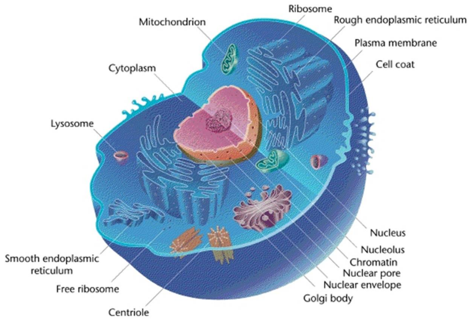
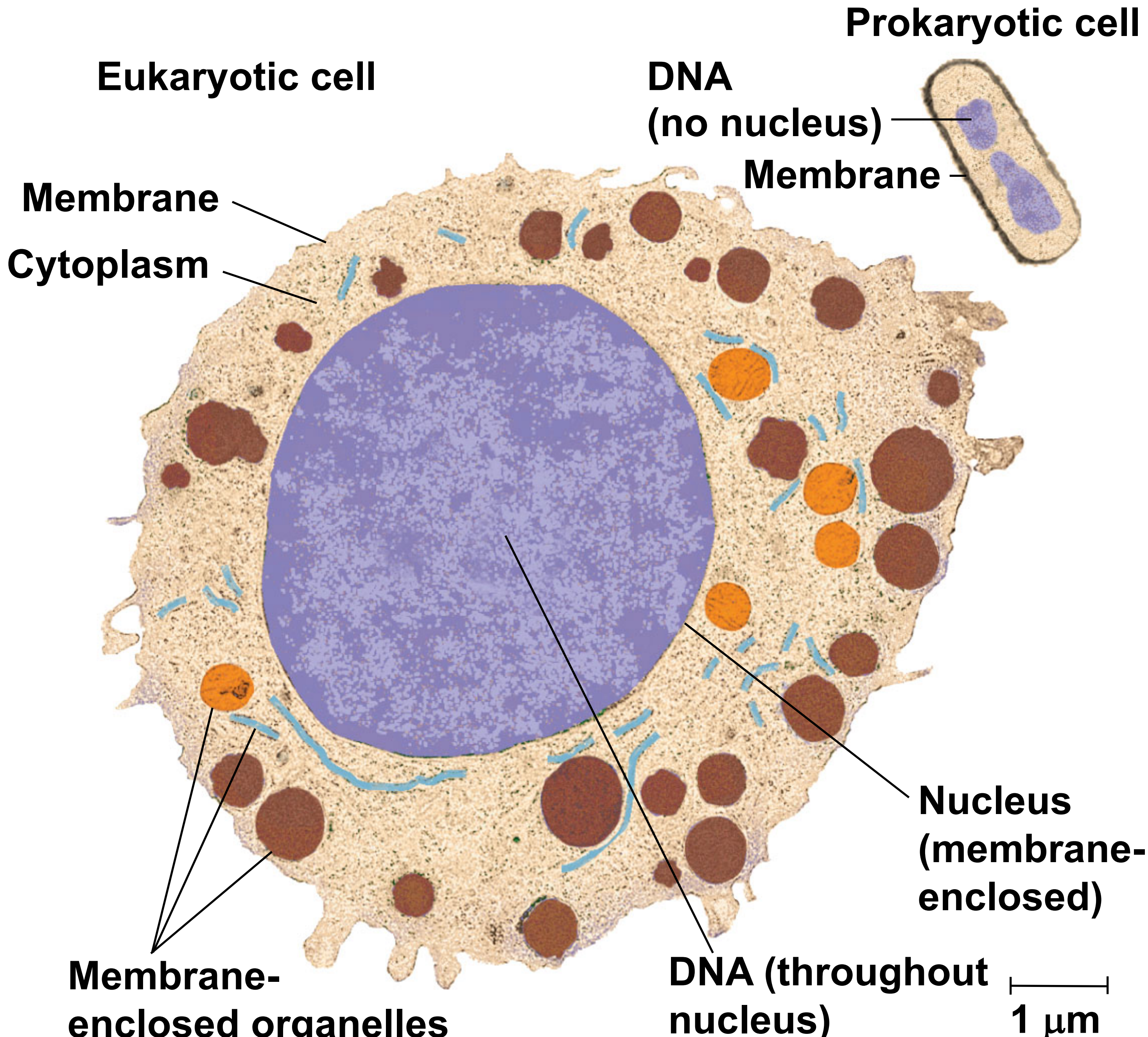
The cell as the basic unit of life



Prokaryotic	Eukaryotic
Bacteria, most single-celled organisms	All multicellular organisms
Simple, smaller	Larger, more complex
Lack membrane-bound organelles such as nucleus	Contain membrane-bound organelles such as a nucleus
Circular DNA	Linear DNA (chromosomes)

