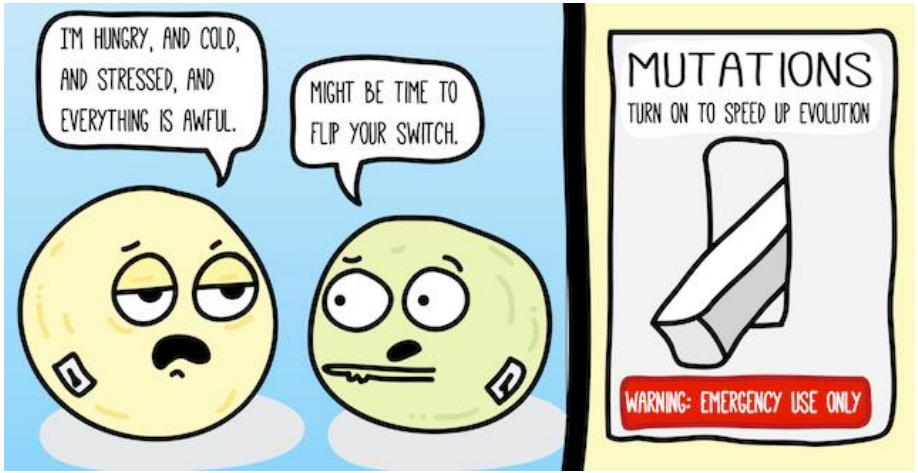
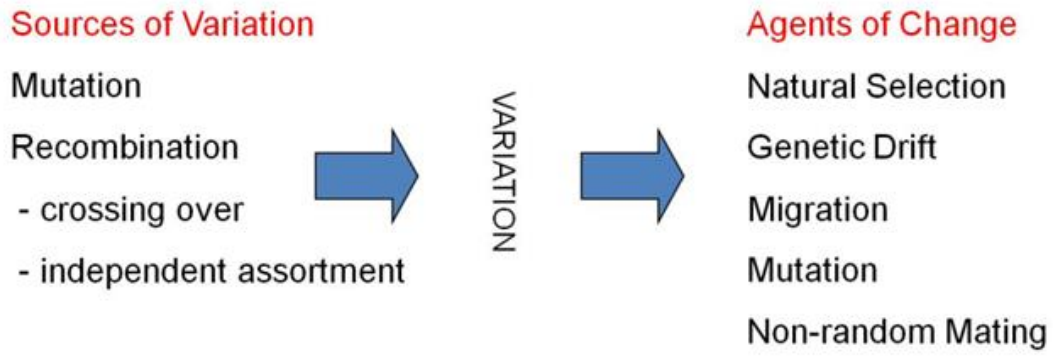
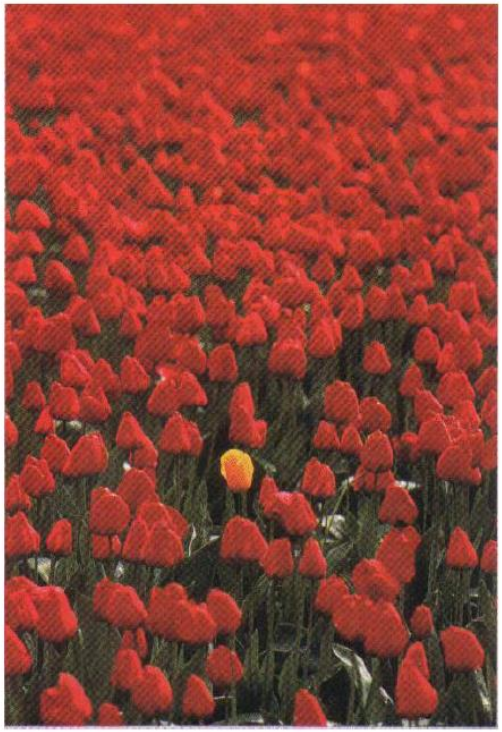
PR_7 Evolucija genomov (Genome evolution)



Molekularni vzroki evolucije genomov



Mutations are the engines of variation



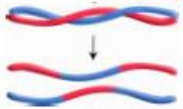
Without new variations, evolution would come to a halt:
all individuals would become the same

Microevolutionary Processes

Generating variation

Mutation: generates genetic variation

Recombination: generates novel combinations of genetic variation



Sorting variation

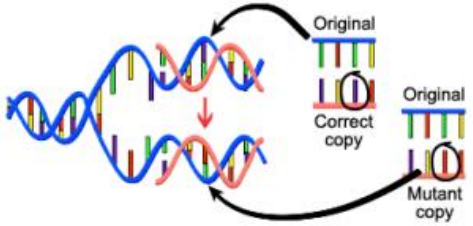
Drift: neutral process

Selection: non-neutral process

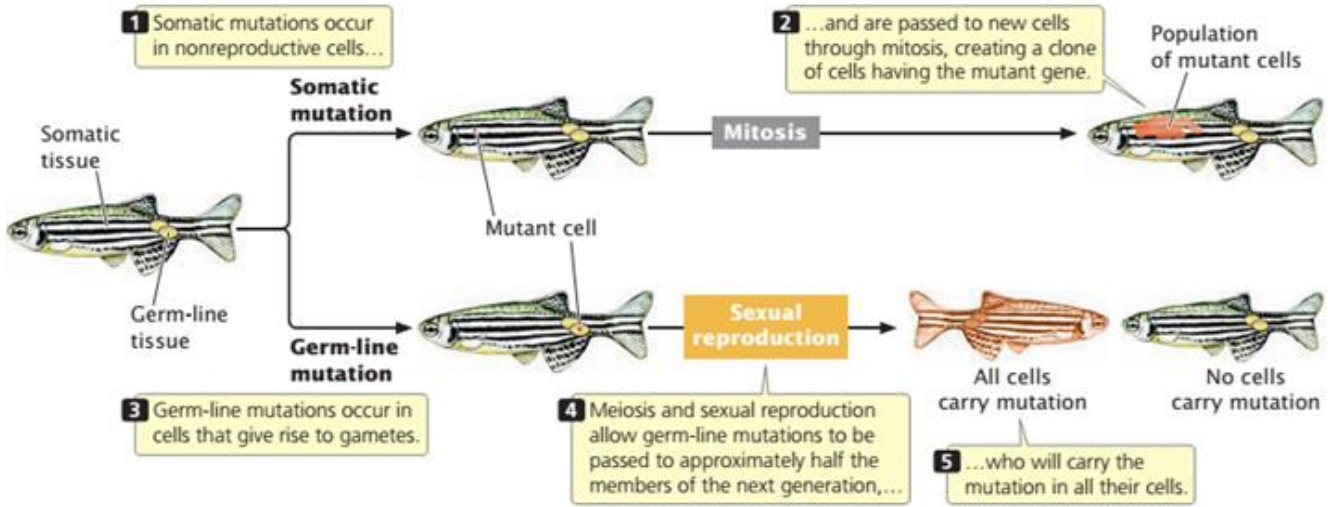
Mutation μ

Mutations: errors in DNA replication or repair

Provide the raw material on which evolution acts.



Mutations can affect a single nucleotide (point mutation) or longer lengths of sequence (segmental mutation).



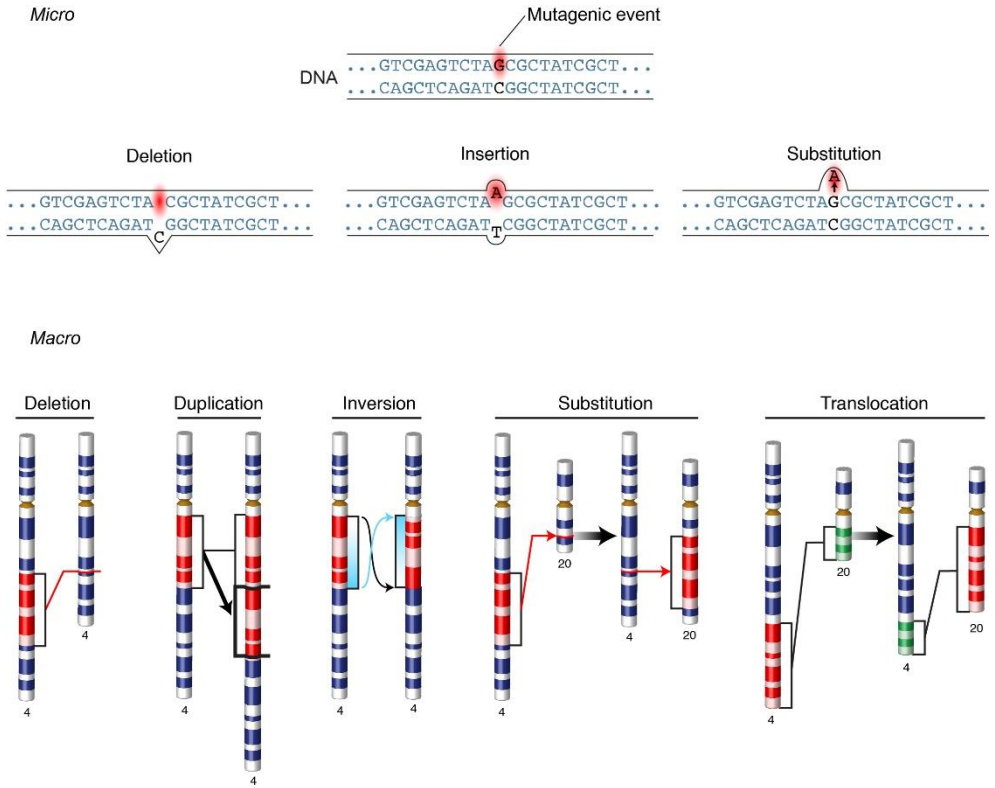
18.1 The two basic classes of mutations are somatic mutations and germ-line mutations.

Which mutations matter for evolution?

Only mutations in the germ line matter for evolution - those in the soma do not.

For example, environmentally induced mutations in human skin will not affect the next generation.

- **Mutation and sexual recombination produce the variation that makes evolution possible**
- Two processes, mutation and sexual recombination
  - Produce the variation in gene pools that contributes to differences among individuals

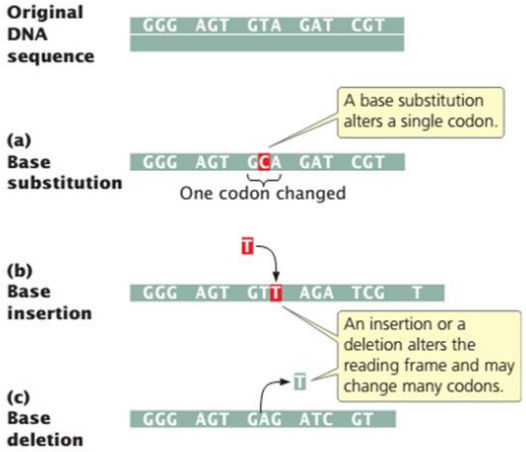


A mutation is a change in a DNA sequence

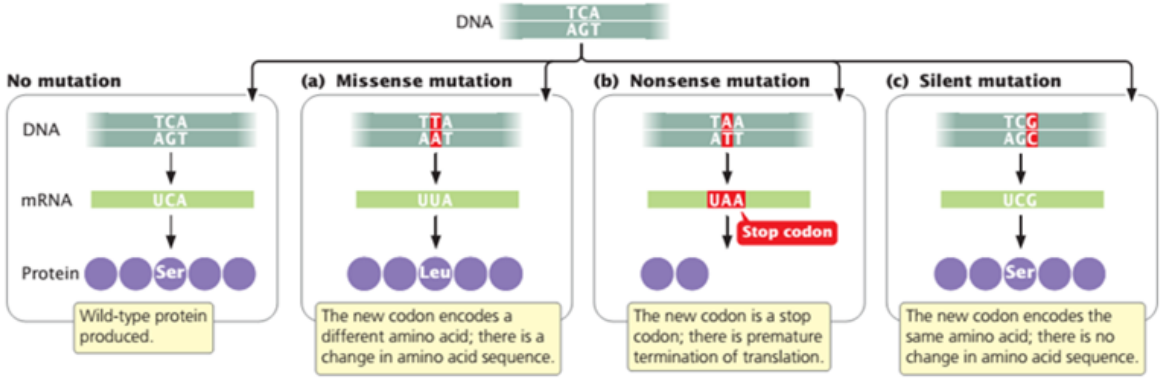
Mutations can result from DNA copying mistakes made during cell division, exposure to ionizing radiation, exposure to chemicals called mutagens, or infection by viruses. Germ line mutations occur in the eggs and sperm and can be passed on to offspring, while somatic mutations occur in body cells and are not passed on.

Table 2 Mutation Types

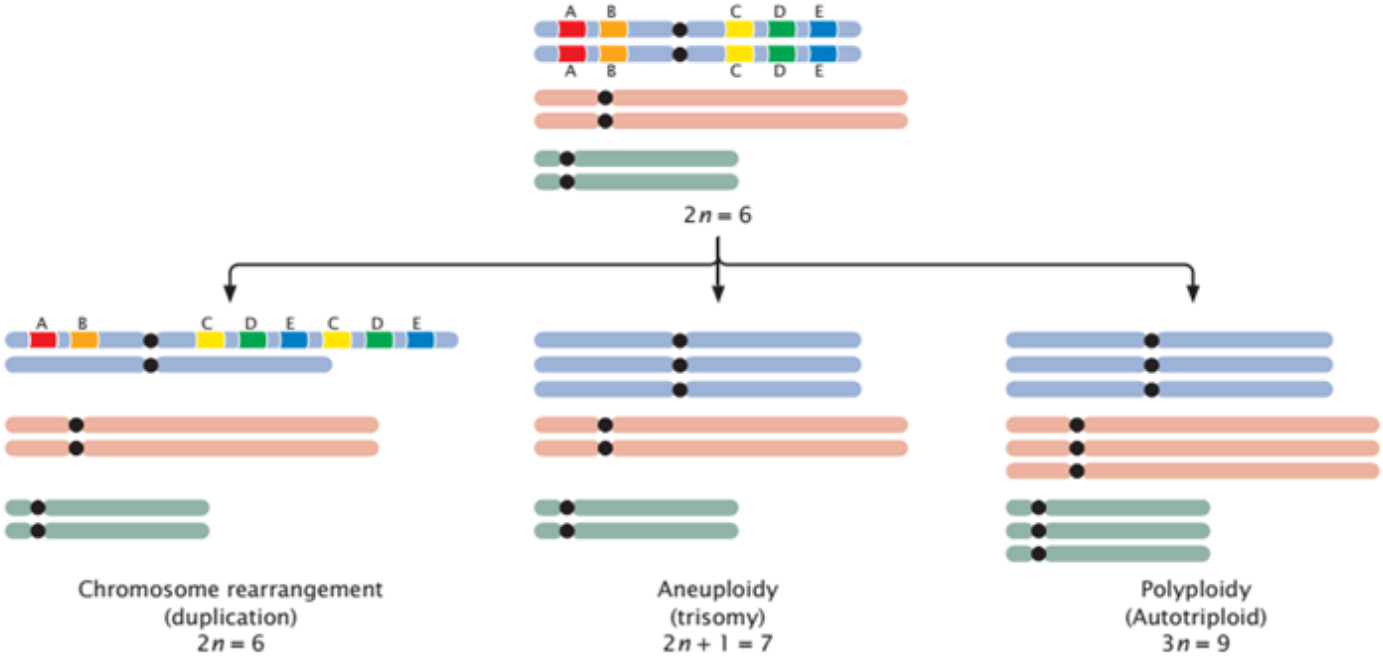
Small-Scale Mutations	
Point mutation	Change in a single nucleotide
Silent mutation	Amino acid sequence is not changed
Missense mutation	Amino acid sequence is changed
Nonsense mutation	Amino acid sequence is changed to a stop codon, therefore truncating the protein
Insertion	Addition of one or more extra nucleotides
Deletion	Removal of one or more nucleotides
Large-Scale Mutations	
Amplification	Multiple copies of chromosomal region
Deletion	Loss of chromosomal region
Translocation	Interchange of regions from different chromosomes
Inversion	Reversal of the orientation of a chromosomal region
Loss of heterozygosity	Loss of one allele



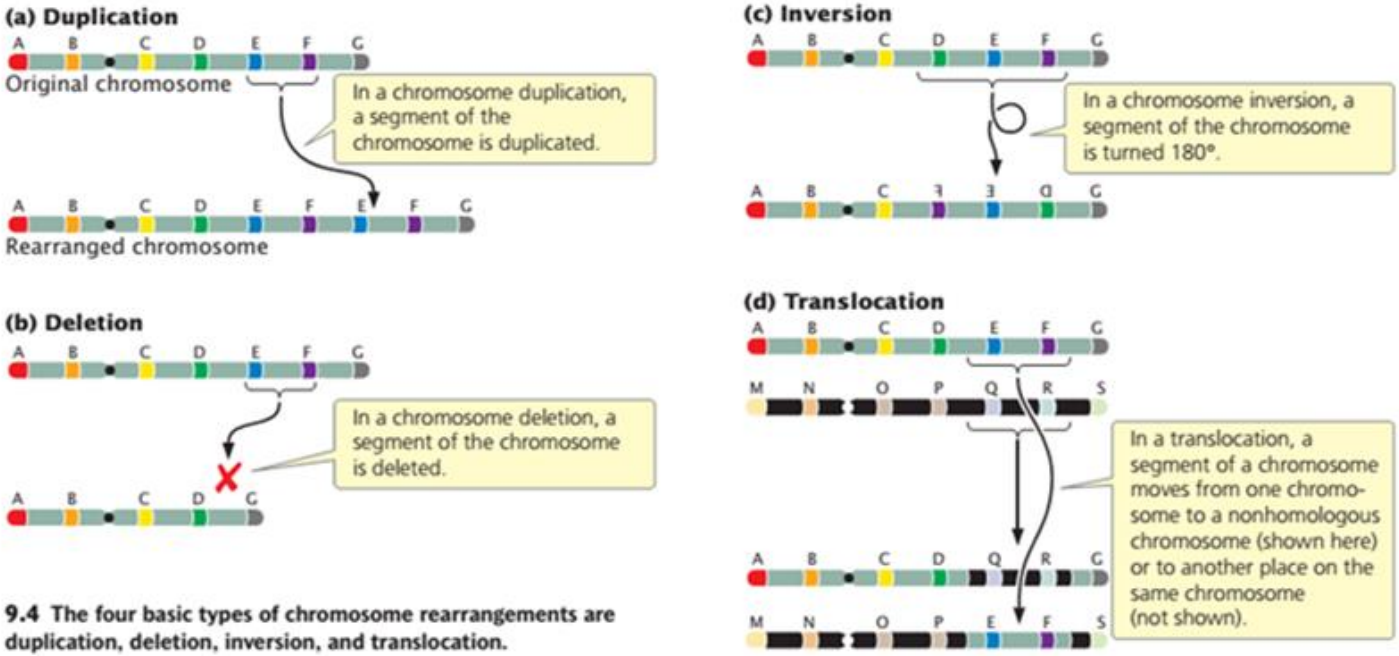
18.2 Three basic types of gene mutations are base substitutions, insertions, and deletions.



18.6 Base substitutions can cause (a) missense, (b) nonsense, and (c) silent mutations.



9.3 Chromosome mutations consist of chromosome rearrangements, aneuploids, and polyploids.
Duplications, trisomy, and autotriploids are examples of each category of mutation.



9.4 The four basic types of chromosome rearrangements are duplication, deletion, inversion, and translocation.

Types of mutation

1) Substitutions

2) Recombinations

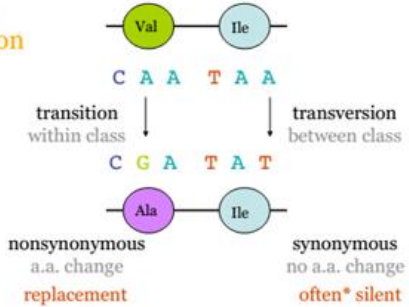
3) Deletions

4) Insertions

5) Inversions

Types of mutation

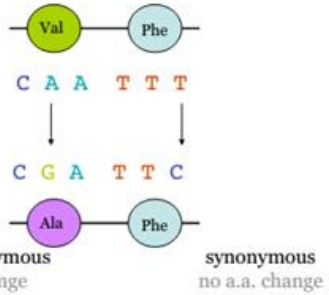
1) Substitutions



* A synonymous substitution is not always silent because can change splice sites

Types of mutation

1) Substitutions

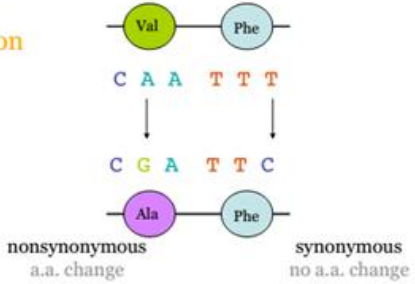


Position in codon matters:

1st position: 96% nonsynonymous
2nd position: 100% nonsynonymous
3rd position: 30% nonsynonymous

Types of mutation

1) Substitutions



How do substitutions happen?

Generally due to mispairing of bases during DNA replication.

Types of mutation

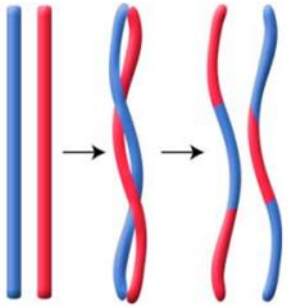
1) Substitutions

2) Recombinations

3) Deletions

4) Insertions

5) Inversions



Types of mutation

1) Substitutions

2) Recombinations

3) Deletions:

4) Insertions:

5) Inversions

Addition or removal of a stretch of sequence (can be a single or thousands of nucleotides).

