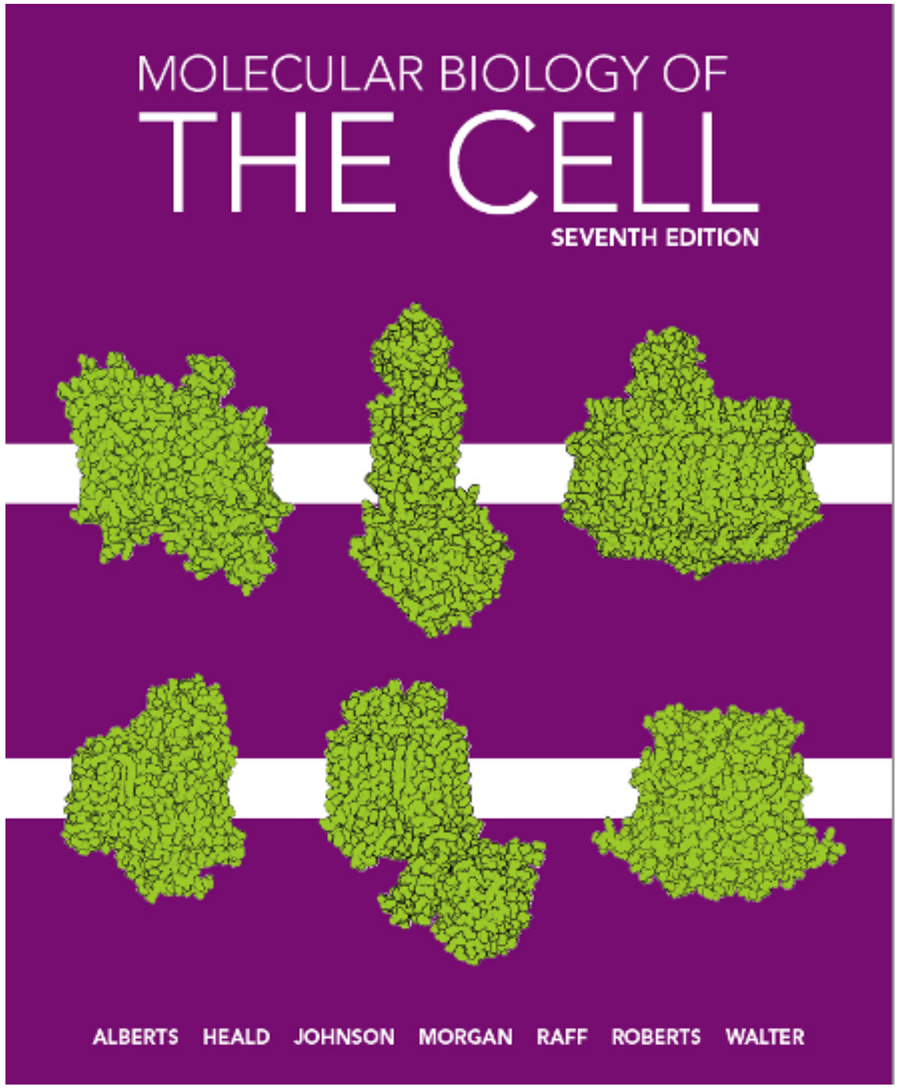
# Cellular and Molecular Biology I

## BIO-205-3

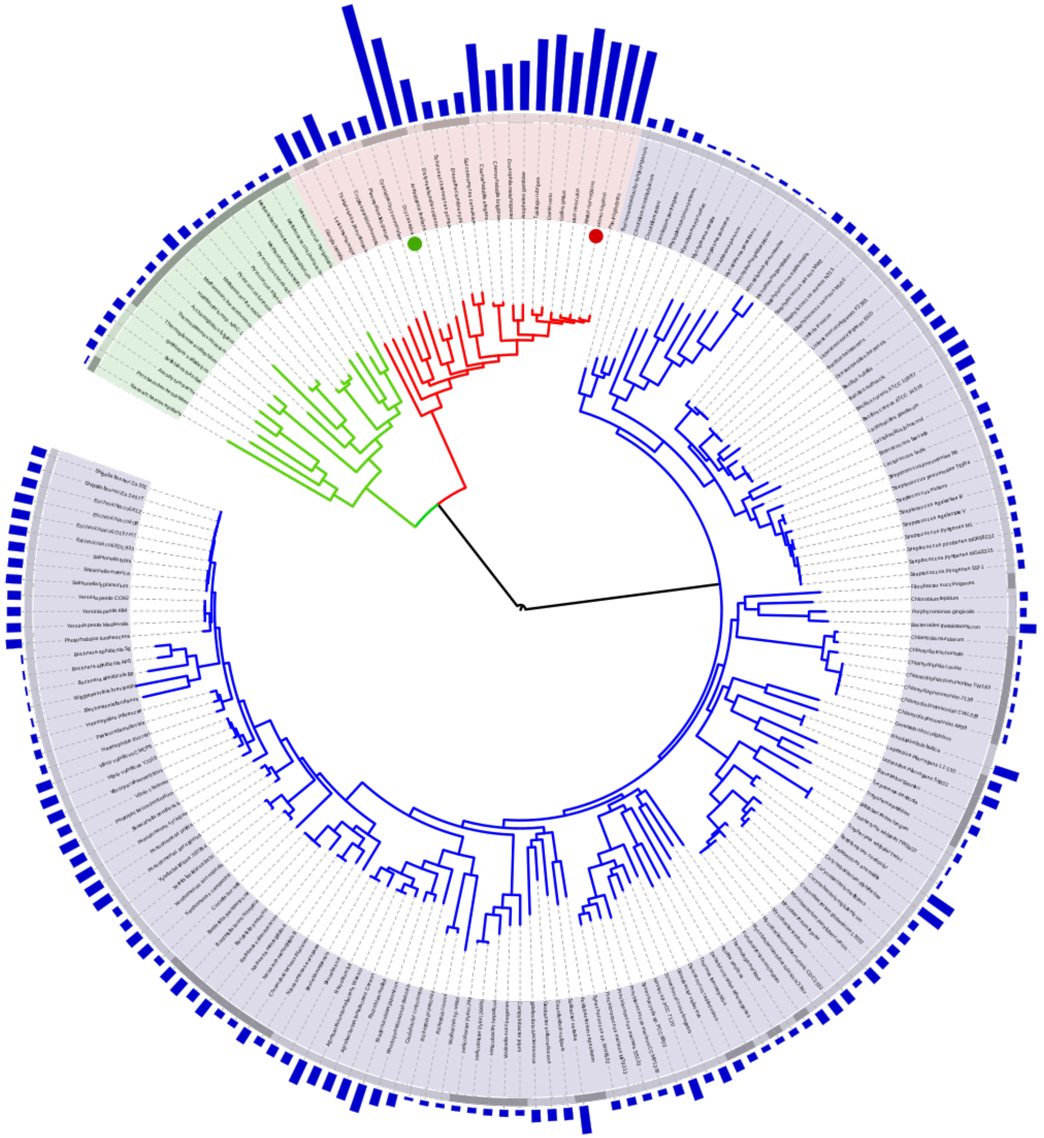
**Camille Goemans - 2024**




# Chapter 4

## DNA, Chromosomes, and Genomes

# Quick recap: genomes



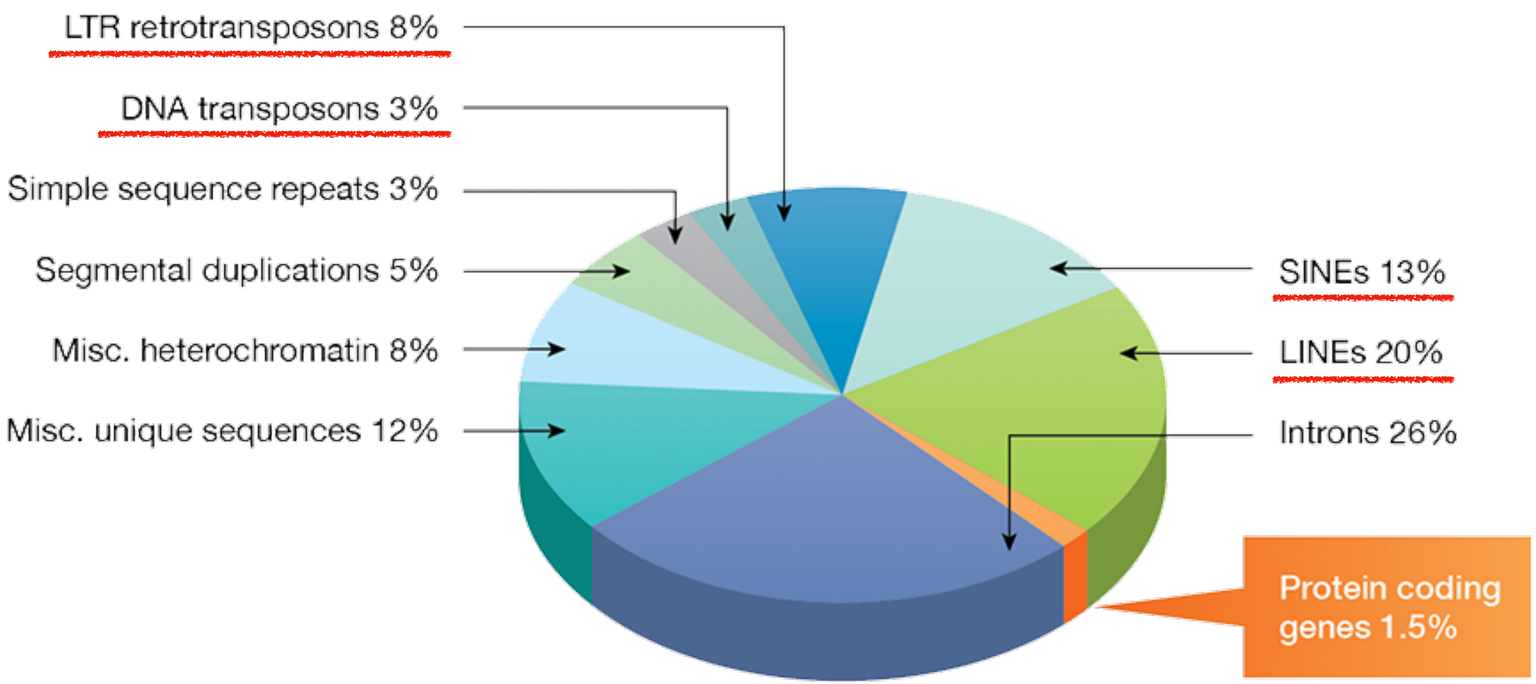
# Quick recap: genome evolution

- Evolution depends on **accident and mistakes** followed by **non-random survival**
- **Failures** in the mechanisms by which genomes are **copied or repaired**
- When errors (mutations) happen in **germ cells**, they are passed on to the next generation
- Errors are **“rare”** events: ~ 1 in $10^8$ per generation (implying that each gamete has in average 30 mutations)

What are the different **types of mutations**?

- Simple, local changes - **point mutations**
- Large-scale genome rearrangements - **deletions, duplications, inversion, translocations**
- **In addition, important role of mobile genetic elements**

# Quick recap: the human genome



# Plan

- Quick recap
- Mobile genetic elements
- Comparing genomes
- The maintenance of DNA sequences
- DNA replication mechanisms

# Mobile genetic elements

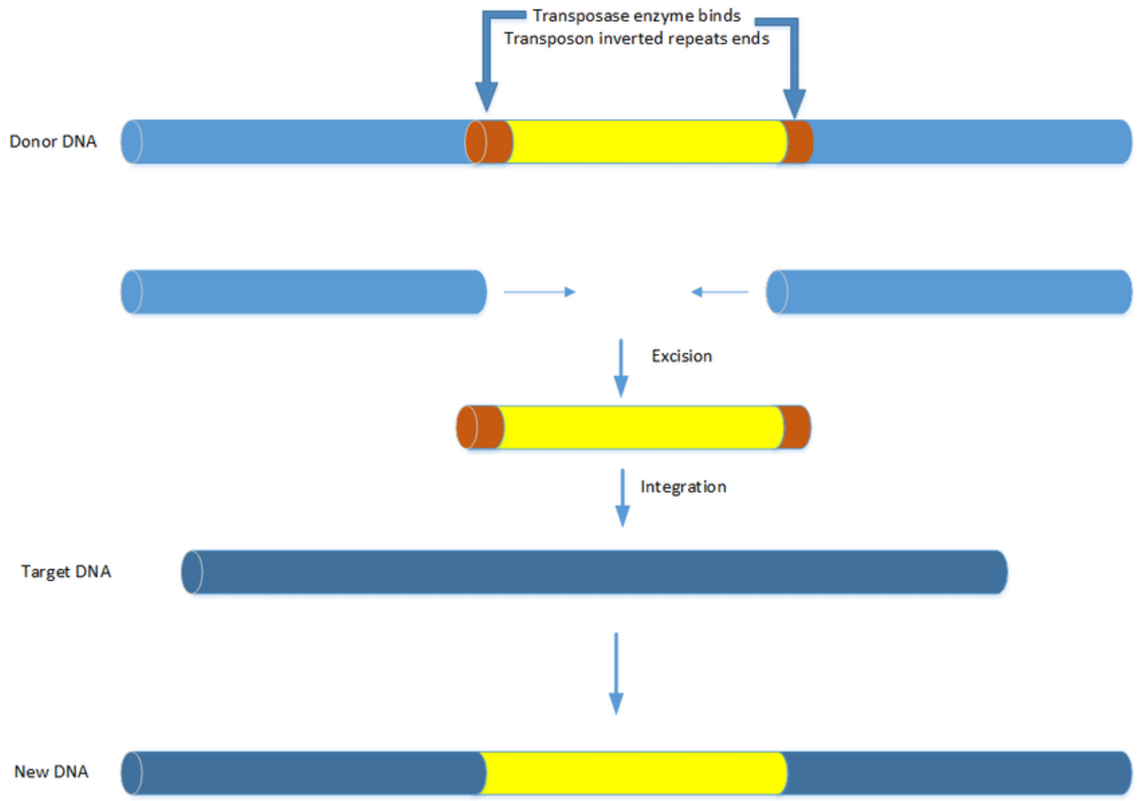
- **All cells** contain mobile genetic elements
- **Mobile genetic elements**
    - Typically between few hundreds to tens of thousands base pairs
    - Each carries a unique set of genes
    - Often, one of these genes catalyses the movement of that element
- They **move** within the genome
- This can involve **replication** or not
- In the process, they might **disrupt the function** or alter the **regulation** of existing genes

# Mobile genetic elements

- They are typically part of the **repeated sequences** in our genome
- Overtime, random mutations affected their sequence and only a **small fraction is still active**
- Considered to be **molecular parasites** (or selfish DNA) that persist in cells because they cannot get rid of them
- The genes they carry can provide an **advantage to the host** (antibiotic resistance in bacteria)
- Over long periods of evolution, transposons are considers as **drivers of evolution and biodiversity**

# Transposons

Transposable DNA elements (=**transposons**) are parasitic DNA sequences that can integrate and spread in the genomes they colonise