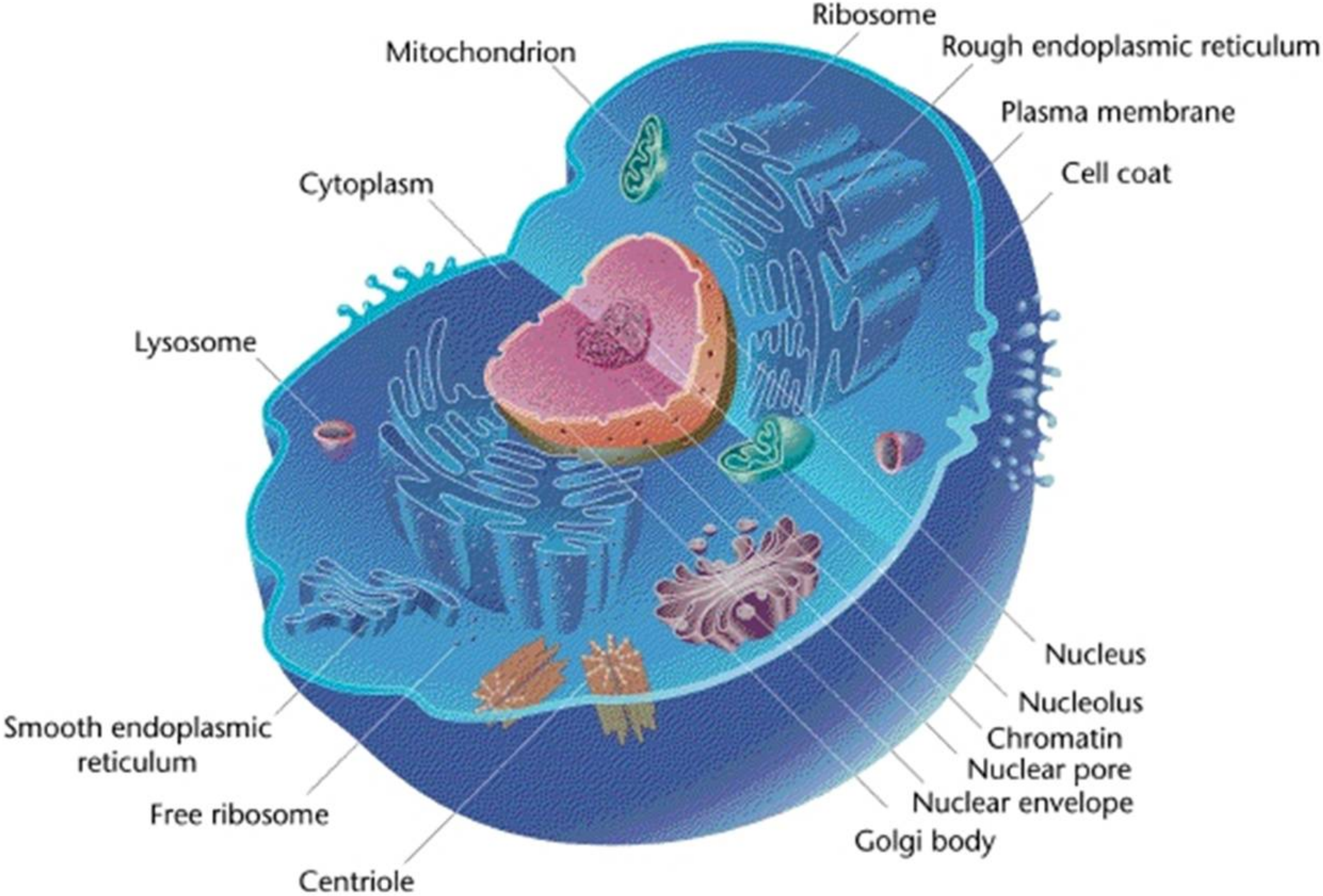
Figure 1.8


Eukaryotic cell
Prokaryotic cell
DNA
(no nucleus)
Membrane
Membrane
Cytoplasm
Nucleus
(membrane-
enclosed)
Membrane-
enclosed organelles
DNA (throughout
nucleus)
1 μm

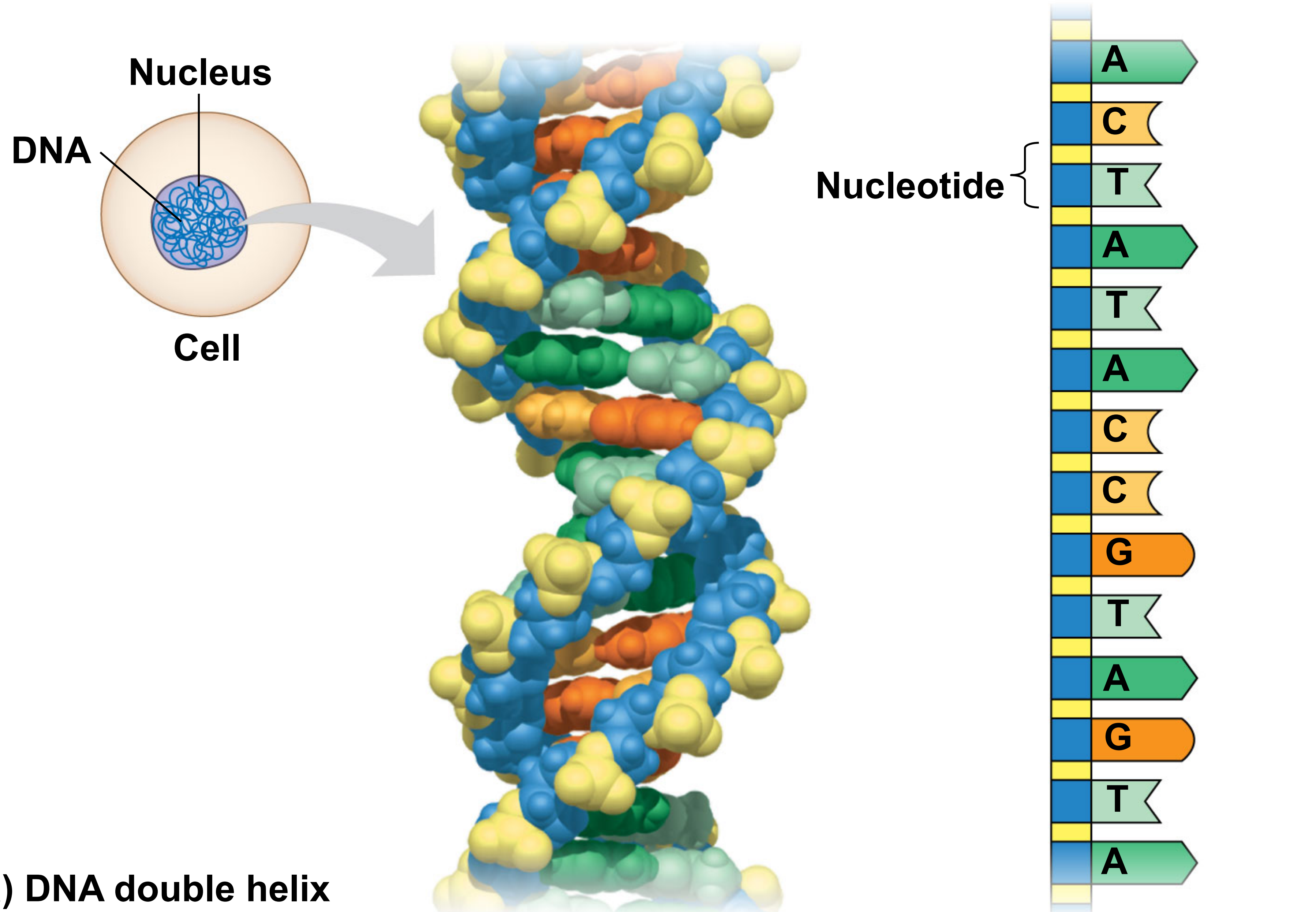
Overview: Cell Structure and Function

- Videos:
  - https://www.youtube.com/watch?v=rABKB5aS2Zg
  - https://www.youtube.com/watch?v=KzMviiBoRtA
- Questions:
  - What is the function of the (cell) *plasma membrane?*
  - What is the function of the *nucleus*?
  - What is the function of the *mitochondria*?
  - What is the function of the *ribosomes*?
  - What is the function of the *Golgi apparatus*?
  - What is the function of the *endoplasmic reticulum (ER)*?