Collectively referred to as indels

Determining whether insertion or deletion can be tricky.

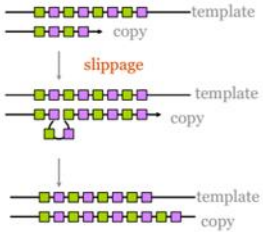
A C G A T T C T C G G
A C G A : T C T C G G
A C G A : T C T C G G
A C G A : T C T C G G

Types of mutation

Indels are more common in repetitive sequences.

3) Deletions

4) Insertions



Can be caused by:

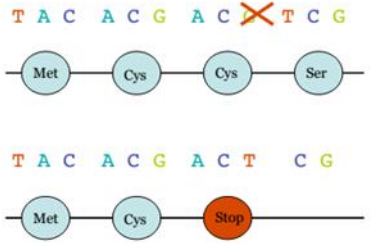
- Unequal crossing over
- Intrastrand deletion
- Replication slippage
- Transposition

Types of mutation

Indels can cause frameshift mutations and cause major alterations to proteins.

3) Deletions

4) Insertions



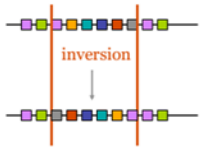
Types of mutation

1) Substitutions

2) Recombinations

3) Deletions

4) Insertions



5) Inversions: end to end chromosomal rearrangement (usually long)

Can be caused by:

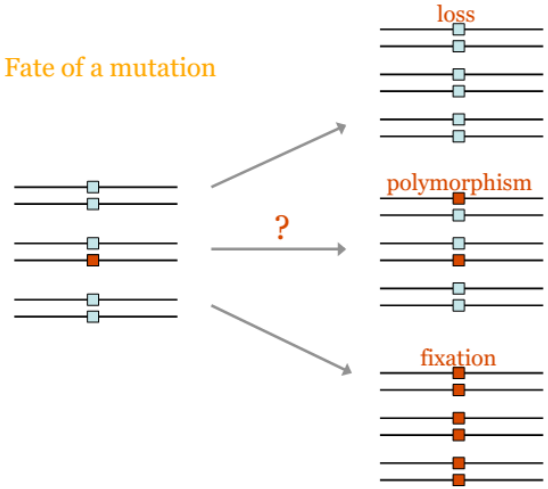
- Chromosome breakage and rejoining
- Intrachromosomal crossing over

Are mutations random?

Mutations are not random with respect to rate.
Mutation rates vary among species, organelles, **genomic locations**, sexes and environmental conditions.

Mutation Hotspots: regions of the genome prone to mutation

Repetitive sequences
Transitions more common than transversions (but not universally so)
Particular nucleotides are more mutable
(e.g., G and C in mammalian genomes)

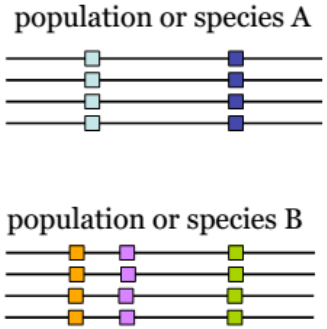


Fate of a mutation

Generating variation
- Mutation
- Recombination

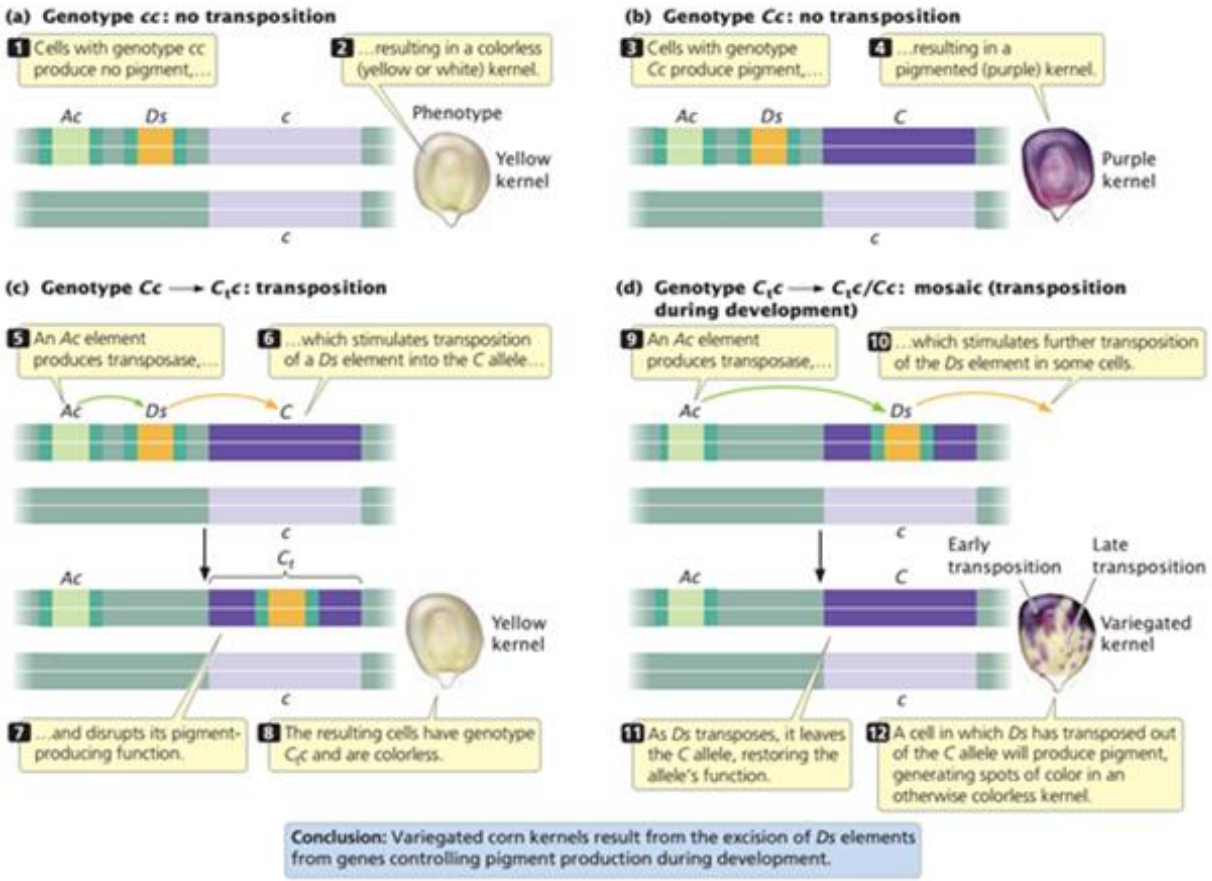
Redistributing variation
- Gene flow
- Hybridization
- Horizontal gene transfer

Sorting variation
- Drift
- Selection



Tomatoes come in a great variety of shapes, sizes, and colors. Elongated tomoatoes in some varieties are produced by a duplication that arose as a result of the presence of the transposable element *Rider*. [Photolibrary.]

11.23 Variegated (multicolored) kernels in corn are caused by mobile genes. The study of variegated corn led Barbara McClintock to discover transposable elements. [Matt Meadows/Peter Arnold.]



11.25 Transposition results in variegated maize kernels.

Table 18.2 Characteristics of different types of mutations

Type of Mutation	Definition
Base substitution	Changes the base of a single DNA nucleotide
Transition	Base substitution in which a purine replaces a purine or a pyrimidine replaces a pyrimidine
Transversion	Base substitution in which a purine replaces a pyrimidine or a pyrimidine replaces a purine
Insertion	Addition of one or more nucleotides
Deletion	Deletion of one or more nucleotides
Frameshift mutation	Insertion or deletion that alters the reading frame of a gene
In-frame deletion or insertion	Deletion or insertion of a multiple of three nucleotides that does not alter the reading frame
Expanding nucleotide repeats	Repeated sequence of a set of nucleotides in which the number of copies of the sequence increases
Forward mutation	Changes the wild-type phenotype to a mutant phenotype
Reverse mutation	Changes a mutant phenotype back to the wild-type phenotype
Missense mutation	Changes a sense codon into a different sense codon, resulting in the incorporation of a different amino acid in the protein
Nonsense mutation	Changes a sense codon into a nonsense codon, causing premature termination of translation
Silent mutation	Changes a sense codon into a synonymous codon, leaving unchanged the amino acid sequence of the protein
Neutral mutation	Changes the amino acid sequence of a protein without altering its ability to function
Loss-of-function mutation	Causes a complete or partial loss of function
Gain-of-function mutation	Causes the appearance of a new trait or function or causes the appearance of a trait in inappropriate tissue or at an inappropriate time
Lethal mutation	Causes premature death
Suppressor mutation	Suppresses the effect of an earlier mutation at a different site
Intragenic suppressor mutation	Suppresses the effect of an earlier mutation within the same gene
Intergenic suppressor mutation	Suppresses the effect of an earlier mutation in another gene

TABLE 16.4
Causes of Mutations

Common Causes of Mutations	Description
Spontaneous	
Aberrant recombination	Abnormal crossing over may cause deletions, duplications, translocations, and inversions (see Chapter 8).
Aberrant segregation	Abnormal chromosomal segregation may cause aneuploidy or polyploidy (see Chapter 8).
Errors in DNA replication	A mistake by DNA polymerase may cause a point mutation (see Chapter 11).
Transposable elements	Transposable elements can insert themselves into the sequence of a gene (see Chapter 17).
Depurination	On rare occasions, the linkage between purines (i.e., adenine and guanine) and deoxyribose can spontaneously break. If not repaired, it can lead to mutation.
Deamination	Cytosine and 5-methylcytosine can spontaneously deaminate to create uracil or thymine.
Tautomeric shifts	Spontaneous changes in base structure can cause mutations if they occur immediately prior to DNA replication.
Toxic metabolic products	The products of normal metabolic processes, such as reactive oxygen species, may be chemically reactive agents that can alter the structure of DNA.
Induced	
Chemical agents	Chemical substances may cause changes in the structure of DNA.
Physical agents	Physical phenomena such as UV light and X-rays can damage the DNA.

TABLE 16.1

Consequences of Point Mutations Within the Coding Sequence

Type of Change	Mutation in the DNA	Example*	Amino Acids Altered	Likely Effect on Protein Function
None	None	5′–A–T–G–A–C–C–G–A–C–C–C–G–A–A–A–G–G–G–A–C–C–3′ Met – Thr – Asp – Pro – Lys – Gly – Thr –	None	None
Silent	Base substitution	5′–A–T–G–A–C–C–G–A–C–C–C–C–A–A–A–G–G–G–A–C–C–3′ Met – Thr – Asp – Pro – Lys – Gly – Thr –	None	None
Missense	Base substitution	5′–A–T–G–C–C–C–G–A–C–C–C–G–A–A–A–G–G–G–A–C–C–3′ Met – Pro – Asp – Pro – Lys – Gly – Thr –	One	Neutral or inhibitory
Nonsense	Base substitution	5′–A–T–G–A–C–C–G–A–C–C–C–G–T–A–A–G–G–G–A–C–C–3′ Met – Thr – Asp – Pro – STOP!	Many	Inhibitory
Frameshift	Addition/deletion	5′–A–T–G–A–C–C–G–A–C–G–C–C–G–A–A–A–G–G–G–A–C–C–3′ Met – Thr – Asp – Ala – Glu – Arg – Asp –	Many	Inhibitory

*DNA sequence in the coding strand. Note that this sequence is the same as the mRNA sequence except that the RNA contains uracil (U) instead of thymine (T). The 3-base codons are shown in alternating black and red colors. Mutations are shown in green.

Substitutions in coding sequences

Synonymous substitution	Nonsynonymous substitution
Val GTC ⇩ GTA Val No fitness effects	Val GTC ⇩ TTC Phe Fitness effects

K_A / K_S

K_A: nonsynonymous substitution per nonsynonymous site

K_S: synonymous substitutions per synonymous site

$K_A / K_S < 1$ **Purifying selection**

$K_A / K_S = 1$ **Neutral (Pseudogene) evolution**

$K_A / K_S > 1$ **Positive selection**

Mis-sense mutations and synonymous mutations

A transversion from an A to a T results in the substitution of glutamic acid to valine in the amino acid sequence of the beta-globin gene.

This is known as a mis-sense mutation.
In this particular case it leads to a disease called sickle cell anemia.

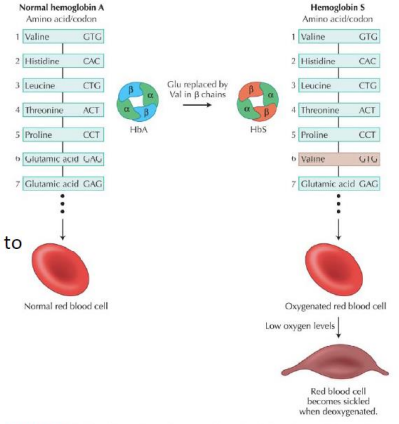
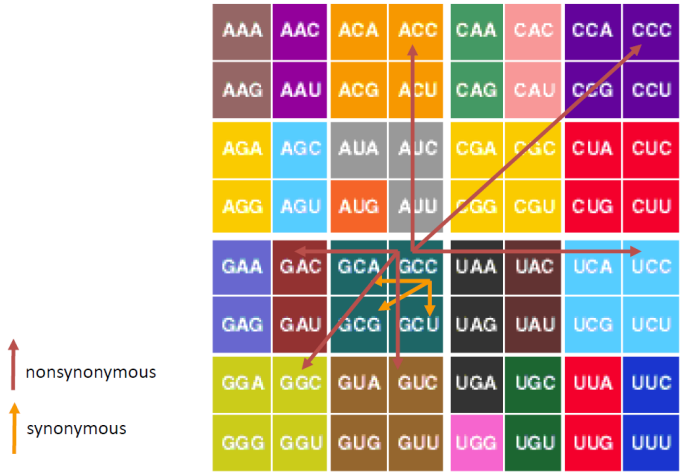
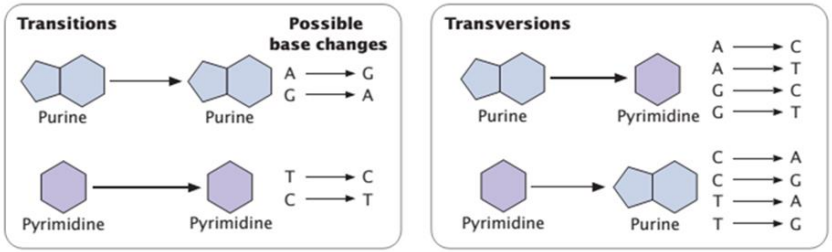


FIGURE 12.1. Sickle-cell anemia can be caused by a single-base change in the coding sequence of the gene that encodes one of the proteins (hemoglobin A) that makes up hemoglobin, resulting in the glutamic acid to valine substitution at the 6th amino acid position in the mutant-form hemoglobin S in sickle-cell disease. A schematic diagram of this is shown in which the normal β subunit (in *blue*) is present in the normal hemoglobin (HbA) and an abnormal β subunit (in *red*) is in the sickle-cell form of hemoglobin (HbS). When present, the altered hemoglobin molecule causes red blood cells to occasionally form a sickle shape, especially when under low oxygen levels. This in turn can lead to major phenotypic consequences such as the blocking of small blood vessels.

Synonymous mutations
due to the degeneracy of the genetic code

Because of the redundancy of the genetic code, point mutations may not change the amino acid.
These are called synonymous mutations





18.3 A transition is the substitution of a purine for a purine or of a pyrimidine for a pyrimidine; a transversion is the substitution of a pyrimidine for a purine or of a purine for a pyrimidine.

TABLE **16.2**

Possible Consequences of Gene Mutations Outside of the Coding Sequence

Sequence	Effect of Mutation
Promoter	May increase or decrease the rate of transcription
Regulatory element/operator site	May disrupt the ability of the gene to be properly regulated
5′-UTR/3′-UTR	May alter the ability of mRNA to be translated; may alter mRNA stability
Splice recognition sequence	May alter the ability of pre-mRNA to be properly spliced

16.1 Consequences of Mutation

- A point mutation is a change in a single base pair. These can be transitions or transversions.
- Silent, missense, nonsense, and frameshift mutations may occur within the coding region of a gene (see Table 16.1, Figure 16.1).
- Mutations may also occur within noncoding regions of a gene and affect gene expression (see Table 16.2).
- Suppressor mutations reverse the phenotypic effects of another mutation. They can be intragenic or intergenic (see Table 16.3).
- Changes in chromosome structure can have a position effect that alters gene expression (see Figures 16.2, 16.3).
- With regard to timing, mutations can occur in the germ line or in somatic cells (see Figure 16.4).

TABLE 24.1
Factors That Govern Microevolution

Source of New Allelic Variation*

Mutation	In this section, we consider allelic variation. Random mutations within preexisting genes introduce new alleles into populations, but at a very low rate. New mutations may be beneficial, neutral, or deleterious. For new alleles to rise to a significant percentage in a population, evolutionary mechanisms (i.e., random genetic drift, migration, natural selection) must operate on them.

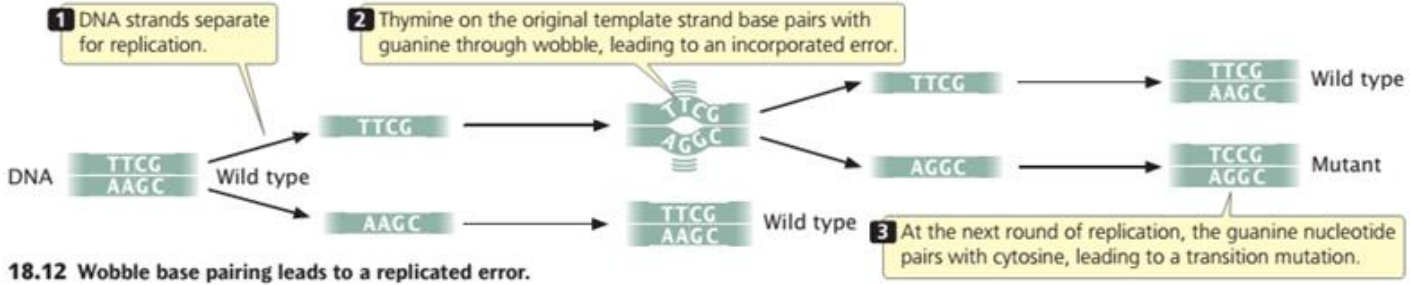
Mechanisms That Alter Existing Genetic Variation

Random genetic drift	This is a change in genetic variation from generation to generation due to random sampling error. Allele frequencies may change as a matter of chance from one generation to the next. This tends to have a greater effect in a small population.
Migration	Migration can occur between two different populations. The introduction of migrants into a recipient population may change the allele frequencies of that population.
Natural selection	This is the phenomenon in which the environment selects for individuals that possess certain traits that favor reproductive success. For example, natural selection may be related to the survival of members to reproductive age.
Nonrandom mating	This is the phenomenon in which individuals select mates based on their phenotypes or genetic lineage. This can alter the relative proportion of homozygotes and heterozygotes predicted by the Hardy-Weinberg equation but does not change allele frequencies.

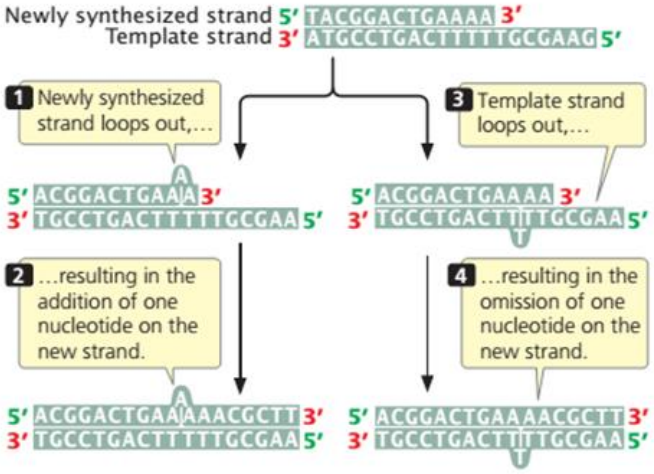
*Allelic variation is just one source of new genetic variation. Section 24.3 considers a variety of mechanisms through which new genetic variation can occur.

