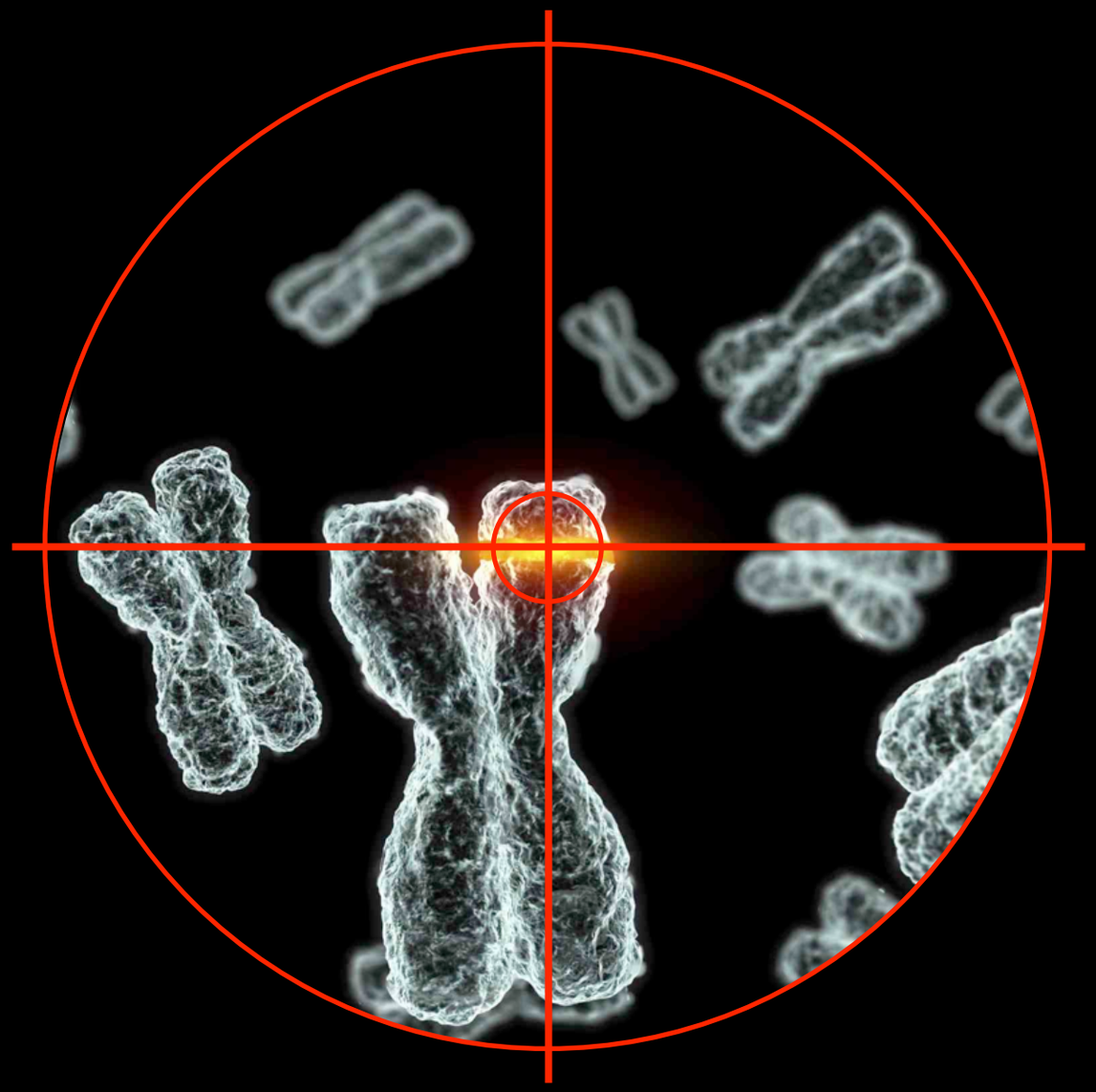
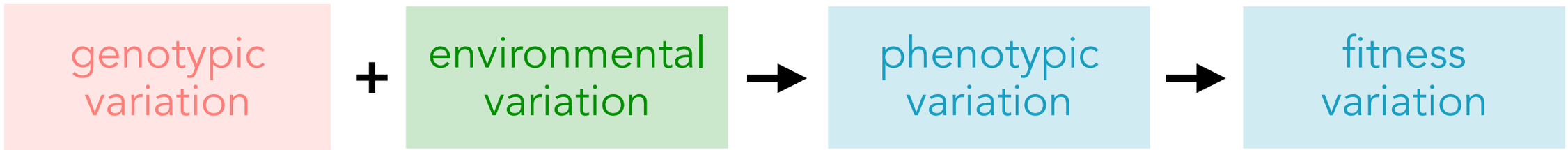
Evolutionary Genetics

LV 25600-01 | Lecture with exercises | 4KP



Mutations



There are **three primary sources of genetic variation**:

1. **Mutations** are the sources of all genetic variation. A single mutation can have a large effect, but in many cases, evolutionary change is based on the accumulation of many mutations.
2. **Gene flow** (migration) is any movement of genes from one population to another and is an important source of genetic variation.
3. **Sex** can introduce new gene combinations into a population. This genetic shuffling is another important source of genetic variation.

Irreversible Mutation

$$A_1 \xrightarrow{\mu} A_2$$

$$P(A_1 \to A_2) : \mu$$
$$P(A_1 \to A_1) : 1-\mu$$

$$p_1 = p_0(1-\mu)$$

p_1 : frequency of allele A_1 in the next generation

μ : muation rate

Irreversible Mutation

For one generation

$$p_t = p_{t-1}(1-\mu)$$

For many generation

$$p_t = p_0(1-\mu)^t$$

Alternative

$$p_t = p_0 e^{-\mu t}$$

Irreversible Mutation

$$p_t = p_{t-1}(1-\mu)$$

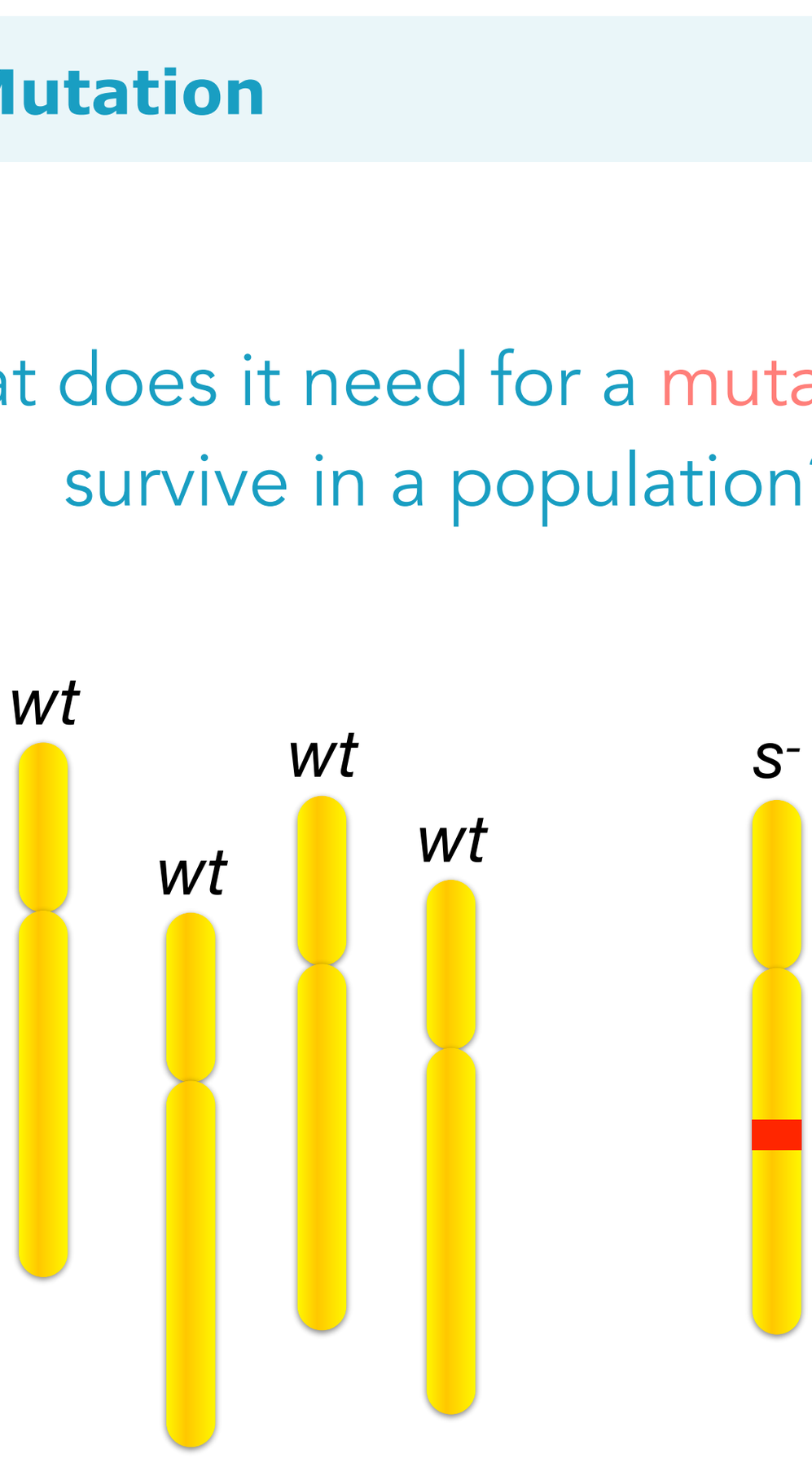
p_t : frequency of allele A after t generations

μ : muation rate

max: $\mu = 1 \rightarrow p_t = 0$

min: $\mu = 0 \rightarrow p_t = p_{t-1}$

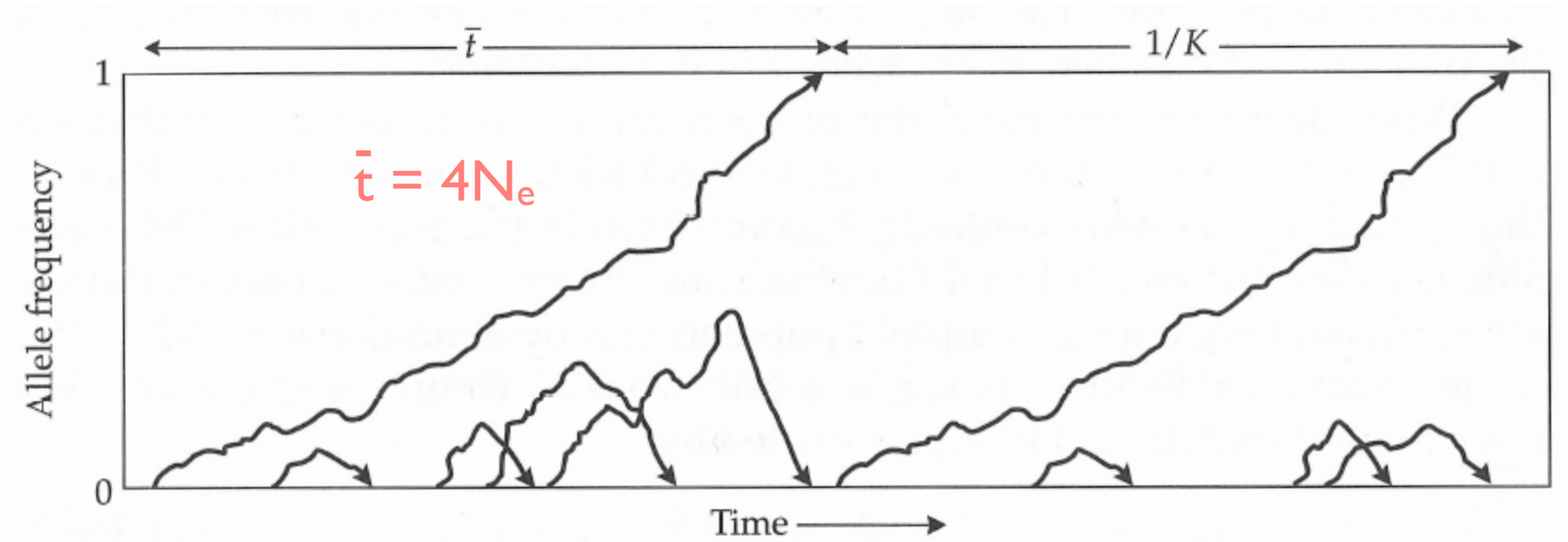
What does it need for a mutation to survive in a population?



The concepts and models of genetic drift can be employed to predict the frequency of a new mutations over time in a population. The first criterial observation is to recognize that the initial frequency of any new mutation is ...