Eukaryotic cell



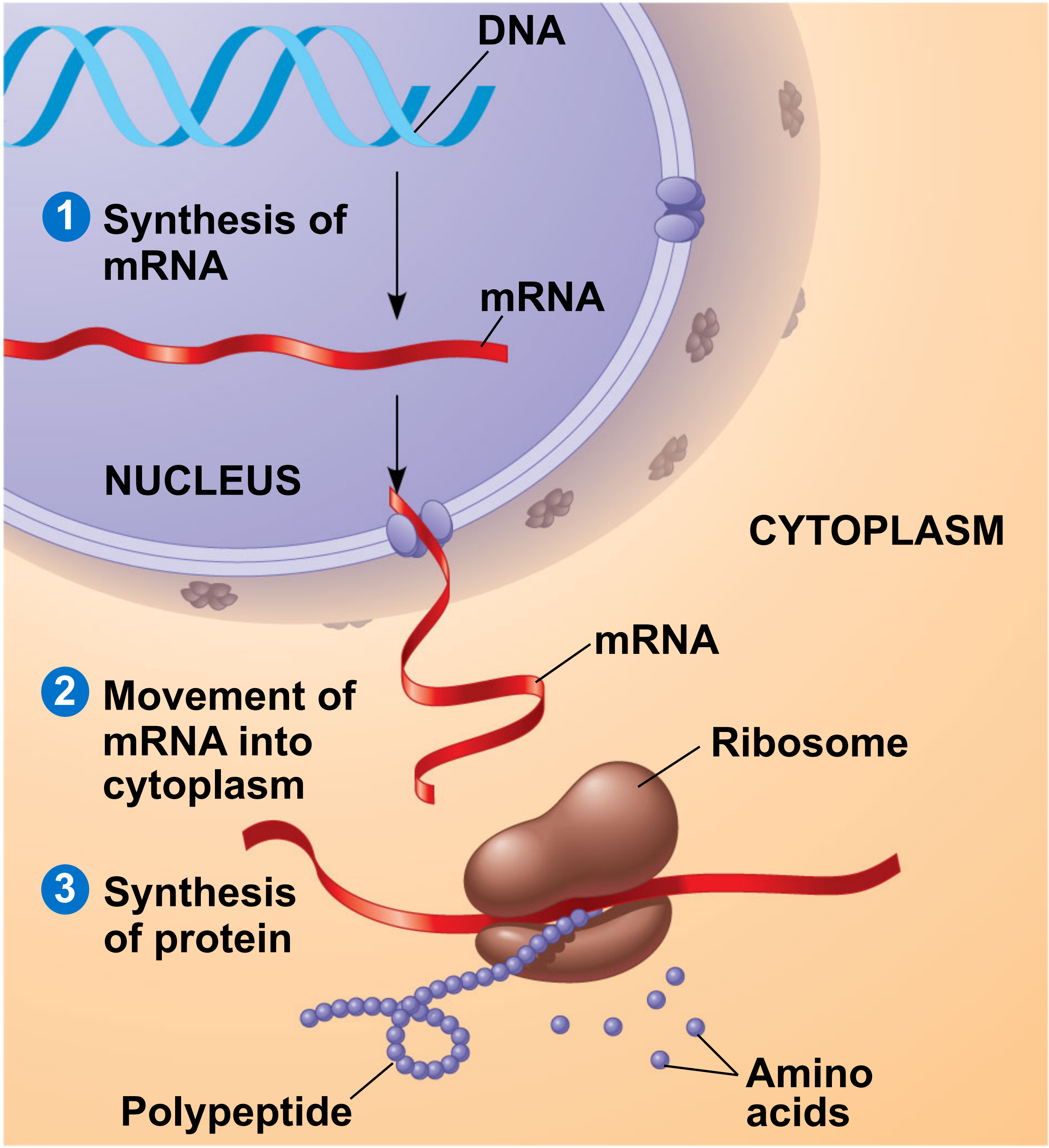
The nucleus contains the DNA



(a) DNA double helix

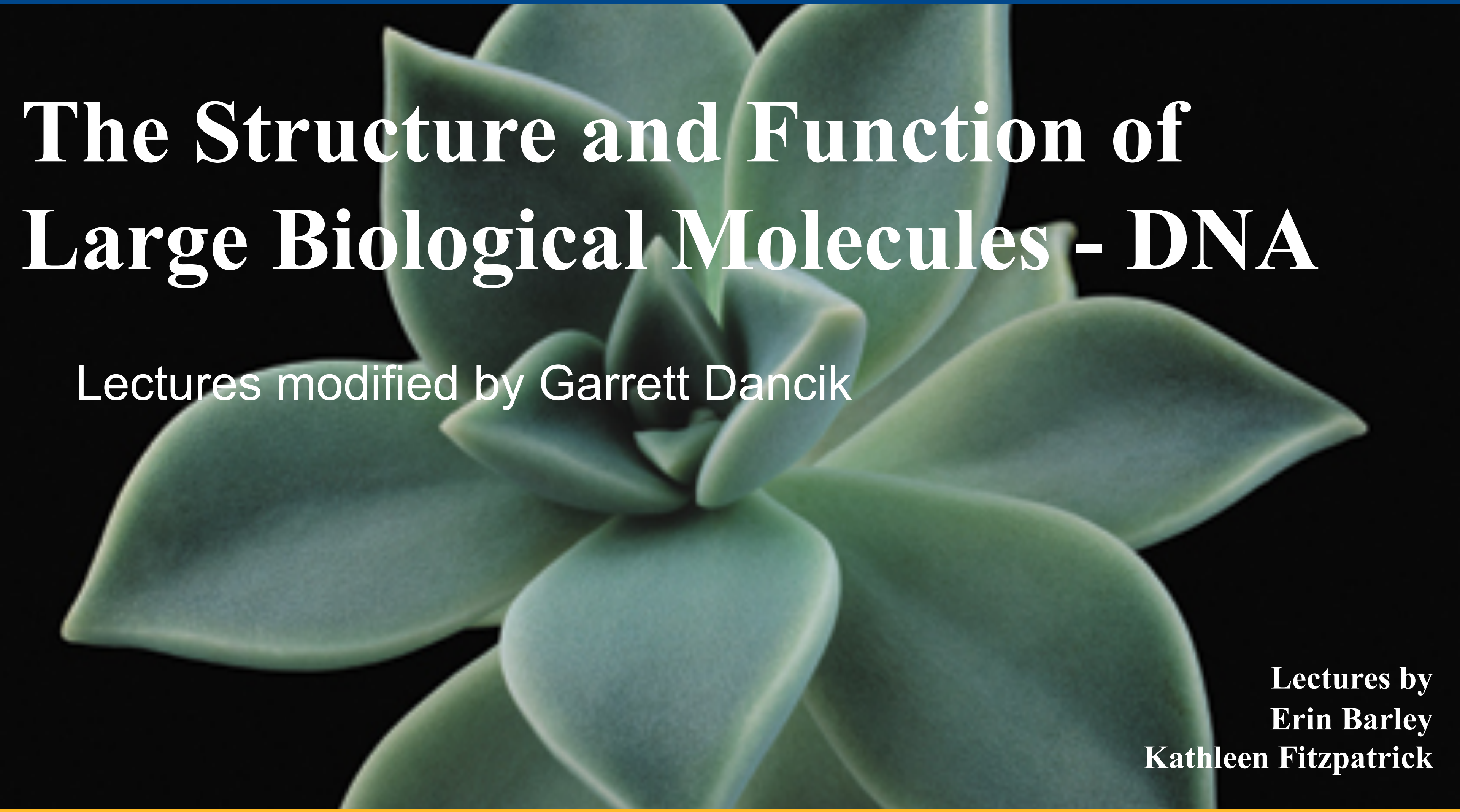
(b) Single strand of DNA

Overview of Gene Expression



LECTURE PRESENTATIONS
For CAMPBELL BIOLOGY, NINTH EDITION
Jane B. Reece, Lisa A. Urry, Michael L. Cain, Steven A. Wasserman, Peter V. Minorsky, Robert B. Jackson