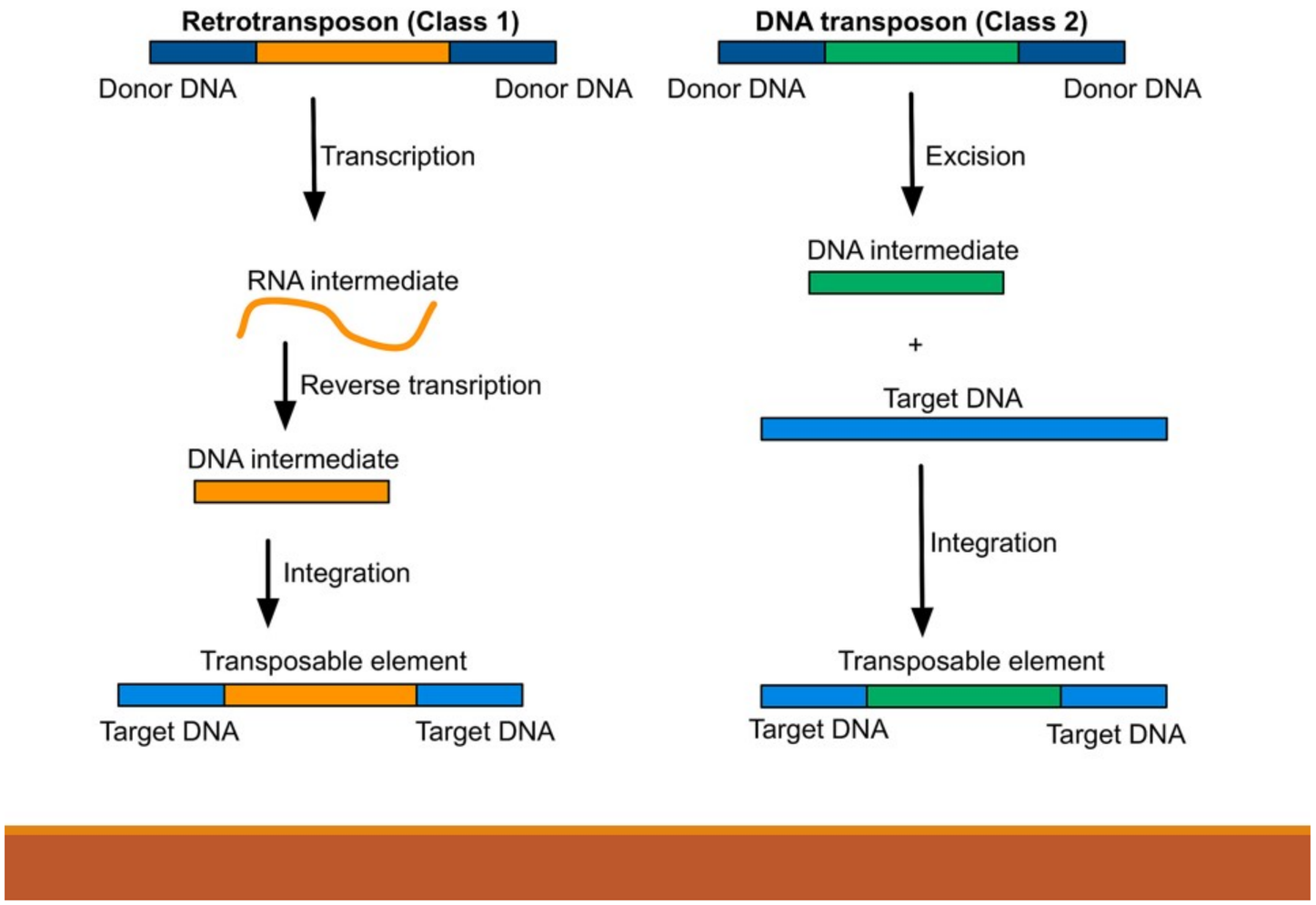



https://www.researchgate.net/publication/311653875_Local_Alignment_on_Highly_Unbalanced_DNA_Sequence_Lengths

# Transposons

- Mobile elements that move by transposition are called **transposons**
- A **transposase** (usually encoded by the transposon) acts on specific DNA sequences at each end of the transposon, causing it to insert at a new DNA location
- Not very **selective** in choosing their target site
- Most transposons move **rarely**
- In bacteria, transposons move every **$10^5$ cell divisions** - more frequent movement would destroy the host genome (not advantageous for the transposon)
- Transposons belong to **3 large classes**, based on their structure and transposition mechanisms

# Transposons




https://www.onlinebiologynotes.com/transposable-elements-in-eukaryotes/

# Transposons

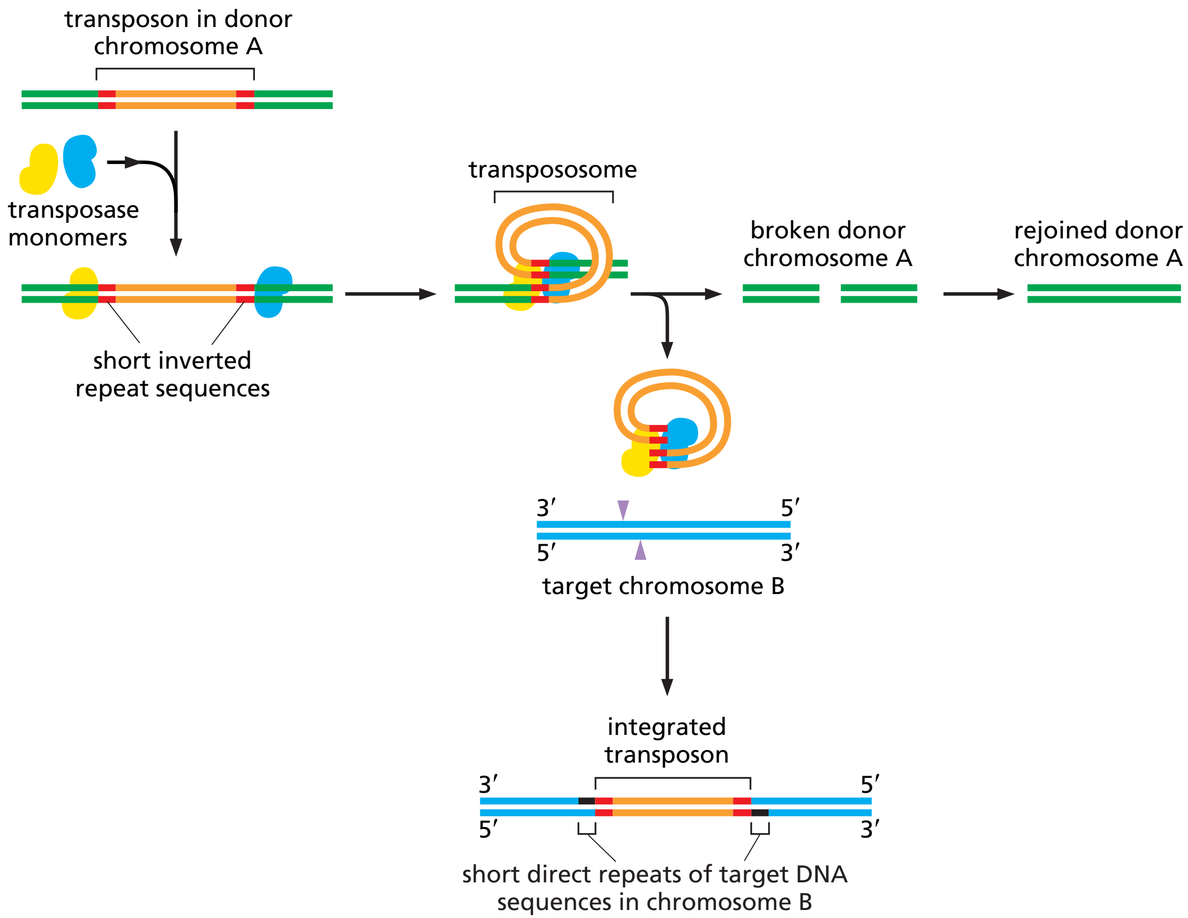
**TABLE 5–4 Three Major Classes of Transposable Elements**

| Class description and structure | Specialized enzymes required for movement | Mode of movement | Examples |
|---|---|---|---|
| **DNA-only transposons** | | | |
|  Short inverted repeats at each end | Transposase | Moves as DNA, either by cut-and-paste or replicative pathways | P element *(Drosophila)*, Ac-Ds (maize), Tn3 and Tn10 *(E. coli),* Tam3 (snapdragon) |
| **Retroviral-like retrotransposons** | | | |
|  Directly repeated long terminal repeats (LTRs) at each end | Reverse transcriptase and integrase | Moves via an RNA intermediate whose production is driven by a promoter in the LTR | Copia *(Drosophila),* Ty1 (yeast), THE1 (human), Bs1 (maize) |
| **Nonretroviral retrotransposons** | | | |
|   Poly A at 3′ end of RNA transcript; 5′ end is often truncated | Reverse transcriptase and endonuclease | Moves via an RNA intermediate that is often synthesized from a neighboring promoter | F element *(Drosophila)*, *L1* (human), Cin4 (maize) |

These elements range in length from 1000 to about 12,000 nucleotide pairs. Each family contains many members, only a few of which are listed here. Some viruses can also move in and out of host-cell chromosomes by transpositional mechanisms. These viruses are related to the first two classes of transposons.

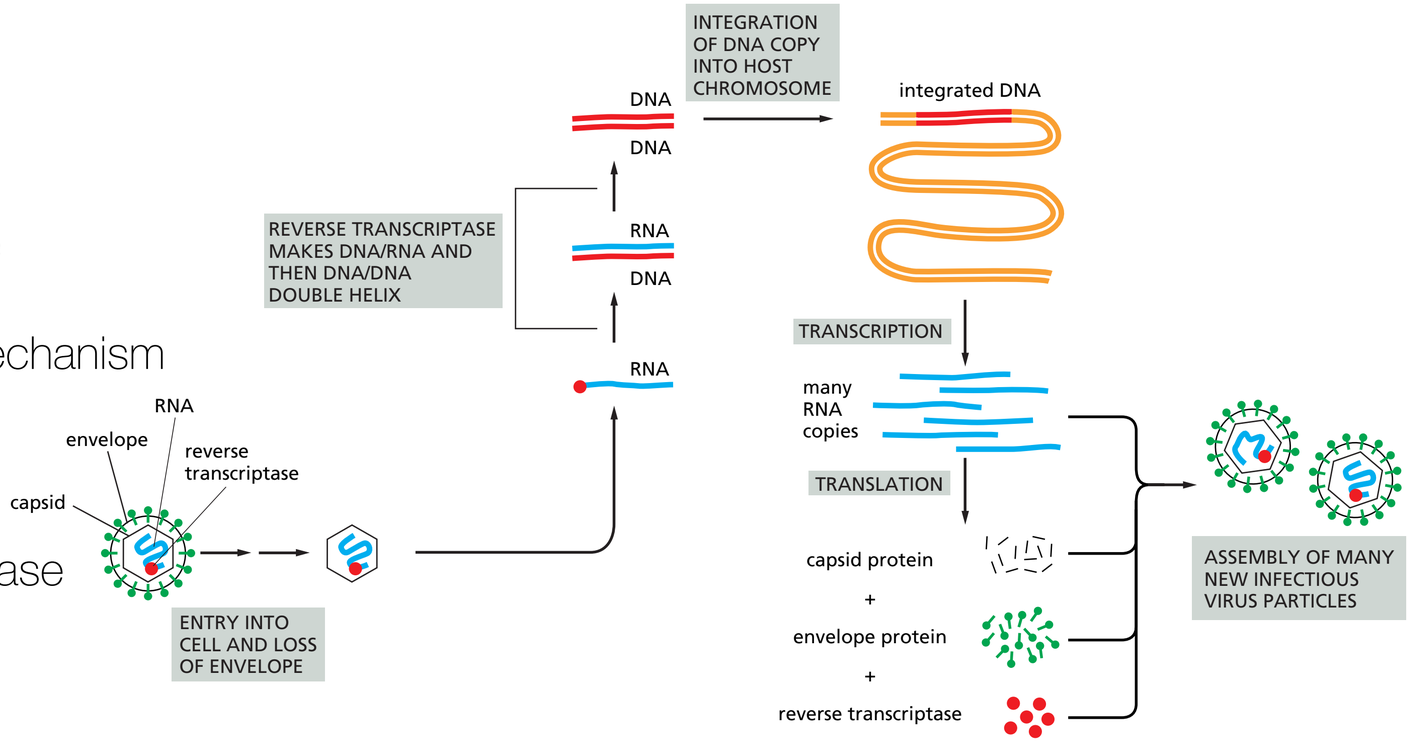
# DNA-only transposons

- Predominate in **bacteria**
- Largely responsible for the spread of **antibiotic resistance** - these elements can be moved from one bacteria to the other by **horizontal gene transfer**
- Once inserted in a cell, it will be passed to the **progeny**
- **Cut-and-paste** transposition
- The “hole” left by excision is repaired by recombinational **double-strand break repair**
- **Short direct repeats** around the transposon allow to track transposition in genomes



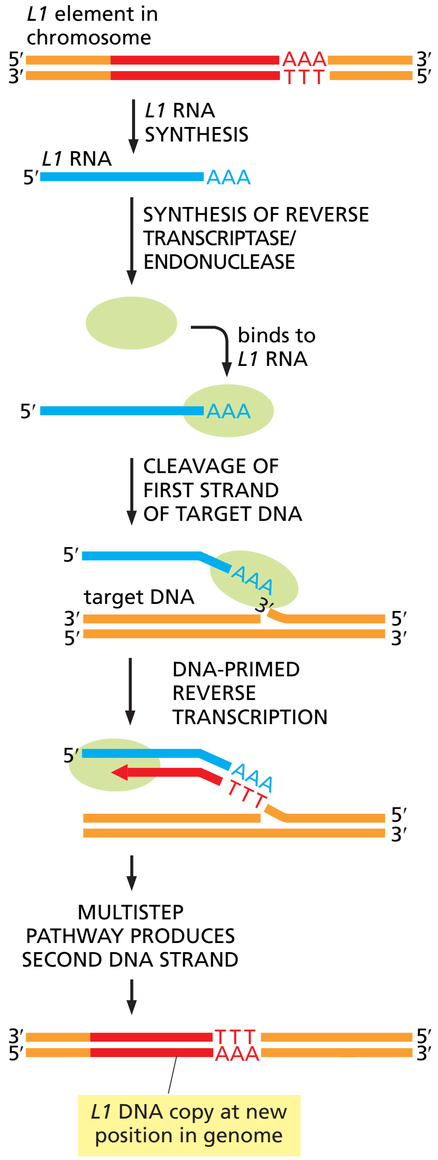
# Transposition used by viruses

- Some viruses are considered mobile genetic elements because they can insert their genome in the host DNA by transposition - but they can also infect other cells
- Transposition has a key role for specific viruses like retroviruses (e.g. HIV)
- Important role of the retrotranscriptase and integrase
- Retroviral-like retrotransposons move by a similar mechanism
    1. Transcription of the whole transposon
    2. Production of the retrotranscriptase and integrase
    3. Double-strand DNA copy from the RNA
    4. Integration in the genome by integrase
- They cannot leave their host cell (unlike viruses)

