Genetics [part 2]

IF3211 Domain Specific Computation

School of Electrical Engineering and Informatics ITB

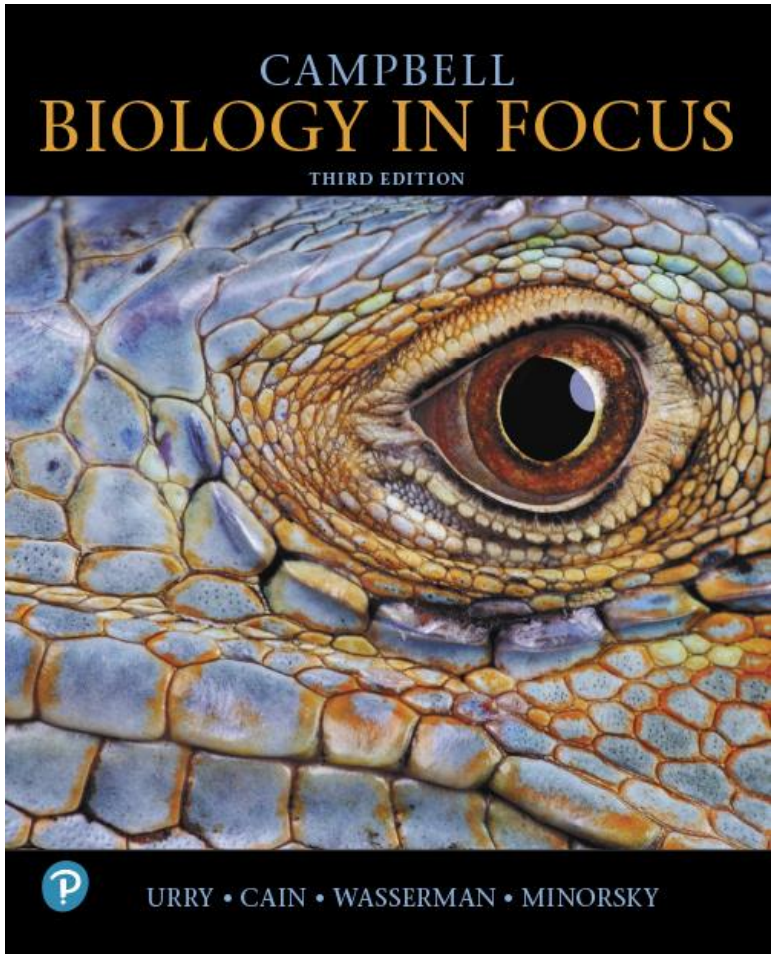
Content

- Molecular Basis of Inheritance
- Gene Expression: From Gene to Protein
- Genome and Their Evolution
- Tools

The Molecular Basis of Inheritance

Campbell Biology in Focus

Third Edition



Chapter 13

The Molecular Basis of Inheritance

Lecture Presentations by
Kathleen Fitzpatrick and Nicole Tunbridge,
Simon Fraser University

Pearson

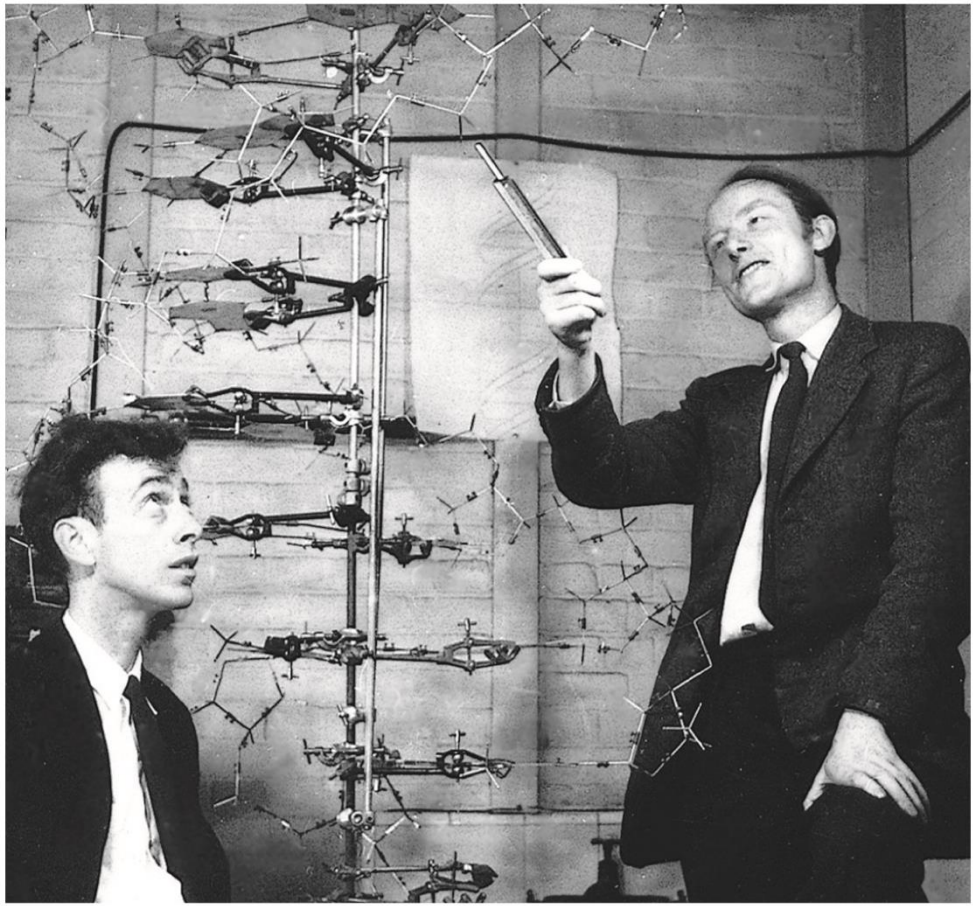
Copyright © 2022, 2020, 2016 Pearson Education, Inc. All Rights Reserved

Overview: Life's Operating Instructions

- In 1953, James Watson and Francis Crick introduced an elegant double-helical model for the structure of deoxyribonucleic acid, or DNA
- Hereditary information in DNA directs the development of your biochemical, anatomical, physiological, and to some extent behavioral traits
- Hereditary information is reproduced in all cells of the body during **DNA replication**

Figure 13.2

James Watson (Left) and Francis Crick with Their DNA Model



Concept 13.1: DNA Is the Genetic Material

- Early in the 20th century, the identification of the molecules of inheritance loomed as a major challenge to biologists

Evidence That DNA Can Transform Bacteria

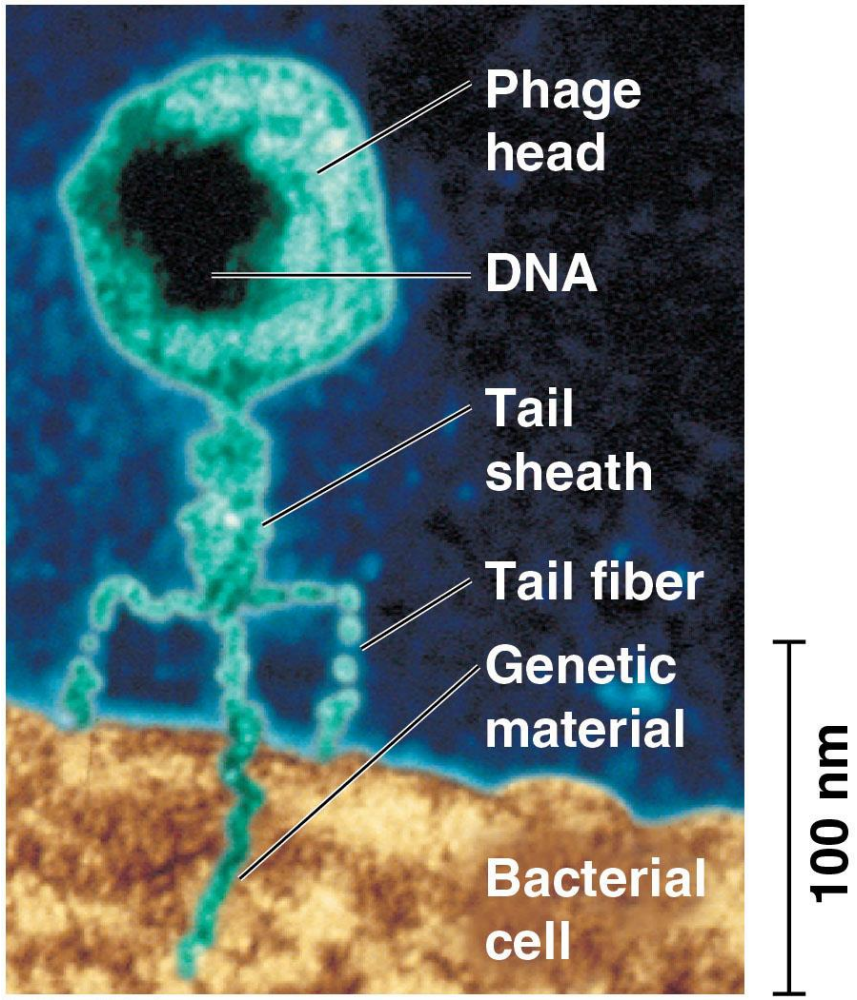
- The discovery of the genetic role of DNA began with research by Frederick Griffith in 1928
- Griffith worked with two strains of a bacterium, one pathogenic and one harmless
- When he mixed heat-killed remains of the pathogenic strain with living cells of the harmless strain, some living cells became pathogenic
- Many biologists remained skeptical, mainly because little was known about DNA

Evidence That Viral DNA Can Program Cells

- More evidence for DNA as the genetic material came from studies of viruses that infect bacteria
- Such viruses, called **bacteriophages** (or **phages**), are widely used in molecular genetics research
- A virus is DNA (or RNA) enclosed by a protective coat, usually made of protein
- Viruses must infect cells and take over the cells' metabolic machinery in order to reproduce

Figure 13.4

A Virus Infecting a Bacterial Cell

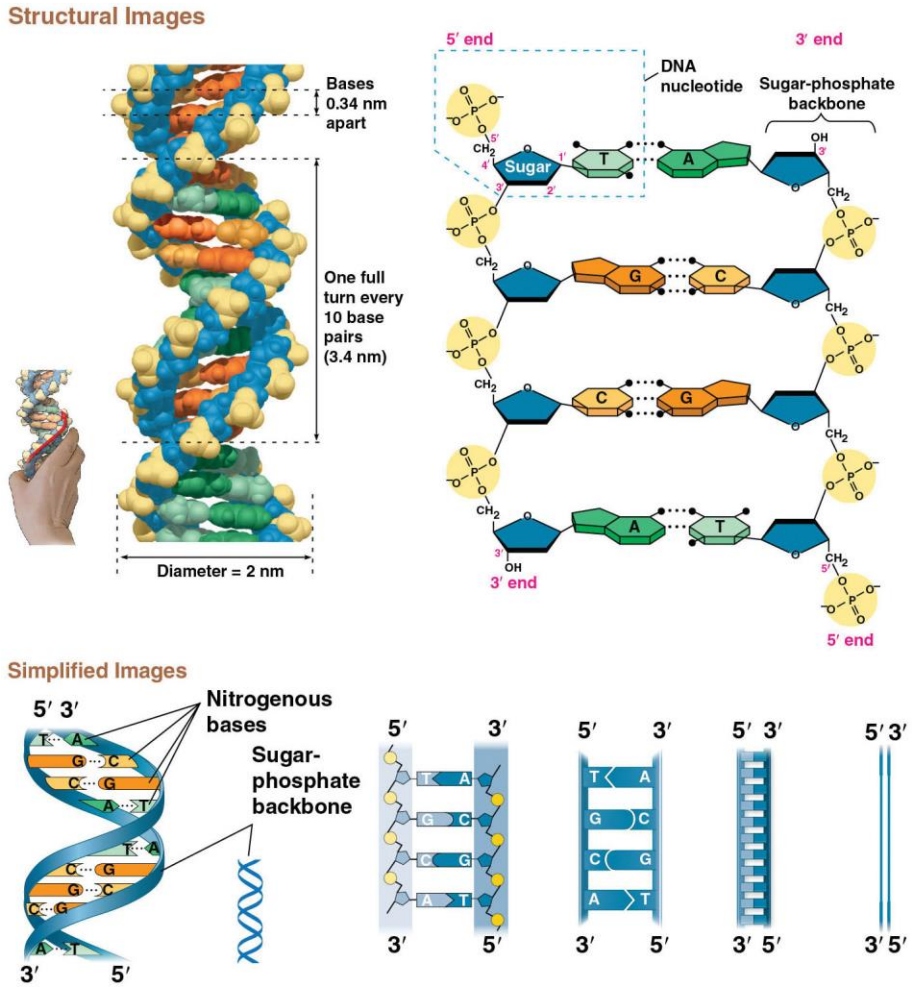


Building a Structural Model of DNA : Scientific Inquiry (1 of 5)