Chapter 5



The Structure and Function of Large Biological Molecules - DNA

Lectures modified by Garrett Dancik

Lectures by
Erin Barley
Kathleen Fitzpatrick

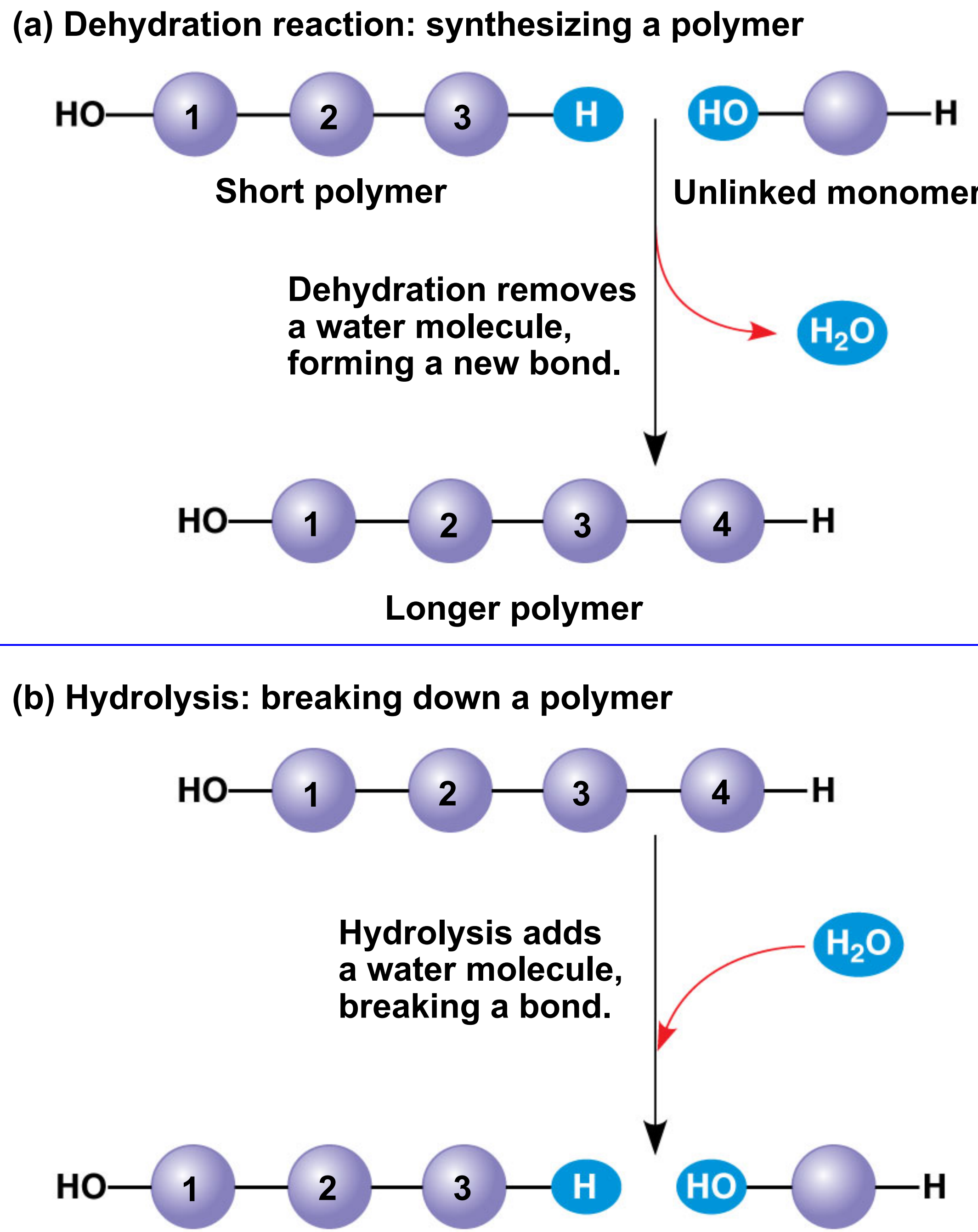
Overview: The Molecules of Life

- All living things are made up of four classes of large biological molecules: carbohydrates, lipids, proteins, and nucleic acids

The Synthesis and Breakdown of Polymers

- A monomer is a building block of a polymer
  - DNA: the nucleotides (characters) A,C,G, and T
  - RNA: the nucleotides (characters) A,C,G, and U
  - Proteins: twenty kinds of amino acids (characters)
- A **dehydration reaction** occurs when two monomers bond together through the loss of a water molecule
- Polymers are disassembled to monomers by **hydrolysis**, a reaction that is essentially the reverse of the dehydration reaction

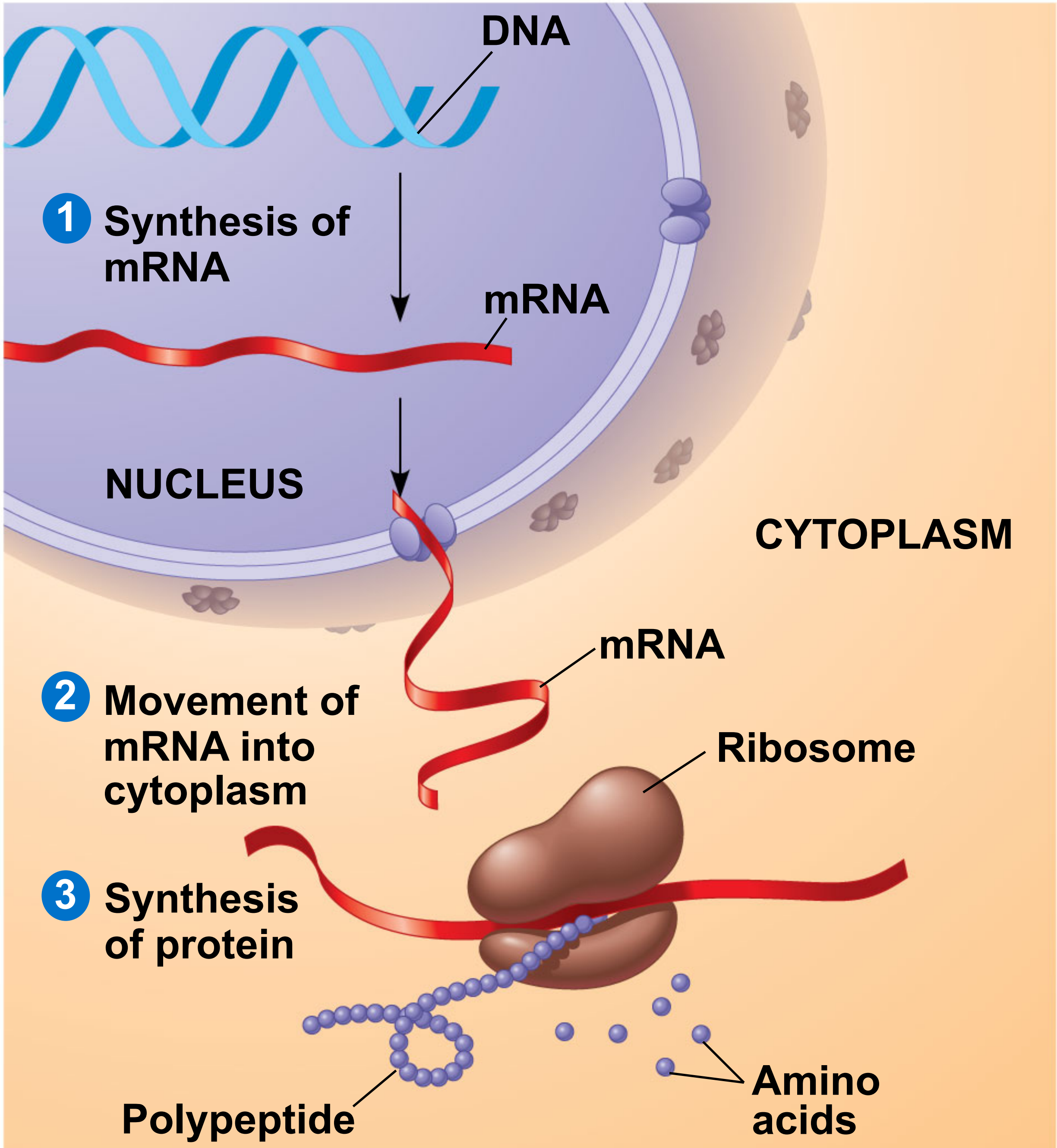
Figure 5.2



(a) Dehydration reaction: synthesizing a polymer
HO
1
2
3
H
HO
H
Short polymer
Unlinked monomer
Dehydration removes a water molecule, forming a new bond.
H_2O
HO
1
2
3
4
H
Longer polymer
(b) Hydrolysis: breaking down a polymer
HO
1
2
3
4
H
Hydrolysis adds a water molecule, breaking a bond.
H_2O
HO
1
2
3
H
HO
H