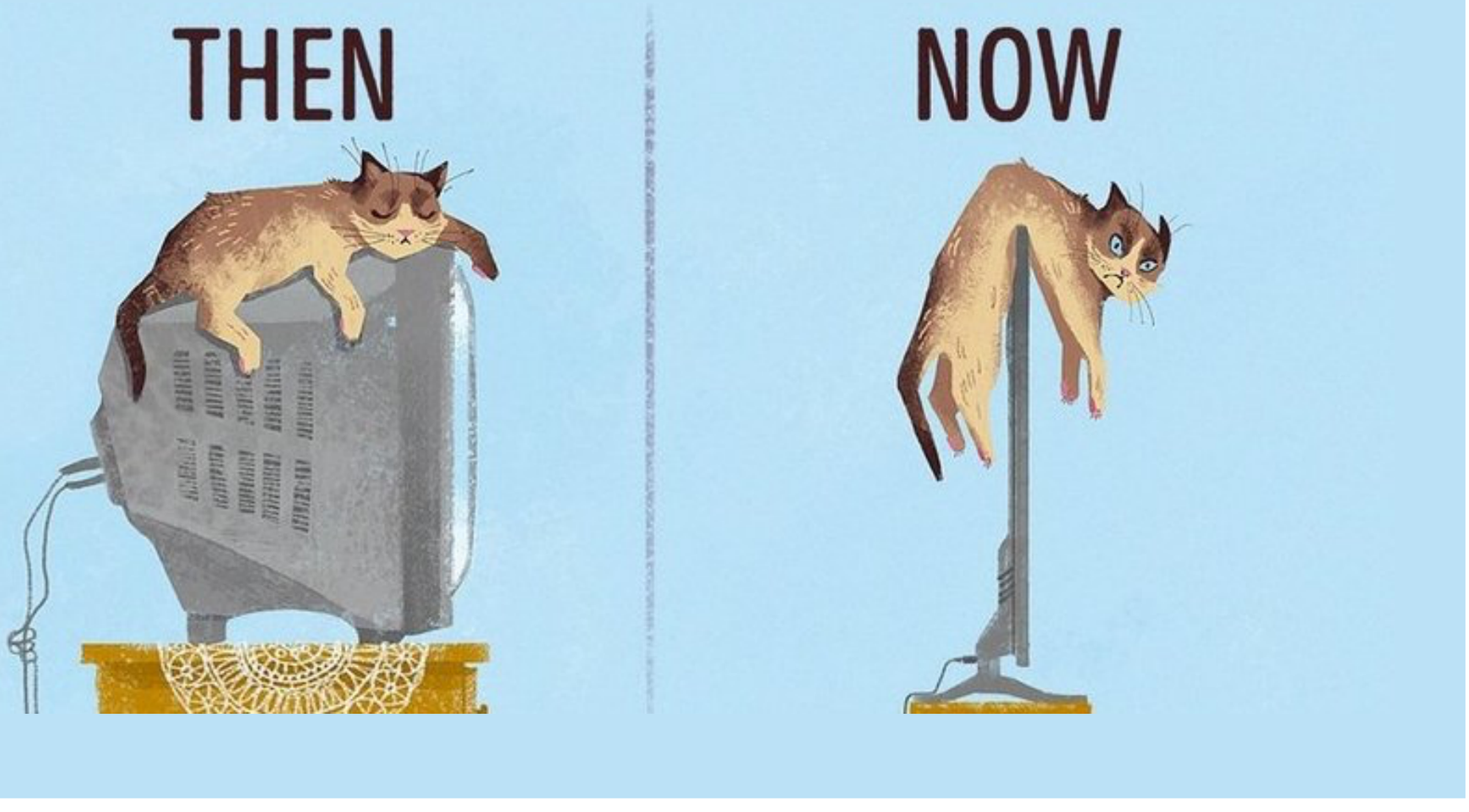
$$p_0(\text{new mutation}) = \frac{1}{2N_e}$$

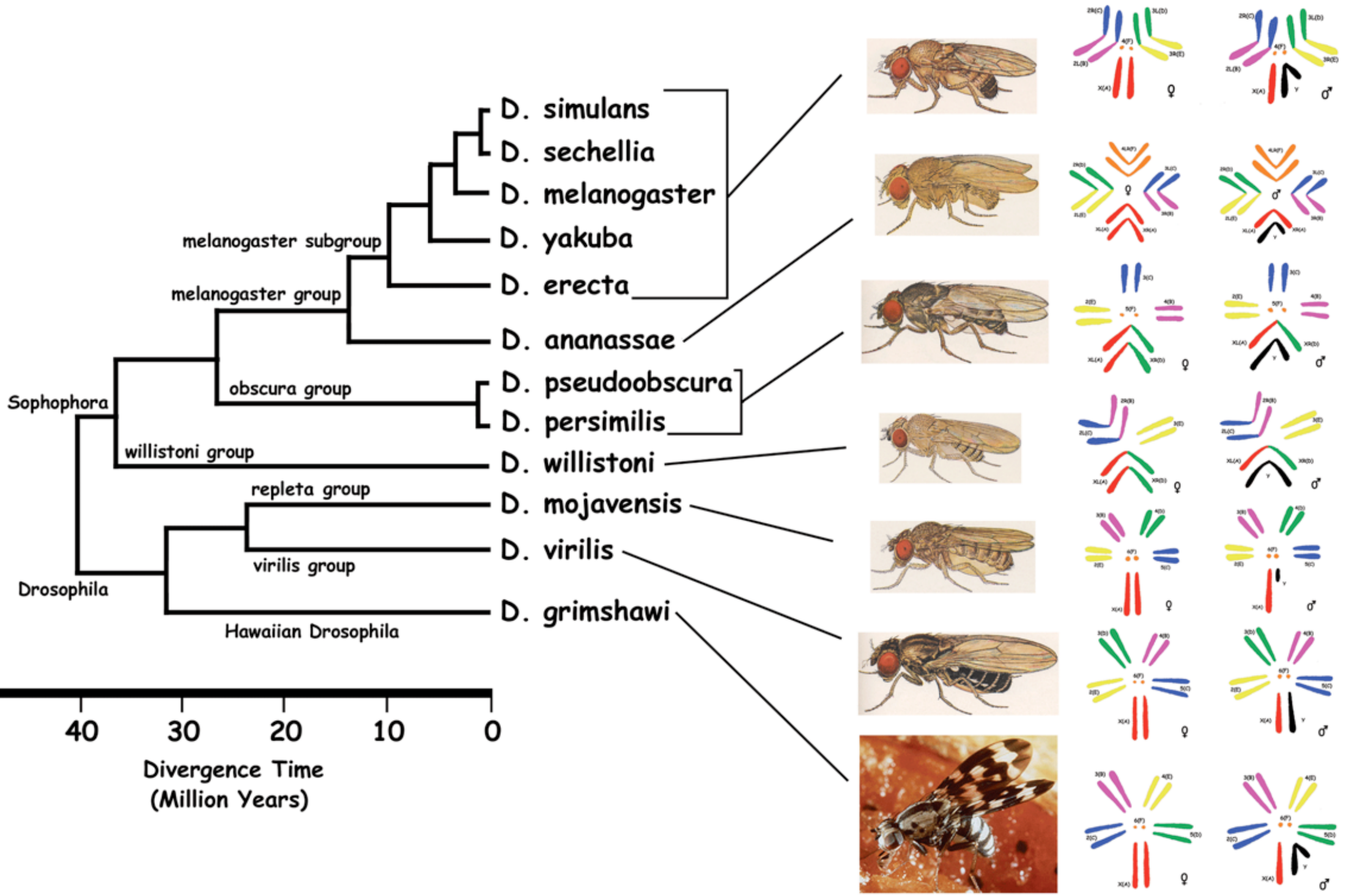
Because a new mutation is present as a single allele copy in a population of N_e alleles copies. If the frequency of the new mutation is determined strictly by genetic drift, then each new mutation has a probability of $1/(2N_e)$ of going to fixation and a probability of $1-1/(2N_e)$ of going to loss. Under the assumption that the effective population size is large, **those alleles that eventually fix do so in average of $4N_e$ generations** (Kimura & Ohta 1969 and Nei 1987).



MORE
THINGS
CONSIDERED

How do genomes change over time?

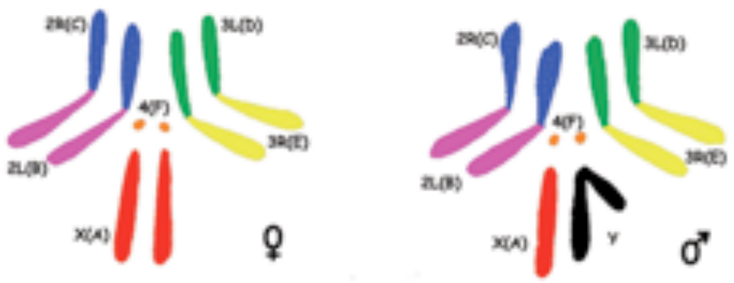




http://insects.eugenes.org/DroSpeGe/

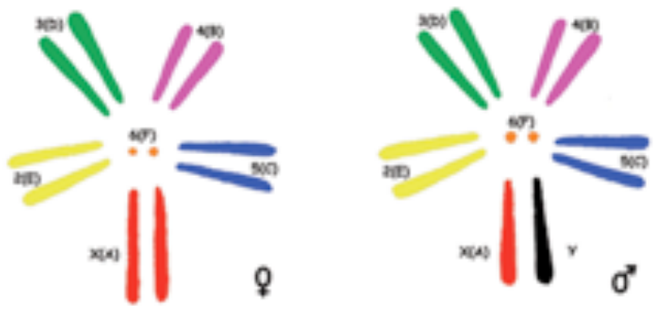
D. melanogaster

n=4



D. virilis

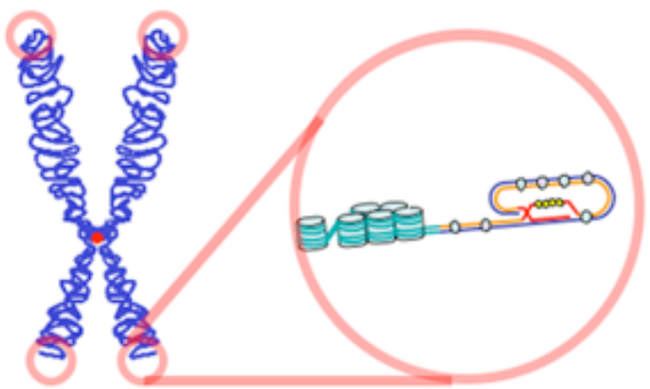
n=6



Telomeres of Human Chromosomes

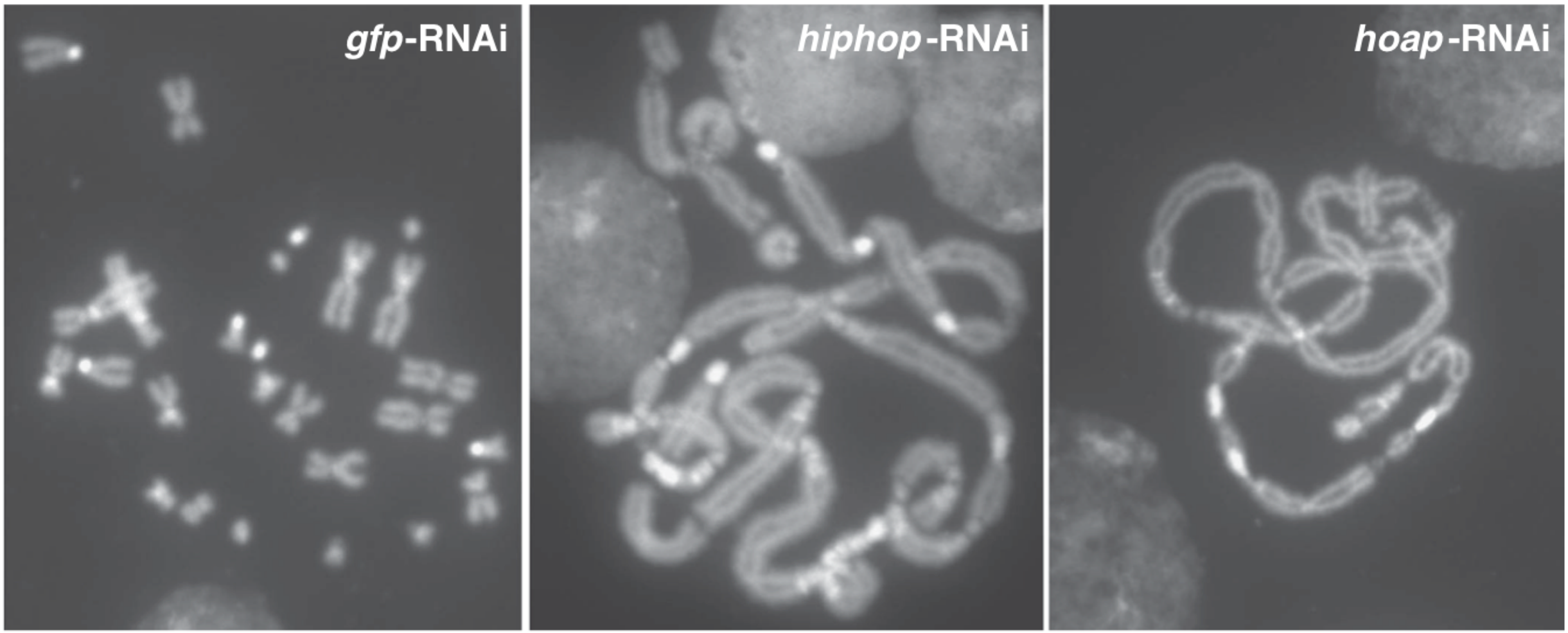
By: Clare O'Connor, Ph.D. (*Biology Department, Boston College*)
Citation: O'Connor, C. (2008) Telomeres of human chromosomes. *Nature Education* 1(1):166

In the 1930s, Hermann Muller was the first researcher to note that the ends of chromosomes had unique properties. Muller named these ends telomeres (from the Greek words *telo*, meaning "end," and *mere*, meaning "part"), based on their position on chromosomes. Later, in his classic mutagenesis experiments in *Drosophila*, Muller used X-rays to generate various mutations, many of which involved chromosome breakage and fusion. However, during these experiments, he was surprised to note that the ends of chromosomes were strangely resistant to the effects of the mutagenic X-rays. Muller therefore hypothesized that "the terminal gene must have a special function, that of sealing the end of the chromosome, so to speak," and that "for some reason, a chromosome cannot persist indefinitely without having its ends thus sealed" (McKnight & Shippen, 2004). We now know that Muller was only partially correct. Telomeres do indeed play an essential role in stabilizing the ends of chromosomes, but they do not contain active genes. Instead, telomeres contain an array of highly repeated DNA sequences and specific binding proteins that form a unique structure at the end of the chromosome.