- James Watson and Francis Crick were first to determine the structure of DNA
- Maurice Wilkins and Rosalind Franklin were using a technique called X-ray crystallography to study molecular structure
- Franklin produced a picture of the DNA molecule using this technique

Figure 13.8

Visualizing DNA

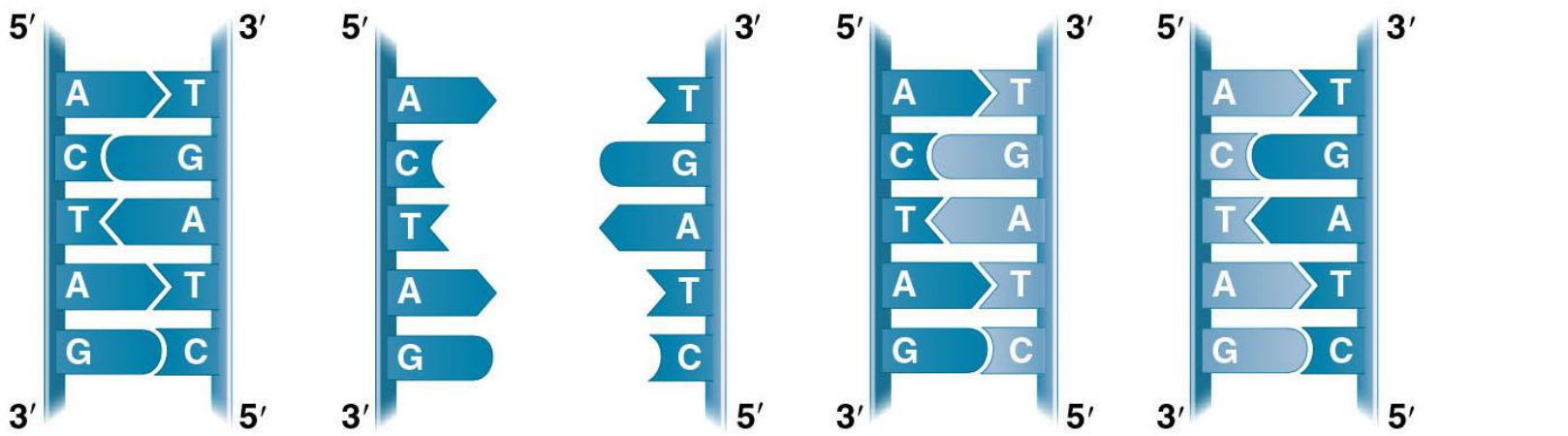


Concept 13.2: Many Proteins Work Together in DNA Replication and Repair

- The relationship between structure and function is apparent in the double helix
- Watson and Crick noted that the specific base pairing suggested a possible copying mechanism for genetic material

Figure 13.11

A Model for DNA Replication: The Basic Concept



(a) Parental molecule

(b) Separation of parental strands into templates

(c) Nucleotides complementary to the parental (dark blue) strand are connected to form the sugar-phosphate backbones of the new "daughter" (light blue) strands.

DNA Replication: A Closer Look

- The copying of DNA is remarkable in its speed and accuracy
- More than a dozen enzymes and other proteins participate in DNA replication
- Much more is known about how this “replication machine” works in bacteria than in eukaryotes
- Most of the process is similar between prokaryotes and eukaryotes

Figure 13.15

Some of the Proteins Involved in the Initiation of DNA Replication

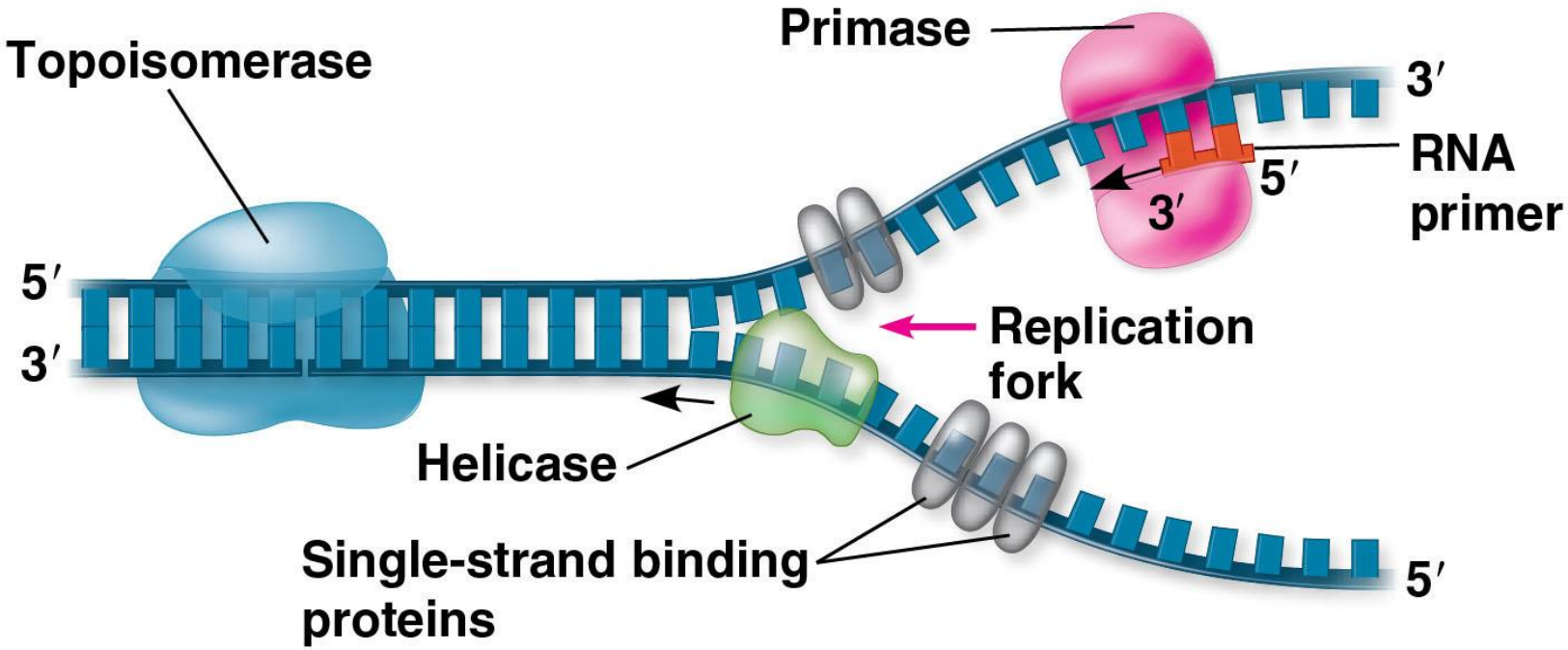
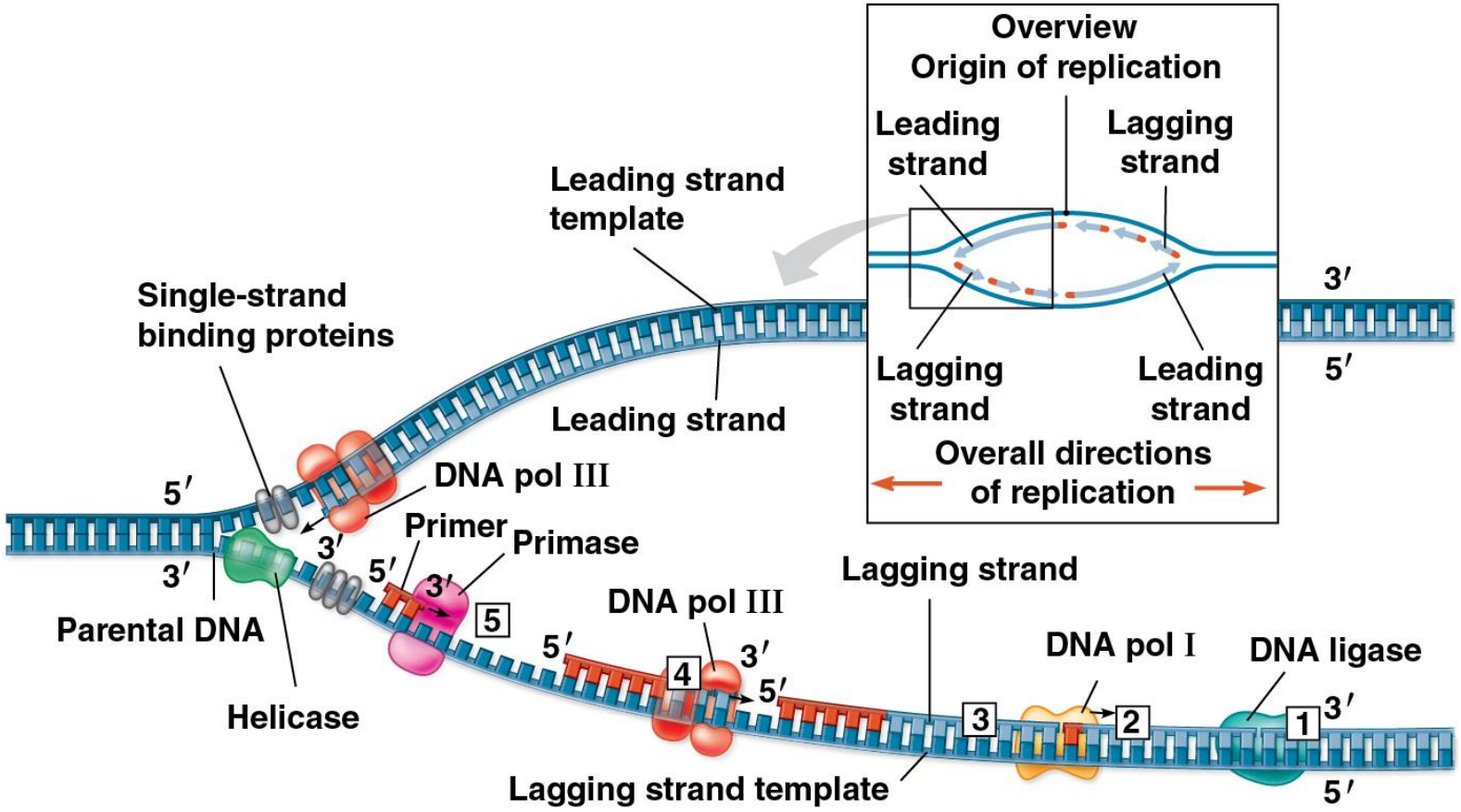


Figure 13.19

A Summary of Bacterial DNA Replication



Proofreading and Repairing DNA (1 of 2)

- Errors in the completed DNA molecule amount to only one in 10 billion
- DNA polymerases proofread newly made DNA, replacing any incorrect nucleotides
- In **mismatch repair** of DNA, other enzymes correct errors in base pairing
- A hereditary defect in one such enzyme is associated with a form of colon cancer
- This defect allows cancer-causing errors to accumulate in DNA faster than normal

Proofreading and Repairing DNA (2 of 2)

- DNA can be damaged by exposure to harmful chemical or physical agents, such as X-rays
- DNA bases can also undergo spontaneous changes
- In many cases a **nuclease** cuts out and replaces damaged stretches of DNA
- One such DNA repair system is called **nucleotide excision repair**
- DNA repair enzymes in our skin repair genetic damage caused by the ultraviolet light of sunlight

Evolutionary Significance of Altered DNA Nucleotides

- The error rate after proofreading repair is extremely low but not zero
- Sequence changes may become permanent and can be passed on to the next generation
- These changes (mutations) are the source of the genetic variation upon which natural selection operates

Concept 13.4: Understanding DNA Structure and Replication Makes Genetic Engineering Possible

- Complementary base pairing of DNA is the basis for **nucleic acid hybridization**, the base pairing of one strand of a nucleic acid to another, complementary sequence
- Nucleic acid hybridization forms the foundation of virtually every technique used in **genetic engineering**, the direct manipulation of genes for practical purposes