Relationship between DNA, RNA, and protein

- Genes are made of DNA, a **nucleic acid** made of monomers called nucleotides
- A gene is a unit of inheritance that codes for the amino acid sequence of a polypeptide (shown) or a functional RNA product (not shown)



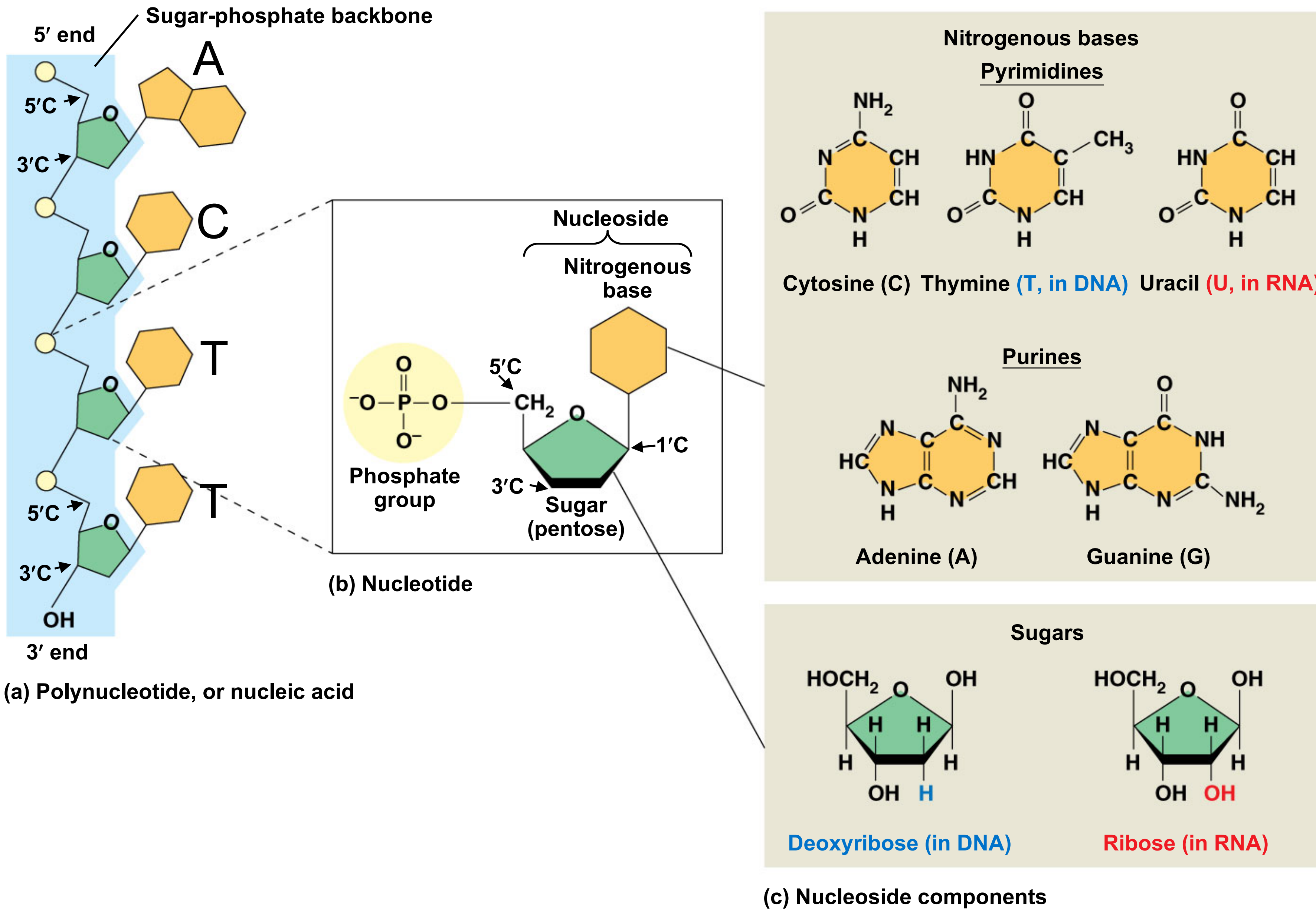
Role of Nucleic Acids

- Nucleic acids store, transmit, and help express hereditary information
- There are two types of nucleic acids
  - **Deoxyribonucleic acid (DNA)**
  - **Ribonucleic acid (RNA)**
- DNA provides directions for its own replication
- DNA directs synthesis of messenger RNA (mRNA) and, through mRNA, controls protein synthesis

The Components of Nucleic Acids

- Nucleic acids are polymers called **polynucleotides**
- Each polynucleotide is made of monomers called **nucleotides**
- Each nucleotide consists of a nitrogenous base, a pentose sugar, and one or more phosphate groups
- The portion of a nucleotide without the phosphate group is called a nucleoside

Figure 5.26



Sugar-phosphate backbone
5′ end
5′C
3′C
A
C
T
T
O
OH
3′ end
(a) Polynucleotide, or nucleic acid
Nucleoside
Nitrogenous base
5′C
CH2
O
1′C
3′C
Phosphate group
Sugar (pentose)
(b) Nucleotide
Nitrogenous bases
Pyrimidines
Cytosine (C)
Thymine (T, in DNA)
Uracil (U, in RNA)
Purines
Adenine (A)
Guanine (G)
Sugars
Deoxyribose (in DNA)
Ribose (in RNA)
(c) Nucleoside components

- Nucleoside = nitrogenous base + sugar
- There are two families of nitrogenous bases
  - **Pyrimidines** (cytosine, thymine, and uracil) have a single six-membered ring
  - **Purines** (adenine and guanine) have a six-membered ring fused to a five-membered ring
- In DNA, the sugar is **deoxyribose**; in RNA, the sugar is **ribose**
- Nucleotide = nucleoside + phosphate group