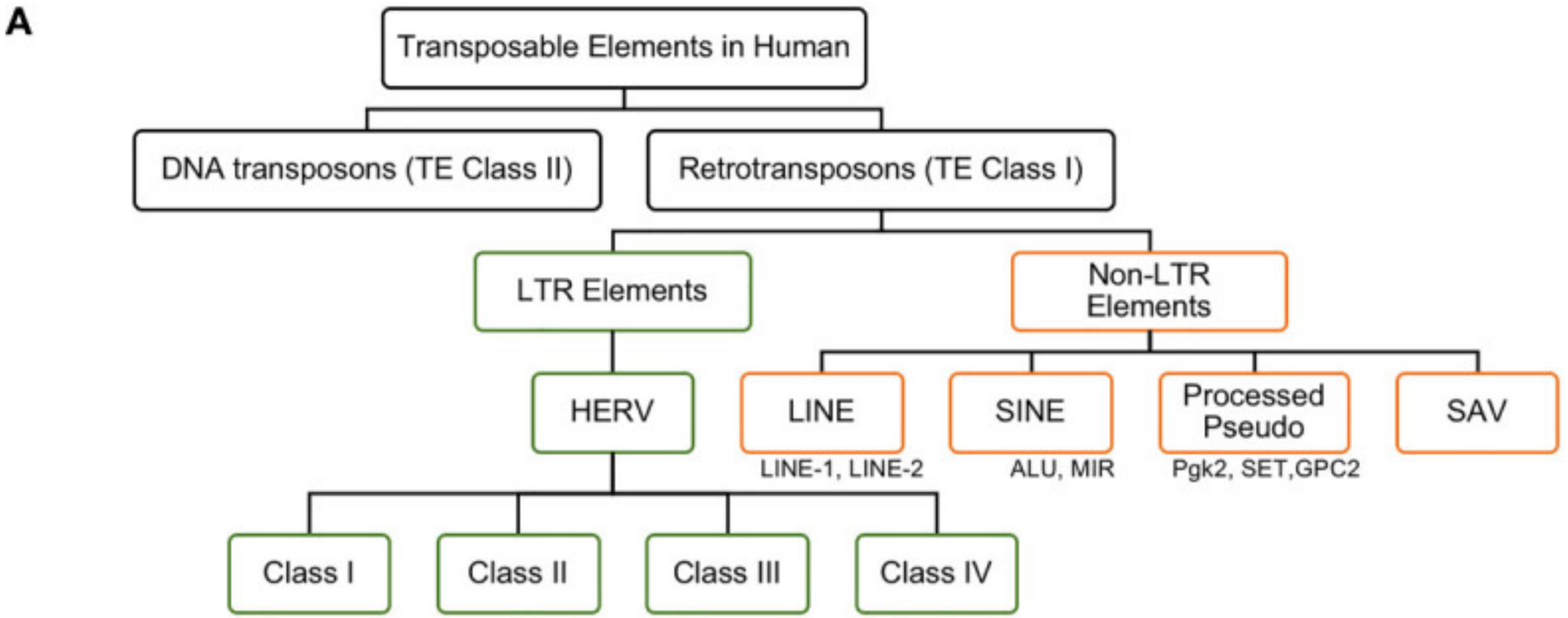


# Nonretroviral retrotransposons

- A significant fraction of vertebrates chromosomes
- A few retained the ability to move
- Need an endonuclease and a reverse transcriptase
- LINEs and SINEs make up over 30% of our genome

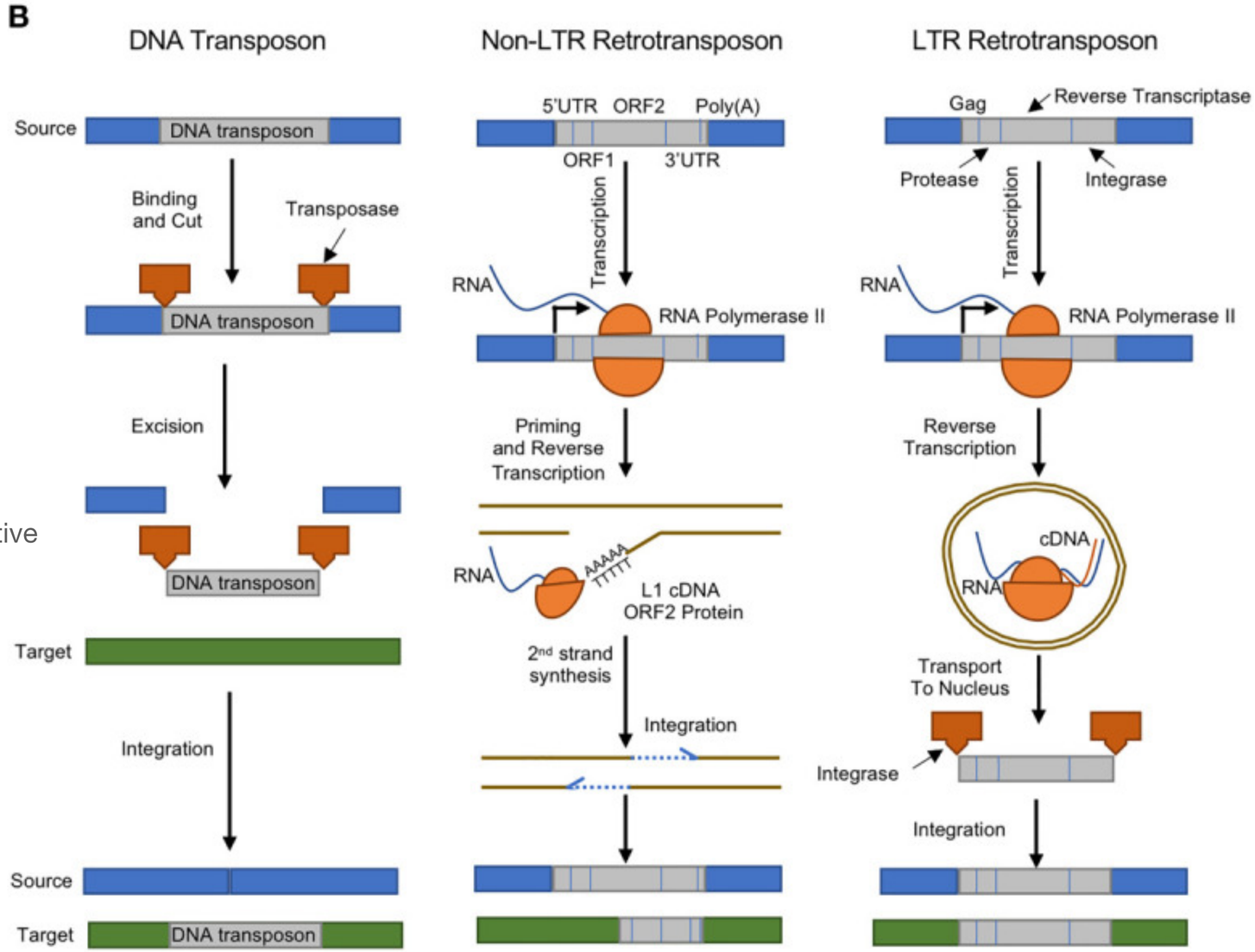


# Mobile genetic elements in Humans




https://europepmc.org/article/med/34867957

# Transposition mechanisms



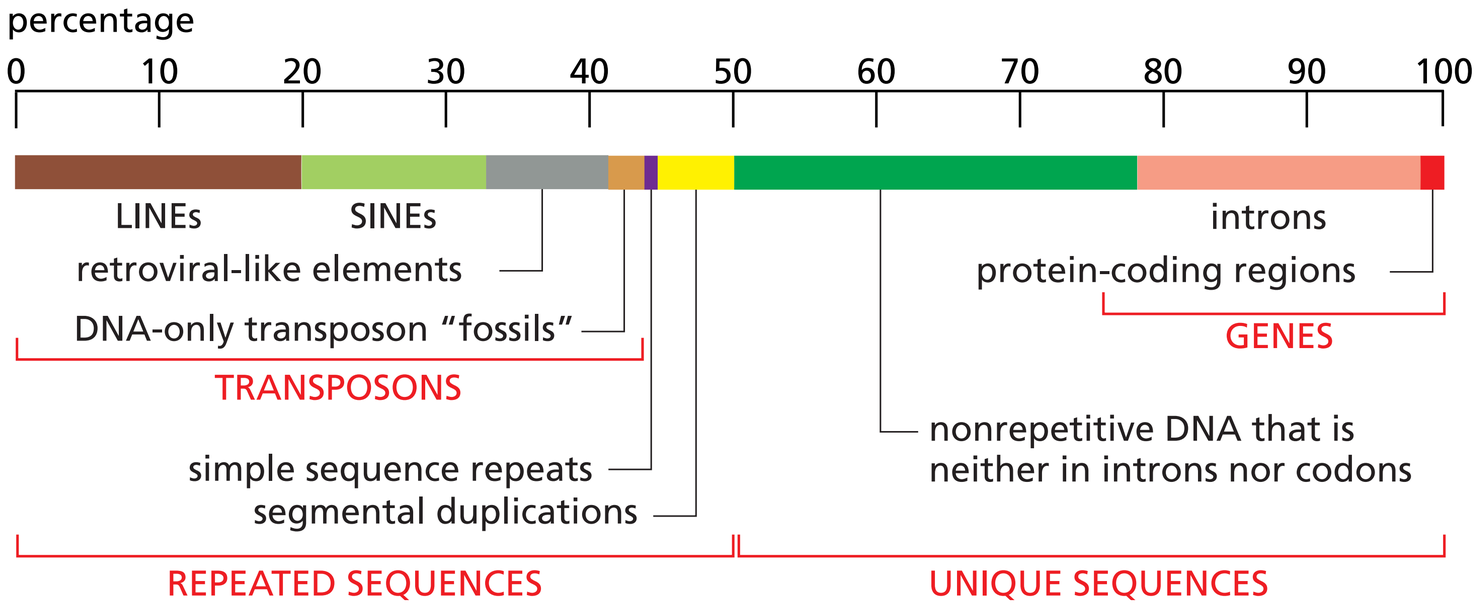

Can be replicative

LTR=long terminal repeats

https://europepmc.org/article/med/34867957

# Mobile genetic elements in the human genome

Transposable DNA elements (=**transposons**) are parasitic DNA sequences that can integrate and spread in the genomes they colonise



