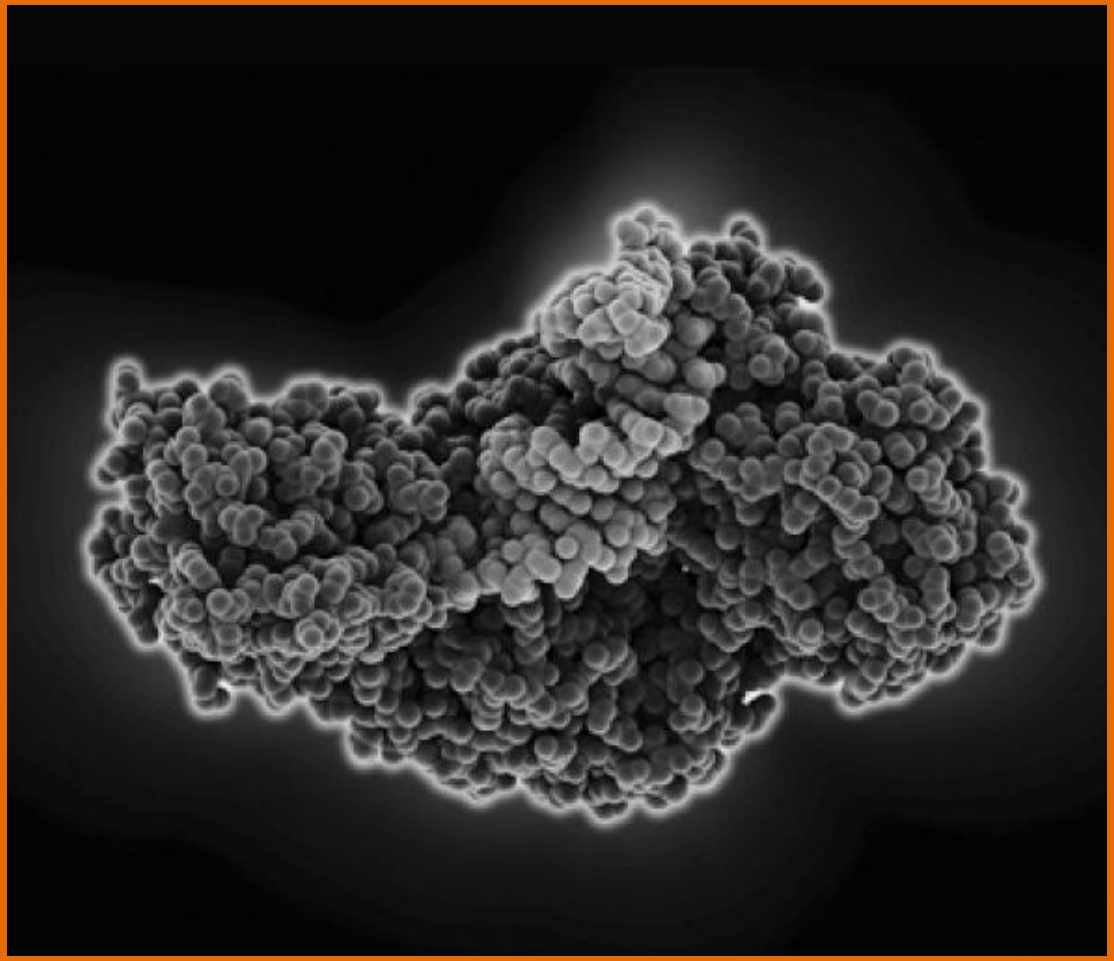