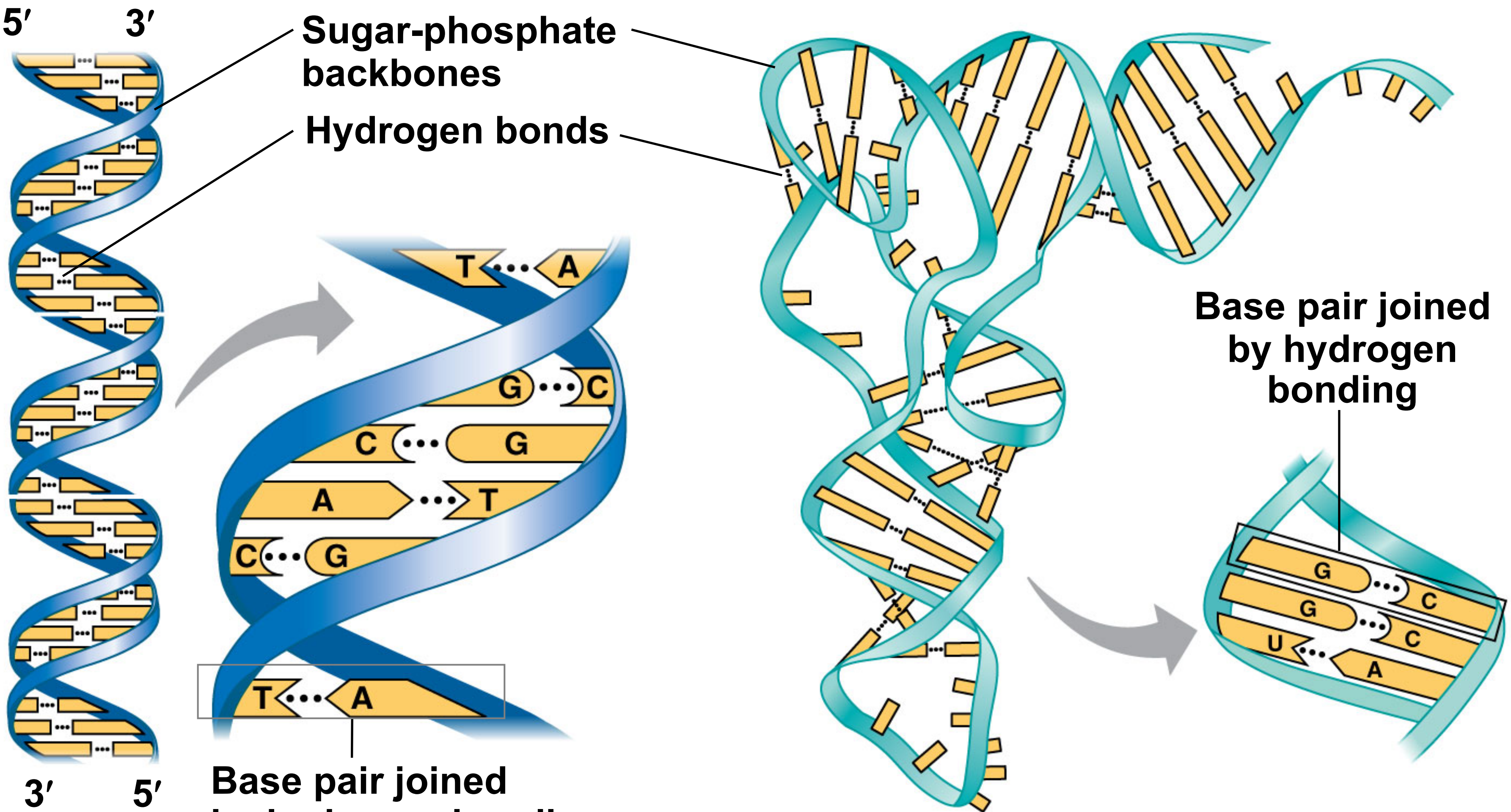
Nucleotide Polymers

- Nucleotide polymers are linked together to build a polynucleotide
- Adjacent nucleotides are joined by covalent bonds that form between the —OH group on the 3′ carbon of one nucleotide and the phosphate on the 5′ carbon on the next
- These links create a backbone of sugar-phosphate units with nitrogenous bases as appendages
- The sequence of bases along a DNA or mRNA polymer is unique for each gene

The Structures of DNA and RNA Molecules

- RNA molecules usually exist as single polypeptide chains
- DNA molecules have two polynucleotides spiraling around an imaginary axis, forming a **double helix**
- In the DNA double helix, the two backbones run in opposite $5' \rightarrow 3'$ directions from each other, an arrangement referred to as **antiparallel**
- One DNA molecule includes many genes

- **Complementary base pairing**
  - The nitrogenous bases in DNA pair up and form hydrogen bonds: adenine (A) always with thymine (T), and guanine (G) always with cytosine (C)
  - Complementary pairing can also occur between two RNA molecules or between parts of the same molecule
- In RNA, thymine is replaced by uracil (U) so A and U pair



(a) DNA

(b) Transfer RNA

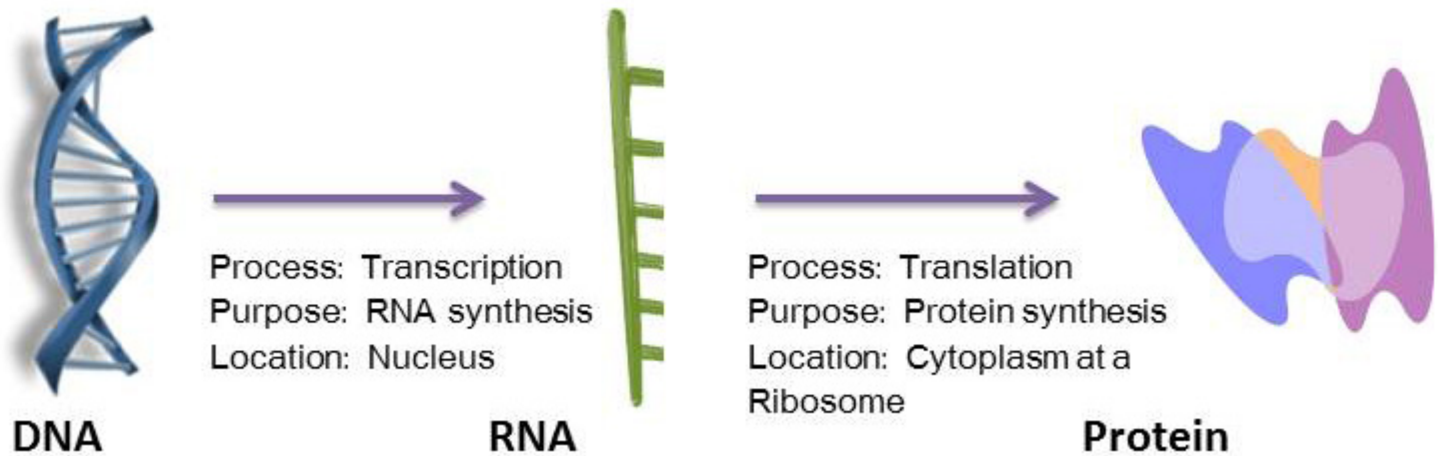
Genome sequencing

- The human genome project took over 13 years to complete and cost ~$3 billion (~$1 / base pair sequenced)
  - Sequence assembly was one of the first bioinformatics challenges

