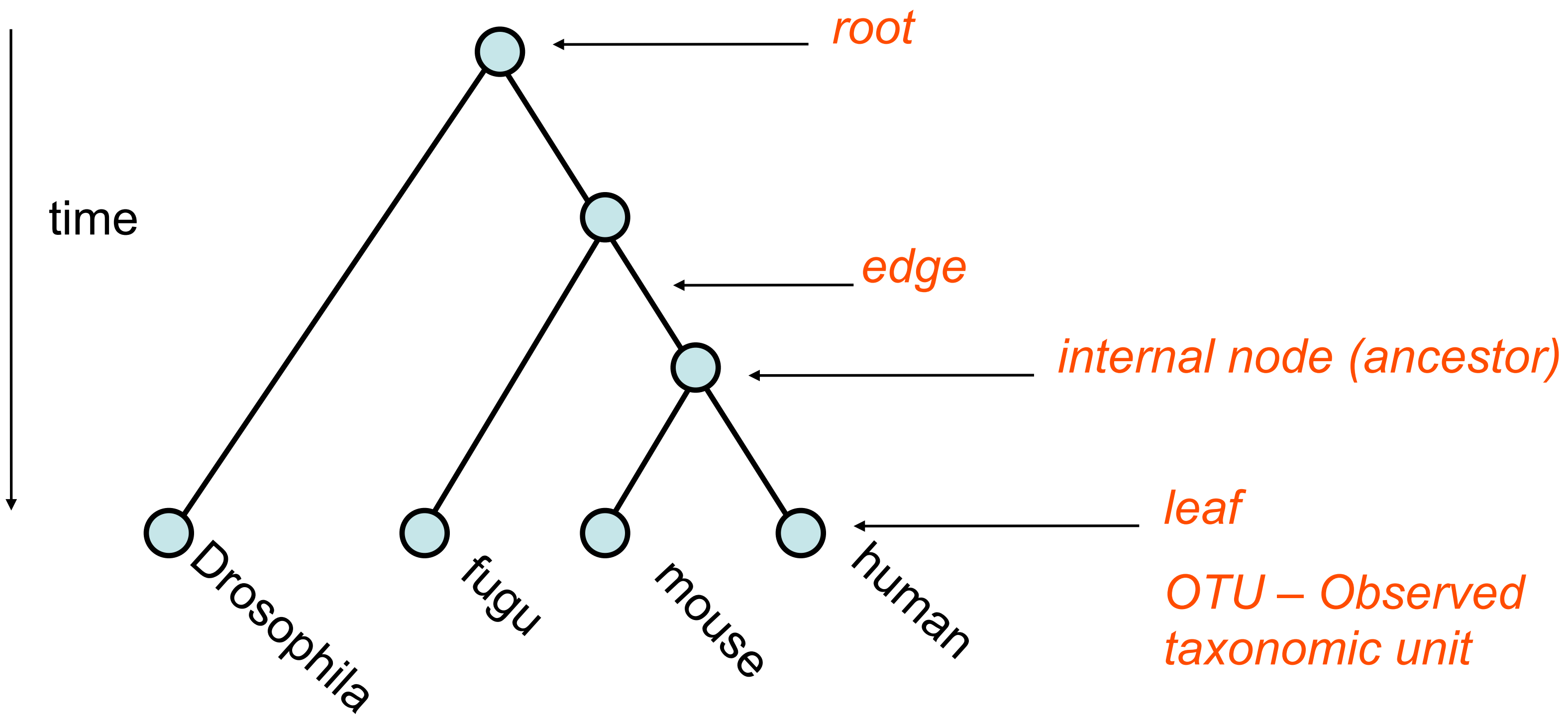
Phylogenetic tree (unrooted)



Phylogenetic tree (unrooted)



Phylogenetic tree (rooted)



time

root

edge

internal node (ancestor)

leaf

OTU – Observed taxonomic unit

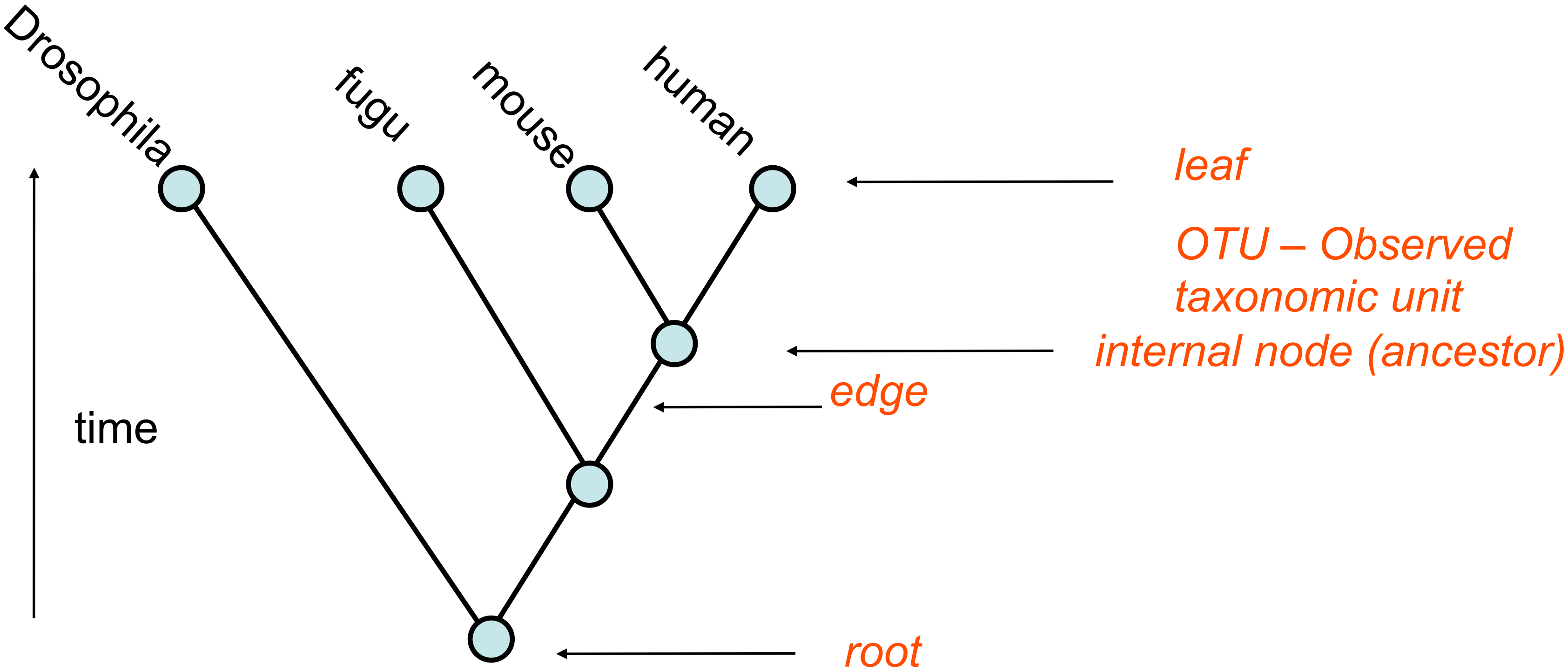
Drosophila

fugu

mouse

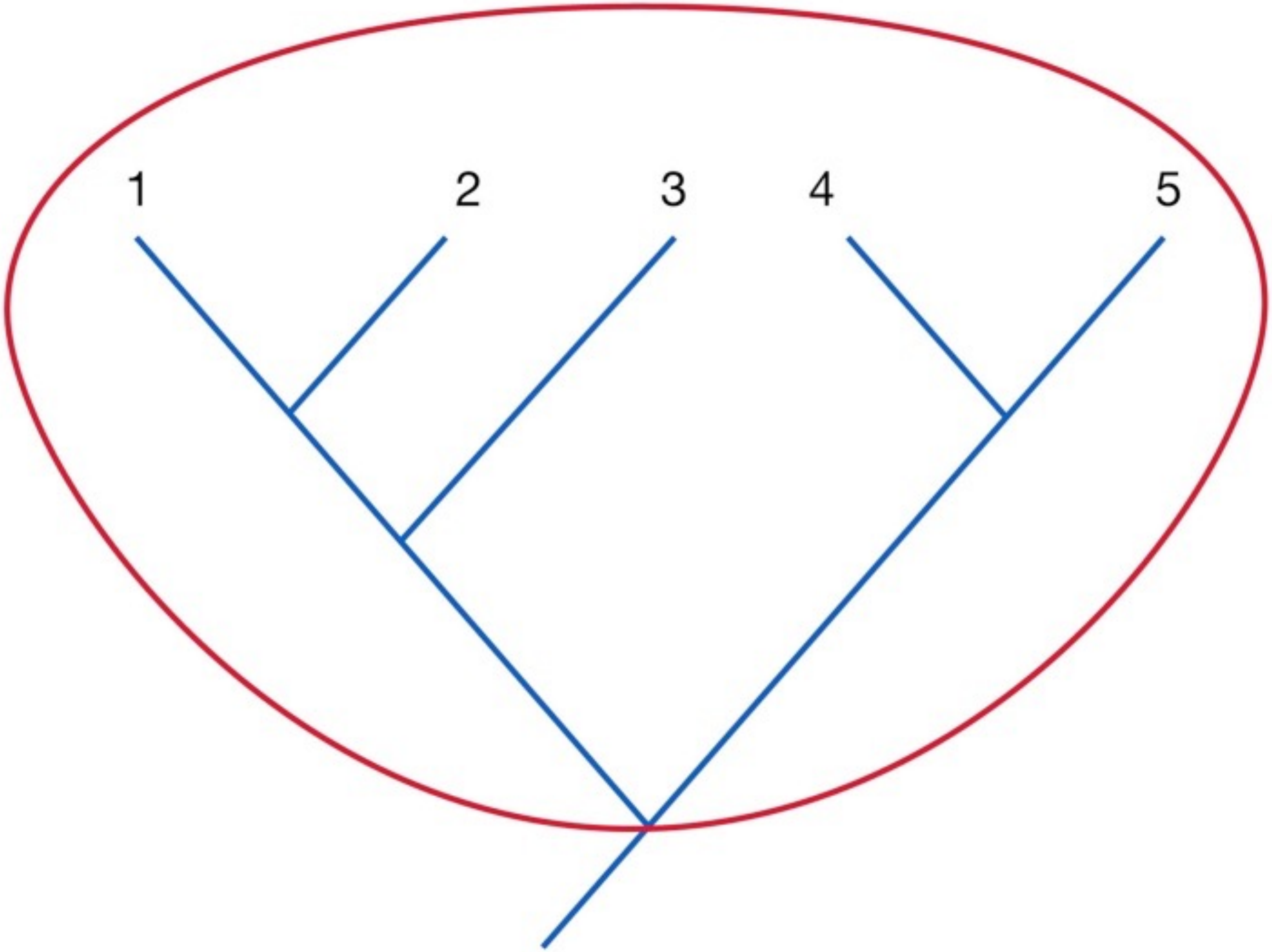
human

Phylogenetic tree (rooted)



(a)

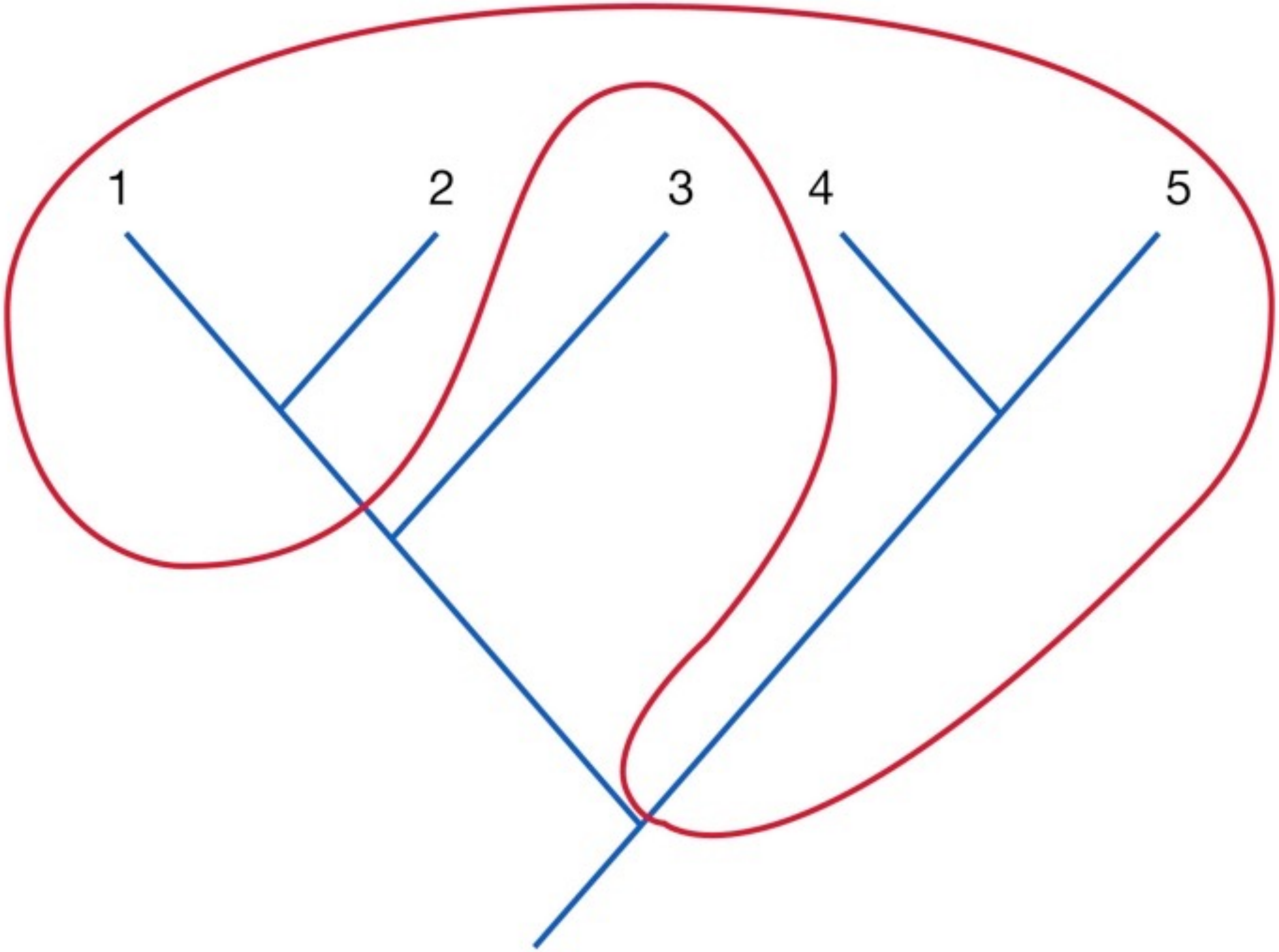
Monophyletic group



Copyright © 2004 Pearson Prentice Hall, Inc.

(b)

Paraphyletic group



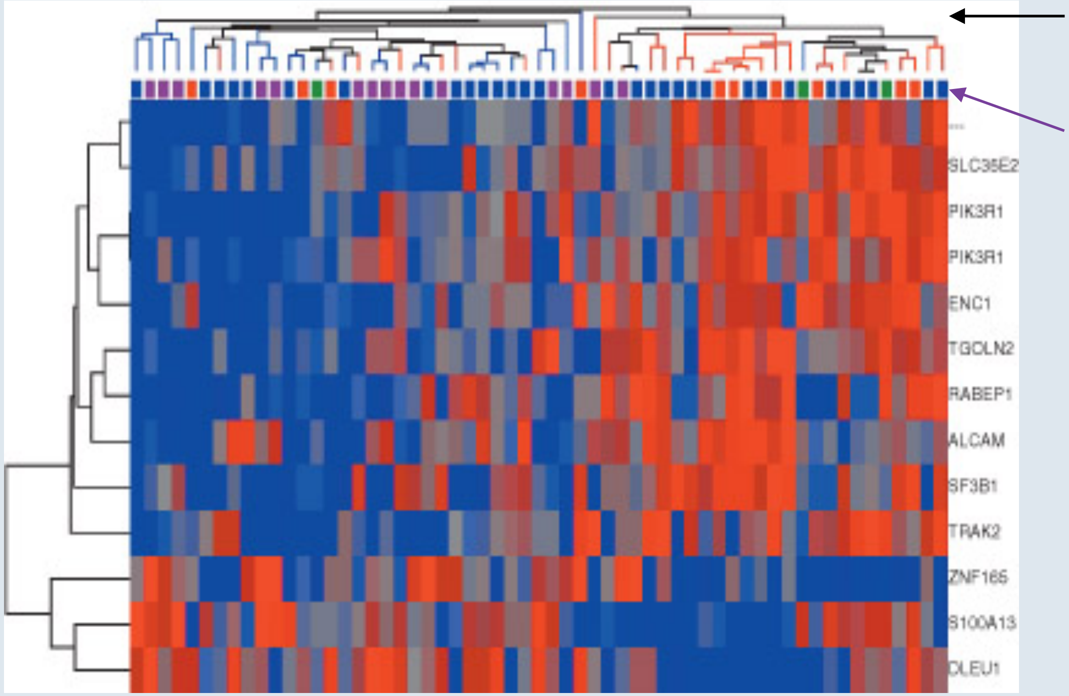
Comparing Characteristics
- Similarity Score -

Many properties can be used:

- Morphological characters
- Isoelectric points
- Molecular weights
- Nucleotides or amino acid composition

Expressed Gene Number Can Be Measured En Masse

- DNA microarray technology allows detailed comparisons of related animal cells to determine (for example) the differences in expression between a normal cell and a cancer cell.



women who breastfed ≥6 months (red lines) or who never breastfed (blue lines)

Different tumor subtypes blue, green, red, and purple bars

Heat map of 59 invasive breast tumors from women who breastfed ≥6 months or who never breastfed with RED - higher expression of tumors and BLUE lower expression of tumors.

Image courtesy of Rachel E. Ellsworth, Clinical Breast Care Project, Windber Research Institute.

- **Phenetics versus Cladistics**

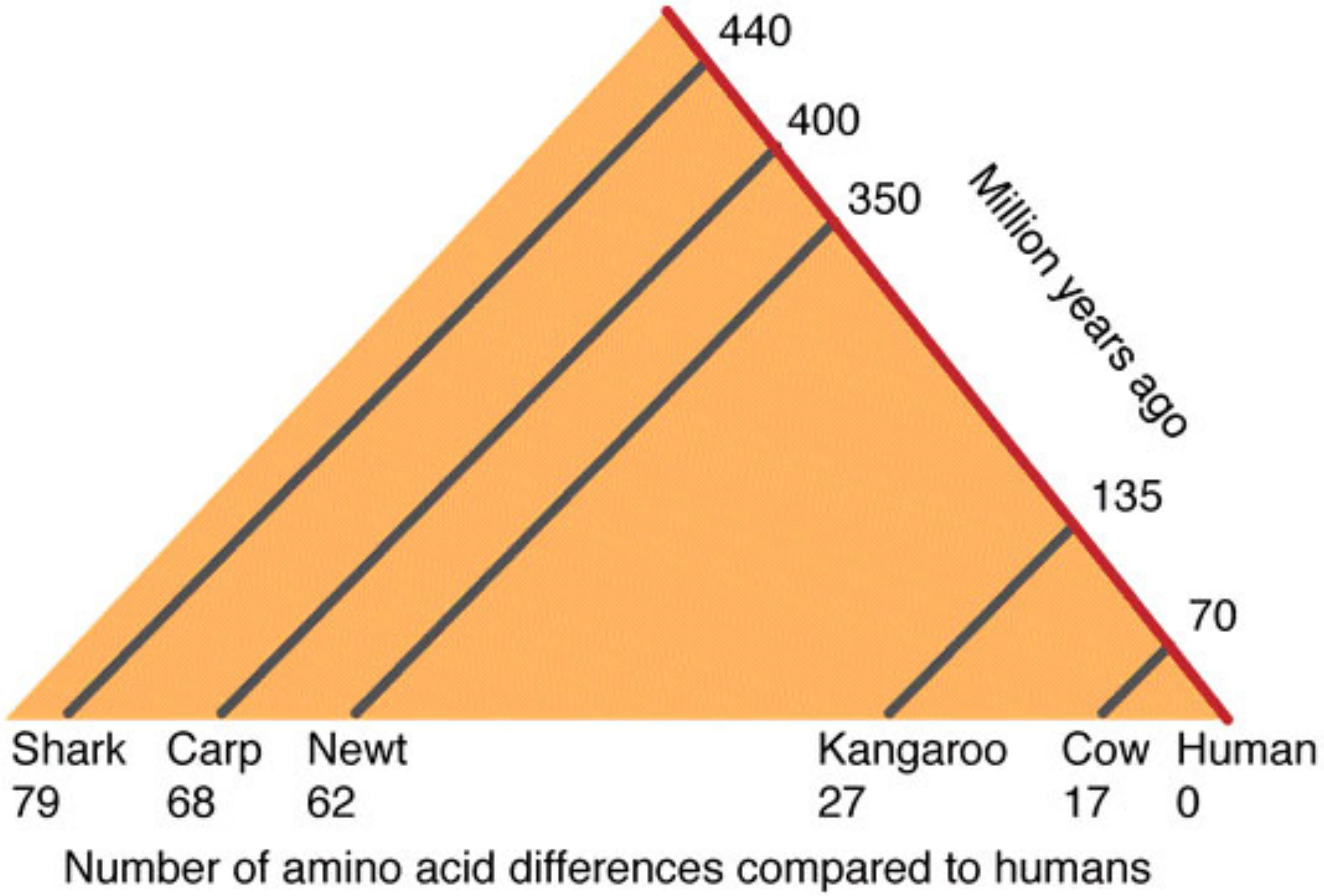
  - **Cladistics** can be defined as the **study of the pathways of evolution.** In other words, **cladists are interested in such questions as:** how many branches there are among a group of organisms; which branch connects to which other branch; and what is the branching sequence.

    A tree-like network that expresses such ancestor-descendant relationships is called a cladogram. Thus, a cladogram refers to the "**topology" of a rooted phylogenetic tree.**

  - **Phenetics** is the study of **relationships among a group of organisms** on the basis of the **degree of similarity** between them, be that similarity **molecular, phenotypic, or anatomical.**

    A tree-like network expressing **phenetic** relationships is called a **phenogram.**

    Choosing which tree is the "most reasonable" or demonstrates the "correct relationship" varies upon a knowledge of any number of factors, and is often resolved through the use of a **"maximium parsimony"** (Cladistic) and **UPGMA** [Unweighted Pair Group



Molecular Clocks..(?)..