Editing Genes and Genomes (1 of 4)

- Over the past 10 years, biologists have developed a powerful new technique called the **CRISPR-Cas9 system**
- Cas9 is a nuclease that cuts double-stranded DNA molecules as directed by a guide RNA that is complementary to the "target" gene
- Researchers have used this system to "knock out" (disable) a given gene in order to determine its function

Editing Genes and Genomes (3 of 4)

- In another application scientists are addressing the global problem of insect-borne diseases
- Altering genes in the insect may prevent it from transmitting the disease
- An extra facet of this work is engineering the new allele to favour its inheritance over the wild type allele
- This is called **gene drive**, because this biased inheritance of the engineered allele rapidly "drives" the allele through the population

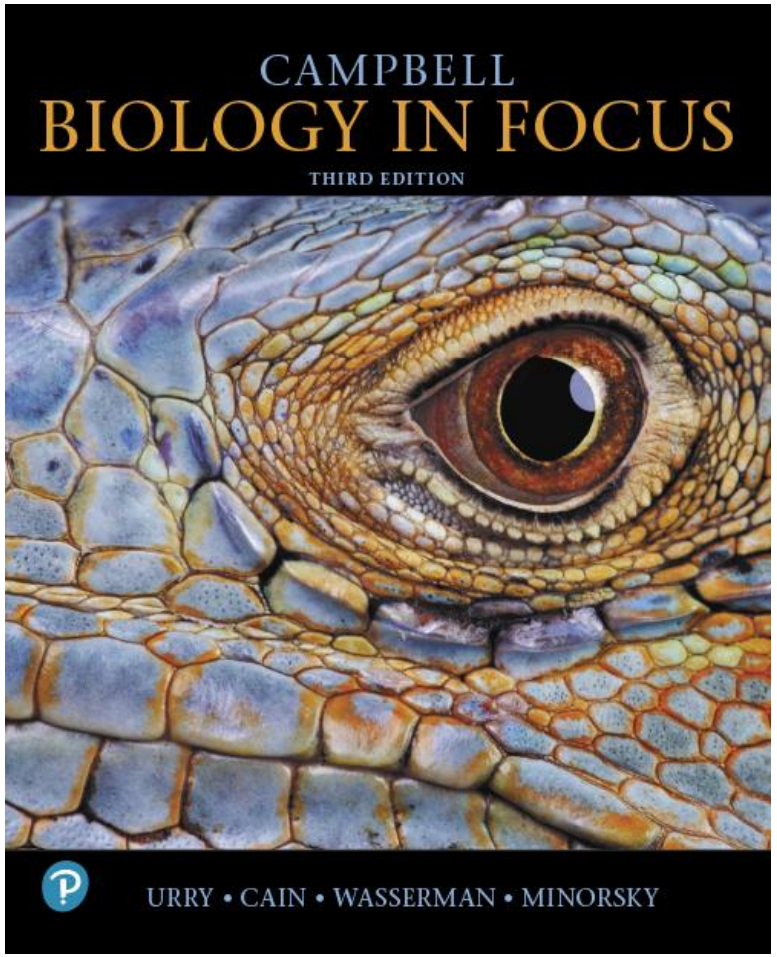
Editing Genes and Genomes (4 of 4)

- Researchers, physicians, and patients alike are excited about the potential of CRISPR technology
- However, Jennifer Doudna, a co-discoverer of the technology recognizes the danger of its misapplication
- She realized the importance of reflecting on the ethical considerations of this technology

Gene Expression:
From Gene to Protein

Campbell Biology in Focus

Third Edition



Chapter 14

Gene Expression: From Gene to Protein

Lecture Presentations by
Kathleen Fitzpatrick and Nicole Tunbridge,
Simon Fraser University

Pearson

Copyright © 2022, 2020, 2016 Pearson Education, Inc. All Rights Reserved

Starter Questions...

- What is the main role of DNA?
- What is a gene?
- What does a gene code for?

This whole process is called ‘protein synthesis’

- Making proteins in a cell – what are proteins and what do we need them for?
- This process is divided into two stages: **Transcription** and **Translation**

1st step: Transcription

- The cell wants to make a protein from a specific section of DNA (gene)
- In nucleus, DNA is unwound so that part of DNA is exposed
- RNA polymerase is an enzyme that helps the RNA nucleotides to build on the DNA to make a piece of mRNA
- This is a ‘transcript’ of the DNA that can now leave the cell to find a ribosome

2nd Step: Translation

- mRNA finds a ribosome and ‘feeds’ through it
- mRNA is read in three nucleotides (codons). When the start codon is identified (by ribosome) it starts the protein building process
- Proteins are made of amino acids. Each codon codes for one amino acid. A chain of amino acids is called a polypeptide (another term for protein)
- Each amino acid is carried on a tRNA (anticodon binds to codon)
- Once it reaches a STOP codon the process stops.

Overview: The Flow of Genetic Information

- The information content of genes is in the form of specific sequences of nucleotides in DNA
- The DNA inherited by an organism leads to specific traits by dictating the synthesis of proteins and of RNA molecules involved in protein synthesis
- Proteins are the link between genotype and phenotype
- **Gene expression**, the process by which DNA directs protein synthesis, includes two stages: transcription and translation

Concept 14.1: Genes Specify Proteins via Transcription and Translation

- How was the fundamental relationship between genes and proteins discovered?

Basic Principles of Transcription and Translation (1 of 5)

- RNA is the bridge between DNA and protein synthesis
- RNA is chemically similar to DNA, but RNA has a ribose sugar instead of deoxyribose and the base uracil (U) rather than thymine (T)
- RNA is usually single-stranded
- Getting from DNA to protein requires two stages: transcription and translation

Basic Principles of Transcription and Translation (2 of 5)

- **Transcription** is the synthesis of RNA using information in DNA
- Transcription produces **messenger RNA (mRNA)**
- **Translation** is the synthesis of a polypeptide, using information in the mRNA
- **Ribosomes** are the sites of translation

Basic Principles of Transcription and Translation (5 of 5)

Evolution of the Genetic Code

- The genetic code is nearly universal, shared by the simplest bacteria and the most complex plants and animals
- Genes can be transcribed and translated after being transplanted from one species to another
- A language shared by all living things must have been operating in the common ancestor of all present-day organisms

Figure 14.7

Evidence for Evolution: Expression of Genes from Different Species

(a) Tobacco plant expressing a firefly gene

(b) Mosquito larva expressing a jellyfish gene

Concept 14.2: Transcription is the DNA-Directed Synthesis of RNA: A Closer Look

- Transcription is the first stage of gene expression

Synthesis of an RNA Transcript

- The three stages of transcription
  - Initiation
  - Elongation
  - Termination

Figure 14.8

The Stages of Transcription: Initiation, Elongation, and Termination

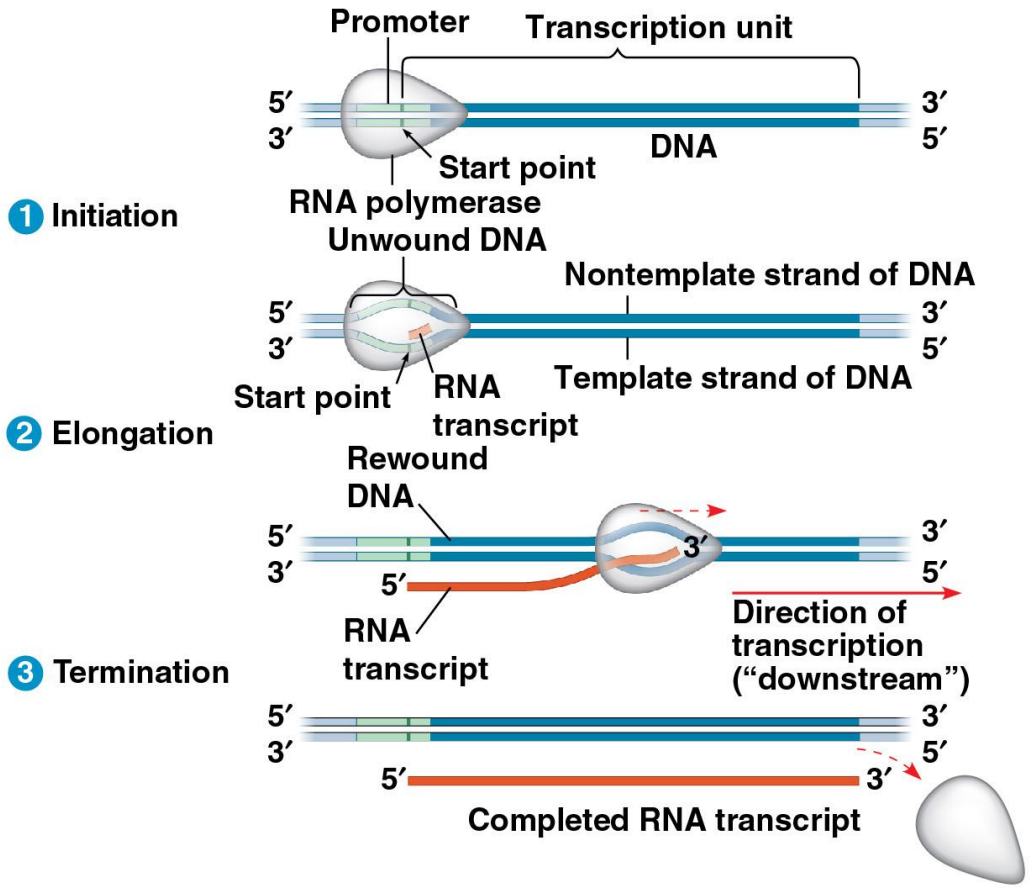
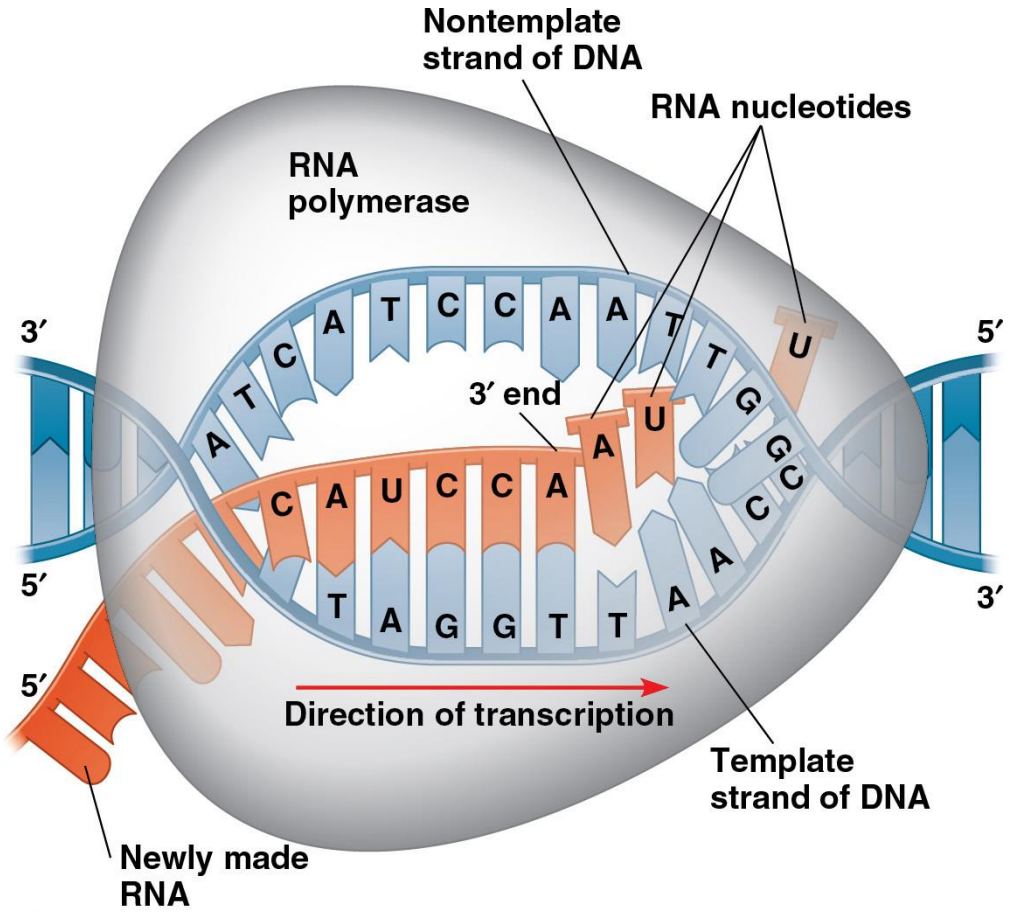