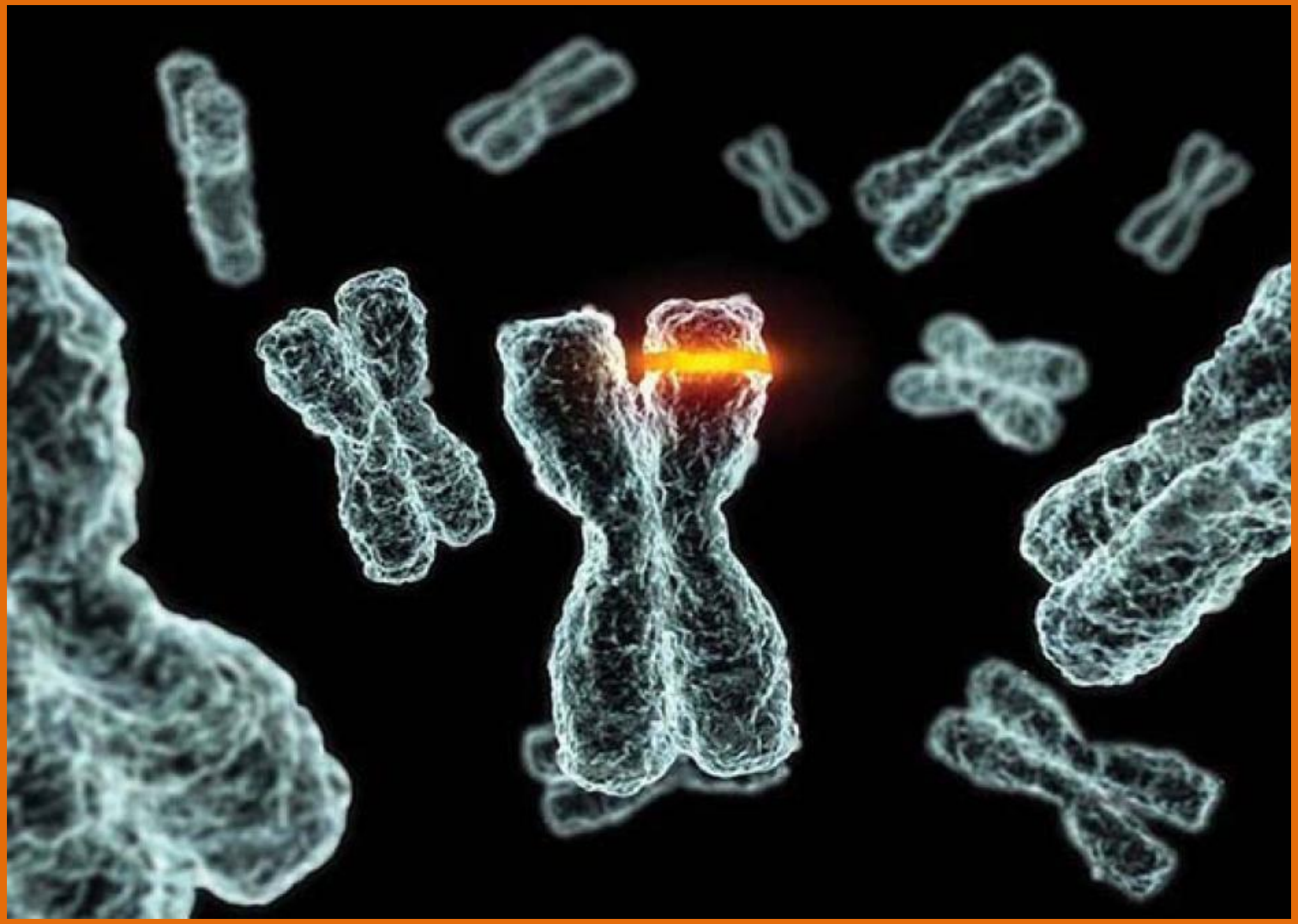
ΤΟ ΠΕΡΙΕΧΟΜΕΝΟ ΤΟΥ ΓΟΝΙΔΙΩΜΑΤΟΣ