The Port Jackson shark..

Heterodontus portusjacksoni

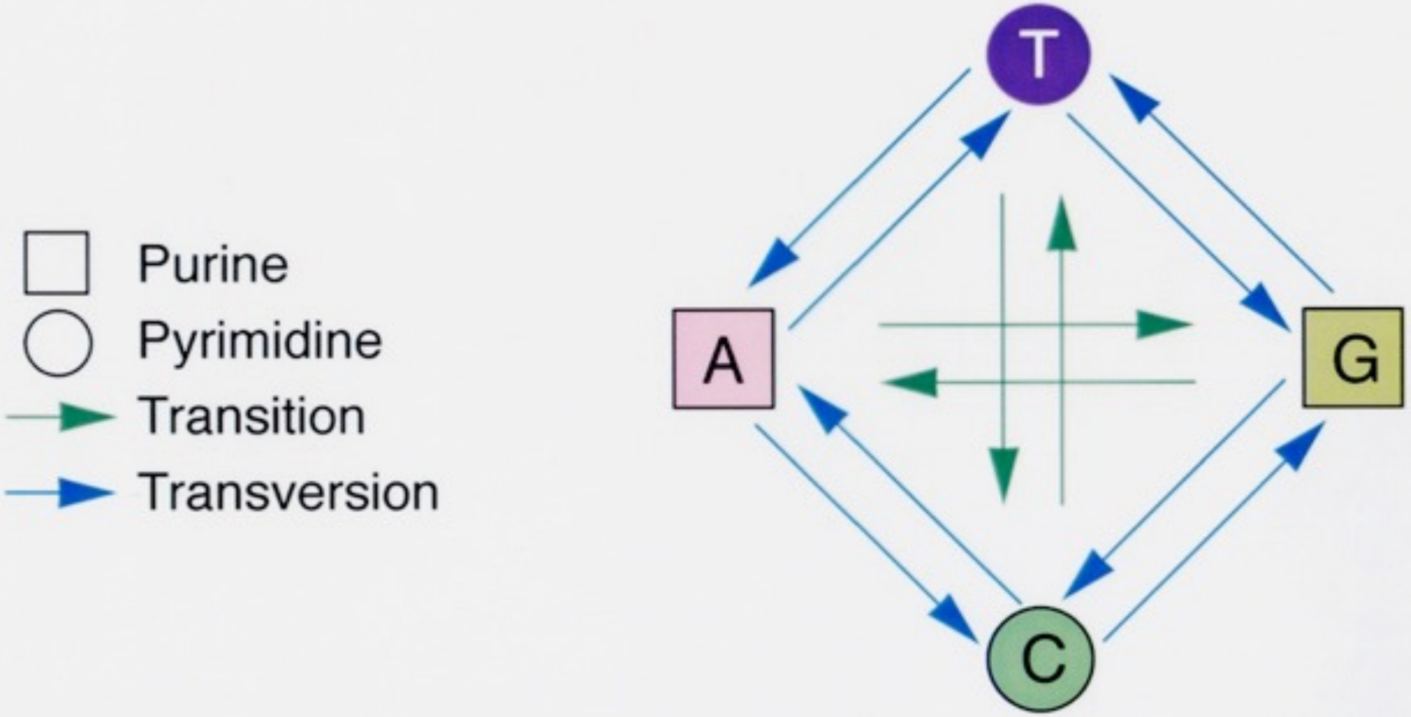
independence of molecular and morphological evolution

Globins	Number of amino acid changes
human alpha vs. human ß	147
carp alpha vs. human ß	149
shark-alpha vs. shark ß	150

Amino acid differences between the - and ß-hemoglobins, for three species pairs.

After Kimura (1983).

DNA Sequences can be envisaged to Evolve by Mutation followed by some some form of “Sorting Mechanism”



(a) Twelve different base substitutions can occur in DNA.

Selective Pressure Can Be Detected by Measuring Sequence Variation

- At the molecular level, the probability of a mutation becoming **fixed** in a population is influenced by the likelihood that the particular error/change will occur **and** the likelihood that it will be repaired.

- **synonymous mutation** – A change in DNA sequence in a coding region that **does not alter** the amino acid that is encoded.

- **non synonymous mutation** – A change in DNA sequence in a coding region that **alters** the amino acid that is encoded.

- Neutral mutation -a change in DNA sequence that gives NO selective advantage or disadvantage

Selective pressure Can Be Detected by Measuring Sequence Variation

- The ratio of **non synonymous** to **synonymous** substitutions in the evolutionary history of a gene is a measure of positive and/or negative selection.

- Low heterozygosity of a gene may indicate recent selective events.

- **genetic hitchhiking –** The change in frequency of a genetic variant due to its linkage to a selected variant at another locus.

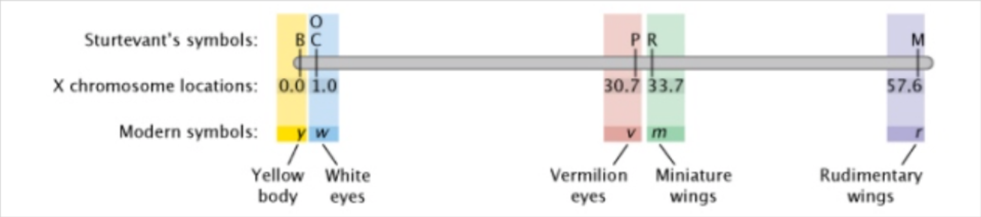


Figure 3: Sturtevant's *Drosophila* gene map.
In Sturtevant's gene map, six traits are arranged along a linear chromosome according to the relative distance of each from trait B. Traits include yellow body (B), white eyes (C, O), Vermillion eyes (P), miniature wings (R), and rudimentary wings (M).
© 2013 Nature Education Adapted from Pierce, Benjamin. *Genetics: A Conceptual Approach*, 2nd ed. All rights reserved.

DNA Sequences can be envisaged to Evolve by Mutation followed by some some form of "Sorting Mechanism"

- **Neo Darwinism:** Natural Selection vs. Genetic Drift

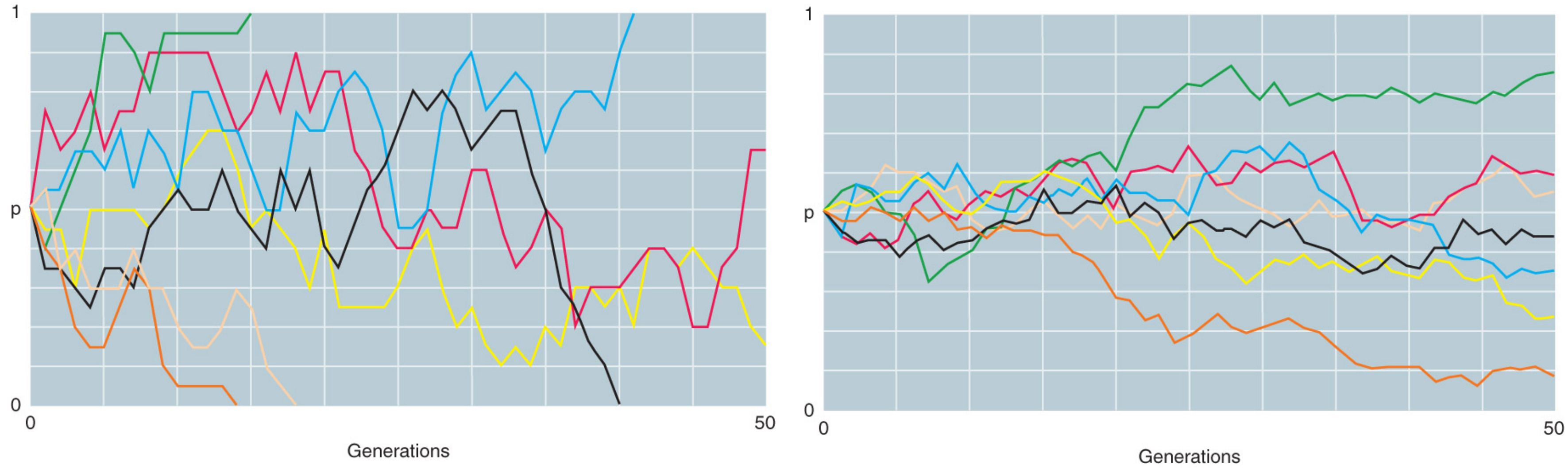
- In small populations, the frequency of a mutation will change randomly and new mutations are likely to be eliminated by selection or chance.

- **fixation –** The process by which a new allele replaces the allele that was previously predominant in a population.

- The frequency of a mutation that affects phenotype will be influenced by negative or positive selection and also population size

- Whereas, the frequency of a **neutral mutation** largely depends on **genetic drift**, the strength of which depends on the size of the population

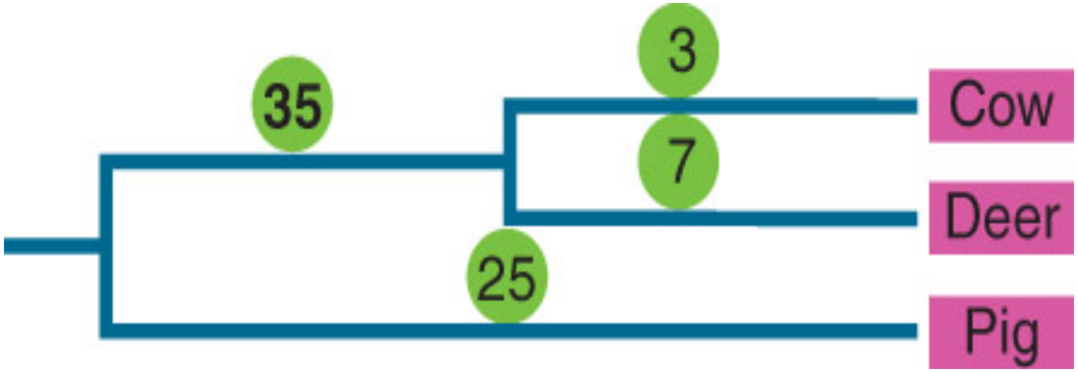
The fixation or loss of alleles by random genetic drift occurs more rapidly in (A) populations of 10 than in (B) populations of 100



Data courtesy of Kent E. Holsinger, University of Connecticut [http://darwin.eeb.uconn.edu]

- Comparing the rates of substitution among related species can indicate whether **selection** on the gene has occurred.

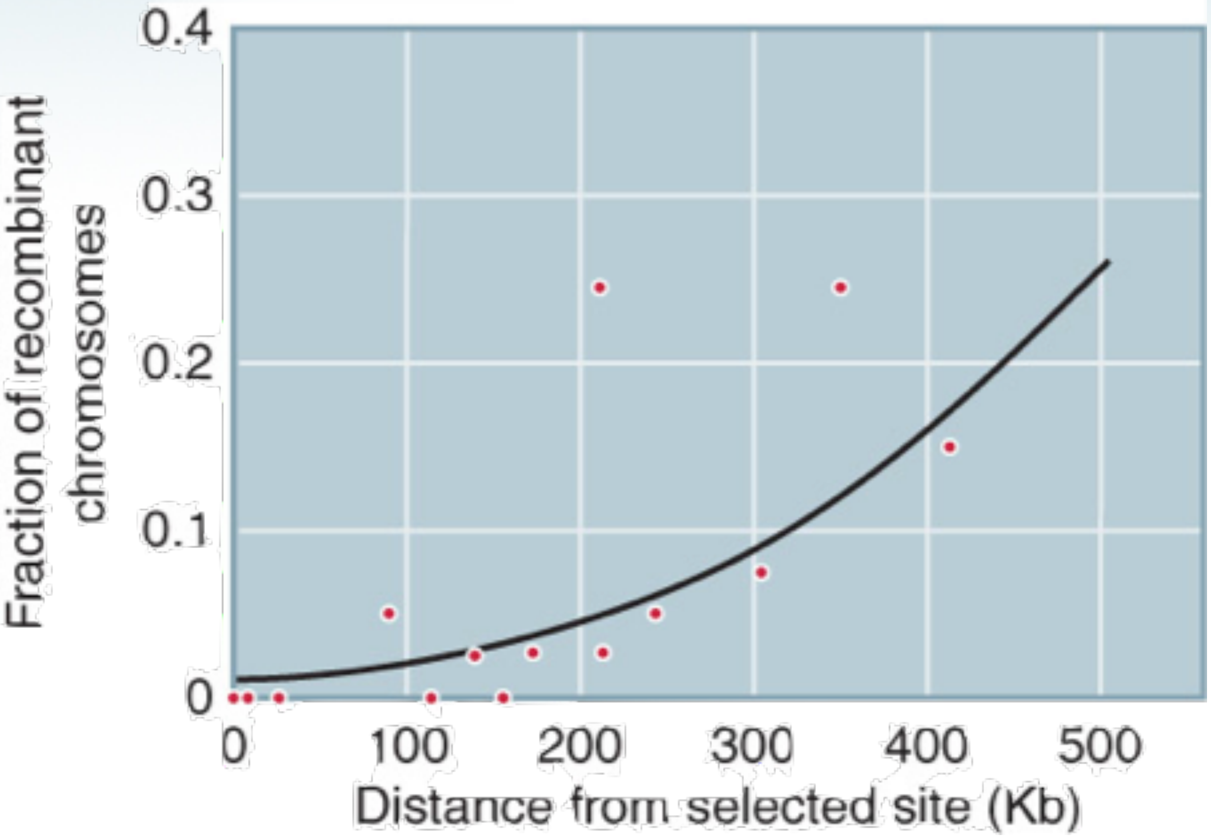
- **linkage disequilibrium –** A nonrandom association between alleles at two different loci, often as a result of linkage.



A higher number of **non synonymous** substitutions in lysozyme sequences in the cow/deer lineage as compared to the pig lineage…

Adapted from N. H. Barton, et al. Evolution. Cold Spring Harbor Laboratory Press, 2007. Original figure appeared in J. H. Gillespie, The Causes of Molecular Evolution. Oxford University Press, 1991.

Selection Can Be Detected by Measuring Sequence Variation



The recently cloned G6PD allele has rapidly increased in frequency

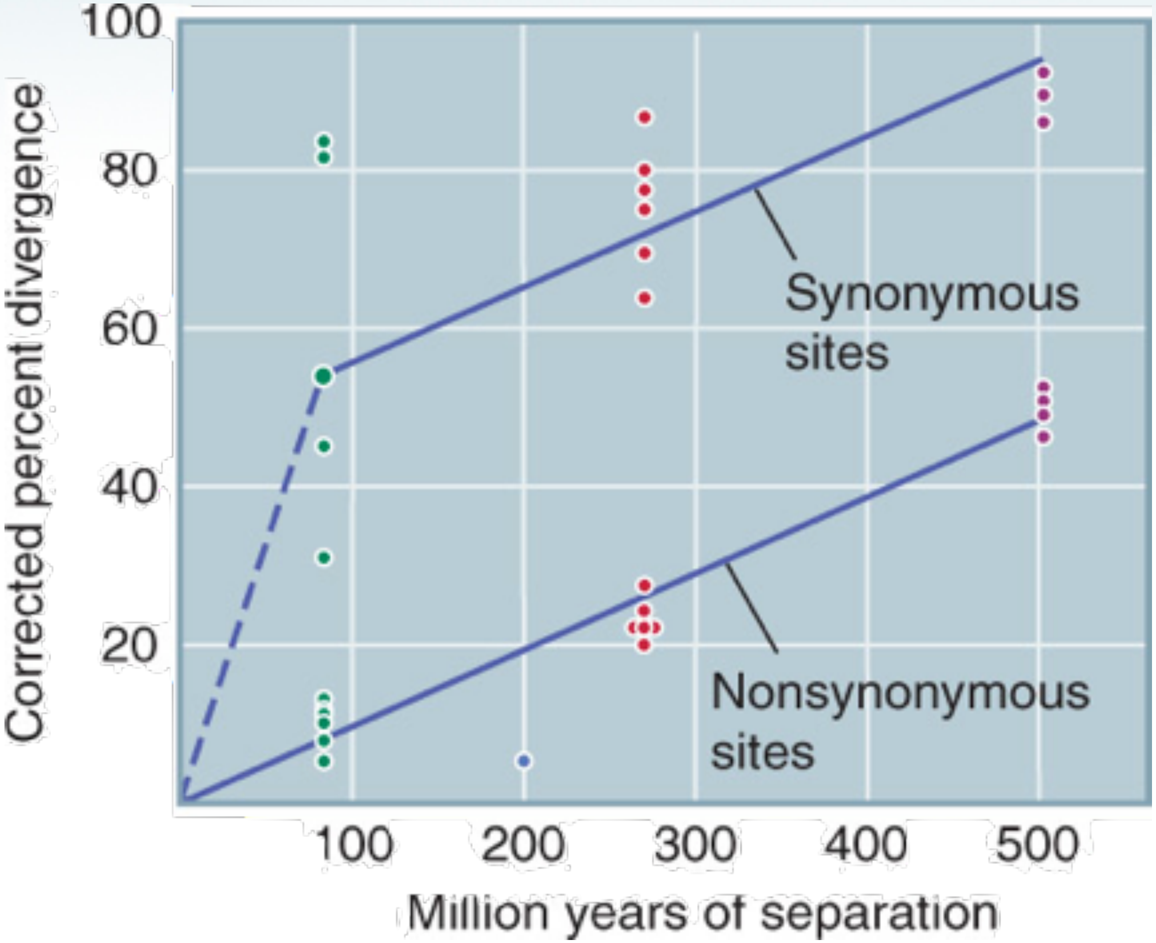
Adapted from E. T. Wang, et al., Proc. Natl. Acad. Sci. USA 103 (2006): 135-140.

A Constant Rate of Sequence Divergence would give rise to a "**Molecular Clock**" -like Sorting Mechanism

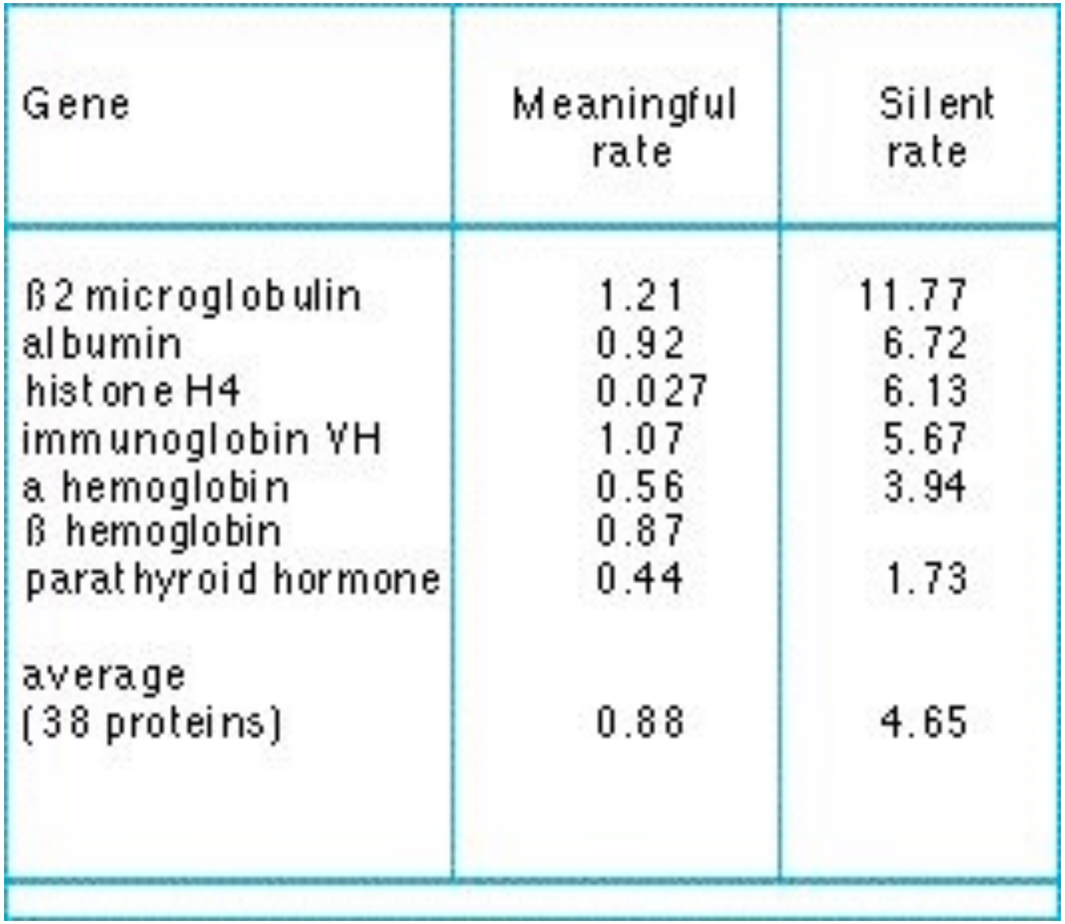
- The sequences of **orthologous** genes in different species vary at **non synonymous** sites (where mutations have caused amino acid substitutions) and **synonymous** sites (where mutation has not affected the amino acid sequence).

- **Synonymous** substitutions accumulate **~10×** ~10× faster than **non synonymous** substitutions.