TABLE 24.2
Sources of New Genetic Variation That Occur in Populations

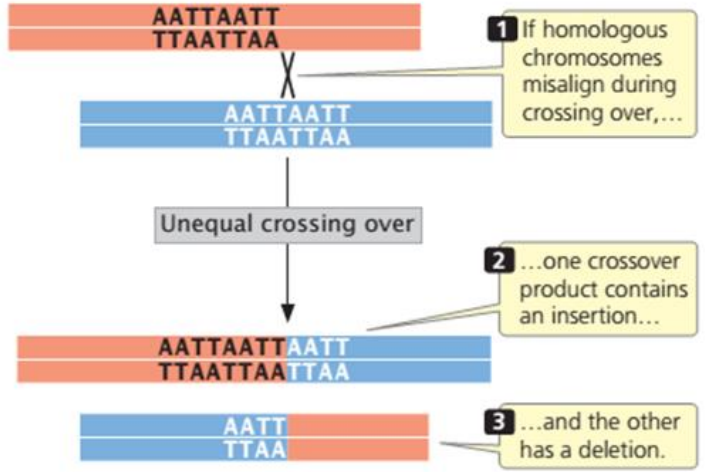
Type	Description
Independent assortment	The independent segregation of different homologous chromosomes may give rise to new combinations of alleles in offspring (see Chapter 3).
Crossing over	Recombination (crossing over) between homologous chromosomes can also produce new combinations of alleles that are located on the same chromosome (see Chapter 6).
Interspecies crosses	On occasion, members of different species may breed with each other to produce hybrid offspring. This topic is discussed in Chapter 26.
Prokaryotic gene transfer	Prokaryotic species possess mechanisms of genetic transfer such as conjugation, transduction, and transformation (see Chapter 7).
New alleles	Point mutations can occur within a gene to create single-nucleotide polymorphisms (SNPs). In addition, genes can be altered by small deletions and additions. Gene mutations are also discussed in Chapter 16.
Gene duplications	Events, such as misaligned crossovers, can add additional copies of a gene into a genome and lead to the formation of gene families. This topic is discussed in Chapter 8.
Chromosome structure and number	Chromosome structure may be changed by deletions, duplications, inversions, and translocations. Changes in chromosome number result in aneuploid, polyploid, and alloploid offspring. These mechanisms are discussed in Chapters 8 and 26.
Exon shuffling	New genes can be created when exons of preexisting genes are rearranged to make a gene that encodes a protein with a new combination of protein domains.
Horizontal gene transfer	Genes from one species can be introduced into another species and become incorporated into that species' genome.
Changes in repetitive sequences	Short repetitive sequences are common in genomes due to the occurrence of transposable elements and due to tandem arrays. The number and lengths of repetitive sequences tend to show considerable variation in natural populations.



18.12 Wobble base pairing leads to a replicated error.



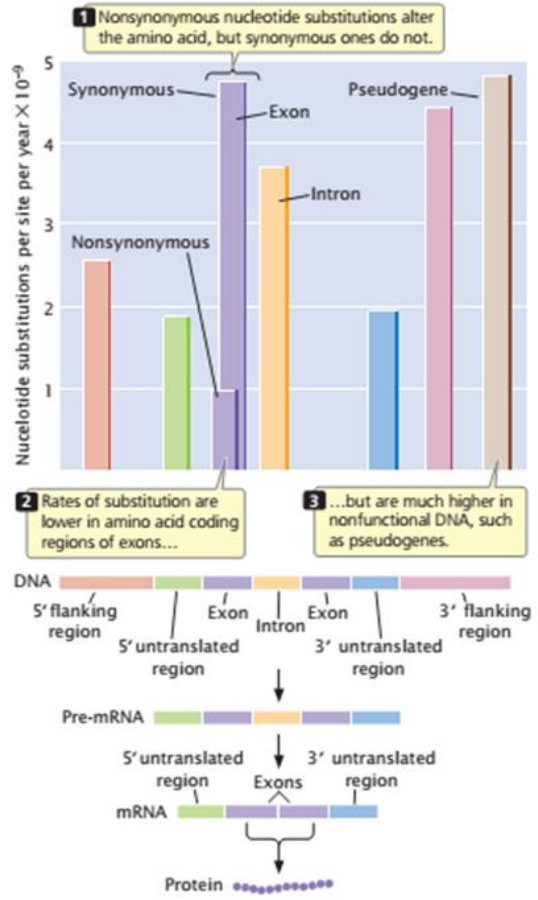
18.13 Insertions and deletions may result from strand slippage.



18.14 Unequal crossing over produces insertions and deletions.

	Within populations	Between populations
Increase genetic variation	**Mutation Migration Some types of natural selection**	**Mutation Genetic drift Some types of natural selection**
Decrease genetic variation	**Genetic drift Some types of natural selection**	**Migration Some types of natural selection**

25.15 Mutation, migration, genetic drift, and natural selection have different effects on genetic variation within populations and on genetic divergence between populations.



26.16 Different parts of genes evolve at different rates. The highest rates of nucleotide substitution are in sequences that have the least effect on protein function.

Table 18.3 Mutation rates of different genes in different organisms

Organism	Mutation	Rate	Unit
Bacteriophage T2	Lysis inhibition	1×10^{-8}	Per replication
	Host range	3×10^{-9}	
Escherichia coli	Lactose fermentation	2×10^{-7}	Per cell division
	Histidine requirement	2×10^{-8}	
Neurospora crassa	Inositol requirement	8×10^{-8}	Per asexual spore
	Adenine requirement	4×10^{-8}	
Corn	Kernel color	2.2×10^{-6}	Per gamete
Drosophila	Eye color	4×10^{-5}	Per gamete
	Allozymes	5.14×10^{-6}	
Mouse	Albino coat color	4.5×10^{-5}	Per gamete
	Dilution coat color	3×10^{-5}	
Human	Huntington disease	1×10^{-6}	Per gamete
	Achondroplasia	1×10^{-5}	
	Neurofibromatosis (Michigan)	1×10^{-4}	
	Hemophilia A (Finland)	3.2×10^{-5}	
	Duchenne muscular dystrophy (Wisconsin)	9.2×10^{-5}	

Mutation rates vary among species

Smaller genomes have higher mutation rates than larger ones.

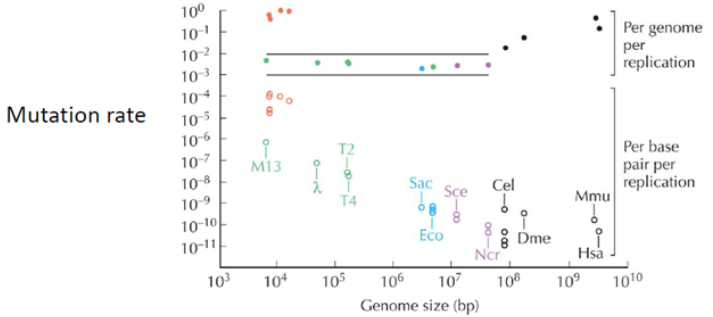
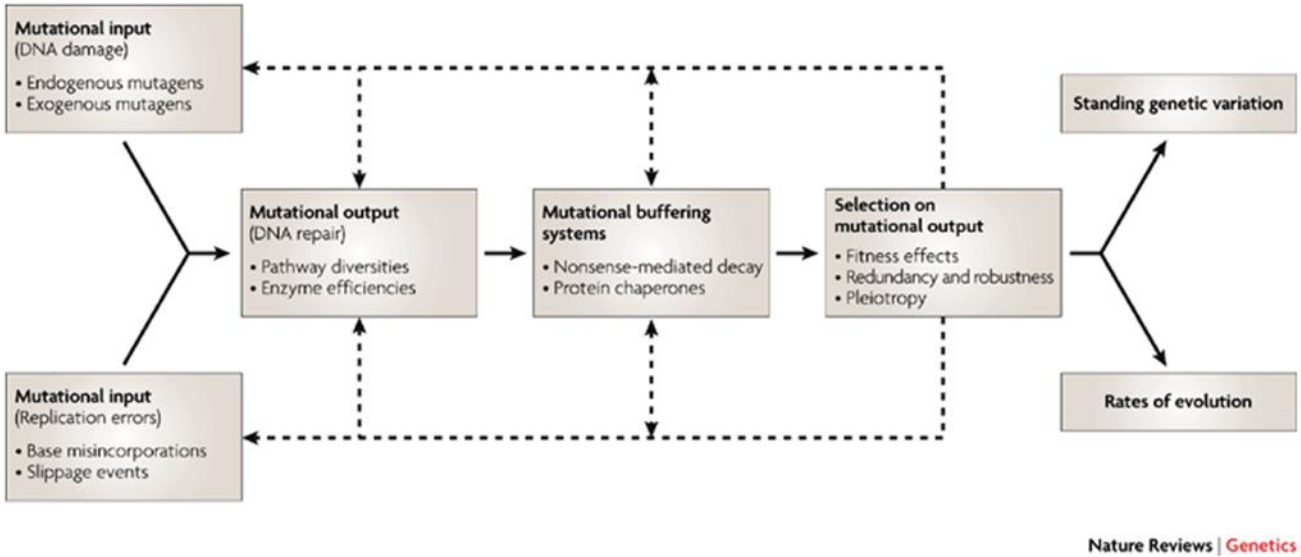


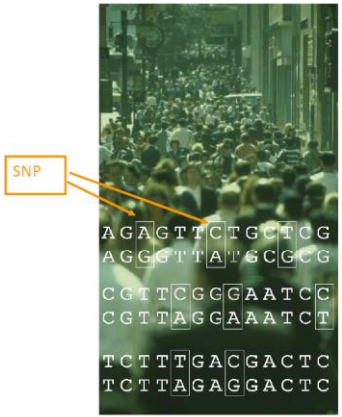
FIGURE 12.23. Mutation rates versus genome size. Mutation rates are shown for a variety of organisms relative to genome size. On the *bottom* portion of the graph, mutation is plotted per base pair per replication event. Note the downward trend, with mutation rate decreasing as genome size increases. On the *top* portion, mutation rate is plotted per genome per replication. Note the relative uniformity in the values across diverse organisms and genome sizes. RNA viruses (*red*): rhinovirus, poliovirus, vesicular stomatitis virus, and measles virus. DNA viruses (*green*): MI3, λ, T1, and T2. Archaea (*blue*): Sac, *Sulfolobus acidocaldarius*. Bacteria (*blue*): Eco, *Escherichia coli*. Eukaryotic microbes (*purple*): Sce, S*accharomyces cerevisiae*; Ncr, *Neurospora crassa*. Metazoa (*black*): Cel, *Caenorhabditis elegans*; Dme, *Drosophila melanogaster*; Mmu, mouse; Hsa, human.



Points of control in mutation rate evolution

Sources of mutational input are processed through DNA repair pathways to yield mutational output. The phenotypic effects of mutations with expressed products are subject to mutational buffering systems before undergoing natural selection. Mutational buffering systems and selective forces have the potential to indirectly affect sources of mutational input and output (dashed arrows). All of these forces ultimately contribute to standing genetic variation in populations and rates of molecular evolution.

How are genomes of individuals different?



Nature **409**, 822 - 823 (2001)

More than 90% of the differences take the form of substitutions at a single base. These are called single nucleotide polymorphisms (**SNP**)

Any 2 human genomes are roughly 99.9% identical

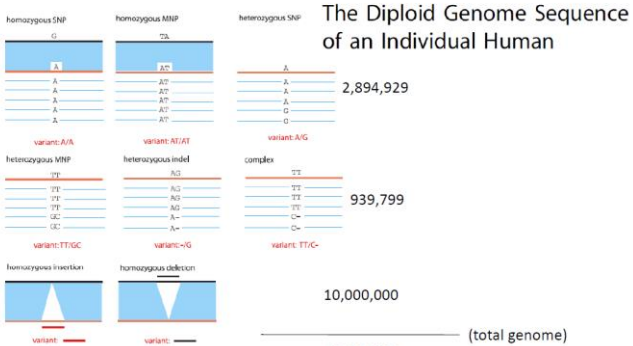
Region	chr.	*n*	bp	*S*	π (%)
LPL[a]	8	142	9700	79	0.166
β-globin	11	349	2670	19	0.157
MC1R	16	242	951	6	0.114
ACE	17	22	24000	74	0.091
APOE[a]	19	192	5491	22	0.053
Dys44	X	250	7622	34	0.093
Xq13.3	X	70	10163	33	0.033
PDHA1	X	35	4200	24	0.178
DMD44[a]	X	41	3000	19	0.141
ZFX	X	336	1089	10	0.082
DMD7	X	41	2389	9	0.034
HPRT	X	10	2485	4	0.038
PLP	X	10	769	2	0.095
GK	X	10	1861	1	0.019
IL2RG	X	10	1147	0	0
IDS	X	10	1909	0	0
VDA survey	–	14	2 000 000	2748	0.046
VDA survey	–	114	196 200	560	0.051[a]
VDA survey	–	148	190 000	874	–[a]

On average ~ 0.1%

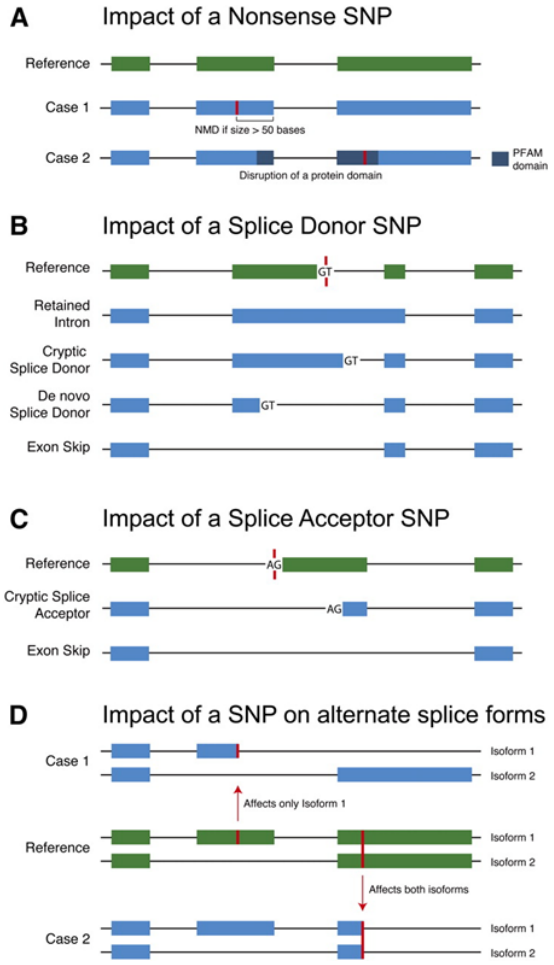
Chr - chromosome
n - Number of samples examined
bp - Number of basepairs sequences
S - Number of polymorphic sites
π - Nucleotide divergence

Przeworski, M., et al. (2000) Trends Genet 16, 296-302.

Humans are 99.5% identical (not 99.9%)



S. Levy, *PLoS Biol* **5** (2007), p. e254



Consequences of nonsense SNPs and SNPs in canonical splice sites.

The SNP is indicated by a red line in A and D, and a SNP in either of the canonical splice site positions is indicated by a red line in B and C.

Types of variations present in a human genome sequence.

The reference genome is shown at the top of each subfigure, with the individual's diploid genome shown below it.

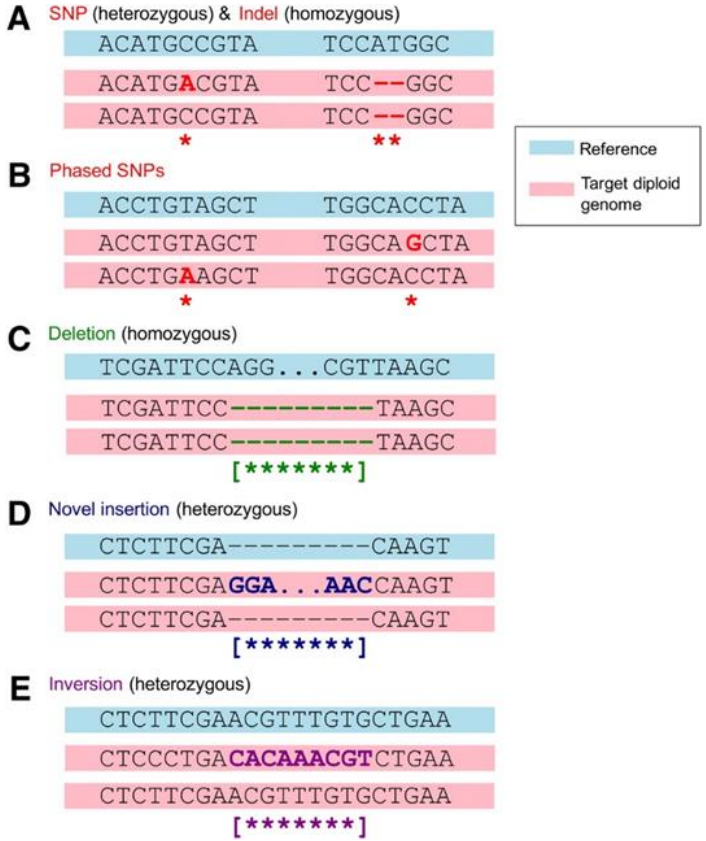
(A) A heterozygous insertion SNP and a homozygous small deletion.

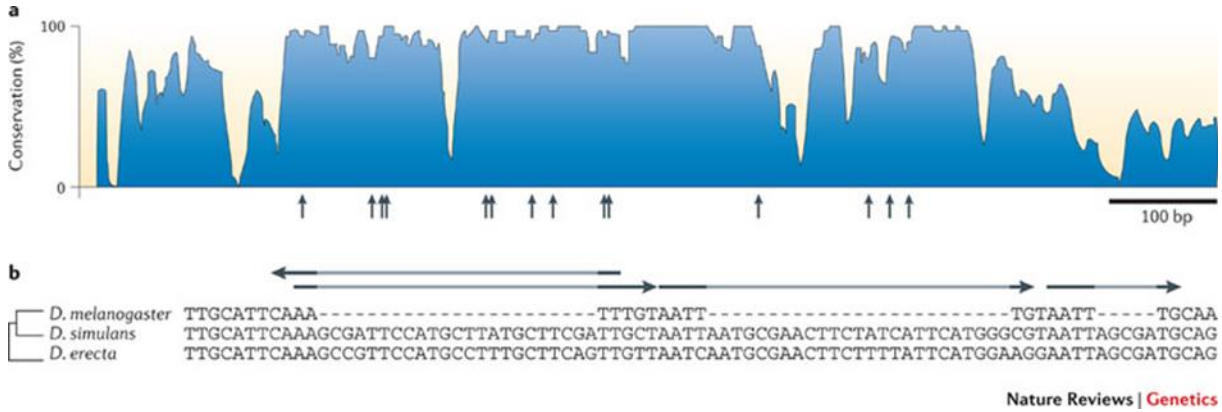
(B) Two phased SNPs.

(C) A homozygous deletion in the target genome.

(D) A heterozygous novel insertion.

(E) A heterozygous inversion event.

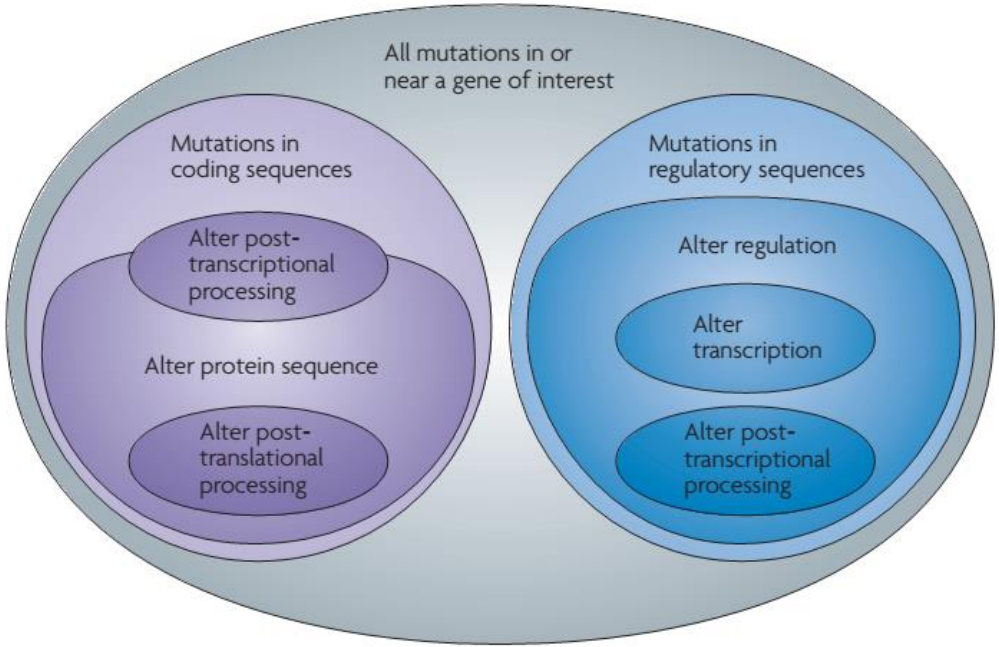




Examples of point mutations and deletions causing divergent cis-regulatory activity

a| A **1,007 bp enhancer** located 5′ of the shavenbaby promoter contains ***14 nucleotide differences that are specific to Drosophila sechellia***. A measure of sequence conservation for this enhancer is plotted, and vertical arrows indicate the location of the 14 D. sechellia-specific alleles.

b| Multiple deletions in an enhancer for the desatF gene (also known as Fad2) (only part of which is shown) have occurred in Drosophila melanogaster and create copies of the hexamer sequence AATTTG — three on one strand and one on the other (indicated with horizontal arrows). The D. melanogaster gene sequence is compared with those of Drosophila simulans and Drosophila erecta. Although no specific TF has yet been shown to bind to these sites, mutating them shows that they are essential for enhancer activity.