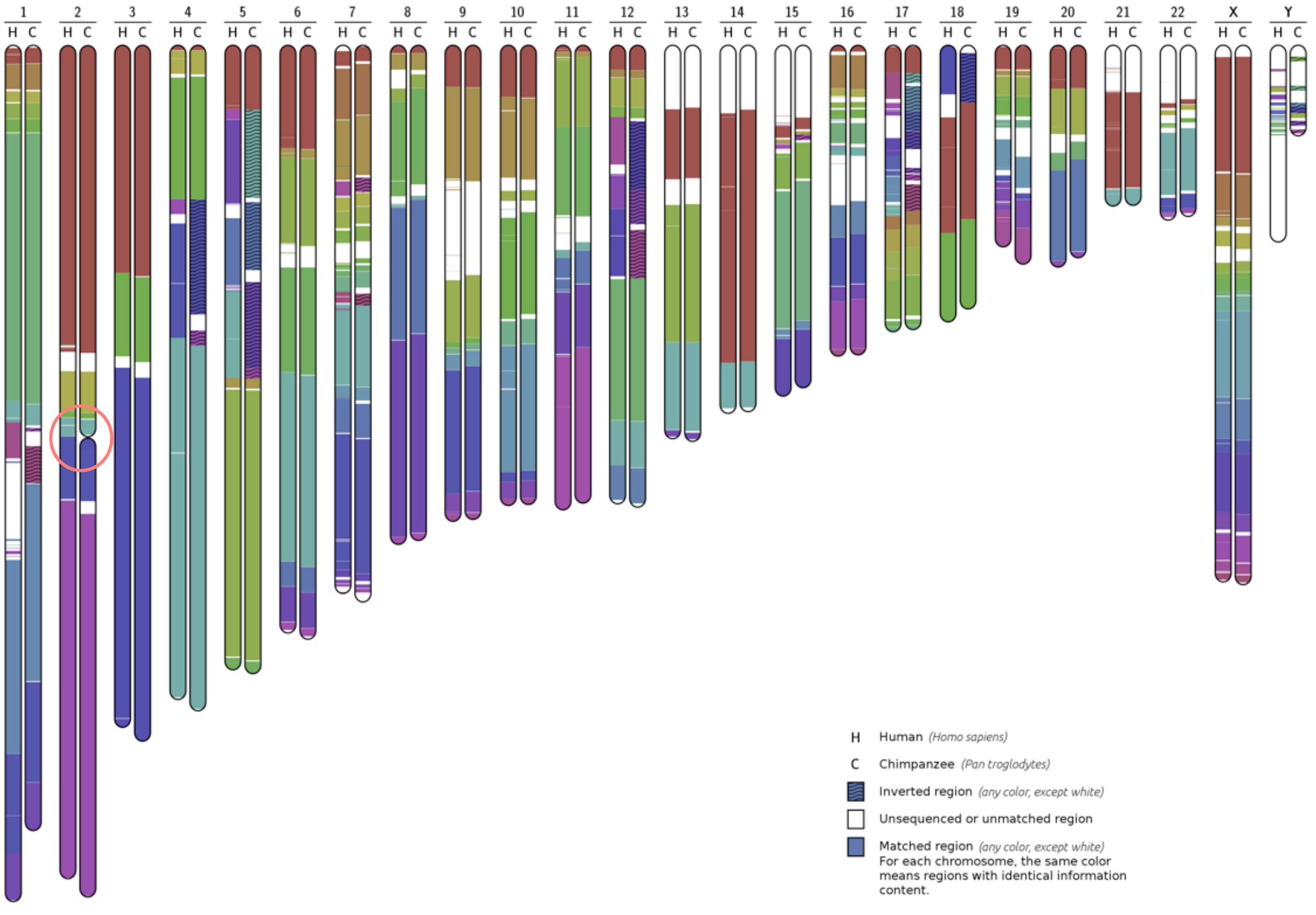
Telomeres are integral in understanding chromosome biology, structure and function. Acting as specialized DNA–protein complexes that cap the ends of eukaryotic chromosomes, telomeres are necessary for **complete chromosomal end replication** and provide chromosomal stability by **protecting chromosome ends from degradation and fusion**.


FEMALE
MALE

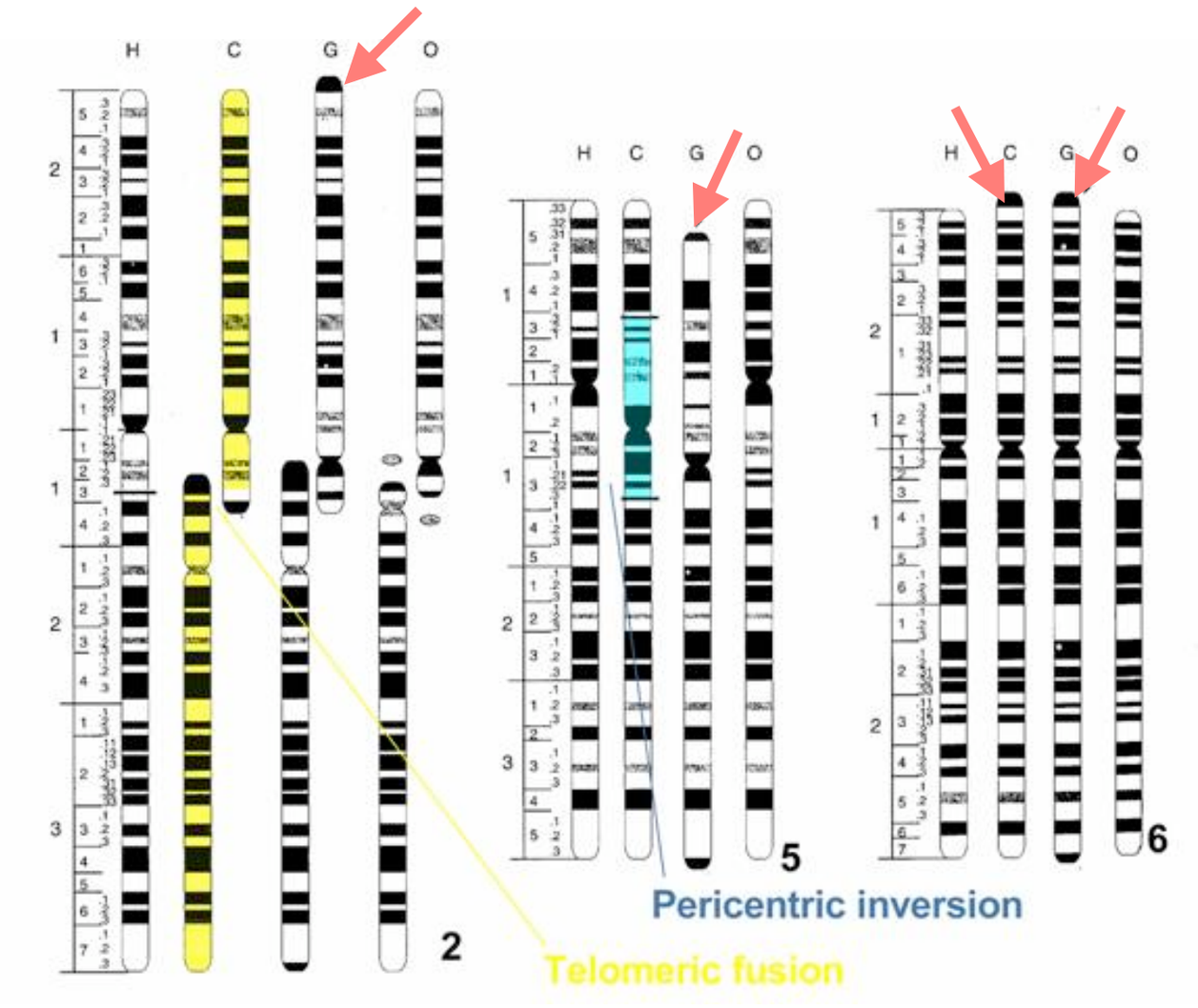

gfp-RNAi
hiphop-RNAi
hoap-RNAi

"The similarity between human and chimpanzee DNA is really in the eye of the beholder. If you look for similarities, you can find them. But if you look for differences, you can find those as well. There are significant differences between the human and chimpanzee genomes that are not easily accounted for in an evolutionary scenario."

David DeWitt (2014)


1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 X Y
H C
H Human (Homo sapiens)
C Chimpanzee (Pan troglodytes)
Inverted region (any color, except white)
Unsequenced or unmatched region
Matched region (any color, except white)
For each chromosome, the same color means regions with identical information content.

Perhaps one of the most often quoted scientific facts has been the one highlighting how genetically close humans are to chimpanzees. “More than 98% of our genes are identical” is typical of the kind of statement found in both popular and scientific literature. Indeed, the sequence homology of both coding and noncoding DNA in humans is very similar to chimpanzee, as it is to bonobo, gorilla and orangutang. This is hardly surprising, however. Considering that perhaps as much as 40% of the genome in every living organism codes for proteins that are essential for general intracellular function, it is little wonder that a group of closely related mammals should have a very high percentage of common genes.



Strickberger (2000) and Yunis and Prakash (1982)

There is another even simpler, numeric, genetic fact about humans and chimpanzees that is not so often cited. Instead of emphasizing our similarities it actually differentiates *Homo sapiens* from the other hominoids: It is the number of chromosomes.