Chapter 4: The Content of the Genome



4.1 Introduction

- The **genome** is the complete set of genes of an organism. Ultimately, it is defined by the complete DNA sequence, although as a practical matter it might not be possible to identify every gene unequivocally solely on the basis of sequence.

- The **transcriptome** is the complete set of genes expressed under particular conditions. It is defined in terms of the set of RNA molecules present in a single cell type, a more complex assembly of cells, or a complete organism. Because some genes generate multiple messenger RNAs (mRNAs), the transcriptome is likely to be larger than the actual number of genes in the genome. The transcriptome includes noncoding RNAs such as transfer RNAs (tRNAs), ribosomal RNAs (rRNAs), microRNAs (miRNAs), and others (see the chapters titled *Noncoding RNA* and *Regulatory RNA*), as well as mRNAs.

4.1 Introduction

- The **proteome** is the complete set of polypeptides encoded by the whole genome or produced in any particular cell or tissue. It should correspond to the mRNAs in the transcriptome, although there can be differences of detail reflecting changes in the relative abundance or stabilities of mRNAs and proteins. There might also be posttranslational modifications to proteins that allow more than one protein to be produced from a single transcript (this is called *protein splicing;* see the *Catalytic RNA* chapter).

- Proteins can function independently or as part of multiprotein or multimolecular complexes, such as holoenzymes and metabolic pathways where enzymes are clustered together. The RNA polymerase holoenzyme (see the *Prokaryotic Transcription* chapter) and the spliceosome (see the *RNA Splicing and Processing* chapter) are two examples. If we could identify all protein–protein interactions, we could define the total number of independent complexes of proteins. This is sometimes referred to as the **interactome**.

4.2 Genome Mapping Reveals That Individual Genomes Show Extensive Variation

- Genomes are mapped by sequencing their DNA and identifying functional genes.
- Polymorphism can be detected at the phenotypic level when a sequence affects gene function, at the restriction fragment level when it affects a restriction enzyme target site, and at the sequence level by direct analysis of DNA.
- The alleles of a gene show extensive polymorphism at the sequence level, but many sequence changes do not affect function.

4.3 SNPs Can Be Associated with Genetic Disorders

- Through genome-wide association studies, researchers can identify SNPs that are more frequently found in patients with a particular disorder.

FIGURE 4.3 shows that a functional gene should consist of a series of exons in which the first exon (containing an initiation codon) immediately follows a promoter, the internal exons are flanked by appropriate splicing junctions, and the last exon has the termination codon and is followed by 3′ processing signals; therefore, a single ORF starting with an initiation codon and ending with a termination codon can be deduced by joining the exons together. Internal exons can be identified as ORFs flanked by splicing junctions. In the simplest cases, the first and last exons contain the beginning and end of the coding region, respectively (as well as the 5′ and 3′ untranslated regions). In more complex cases, the first or last exons might have only untranslated regions and can therefore be more difficult to identify.

A hypothetical example is shown in **FIGURE 4.1**. This shows the basic approach of a **genome-wide association study (GWAS)** in which entire genomes of both patients and nonpatients are scanned for SNPs (see the chapter titled *Methods in Molecular Biology and Genetic Engineering*) and those SNPs that are associated with the disorder are identified. The disorder does not need to be determined by a single gene; it can be a polygenic or multifactorial (with nongenetic influences) disorder, as well. Although some associated SNPs might have no functional relevance to the disorder, others might.