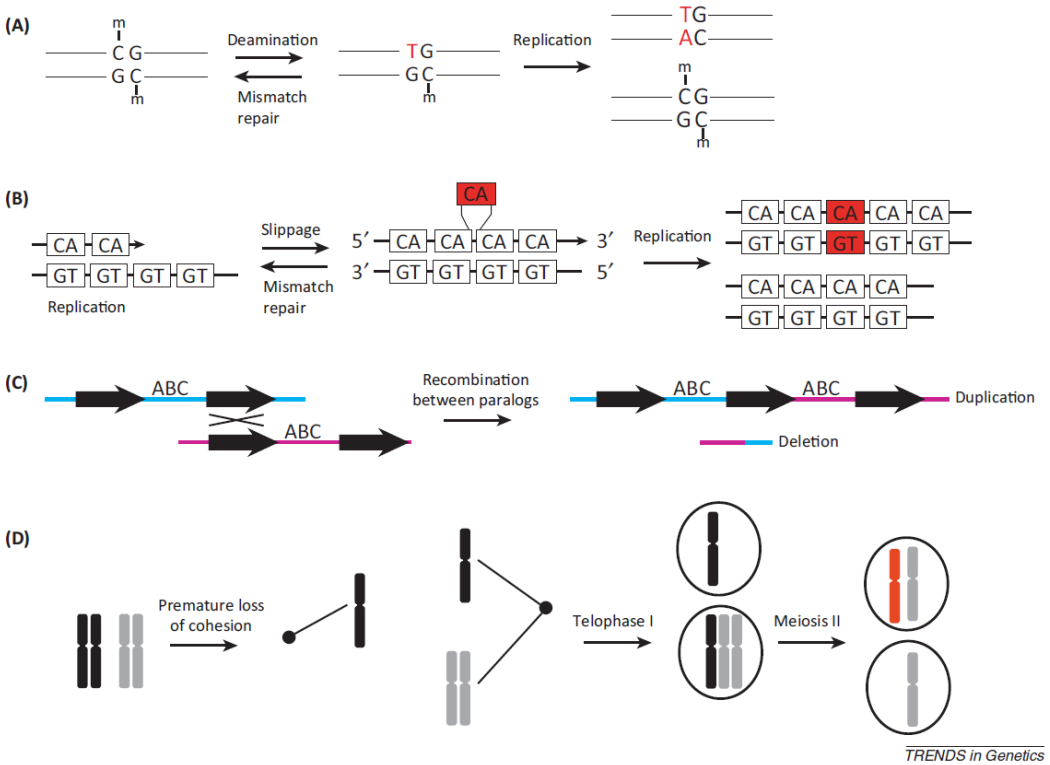
Functional classes of mutations

Although most mutations in and around a gene have no functional consequence (grey area), a subset occur in coding or regulatory regions (purple and blue areas, respectively). Only some of these mutations have functional consequences at a molecular level. Of these, the only mutations that can be reliably and exhaustively identified through sequence comparisons are those that result in an amino-acid substitution. Mutations that affect regulation of all kinds (transcription, mRNA splicing and stability, and posttranslational modifications) are generally difficult to identify without functional tests. As a result, the vast majority of evolutionary analyses focus on non-synonymous substitutions, while the evolutionary consequences of other functional classes of mutations remain poorly studied.

Table 1 The spectrum of variation in the human genome

Variation	Rearrangement type	Size range[a]	References
Single base-pair changes	Single nucleotide polymorphisms, point mutations	1 bp	(3)
Small insertions/deletions	Binary insertion/deletion events of short sequences (majority <10 bp in size)	1–50 bp	(18, 143)
Short tandem repeats	Microsatellites and other simple repeats	1–500 bp	(39)
Fine-scale structural variation	Deletions, duplications, tandem repeats, inversions	50 bp to 5 kb	(34, 54, 87)
Retroelement insertions	SINEs, LINEs, LTRs, ERVs[b]	300 bp to 10 kb	(17)
Intermediate-scale structural variation	Deletions, duplications, tandem repeats, inversions	5 kb to 50 kb	(34, 54, 87, 136)
Large-scale structural variation	Deletions, duplications, large tandem repeats, inversions	50 kb to 5 Mb	(34, 35, 54, 57, 87, 123, 124, 136)
Chromosomal variation	Euchromatic variants, large cytogenetically visible deletions, duplications, translocations, inversions, and aneuploidy	~5 Mb to entire chromosomes	(61, 62)

[a]Size ranges quoted are indicative only of the scale of each type of rearrangement, and are not definitive.
[b]SINE, short interspersed element; LINE, long interspersed element; LTR, long terminal repeat; ERV, endogenous repeat virus.



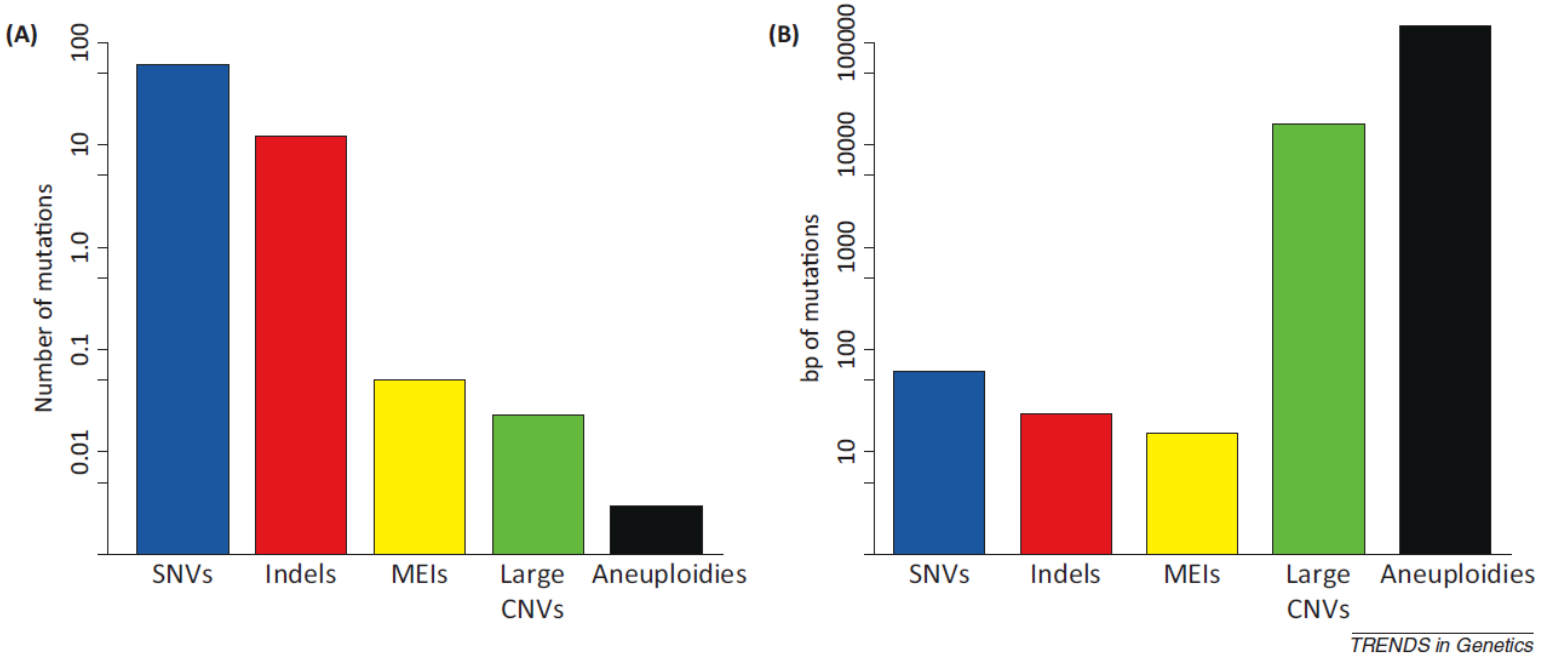
Common mechanisms leading to biases in mutation

(A) CpG dinucleotides are the sites of cytosine methylation and frequent mutation. 5-methyl-cytosine can be deaminated to thymine (red). This mutation can either be repaired by mismatch repair pathways or be replicated to yield a cytosine to thymine mutation.

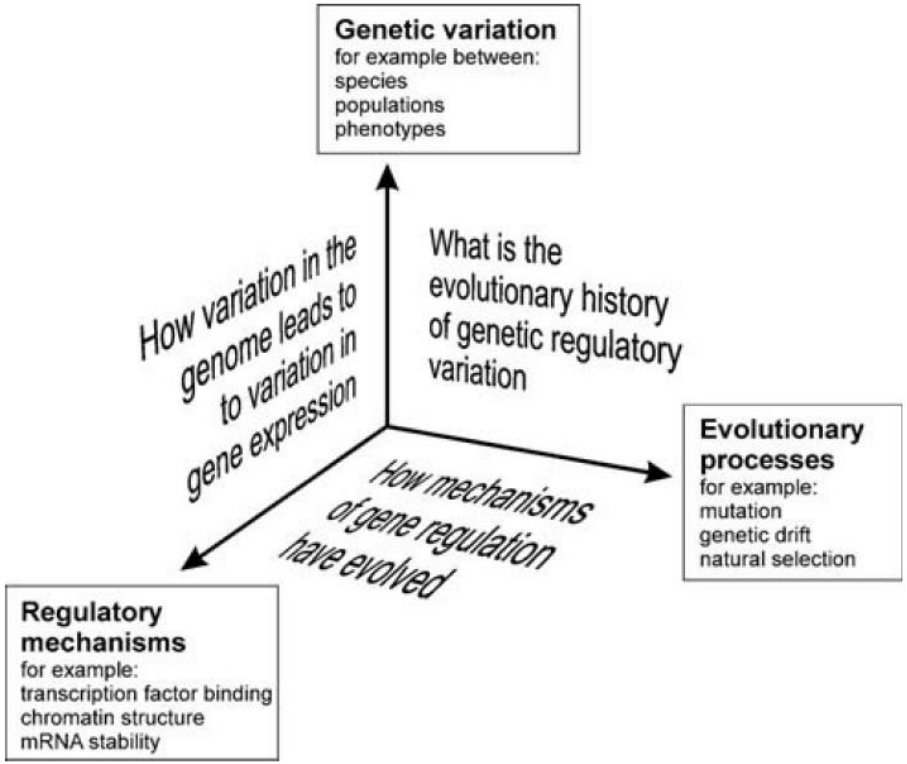
(B) Indels can occur by polymerase slippage during replication if these events are not repaired by mismatch repair, especially in regions of low complexity, such as microsatellites. Replication slippage is shown (red) on the newly synthesized strand leading to an insertion.

(C) Regions flanked by highly identical segmental duplications (SDs; black boxes) are prone to ***nonallelic homologous recombination (NAHR).*** Recombination between homologous chromosomes (blue and magenta) occurs in paralogous regions, leading to duplication of genes ABC in one of the recombined chromosomes and deletion on the other.

(D) Replicated homologous chromosomes are shown in black and gray. ***Premature loss of cohesion*** between sister chromatids ***can lead to separation of chromatids in meiosis*** I (black), ***leading to cells with only one chromatid or three chromatids.*** Trisomy results after meiosis II, when one gamete ends up with an extra chromatid (red).

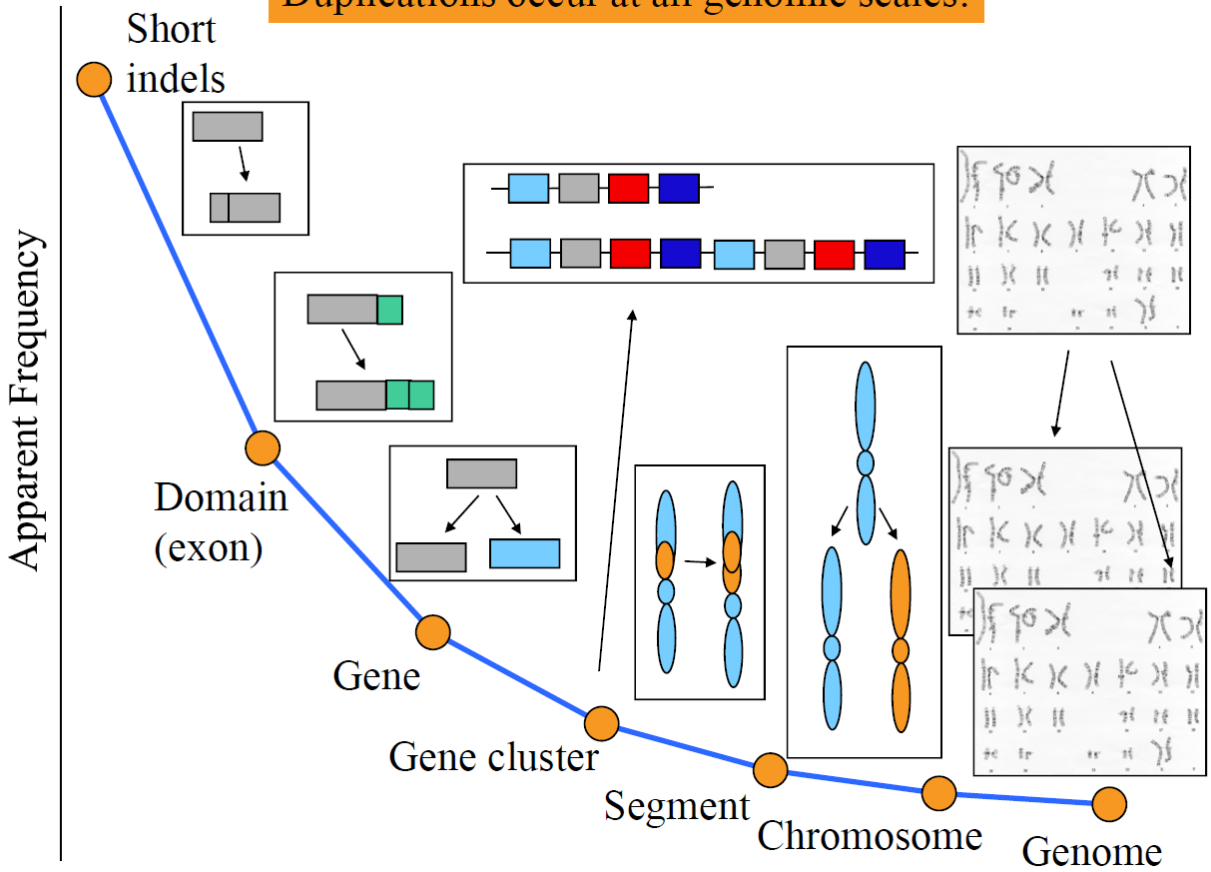


Comparison of the frequency and scale of different forms of genetic variation. There is an inverse relation between mutation size and frequency. Although single nucleotide variants (SNVs) occur more frequently, each mutation affects only a single base pair. By contrast, large mutations, such as copy number variants (CNVs) or chromosomal aneuploidy, are rare, yet affect thousands to millions of base pairs. In addition, although these mutations are rare, they affect more base pairs per birth on average than do SNVs.
(A) Average number of mutations of each type of variant per birth.
(B) Average number of mutated bases contributed by each type of variant per birth.
Y-axis is log10 scaled in both (A) and (B). Abbreviation: MEI, mobile element insertion.

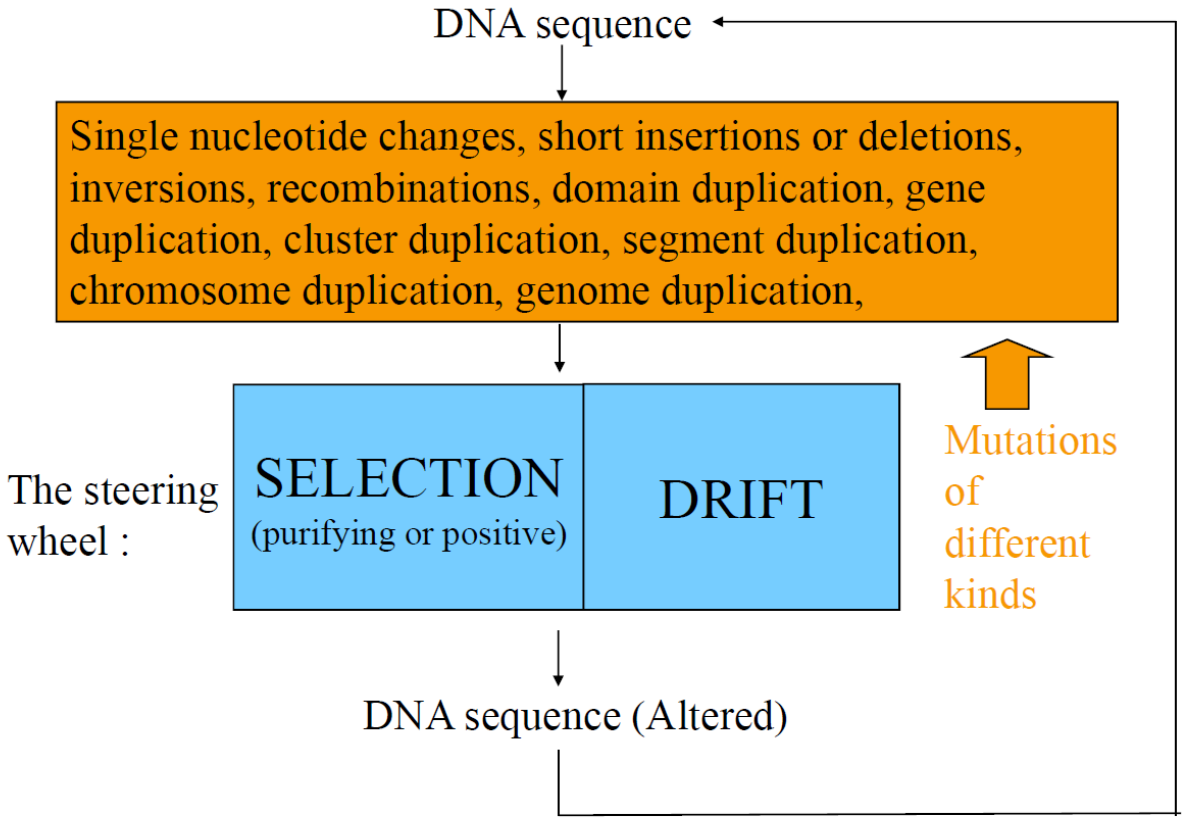


Multiple dimensions of the analysis of the evolutionary history of regulatory variation

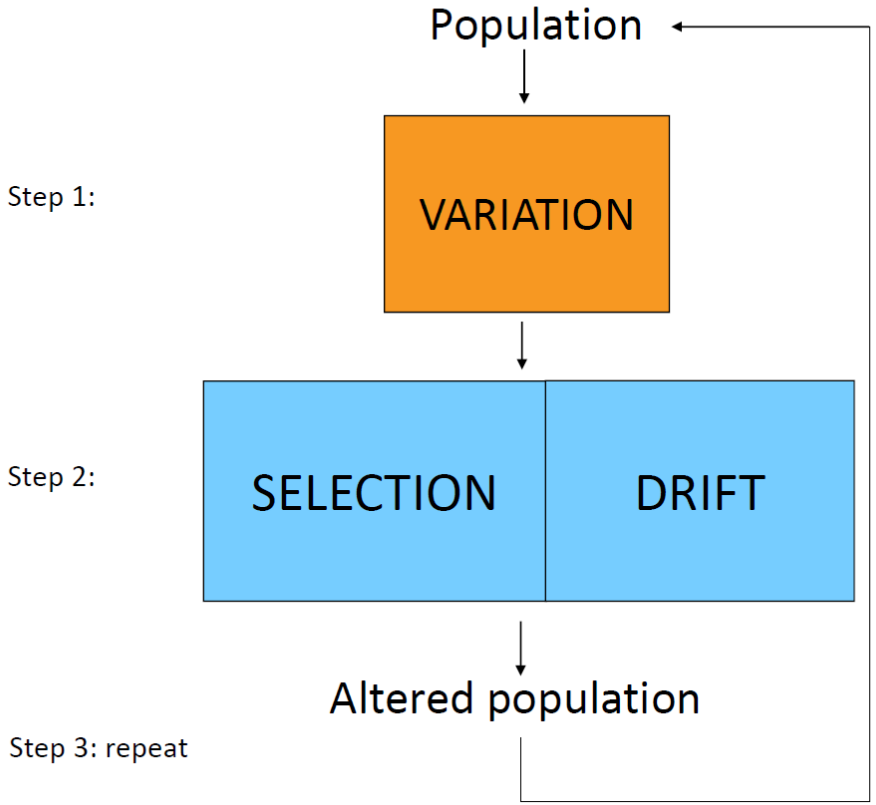
Duplications occur at all genomic scales!