Figure 14.10

Transcription Elongation



Concept 14.4: Translation is the RNA-Directed Synthesis of a Polypeptide: A Closer Look

- Genetic information flows from mRNA to protein through the process of translation

Molecular Components of Translation

- A cell translates an mRNA message into protein with the help of **transfer RNA** (**tRNA**)
- tRNAs transfer amino acids to the growing polypeptide in a ribosome
- Translation is a complex process in terms of its biochemistry and mechanics

Figure 14.14

Translation: The Basic Concept

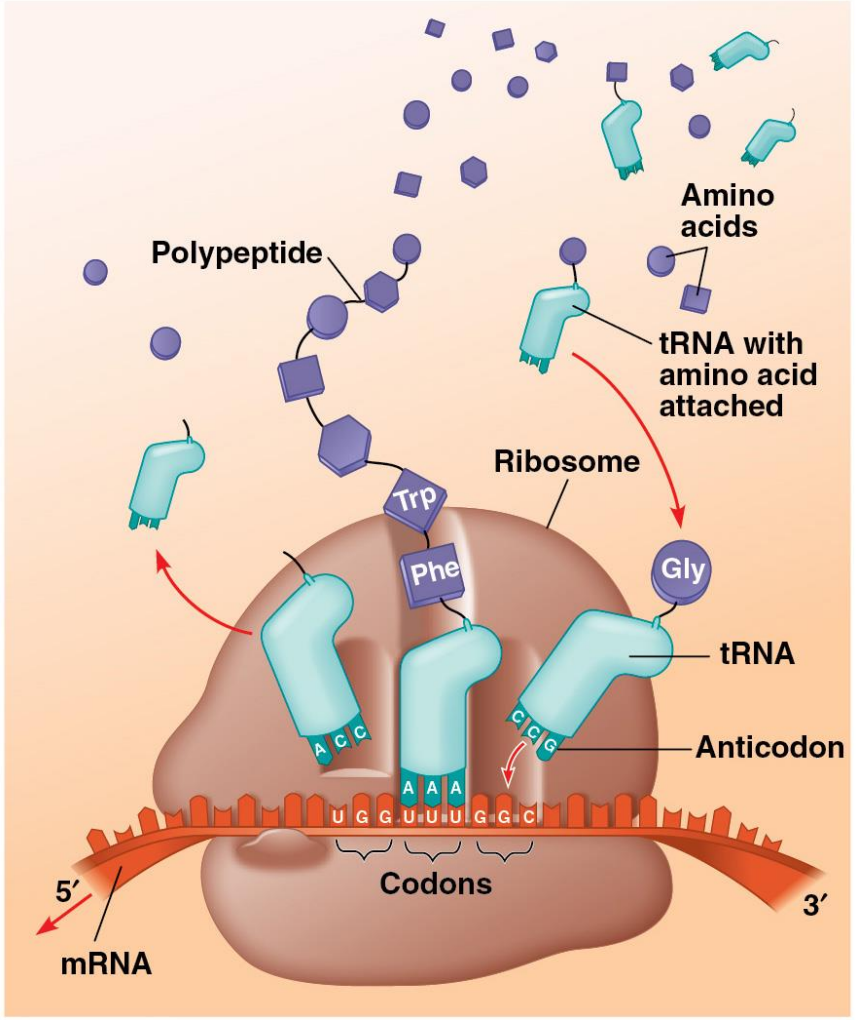


Figure 14.18

The Initiation of Translation

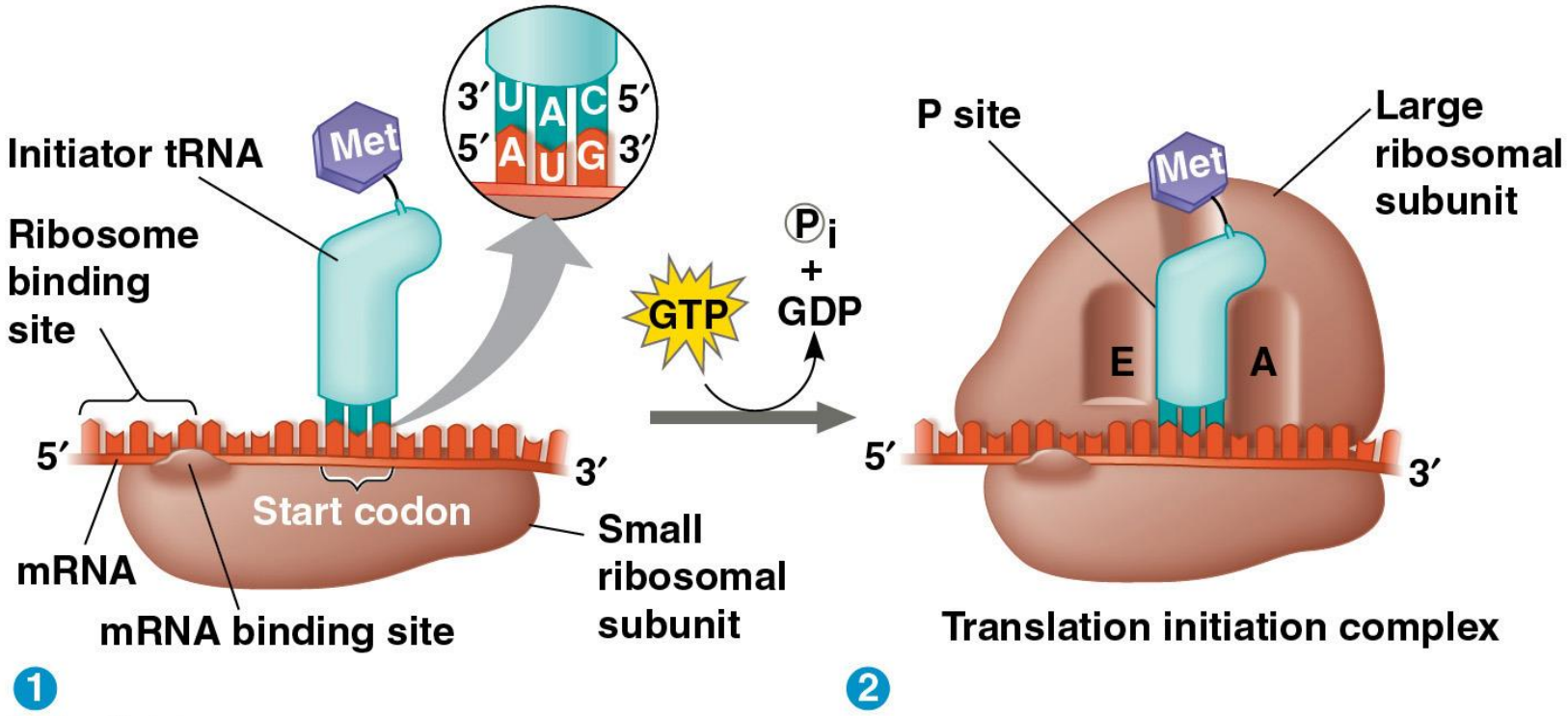


Figure 14.19

The Elongation Cycle of Translation

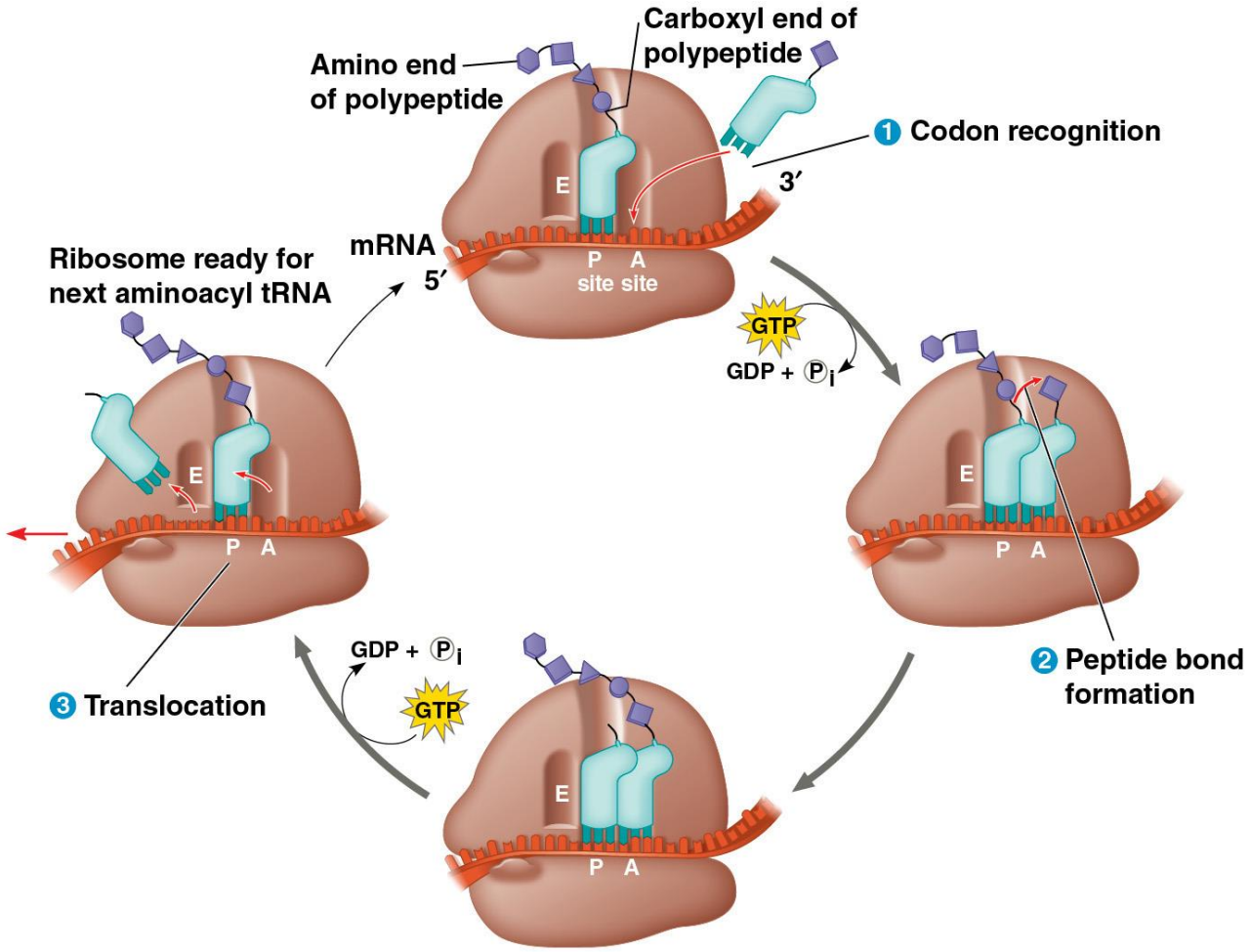
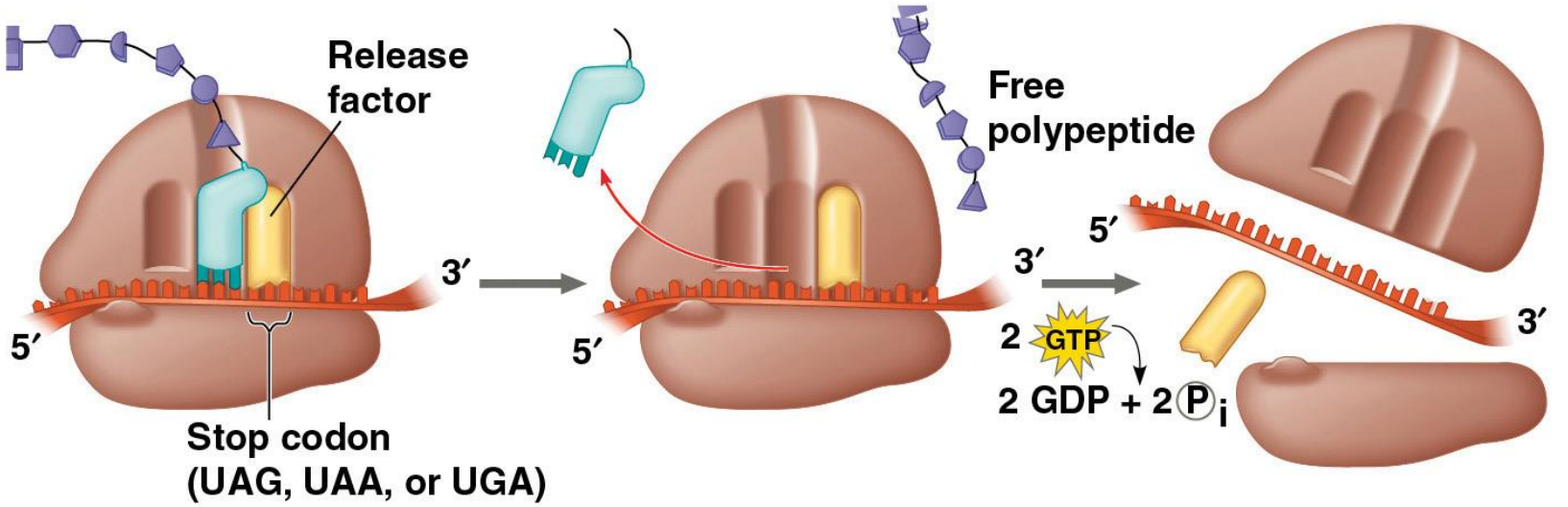