# Using transposons in research

Building Tn-libraries and Tn-Seq

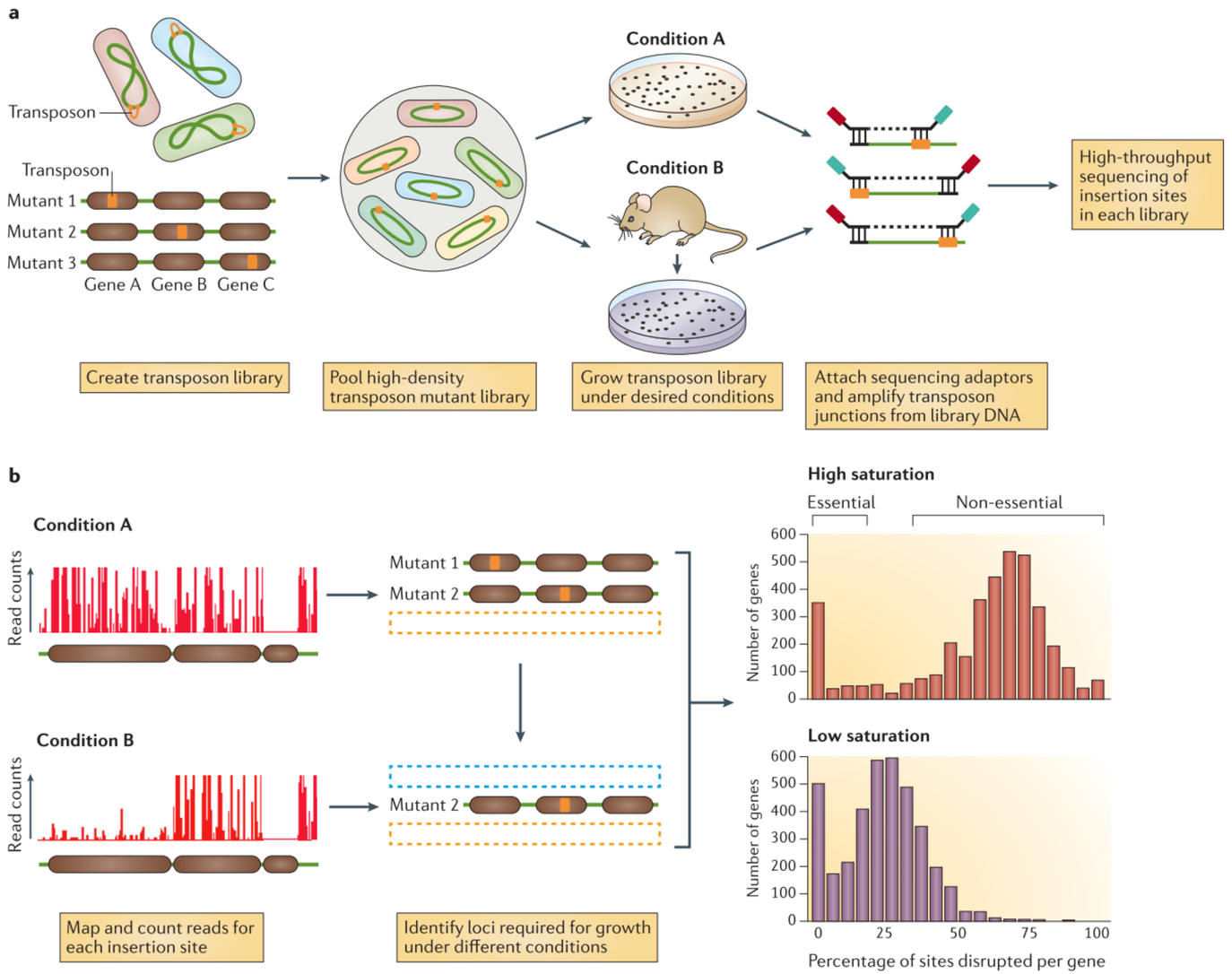


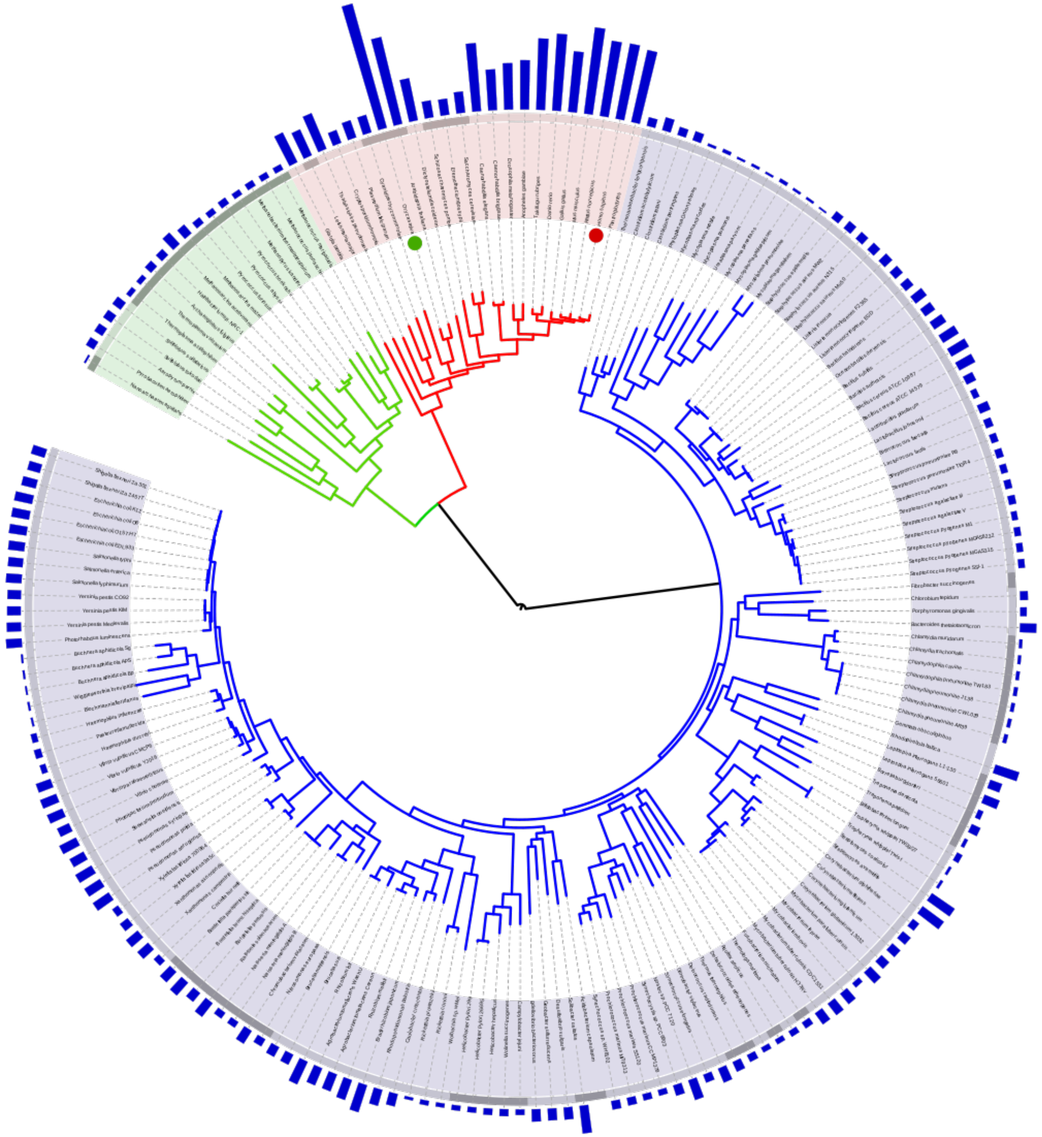

**Figure 1**: Transposon insertion sequencing method (from Chao *et al.* 2016)

# Plan

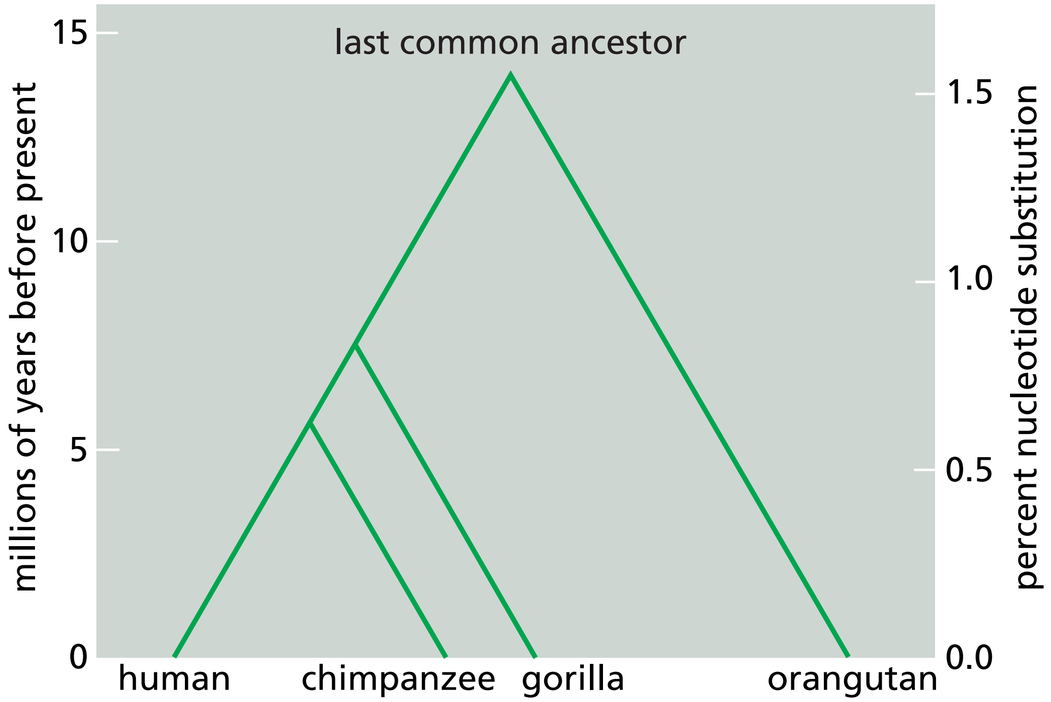
- Quick recap
- Mobile genetic elements
- Comparing genomes
- The maintenance of DNA sequences
- DNA replication mechanisms

# How can we reconstruct evolution?

- Differences in genomes have accumulated over 3 billon years
- **Nucleotide substitution rate** reflects the time available for the accumulation of mutations
- Comparative genomics use **phylogenetic trees**, built using comparison of genes or protein sequences
- Timing has been calibrated with fossil record and mutations occur at a nearly constant rate (**molecular clock** for evolution)
- Some clocks run faster than others



# How can we reconstruct evolution?



# How can we reconstruct evolution?

- **Mutation rate** is different in different parts of the genome

**Negative selection or purifying selection** is the selective removal of alleles that are deleterious

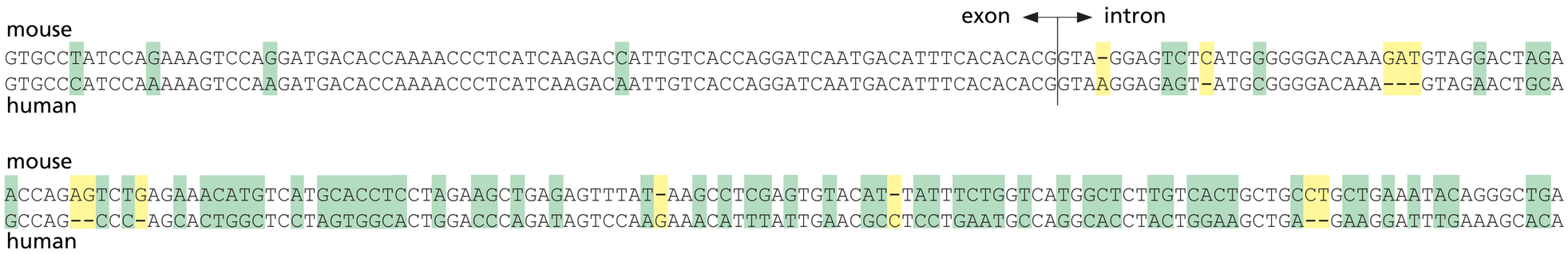
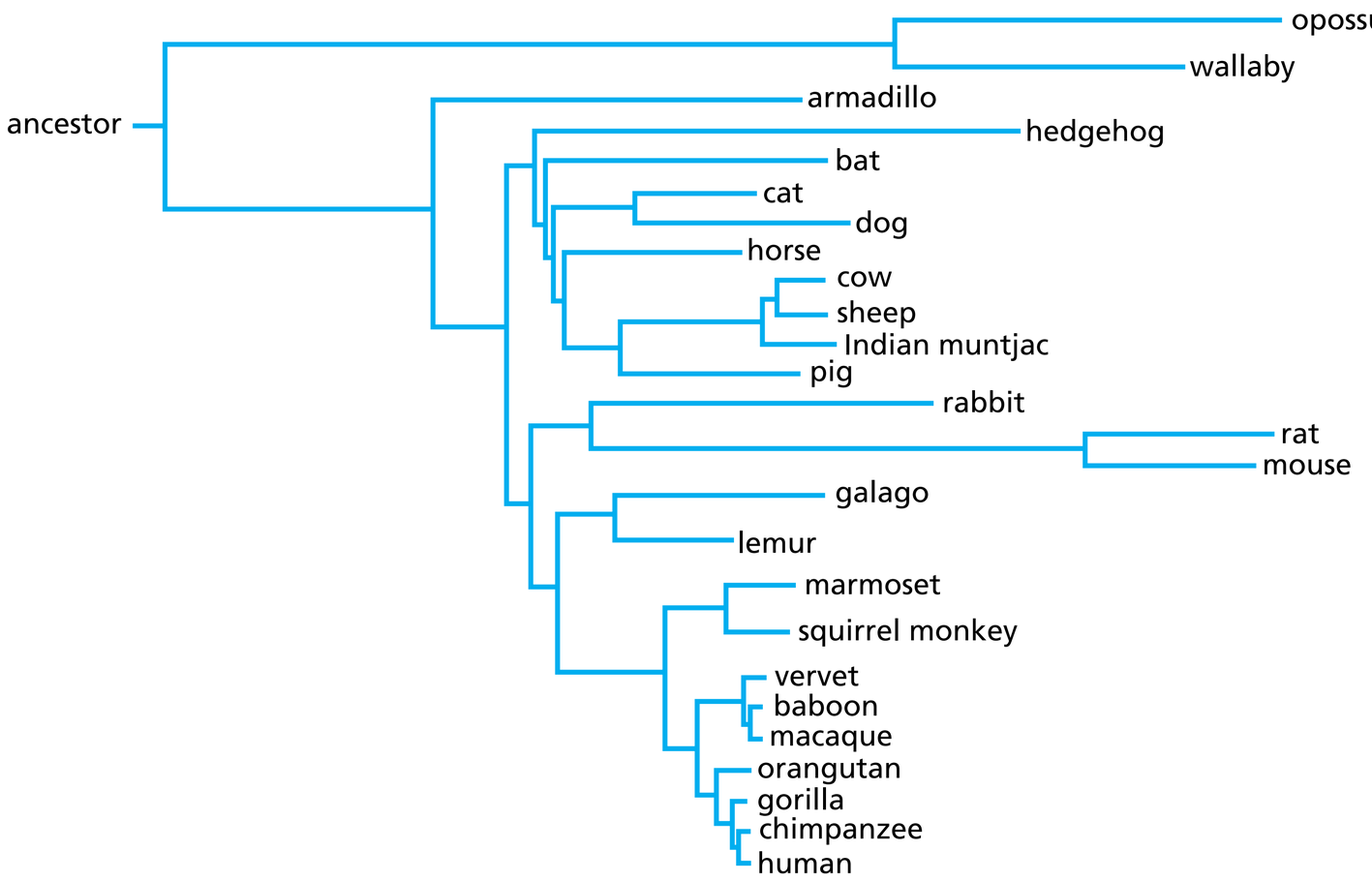



Figure 4–65 **The very different rates of evolution of exons and introns, as illustrated by comparing a portion of the mouse and human leptin genes.** Positions where the sequences differ by a single nucleotide substitution are boxed in *green*, and positions that differ by the addition or deletion of nucleotides are boxed in *yellow*. Note that, thanks to purifying selection, the coding sequence of the exon is much more conserved than is the adjacent intron sequence.

# Comparing human and mouse genomes

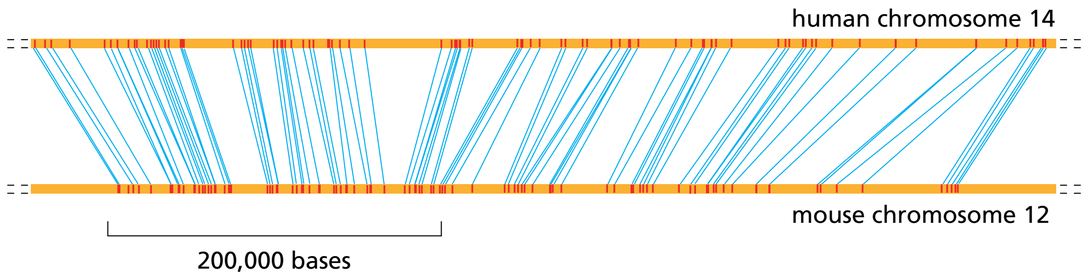


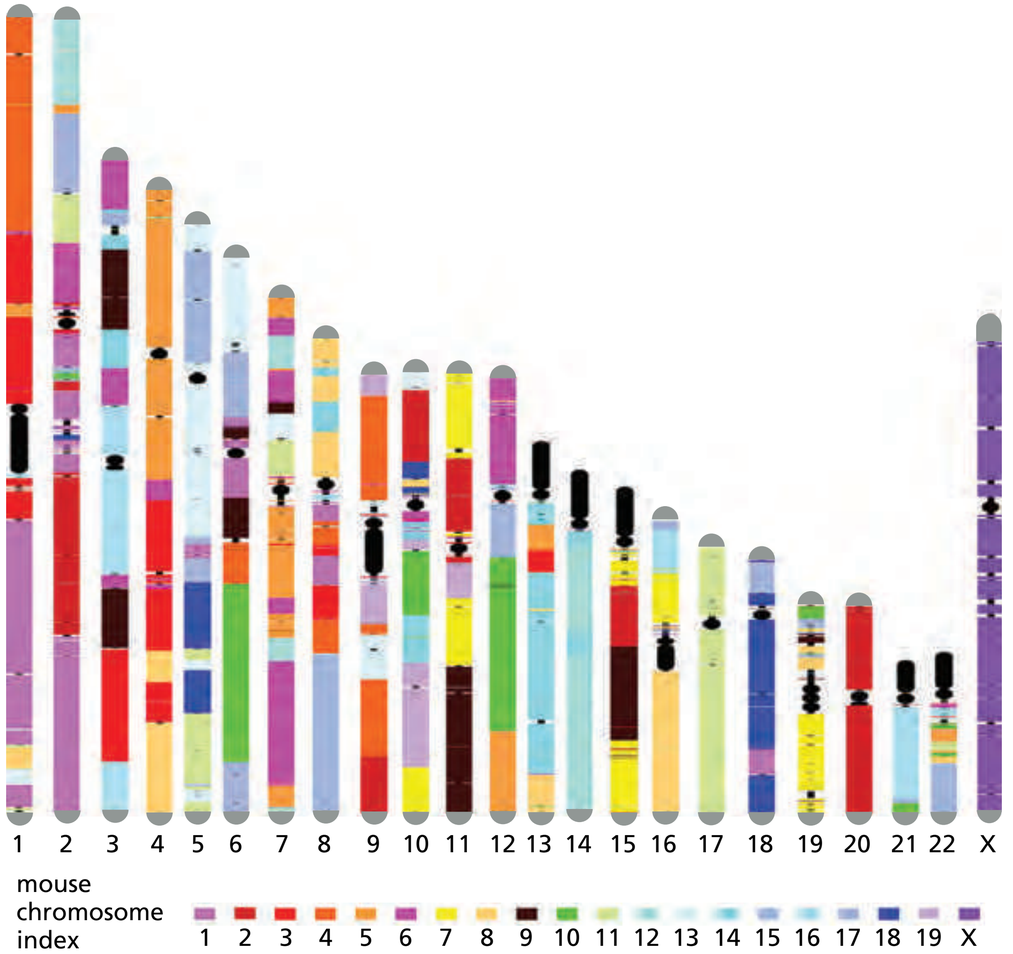

- Human and chimpanzee genomes are much more alike than mouse genome, although they have similar sizes and carry a nearly identical set of genes
- Mouse and human have diverged 80 million years ago (only 6 million for human and chimpanzees)
- Rodents have fast molecular clocks (generation time is shorter)

# Comparing human and mouse genomes

- Mouse (20 chromosome pairs); Human (23 chromosome pairs)
- Heavy **chromosomal rearrangement** during evolution (e.g. 180 breakage/fusion events)
- DNA was lost over evolution in mice

Stretches of DNA with conserved gene order are called **synteny**



# Short introns do not affect gene function

- The size of vertebrate genomes reflects the rate of **deletion and addition of DNA**
- Large difference in genome size between similar organisms