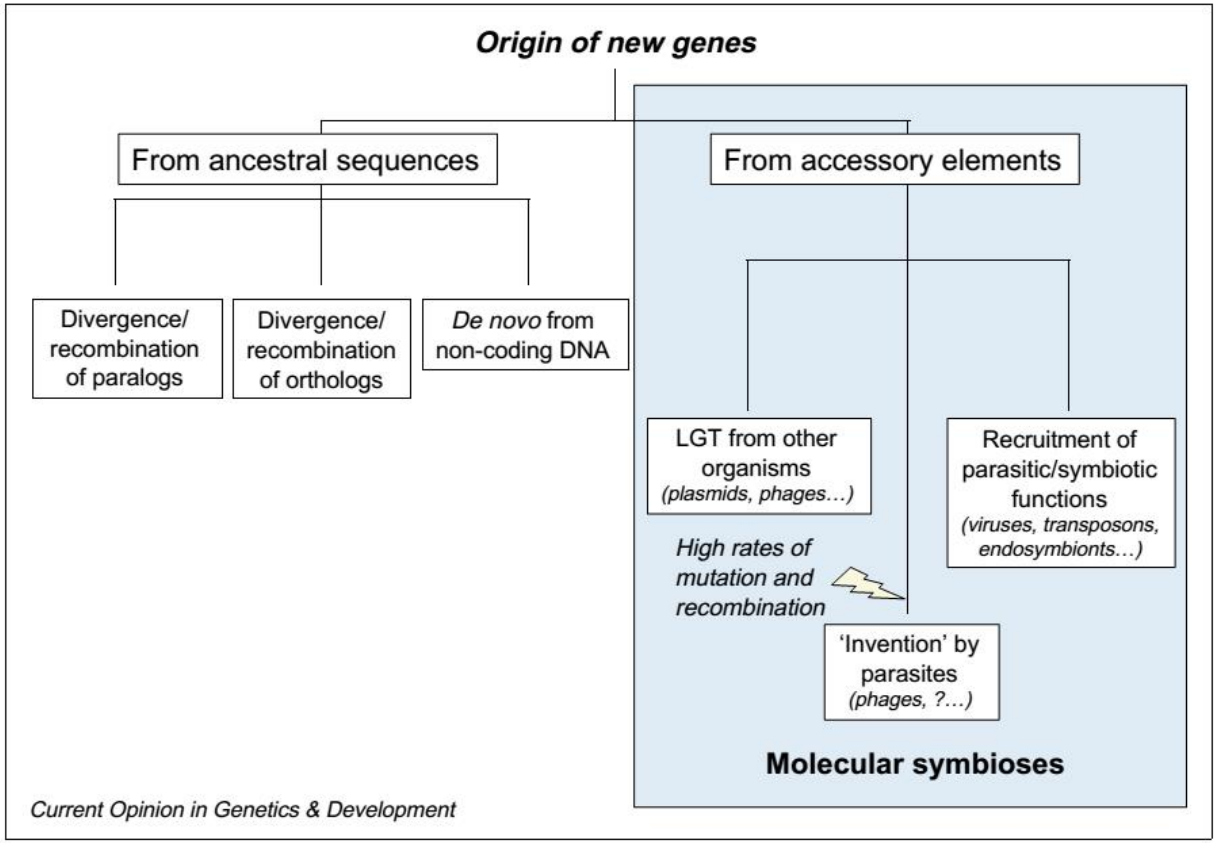
A modified model of genome evolution



The standard model of evolution

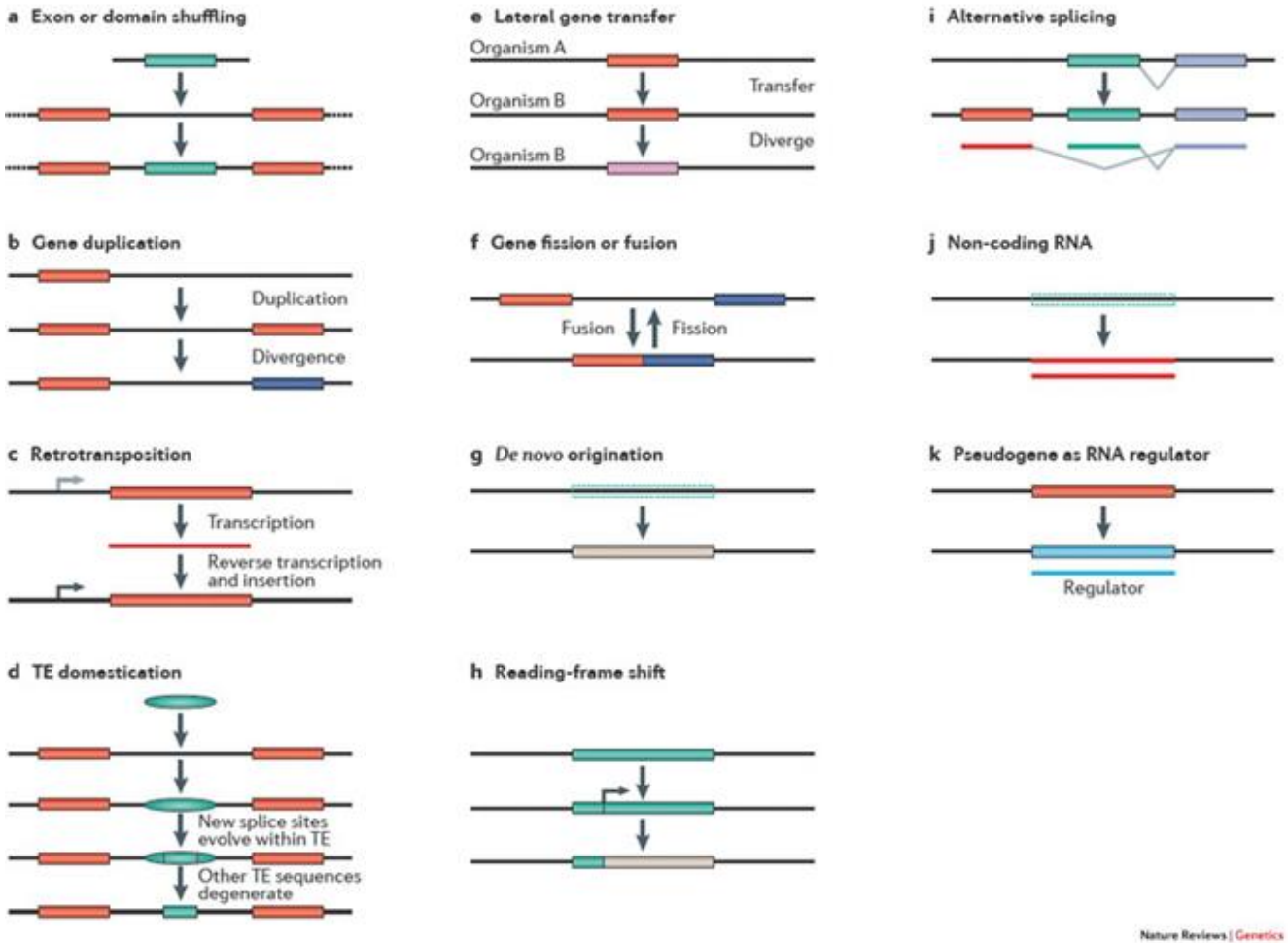


Vzorci genomske evolucije



Mechanisms of the origin of new genes

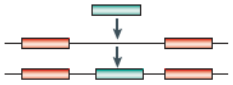
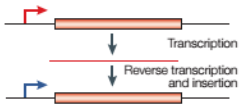
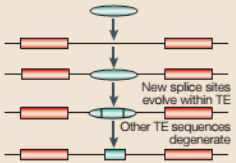
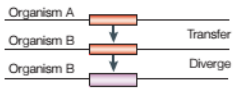
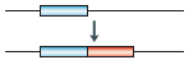
Eukaryotic genome repertoires are usually thought to evolve from ancestral sequences that are already present in the genome (left), although contributions by accessory elements (right), such as transposons, have been reported. ***In prokaryotes, selfish elements (right), and particularly bacteriophages, are likely to have played a much more significant role.*** Abbreviation: LGT, lateral gene transfer.



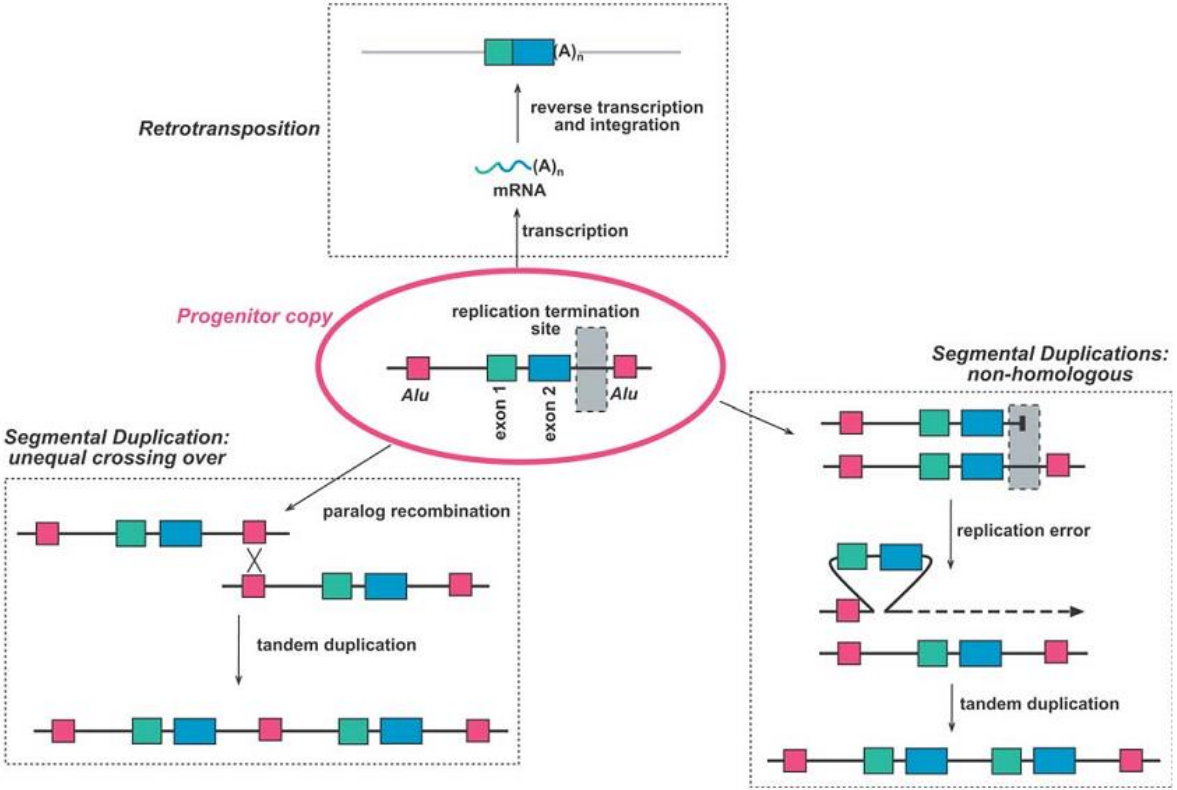
Mechanisms of new gene origination

Nature Reviews | Genetics

Table 1 | **Molecular mechanisms for creating new gene structures**

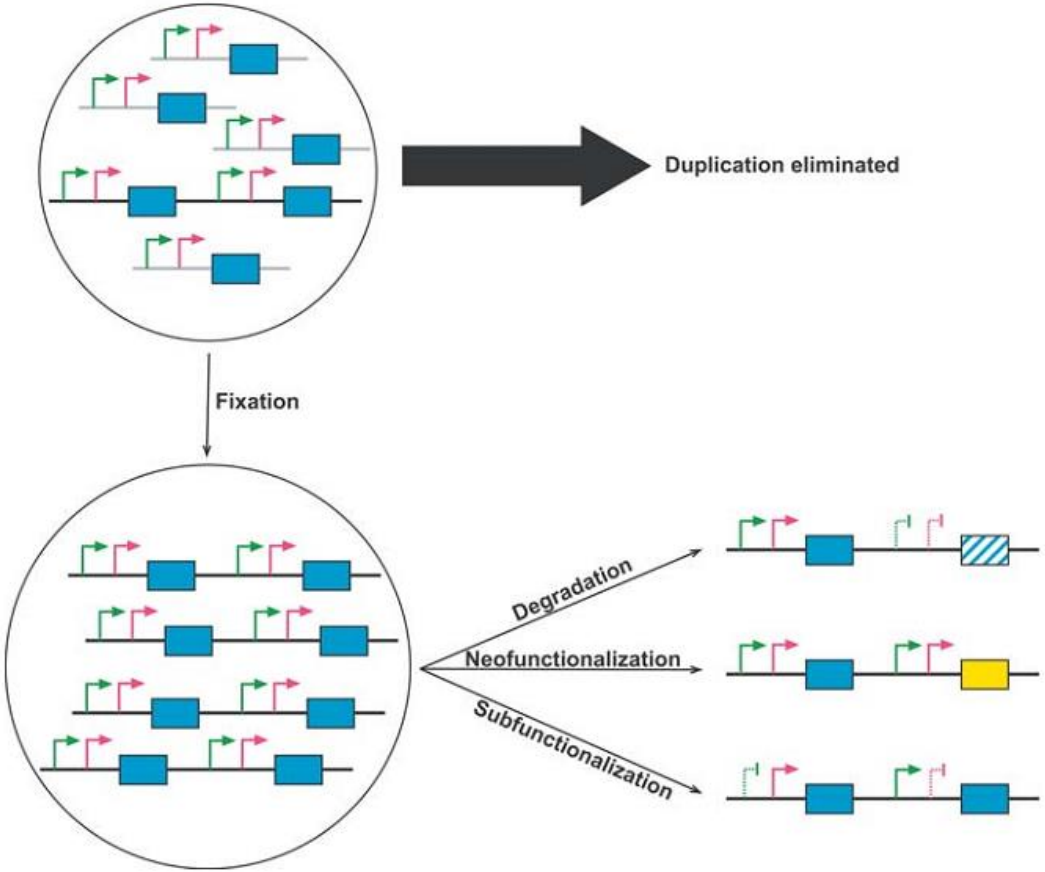
Mechanism	Process	Examples	Comments	References
Exon shuffling: ectopic recombination of exons and domains from distinct genes		*fucosyltransferase*, *jingwei*, *Tre2*	~19% of exons in eukaryotic genes have been formed by exon shuffling	8,32,40,62, 65–68,105
Gene duplication: classic model of duplication with divergence		*CGβ*, *Cid* , *RNASE1B*	Many duplicates have probably evolved new functions	9–11,29,35,39, 47,48,106
Retroposition: new gene duplicates are created in new genomic positions by reverse transcription or other processes		*PGAM3*, *Pgk2*, *PMCHL1*, *PMCHL2*, *Sphinx*	1% of human DNA is retroposed to new genomic locations	23,43,61,76, 80–82,107–110
Mobile element: a mobile element, also known as a transposable element (TE), sequence is directly recruited by host genes		*HLA-DR-1*, human *DAF*, *lungerkine* mRNA, *mNSC1* mRNA	Generates 4% of new exons in human protein-coding genes	16,78,111,112
Lateral gene transfer: a gene is laterally (horizontally) transmitted among organisms		*acytylneuraminate lysase*, *Escherichia coli* *mutU* and *mutS*	Most often reported in prokaryotes and recently reported in plants	18–20,113
Gene fusion/fission: two adjacent genes fuse into a single gene, or a single gene splits into two genes		Fatty-acid synthesis enzymes, *Kua-UEV*, *Sdic*	Involved in the formation of ~0.5% of prokaryotic genes	21,22,42, 114,115
De novo origination: a coding region originates from a previously non-coding genomic region		*AFGPs*, *BC1RNA*, *BC200RNA*	Rare for whole gene origination; might not be rare for partial gene origination	52–53,116,117

AFGP, antifreeze glycoprotein; *CGβ*, chorionic gonadotropin β polypeptide; *Cid*, centromere identifier; *DAF*, decay-accelerating factor; *HLA-DR-1*, major histocompatibility complex DR1; *PGAM3*, phosphoglycerate mutase 3; *Pgk2*, phosphoglycerate kinase 2; *PMCHL*, pro-melanin-concentrating hormone-like; *RNASE*, ribonuclease; *Sdic*, sperm-specific dynein intermediate chain; *UEV*, tumour susceptibility gene.



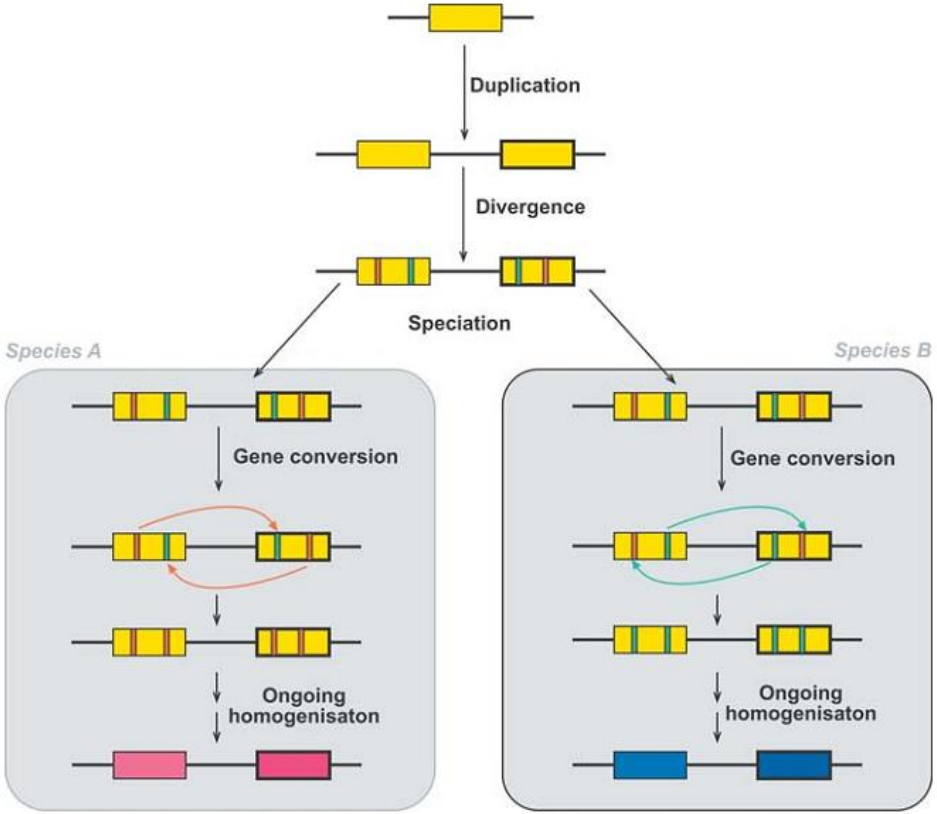
Mechanism of Gene Duplication

A two-exon gene is flanked by two Alu elements and a neighbouring replication termination site. Recombination between the two Alu elements leads to a tandem duplication event, as does a replication error instigated by the replication termination site. Retrotransposition of the mRNA of the gene leads to the random integration of an intron-less paralogue at a distinct genomic location.



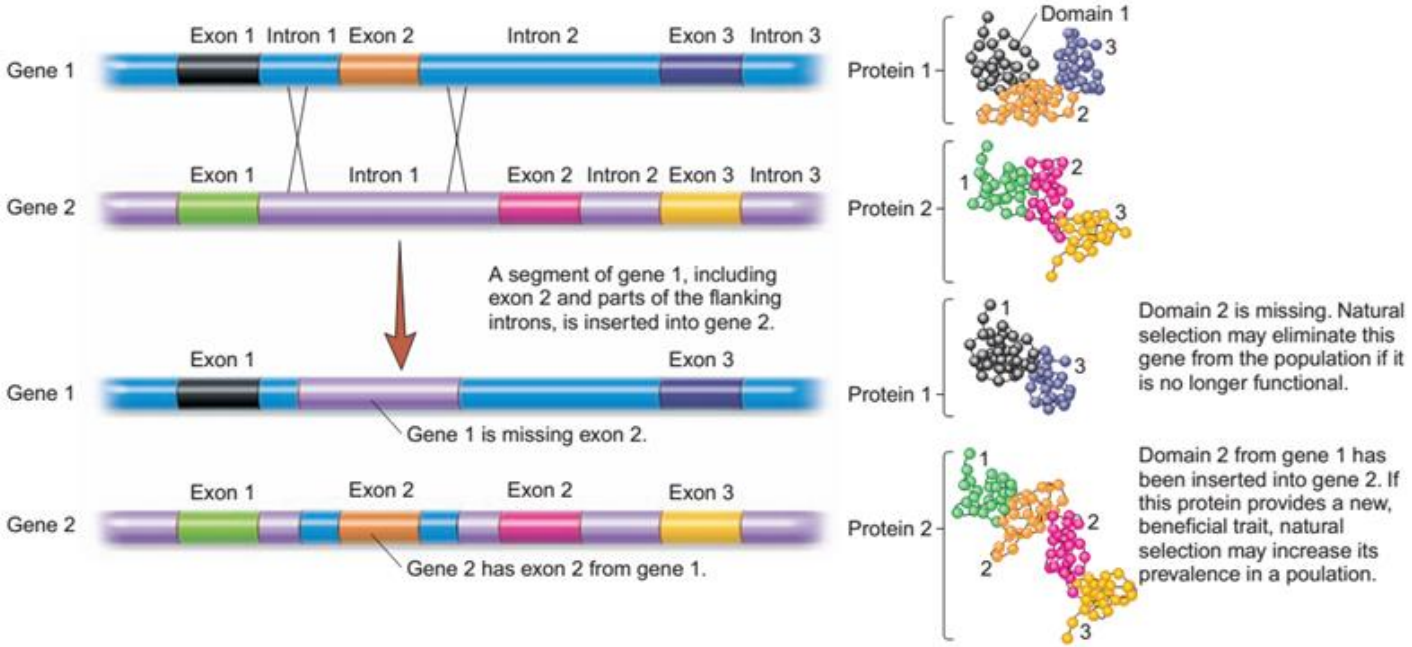
Fates of Duplicate Genes

A new duplication in a gene (blue) with *two tissue-specific promoters* (arrows) arises in a population of single copy genes. **Fixation within the population** results in a minority of cases. After fixation, one gene is inactivated (**degradation**) or assumes a new function (**neofunctionalization**), or the expression pattern of the original gene is partitioned between the two duplicates as one promoter is silenced in each duplicate in a complementary manner (**subfunctionalization**).