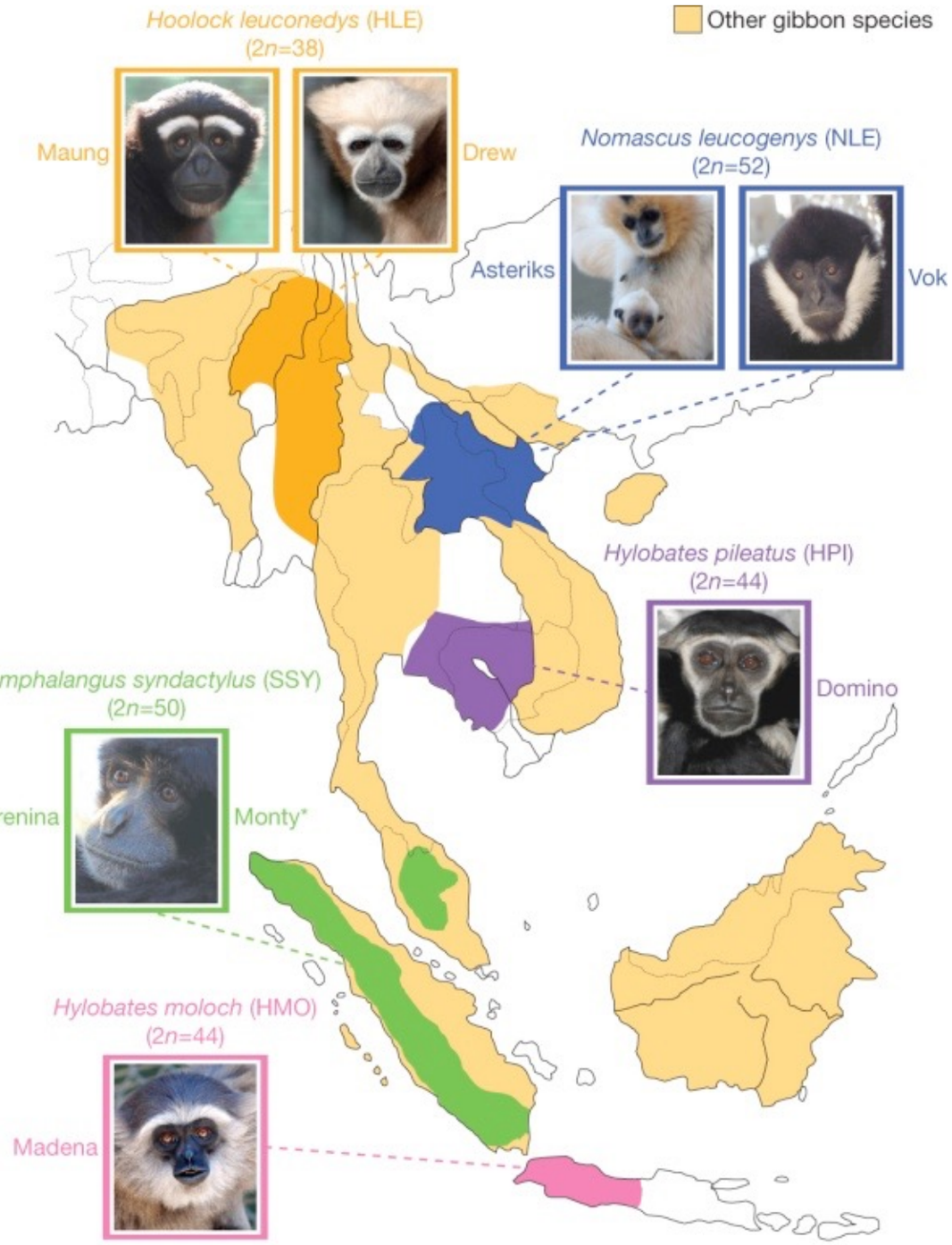
Chr2: A telomere-telomere fusion of ancestral chromosomes 2 & 3

Chr5: A pericentric inversion and a reciprocal translocation

Chr6: A small terminal deletion

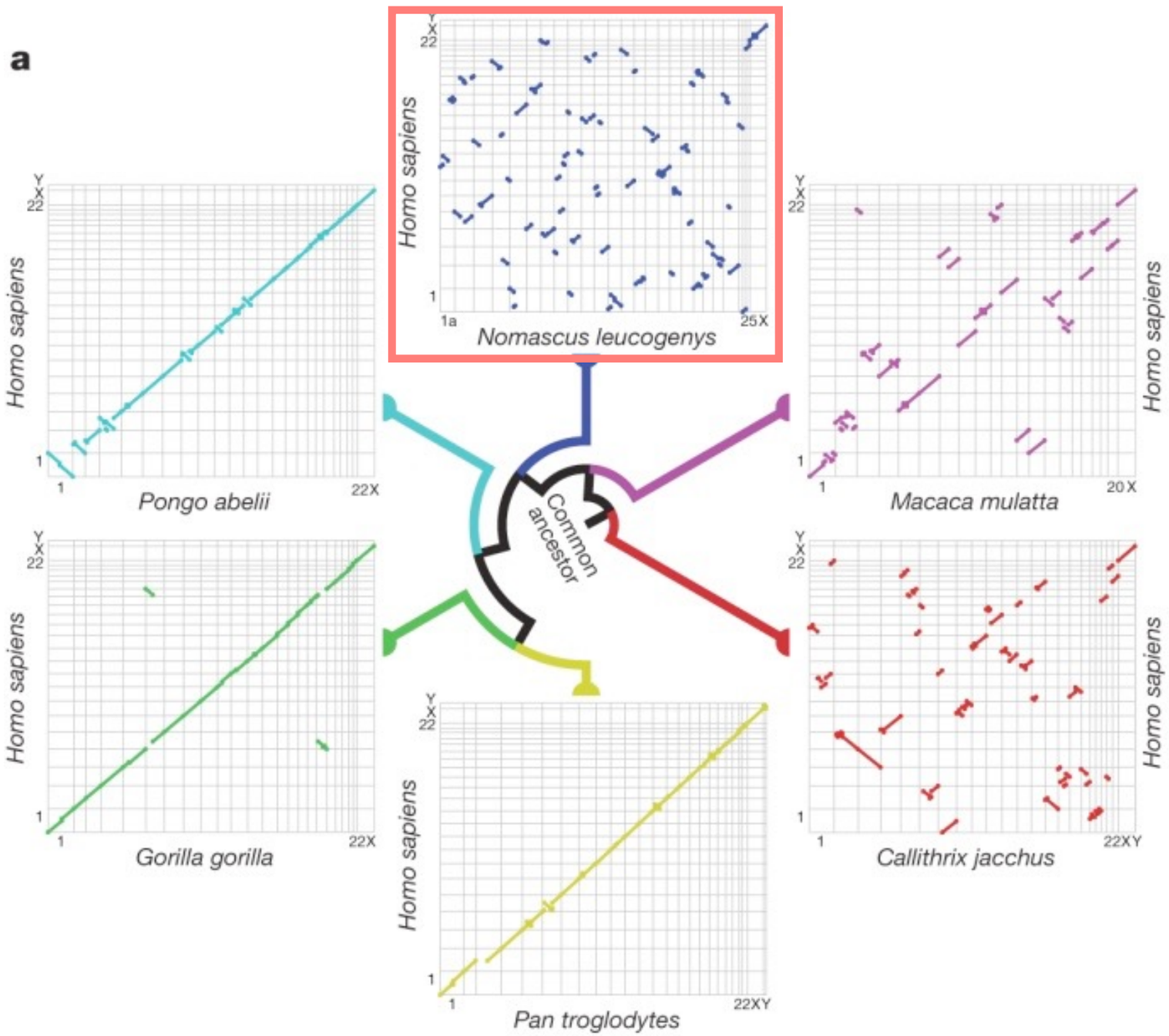
Gibbon genome and the fast karyotype evolution of small apes

Carbone *et al. Nature* **513**, 195-201 (2014) doi:10.1038/nature13679



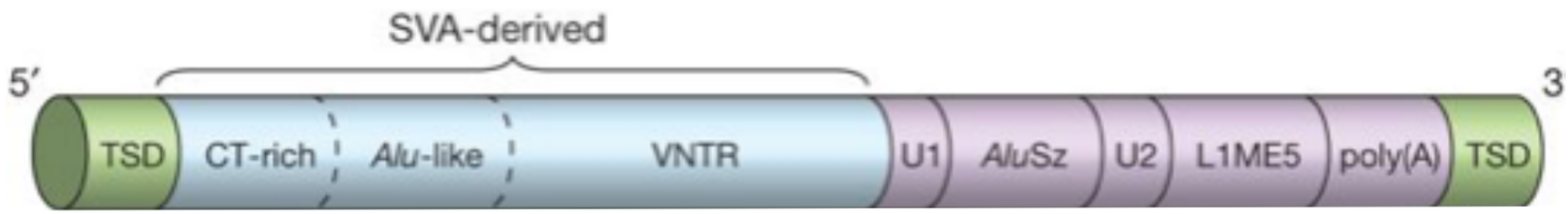
Gibbons (Hylobatidae) are critically endangered small apes that inhabit the tropical forests of southeast Asia

Analysis of gibbon–human **synteny** and **breakpoints**.



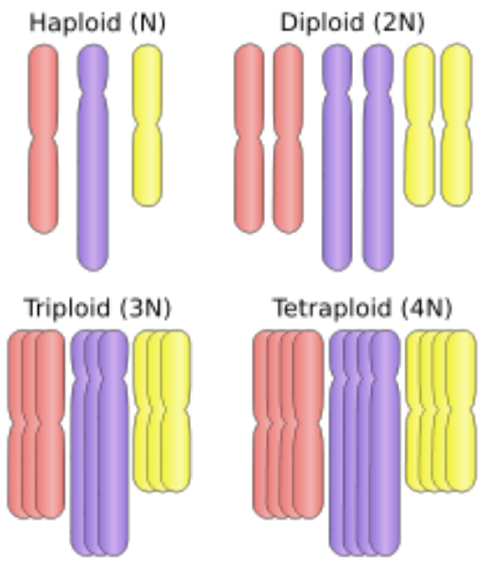
Carbone *et al. Nature* **513**, 195-201 (2014) doi:10.1038/nature13679

The gibbon genome contains all previously described classes of transposable elements that are mostly also present in other primates. One exceptional addition is the **LAVA element**, a novel retrotransposon that emerged exclusively in gibbons and has a composite structure comprised of portions of other repeats.



Accelerated **rearrangement** was confined to large-scale chromosomal events, pointing to a mechanism responsible for causing **gross chromosomal changes**, rather than global genomic instability. This is in line with our hypothesis that the high rate of chromosomal rearrangements may have been due to **LAVA-induced premature transcription termination of chromosome segregation genes**.

Carbone *et al. Nature* **513**, 195-201 (2014) doi:10.1038/nature13679



Polyploidy is the state of a cell or organism having more than two paired (homologous) sets of chromosomes. Most species whose cells have nuclei (eukaryotes) are **diploid**, meaning they have two sets of chromosomes -one set inherited from each parent. However, some organisms are polyploid, and polyploidy is especially common in plants. Polyploidy may occur due to **abnormal cell division**, either during mitosis, or commonly during metaphase I in meiosis. In addition, it can be induced in plants and cell cultures by some chemicals (e.g. colchicine).

"Mutation" is a change in DNA, the hereditary material of life. An organism's DNA affects how it looks, how it behaves, and its physiologys. So a change in an organism's DNA can cause changes in all aspects of its life.

Mutations are random - Mutations can be beneficial, neutral, or harmful for the organism, but mutations do not "try" to supply what the organism "needs." In this respect, mutations are random—whether a particular mutation happens or not is unrelated to how useful that mutation would be.

Not all mutations matter to evolution - Since all cells in our body contain DNA, there are lots of places for mutations to occur; however, not all mutations matter for evolution. Somatic mutations occur in non-reproductive cells and won't be passed onto offspring.

The only mutations that matter to large-scale evolution are those that can be passed on to offspring. These occur in reproductive cells like eggs and sperm and are called **germ line mutations**.

A single germ line mutation can have a range of effects:

No change occurs in phenotype - Some mutations don't have any noticeable effect on the phenotype of an organism. This can happen in many situations: perhaps the mutation occurs in a stretch of DNA with no function, or perhaps the mutation occurs in a protein-coding region, but ends up not affecting the amino acid sequence of the protein.

Small change occurs in phenotype - A single mutation caused this cat's ears to curl backwards slightly.

Big change occurs in phenotype - Some really important phenotypic changes, like DDT resistance in insects are sometimes caused by single mutations (Williamson et al., 1996). A single mutation can also have strong negative effects for the organism. Mutations that cause the death of an organism are called lethal—and it doesn't get more negative than that.

There are some sorts of changes that a single mutation, or even a lot of mutations, could not cause. Neither mutations nor wishful thinking will make pigs have wings.

Mutations may be caused by **external factors** (UV light, chemical agents, etc.) or **spontaneous cellular processes** (replication errors, accidental deamination, etc.).

Replication errors are the main source of mutations. It has been estimated that uncorrected replication errors occur with a frequency of 10^{-9} - 10^{-11} for each nucleotide added by DNA polymerases.

Organism	Effective genome size (Ge)	Mutations per genome per replication
bacteriophage M13	6.4×10^3	0.0046
bacteriophage lambda	4.9×10^4	0.0038
bacteriophages T2 & T4	1.7×10^5	0.0040
E. coli	4.6×10^6	0.0025
Saccharomyces cerevisiae	1.2×10^7	0.0027
Neurospora crassa	4.2×10^7	0.0030
C. elegans	1.8×10^7	0.004
Drosophila	1.6×10^7	0.005
Mouse	8.0×10^7	0.014
Human	8.0×10^7	0.004

Rates of spontaneous mutation.

J W Drake, B Charlesworth, D Charlesworth, and J F Crow

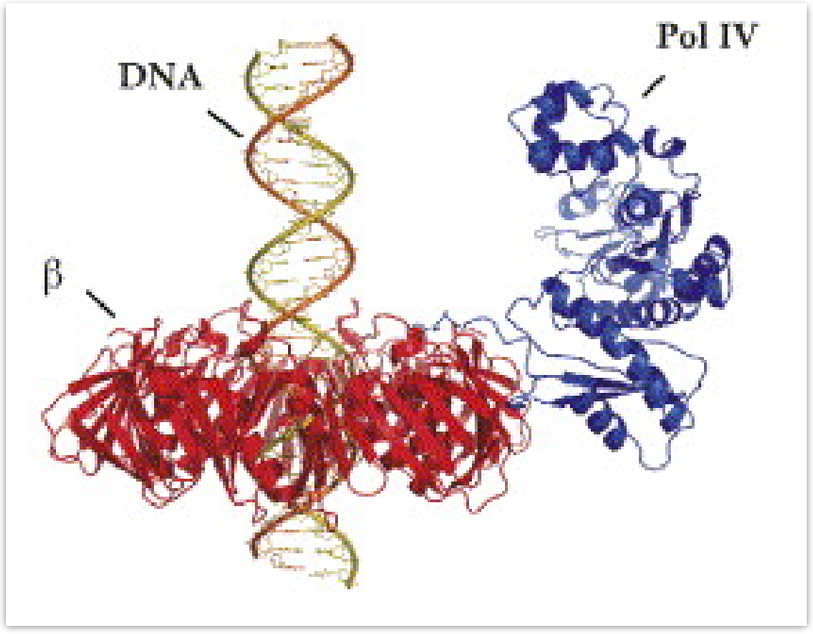
Genetics. 1998 April; 148(4): 1667–1686.