Figure 14.20

The Termination of Translation



1 Ribosome reaches a stop codon on mRNA. A site accepts release factor.

2 Release factor promotes hydrolysis, freeing polypeptide.

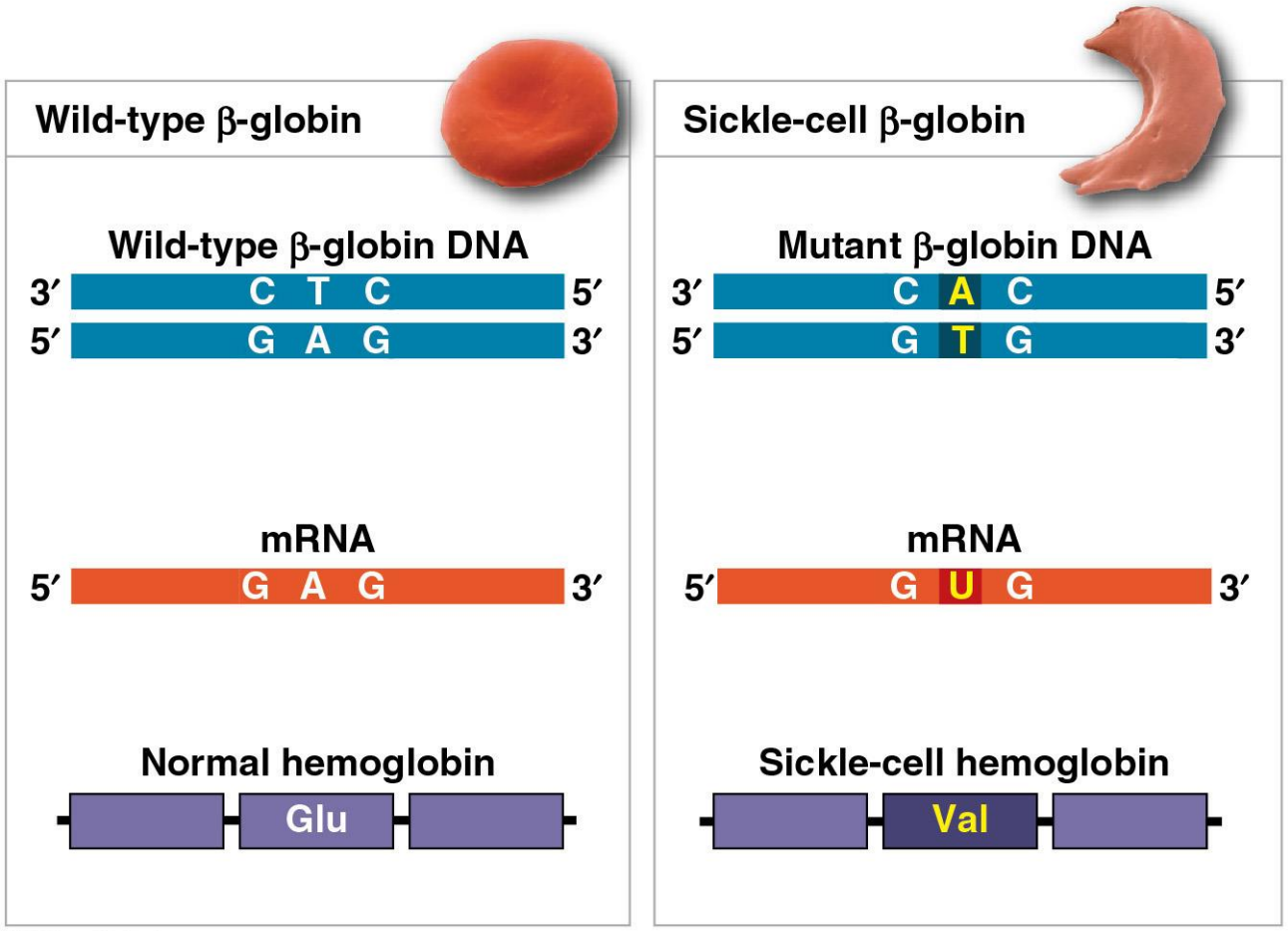
3 Ribosomal subunits and other components dissociate.

Concept 14.5: Mutations of One or a Few Nucleotides Can Affect Protein Structure and Function

- **Mutations** are changes in the genetic material of a cell
- **Point mutations** are chemical changes in just one nucleotide pair of a gene
- The change of a single nucleotide in a DNA template strand can lead to the production of an abnormal protein
- If a point mutation occurs in a gamete, it may be transmitted to offspring

Figure 14.25

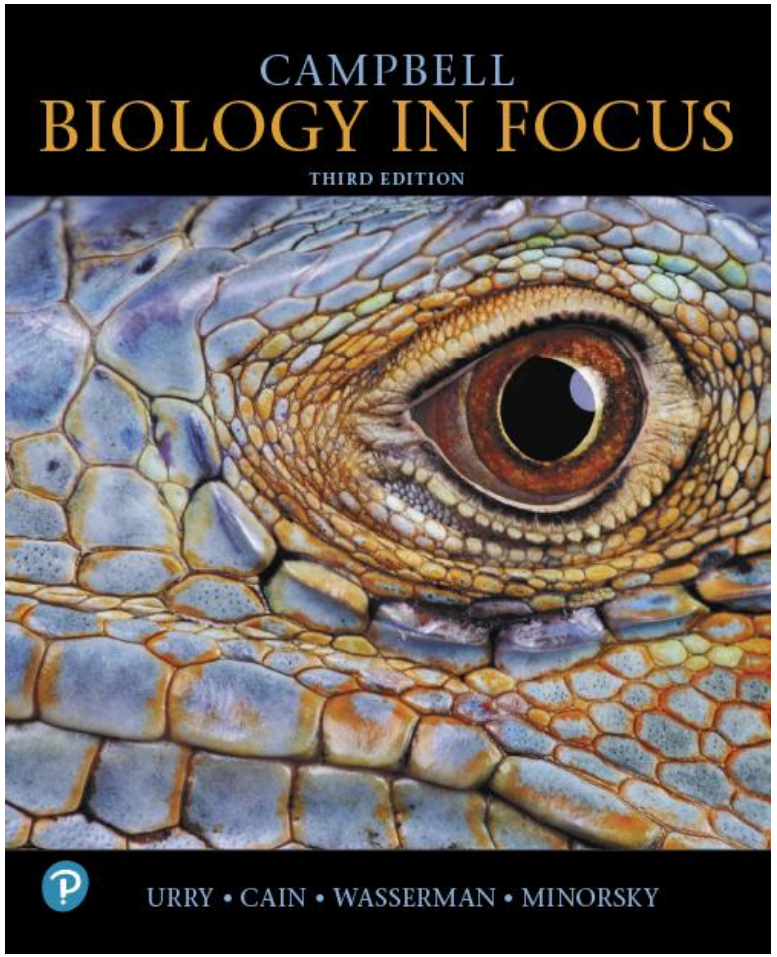
the Molecular Basis of Sickle-Cell Disease: A Point Mutation



Genomes and Their Evolution

Campbell Biology in Focus

Third Edition



Chapter 18

Genomes and Their Evolution

Lecture Presentations by
Kathleen Fitzpatrick and Nicole Tunbridge,
Simon Fraser University

Pearson

Copyright © 2022, 2020, 2016 Pearson Education, Inc. All Rights Reserved

Overview: Mining the Genome (1 of 2)

- Complete genome sequences exist for *E. coli* and other prokaryotes, as well as for many eukaryotes, including humans, corn, chimpanzees, and even *Homo neanderthalensis*
- These genomes provide important insights into evolution
- Comparing the human genome sequence to that of the chimpanzee and other primates should reveal sets of genes that define group characteristics

Overview: Mining the Genome (2 of 2)

- **Genomics** is the study of whole sets of genes and their interactions
- **Bioinformatics** is the application of computational methods to the storage and analysis of biological data

Concept 18.1: The Human Genome Project Fostered Development of Faster, Less Expensive Sequencing Techniques (1 of 4)

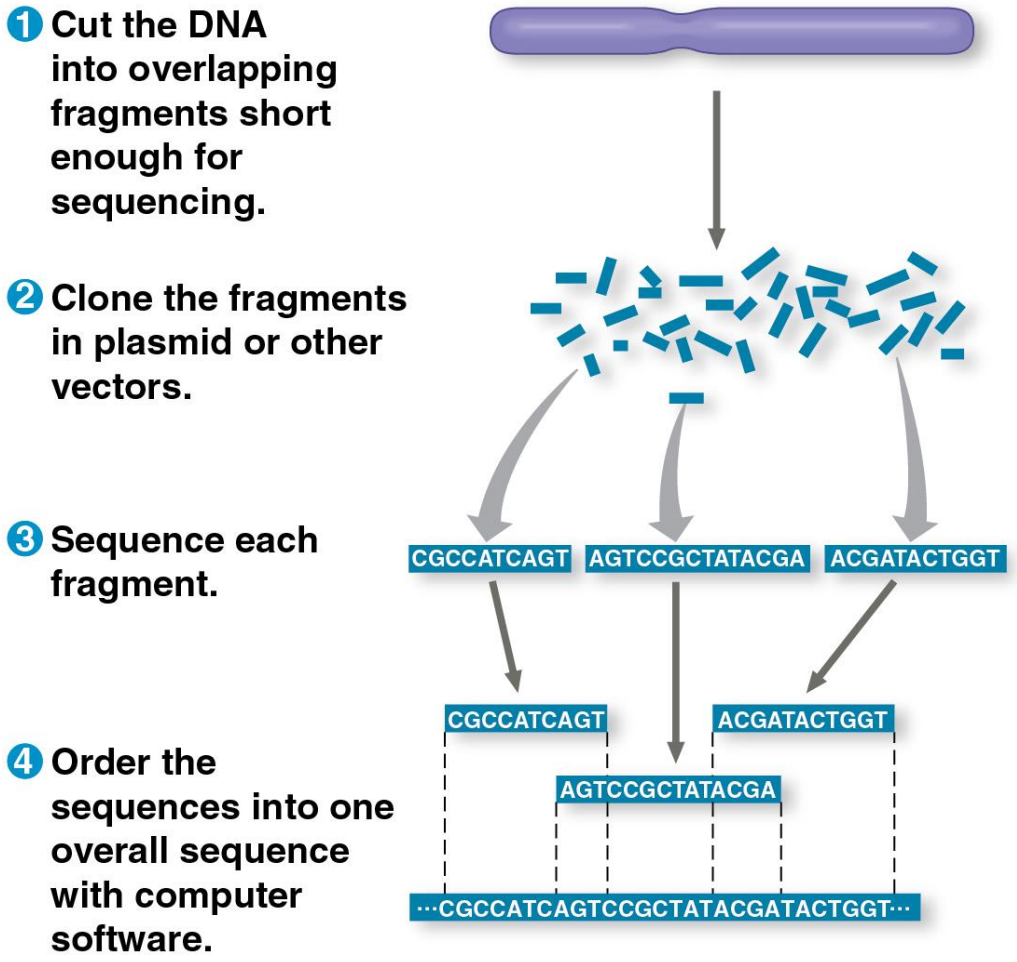
- The **Human Genome Project** officially began in 1990, and the sequencing was largely completed by 2003
- A major thrust of the Human Genome Project was the development of sequencing machines with automated technology for faster sequencing
- Methods that can analyze biological materials rapidly and produce large amounts of data are called "high through-put"