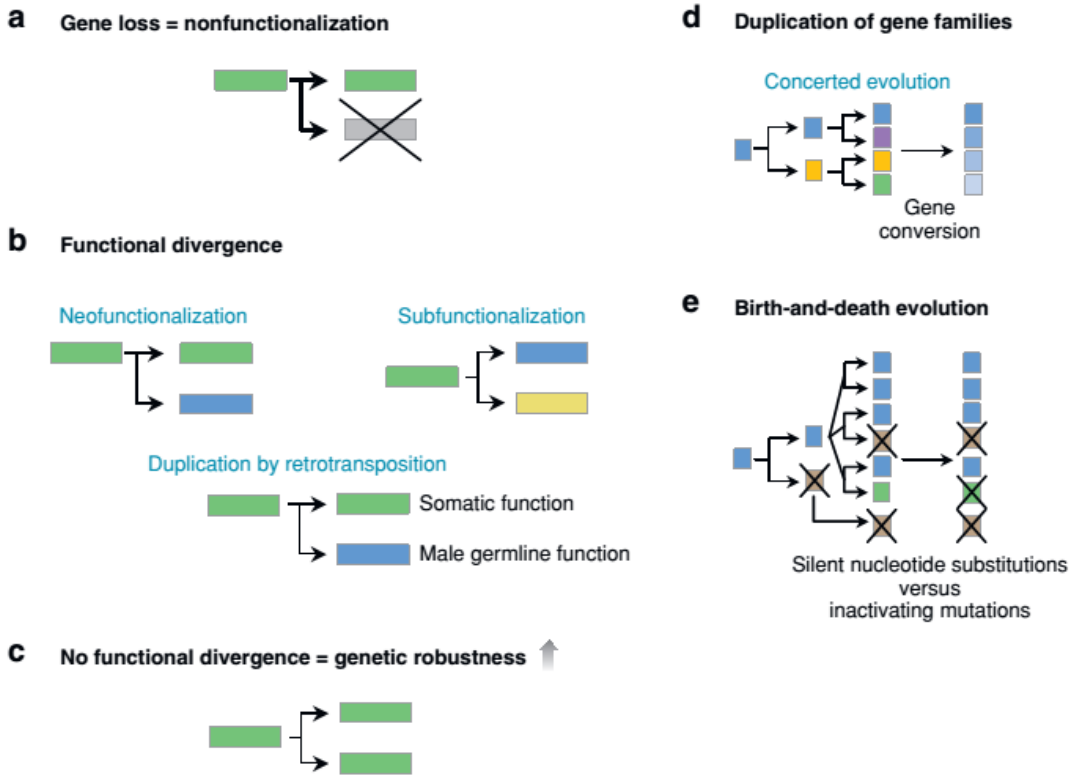
Concerted Evolution

Different gene conversion events ***homogenize minimally diverged duplicate genes*** in each daughter species (A and B), with the result that while ***paralogues are highly similar, orthologues diverge over time***.

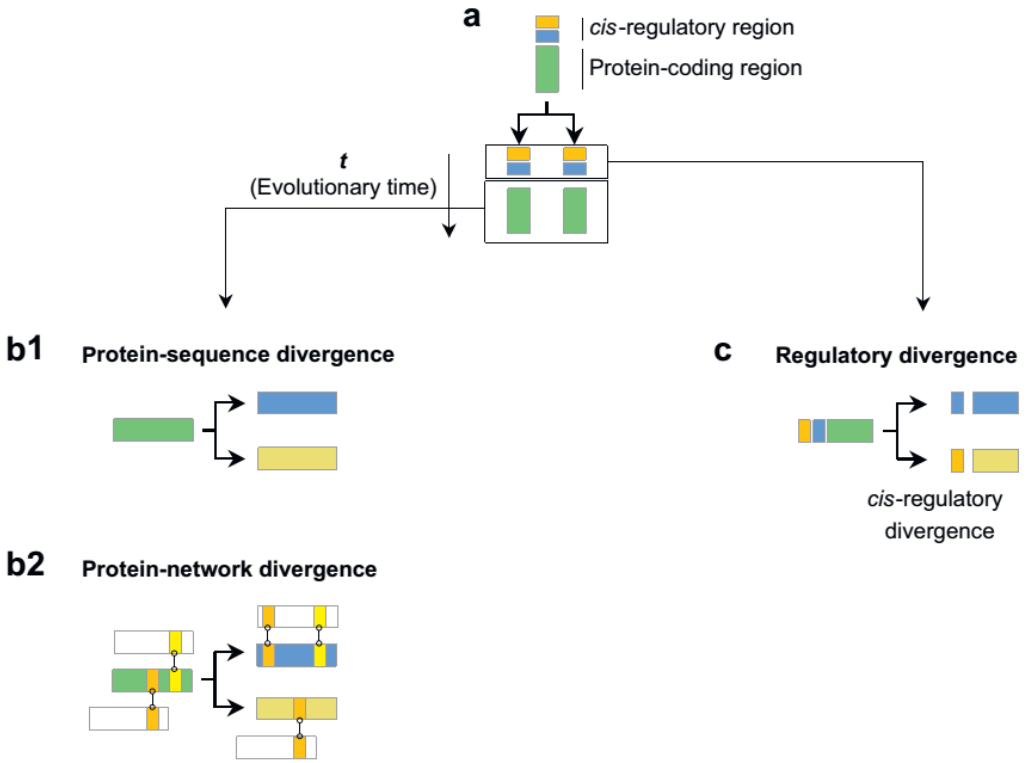


The process of exon shuffling

In this example, a segment of one gene containing an exon and its flanking introns has been inserted into another gene. A rare, abnormal crossing-over event called nonhomologous recombination may cause this to happen. This results in proteins that have new combinations of domains and possibly new functions.



Evolutionary fate of single gene duplications (a–c), and duplication of multigene families (d–e) Single gene duplication most often results in a nonfunctional duplicate gene copy (a, ***nonfunctionalization***). (b) In rare instances, the functional duplicate gene copy and the ancestral gene diverge in function; ***neofunctionalization*** means that one of the two genes retains the original function, while the other evolves a new, often beneficial function. ***Subfunctionalization*** implies that both the original and the duplicate genes mutate and evolve to fulfill complementary functions already present in the original gene. Duplication via retrotransposition represents a particular case of sub- or neofunctionalization. ***Multigene families*** evolve in a coordinated fashion, such that the DNA coding sequences and function of the single members of a family remain close to that of the ancestral gene (d–e). (d) ***Concerted evolution***: After multiple rounds of duplication, gene conversion homogenizes the DNA sequences of the individual members. (e) ***Birth-and-death evolution*** invokes a process of equilibrium between inactivating mutations and ongoing duplication of functional gene copies.

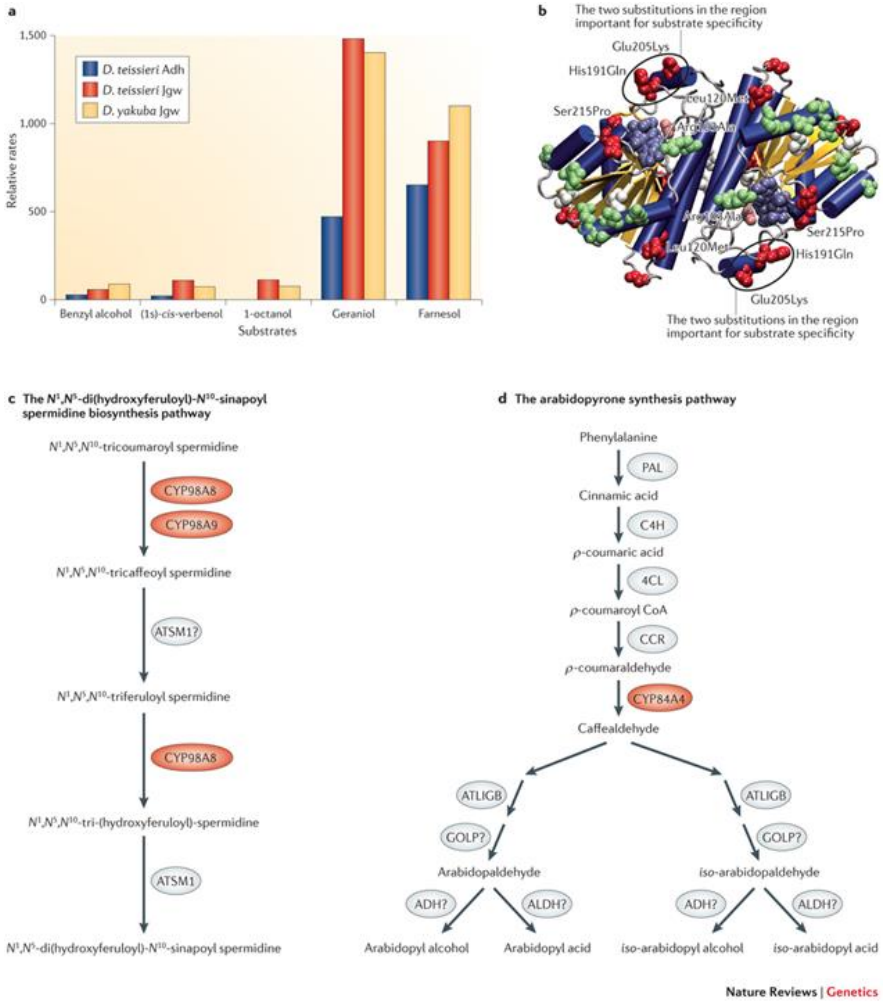


Functional divergence of duplicated genes

(a) The cis-regulatory and the protein-coding regions evolve independently after duplication. Divergence increases with evolutionary distance.

(b) Schematic representation of ***(b1) protein sequence divergence after duplication*** (green sequence diverges into blue or orange); ***(b2) protein network divergence***, where the protein interaction domains of the original green sequence evolve by maintenance, gain, or loss of interacting partners.

(c) Schematic representation of DNA sequence regulatory divergence after duplication (certain regulatory motifs are lost in one copy of the duplicated gene sequence).



Several new genes evolved novel biochemical functions

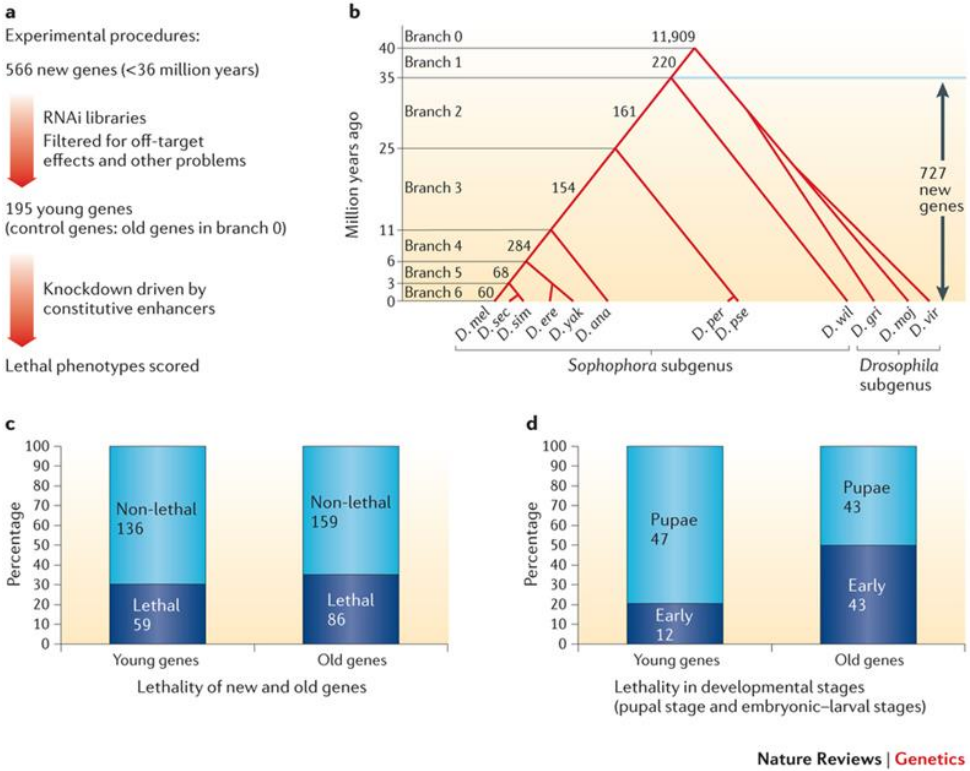
a| The ***protein Jingwei (Jgw)*** encoded by African Drosophila spp. is a dehydrogenase enzyme that has evolved altered substrate specificity compared with the ancestral Alcohol dehydrogenase (Adh). Compared with the parental enzyme, Jgw can use long-chain primary alcohols more efficiently; in particular, Jgw has greater activities than the ancestral Adh towards farnesol (which is involved in juvenile hormone biosynthesis) and geraniol (which is involved in recruitment pheromone biosynthesis).

b| The locations of the substitutions on the structure of the ***dimeric Adh***.

c,d| Three *recent duplicates in the cytochrome P450 family in Arabidopsis* thaliana led to the assembly of a new pathway of N1,N5-di(hydroxyferuloyl)-N10-sinapoyl spermidine biosynthesis (the new duplicates CYP98A8 and CYP98A9 are shown in red) (part c) and a pathway for arabidopyrone biosynthesis (the young duplicated enzyme CYP84A4 is shown in red) (part d).

D. teissieri, Drosophila teissieri; D. yakuba, Drosophila yakuba.

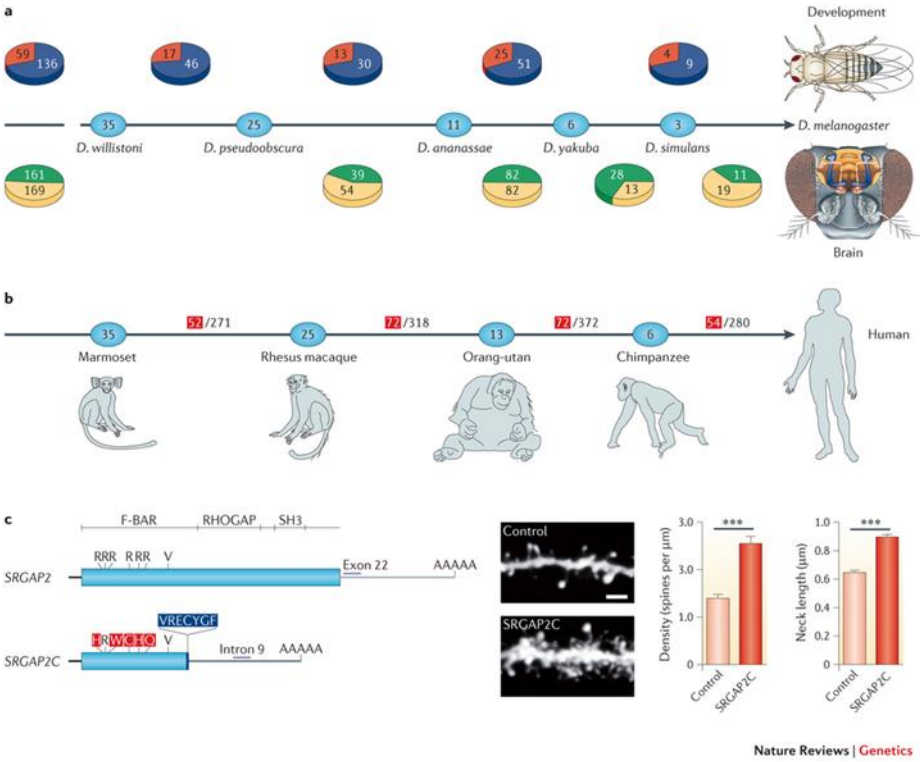
New genes participate in developmental processes in Drosophila



a,b | ***The experimental procedures (part a) assessed the developmental importance of the new genes that have originated during Drosophila spp. speciation*** (as mapped in the evolutionary tree by comparing the genomes of different Drosophila spp. (part b)). The values in the tree indicate the numbers of new genes in the Drosophila melanogaster (D. mel) lineage at each evolutionary branchpoint; 566 genes were chosen for designing knockdown experiments from the total of 727 new genes that have originated since branch 1 in the tree.

c,d | ***A summary of the main experimental results from the 195 new genes tested for the effect on development following gene knockdown.*** First, knockdown of approximately one-third of either new genes or old genes resulted in lethality, indicating that the proportion of essential new genes is similar to the proportion of essential old genes in branch 0 of the evolutionary tree (part c). Second, knockdown of older genes led to the termination of development at earlier stages than knockdown of new genes, indicating that old genes function earlier in development than new genes, most of which have essential functions in the pupal stages (part d).

D. ana, Drosophila ananassae; D. ere, Drosophila erecta; D. gri, Drosophila grimshawi; D. mel, Drosophila melanogaster; D. moj, Drosophila mojavensis; D. per, Drosophila persimilis; D. pse, pseudoobscura; D. sec, Drosophila sechellia; D. sim, Drosophila simulans; D. wil, Drosophila willistoni; D. vir, Drosophila virilise; D. yak, Drosophila yakuba.



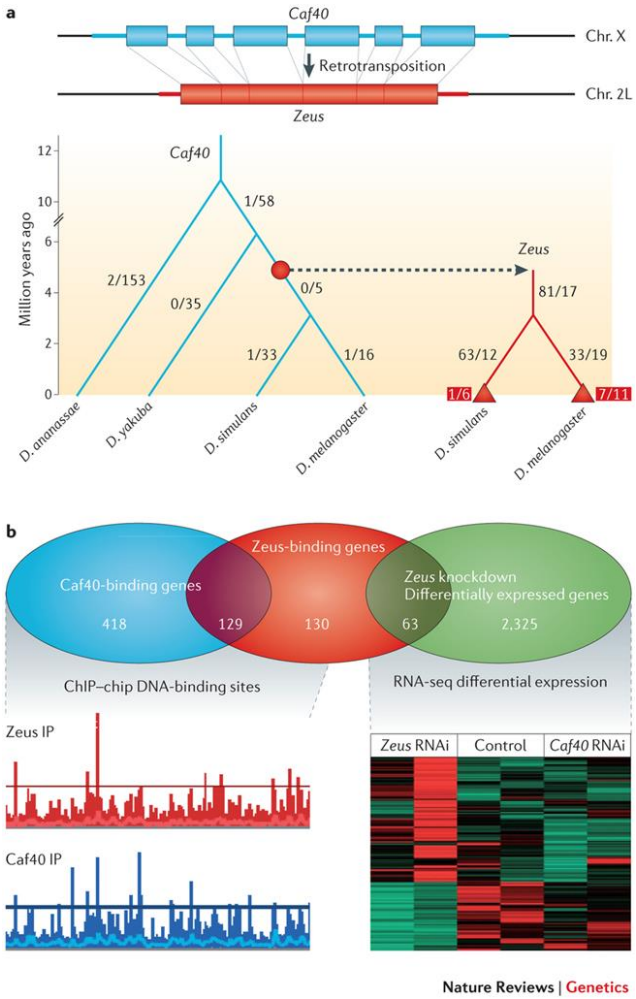
The involvement of new genes in the development of Drosophila spp., and in the brains of Drosophila and humans.

a | The phylogeny shows the distributions of new Drosophila spp. genes involved in development (above) and in the brain (below) in various evolutionary stages within the past 36 million years. The numbers in the ovals show the divergence time (in millions of years) between Drosophila melanogaster and various Drosophila spp. **Red represents the number of new genes that were found to have essential developmental functions,** whereas blue shows the number of new genes that were non-essential in development. **Green represents the number of new genes that were expressed in the brain**, and yellow shows the number of the non-brain-expressed new genes in reverse transcription PCR screening experiments. The left-most pie charts show the total numbers of new genes across all analysed stages. For the developmental data, the origination events at 3–6 million years ago (MYA) and 0–3 MYA are pooled. **These data reveal that older genes are no more likely than newer genes to have evolved essential developmental functions or brain expression, suggesting the rapid evolution of phenotypic effects in new genes.**

b | The phylogeny shows the numbers of new genes in the human genome that originated at various stages during the divergence of human ancestors from the ancestors of other primates through duplication (DNA based and RNA based) and de novo mechanisms. The numbers within the ovals show divergence time (MYA). The numbers in the denominators are the total numbers of new genes that originated in each evolutionary branch, as identified previously, and the ***red numerators are the numbers of those new genes that are expressed in the prefrontal cortex (PFC)***, as detected based on the available microarray expression data. The PFC is the anterior part of the frontal lobe of the neocortex, which is **implicated in cognitive functions in the developing human brain**. These data suggest that **new genes have been frequently acquired into the PFC transcriptome**.

c | The **SLIT–ROBO RHO GTPase-activating protein 2C (SRGAP2C) gene** is one of the 54 PFC-expressed human-specific new genes (the end branch of the tree in part b). It was formed by DNA-based duplication and has been subjected to extensive genetic and evolutionary analyses and functional characterization. The gene structure shows that SRGAP2C (bottom) is a duplicate of the amino-terminal part of the parental gene, SRGAP2 (top). The transgenic expression of SRGAP2C in cultured mouse cortical neurons induces denser and longer spines, as shown in the **dendrite** images (scale bar represents 2 µm), and the measured spine density and neck length, shown in the graphs on the right. *** indicates a significance of P<0.001.

D. ananassae, Drosophila ananassae; D. pseudoobscura, Drosophila pseudoobscura; D. simulans, Drosophila simulans; D. willistoni, Drosophila willistoni; D. yakuba, Drosophila yakuba.