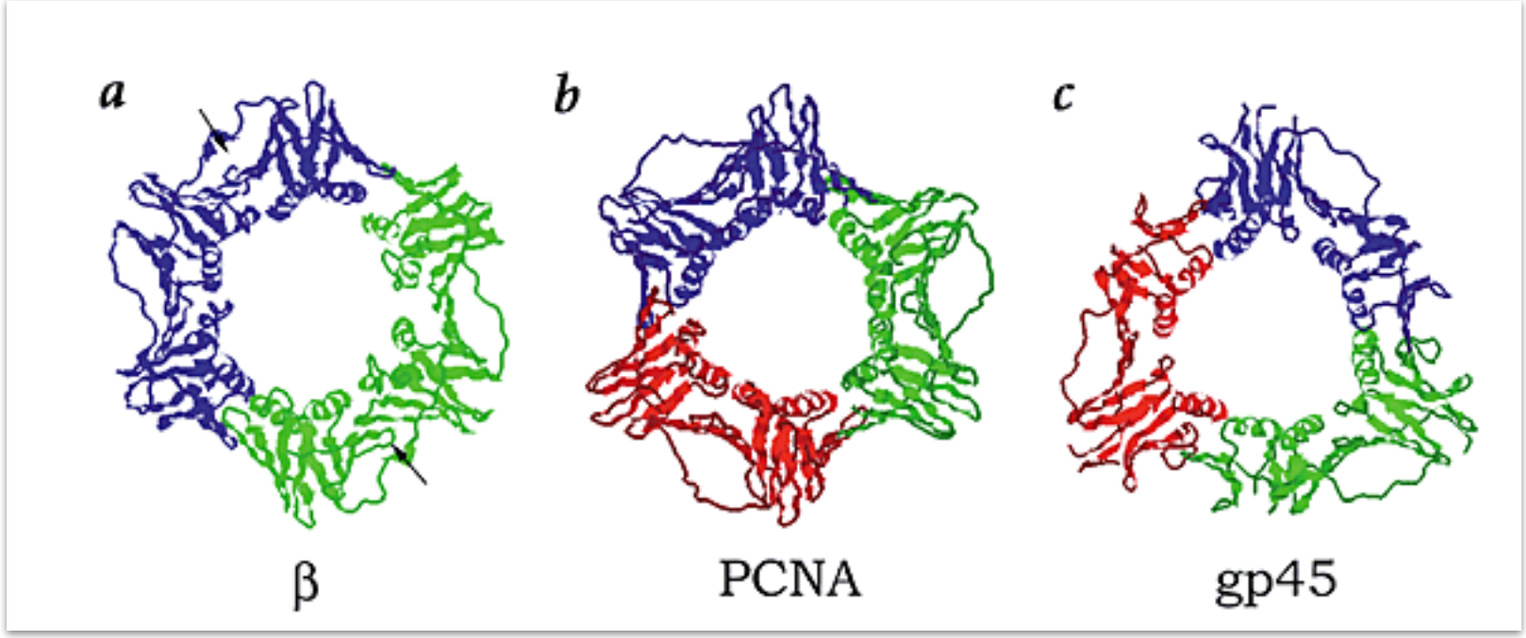
Mutations may be caused by **external factors** (UV light, chemical agents, etc.) or **spontaneous cellular processes** (replication errors, accidental deamination, etc.).

Replication errors are the main source of mutations. It has been estimated that uncorrected replication errors occur with a frequency of 10^{-9} - 10^{-11} for each nucleotide added by DNA polymerases.

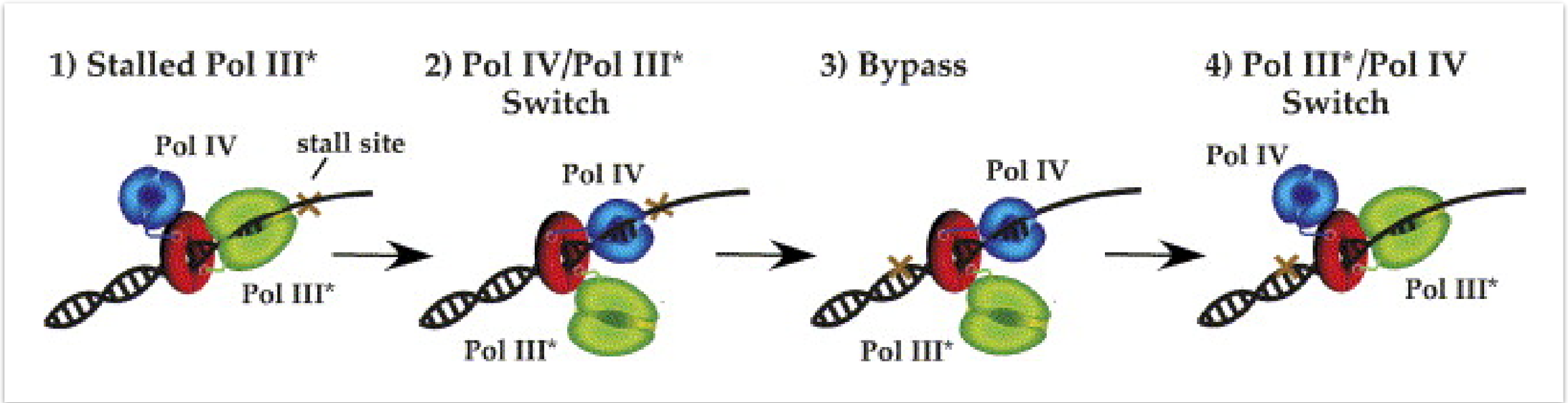
A commonly observed replication error is the **replication slippage**, which occurs at the repetitive sequences when the new strand mispairs with the template strand. The microsatellite polymorphism is mainly caused by the replication slippage. If the mutation occurs in a coding region, it could produce abnormal proteins, leading to diseases. The Huntington's disease is a well known example.

On some occasions, DNA damage is not repaired, or is repaired by an **error-prone mechanism** which results in a change from the original sequence. When this occurs, mutations may propagate into the genomes of the cell's progeny. The rate of evolution in a particular species (or, more narrowly, in a particular gene) is a function of the rate of mutation. Consequently, the rate and accuracy of DNA repair mechanisms have an influence over the process of evolutionary change.





Sliding-clamp tool-belt



β sliding clamp of *E.coli* binds high- and low-fidelity DNA polymerases at the same time

$v_{Pol\ III}$ = ~ 400 bp/sec (*E. coli* replicates its entire genome in ~40 minutes => ~15,000 bp/sec)
$v_{Pol\ IV}$ = ~ 10 bp/sec

Source: Indiani et al. (2005) Molecular Cell 19, 805-815.

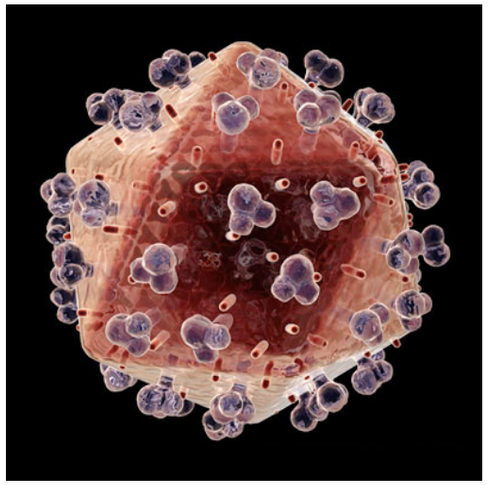
Eukaryotic DNA polymerase

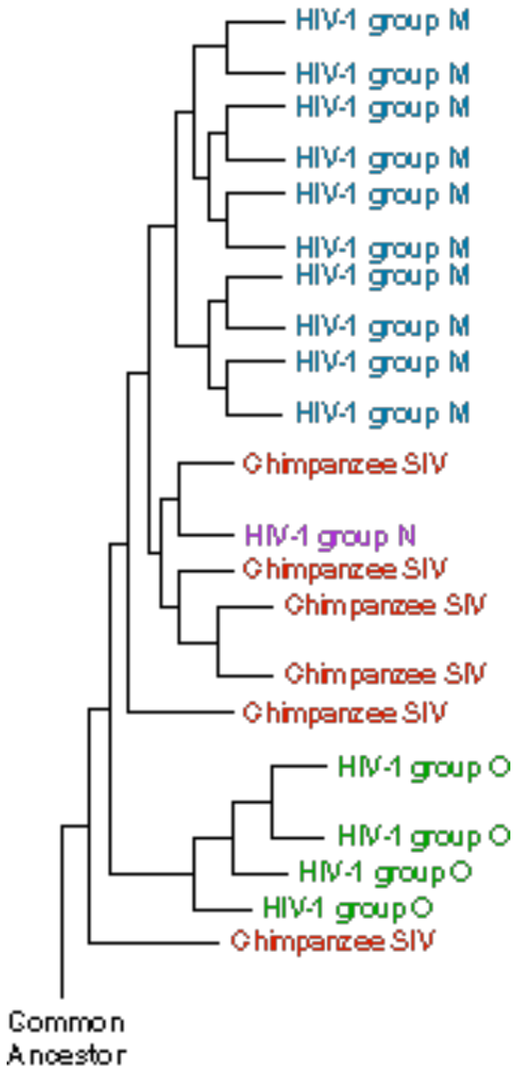
Family	Example	Fidelity on undamaged DNA
A	*POLG*	10^{-4} - 10^{-6}
B	*POL5*	
X	*POLL, TDT*	
Y	*POLI*	~ $2x10^{-1}$

Goodman (2001) Error-prone repair DNA polymerases in prokaryotes and eukaryotes. Annu Rev Biochem, 71: 17-50.

Yang (2003) Damage repair DNA polymerases Y. Curr Opin Struct Biol, 71: 17-50.

Rattray & Strathern (2003) Error-prone DNA polymerases: When making a mistake is the only way to get ahead. Annu Rev Gen, 13(1): 23-30.



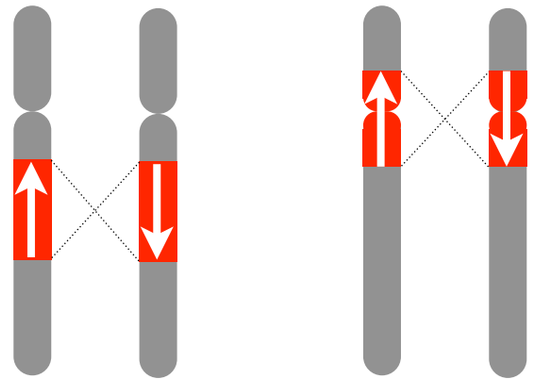


The **human immunodeficiency virus** is one of the fastest evolving entities known. It reproduces sloppily, accumulating lots of mutations when it copies its genetic material. It also reproduces at a lightning-fast rate — a single virus can spawn billions of copies in just one day. To fight HIV, one must understand its evolution within the human body and then ultimately find a way to control its evolution.

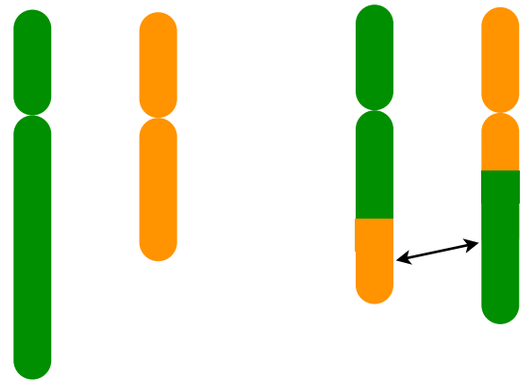
HIV is by no means the first plague that human populations have weathered. Many pathogens have deeply affected our evolutionary history. In fact, the human genome is littered with the remnants of our past battles with pathogens — and one of these remnants, a mutation to a gene called CCR5, may lead researchers to a new treatment for HIV.

The mutant CCR5 allele probably began to spread in northern Europe during the past 700 years when the population was ravaged by a plague. (It may have been bubonic plague or some other pathogen; research on this topic continues.) The mutant CCR5 probably made its bearers resistant to the disease, and so its frequency increased. We now know that the mutant CCR5 allele has an unexpected side effect: it confers resistance to HIV.