A Constant Rate of Sequence Divergence Is a Molecular Clock



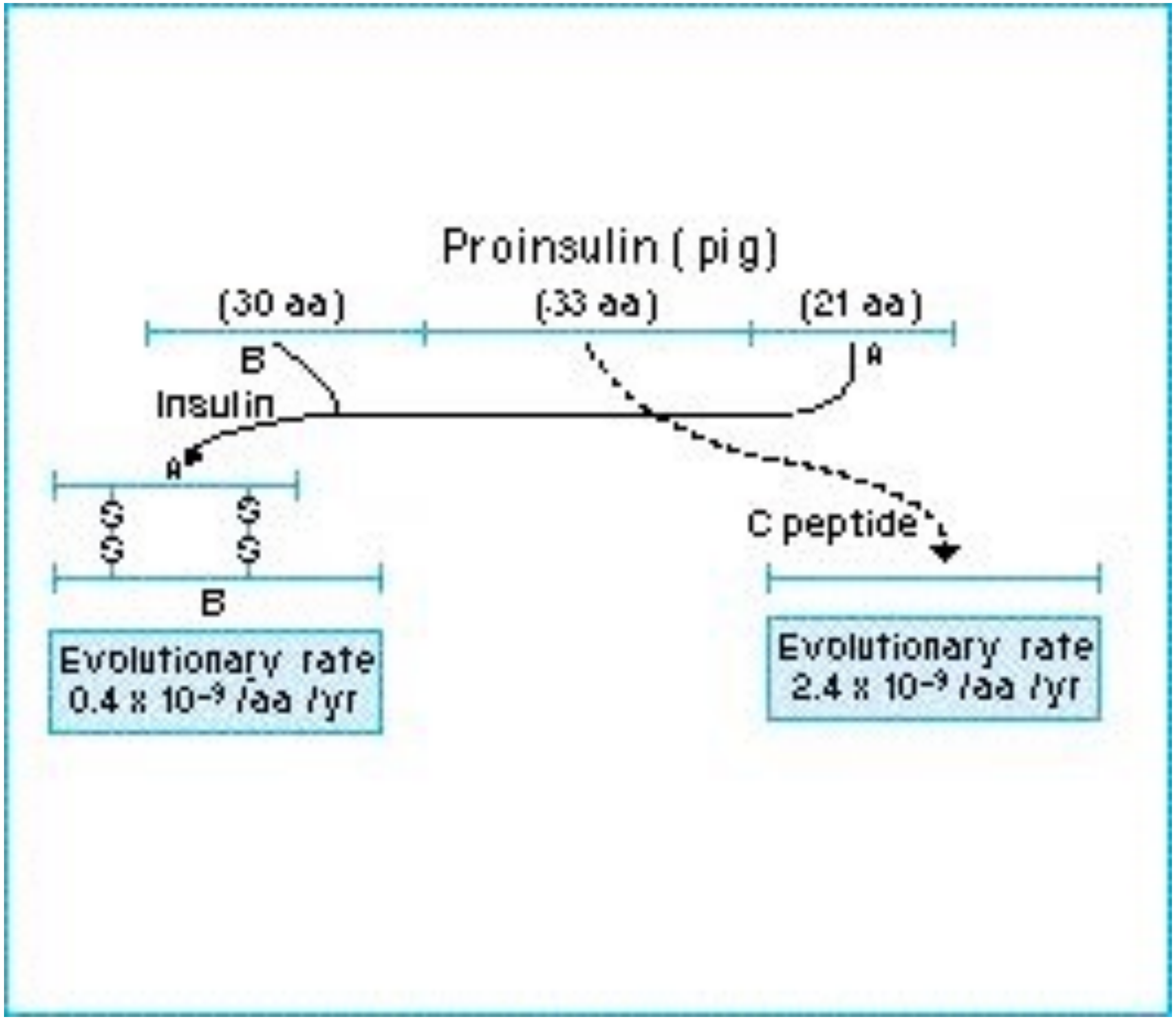
Divergence of DNA sequences depends on evolutionary separation



Gene	Meaningful rate	Silent rate
ß2 microglobulin	1.21	11.77
albumin	0.92	6.72
histone H4	0.027	6.13
immunoglobin VH	1.07	5.67
a hemoglobin	0.56	3.94
ß hemoglobin	0.87	
parathyroid hormone	0.44	1.73
average (38 proteins)	0.88	4.65

Rates of evolution for “meaningful “ (i.e. amino acid changing) and silent base changes in various genes.

Rates are expressed as inferred number of base changes per 109 years. Simplified from Li, Wu & Luo (1985).


Proinsulin (pig)
(30 aa)
(33 aa)
(21 aa)
B
A
Insulin
A
S S
S S
B
C peptide
Evolutionary rate 0.4 x 10⁻⁹ /aa /yr
Evolutionary rate 2.4 x 10⁻⁹ /aa /yr

The insulin protein is made by snipping the center out of a larger proinsulin peptide. The rate of evolution in the central part, which is discarded, is found to be **slightly higher** than that of the functional extremities. From Kimura (1983).

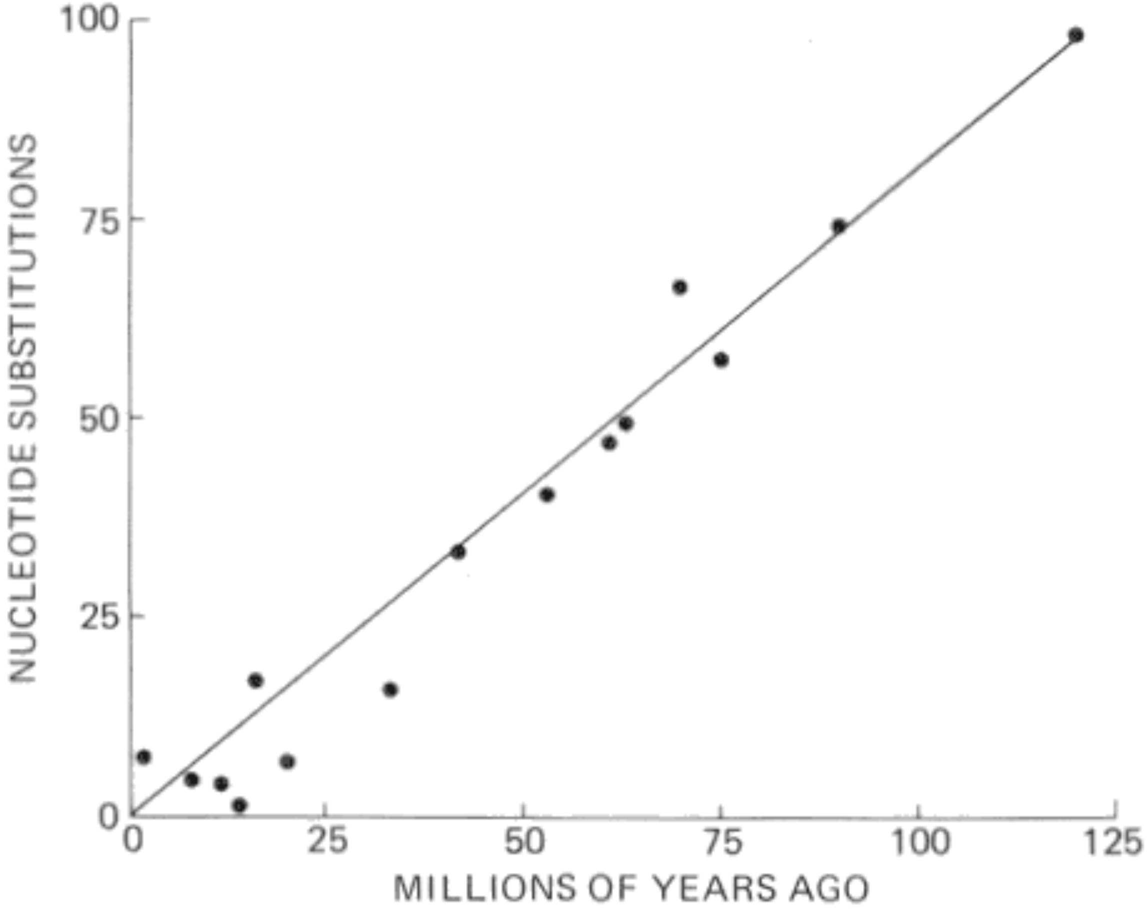


Figure: the rate of evolution of hemoglobin. Each point on the graph is for a pair of species, or groups of species. Some of the points are for a-hemoglobin, others for ß -hemoglobin. From Kimura (1983).

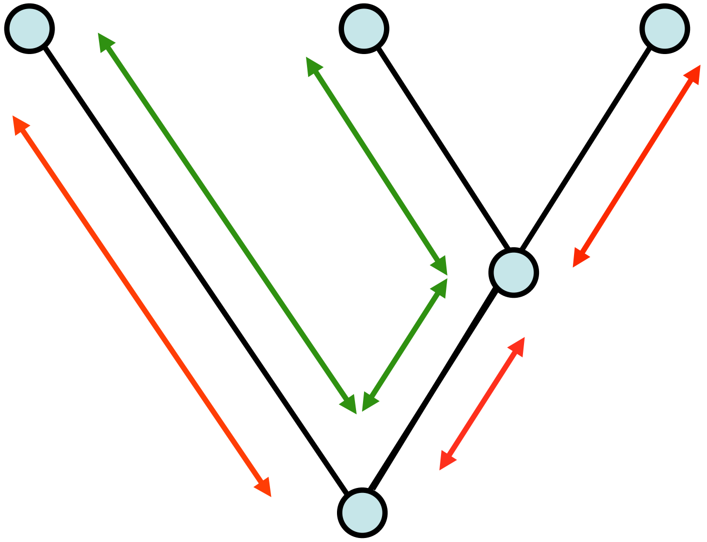
- The evolutionary **divergence** between two DNA sequences is measured by the “corrected” percent of positions at which the corresponding nucleotides differ.

- Substitutions may appear to accumulate at a more or less constant rate after genes separate, so that the divergence between any pair of **globin sequences** (for example) is proportional to the time since they last shared a common ancestry.

sequence C (functional)

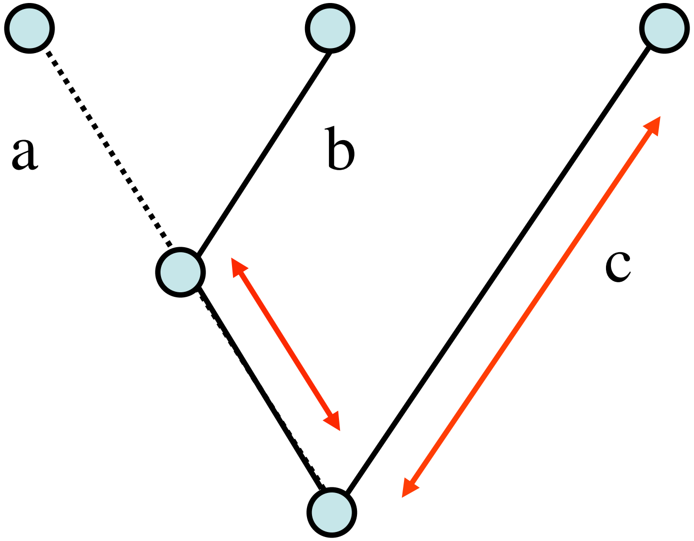
sequence A (functional)

sequence B (functional)



root

To test a MC, a relative rate test that does not depend on absolute divergence times can be used.

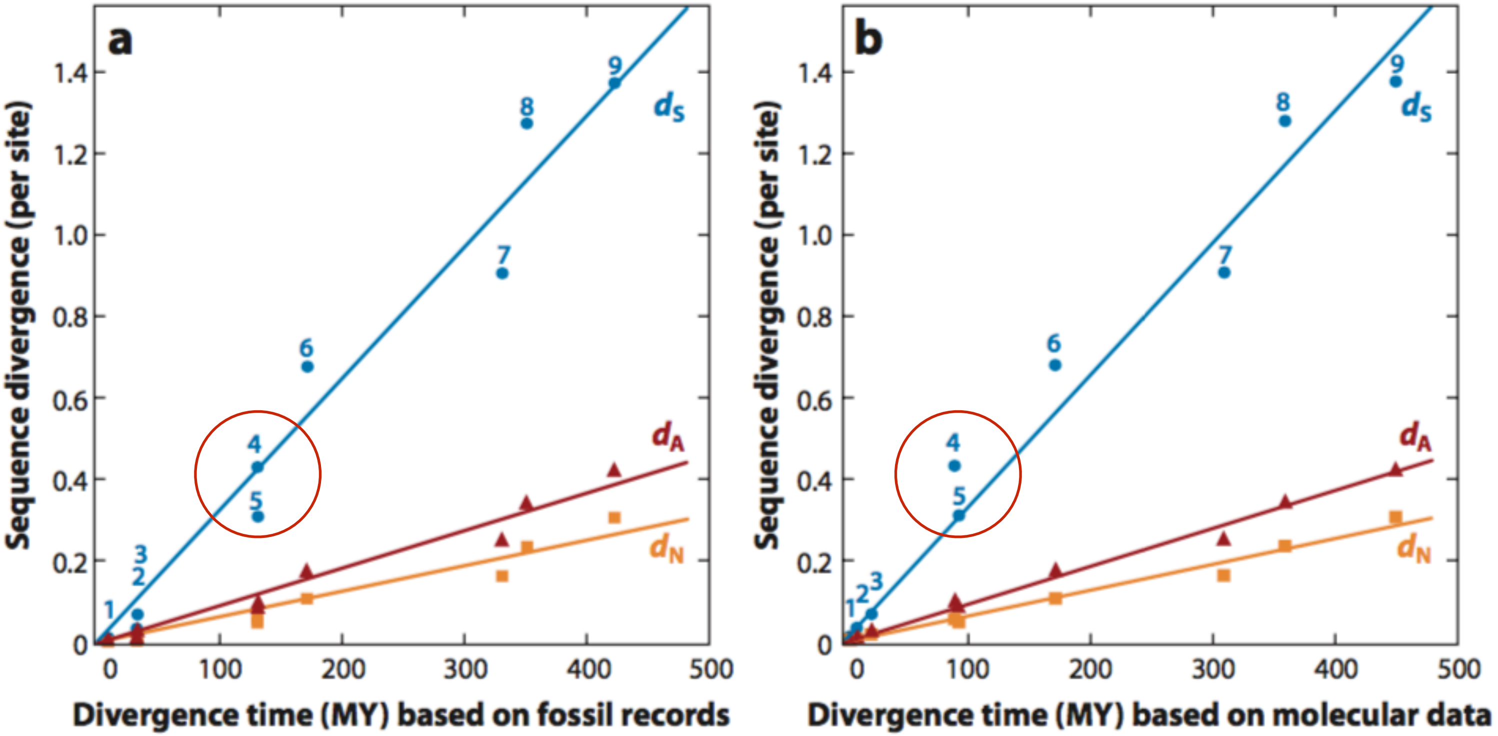


Short generation species (eg. mouse)

Long generation (eg. whale)

outgroup

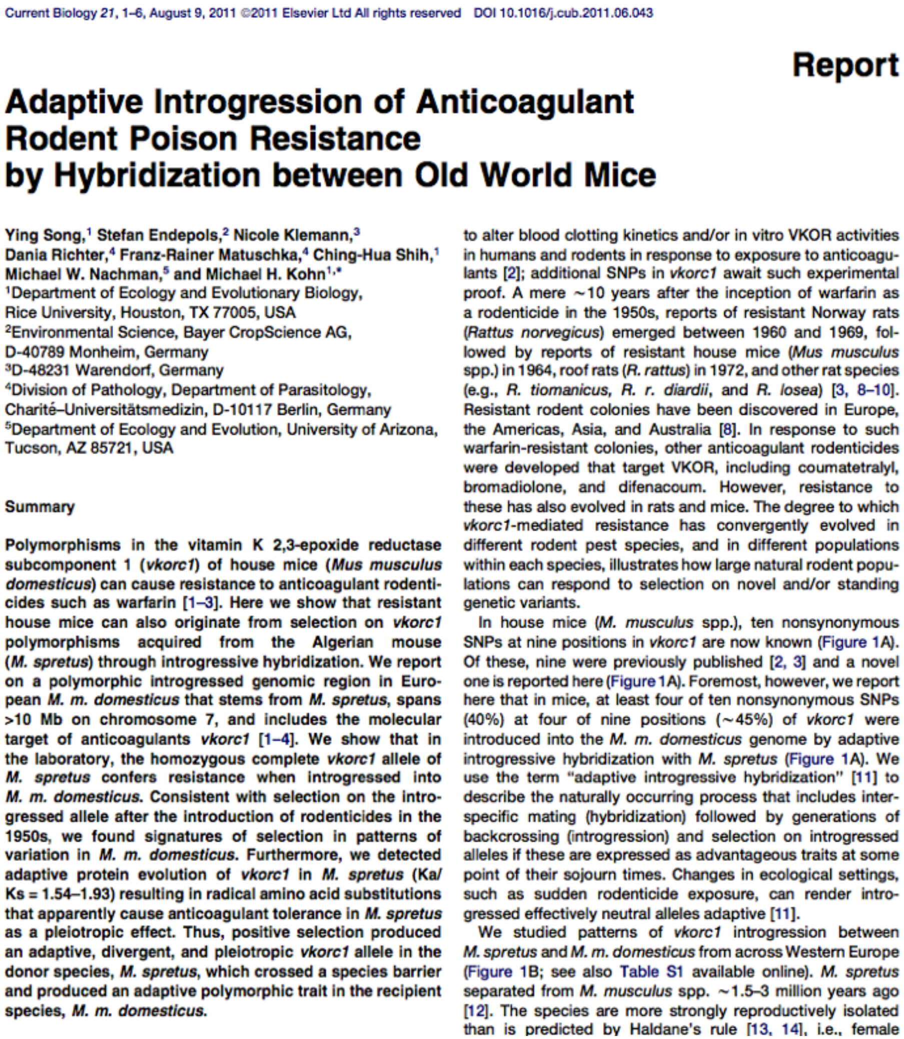
Where a, b and c are the numbers of evolutionary changes in the three segments of the tree. The “out group” can be any species known to have a much greater distant common ancestor between each of the pair of species being compared. The evidence suggests that “a” approximately equals “b” for many species, whereas a would be less than b if generation time influenced evolutionary rate…



Linear relationships of the number of amino acid substitutions per residue (d_A) and the numbers of synonymous (d_S) and nonsynonymous (d_N) nucleotide substitutions per site, with divergence times based on the **fossil record** (a) **and molecular data** *b*).

Each point represents the average sequence divergence of 4,198 nuclear genes with ≥100 codons from 10 vertebrate species (**human** versus1 = **chimpanzee**, 2 = **orangutan,** 3 = **macaque**, 4 = **mouse**, 5 = **cow**, 6 = **opossum**, 7 = **chicken**, 8 = **western clawed frog**, 9 = **zebrafish**). Sequence and orthology data are from Ensembl (147). The d_A distance was computed by the Poisson correction method, whereas d_S and d_N were computed by the modified Nei–Gojobori method (178) with a transition/transversion ratio of 2.

Masatoshi Nei, Yoshiyuki Suzuki, and Masafumi Nozawa Annu. Rev. Genomics Hum. Genet. 2010. 11:265–89



Current Biology *21*, 1–6, August 9, 2011 ©2011 Elsevier Ltd All rights reserved DOI 10.1016/j.cub.2011.06.043

Report

Adaptive Introgression of Anticoagulant Rodent Poison Resistance by Hybridization between Old World Mice

Ying Song,[1] Stefan Endepols,[2] Nicole Klemann,[3] Dania Richter,[4] Franz-Rainer Matuschka,[4] Ching-Hua Shih,[1] Michael W. Nachman,[5] and Michael H. Kohn[1,*]

[1]Department of Ecology and Evolutionary Biology, Rice University, Houston, TX 77005, USA
[2]Environmental Science, Bayer CropScience AG, D-40789 Monheim, Germany
[3]D-48231 Warendorf, Germany
[4]Division of Pathology, Department of Parasitology, Charité–Universitätsmedizin, D-10117 Berlin, Germany
[5]Department of Ecology and Evolution, University of Arizona, Tucson, AZ 85721, USA

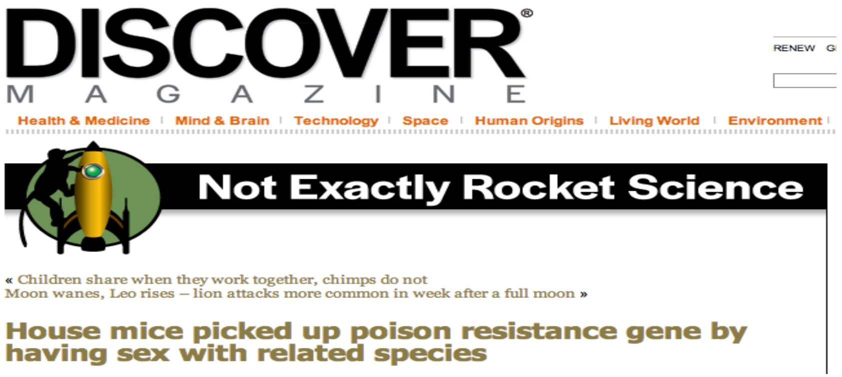
Summary

Polymorphisms in the vitamin K 2,3-epoxide reductase subcomponent 1 (*vkorc1*) of house mice (*Mus musculus domesticus*) can cause resistance to anticoagulant rodenticides such as warfarin [1–3]. Here we show that resistant house mice can also originate from selection on *vkorc1* polymorphisms acquired from the Algerian mouse (*M. spretus*) through introgressive hybridization. We report on a polymorphic introgressed genomic region in European *M. m. domesticus* that stems from *M. spretus*, spans >10 Mb on chromosome 7, and includes the molecular target of anticoagulants *vkorc1* [1–4]. We show that in the laboratory, the homozygous complete *vkorc1* allele of *M. spretus* confers resistance when introgressed into *M. m. domesticus*. Consistent with selection on the introgressed allele after the introduction of rodenticides in the 1950s, we found signatures of selection in patterns of variation in *M. m. domesticus*. Furthermore, we detected adaptive protein evolution of *vkorc1* in *M. spretus* (Ka/Ks = 1.54–1.93) resulting in radical amino acid substitutions that apparently cause anticoagulant tolerance in *M. spretus* as a pleiotropic effect. Thus, positive selection produced an adaptive, divergent, and pleiotropic *vkorc1* allele in the donor species, *M. spretus*, which crossed a species barrier and produced an adaptive polymorphic trait in the recipient species, *M. m. domesticus*.

to alter blood clotting kinetics and/or in vitro VKOR activities in humans and rodents in response to exposure to anticoagulants [2]; additional SNPs in *vkorc1* await such experimental proof. A mere ~10 years after the inception of warfarin as a rodenticide in the 1950s, reports of resistant Norway rats (*Rattus norvegicus*) emerged between 1960 and 1969, followed by reports of resistant house mice (*Mus musculus* spp.) in 1964, roof rats (*R. rattus*) in 1972, and other rat species (e.g., *R. tiomanicus*, *R. r. diardii*, and *R. losea*) [3, 8–10]. Resistant rodent colonies have been discovered in Europe, the Americas, Asia, and Australia [8]. In response to such warfarin-resistant colonies, other anticoagulant rodenticides were developed that target VKOR, including coumatetralyl, bromadiolone, and difenacoum. However, resistance to these has also evolved in rats and mice. The degree to which *vkorc1*-mediated resistance has convergently evolved in different rodent pest species, and in different populations within each species, illustrates how large natural rodent populations can respond to selection on novel and/or standing genetic variants.

In house mice (*M. musculus* spp.), ten nonsynonymous SNPs at nine positions in *vkorc1* are now known (Figure 1A). Of these, nine were previously published [2, 3] and a novel one is reported here (Figure 1A). Foremost, however, we report here that in mice, at least four of ten nonsynonymous SNPs (40%) at four of nine positions (~45%) of *vkorc1* were introduced into the *M. m. domesticus* genome by adaptive introgressive hybridization with *M. spretus* (Figure 1A). We use the term "adaptive introgressive hybridization" [11] to describe the naturally occurring process that includes interspecific mating (hybridization) followed by generations of backcrossing (introgression) and selection on introgressed alleles if these are expressed as advantageous traits at some point of their sojourn times. Changes in ecological settings, such as sudden rodenticide exposure, can render introgressed effectively neutral alleles adaptive [11].