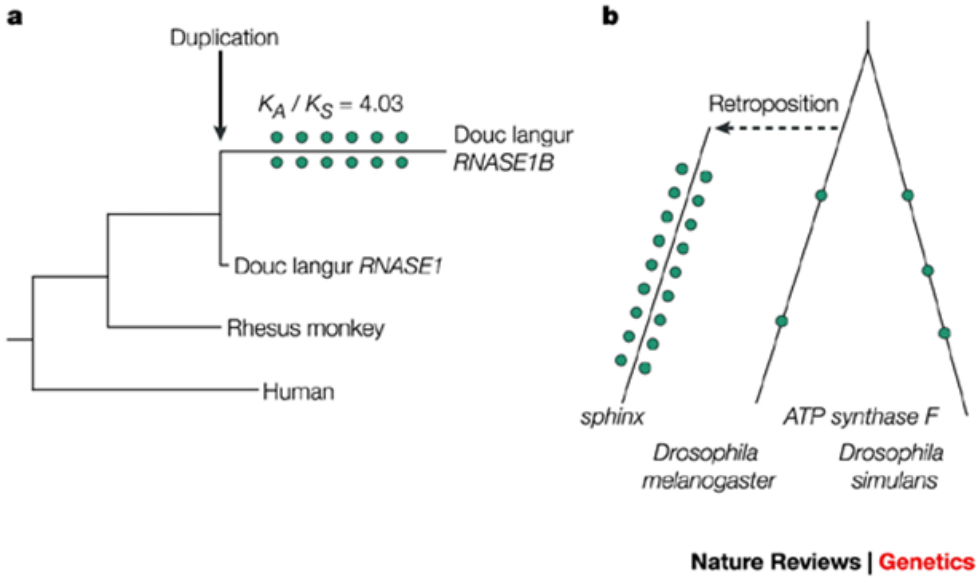


The rapid and extensive evolution of gene expression networks by integration of the new gene Zeus in Drosophila

a | Zeus, a new gene derived from the retrotransposition of the parental Caf40 gene, evolved under positive selection, as detected by the McDonald–Kreitman test and by the Ka/Ks ratio (defined as the ratio of the number of substitutions at non-synonymous sites to the number of substitutions at synonymous sites). In contrast to the slow evolution of Caf40 (blue), Zeus evolved rapidly in its protein sequences (red). The ratios displayed throughout the tree are Ka/Ks ratios, except those in red, which show the ratio of the number of non-synonymous sites to the number of synonymous polymorphic sites of nucleotides in the alleles of the natural population in Drosophila melanogaster and Drosophila simulans.

b | The left panels show a ***comparison of the DNA-binding sites of Caf40 (blue) and Zeus (Red)***, as determined by chromatin immunoprecipitation followed by microarray (ChIP–chip). Separate analyses revealed that almost all of these sites have protein-coding genes downstream (not shown). The right panels show the genes for which expression correlates with the expression of Zeus, as determined by high-throughput RNA sequencing (RNA-seq) following RNA interference (RNAi)-mediated knockdown of Zeus, and the raw data is shown in the heatmap. **The extensive changes in DNA-binding sites revealed that the integration of new genes reshaped the gene expression networks.**

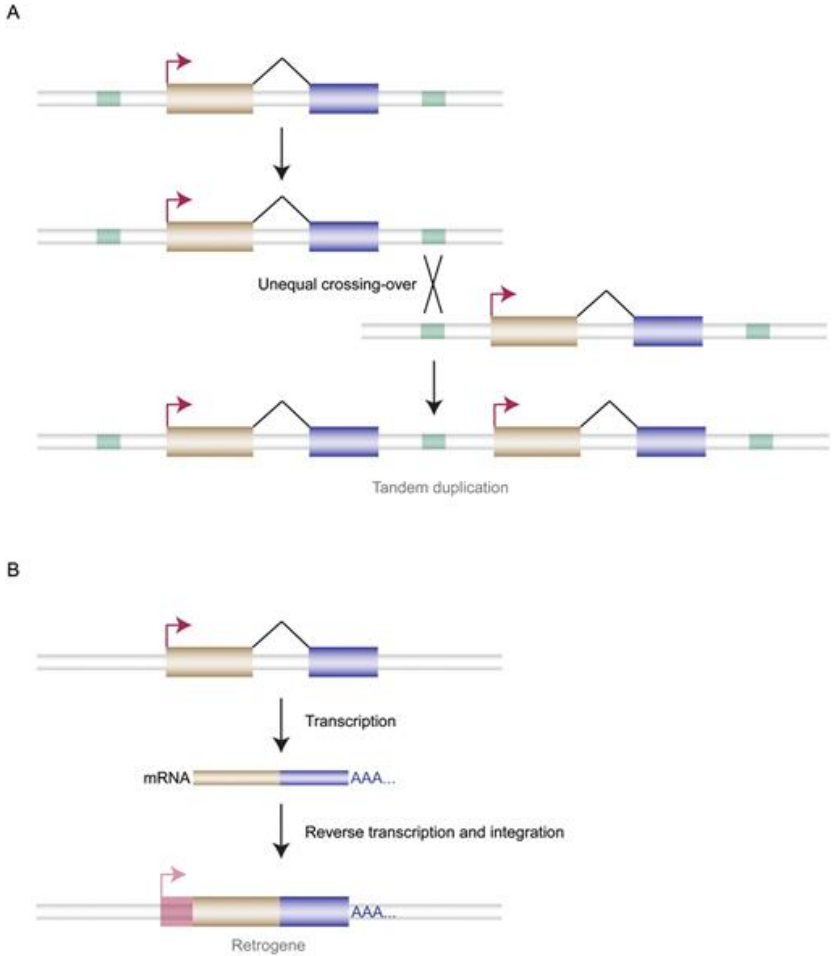
D. ananassae, Drosophila ananassae; D. yakuba, Drosophila yakuba; IP, immunoprecipitation.



Two examples of the accelerated evolution of new genes with new functions

a| The ribonuclease RNASE1B gene in the leaf-eating colobine monkey Pygathrix nemaeus (the douc langur). The green circles indicate nucleotide substitutions in the coding regions of the genes. In the RNASE1B lineage, 12 sites have been substituted, most of which are REPLACEMENT CHANGES; however, its sibling copy RNASE1 has no sequence changes.

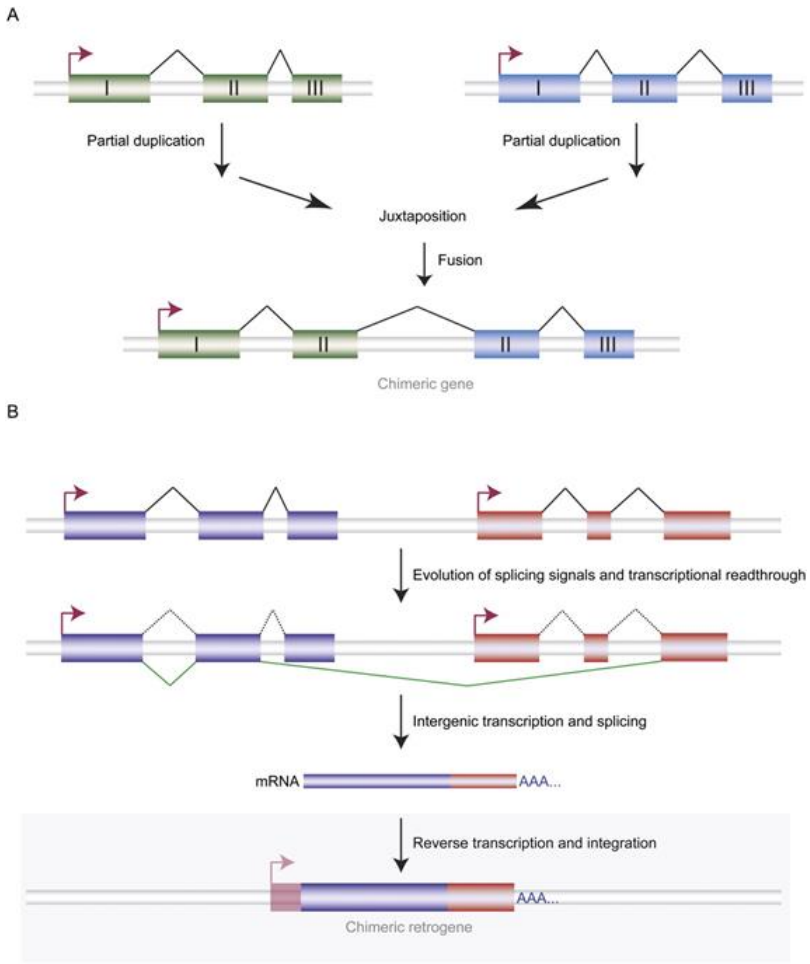
b| The sphinx RNA gene in Drosophila melanogaster. Similar to the protein-coding gene shown in part a, the RNA gene sphinx has also undergone accelerated substitution since its formation: there are 18 substitutions in sphinx versus 2 substitutions in the parental gene ATP synthase F.



Origin of new gene copies through gene duplication

(A) DNA-based duplication. A common type of segmental duplication—tandem duplication—is shown. It may occur via unequal crossing-over that is mediated by transposable elements (light green). There are different fates of the resulting duplicate genes. For example, one of the duplicates may acquire new functions by evolving new expression patterns and/or novel biochemical protein or RNA functions. (Gold and blue boxes) Exons, (black connecting lines) exon splicing, (red right-angled arrows) transcriptional start sites (TSSs), (gray tubes) nonexonic chromatin.

(B) RNA-based duplication (termed retroposition or retroduplication). New retroposed gene copies may arise through the reverse transcription of messenger RNAs (mRNAs) from parental source genes. Functional retrogenes with new functional properties may evolve from these copies after acquisition or evolution of promoters in their 5′ flanking regions that may drive their transcription. (Pink right-angled arrow) TSS, (transparent pink box) additionally transcribed flanking sequence at the insertion site.

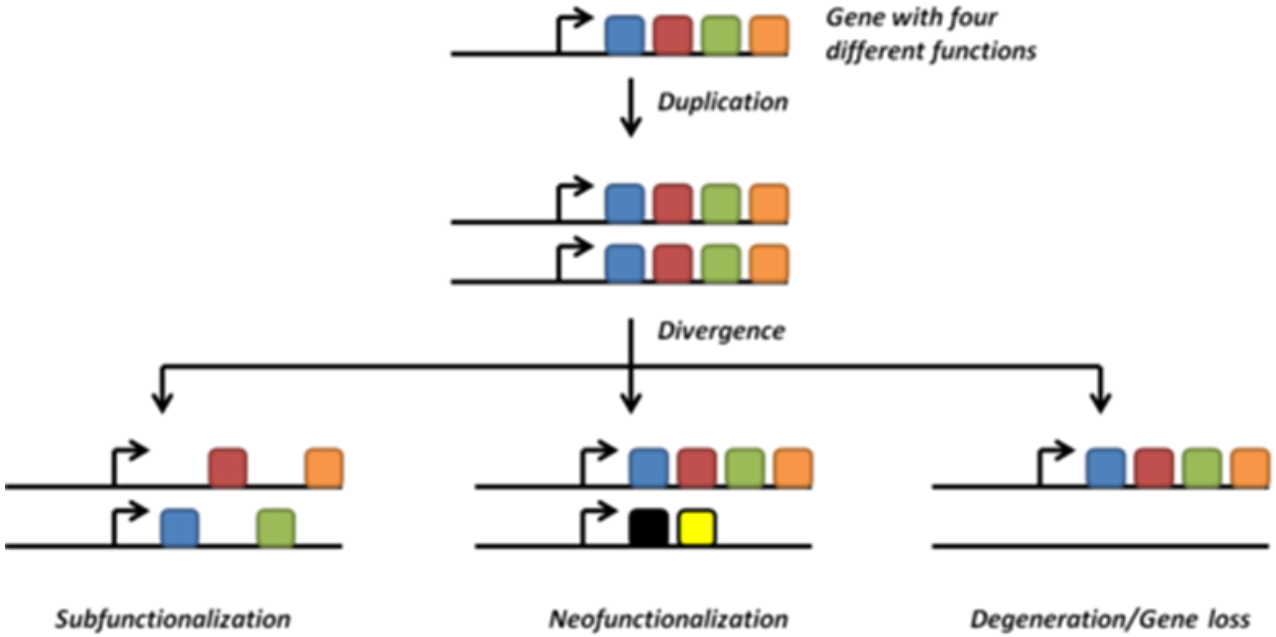


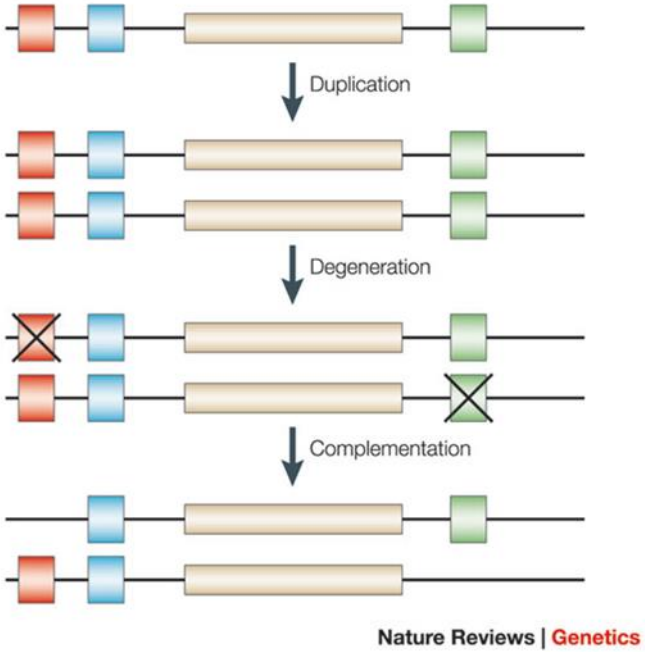
Origin of new chimeric gene or transcript structures

(A) DNA-based (genomic) gene fusion. Partial duplication (and hence fission) of ancestral source genes precedes juxtaposition of partial duplicates and subsequent fusion (presumably mediated by the evolution of novel splicing signals and/or transcription termination/polyadenylation sites).

(B) Transcription-mediated gene fusion. Novel transcript structures may arise from intergenic splicing after evolution of novel splicing signals and transcriptional readthrough from the upstream gene. New chimeric mRNAs may sometimes be reversed transcribed to yield new chimeric retrogenes. (Green, blue, red large boxes) Exons, (red right-angled arrows) transcriptional start sites (TSSs), (black connecting lines) constitutive splicing, (dotted lines) splicing of ancestral gene structures, (green lines) intergenic splicing that results in new chimeric transcripts.

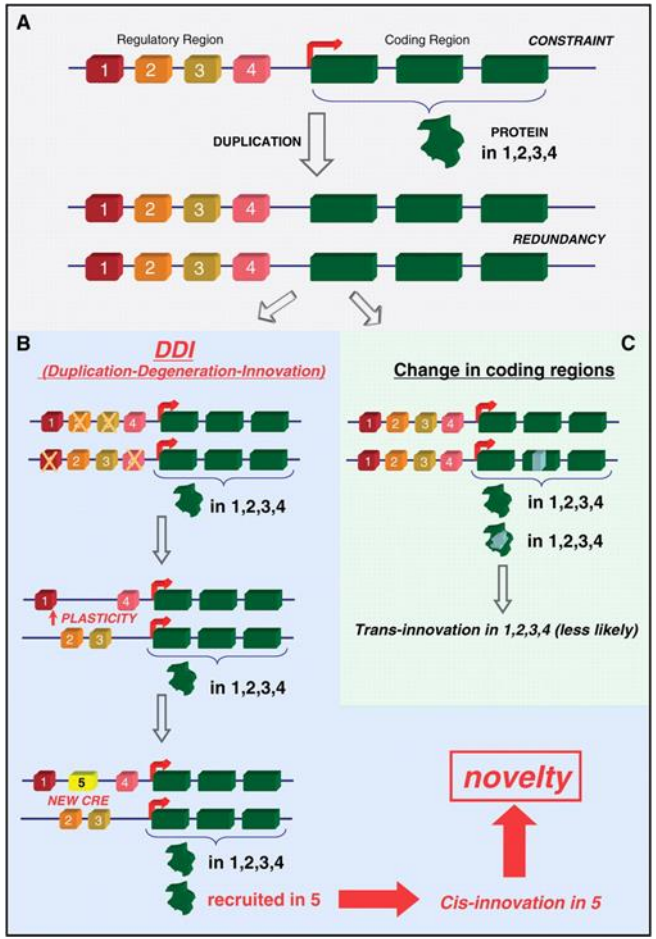
Evolutionary fate of duplicate genes





The duplication–degeneration–complementation model

The duplication–degeneration–complementation (DDC) model relies on complementary degenerative changes in a pair of duplicate genes, such that the duplicates together retain the original functions of their single ancestor. The red, blue and green boxes denote cis-regulatory elements, although degenerative mutations in any functionally discrete, independently mutable portion of a locus (a protein domain or alternative splice site, for example) could participate in sub-functionalization.



The DDI model for regulatory evolution or how simplicity after duplication and degeneration turns into plasticity and innovation

The scheme shows a **pleiotropic gene (A) regulated by four enhancers that drive expression in distinct temporal–spatial territories.** The gene is evolutionary constrained, as modifications of the coding region or the complex regulatory region may affect the functionality of the gene in multiple territories. **After gene duplication, changes in coding regions (C), as suggested by Sosumo Ohno, would allow the expression of a new protein in all the territories in which the original gene was expressed.** For high pleiotropic genes, this will likely be negatively selected, as it will affect too many processes.

(B) DDI: after the differential degeneration of regulatory enhancers in the duplicated copies (subfunctionalization), the resulting genes will had its pleiotropy reduced, and the regulation of a particular copy will be simpler than originally. This will turn into higher plasticity or evolvability of these regions, thus facilitating the appearance of new regulatory regions which will allow the recruitment of the gene and its associated gene network in a new territory, paving the path to morphological innovations.

Polyploidisation

- Global increase in genome
- Addition of one or more complete chromosome sets
- 2 copies : diploid
- 3 : triploid (sterile)
- 4 : tetraploid
- 6 : hexaploid

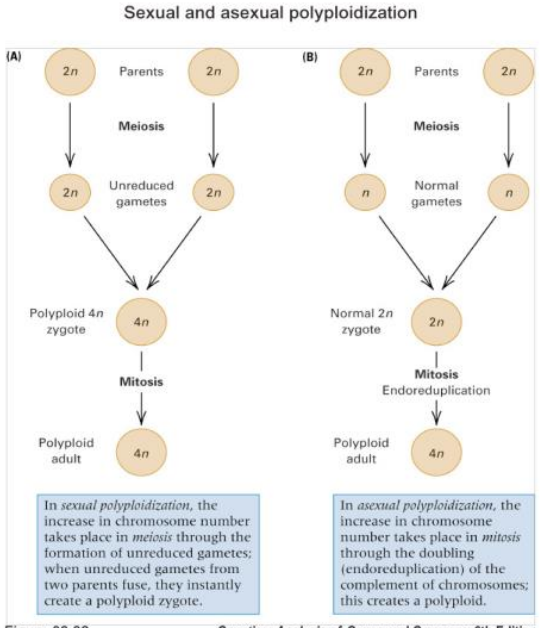


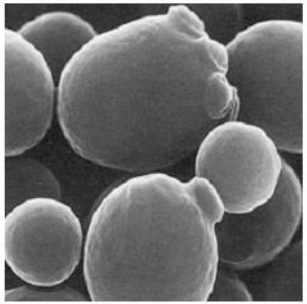
Figure 08.32 *Genetics: Analysis of Genes and Genomes*, 6th Edition
Hartl, Jones

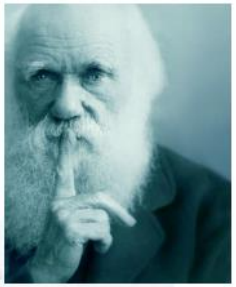
Duplication

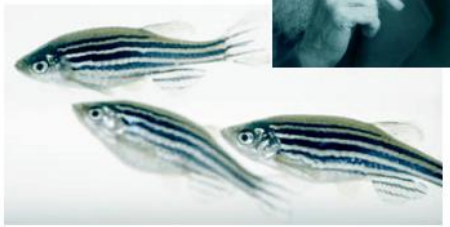
1. Polyploidy – whole genome duplication
2. Aneuploidy – chromosomal duplication
3. Partial chromosome duplication
4. Gene duplication
5. Partial gene duplication

Examples of Paleopolyploids

- Yeast
- Arabidopsis
- Wheat
- Fish
- Ancestral vertebrate (2R)







a) repeated hybridization and polyploidization in the origin of wheat b) spike of *T. turgidum*