Concept 18.1: The Human Genome Project Fostered Development of Faster, Less Expensive Sequencing Techniques (2 of 4)

- The **whole-genome shotgun approach** was developed by J. Craig Venter and colleagues
- This approach starts with cloning and sequencing random DNA fragments
- Powerful computer programs are used to assemble the resulting short overlapping sequences into a single continuous sequence

Figure 18.2

Whole-Genome Shotgun Approach to Sequencing



Concept 18.2: Scientists Use Bioinformatics to Analyze Genomes and Their Functions

- The Human Genome Project established databases and refined analytical software to make data available on the Internet

Centralized Resources for Analyzing Genome Sequences (1 of 3)

- Bioinformatics resources are provided by a number of sources
  - National Library of Medicine and the National Institutes of Health (NIH) maintain the National Center for Biotechnology Information (NCBI)
  - European Molecular Biology Laboratory
  - DNA Data Bank of Japan
  - BGI in Shenzhen, China

Centralized Resources for Analyzing Genome Sequences (2 of 3)

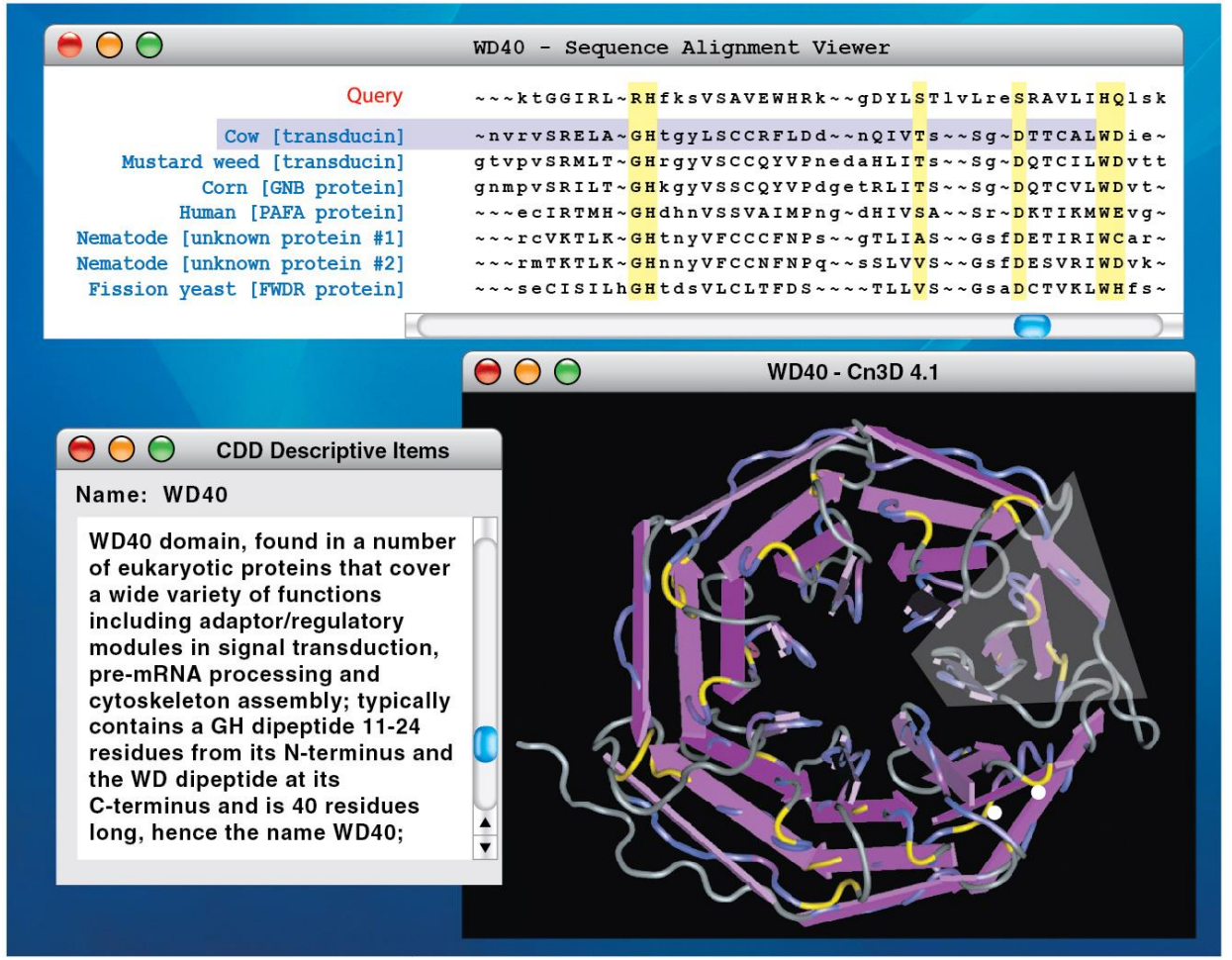
- GenBank, the NCBI database of sequences, is constantly updated, and the data it contains increase rapidly
- Software is available that allows online visitors to search GenBank for matches to
  - A specific DNA sequence
  - A predicted protein sequence
  - Common stretches of amino acids in a protein and a three-dimensional model of the domain

Centralized Resources for Analyzing Genome Sequences (3 of 3)

- Sequences of DNA or proteins can be diagrammed as an evolutionary tree based on sequence relationships
- The Protein Data Bank contains all three-dimensional protein structures that have been determined
- This vast array of resources can be used by researchers anywhere in the world, free of charge

Figure 18.3

National Center for Biotechnology Information (NCBI) Website



Systems Biology

- **Proteomics** is the systematic study of the full protein sets (**proteomes**) expressed by cells
- We must study when and where proteins are produced in an organism in order to understand the function of cells and organisms
- **Systems biology** aims to model the dynamic behavior of whole biological systems based on the study of interactions among the system's parts

Application of Systems Biology to Medicine (1 of 2)

- A systems biology approach has several medical applications
  - The Cancer Genome Atlas project began in 2007 and culminated in 2018 with publications called the Pan-Cancer Atlas
  - This project aimed to identify all the common mutations in several types of cancer by comparing gene sequences and expression in cancer cells versus normal cells
  - Silicon and glass “chips” have been produced that hold a microarray of most known human genes

Application of Systems Biology to Medicine (2 of 2)

- Analyzing which genes are over- or underexpressed in cancers may allow physicians to tailor treatment to particular patients and the specifics of their cancers
- Ultimately, medical records may include an individual's DNA sequence
- The use of such sequences for personalized medicine has great potential

Figure 18.4

Human Gene Microarray Chip

Concept 18.3: Genomes Vary in Size, Number of Genes, and Gene Density

- To date, the sequences of thousands of genomes have been completed
- Tens of thousands of genomes are in progress or considered permanent drafts
- As of 2021, more than 30,000 metagenomes were also in progress

Genome Size

- Genomes of most bacteria and archaea range from 1 to 6 million base pairs (Mb); genomes of eukaryotes are usually larger
- Most plants and animals have genomes greater than 100 Mb; humans have 3,000 Mb
- Within each domain there is no systematic relationship between genome size and phenotype

Table 18.1 (1 of 2)

Genome Sizes and Estimated Numbers of Genes*

Organism	Haploid Genome Size (Mb)†	Number of Genes	Genes per Mb†
Bacteria			
Haemophilus influenzae	1.8	1,700	940
Escherichia coli	4.6	4,400	950
Archaea			
Archaeoglobus fulgidus	2.2	2,500	1,130
Eukaryotes			
Saccharomyces cerevisiae (yeast, a fungus)	12	6,300	525
Caenorhabditis elegans (nematode)	100	20,100	200
Arabidopsis thaliana (mustard family plant)	120	27,000	225

Table 18.1 (2 of 2)

Genome Sizes and Estimated Numbers of Genes*

Organism	Haploid Genome Size (Mb)†	Number of Genes	Genes per Mb†
Drosophila melanogaster (fruit fly)	165	14,000	85
Zea mays (corn)	2,300	32,000	14
Ailuropoda melanoleuca (giant panda)	2,400	21,000	9
Homo sapiens (human)	3,000		7

*Some values given here are likely to be revised as genome analysis continues.

†**Mb = million** base pairs; the haploid number is used because it represents a complete set of genetic information.

Concept 18.6: Comparing Genome Sequences Provides Clues to Evolution and Development

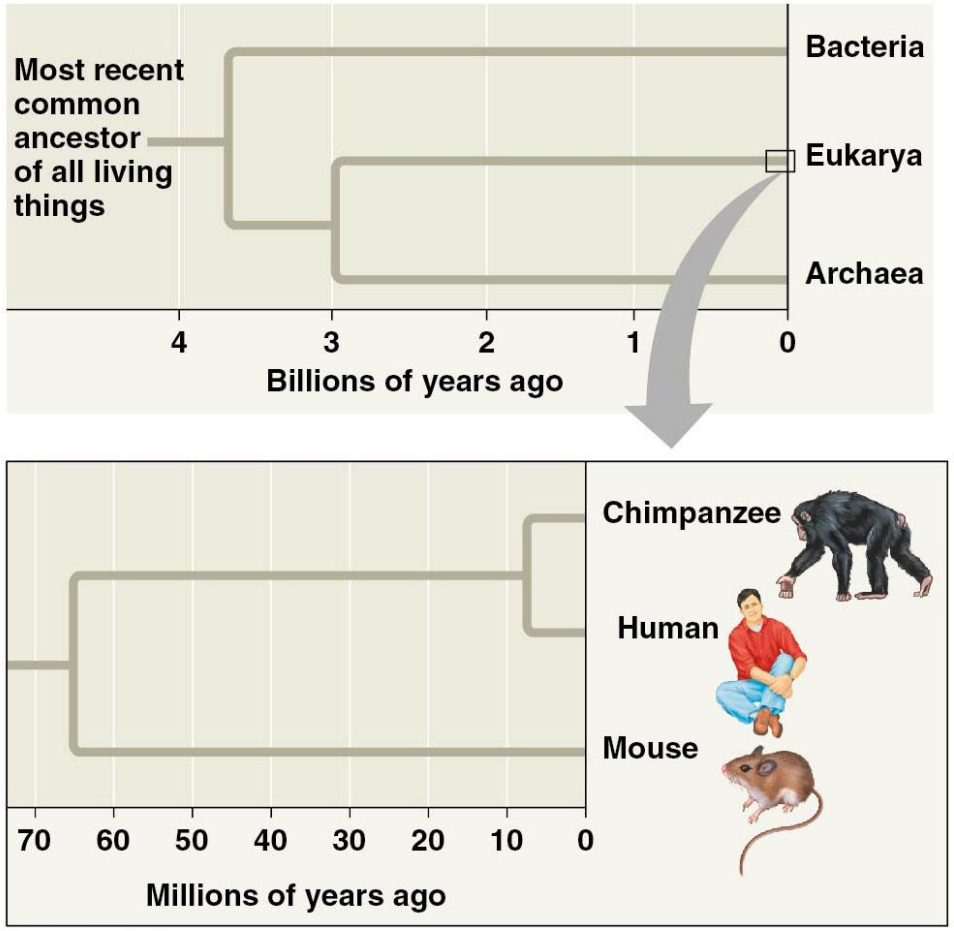
- Genome sequencing and data collection have advanced rapidly in the last 30 years
- Comparative studies of genomes
  - Reveal much about the evolutionary history of life
  - Uncover the mechanisms that generated the great diversity of present-day life-forms

Comparing Genomes

- Genome comparisons of closely related species shed light on recent evolutionary events
- Genome comparisons of distantly related species help us understand ancient evolutionary history
- Evolutionary relationships among species can be represented by a tree-shaped diagram

Figure 18.15

Evolutionary Relationships of the Three Domains of Life



Comparing Distantly Related Species

- Highly conserved genes have remained similar over time
- These help clarify relationships among species that diverged from each other long ago
- Bacteria, archaea, and eukaryotes diverged from each other between 2 and 4 billion years ago
- Comparative genomic studies confirm the relevance of research on model organisms to our understanding of biology in general and human biology in particular

Comparing Closely Related Species
(1 of 5)

- The genomes of two closely related species are likely to be organized similarly
- Particular genetic differences between the two species can be easily correlated with phenotypic differences between them

Comparing Closely Related Species
(2 of 5)

- Human and chimpanzee genomes differ by 1.2% at single base pairs and by 2.7% because of insertions and deletions
- Bonobos are the other African ape species that are the closest living relatives to humans
- Sequencing of the bonobo genome reveals that in some regions human sequences are more closely related to either chimpanzee or bonobo sequences
- This comparison allows a detailed evolutionary history to be reconstructed

Comparing Closely Related Species
(3 of 5)

- Several genes are evolving faster in humans than in chimpanzees
- These include genes involved in defense against malaria and tuberculosis and in regulation of brain size
- Genes that seem to be evolving fastest code for transcription factors

Copyright

This work is protected by United States copyright laws and is provided solely for the use of instructors in teaching their courses and assessing student learning. Dissemination or sale of any part of this work (including on the World Wide Web) will destroy the integrity of the work and is not permitted. The work and materials from it should never be made available to students except by instructors using the accompanying text in their classes. All recipients of this work are expected to abide by these restrictions and to honor the intended pedagogical purposes and the needs of other instructors who rely on these materials.