We studied patterns of *vkorc1* introgression between *M. spretus* and *M. m. domesticus* from across Western Europe (Figure 1B; see also Table S1 available online). *M. spretus* separated from *M. musculus* spp. ~1.5–3 million years ago [12]. The species are more strongly reproductively isolated than is predicted by Haldane's rule [13, 14], i.e., female



DISCOVER®
MAGAZINE
Health & Medicine | Mind & Brain | Technology | Space | Human Origins | Living World | Environment |
RENEW G
Not Exactly Rocket Science

« Children share when they work together, chimps do not
Moon wanes, Leo rises – lion attacks more common in week after a full moon »

House mice picked up poison resistance gene by having sex with related species

Warfarin works by acting against vitamin K. This vitamin activates a number of genes that create clots in blood, but it itself has to be activated by a protein called VKORC1. Warfarin stops **VKORC1** from doing its job, thereby **suppressing vitamin K**. The clotting process fails, and bleeds continue to bleed.

Rodents can evolve to shrug off warfarin by tweaking their **vkorc1 gene**, which encodes the protein of the same name.

In European house mice, scientists have found at least 10 different genetic changes (mutations) in **vkorc1** that change how susceptible they are to warfarin. But only six of these changes were the house mouse's own innovations. The other four came from a close relative – the Algerian mouse, which is found throughout northern Africa, Spain, Portugal, and southern France.

The two species separated from each other between 1.5 and 3 million years ago. They rarely met, but when they have, they can breed with one another. The two species have identifiably different versions of vkorc1. But Song found that virtually all **Spanish house mice carry a copy of vkorc1 that partially or totally matches the Algerian mouse version**.

Even in Germany, where the two species don't mingle, a third of house mice now copies of **vkorc1** that descended from Algerian peers.

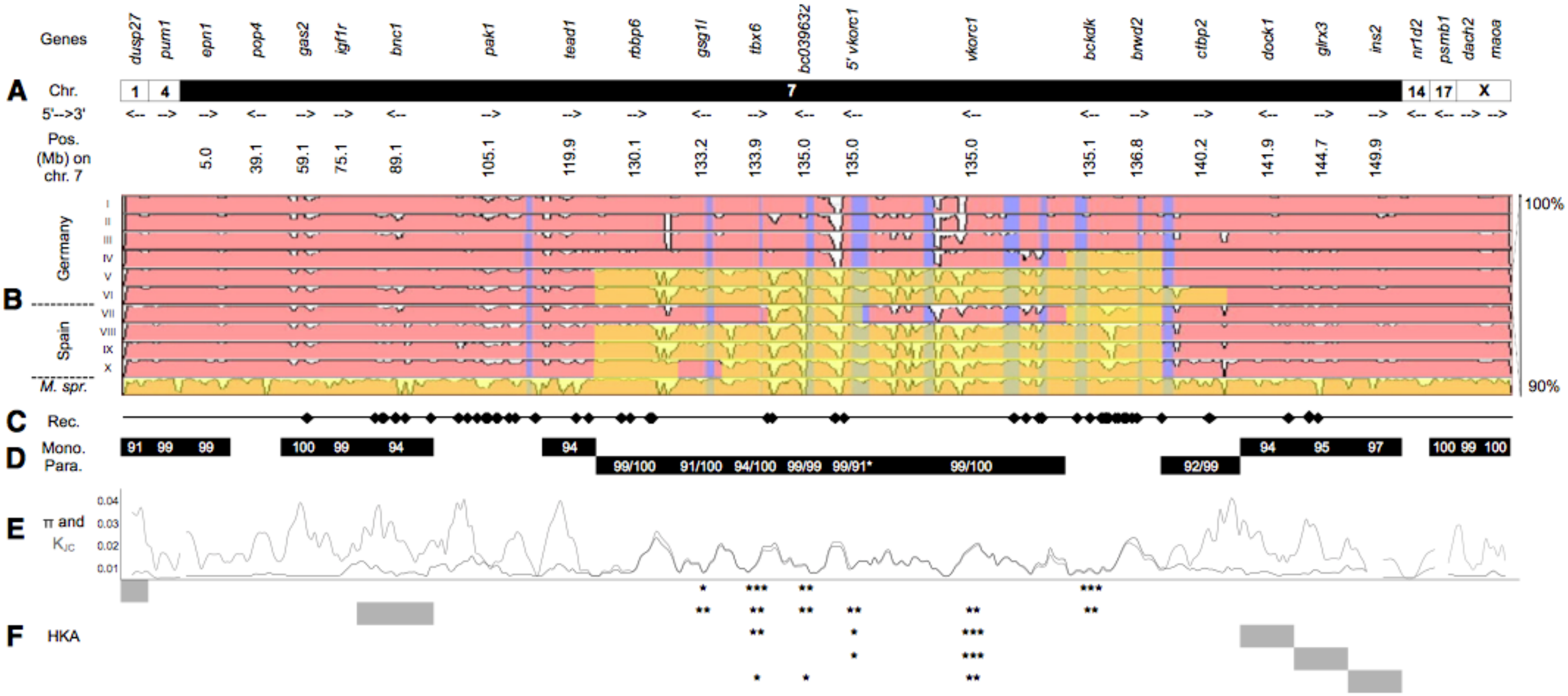


Figure 2. Genome Profiling of Ten *M. m. domesticus* from Germany and Spain

(A) Coverage of genes, their transcript orientation, and their chromosomal positions (in megabases) (see Table S2 for gene and PCR/sequencing primer information).

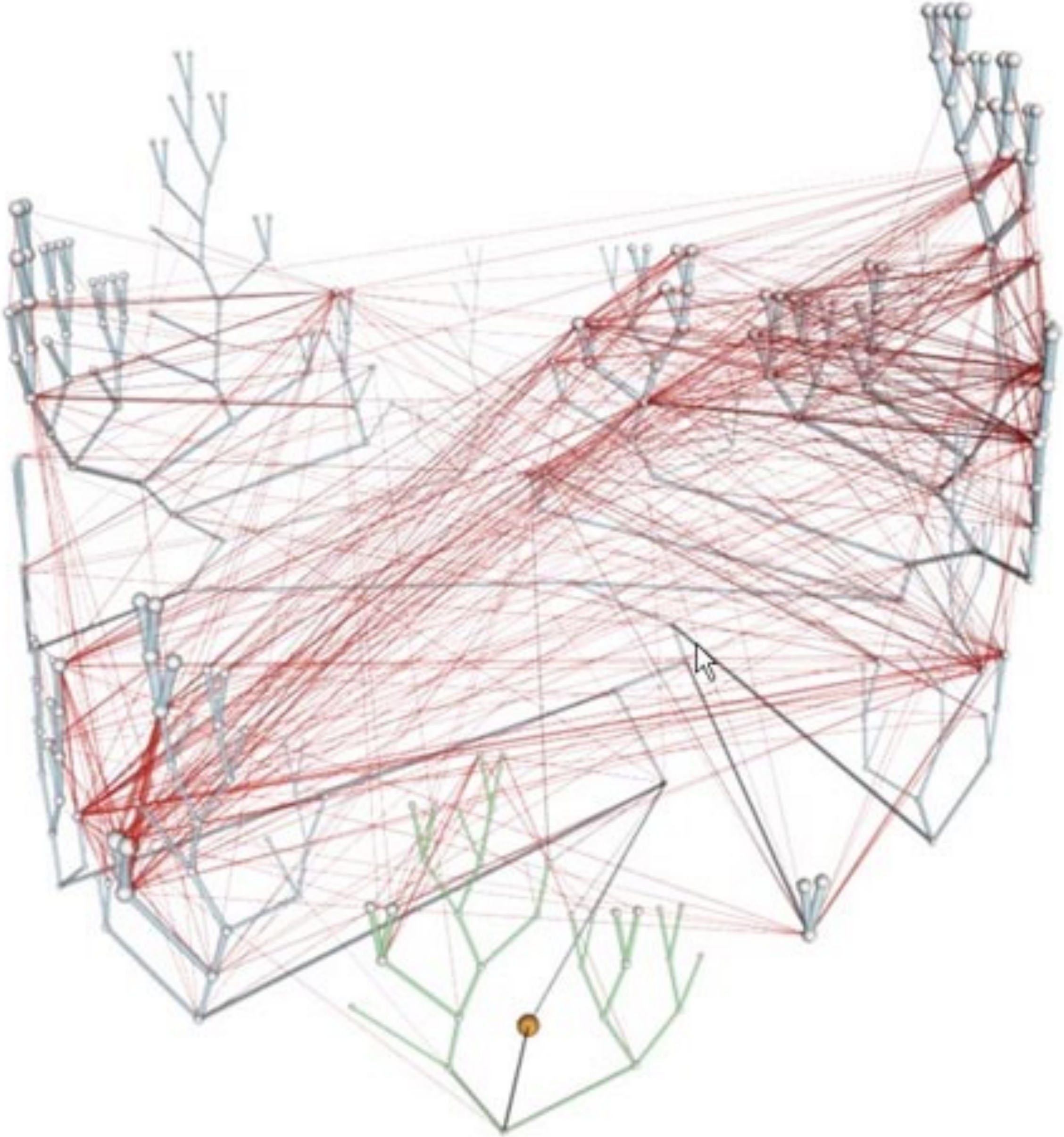
(B) VISTA plot depicting pairwise DNA sequence similarity scores (y axes, right, scaled between 90% and 100%) between C57BL/6J and six *M. m. domesticus* from Germany (genome profiles I–VI) and four *M. m. domesticus* from Spain (genome profiles VII–X). Exons are shown in purple; the coloring scheme is as in Figure 1 indicating, at a coarse resolution, regions comprised of predominantly *M. m. domesticus* sequences (pink) and *M. spretus* (*M. spr.*) sequences (yellow).

(C) Minimum number of recombination events (black diamonds) within chromosome 7 among *M. m. domesticus* (excluding *M. spretus* and C57BL/6J). See also the analysis of linkage disequilibrium in Figure S1B.

(D) Gene genealogies of *M. m. domesticus* identified as monophyletic (Mono.) or paraphyletic (Para.) with respect to *M. spretus* using 90% support for nodes as cutoff (Figures S1C and S1D). Significance of topologies is given in percent bootstrap values supporting monophyly of *M. m. domesticus* samples (top) or both clusters in paraphyletic topologies (bottom; first number *M. m. domesticus*, second number *M. spretus*). Asterisk indicates significance for *vkorc1* 5′ region taken from genealogy constructed using C57BL/6J as outgroup.

(E) Plot of polymorphism (expected hereozygosity; π) in *M. m. domesticus* relative to divergence (Jukes Cantor corrected K) to *M. spretus*.

(F) Asterisks mark significance (at α = 0.05, 0.01, and 0.001) of rejection of Hudson-Kreitman-Aguade (HKA) testing performed on select nonrecombining segments representing reference genes (gray boxes; see A for gene identifiers).



A bird's-eye view of the tree of life, showing the vines in red and the tree's branches in grey [Bacteria] and green [Archaea]. The last universal common ancestor is shown as a yellow sphere.

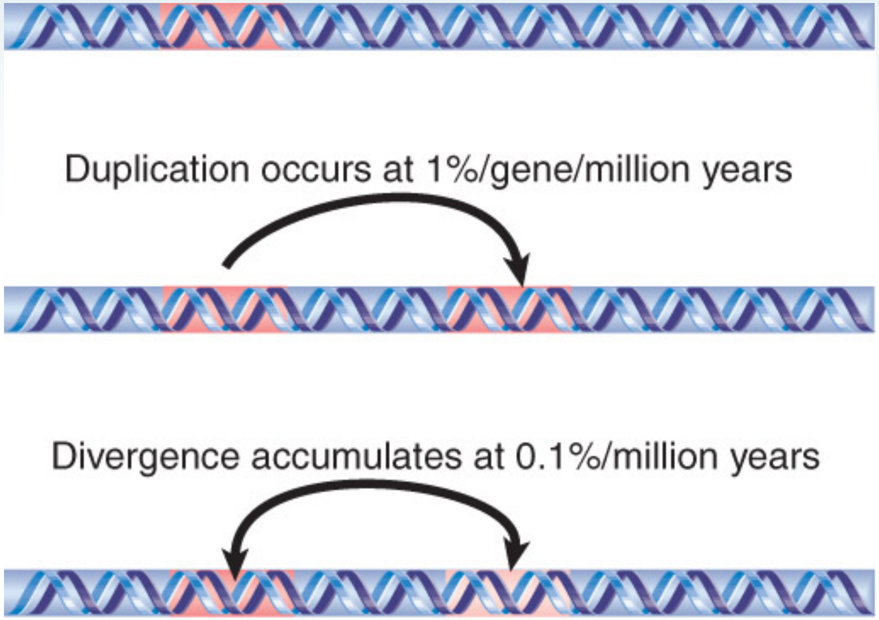
Gene Duplication Provides a Major Force in Evolution change of the different genomes

- Most of the genes that are unique to vertebrates are concerned with the immune or nervous systems.

- Duplicated genes may diverge to generate different genes, or one copy may become an inactive or *pseudogene*.

- …"nothing in evolution makes sense except in the light of the genome and development".

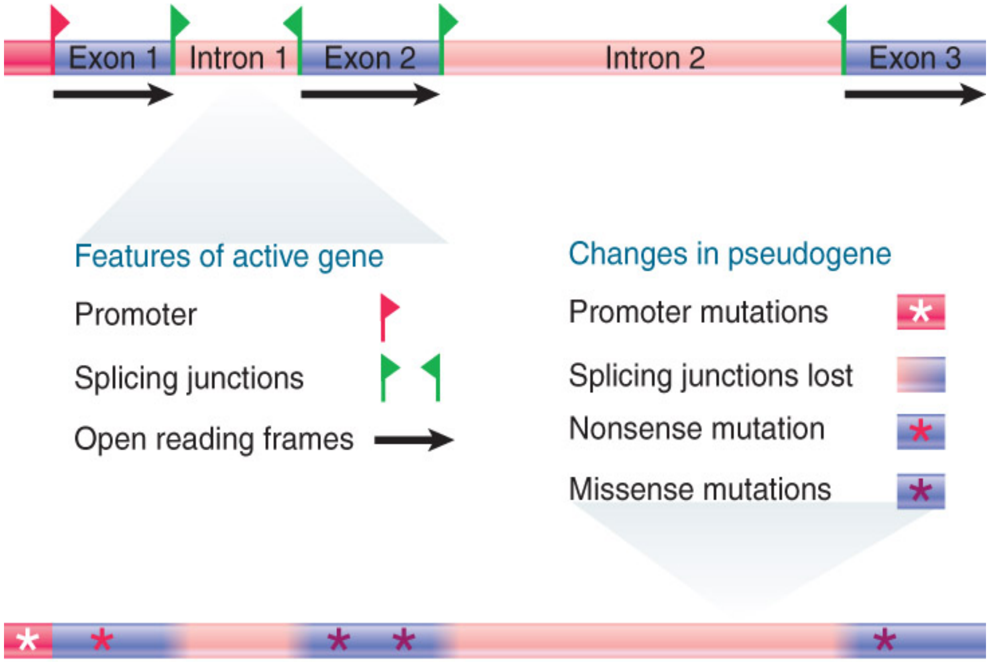
Gene Duplication is a Major Force in Genome Evolution


Duplication occurs at 1%/gene/million years
Divergence accumulates at 0.1%/million years

After a globin gene has been duplicated, differences may accumulate between the copies

Pseudogenes Are Nonfunctional Gene Copies

- **Processed pseudogenes** result from reverse transcription and integration of mRNA transcripts.
- **Nonprocessed pseudogenes** result from incomplete duplication or second-copy mutation of functional genes.
- Some pseudogenes:
- may gain functions
- different from those
- of their parent genes,
- such as regulation of
- gene expression, and
- take on different names.

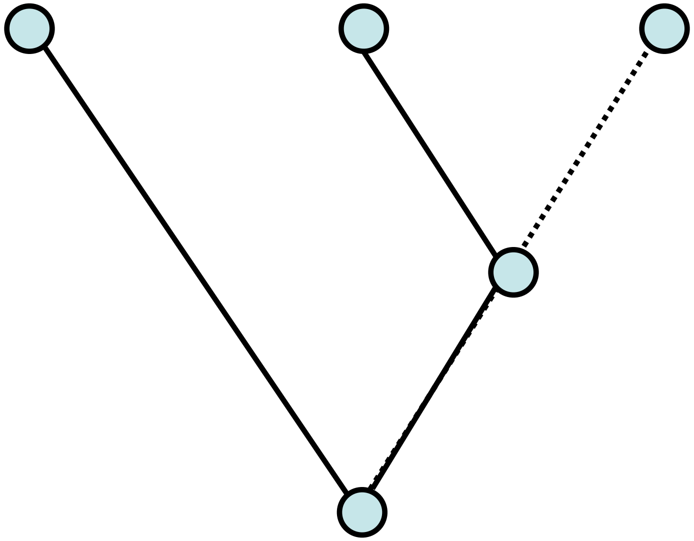


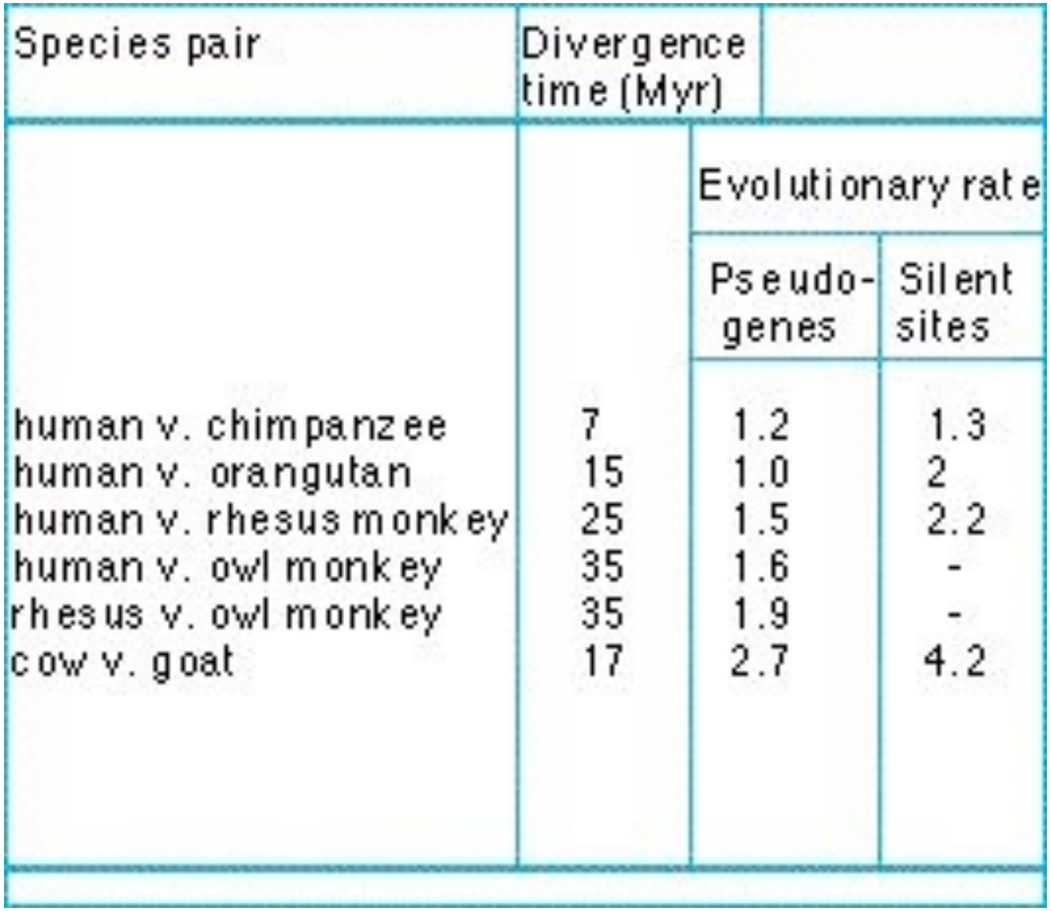
eg. Many changes have occurred in a beta-globin gene since it first became a pseudogene through duplication

sequence1
(functional)

sequence 2
(functional)

sequence 3
(pseudogenes)





Species pair	Divergence time (Myr)	Evolutionary rate	
		Pseudo-genes	Silent sites
human v. chimpanzee	7	1.2	1.3
human v. orangutan	15	1.0	2
human v. rhesus monkey	25	1.5	2.2
human v. owl monkey	35	1.6	-
rhesus v. owl monkey	35	1.9	-
cow v. goat	17	2.7	4.2

Curiously, **pseudogenes** evolve at about the same rate as **silent base changes.** Rates are expressed in numbers of base changes per 109 years. The comparisons are for various genes and pseudogenes in the globin gene family.