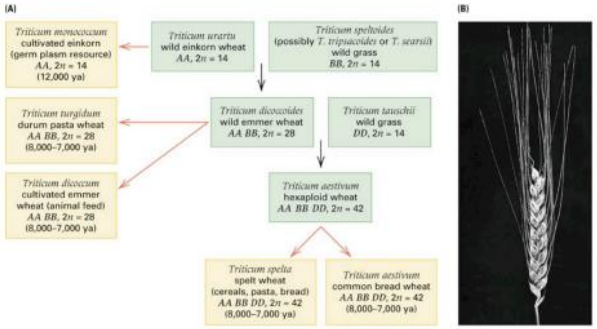


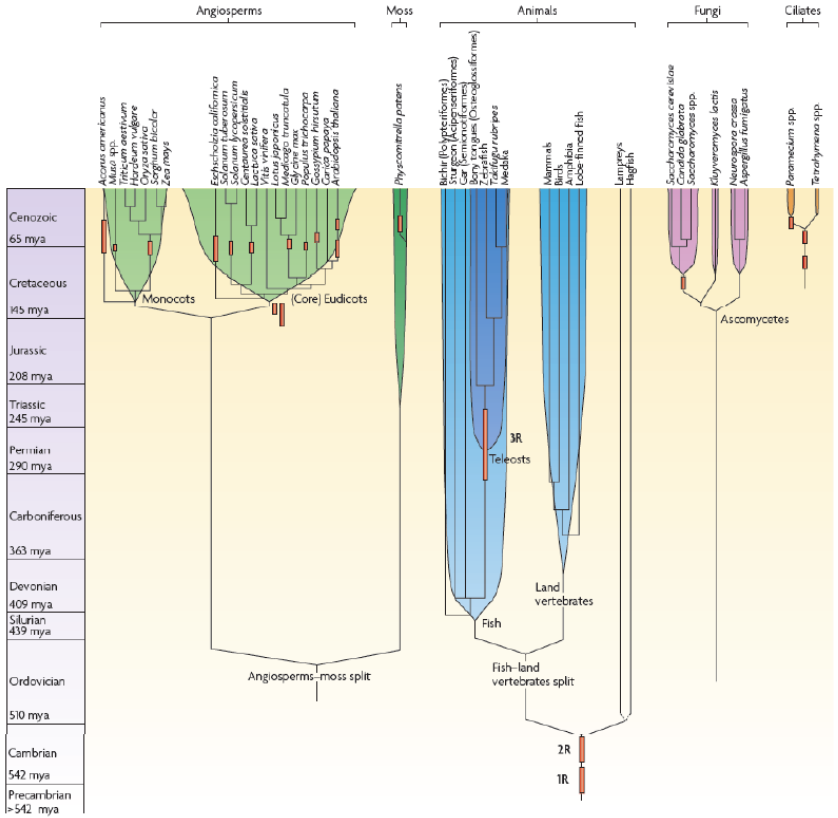
Figure 08.34

Genetics: Analysis of Genes and Genomes, 6th Edition
Hartl, Jones

Loss or retention of genes duplicated by WGD (ohnologs)

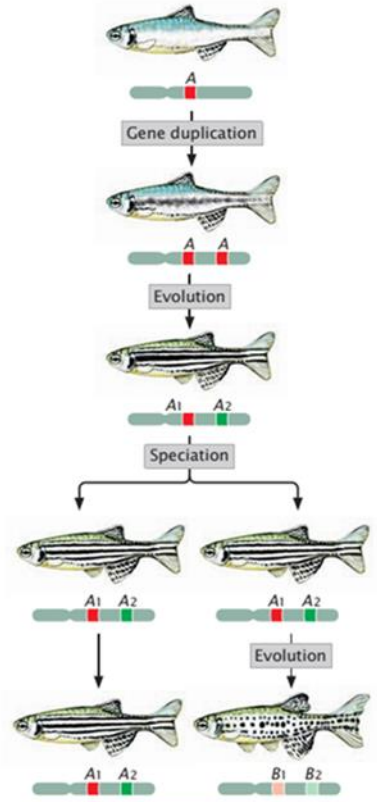
- Most duplicates are subsequently lost
- Biased retention of certain classes of genes
- Retained duplicates are enriched for:
  - Developmental genes
  - Transcription factors
  - Metabolic genes
  - Protein complex membership

Detected genome duplication events

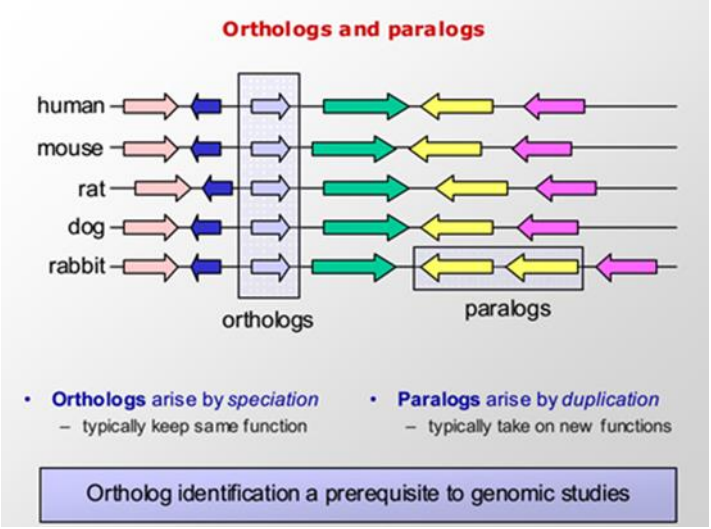


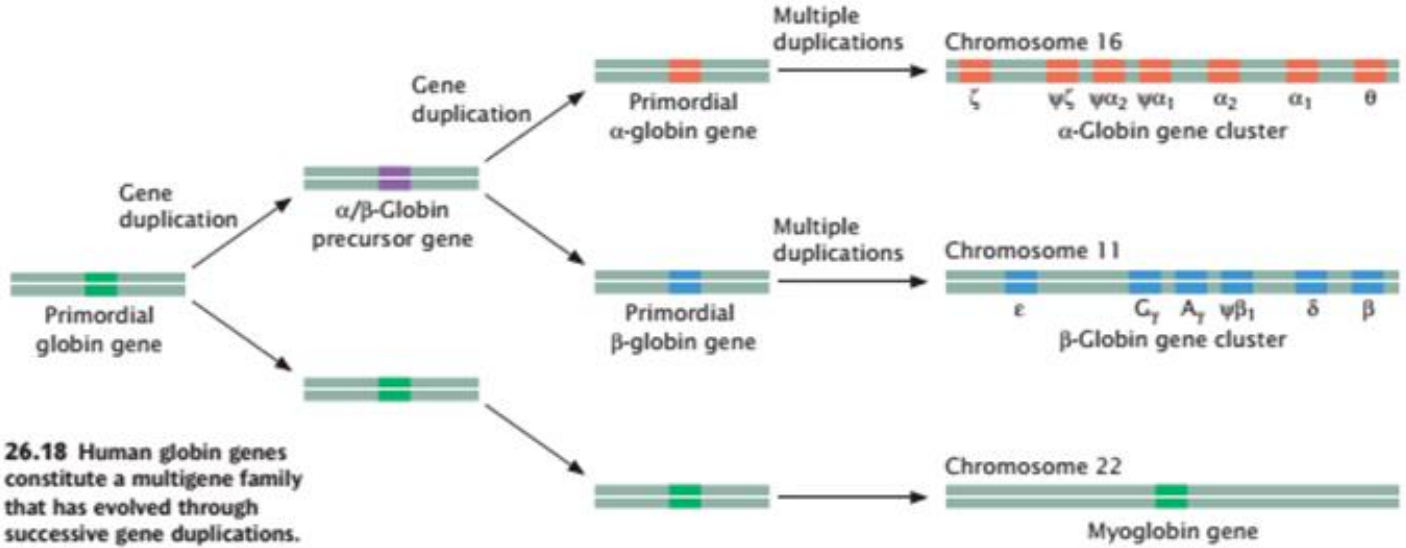
genome duplication

Van de Peer *et al*. *Nature Reviews Genetics* (2009)



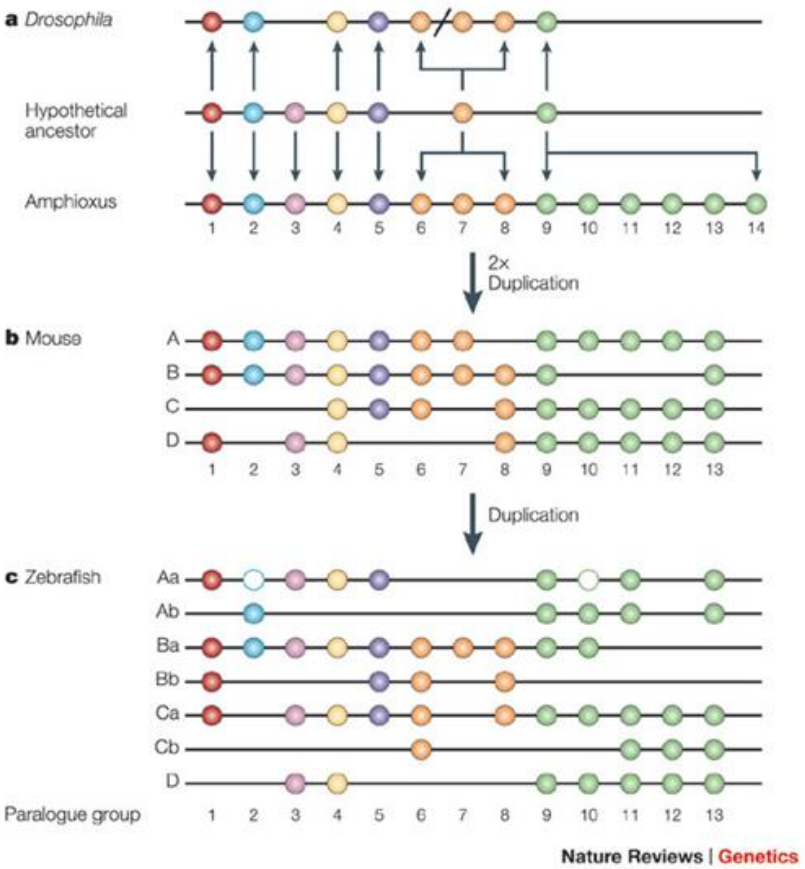
20.10 Homologous sequences are evolutionarily related. Orthologs are homologous sequences found in different species; paralogs are homologous genes in the same species and arise from gene duplication.



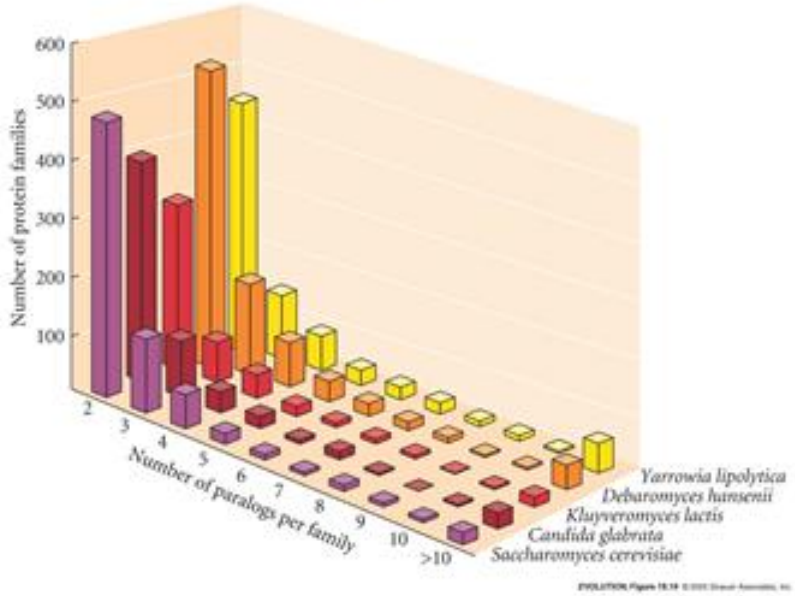


26.18 Human globin genes constitute a multigene family that has evolved through successive gene duplications.

Hox cluster evolution through gene duplication

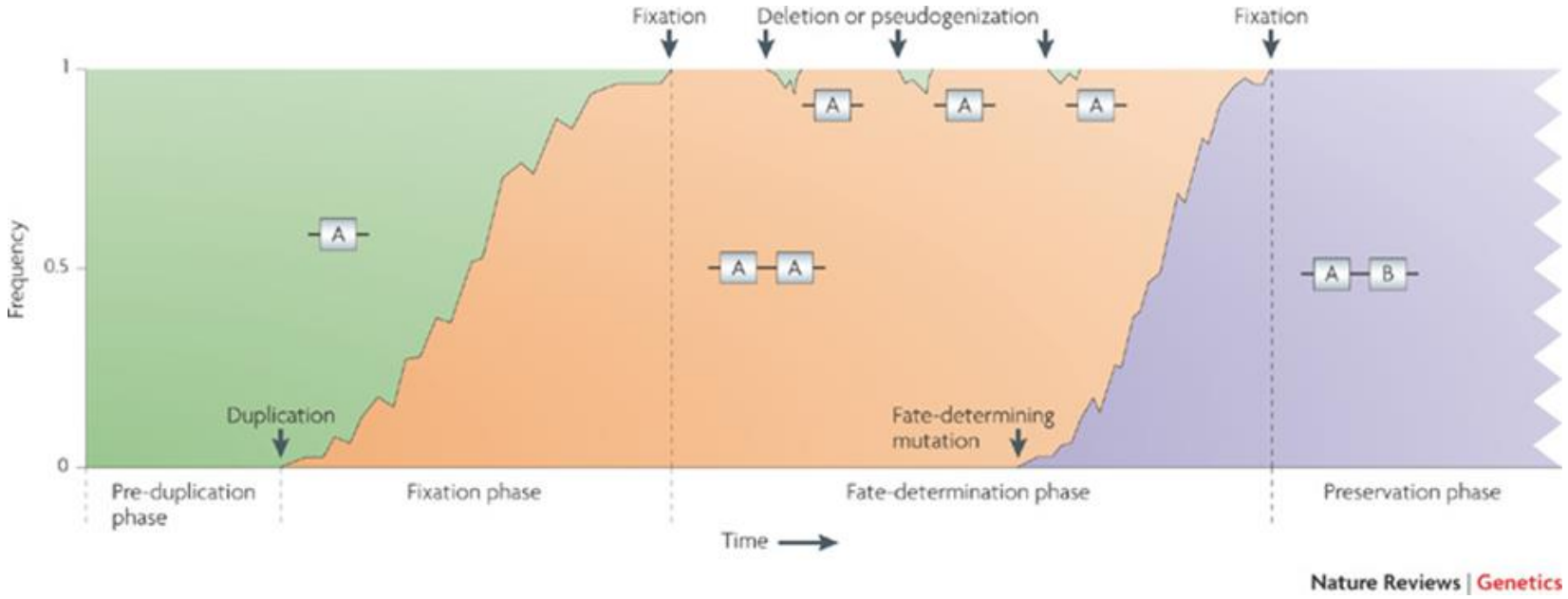


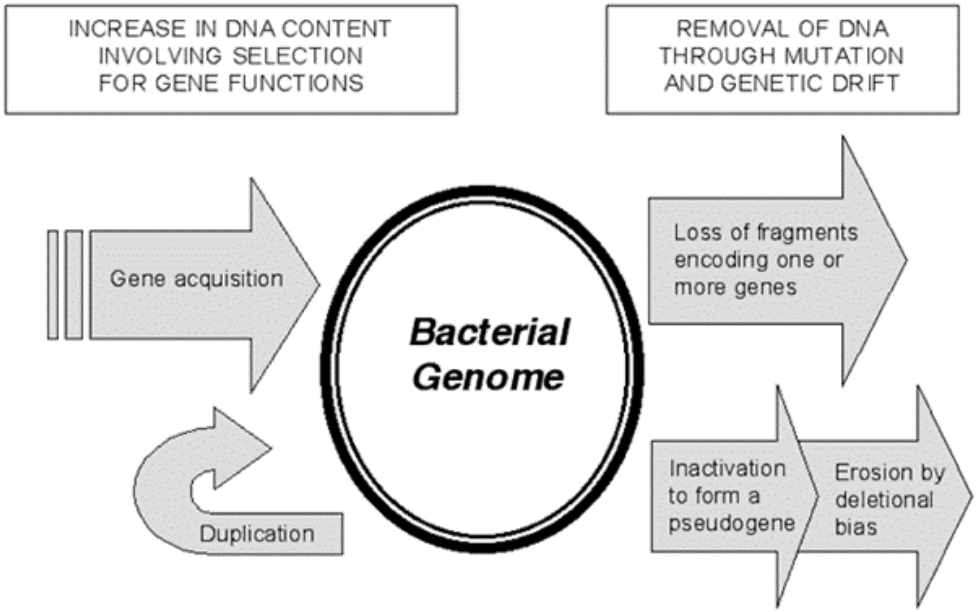
Gene Duplication

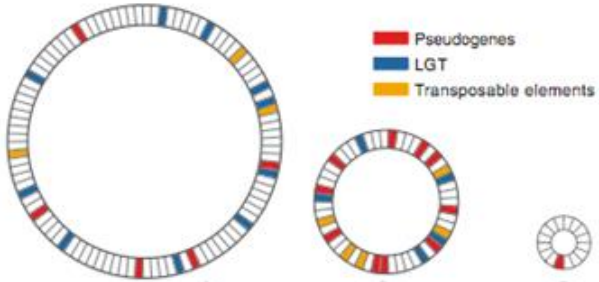


Most gene families are small; exceptions often have an adaptive basis: immunoglobulin genes (1000 copies in humans), olfactory receptor genes (100's of copies in mammals)

Phases leading to the stable preservation of a duplicated gene







	Free-living	Recent or facultative pathogen	Obligate symbiont or pathogen
Genome size	Large (5-10 MB)	Intermediate (2-5 MB)	Small (0.5-1.5 MB)
Number of pseudogenes	Few	Many	Rare
Incidence of LGT	Frequent	Frequent to rare	Rare to none
Selfish genetic elements	Few	Common	Rare
Genome organization	Stable or unstable	Unstable	Stable
Effective population size	Large	Small	Small

Fig. 1. Trends in the size and contents of bacterial genomes. Upon classifying bacteria according to their degree of association or dependence on a eukaryotic host, several general features emerge (*8*). Because of their relatively large population sizes, selection operates most effectively in free-living species—i.e., those that replicate in the environment independently of a host. Because selection is effective in removing deleterious sequences, these species usually possess large genomes containing relatively few pseudogenes (red) or mobile genetic elements (yellow). In recently derived pathogens, the availability of host-supplied nutrients combined with decreases in effective population sizes allows for the accumulation of pseudogenes and of transposable elements. In long-term host-dependent species, the ongoing mutational bias toward deletions has removed all superfluous sequences, resulting in a highly reduced genome containing few, if any, pseudogenes or transposable elements. LGT, lateral gene transfer.

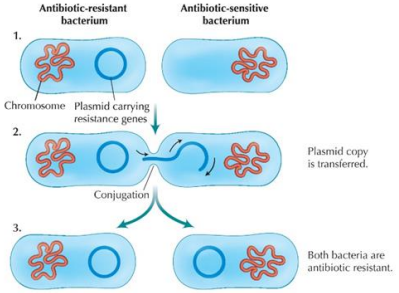


FIGURE 7.19. Lateral transfer of antibiotic resistance via plasmids. (*1*) Antibiotic-resistant and -sensitive bacteria are shown. (*2*) Bacteria "mate" via conjugation, during which a copy of the plasmid carrying antibiotic resistance genes is transferred. (*3*) Both bacteria are now antibiotic resistant.

7.19, adapted from Collignon P.J., *Med. J. Australia* **177:** 325–329, © 2002 Australasian Medical Publishing Co.

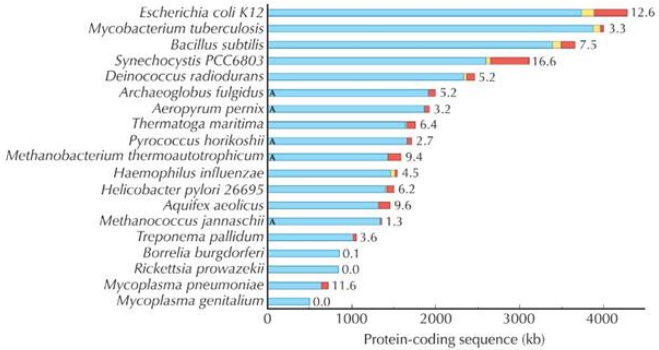


FIGURE 7.23. Estimation of the percentage of the genome that has been recently acquired by lateral transfer for different species of bacteria and archaea. *Blue,* "native" DNA (not acquired by transfer); *yellow,* known mobile DNA elements; *red,* other foreign DNA; and A, archaeal species.

7.23, redrawn from Ochman H. et al., *Nature* **405:** 299–304, © 2000 Macmillan, www.nature.com

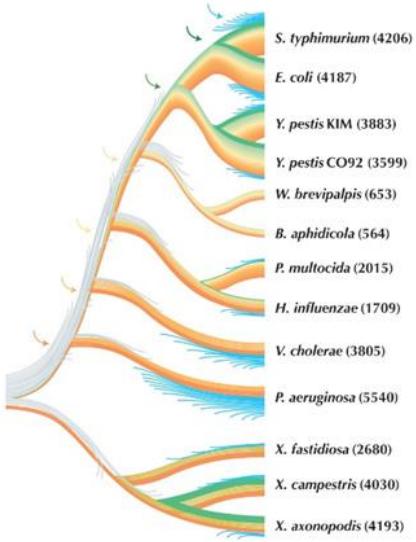


FIGURE 7.25. Lateral gene transfer and genome evolution in γ-proteobacteria. Only a small proportion of genes have been retained since the common ancestor of γ-proteobacteria (in *orange*). Under the assumption that ancestral and contemporary genome sizes are similar, most of the genes present in this ancestral genome (in *gray*) have been replaced by nonhomologous genes (*yellow* to *green*), usually via lateral gene transfer from organisms outside of this clade. Once a new gene is acquired, its transmission follows vertical inheritance. The abundance of genes unique to a species (in *blue*) indicates that these bacteria (with the exception of the endosymbionts *W. brevipalpis* and *B. aphidicola*) constantly acquire new genes, most of which do not persist in the long term within lineages. (Numbers of protein-coding genes, excluding those corresponding to known insertion sequence elements and phage, are in parentheses for each genome.)

7.25, redrawn from Lerat E. et al., *PLoS Biol.* **3:** E130, © 2005 Public Library of Science