Genetics Tools

Comparing CS and Biological information

Biology	Computer science
1. Digital alphabet consists of bases A, C, T, G	1. Digital alphabet consists of 0, 1
2. Codons consist of three bases	2. Computer bits form bytes
3. Genes consist of codons	3. Files consist of bytes
4. Promoters indicate gene locations	4. File-allocation table indicates file locations
5. DNA information is transcribed into hnRNA and processed into mRNA	5. Disc information is transcribed into RAM
6. mRNA information is translated into proteins	6. RAM information is translated onto a screen or paper
7. Genes may be organized into operons or groups with similar promoters	7. Files are organized into folders
8. "Old" genes are not destroyed; their promoters become nonfunctional	8. "Old" files are not destroyed; references to their location are deleted
9. Entire chromosomes are replicated	9. Entire discs can be copied
10. Genes can diversify into a family of genes through duplication	10. Files can be modified into a family of related files
11. DNA from a donor can be inserted into host chromosomes	11. Digital information can be inserted into files
12. Biological viruses disrupt genetic instructions	12. Computer viruses disrupt software instructions
13. Natural selection modifies the genetic basis of organism design	13. Natural selection procedures modify the software that specifies a machine design
14. A successful genotype in a natural population outcompetes others	14. A successful website attracts more "hits" than others

https://www.cl.cam.ac.uk/~pl219/Bioinformatics2015.pdf

Several Bioinformatics Tasks

- Sequence Analysis (similarity searching)
- Sequence Assembly
- Genome Annotation
- Comparative Genomics
- Analysis Mutation in Cancer
- Biomedical Data Mining

Computing Tasks Related

- Searching and Pattern Matching: Sequence Analysis
- Machine Learning for detection/prediction: Disease prediction, Gene clustering
- Simulation and modeling: Molecular Dynamics Simulations, Pathway and Metabolic Network Modeling

Sequence Alignment

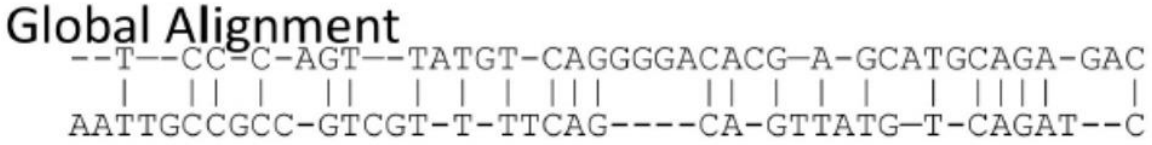
- Alignment is a way of arranging two DNA or protein sequences to identify regions of similarity that are conserved among species. Each aligned sequence appears as a row within a matrix.
- Given two strings x = x1, x2, ..., xM , and y = y1, y2, ..., yN , An alignment is an assignment of gaps to positions 0, ..., M in x, and 0, ..., N in y, so as to line up each letter in one sequence with either a letter, or a gap in the other sequence.

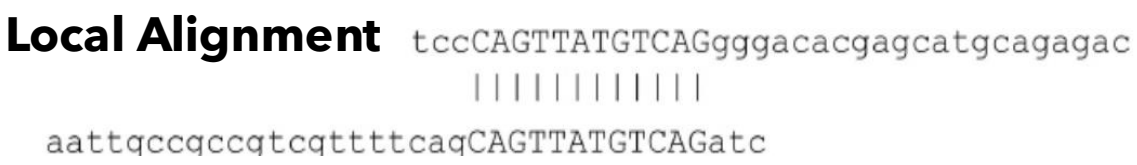
Purpose:

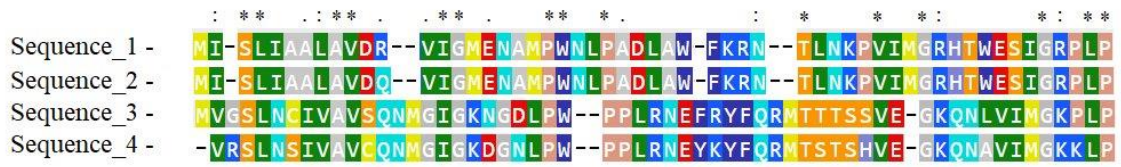
- Finding genetic similarities between organisms.
- Analyzing evolutionary relationships (phylogenetics).
- Identifying mutations such as substitutions, insertions, and deletions.
- Predicting the function of a gene or protein by comparing it to known sequences.

Types and Algorithms

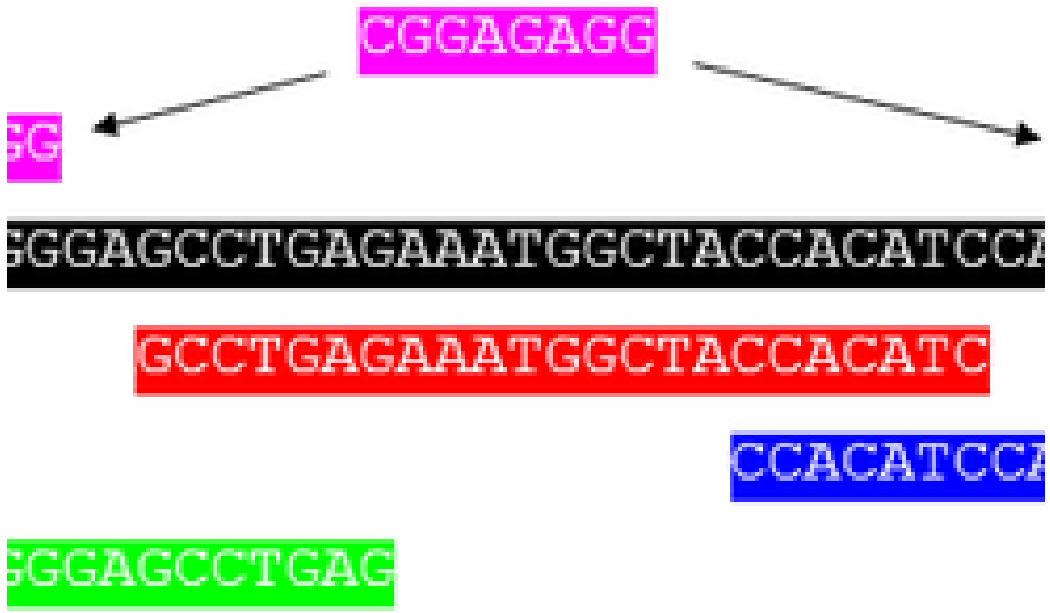
1. Global Alignment: Align the entire sequence from start to finish. Algorithm: Needleman-Wunsch
2. Local Alignment: Find the best alignment between two sequences. Algorithm: Smith-Waterman
3. Multiple Sequence Alignment: Align more than two sequences at once. Algorithms/Tools: ClustalW, MUSCLE







Genome Assembly

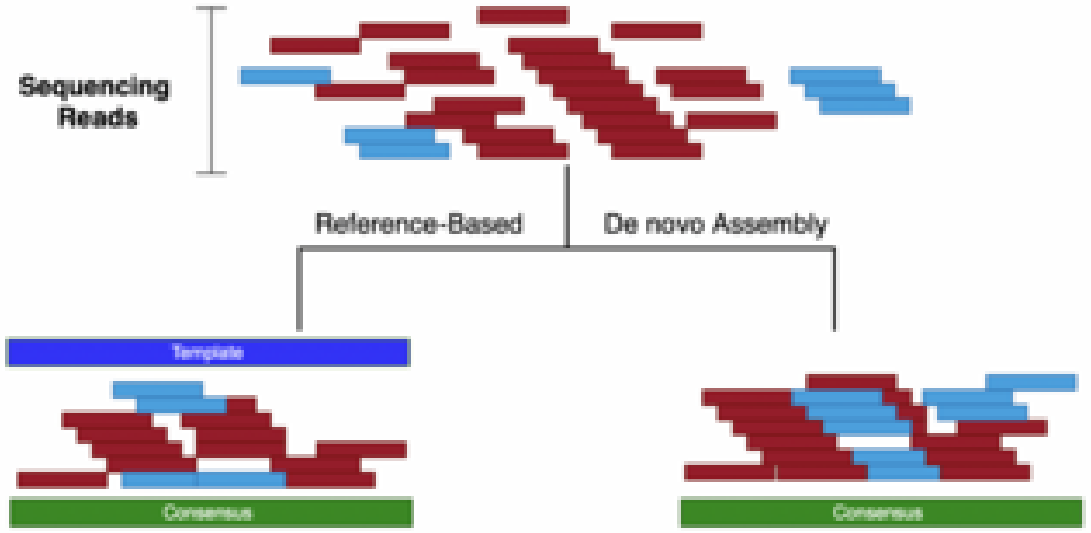


- Genome assembly refers to the process of putting nucleotide sequence into the correct order. Assembly is required, because sequence read lengths – at least for now – are much shorter than most genomes or even most genes.
- How do genome assembly algorithms work?

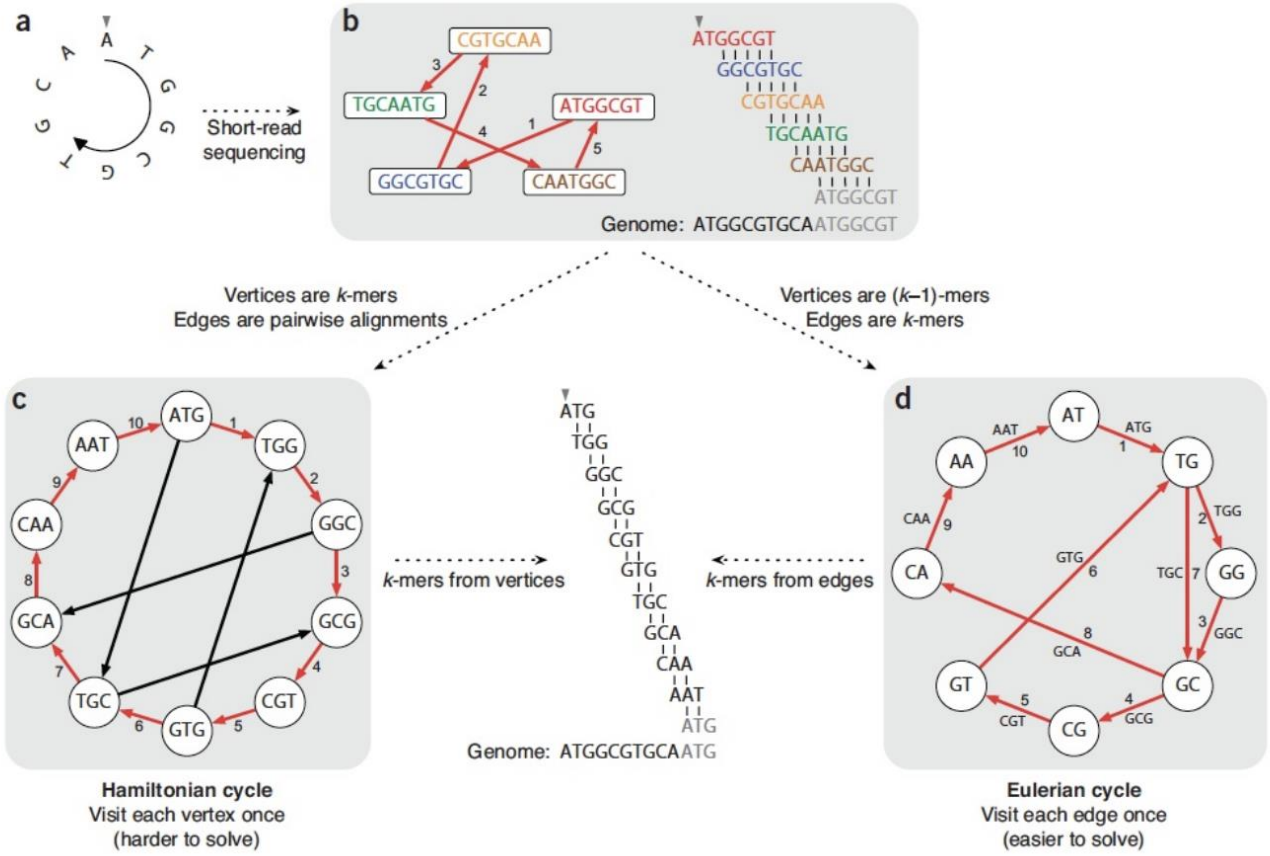
1. Align DNA fragments against each other
2. Identify overlapping sections
3. Merge the overlapping sections