Simplified from Li, Tanimura & Sharp (1987)

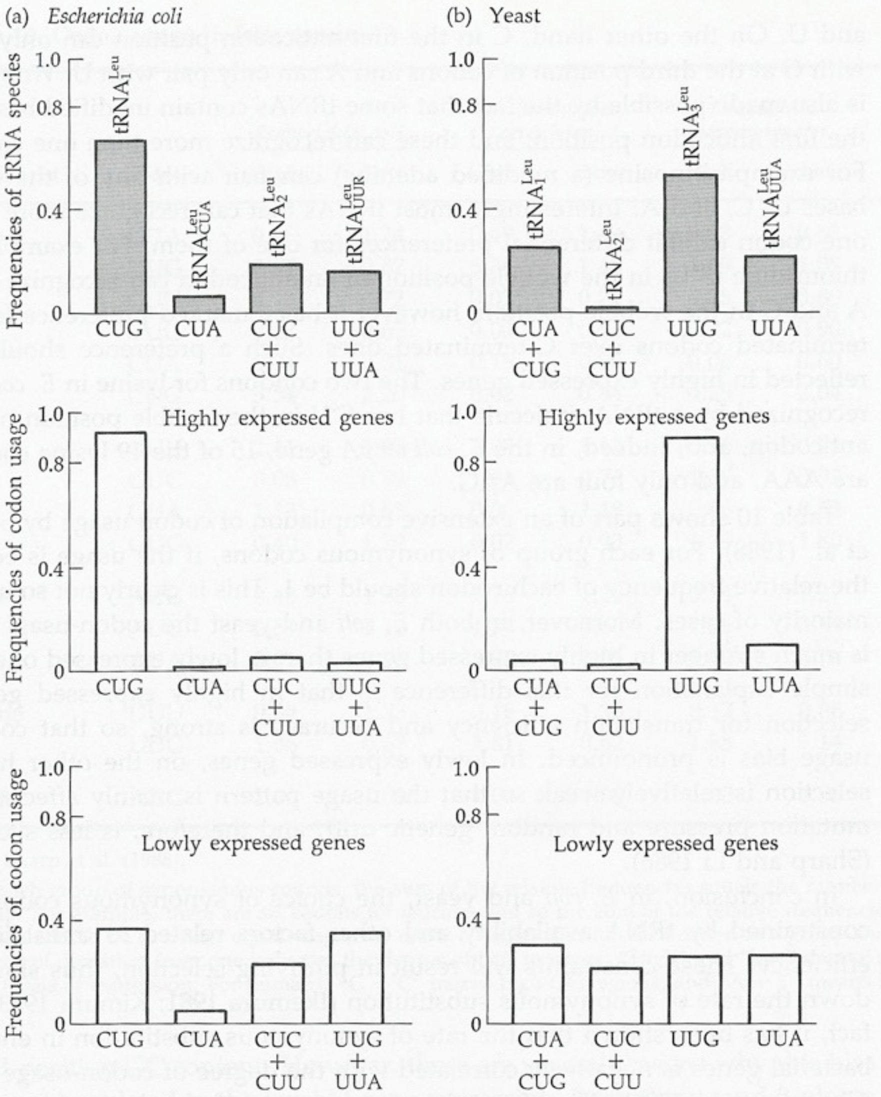
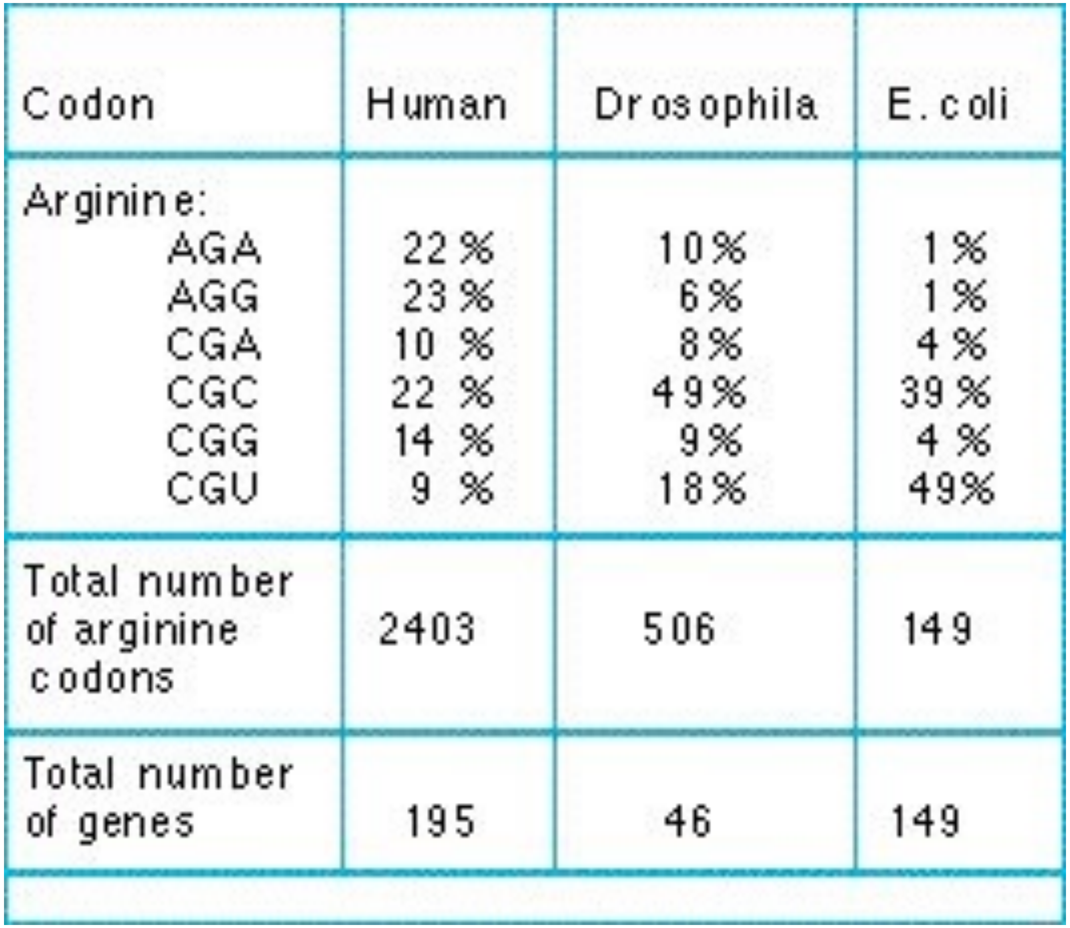


Figure 8. Diagram illustrating the relationship between the relative frequency of codon usage for leucine (open bars) and the relative abundance of the corresponding cognate tRNA species (solid bars) in (a) *Escherichia coli* and (b) *Sacharomyces cerevisiae*. The plus signs (e.g., between codons CUC and CUU for *E. coli*) indicate that each of these pairs of codons is recognized by a single tRNA species (e.g., tRNA$_2^{Leu}$ for CUC and CUU in *E. coli*).



Codon	Human	Drosophila	E. coli
Arginine:			
AGA	22 %	10%	1 %
AGG	23 %	6%	1 %
CGA	10 %	8%	4 %
CGC	22 %	49%	39 %
CGG	14 %	9%	4 %
CGU	9 %	18%	49%
Total number of arginine codons	2403	506	149
Total number of genes	195	46	149

Frequencies of six arginine codons in the DNA of three species.

The table gives the percentages of arginine amino acids that are encoded by each of the six codons in various numbers of genes in species.

Simplified from Grantham, Perrin & Mouchiroud (1986).

A Maximum Likelihood Method for Analyzing Pseudogene Evolution: Implications for Silent Site Evolution in Humans and Rodents

1. **Carlos D. Bustamante, Rasmus Nielsen and Daniel L. Hartl**

+

Author Affiliations

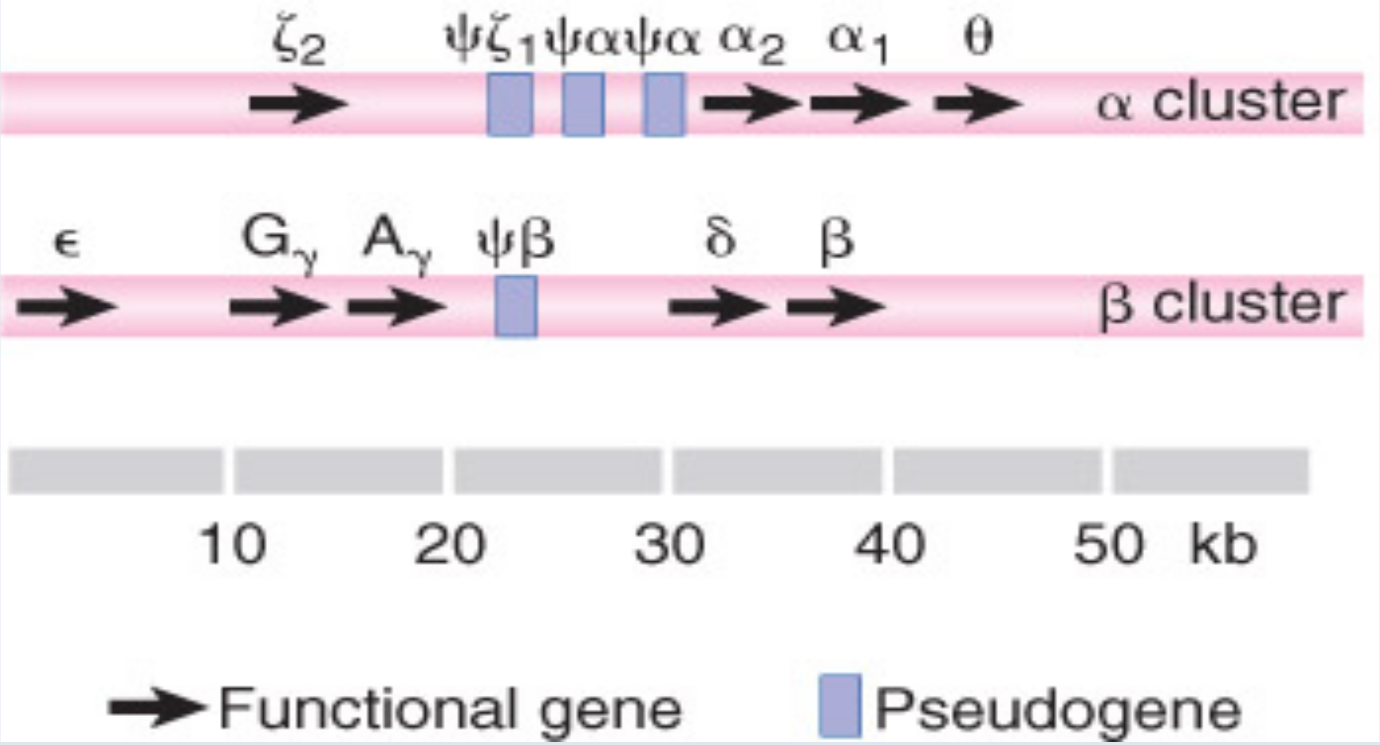
We present a new likelihood method for detecting constrained evolution at synonymous sites and other forms of nonneutral evolution in putative pseudogenes. The model is applicable whenever the DNA sequence is available from a protein-coding functional gene, a pseudogene derived from the protein-coding gene, and an orthologous functional copy of the gene. Two nested likelihood ratio tests are developed to test the hypotheses that (1) the putative pseudogene has equal rates of silent and replacement substitutions; and (2) the rate of synonymous substitution in the functional gene equals the rate of substitution in the pseudogene. The method is applied to a data set containing 74 human processed-pseudogene loci, 25 mouse processed-pseudogene loci, and 22 rat processed-pseudogene loci. Using the informatics resources of the Human Genome Project, we localized 67 of the human-pseudogene pairs in the genome and estimated the GC content of a large surrounding genomic region for each. **We find that, for pseudogenes deposited in GC regions similar to those of their paralogs, the assumption of equal rates of silent and replacement site evolution in the pseudogene is upheld;** in these cases, **the rate of silent site evolution in the functional genes is ~70% the rate of evolution in the pseudogene**. On the other hand, for pseudogenes located in genomic regions of much **lower GC than their functional gene, we see a sharp increase in the rate of silent site substitutions, leading to a large rate of rejection for the pseudogene equality likelihood ratio test.**

http://mbe.oxfordjournals.org/content/19/1/110.full 2002

Globin Clusters Are Formed by Duplication and Divergence

- All globin genes are descended from duplications and mutations from an ancestral gene that had three exons.
- …”nothing in evolution makes sense except in the light of the genome **and development”**.

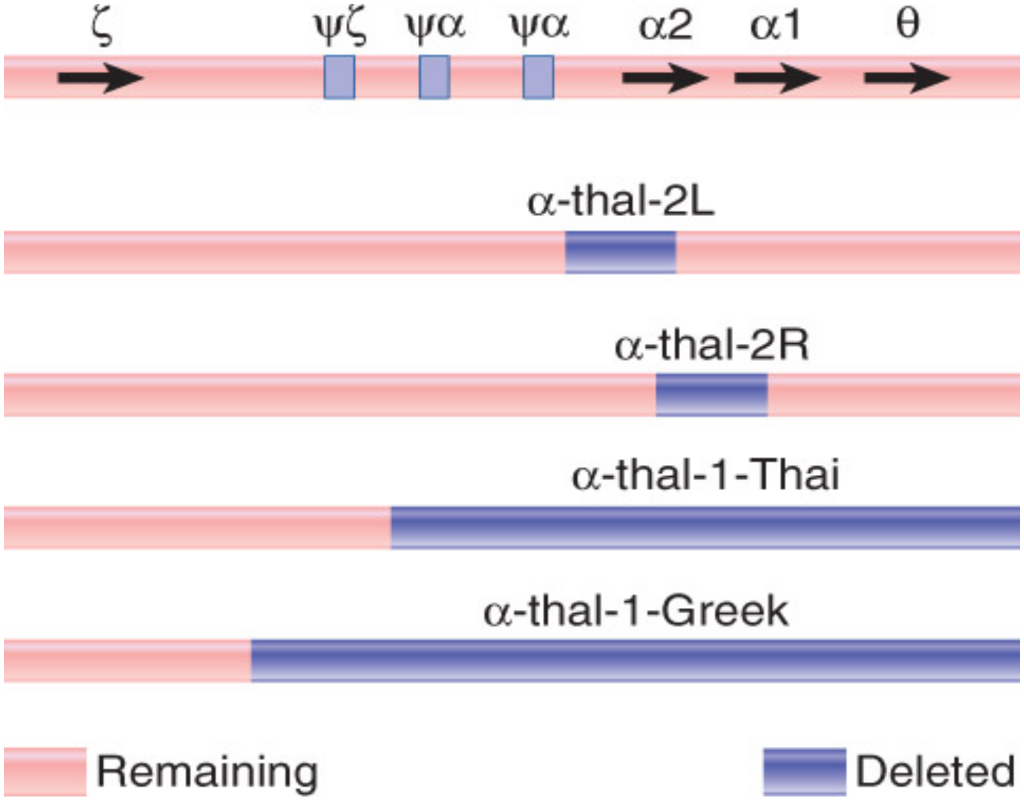
Globin Clusters Are Formed by Duplication followed by Divergence



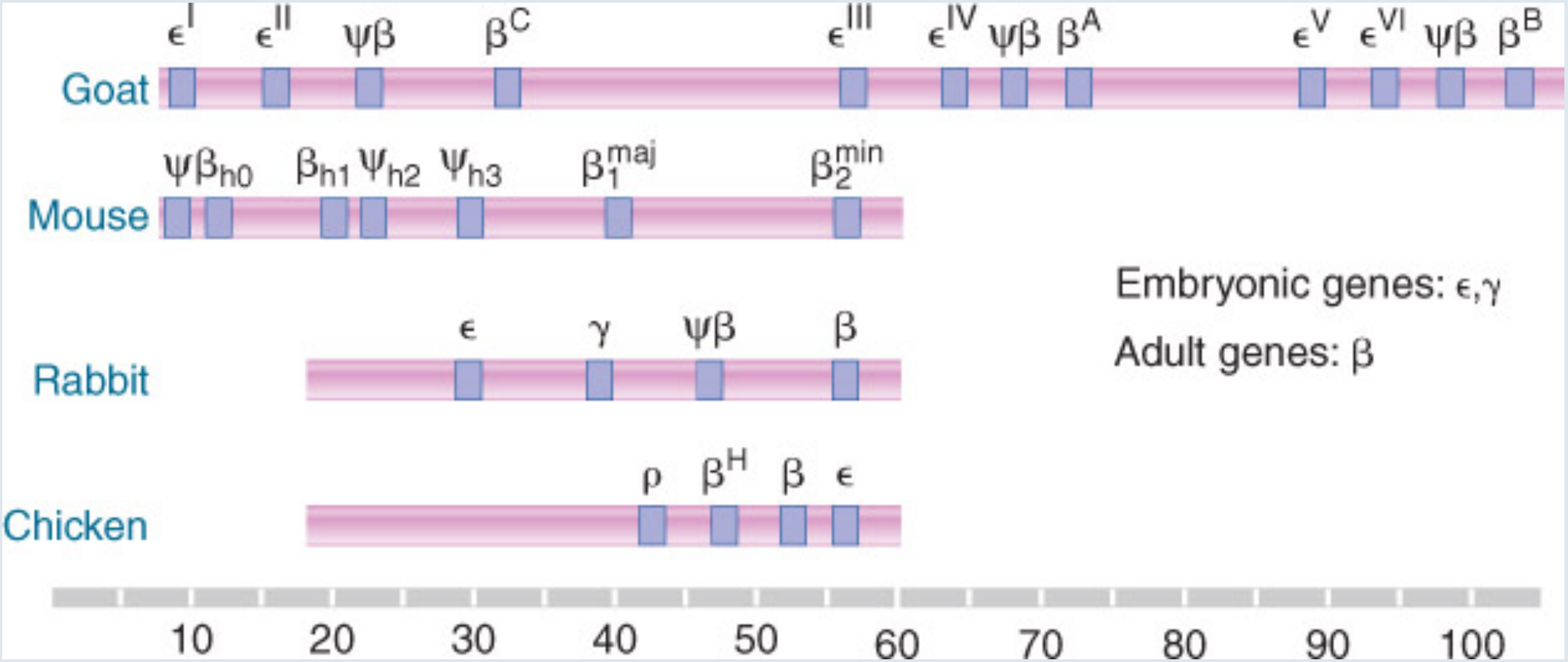
Each of the α-like and β-like globin gene families is organized into a single cluster, which includes functional genes and pseudogenes.

- All globin genes are descended from duplications and mutations from an ancestral gene that had three exons.

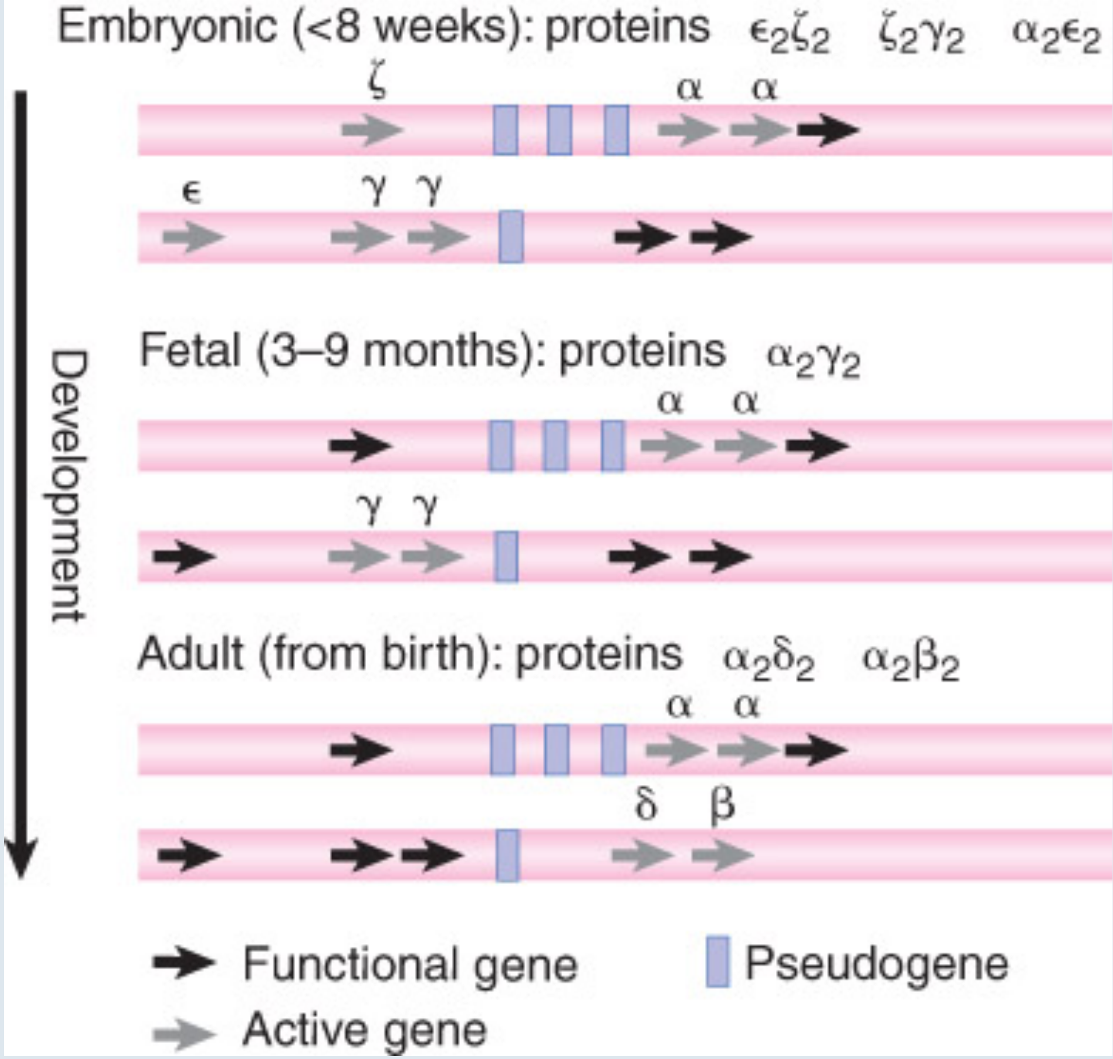
- Different **thalassaemias** are caused by various deletions that eliminate α- or β-globin genes.
  - The severity of the disease depends on the individual deletion.