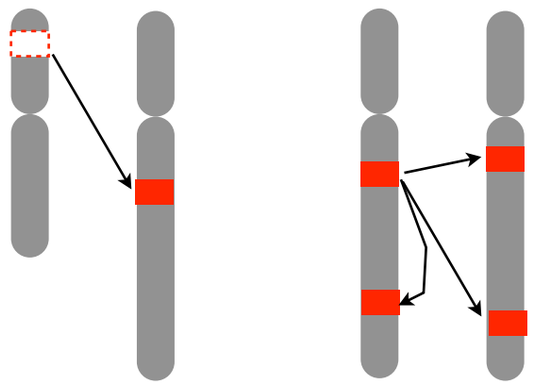
In an **inversion**, a piece of chromosome is lifted out, turned around and reinserted. If this includes the centromere then the inversion is termed pericentric. If it excludes the centromere then it is a paracentric inversion. The two have slightly different genetic consequences. 1% of the UK population are heterozygous for a pericentric inversion of chromosome 9. This is absolutely without genetic consequences.



In a **balanced translocation** there is no net gain or loss of chromosomal material, two chromosomes have been broken and rejoined in the wrong combination. The figure shows a translocation between the imaginary chromosomes. Balanced reciprocal translocation is unlikely to have any severe consequence for the cell because, even if one of the breakpoints lies within a gene, most mutations are recessive



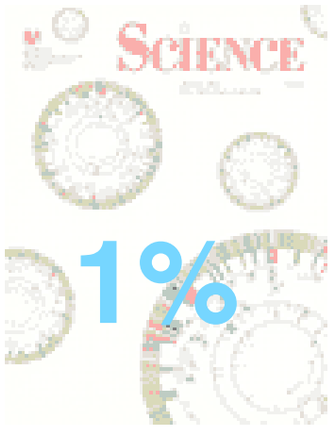
Transposable elements are sequences of DNA that can move around to different positions within the genome of a single cell, a process called (retro)transposition. In the process, they can cause mutations and change the amount of DNA in the genome. Transposons were also once called jumping genes, and are examples of mobile genetic elements. They were discovered by Barbara McClintock early in her career, for which she was awarded a Nobel prize in 1983.



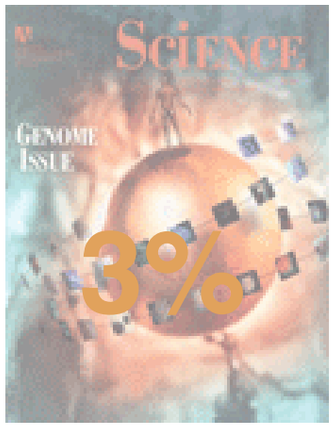
"The outstanding work in maize (review: McClintock, 1956) ... is probably very relevant here. Unfortunately I do not understand the details of this work well enough to put my finger on what may be particularly significant."

Guido Pontecorvo (1958) Trends in Genetic Analysis.