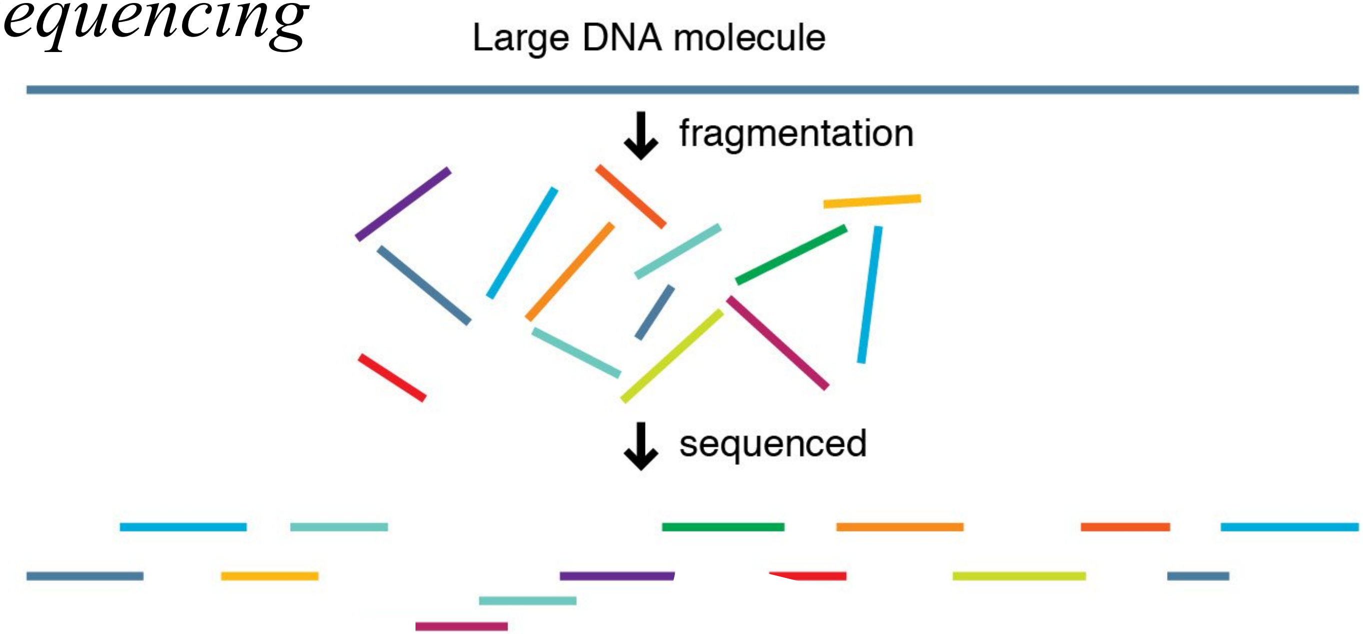
The genomic revolution

- The $1000 genome arrived in 2014
- https://www.forbes.com/sites/matthewherper/2014/01/14/the-1000-genome-arrives-for-real-this-time/
  - Sequencing machines cost $10 million
    - Can sequence 18,000 genomes / year
- Implications of cheap genomic sequencing
  - http://www.ted.com/talks/richard_resnick_welcome_to_the_genomic_revolution
  - What are they????

Gene Expression



Genomic sequencing



http://knowgenetics.org

TAGACGTAGC

GAATAGCTAG

GTCGAGCGTA

CCTCATAAGA

CGAGAATAGC

....

- ~ 1 billion reads
- Each read is ~ 100 bp

Genome assembly when a reference genome is available

Reference Genome Sequence (~3 billion bp for humans)

-----ACGTCGAGCGTAGACGTAGCGAGAATAGCTAGCTATAAAGGCCTCGTAAGA---

Align fragments to reference genome; must allow for variation

TAGACGTAGC

GAATAGCTAG

GTCGAGCGTA

CCTCATAAGA

CGAGAATAGC

....

- ~ 1 billion reads
- Each read is ~ 100 bp

Genome assembly when a reference genome is available

Reference Genome Sequence (~3 billion bp for humans)

-----ACGTCGAGCGTAGACGTAGCGAGAATAGCTAGCTATAAAGGCCTCGTAAGA---

Align fragments to reference genome; must allow for variation

GTAGACGTAGCGAGAATAGCTAG

(Inferred sequence)

TAGACGTAGC

GAATAGCTAG

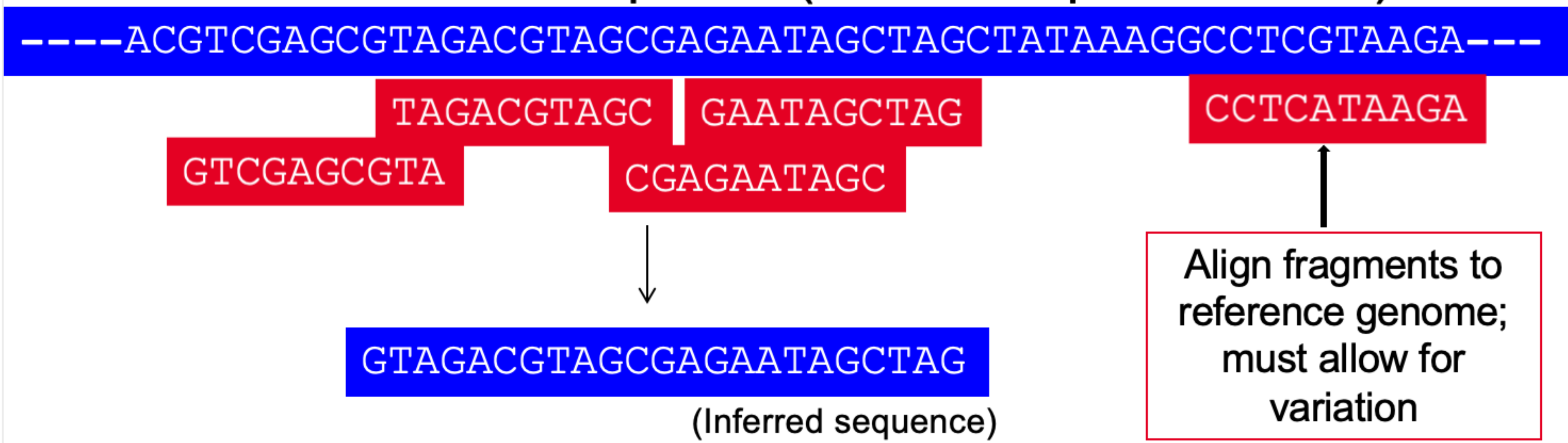
GTCGAGCGTA

CCTCATAAGA

CGAGAATAGC

Genome assembly when a reference genome is available

Reference Genome Sequence (~3 billion bp for humans)