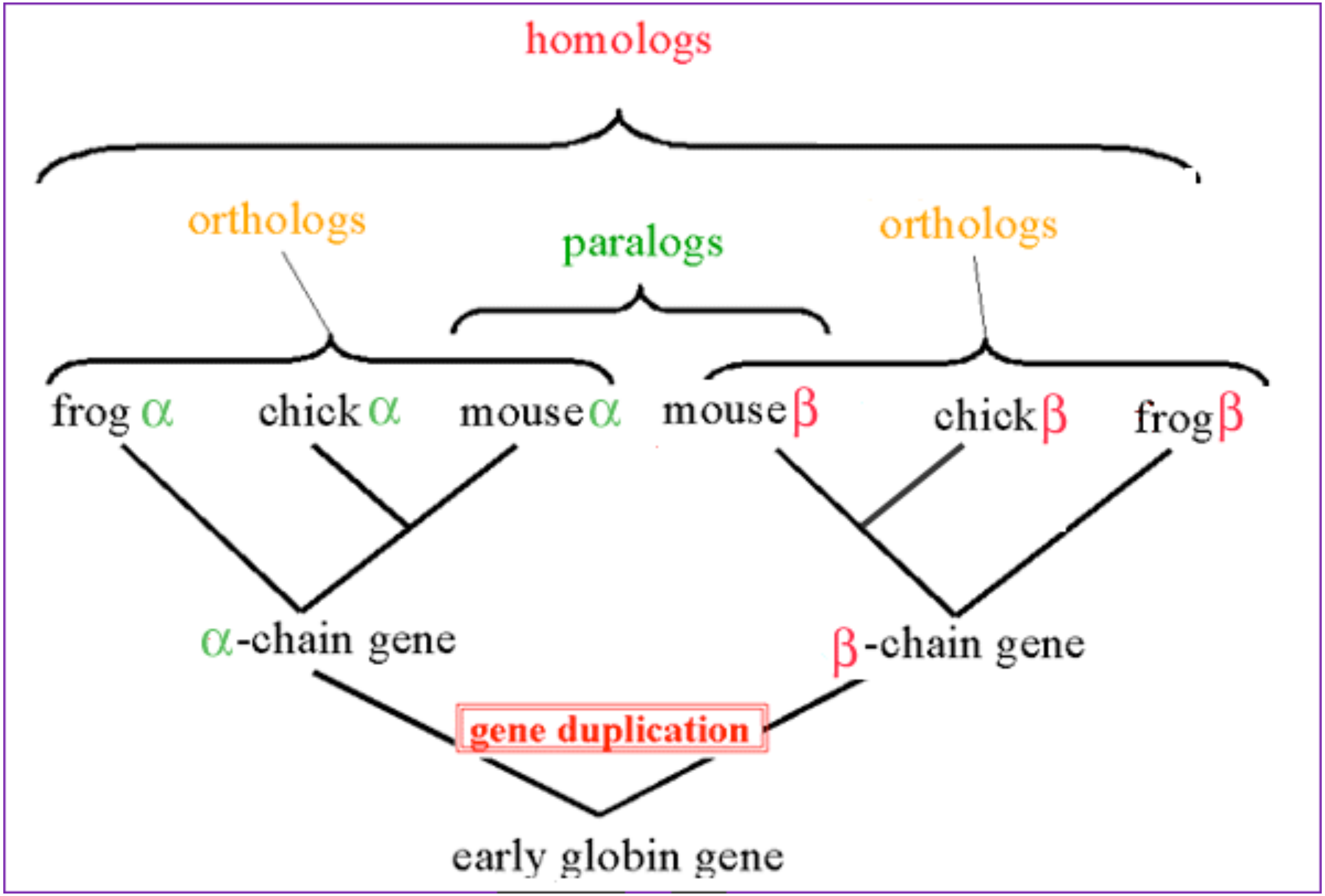
α thalassaemias result from various deletions in the a-globin gene cluster



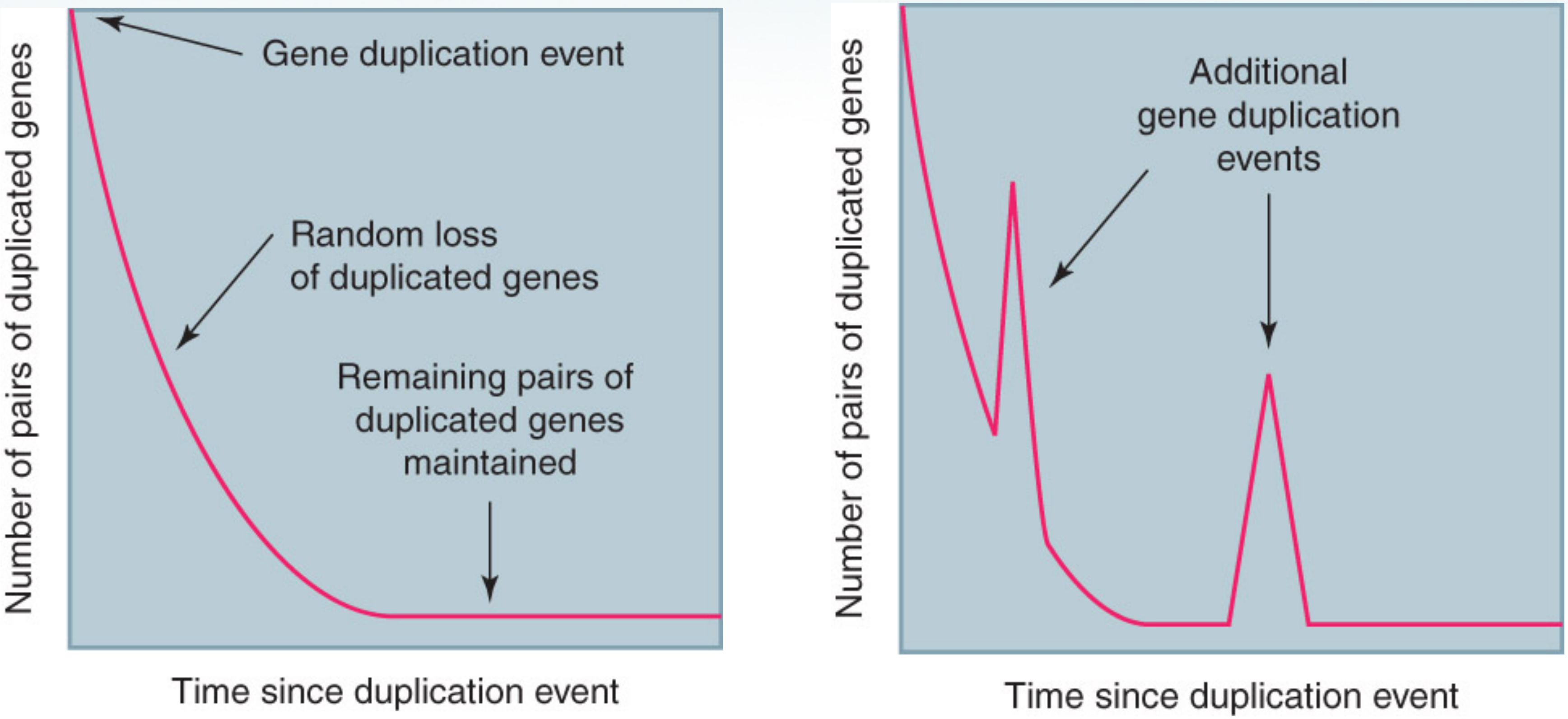
Some of the clusters of β-globin genes and pseudogenes that are found in vertebrates.


Embryonic (<8 weeks): proteins $\epsilon_2\zeta_2$ $\zeta_2\gamma_2$ $\alpha_2\epsilon_2$
ζ
α α
ϵ
γ γ
Development
Fetal (3–9 months): proteins $\alpha_2\gamma_2$
α α
γ γ
Adult (from birth): proteins $\alpha_2\delta_2$ $\alpha_2\beta_2$
α α
δ β
Functional gene
Pseudogene
Active gene

Different hemoglobin genes are expressed during embryonic, fetal, and adult periods of human development.

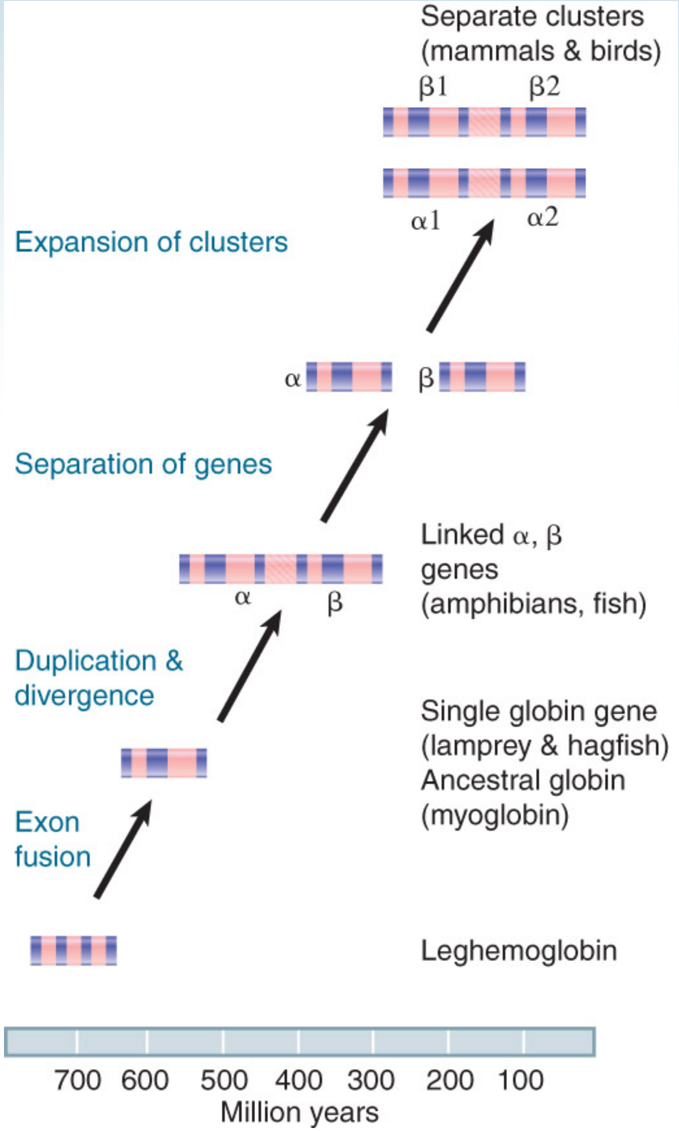

homologs
orthologs
paralogs
orthologs
frog α
chick α
mouse α
mouse β
chick β
frog β
α-chain gene
β-chain gene
gene duplication
early globin gene

Genome Duplication Has Potentially Played a Role in…..Bacterial, Plant and Vertebrate Evolution



Gene and genome duplication

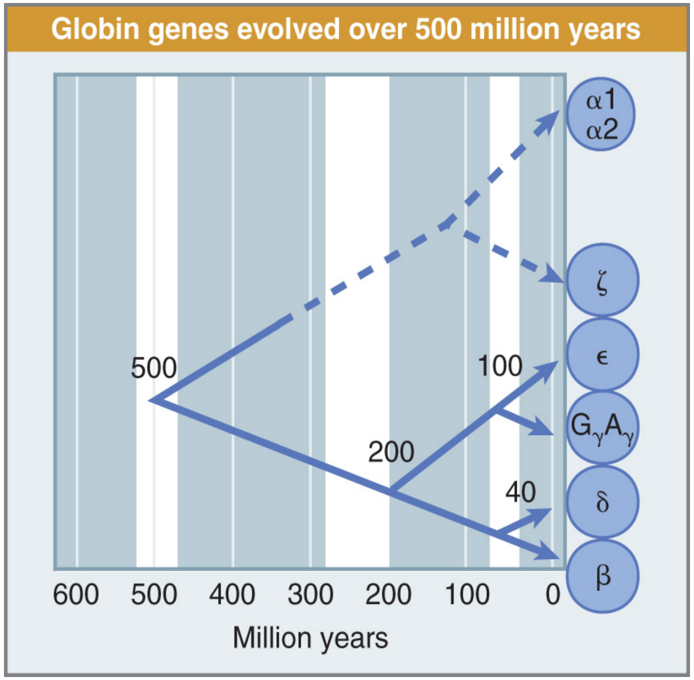
Adapted from G. Blanc and K. H. Wolfe, Plant Cell 16 (2004): 1667-1678.



All globin genes appear to have evolved by a series of duplications, transpositions, and subsequent mutations -from a single ancestral gene

- The evolutionary divergence between two proteins can be measured by:
  - The percent of positions at which the corresponding amino acids differ.

- Mutations accumulate at a “more or less” clock like rate AFTER genomes diverge and then separate.
  - The divergence between any pair of g**lobin sequences** is approximately proportional to the time since their genes separated.

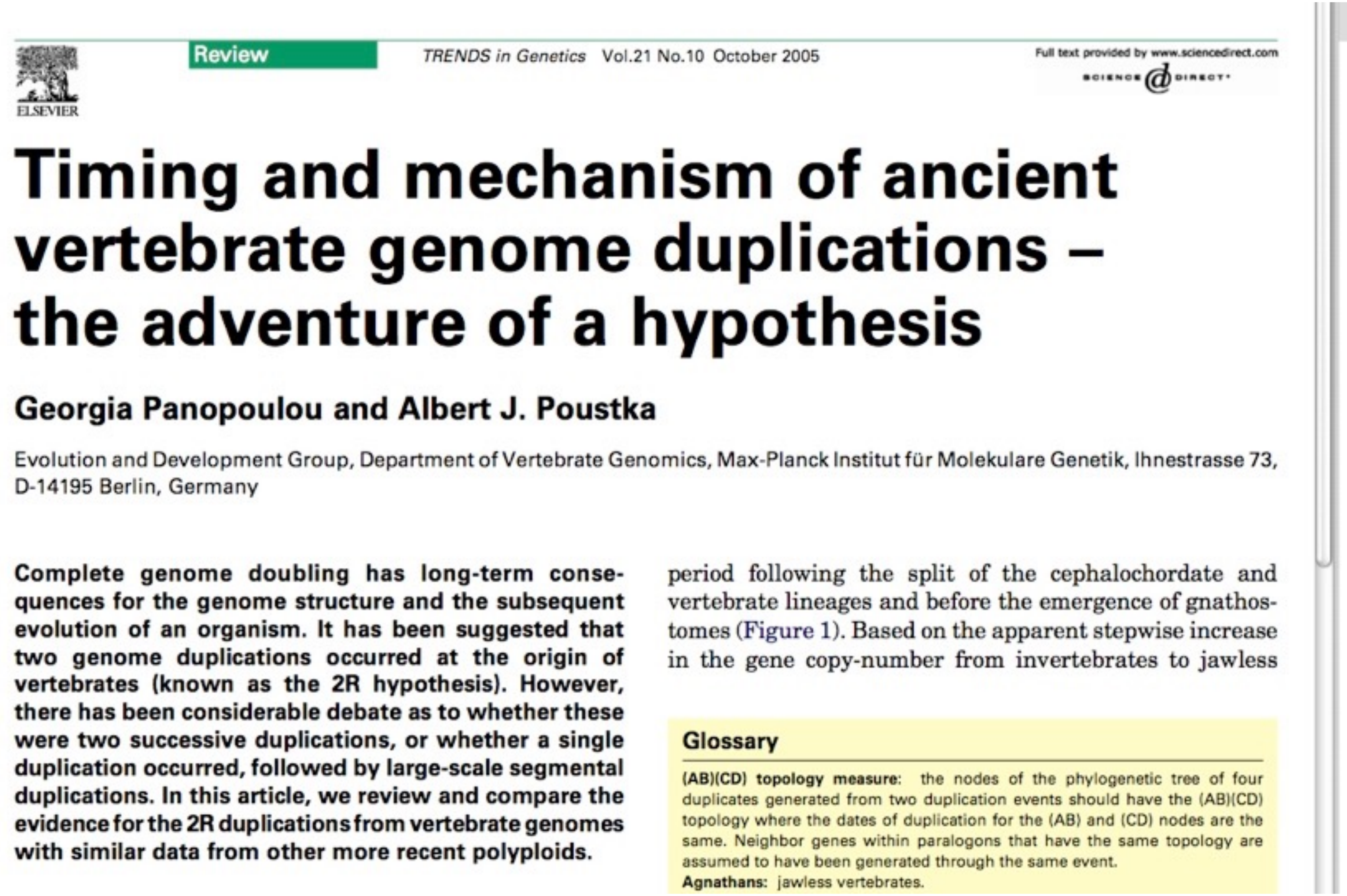


Genome Duplication Has Played a Role in Plant and Vertebrate Evolution

- Genome duplication events can be obscured by the evolution and/or loss of duplicates as well as by chromosome rearrangements.

- Genome duplication has been detected in the evolutionary history of many flowering plants and of vertebrate animals.

- **2R hypothesis –** The hypothesis that the early vertebrate genome has actually undergone at least **two rounds** of duplication.



Review TRENDS in Genetics Vol.21 No.10 October 2005

Full text provided by www.sciencedirect.com

Timing and mechanism of ancient vertebrate genome duplications – the adventure of a hypothesis

Georgia Panopoulou and Albert J. Poustka

Evolution and Development Group, Department of Vertebrate Genomics, Max-Planck Institut für Molekulare Genetik, Ihnestrasse 73, D-14195 Berlin, Germany

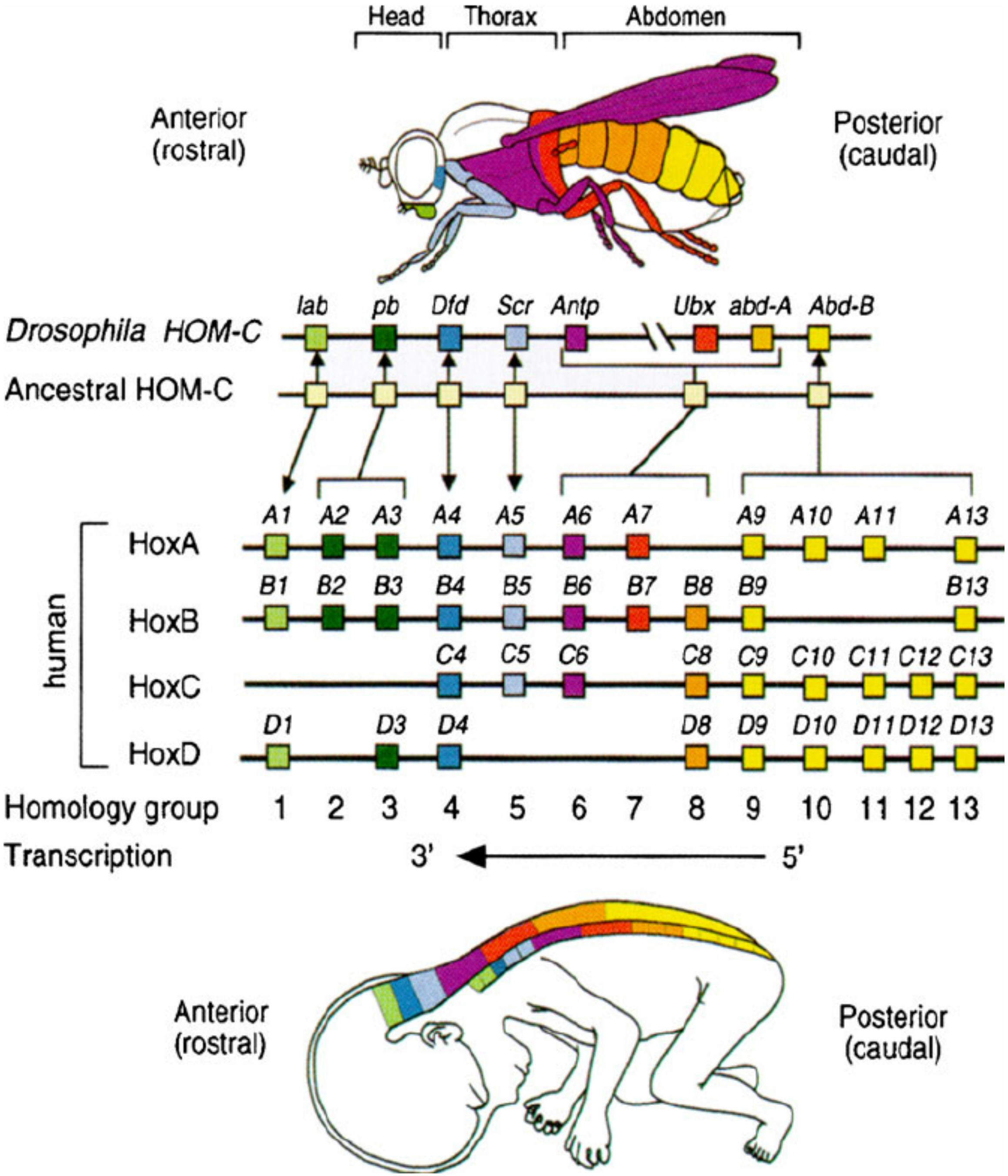
Complete genome doubling has long-term consequences for the genome structure and the subsequent evolution of an organism. It has been suggested that two genome duplications occurred at the origin of vertebrates (known as the 2R hypothesis). However, there has been considerable debate as to whether these were two successive duplications, or whether a single duplication occurred, followed by large-scale segmental duplications. In this article, we review and compare the evidence for the 2R duplications from vertebrate genomes with similar data from other more recent polyploids.

period following the split of the cephalochordate and vertebrate lineages and before the emergence of gnathostomes (Figure 1). Based on the apparent stepwise increase in the gene copy-number from invertebrates to jawless

Glossary

(AB)(CD) topology measure: the nodes of the phylogenetic tree of four duplicates generated from two duplication events should have the (AB)(CD) topology where the dates of duplication for the (AB) and (CD) nodes are the same. Neighbor genes within paralogons that have the same topology are assumed to have been generated through the same event.

Agnathans: jawless vertebrates.