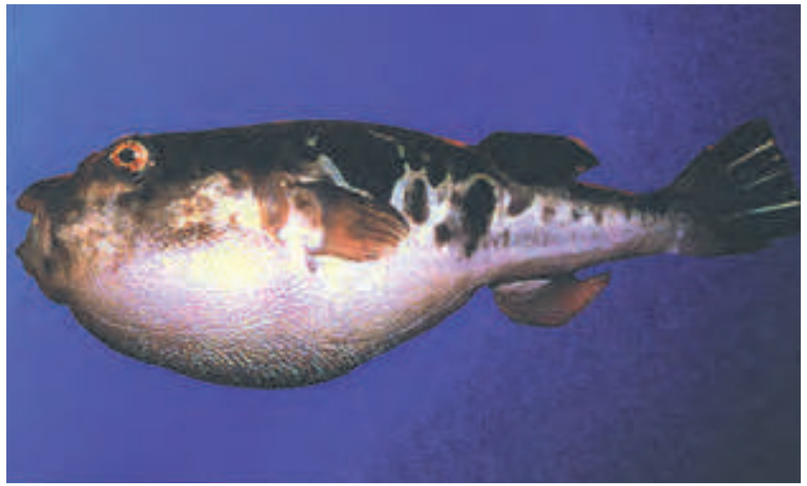
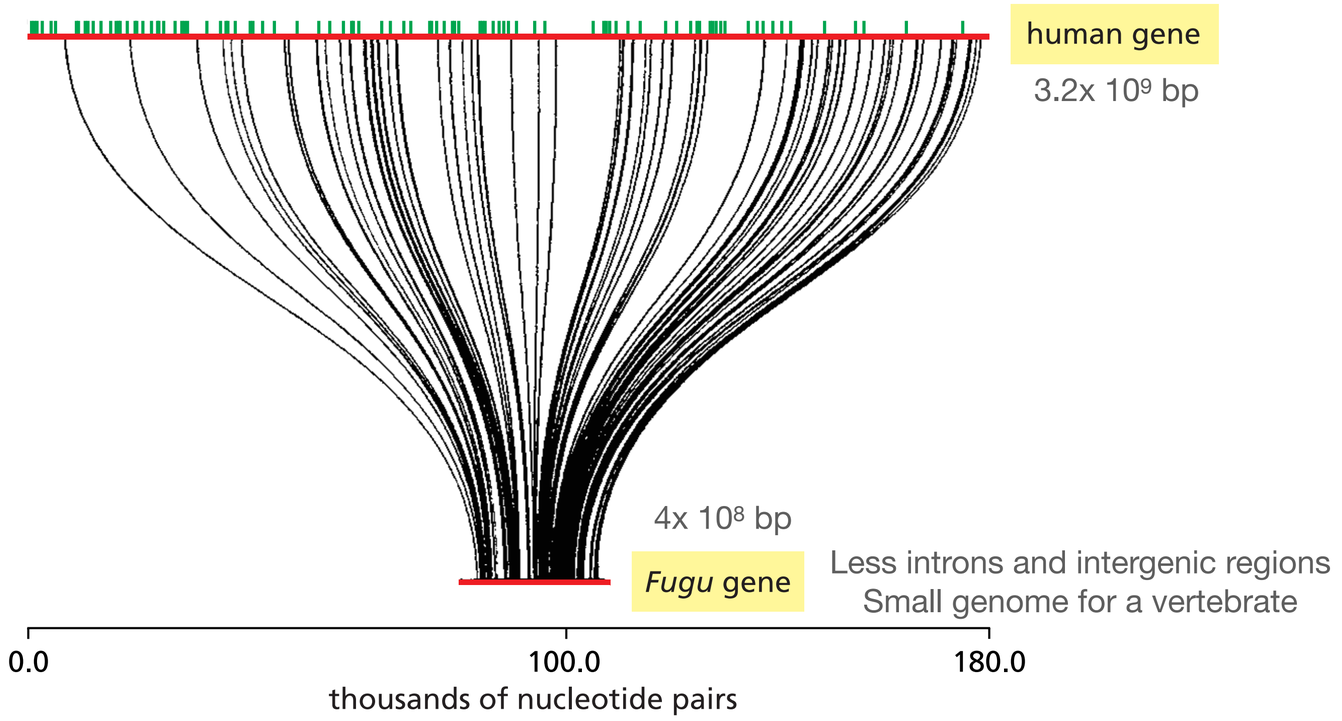
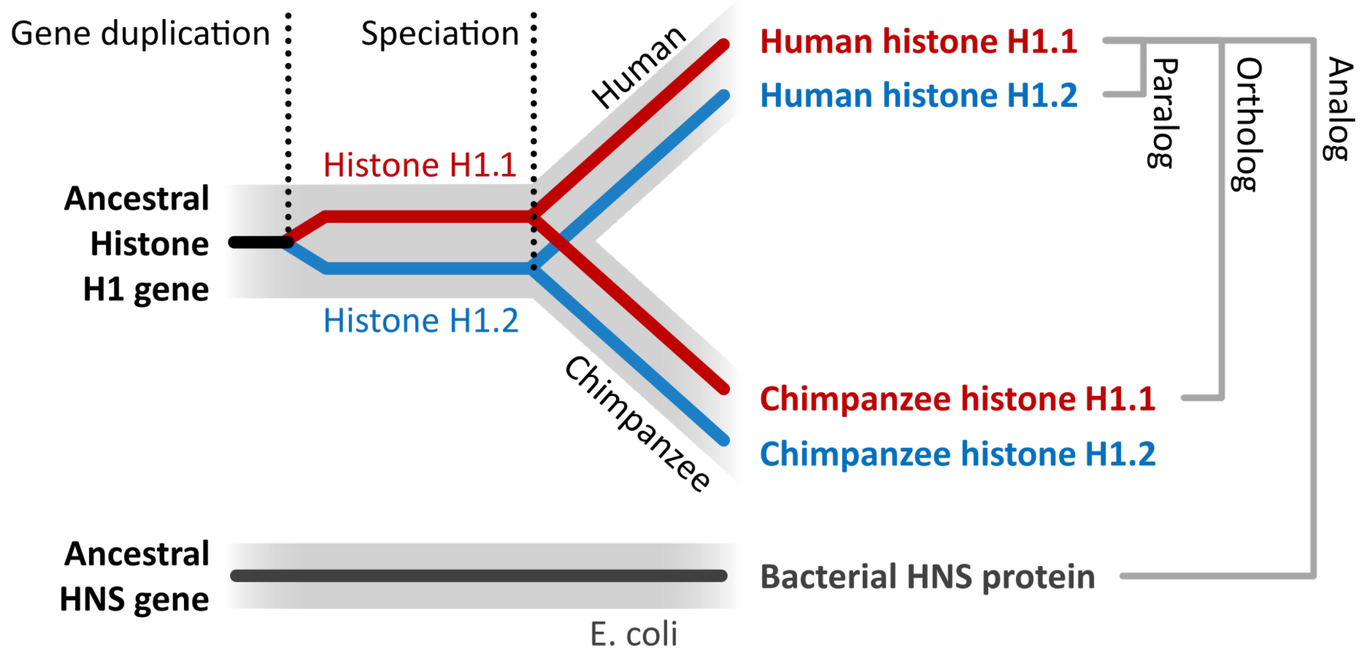


**Figure 4–70 The puffer fish, *Fugu rubripes*.** (Courtesy of Byrappa Venkatesh.)



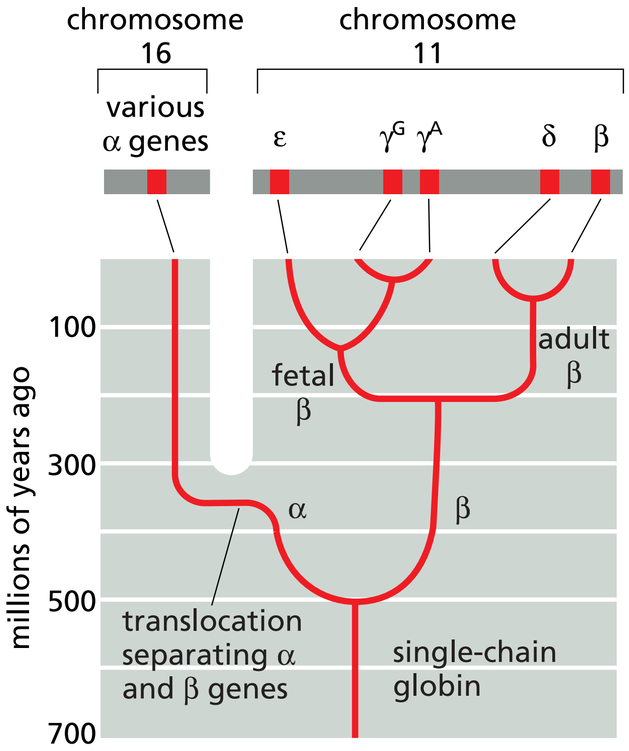
# Conservation of genes across evolution

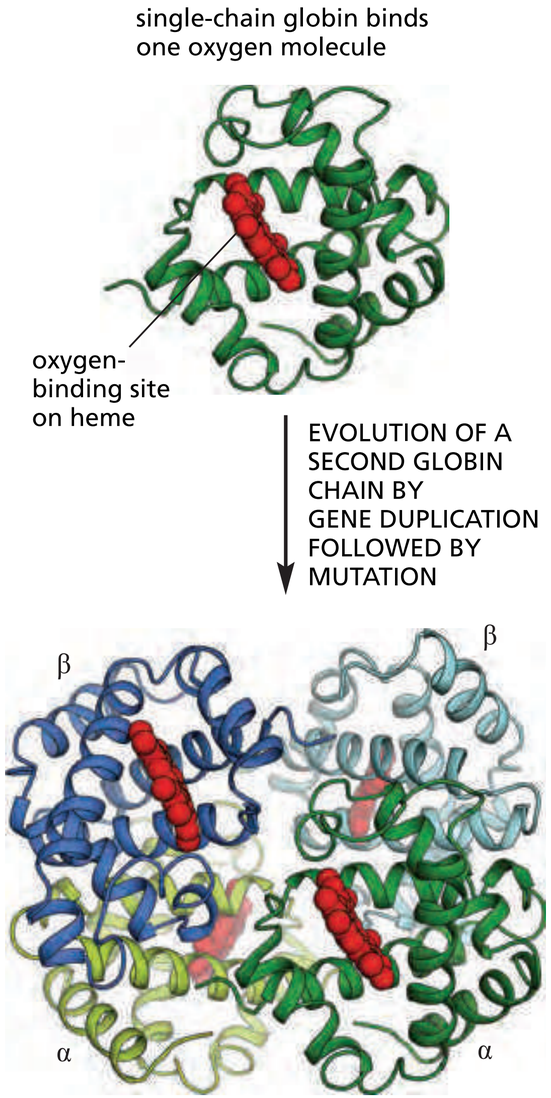
- **Homologous genes** are genetic sequences inherited in two species from a common ancestor. They are similar in sequence and can perform similar functions. **Paralogues** and **orthologues** are homologues.
- **Paralogues** are homologues in the same genome that arose from a gene duplication event.
- **Orthologues** are homologues in different genomes derived from a common ancestor.
- **Analogues** are similar sequences in different genomes without a common ancestor.



# Example of the globin gene family

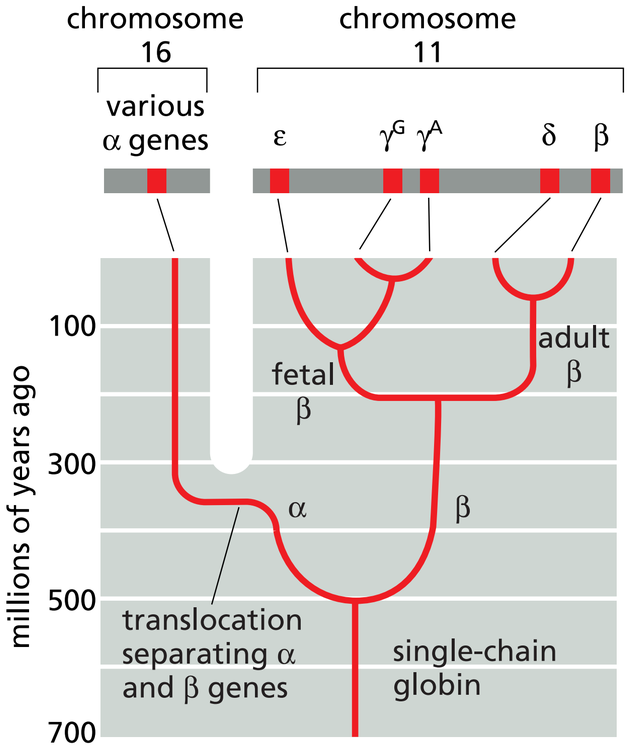
- DNA duplication contributes to the evolution of organisms
- In animals, the most primitive oxygen-carrying molecule is a single-chain globin (in primitive fish, worms, insects)
- 500 million years ago, gene mutations and duplications occurred
- Establishment of two globin genes in the genome of each individual (alpha and beta globin chains which associate to form hemoglobin, necessary for the survival of large animals)



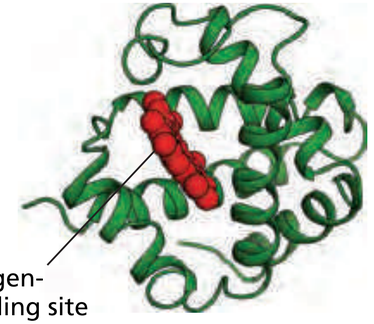



# Example of the globin gene family

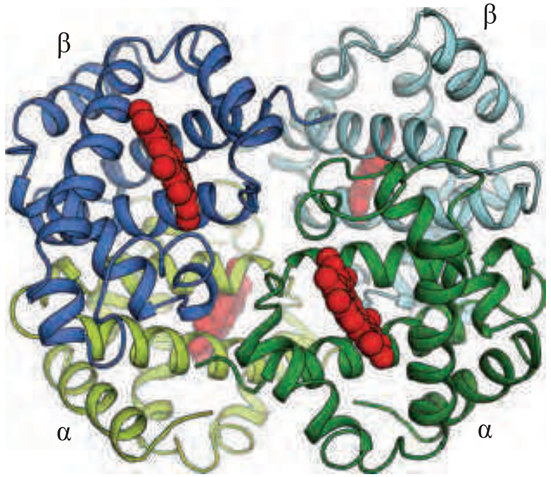
- Second duplication event of the beta-chain gene to give rise to a second one specifically synthesised in the foetus (higher affinity for oxygen for mother-foetus transfer)
- This duplicated and mutated again to produce two new genes (epsilon and gamma), both produced at different stage of development
- Duplication also in adult beta-chain