De novo sequence assembly

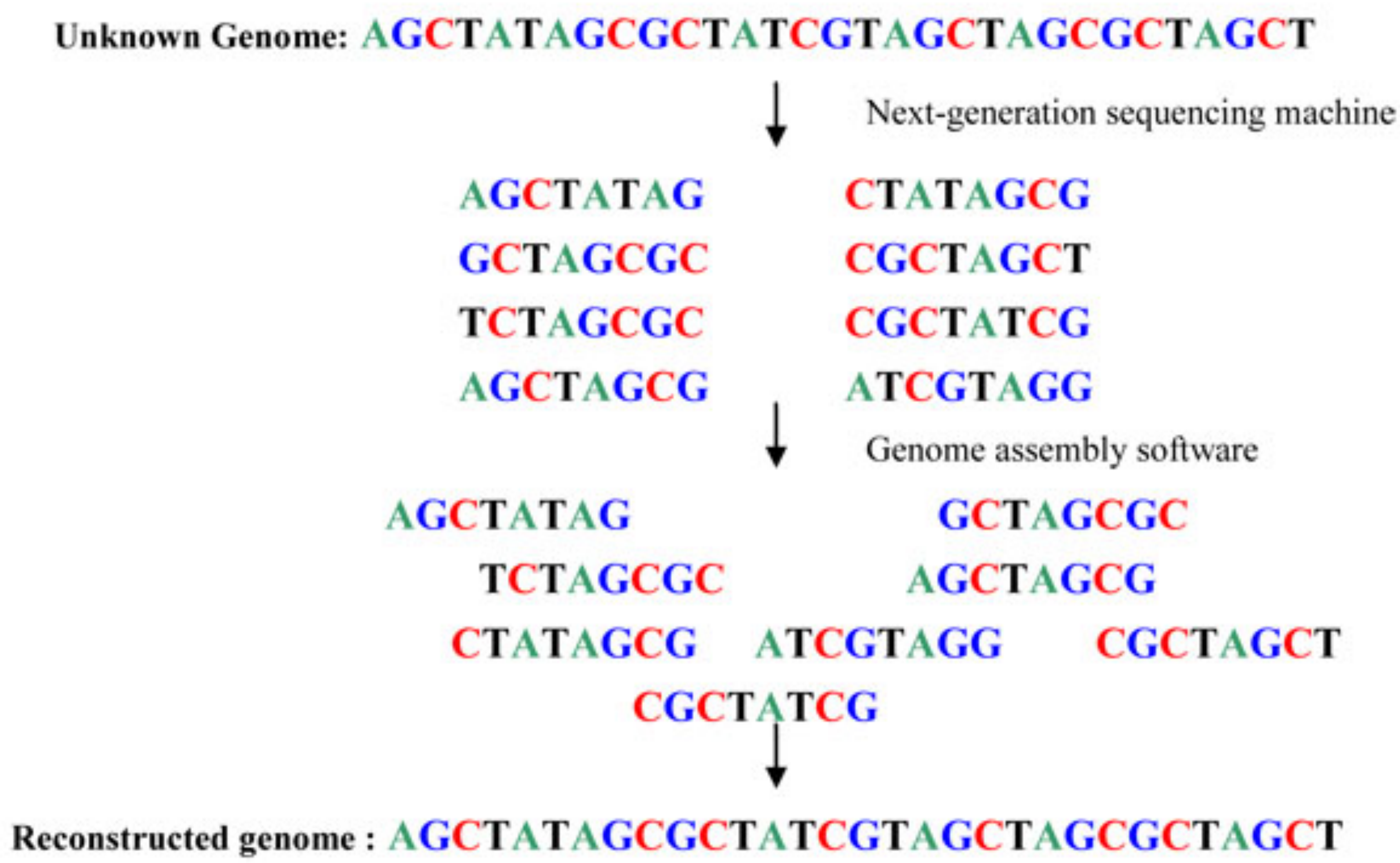
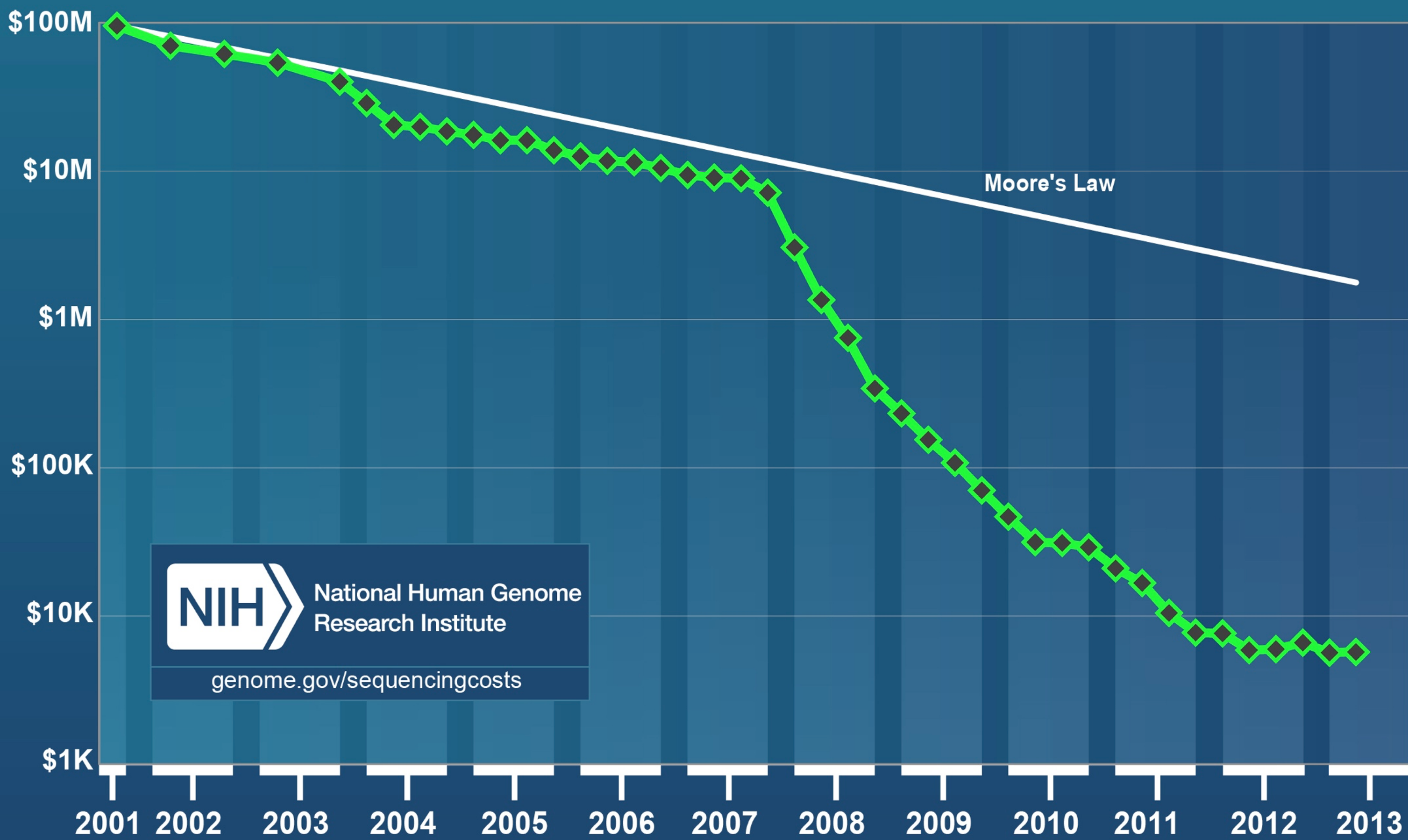


Figure 1. Workflow of discovering the genome of a species

De novo assembly when a reference genome is not available

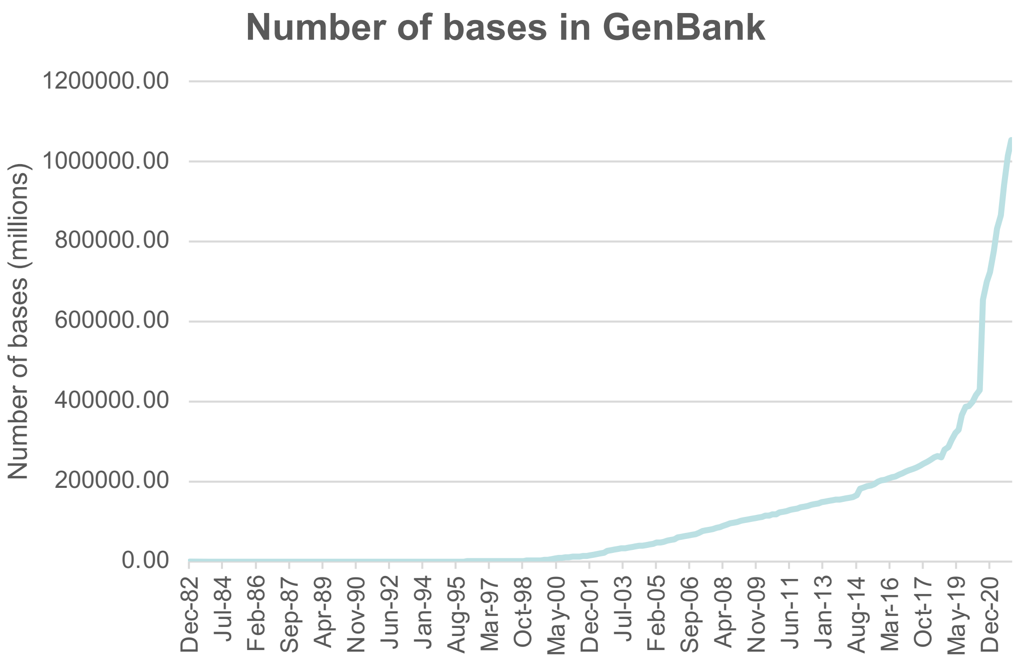
Cost per Genome



The number of DNA nucleotides sequenced has grown exponentially

Genbank statistics (December 2021)

- 1+ trillion bases in nucleotide database
- 14+ trillion additional bases processed for whole genome shotgun sequencing projects



https://www.ncbi.nlm.nih.gov/genbank/statistics/