Head
Thorax
Abdomen
Anterior (rostral)
Posterior (caudal)
lab
pb
Dfd
Scr
Antp
Ubx
abd-A
Abd-B
Drosophila HOM-C
Ancestral HOM-C
A1
A2
A3
A4
A5
A6
A7
A9
A10
A11
A13
HoxA
B1
B2
B3
B4
B5
B6
B7
B8
B9
B13
HoxB
human
C4
C5
C6
C8
C9
C10
C11
C12
C13
HoxC
D1
D3
D4
D8
D9
D10
D11
D12
D13
HoxD
Homology group
1
2
3
4
5
6
7
8
9
10
11
12
13
Transcription
3'
5'
Anterior (rostral)
Posterior (caudal)

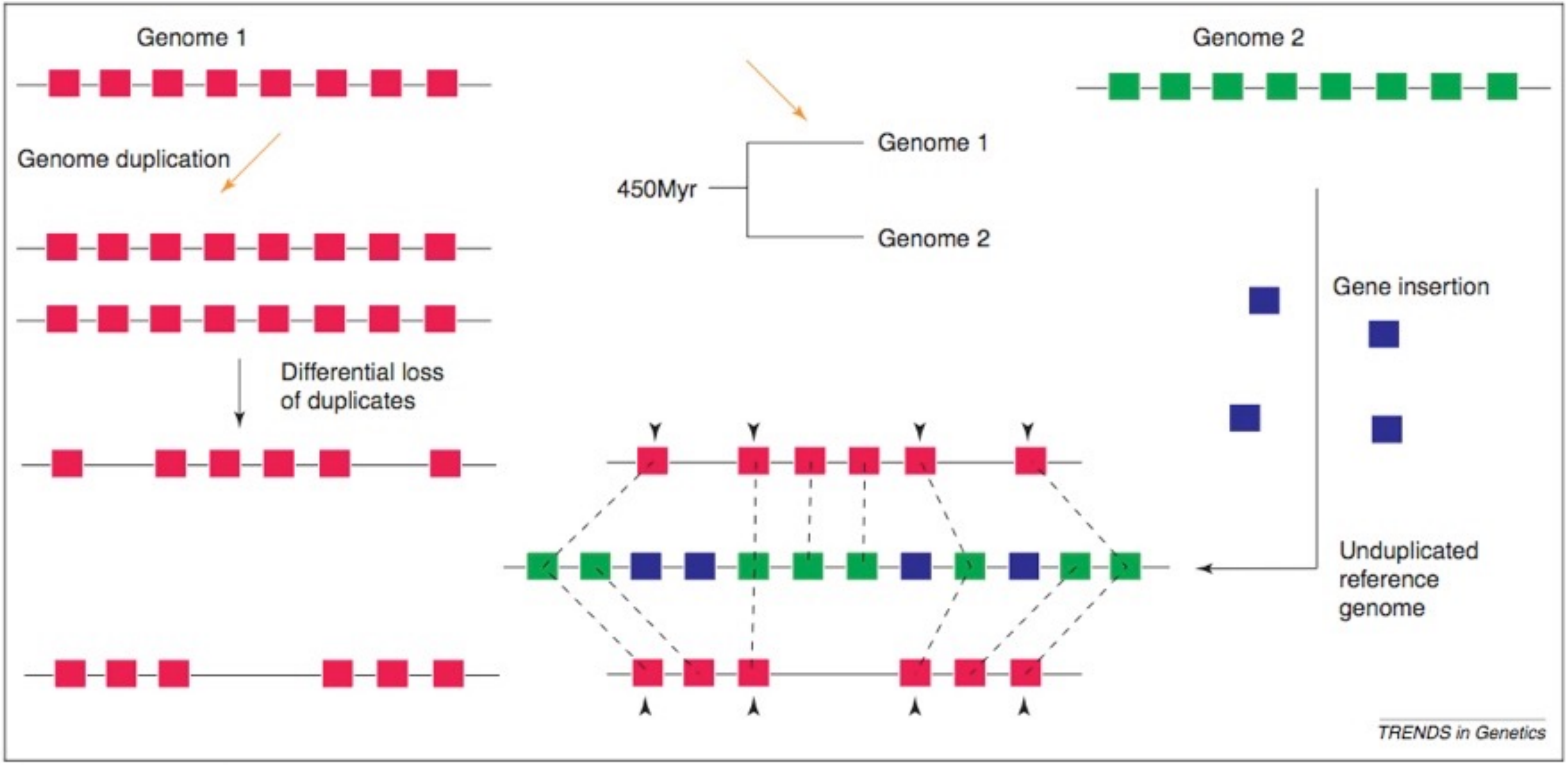


Figure 4. Illustration of the comparative approach used to prove genome duplications in yeast and *Tetraodon*. Genome 1 undergoes a Genome duplication (e.g. *Tetraodon*) creating two identical sets of chromosomes and genes followed by gene loss (left side). Genome 2 (e.g. human) experiences only some gene insertions and serves as 'unduplicated' reference genome. In most cases, large regions of 'double conserved synteny' can be identified (i.e. every chromosome of Genome 2 maps to two chromosomes of Genome 1 in an interleaving pattern; (middle lower panel). Genes that have been retained in two copies (arrowheads) would function as anchor points to identify a paralogon. The approach has been shown to be effective in detecting 'double conserved segments' in a genome that has undergone a WGD around 200–300 Mya and it has separated from its reference genome ∼450 Mya.

Genome Duplication Has Played a Role in Plant and Vertebrate Evolution

….more so in plants

- Genome duplication occurs when **polyploidization** increases the chromosome number by multiples of… **TWO**.

- **autopolyploidy –** Polyploidization resulting from mitotic or meiotic errors within a species.

- **allopolyploidy –** Polyploidization resulting from hybridization between two different but reproductively compatible species.

Genome Duplication Has Played a Role in Plant and Vertebrate Evolution….more so in plants

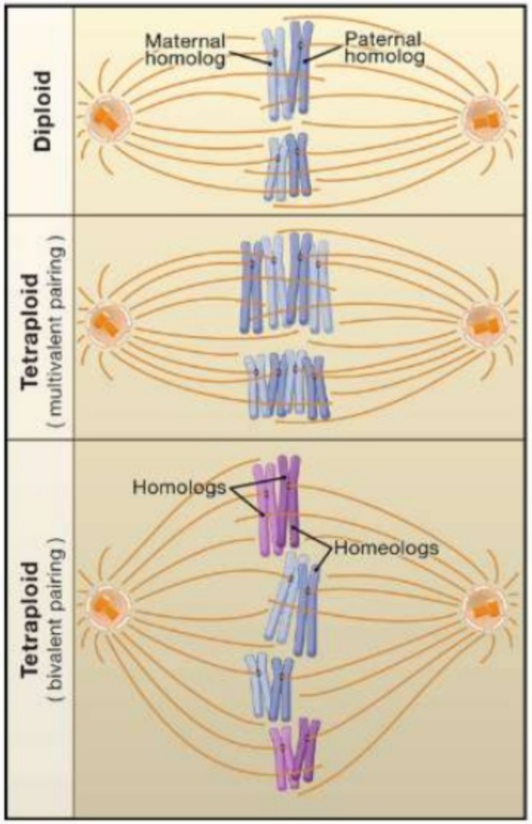
Autopolyploids typically have multivalent pairing

- chromosomes are more or less identical

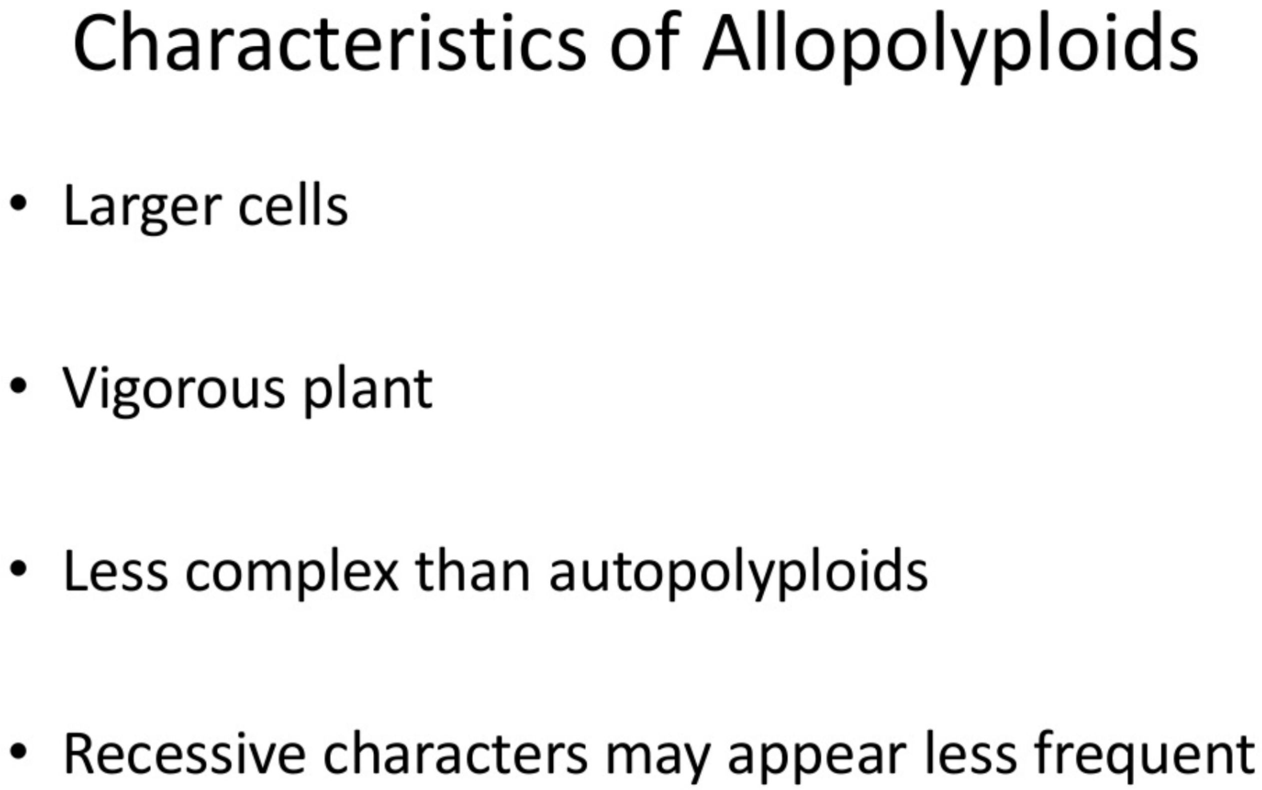
Allopolyploids are variable

- bivalent pairing with more genetic divergence

- multivalent pairing when closely related



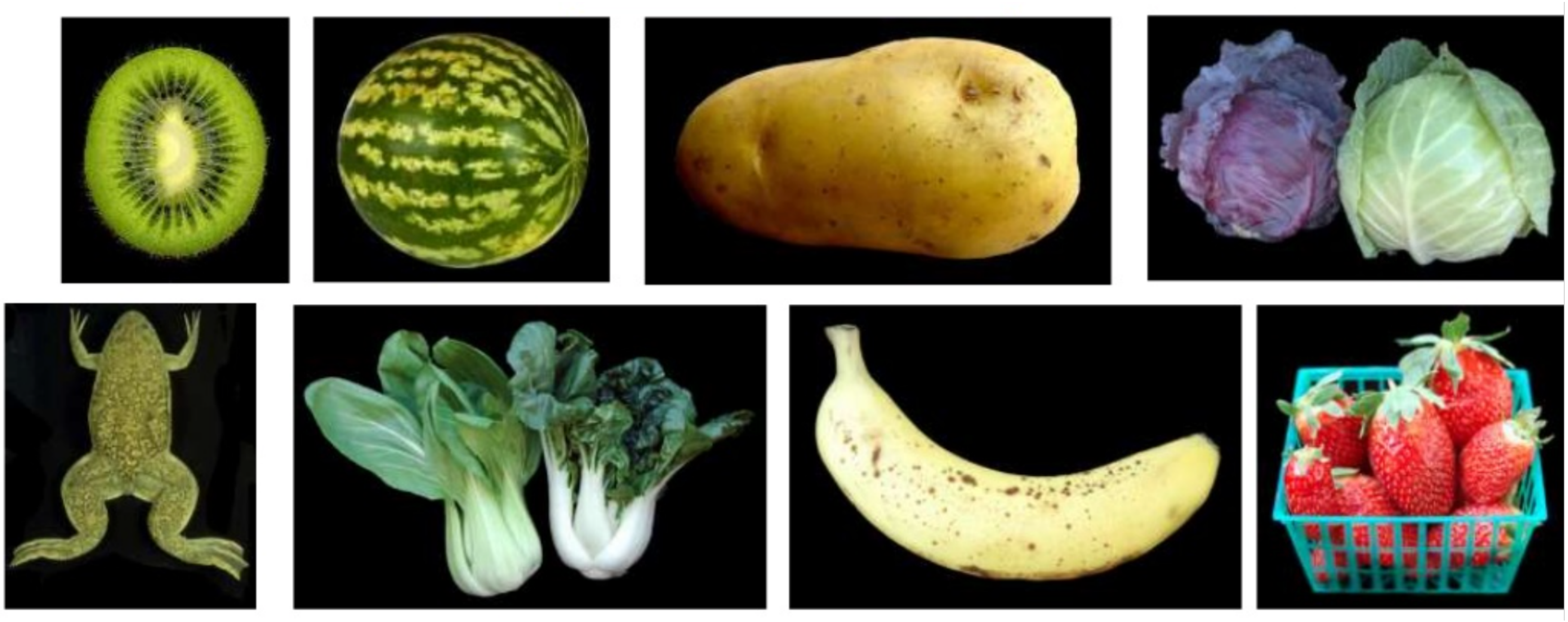
Allopolyploidy – Polyploidization resulting from hybridization between **two different but reproductively compatible species**.



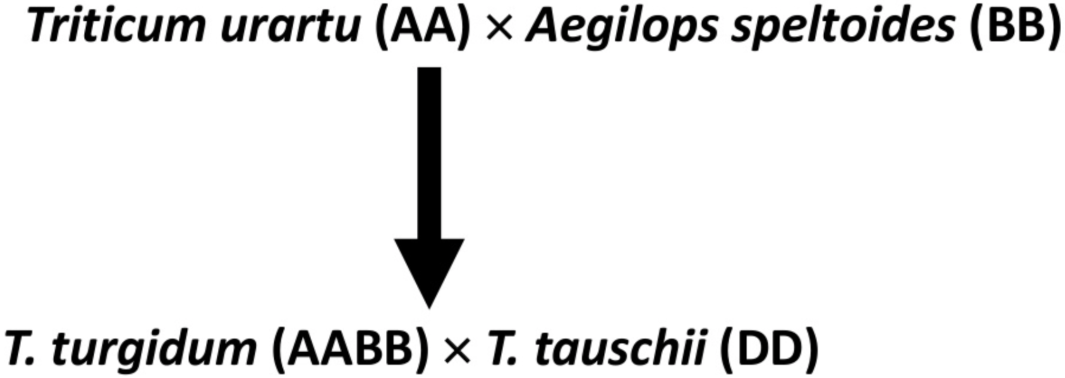
Characteristics of Allopolyploids

- Larger cells
- Vigorous plant
- Less complex than autopolyploids
- Recessive characters may appear less frequent

Allopolyploidy – Polyploidization resulting from hybridization between **two different but reproductively compatible species**.



Allopolyploidy – Polyploidization resulting from hybridization between **two different but reproductively compatible species**.



Triticum urartu (AA) × *Aegilops speltoides* (BB)

T. turgidum (AABB) × *T. tauschii* (DD)

The common bread wheat (*Triticum aestivum*) is an allohexaploid containing three distinct sets of chromosomes derived from three different diploid species of goat-grass (*Aegilops*) through a tetraploid intermediary (durum wheat).

T.astivum

AABBDD

Gene Duplication Provides a Major Force in Evolution CHANGE in different genomes

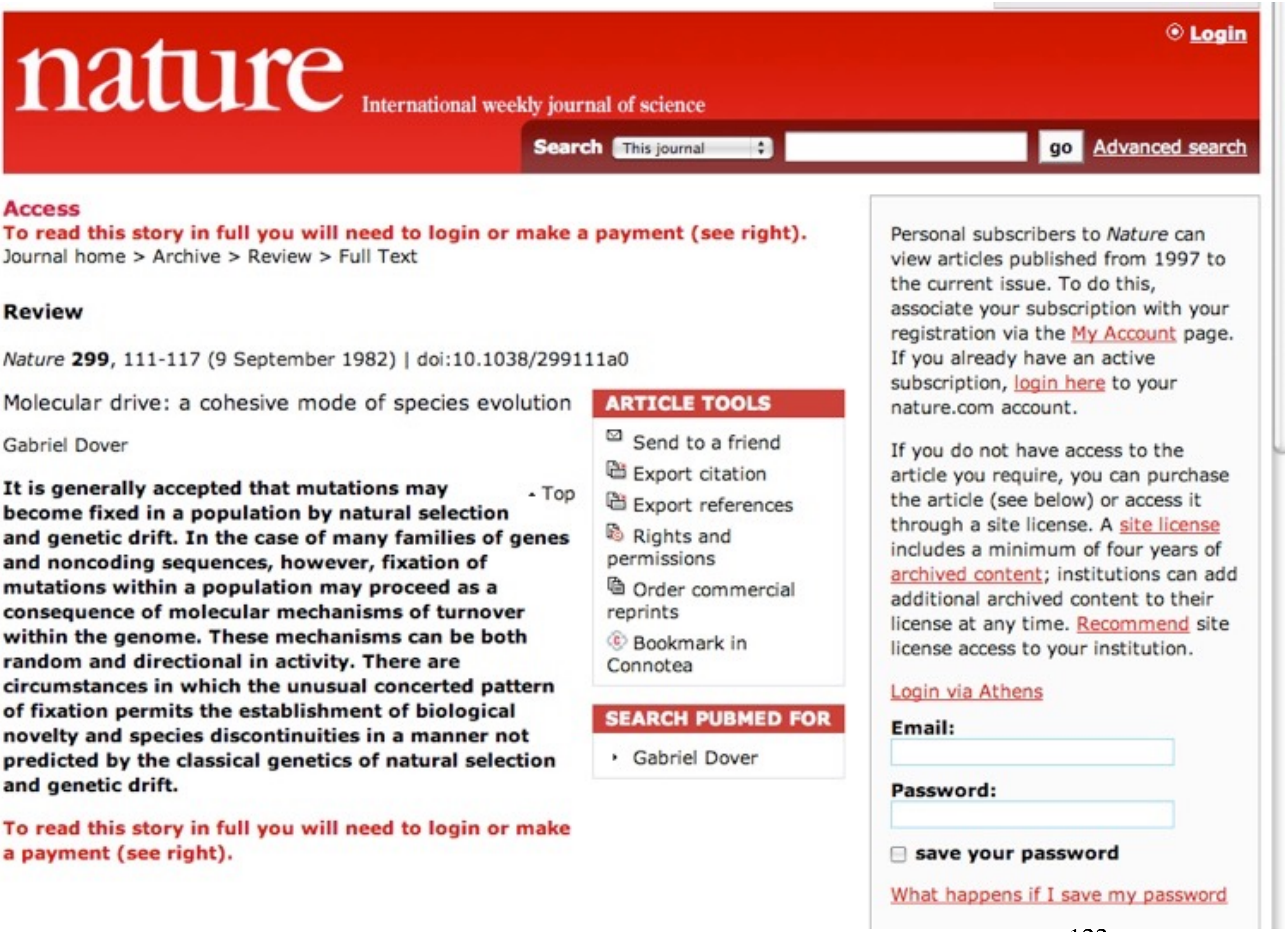
- Most of the genes that are unique to vertebrates are concerned with the immune or nervous systems.

- Duplicated genes may diverge to generate different genes, or one copy may become an inactive or *pseudogene*.

Gene Duplication Provides a Major Force in Evolution CONSTANCY within gene families

- Most of the genes that are unique to vertebrates are concerned with the immune or nervous systems.

- Duplicated genes may diverge yet converge with respect to their orthologues within gene families…



Access

To read this story in full you will need to login or make a payment (see right).

Journal home > Archive > Review > Full Text

Review

Nature **299**, 111-117 (9 September 1982) | doi:10.1038/299111a0

Molecular drive: a cohesive mode of species evolution

Gabriel Dover

It is generally accepted that mutations may become fixed in a population by natural selection and genetic drift. In the case of many families of genes and noncoding sequences, however, fixation of mutations within a population may proceed as a consequence of molecular mechanisms of turnover within the genome. These mechanisms can be both random and directional in activity. There are circumstances in which the unusual concerted pattern of fixation permits the establishment of biological novelty and species discontinuities in a manner not predicted by the classical genetics of natural selection and genetic drift.

To read this story in full you will need to login or make a payment (see right).

Personal subscribers to *Nature* can view articles published from 1997 to the current issue. To do this, associate your subscription with your registration via the My Account page. If you already have an active subscription, login here to your nature.com account.

If you do not have access to the article you require, you can purchase the article (see below) or access it through a site license. A site license includes a minimum of four years of archived content; institutions can add additional archived content to their additional site license at any time. Recommend site license access to your institution.

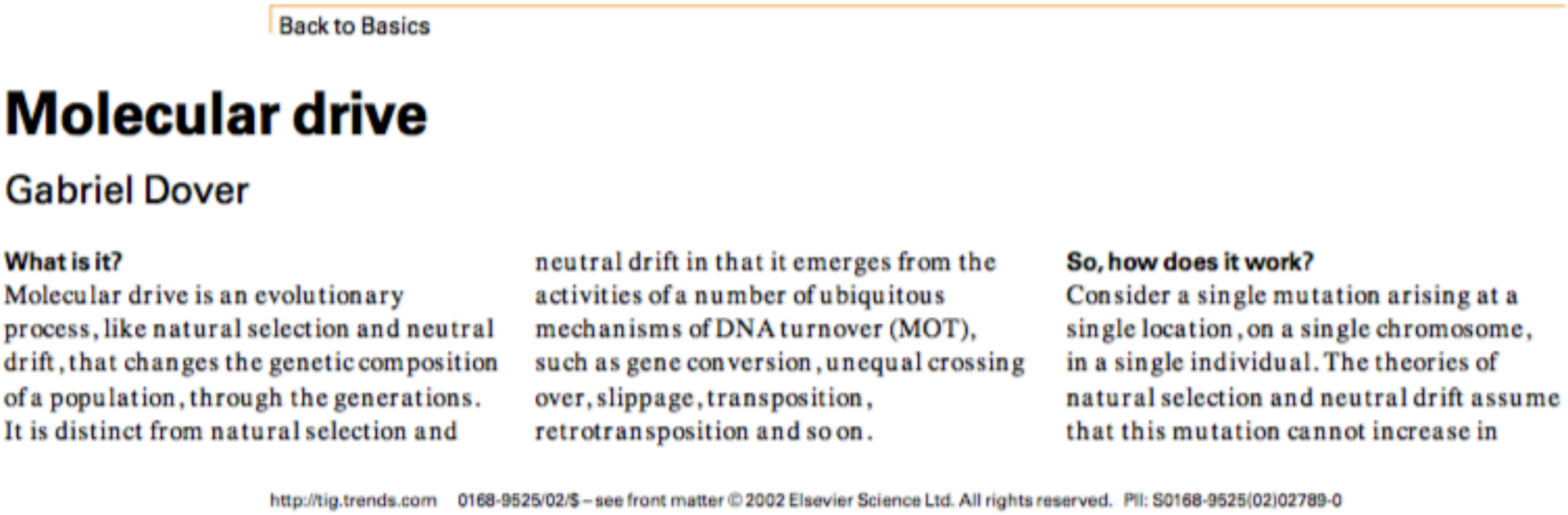
Login via Athens

Email:

Password:

□ **save your password**

What happens if I save my password



Molecular drive

Gabriel Dover

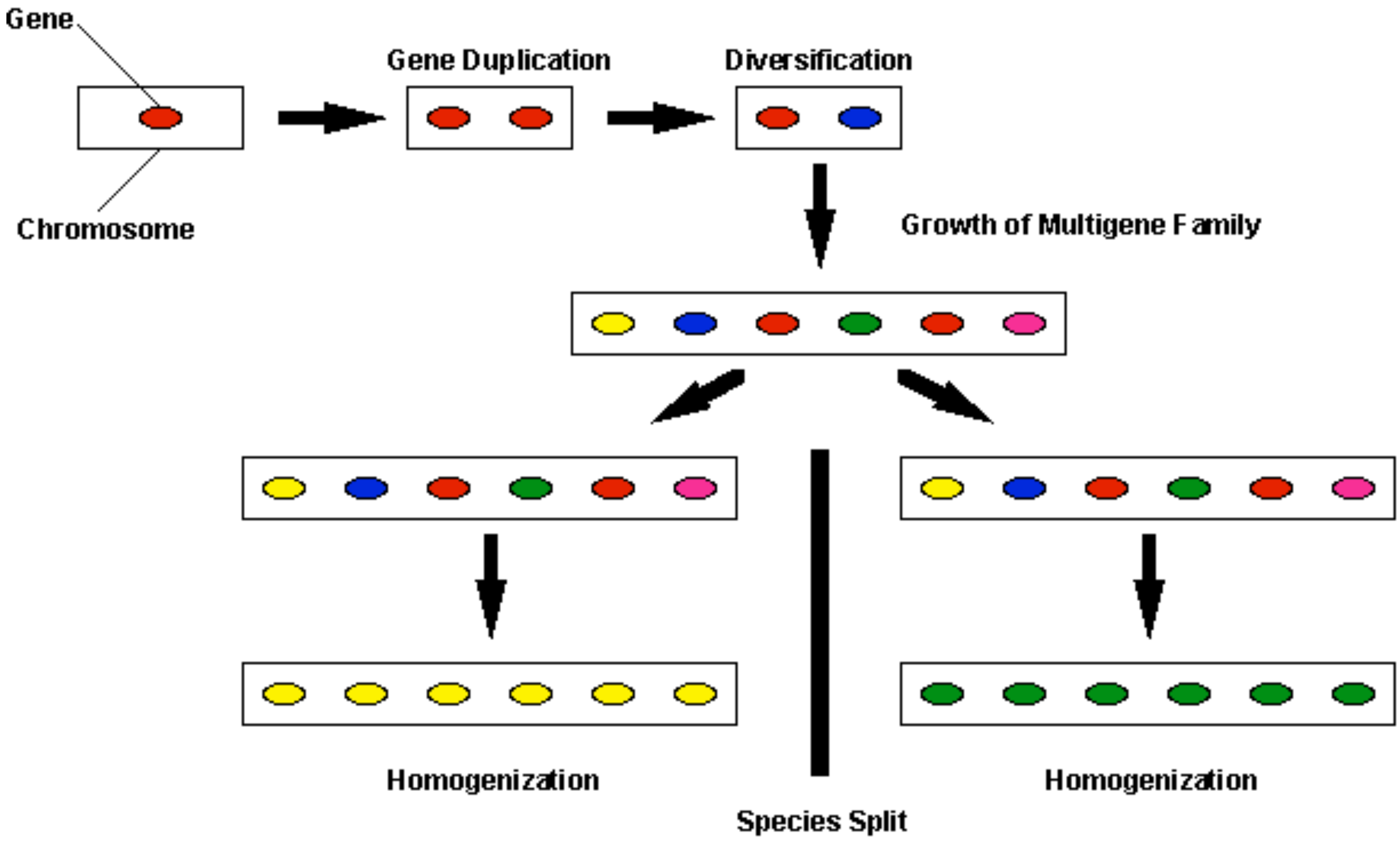
What is it?