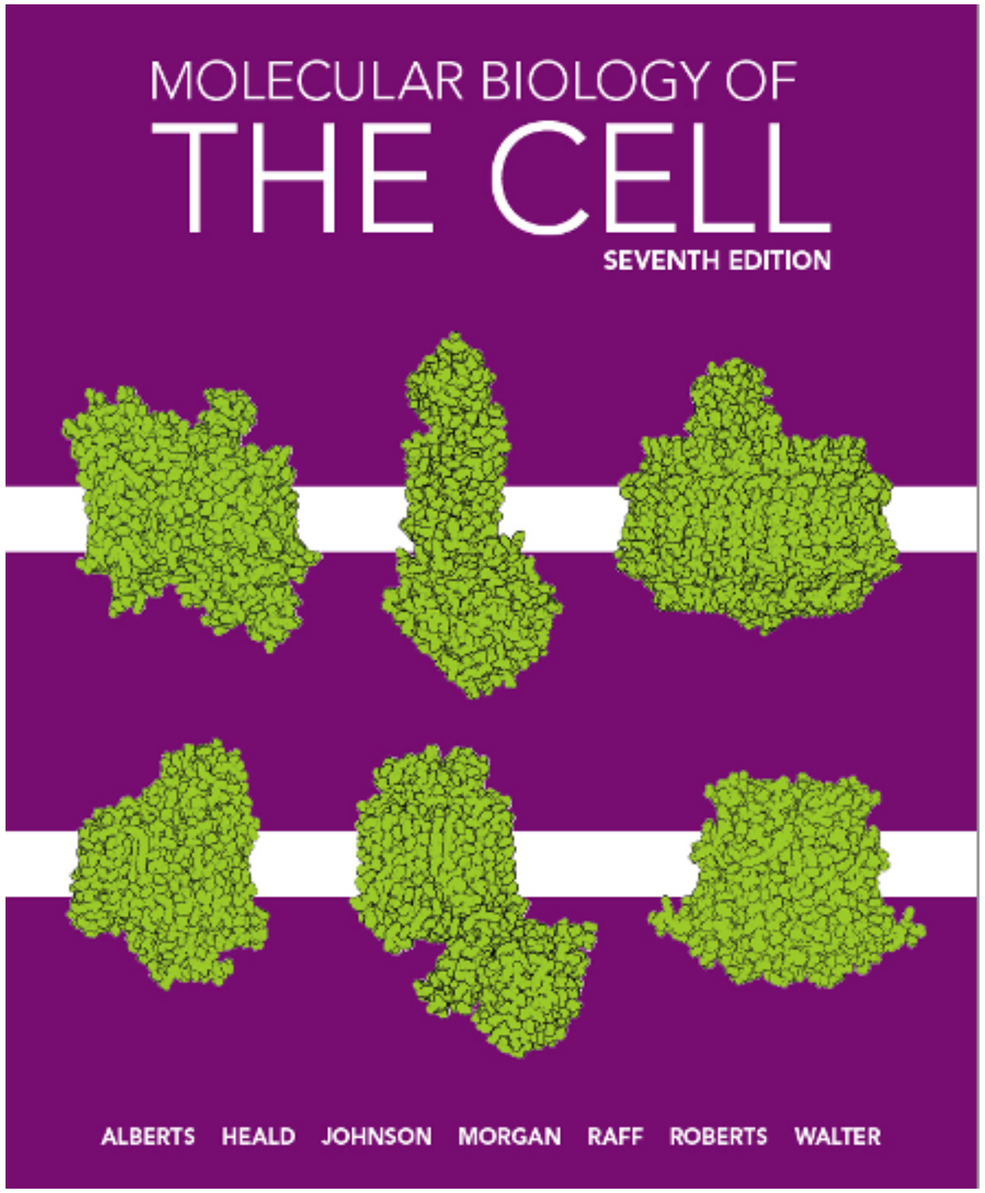

single-chain globin binds one oxygen molecule




EVOLUTION OF A SECOND GLOBIN CHAIN BY GENE DUPLICATION FOLLOWED BY MUTATION




four-chain globin binds four oxygen molecules in a cooperative manner




# Chapter 5

## DNA Replication, Repair, and Recombination

# Plan

- Quick recap
- Mobile genetic elements
- Comparing genomes
- The maintenance of DNA sequences
- DNA replication mechanisms

# Maintaining DNA

**DNA replication**

- Accurate **duplication** of vast quantities of DNA
- Occurs before a cell can produce **two genetically-identical** daughter cells

**DNA repair**

- DNA is **continuously damaged** by chemicals, radiation, thermal accidents or reactive molecules inside the cells
- Protein machineries that **repair** DNA

# Maintaining DNA

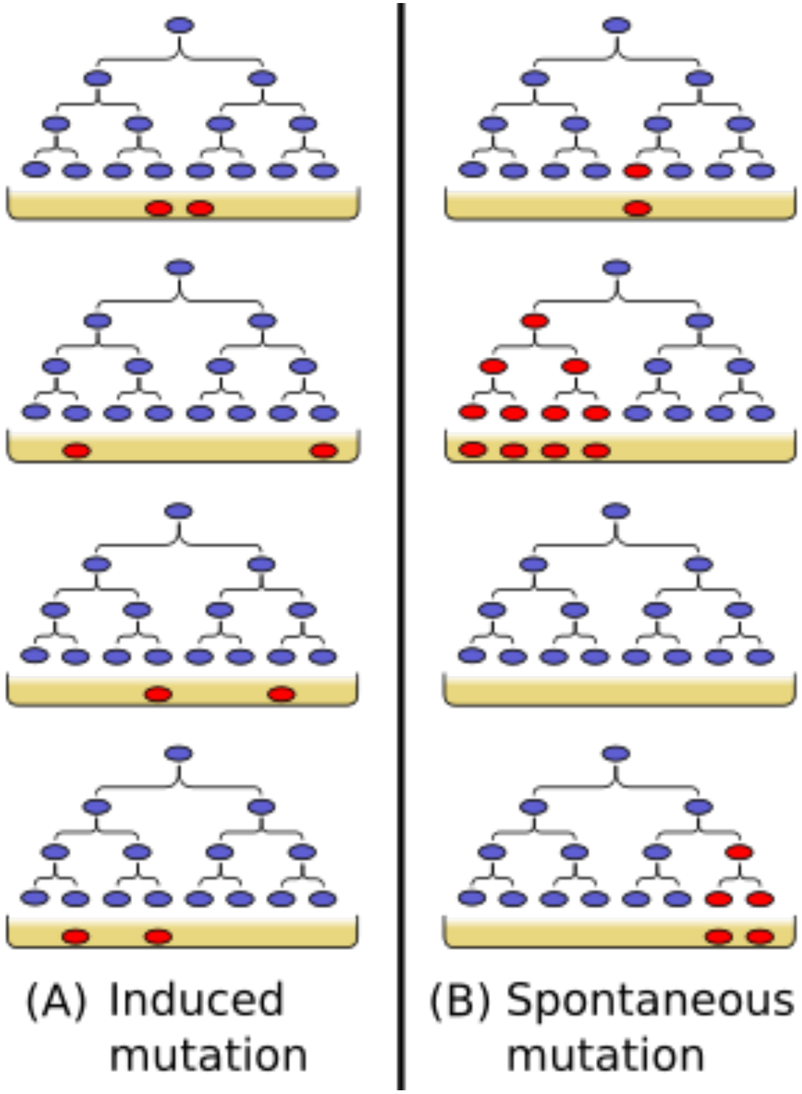
- Occasional genetic changes enhance the **long-term survival** of the species through evolution
- Mutation rates are extremely **low**
- Can be determined with **bacteria** (e.g. *Escherichia coli*) as their doubling time in the lab is ~ 20 minutes
- Bacterial mutation rate is about **3 nucleotides per $10^{10}$ nucleotides** per cell generation (20 minutes) - *E. coli* genome is ~ 5 Mb (megabases, so $5x10^6$ bases)
- Some mutations are **silent**
- In human, the estimate is **1 nucleotides per $10^{10}$ nucleotides** per cell division (not human generation)

# Fluctuation test

➡ How do we know if mutations are **induced by a given condition** or if they are **random** and **later selected** by the given condition?

Luria and Delbrück (1943)

# Fluctuation test



# Maintaining DNA

- Sexually-reproducing animals or plants have two types of cells: **germ cells and somatic cells**
- Both need to **protect their DNA**: germ-cells to maintain the species and somatic cells, to maintain the structure of the body
- Uncontrolled mutant proliferation in somatic cells = **cancer**

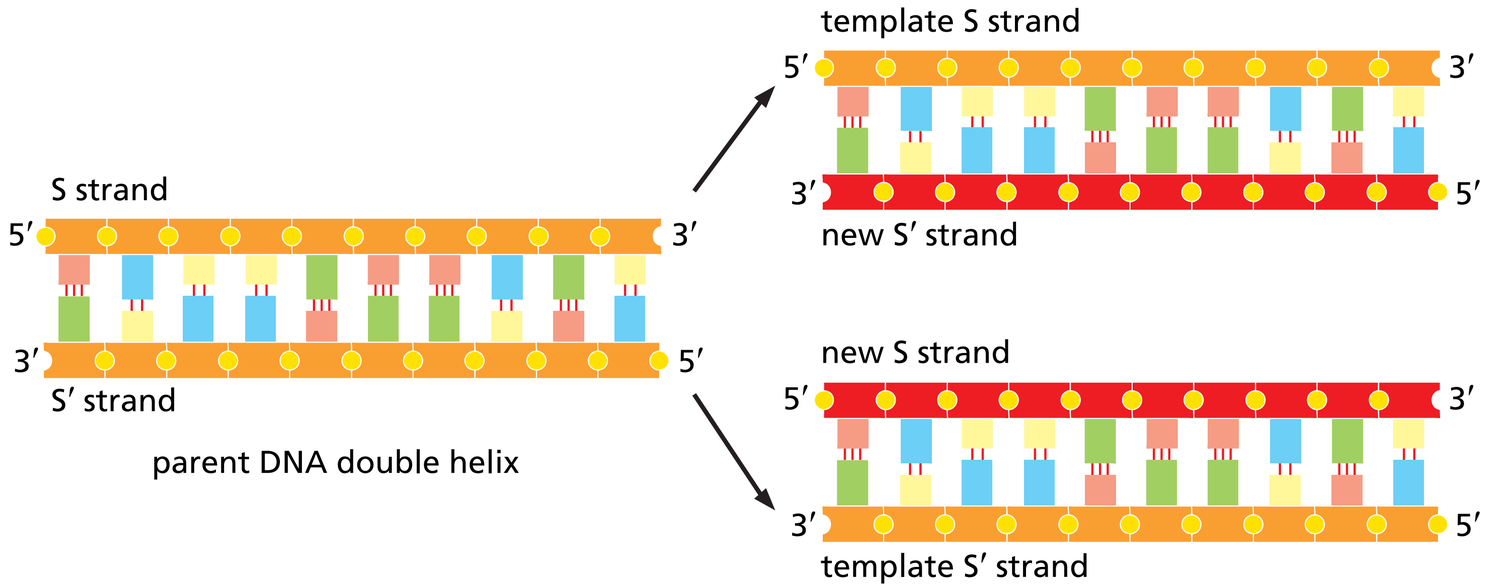
**Germ-cells** transmit genetic information from parent to offspring while **somatic cells** form the body of the organism

# Plan

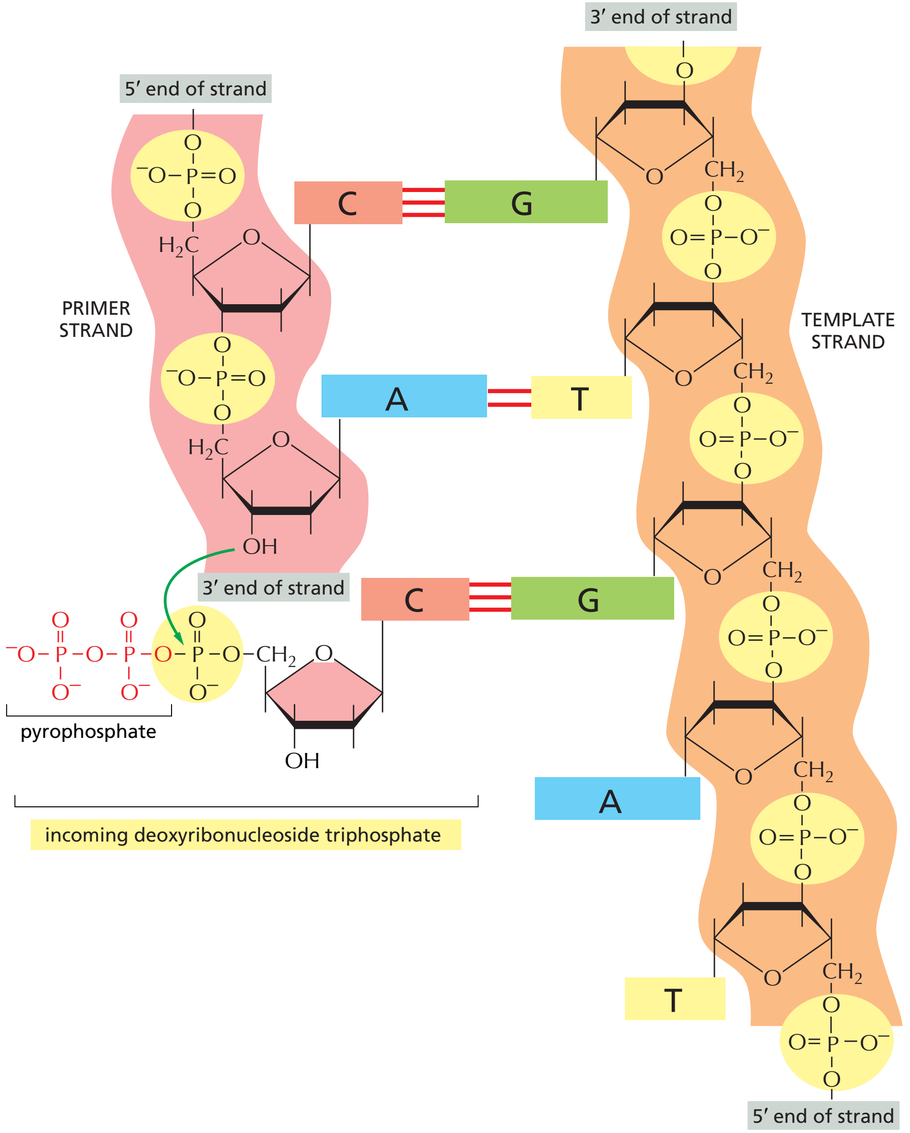
- Quick recap
- Mobile genetic elements
- Comparing genomes
- The maintenance of DNA sequences
- DNA replication mechanisms

# DNA replication principles

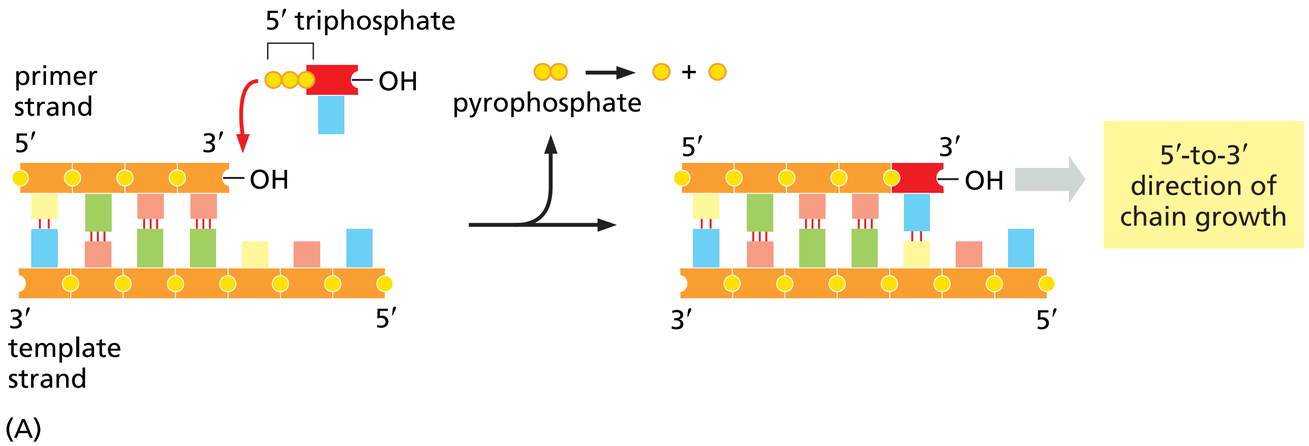
- **Separation** of the DNA helix into two strands
- **Recognition** of each template nucleotide by a free complementary nucleotide (deoxyribonucleotide triphosphate)
- Polymerisation of the nucleotide by the **DNA polymerase**
- Most mechanisms uncovered in **bacteria** and viruses