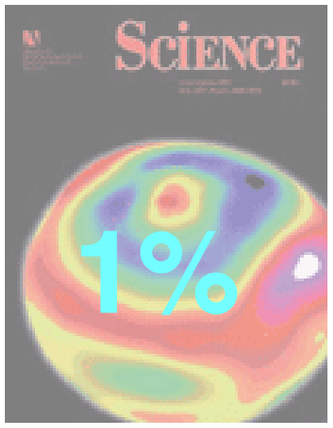
1995



1996



1997

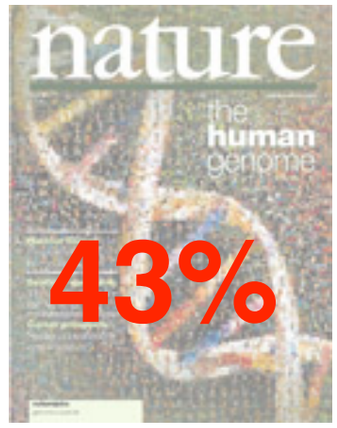


1998

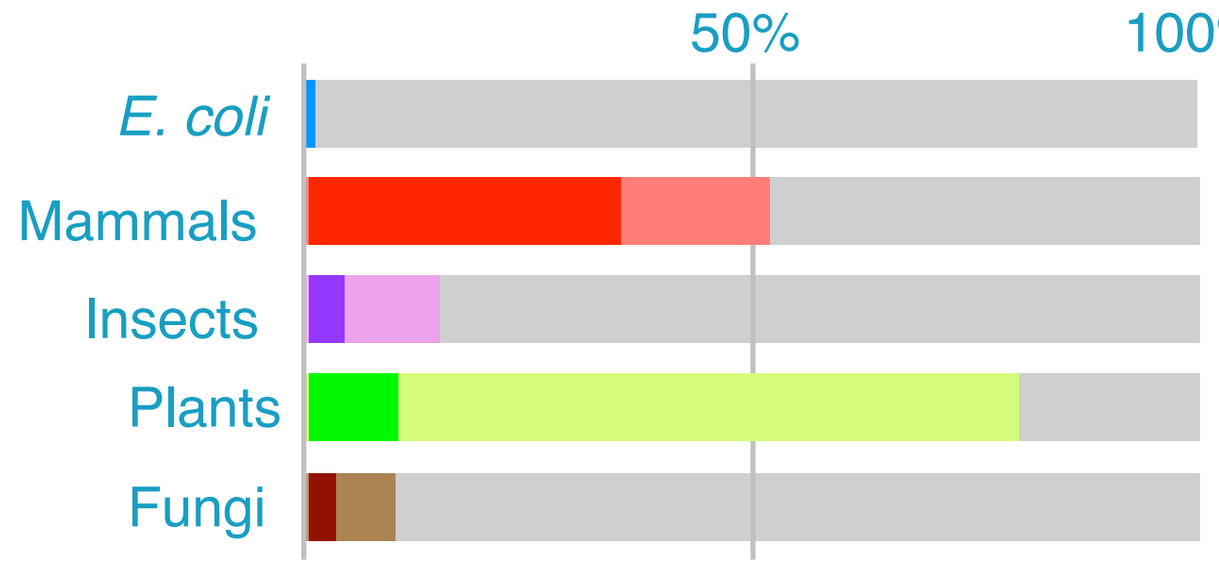
2000

2000

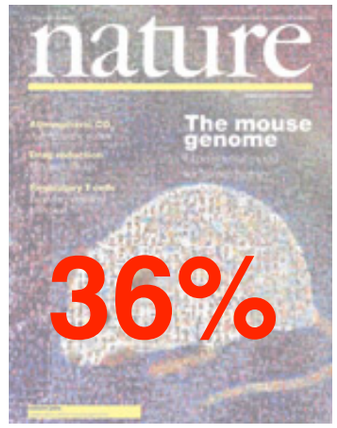
2001



2002



2002



2004

2004

2005

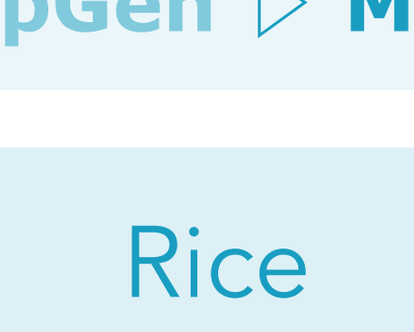
2005

2006

2007

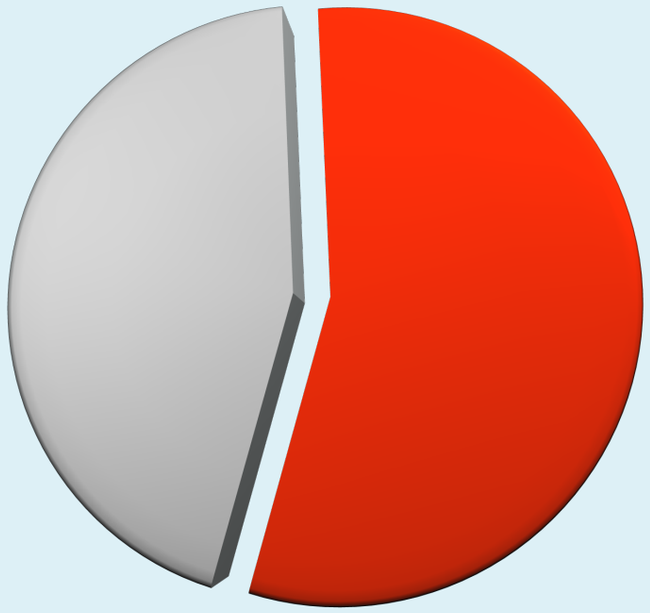
2007

Rice
430 MB



~14%

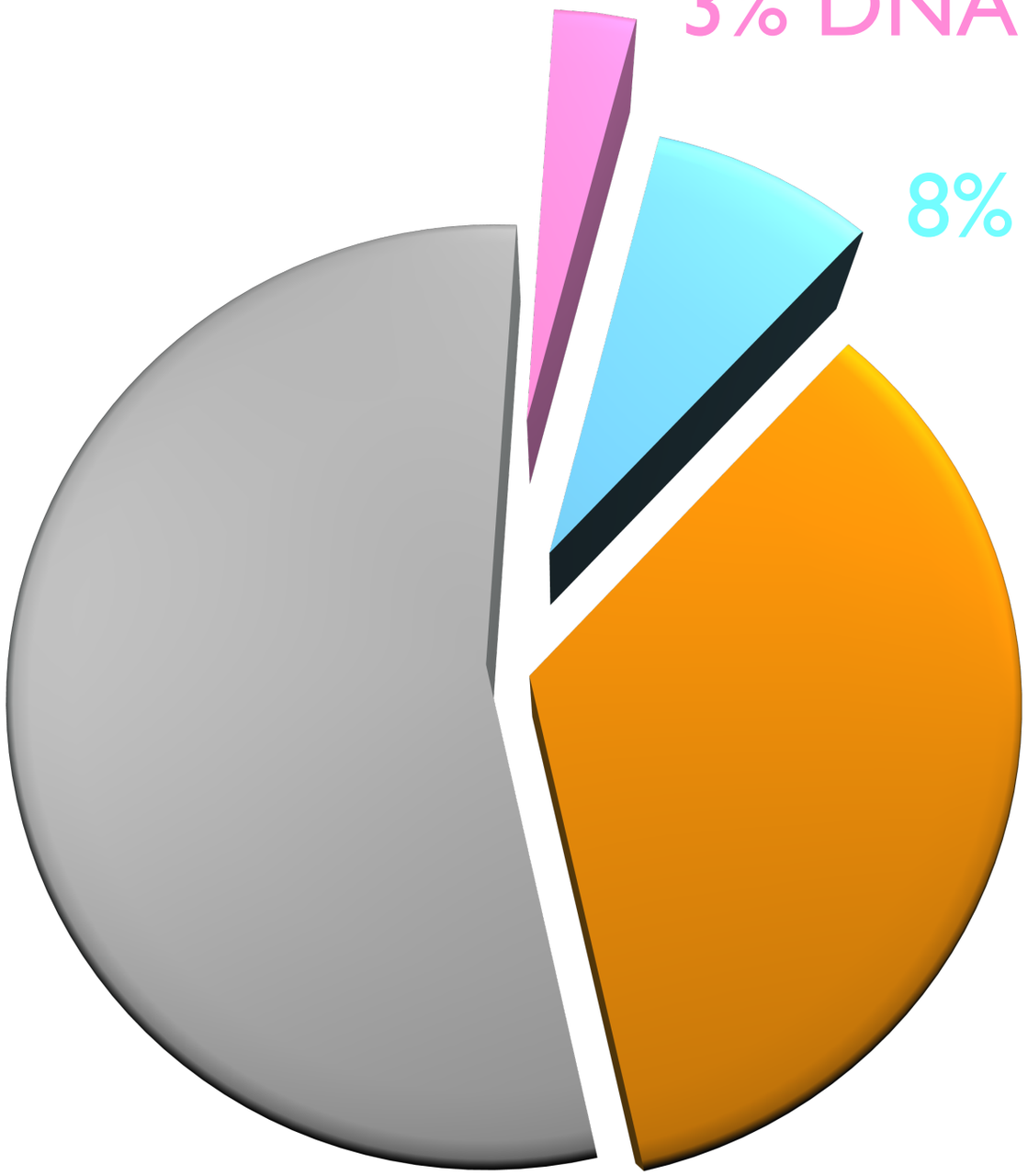
Maize
~2,500 MB



50-60%

Barley
~4,800 MB

>70%



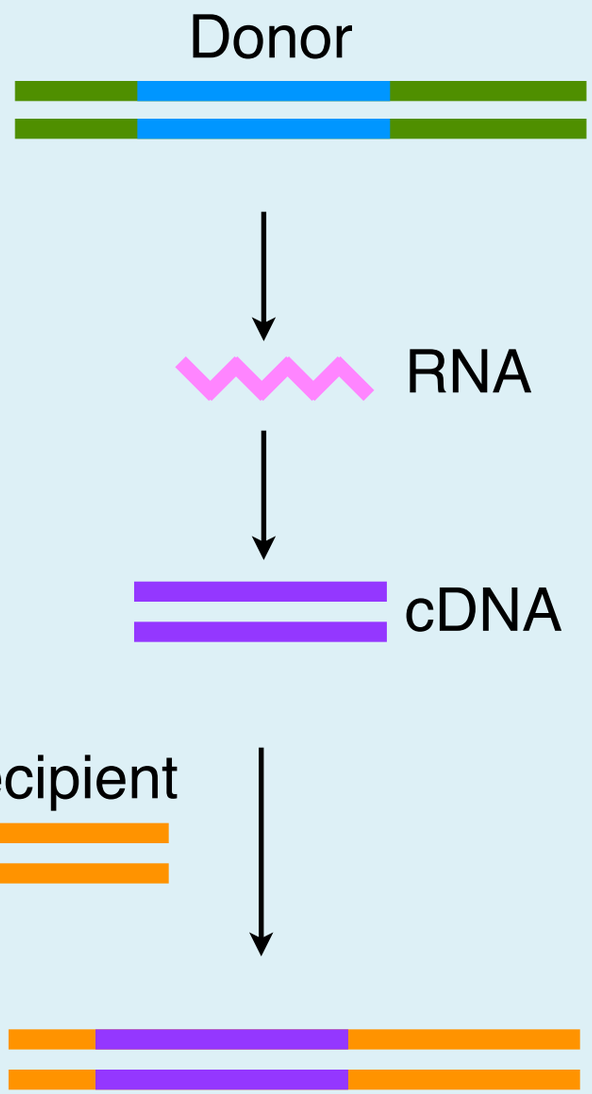
3% DNA Transposons

8% LTR

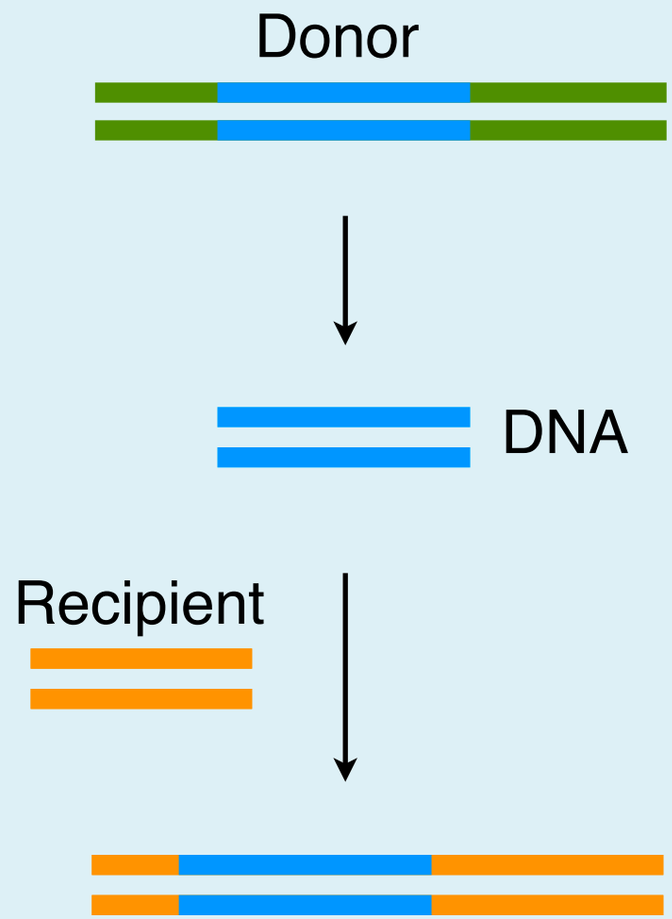
35% nonLTR { LINE: 21%
SINE: 14%

Total: 46%

Type I
Retrotransposons



Type II
DNA transposons



Proposed classification system for transposable elements (TEs)

Classification		Structure	TSD	Code	Occurrence
Order	**Superfamily**				
Class I (retrotransposons)					
LTR	*Copia*	GAG AP INT RT RH	4–6	RLC	P, M, F, O
	Gypsy	GAG AP RT RH INT	4–6	RLG	P, M, F, O
	Bel–Pao	GAG AP RT RH INT	4–6	RLB	M
	Retrovirus	GAG AP RT RH INT ENV	4–6	RLR	M
	ERV	GAG AP RT RH INT ENV	4–6	RLE	M
DIRS	*DIRS*	GAG AP RT RH YR	0	RYD	P, M, F, O
	Ngaro	GAG AP RT RH YR	0	RYN	M, F
	VIPER	GAG AP RT RH YR	0	RYV	O
PLE	*Penelope*	RT EN	Variable	RPP	P, M, F, O
LINE	*R2*	RT EN	Variable	RIR	M
	RTE	APE RT	Variable	RIT	M
	Jockey	ORF1 APE RT	Variable	RIJ	M
	L1	ORF1 APE RT	Variable	RIL	P, M, F, O
	I	ORF1 APE RT RH	Variable	RII	P, M, F
SINE	*tRNA*		Variable	RST	P, M, F
	7SL		Variable	RSL	P, M, F
	5S		Variable	RSS	M, O
Class II (DNA transposons) - Subclass 1					
TIR	*Tc1–Mariner*	Tase*	TA	DTT	P, M, F, O
	hAT	Tase*	8	DTA	P, M, F, O
	Mutator	Tase*	9–11	DTM	P, M, F, O
	Merlin	Tase*	8–9	DTE	M, O
	Transib	Tase*	5	DTR	M, F
	P	Tase	8	DTP	P, M
	PiggyBac	Tase	TTAA	DTB	M, O
	PIF–Harbinger	Tase* ORF2	3	DTH	P, M, F, O
	CACTA	Tase ORF2	2–3	DTC	P, M, F
Crypton	*Crypton*	YR	0	DYC	F
Class II (DNA transposons) - Subclass 2					
Helitron	*Helitron*	RPA Y2 HEL	0	DHH	P, M, F
Maverick	*Maverick*	C-INT ATP CYP POL B	6	DMM	M, F, O

Source: Wicker et atl. (2007) A unified classification system for eukaryotic transposable elements. Nature Reviews Genetics.

Structural features

Long terminal repeats · Terminal inverted repeats · Coding region · Non-coding region · Diagnostic feature in non-coding region · Region that can contain one or more additional ORFs

Protein coding domains

AP, Aspartic proteinase · APE, Apurinic endonuclease · ATP, Packaging ATPase · C-INT, C-integrase · CYP, Cysteine protease · EN, Endonuclease · ENV, Envelope protein · GAG, Capsid protein · HEL, Helicase · INT, Integrase · ORF, Open reading frame of unknown function · POL B, DNA polymerase B · RH, RNase H · RPA, Replication protein A (found only in plants) · RT, Reverse transcriptase · Tase, Transposase (* with DDE motif) · YR, Tyrosine recombinase · Y2, YR with YY motif

Species groups

P, Plants · M, Metazoans · F, Fungi · O, Others

Genome	RoundGoby_Haploid_Assembly_V2_160614_NEME.fasta							
Sequences	1'364							
GC level	41.60%							
Total length	1'003'738'563	bp						
Repeat type	class/family		Occurrence	Coverage [bp]	Coverage [%]	Total Coverage		
Simple repeats			307'646	16'115'088	1.606			
Tadem repeats*			122'350	16'325'172	1.626			
Satellite repeats			5'805	756'612	0.075			
Low_complexity			45'040	3'004'359	0.299			
rRNA			11'743	1'108'558	0.110			
Class I (retrotransposons)								
	LTR	Copia	311	301'375	0.030	0.892	6.019	
		Gypsy	9'729	6'021'751	0.600			
		Pao	2'624	2'254'480	0.225			
		ERV	2'602	373'700	0.037			
	DIRS	DIRS	1'033	437'708	0.044	0.123		
		Ngaro	8'163	801'119	0.080			
	PLE	Penelope	2'286	420'654	0.042	0.042		
	LINE	R2	479	214'481	0.021	4.468		
		RTE	44'496	10'125'330	1.009			
		L1	10'790	4'303'403	0.429			
		I	1'670	360'464	0.036			
		L2	92'660	24'236'084	2.415			
		Rex-Babar	27'235	5'609'430	0.559			
	SINE	tRNA	28'357	3'633'378	0.362	0.494		
		5S	19'392	1'322'156	0.132			
		L2	2	28	0.000			
		Retroposon	530	115'135	0.011			
Class II (DNA transposons) - Subclass 1								
	TIR	TcMar	52'612	17'561'202	1.750	3.058	3.079	
		hAT	100'002	8'613'300	0.858			
		MULE / Mutator	354	51'894	0.005			
		Merlin	3'383	256'287	0.026			
		P	172	42'157	0.004			
		PiggyBac	1'252	240'835	0.024			
		PIF-Harbinger	4652	610994	0.061			
		Dada	957	134'365	0.013			
		CMC-EnSpm	20301	1918104	0.191			
		Ginger	243	21'205	0.002			
		IS3EU	963	142'136	0.014			
		Kolobok	1110	161990	0.016			
		PIG	4'652	610'994	0.061			
		Sola	2'665	240'885	0.024			
		Zisupton	697	78'692	0.008			
		Academ	101	8915	0.001			
	Crypton	Crypton	2'131	214'193	0.021	0.021		
Class II (DNA transposons) - Subclass 2								
	Helitron	Helitron	4'405	710'090	0.071	0.071	0.126	
	Maverik	Maverick	5'269	550'018	0.055	0.055		
Unknown	Unknown	Unknown	3'779	804'242	0.080	0.080	0.080	

Note: Results are based on Repeat Masker output files. Classification is according to Wicker et al. (2007). For more details see report.
* Tandem Repeats Finder was used to estimate tadem repeats. (see report for more details)

NATURE|Vol 460|27 August 2009

DEVELOPMENTAL BIOLOGY

Jumping-gene roulette

Sandra L. Martin

Jumping genes, which make DNA copies of themselves through an RNA middleman, provide a stochastic process for generating brain diversity among humans. The effect of their random insertion, however, is a bit of a gamble.