Molecular drive is an evolutionary process, like natural selection and neutral drift, that changes the genetic composition of a population, through the generations. It is distinct from natural selection and neutral drift in that it emerges from the activities of a number of ubiquitous mechanisms of DNA turnover (MOT), such as gene conversion, unequal crossing over, slippage, transposition, retrotransposition and so on.

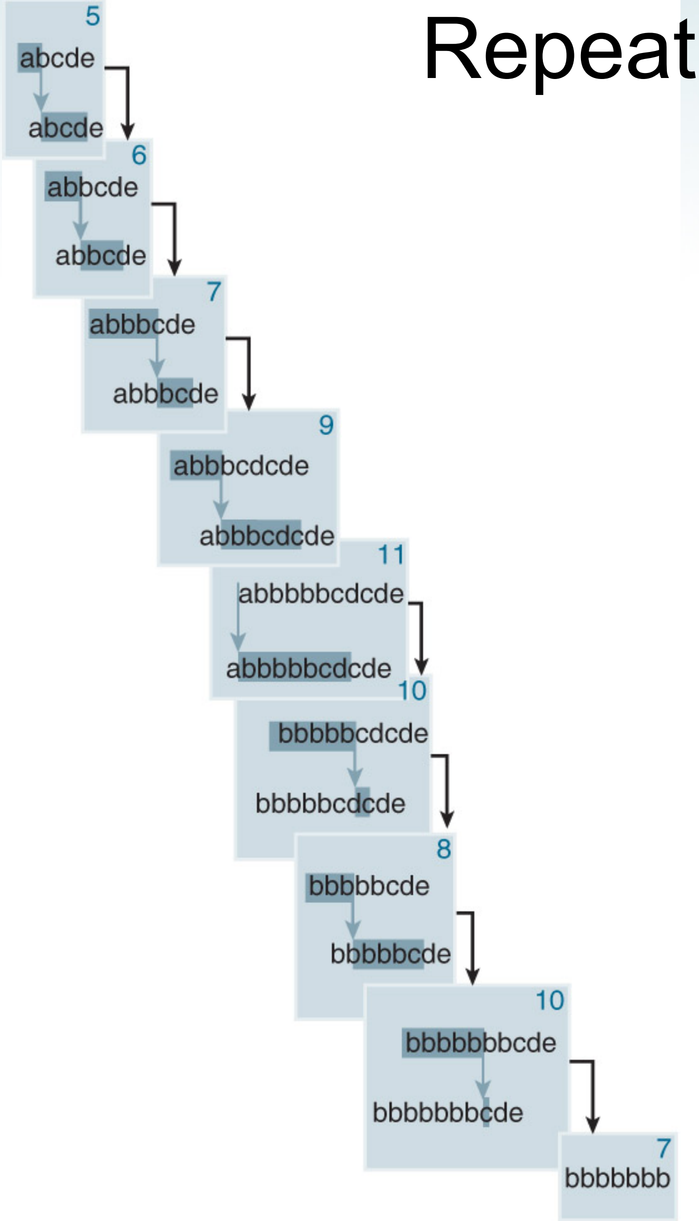
So, how does it work?

Consider a single mutation arising at a single location, on a single chromosome, in a single individual. The theories of natural selection and neutral drift assume that this mutation cannot increase in

http://tig.trends.com 0168-9525/02/$ – see front matter © 2002 Elsevier Science Ltd. All rights reserved. PII: S0168-9525(02)02789-0

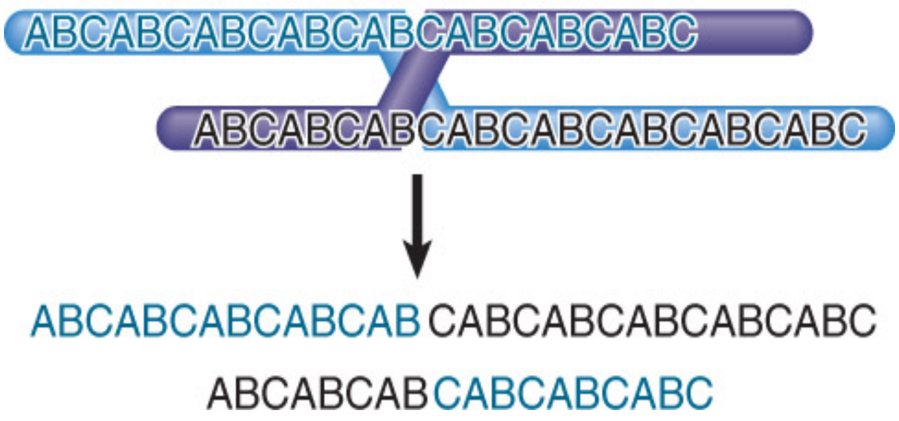

Gene
Gene Duplication
Diversification
Chromosome
Growth of Multigene Family
Homogenization
Species Split
Homogenization

Crossover Fixation Could Maintain Identical Repeats

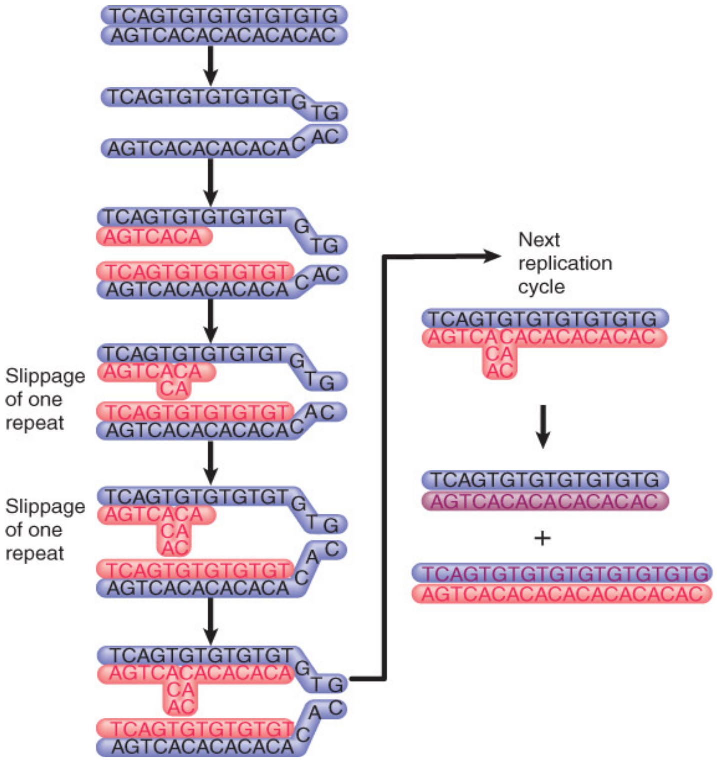


Unequal recombination allows one particular repeating unit to occupy the entire cluster

- **Unequal crossing-over (nonreciprocal recombination) –** Unequal crossing-over results from an error in pairing and crossing-over in which nonequivalent sites are involved in a recombination event.



Unequal crossing-over results from pairing between nonequivalent repeats in regions of DNA consisting of repeating units



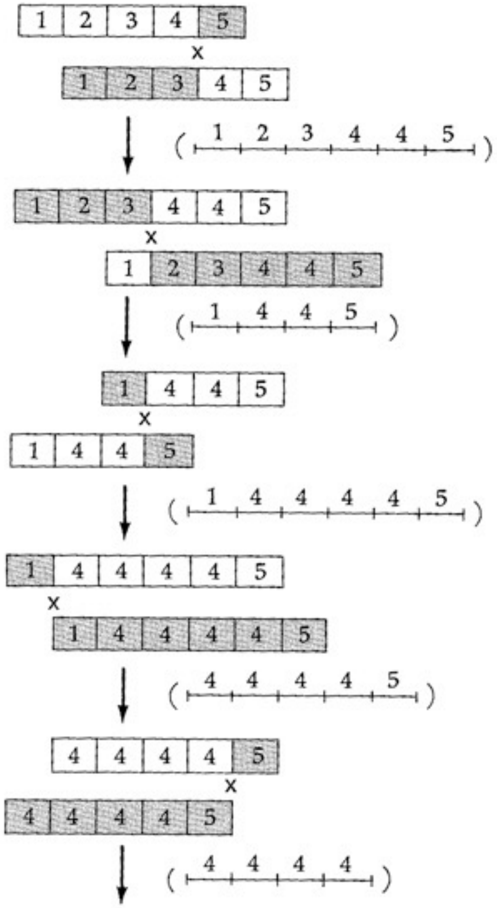
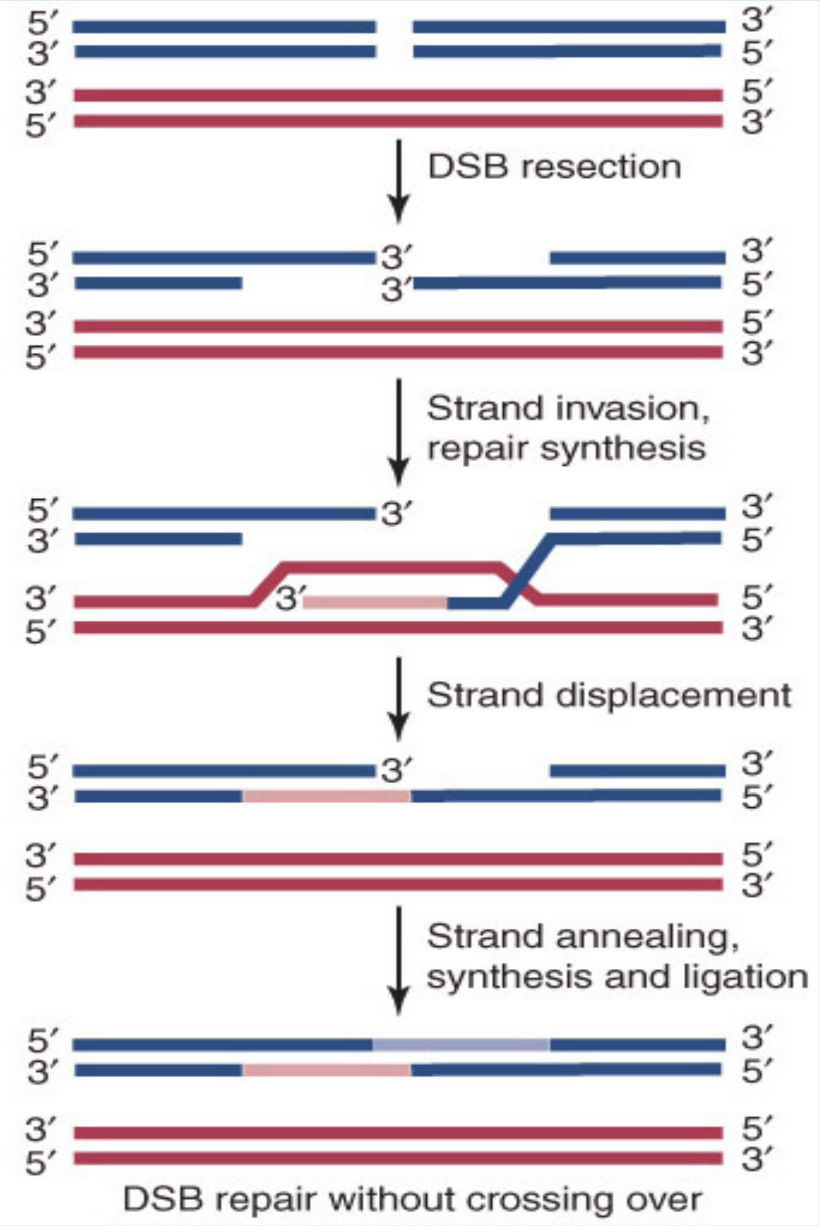
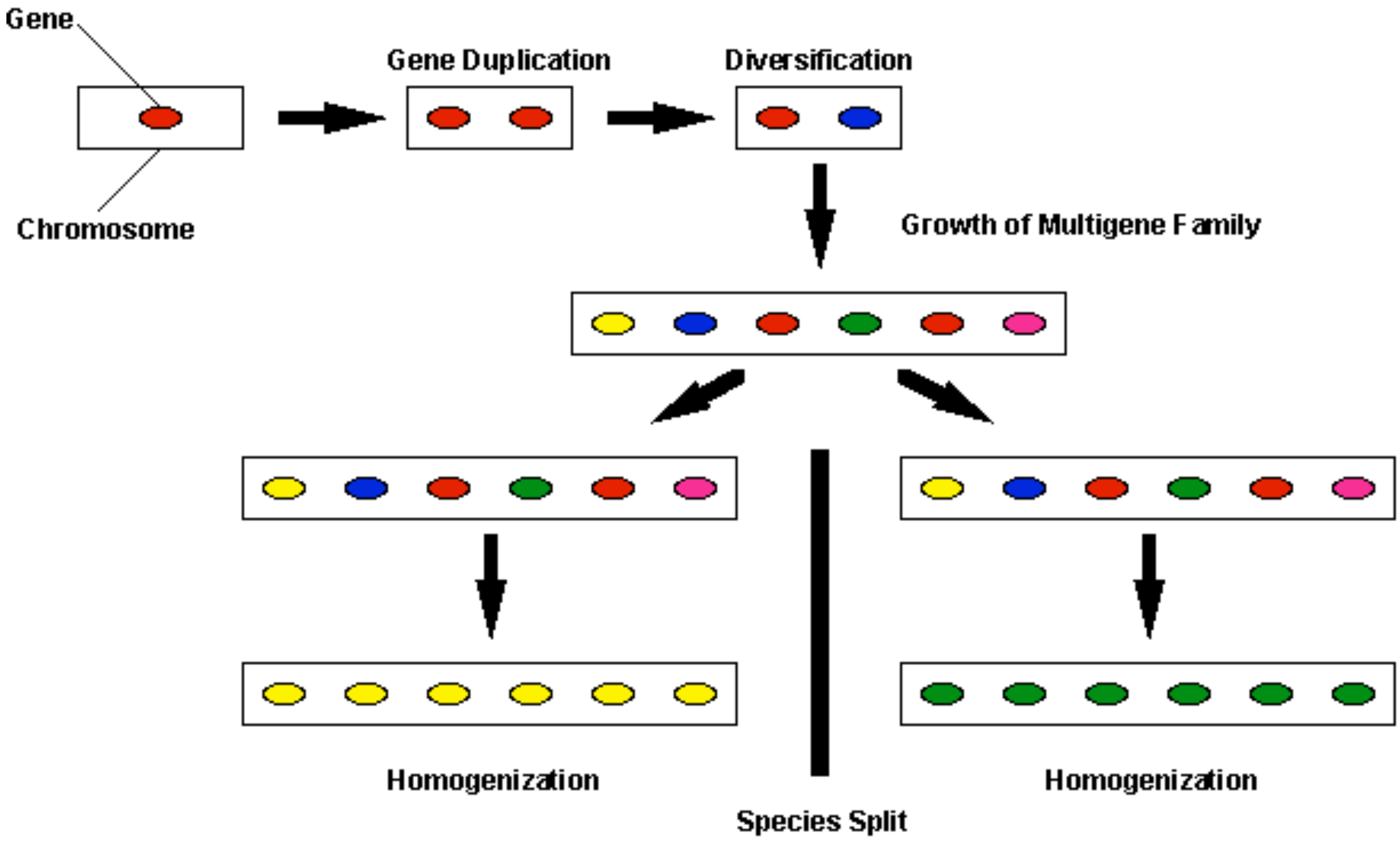


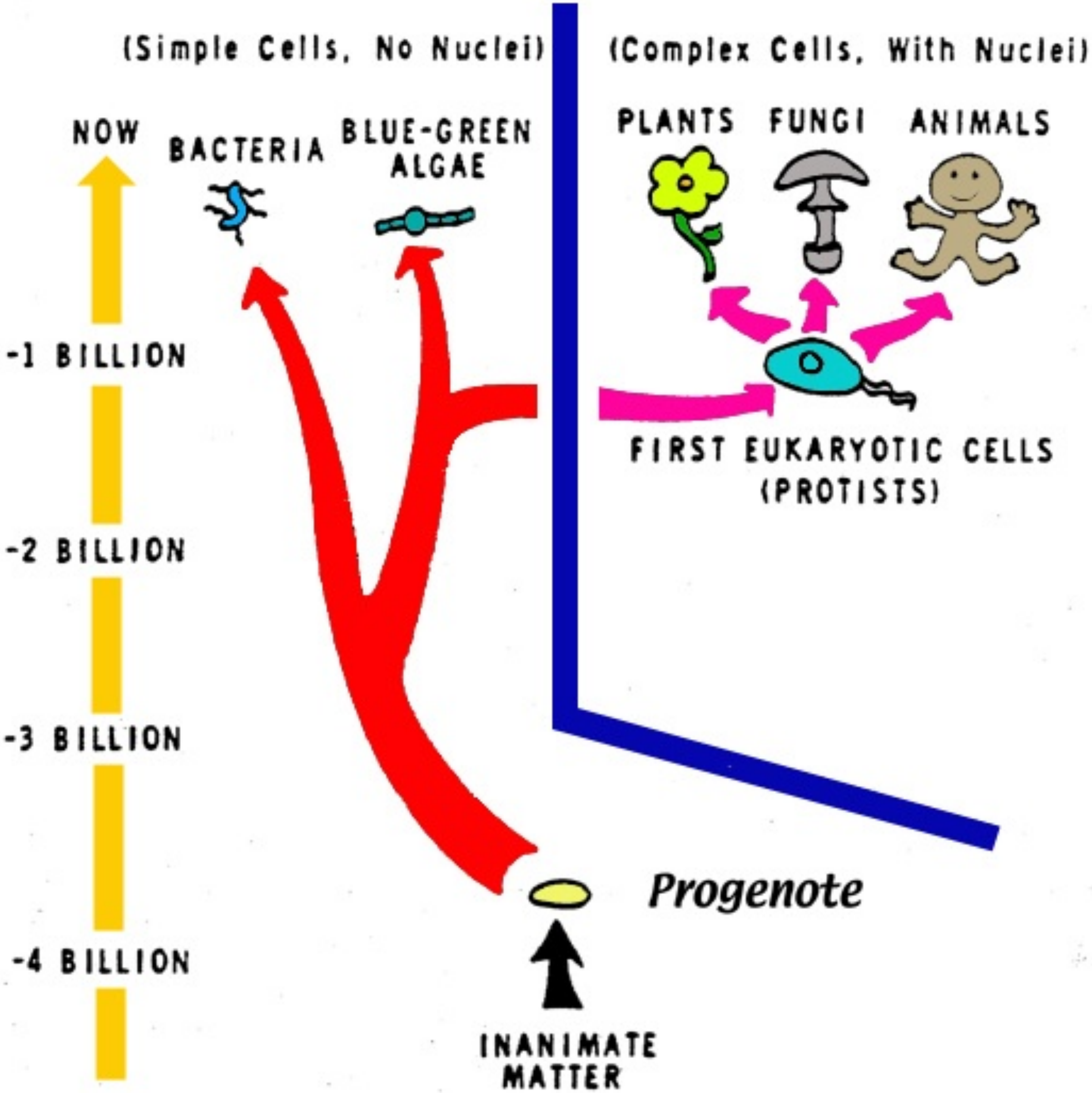
Figure 19. Concerted evolution by unequal crossing-over. Repeated cycles of unequal crossover events cause the duplicated genes on each chromosome to become progressively more homogenized. The process also affects the number of repeated sequences on each chromosome. From Ohta (1980).



- The synthesis-dependent strand-annealing model (SDSA) is relevant for mitotic recombination, as it produces gene conversions from double-strand breaks without associated crossovers.


Gene
Gene Duplication
Diversification
Chromosome
Growth of Multigene Family
Homogenization
Homogenization
Species Split

PROKARYOTES *EUKARYOTES*



Progenote