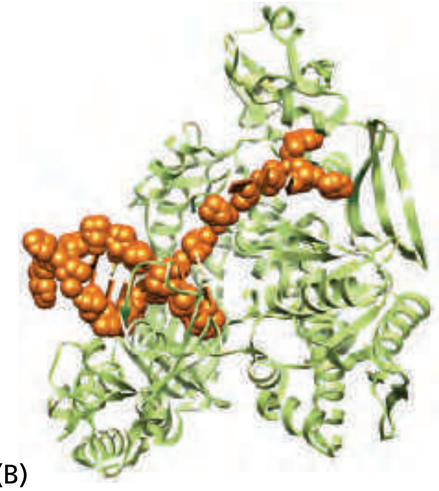
# DNA replication mechanism



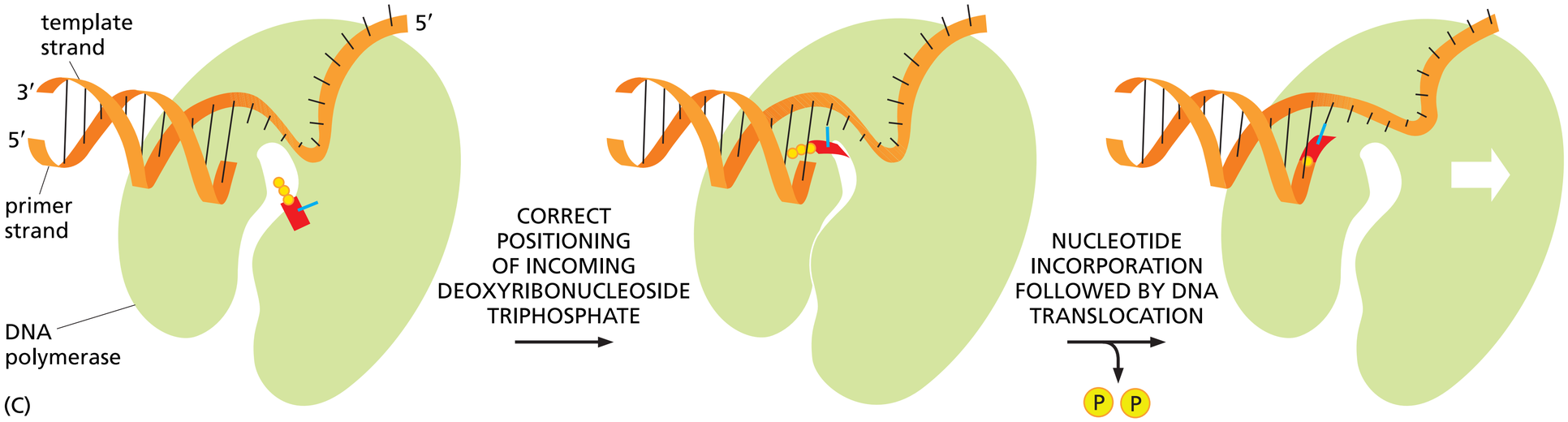
# DNA replication mechanism




(A)



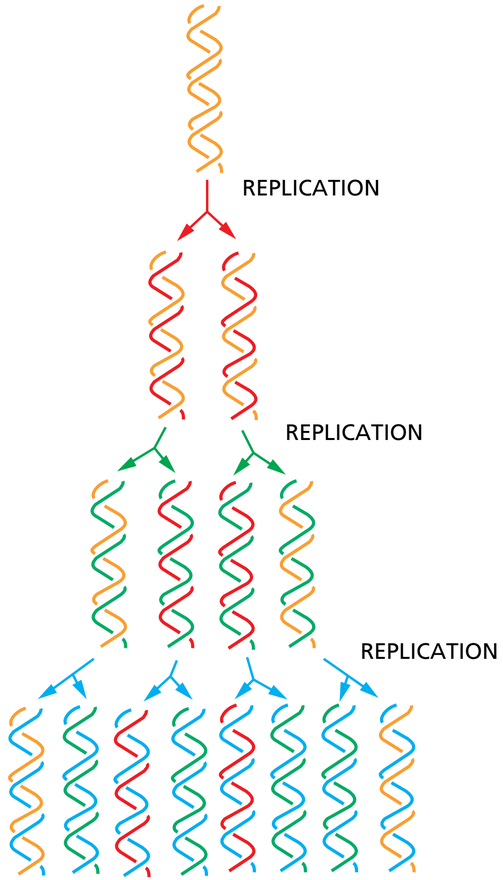
(B)




(C)

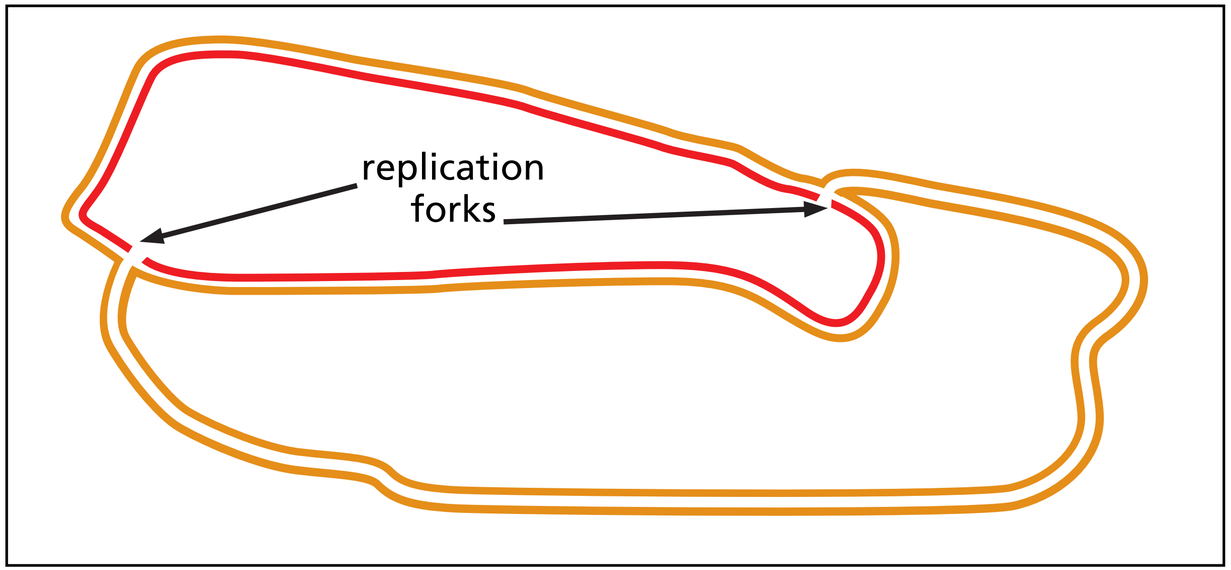
# DNA replication principles

- DNA replication is **semiconservative** as the two daughter cells will inherit a double helix that contains one “original” and one “new” strand



# DNA replication principles

- Localised region of replication that moves along the DNA = **replication fork**
- At the replication fork, **a multienzyme complex** (with DNA polymerase) synthesises the DNA of both new daughter strands

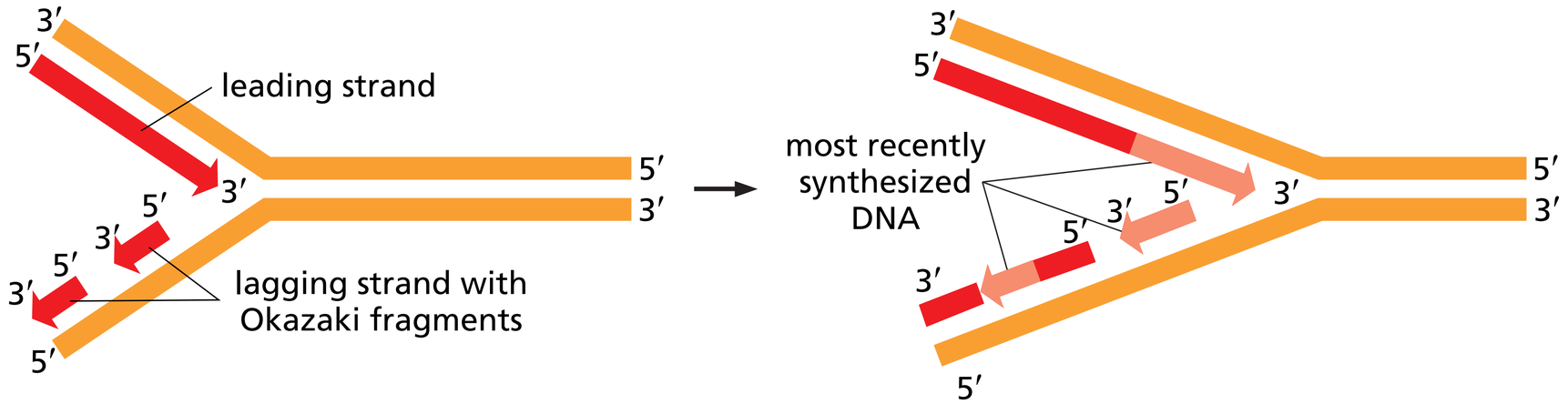


# DNA replication principles

**DNA polymerase** can only polymerise **from 5’ to 3’ —>** How does it work for the **other strand**?

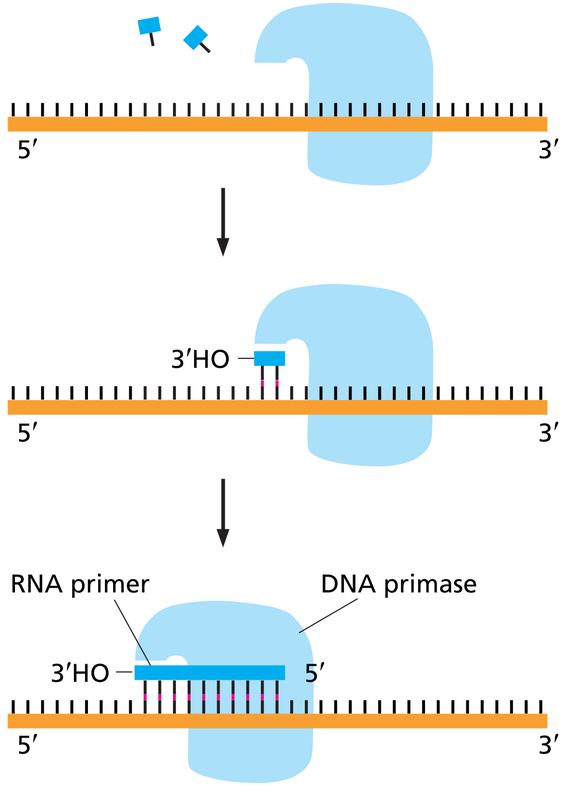
# DNA replication principles

- Transient existence of pieces of DNA that are ~ 100-1000 nucleotide long at the replication fork = **okazaki fragments**
- They are synthesised **from 5' to 3'** and **joined together after synthesis**
- The replication fork has therefore an **asymmetric structure** with a **leading strand** (synthesised continuously) and a **lagging strand** (synthesised non-continuously)



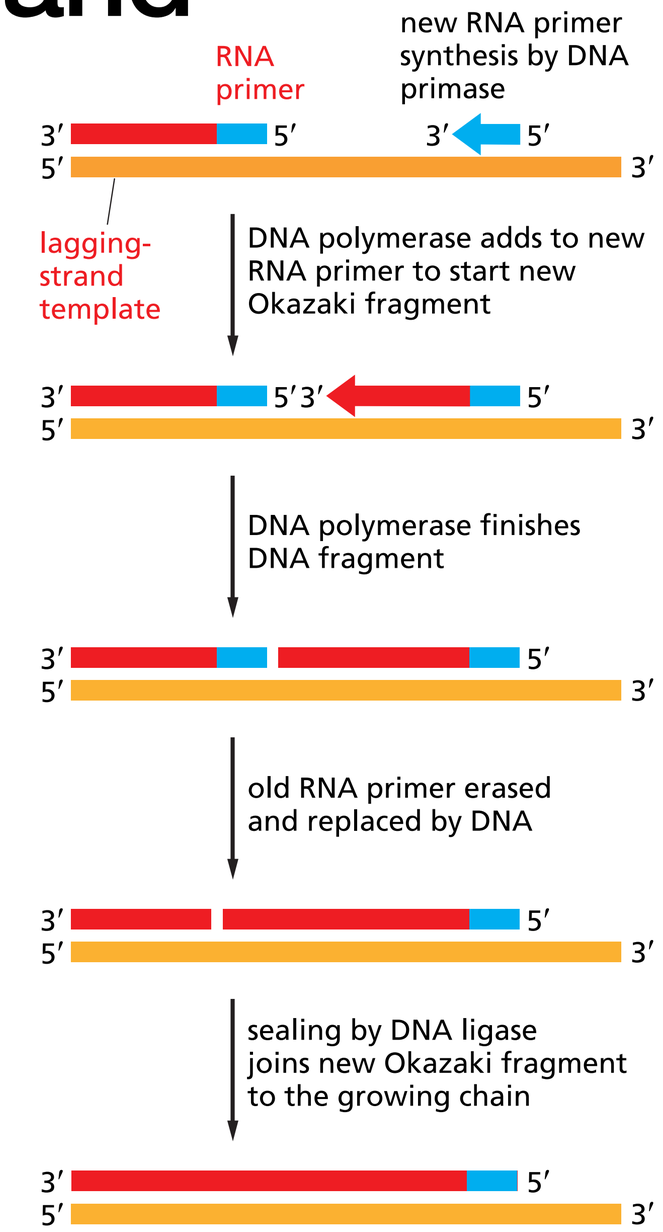
# Building an RNA primer

- DNA polymerase **cannot start synthesis *de novo* without a primer**
- For the leading strand, a primer is needed at the start of replication
- For the lagging strand, a primer is needed at each new Okazaki fragment
- This mechanism depends on the **DNA primase** which uses ribonucleotide triphosphate to synthesise short **RNA primers** on the lagging strand
- In eukaryotes, those are about **10 nucleotide long** and are made at intervals of 100-200 nucleotides on the lagging strand



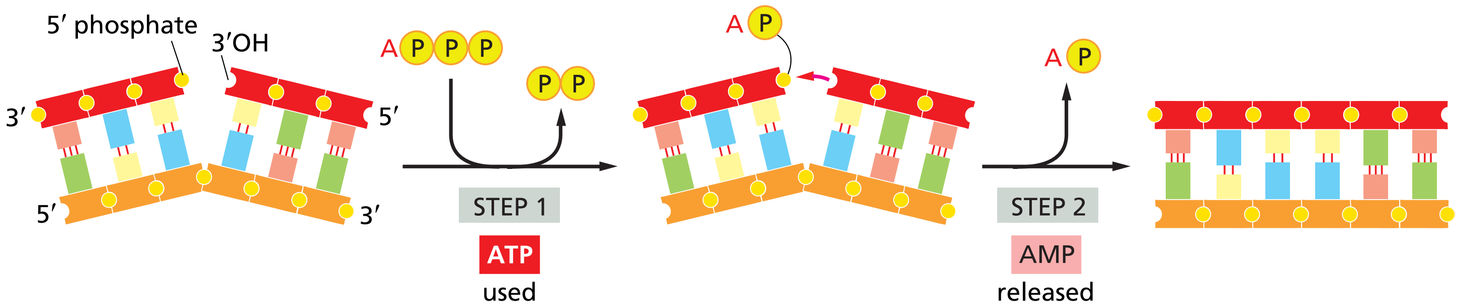
# DNA synthesis on the lagging strand

- On the lagging strand, DNA synthesis **stops when the DNA polymerase runs into the next RNA primer**
- A special **DNA repair system** acts quickly to **erase the RNA primer** and **replace it with DNA**
- A **DNA ligase** then **joins** the 3’ end of the new fragment to the 5’ end of the old fragment