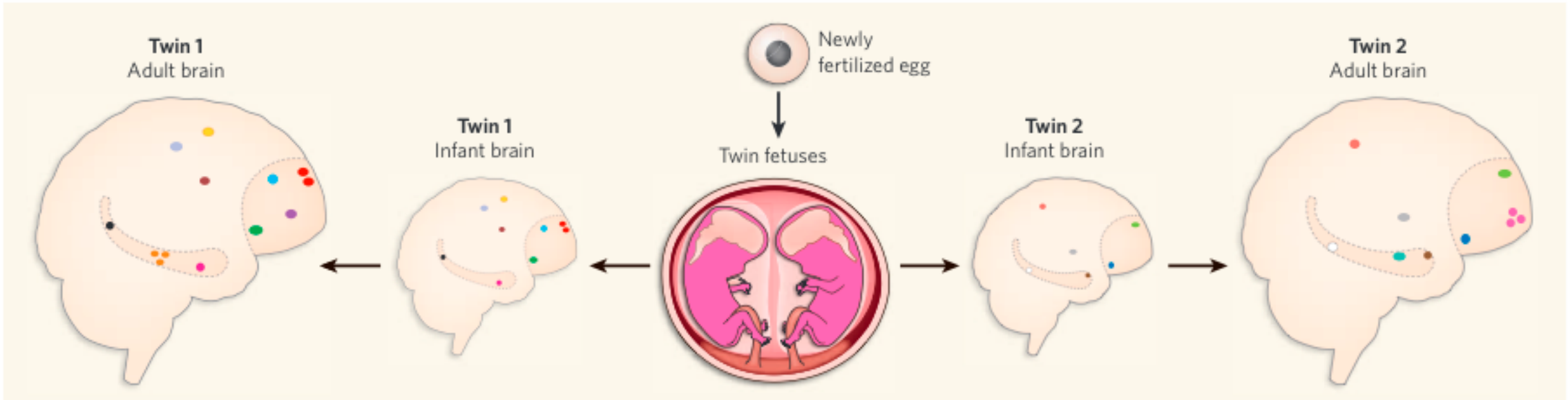
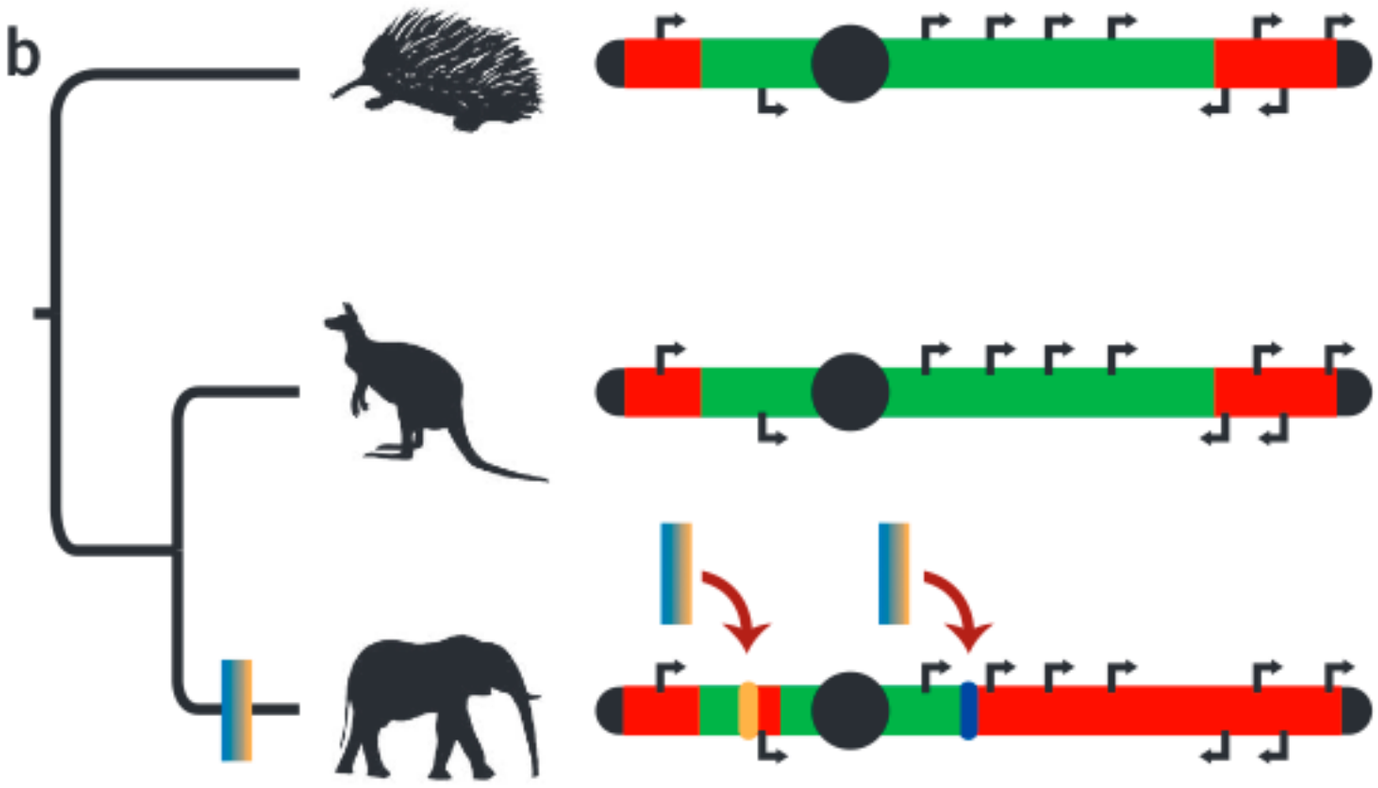


Figure 1 | Human brain variation by retrotransposition. These twins are genetically identical at conception, but at birth their brains differ because of new L1 insertions that take place during the development of the nervous system in the fetus. Ongoing retrotransposition in neural progenitor cells as shown to occur by Coufal *et al.*[1] will further diversify the genetic make-up of their brains in adulthood. Depending on the target genes and the neurons affected by L1 insertions, the twins may differ in brain function or dysfunction. Each unique insertion is represented by a different colour. Darker-shaded areas highlight regions of the brain where L1 retrotransposition may be more likely to occur after birth.

nature genetics

Transposon-mediated rewiring of gene regulatory networks contributed to the evolution of pregnancy in mammals

Vincent J Lynch, Robert D Leclerc, Gemma May & Günter P Wagner

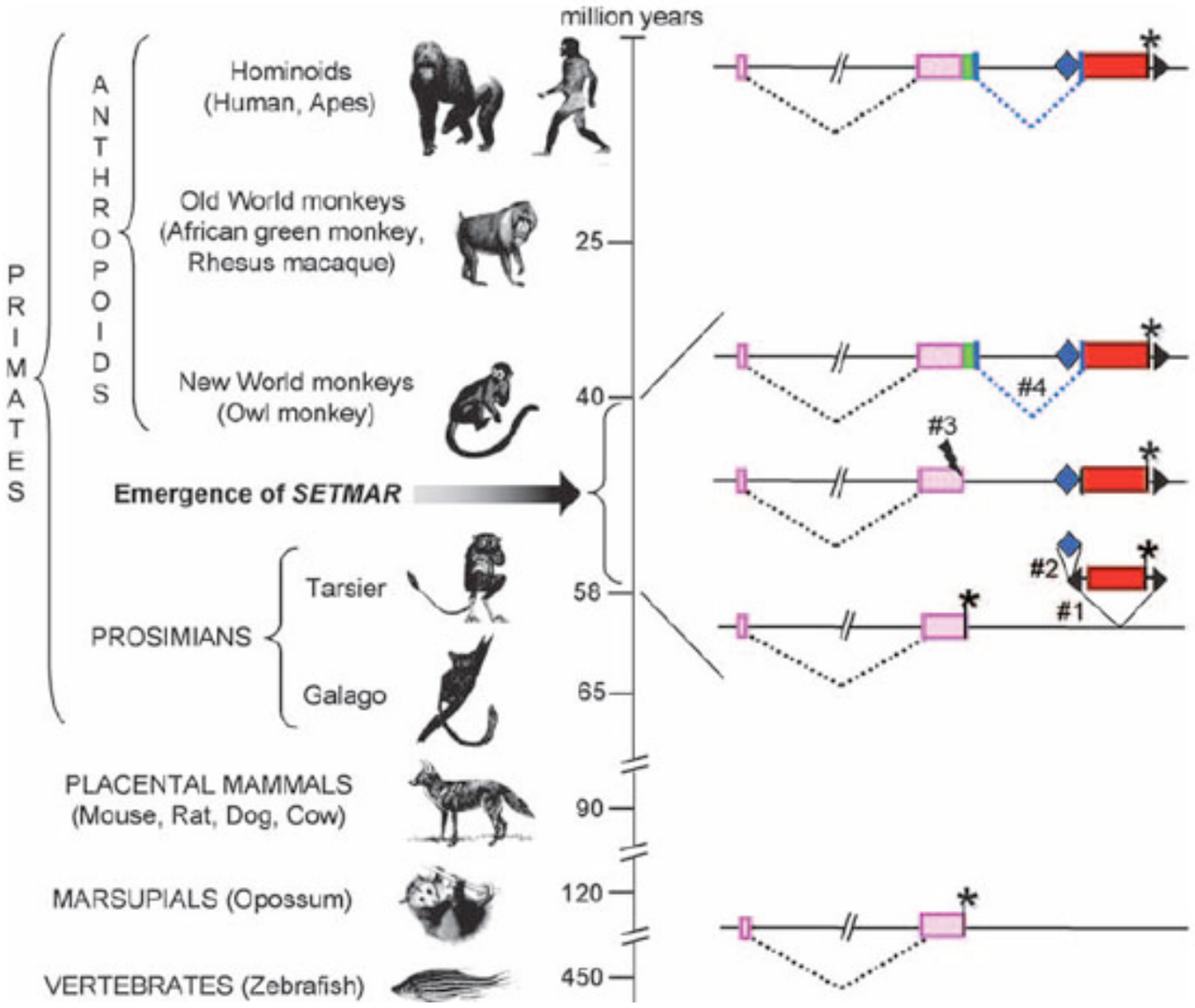


Model of gene regulatory rewiring by MER20s. Ancestrally, numerous genes (black arrows) were not expressed in endometrial stromal cells because they were repressed by epigenetic modifications of chromatin and direct silencing by transcriptional repressors. MER20s inserted into the genome in the placental mammal lineage (blue/yellow box on phylogeny), which prevented the spread of silent chromatin, establishing new borders between transcriptionally silent (green) and active (red) chromatin

Birth of a chimeric primate gene by capture of the transposase gene from a mobile element

Richard Cordaux*, Swalpa Udit†, Mark A. Batzer*, and Cédric Feschotte†‡

PNAS | 2006 | vol. 103 | no. 21 | 8101-8106



A gene captured from a mobile element (Hsmar1) fused with another gene (SET) to make a new primate gene (SETMAR).

#1: Hsmar1 transposon is inserted in the primate lineage

#2: secondary AluSx insertion within the TIR of Hsmar1

#3: deletion removing the stop codon of the SET gene

#4: de novo conversion from noncoding to exonic sequence is shown (green) and creation of the second intron (blue line)

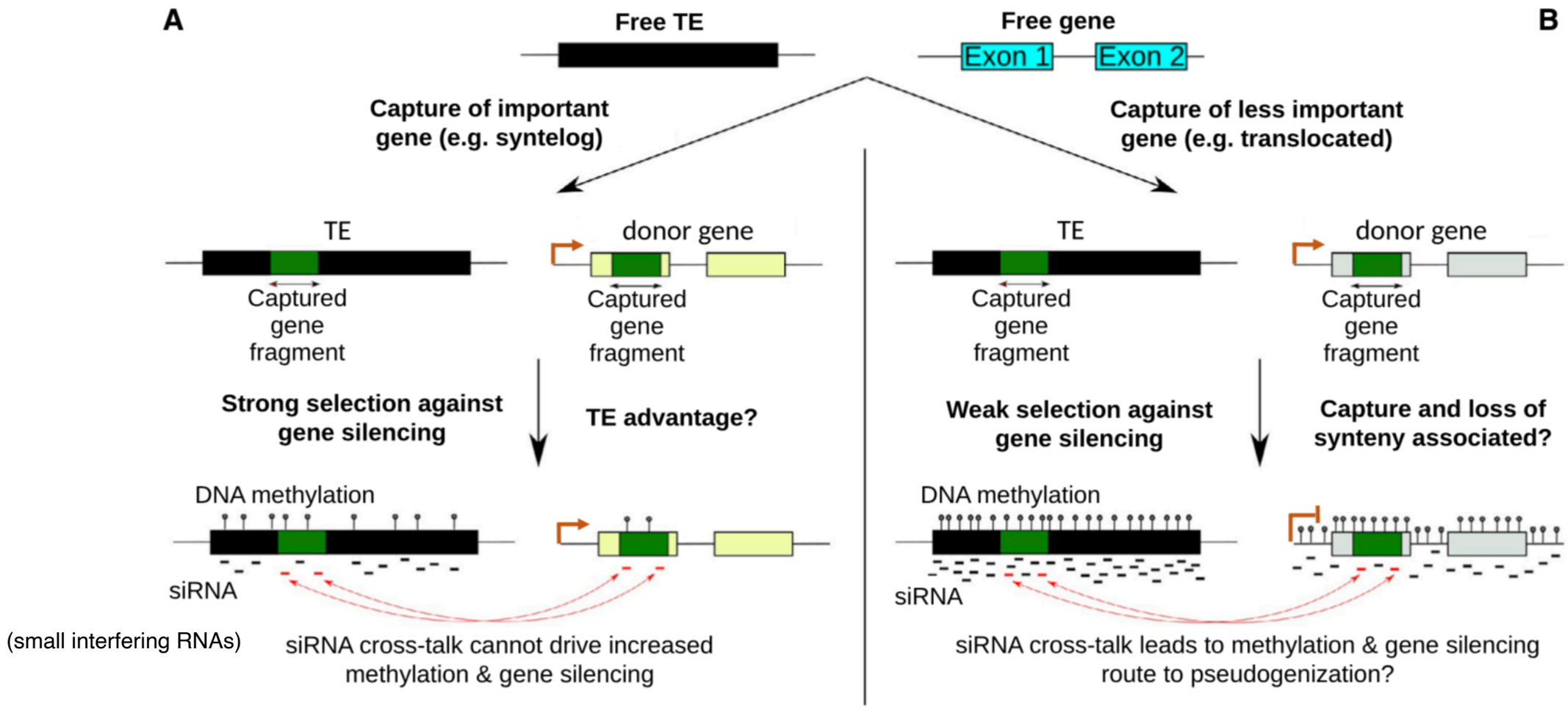


Figure 5. The Epigenetic Conflict Model of Gene Capture.

When TEs capture fragments of genes, siRNAs derived by the TEs may act in *trans* to accidentally mediate an epigenetic response against the gene, leading to increased methylation and reduced expression. The conflict comes from evolutionary pressure to silence TEs without simultaneously silencing functionally important genes, syntelogs in our example. As a result, (**A**) epigenetic effects on these genes are moderated by natural selection and expression is not affected. TEs may benefit from this moderation, although this remains unclear. In contrast, (**B**) for genes that are not under strong selective constraint, methylation can increase in the absence of conflict, leading to loss of expression and potential pseudogenization. This profile is characteristic of genes that have moved from their syntenic loci, which are overrepresented among donor genes, suggesting that capture may trigger movement.

Source: Muyle et al. (2021). Gene capture by transposable elements leads to epigenetic conflict in maize. Mol. Plant. 14, 237–252.