Assembly Approaches and Algorithms

- Approaches to assembling sequencing data:
  - De-novo: assembling sequencing reads to create full-length (sometimes novel) sequences, without using a template
  - Mapping/Aligning: assembling reads by aligning reads against a template (AKA reference).
  - Reference-guided: grouping of reads by similarity to the most similar region within the reference (step wise mapping). A typical method to do so is the k-mer approach.
- Algorithms:
  - Overlap relationships between sequences: Overlap-Layout-Concensus (OLC)
  - Graph Assembly: de Brujin Graph (DBG)
  - Greedy strategy



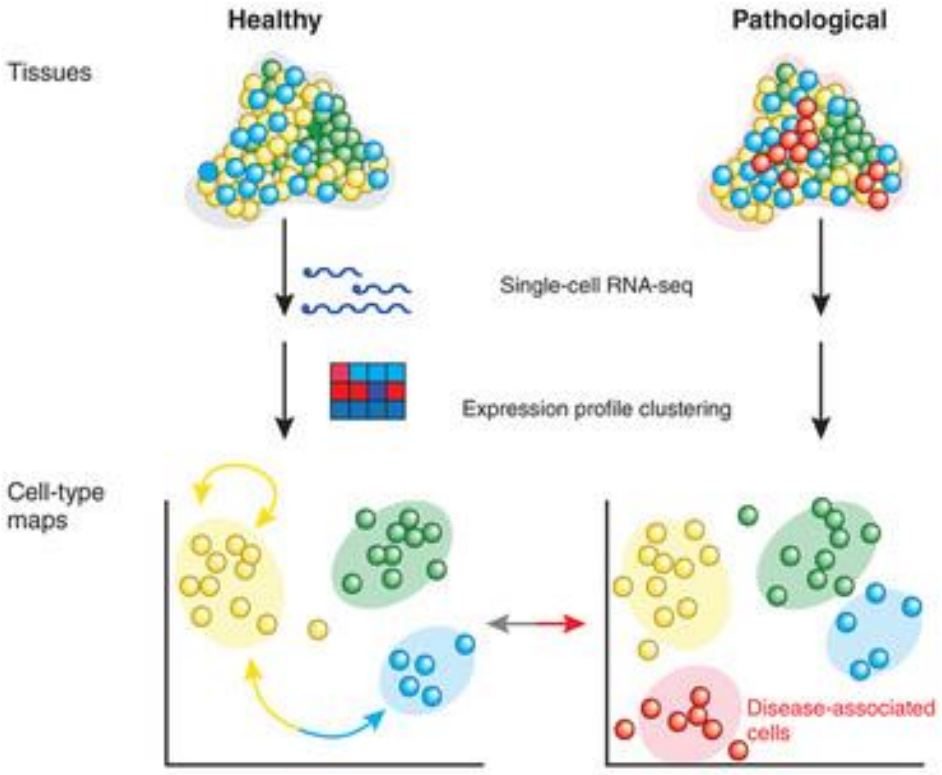
Three Graph Approaches in Genome Alignment



https://www.cl.cam.ac.uk/~pl219/Bioinformatics2015.pdf

Gene Expression Clustering

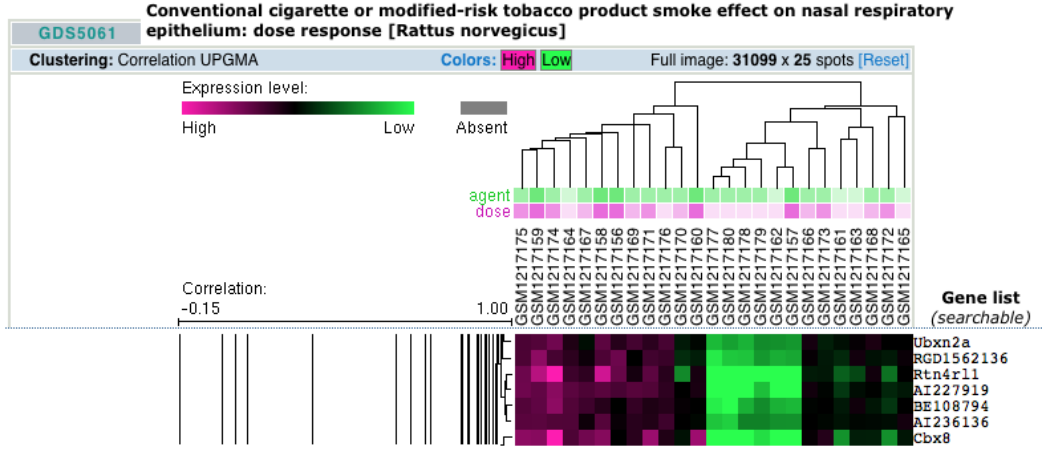
- Gene expression clustering is a computational technique used to group genes together based on their similar expression patterns across different conditions
- It allows researchers to identify sets of genes that behave similarly and potentially share related functions within a biological system.



https://genetics564.weebly.com/gene-expression-analysis.html

Objectives of Gene Expression Clustering

1. Identifying groups of genes that work together in a biological process.
2. Detecting biomarkers for specific diseases (e.g., cancer).
3. Clustering patient samples based on gene expression patterns for more personalized therapy.
4. Understanding relationships between genes in metabolic pathways or genetic regulation.

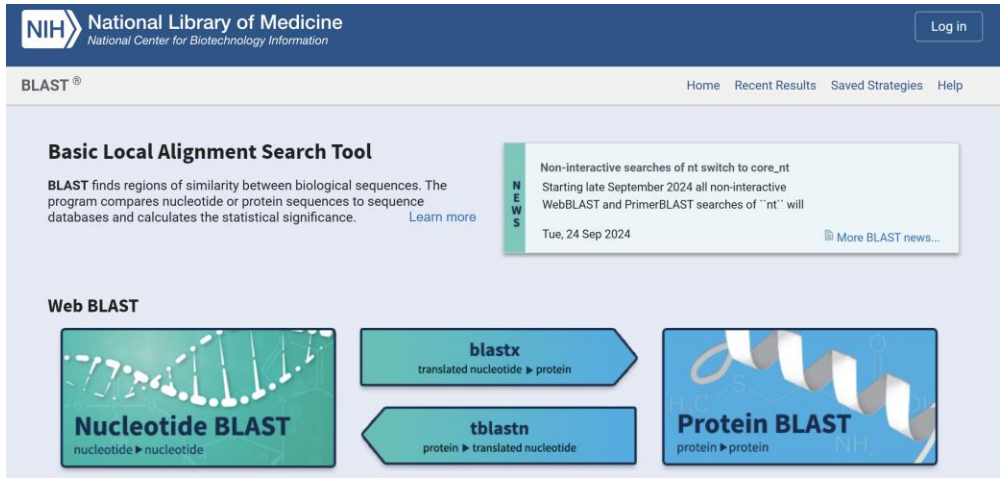


https://genetics564.weebly.com/gene-expression-analysis.html

Genetics Repositories and Tools

Repositories:

- https://www.ebi.ac.uk/
- https://www.ncbi.nlm.nih.gov/
- https://www.ensembl.org/index.html
- https://www.ddbj.nig.ac.jp/index-e.html



Tools:

- Basic Local Alignment Search Tool: https://blast.ncbi.nlm.nih.gov/Blast.cgi
- Gene Expression Clustering Tool: https://docs.gdc.cancer.gov/

Tool Exploration: Bioinformatics Tools for Genetic Analysis

Purpose: to understand how these tools work and how genetic data can be analyzed using computational methods.

1. Choose one of the following tools to explore (or another relevant tool):
   - NCBI BLAST (matches DNA/protein sequences with a database)
   - UCSC Genome Browser (genome visualization)
   - Ensembl Genome Browser (genome annotation exploration)
2. Explore the following datasets:
   - DNA Sequences: Use human DNA sequences from the NCBI or Ensembl databases.
   - Protein Sequences: Retrieve protein sequences from the UniProt database.
   - Genetic Mutation Data: Use mutation datasets from COSMIC or dbSNP.

Tool Exploration [2]

3. Perform one of the following analyses based on the chosen tool:
    - BLAST: Search for homologs of DNA/protein sequences using NCBI BLAST.
    - Genome Browser: Locate and visualize a specific gene in the human genome.
4. Prepare an exploration report (pdf 1-2 pages) containing:
    - Description of the selected tool (main function, working principles, advantages).
    - Step-by-step analysis (including screenshots or code).
    - Analysis results (e.g., BLAST output, gene annotation).
    - Conclusions and reflections on the usefulness of the tool in bioinformatics.

Deadline: 1 week, 19/03/2025. Submission Format: PDF, submitted via Edunex.

The use of generative-AI tools in any form is not permitted.

References

- University of Cambridge, Bioinformatics Algorithms, https://www.cl.cam.ac.uk/~pl219/Bioinformatics2015.pdf
- "How does gene expression clustering work?", https://www.gene-quantification.de/haeseleer-bioinf-2005.pdf
- "Gene Expression & Transcriptome Analysis", https://genetics564.weebly.com/gene-expression-analysis.html

Project Assignment

Students will work in groups (2-3 people) to develop computational biology-based solutions. Each group chooses one of the following topics, or proposes another related topic:

1. DNA/RNA Sequence Analysis: Using sequence matching algorithms to analyze genomic data from public databases.
2. Protein Structure Prediction: Developing a computational model to predict protein structure based on amino acid sequences.
3. Biological Network Simulation: Creating a simulation model of protein interactions or metabolic networks using bioinformatics software.
4. Application of Machine Learning in Biology: Training a machine learning model for DNA sequence classification, mutation detection, or gene expression analysis.
5. Metagenomic Analysis: Using computational techniques to identify and classify microbial species in environmental samples.

Project Assignment [2]

6. Evolution: Using computational models to simulate the process of evolution and natural selection in various organisms.

7. Biological Diversity: Analyzing and visualizing biodiversity using genomic and bioinformatics data.

8. Plant Form and Function: Developing computational models to understand genetic regulation in plant growth and development.

9. Animal Form and Function: Using computational biology techniques to analyze the physiological and morphological adaptations of animals to the environment.

10. Ecology: Modeling population dynamics and ecosystem interactions using computational methods.

- Expected Outputs:
  1. Program Code: Implementation of algorithms in Python.
  2. Report in the form of a Scientific Paper: Complete presentation of methods, results, and analysis.
  3. Presentation: Presentation of project results at the end of the semester.
- Project assessment criteria include accuracy of analysis (30%), solution innovation (20%), report quality (20%), teamwork (10%), and presentation (20%).
- Timeline:

  Week IV 12 March 2025 : Assignment release

  Week VI 26 March 2025 : Submission of topic proposals (topic distribution will be discussed in class)

  Week VII 9 April 2025 : Initial project design

  Week VIII - XIV : Project development, Testing and Analysis, Paper preparation (outside class). Lecturers can check progress in class.

  Week XV 4 June 2025 : Project presentation and Paper report