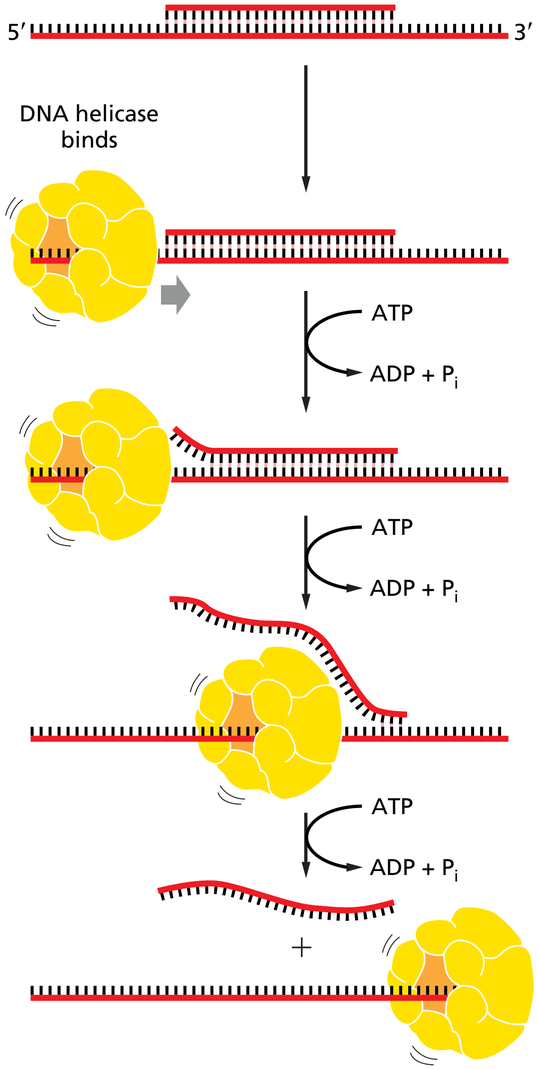


# DNA synthesis on the lagging strand

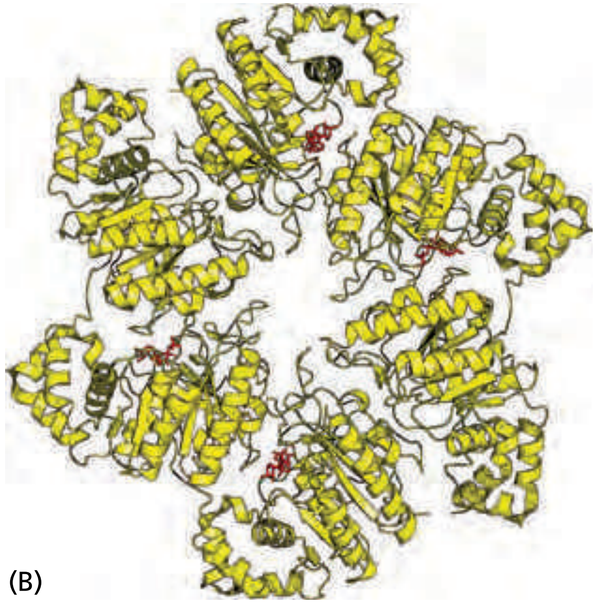
- On the lagging strand, DNA synthesis **stops when the DNA polymerase runs into the next RNA primer**
- A special **DNA repair system** acts quickly to **erase the RNA primer** and **replace it with DNA**
- A **DNA ligase** then **joins** the 3’ end of the new fragment to the 5’ end of the old fragment



# Opening up the double helix

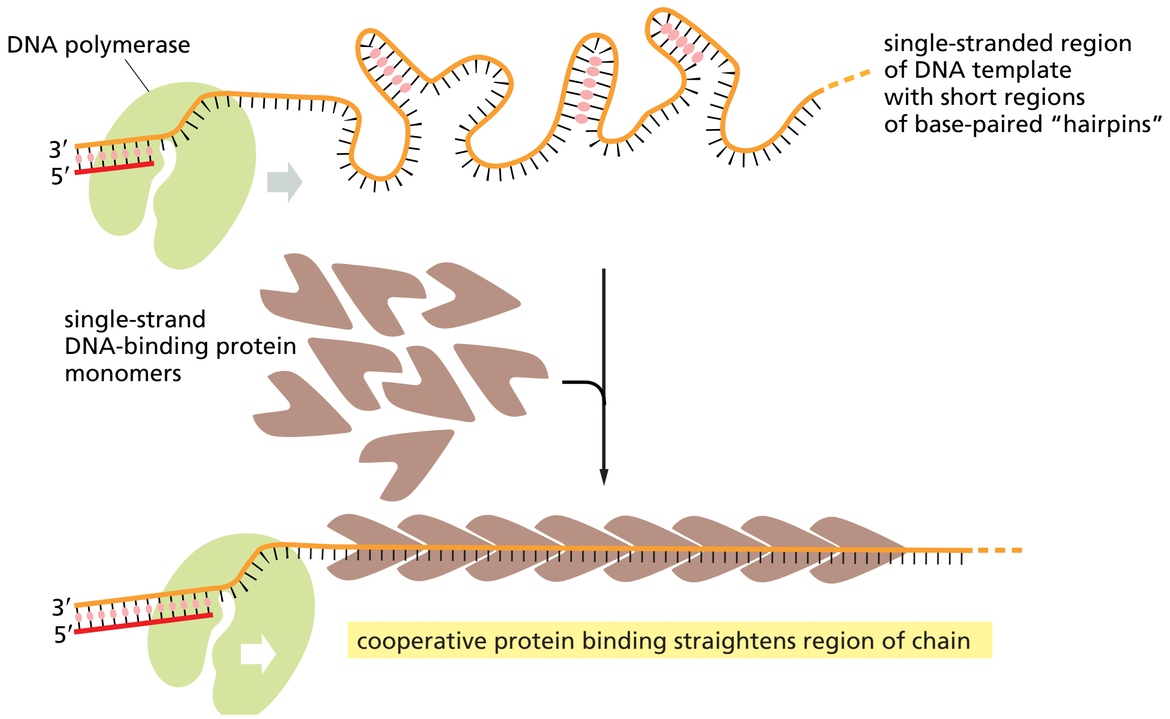



- The double helix should be **opened** ahead of the replication fork
- DNA is very stable, in the lab, we use **very high temperatures** to separate two DNA strands
- **DNA helicases** hydrolyze ATP when they are bound to single strands of DNA, they move along the strand and open the helix once they encounter double-stranded DNA (1000 nt/sec)

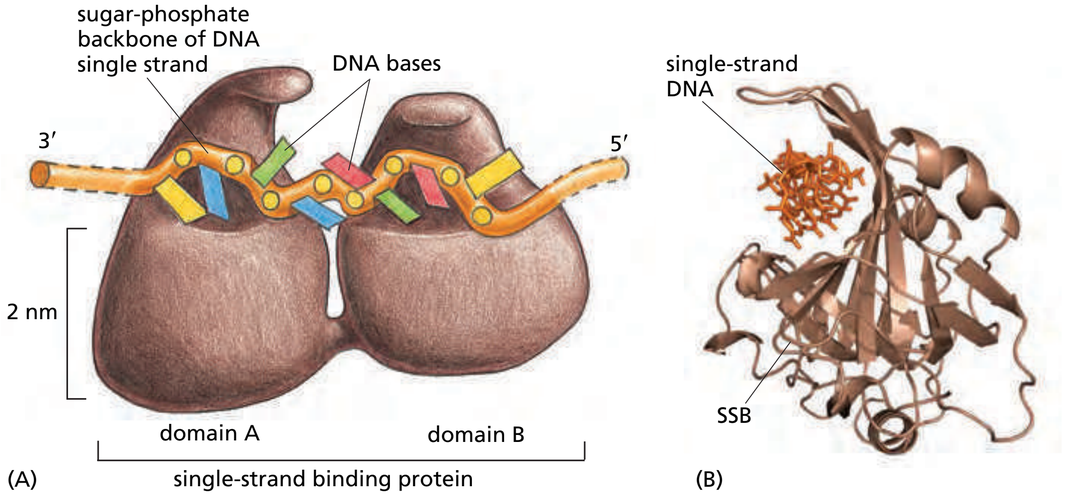


(B)

# Opening up the double helix

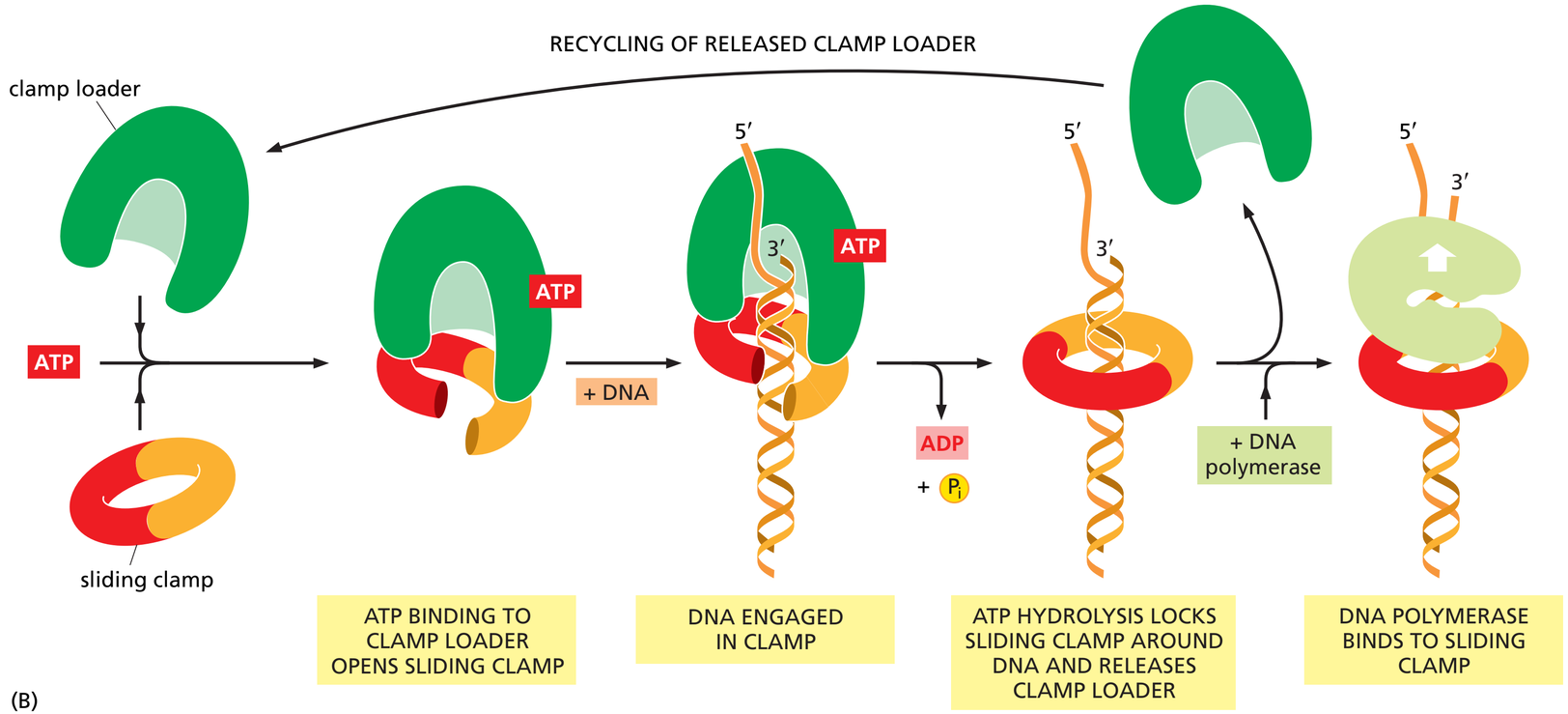



- **Single-strand DNA binding proteins** (SSB) bind to single-stranded DNA without covering the bases (available as templates)
- They **help** helices by **stabilising and straightening** the DNA
- They are **unable to open** a long DNA helix



# Holding the DNA polymerase onto the DNA

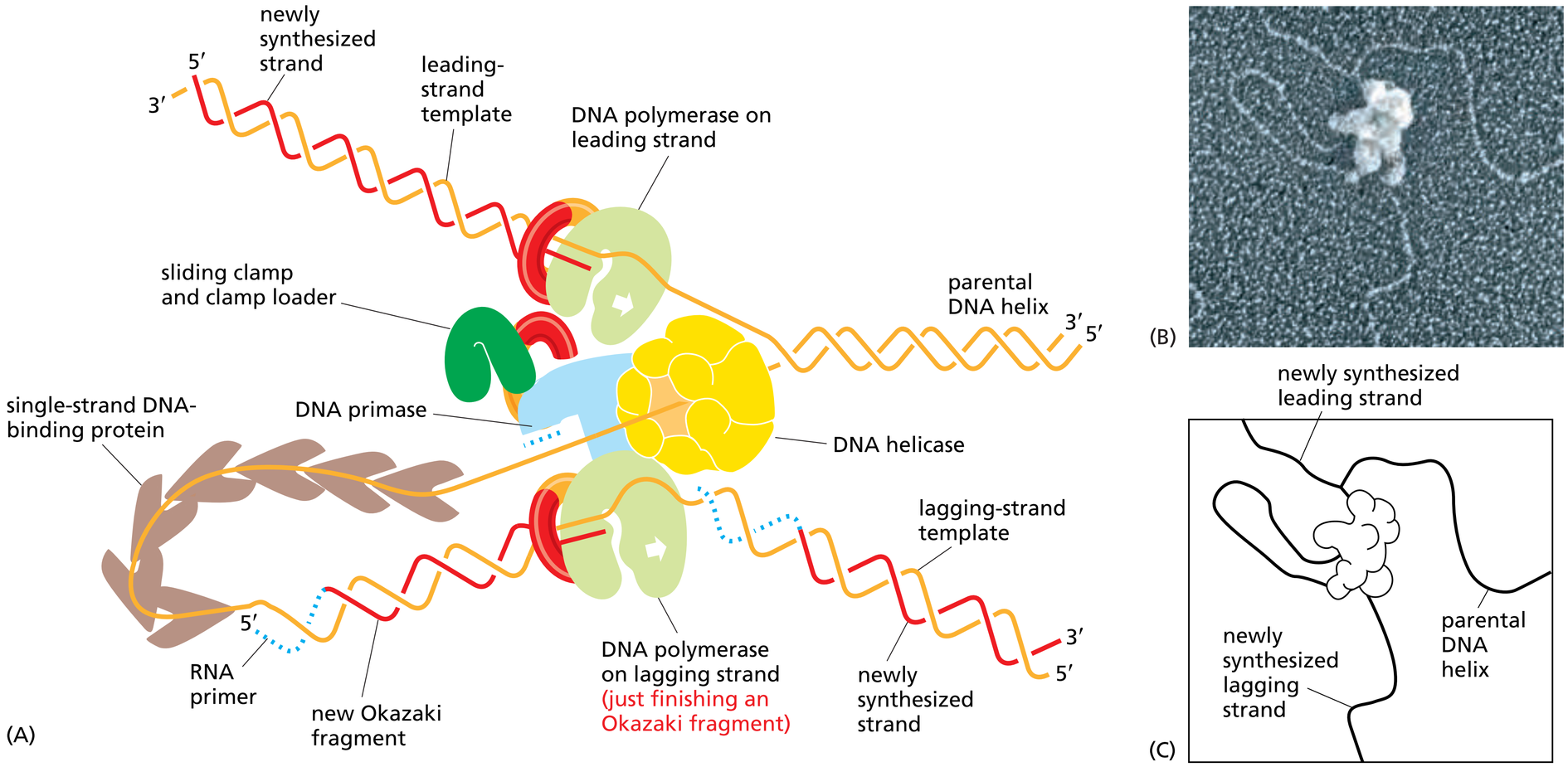
- DNA polymerases need to be able to **“fall off” the DNA** to synthesise the lagging strand
- On the leading strand, it needs a **sliding clamp** (proteins) and **clamp loader**
- The clamp **keeps the polymerase onto the DNA** until running into double-strand regions




(B)

# The replication machine

- Most of the proteins discussed before act together as a large **multi-enzyme complex** that rapidly synthesizes DNA




Bacterial replication fork

# Have a nice day!