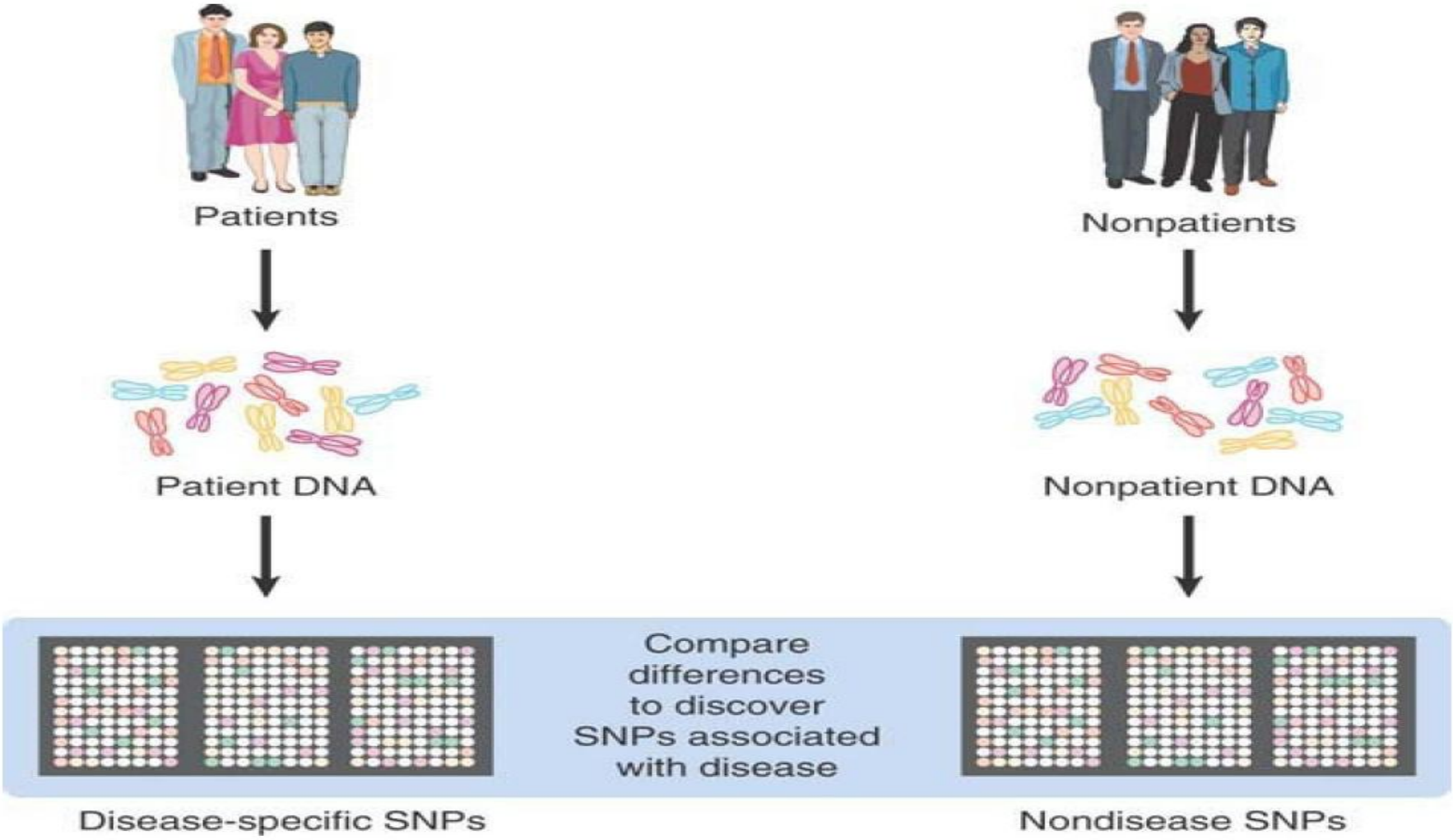


FIGURE 4.1 In a genome-wide association study, both patients and nonpatient controls for a particular disorder (such as heart disease, schizophrenia, or a single-gene disorder) are screened for SNPs across their genomes. Those SNPs that are statistically more frequently found in patients than in nonpatients can be identified.

4.4 Eukaryotic Genomes Contain Nonrepetitive and Repetitive DNA Sequences

- The kinetics of DNA reassociation after a genome has been denatured distinguish sequences by their frequency of repetition in the genome.
- Polypeptides are generally encoded by sequences in nonrepetitive DNA.
- Larger genomes within a taxonomic group do not contain more genes but have large amounts of repetitive DNA.
- A large part of moderately repetitive DNA can be made up of transposons.

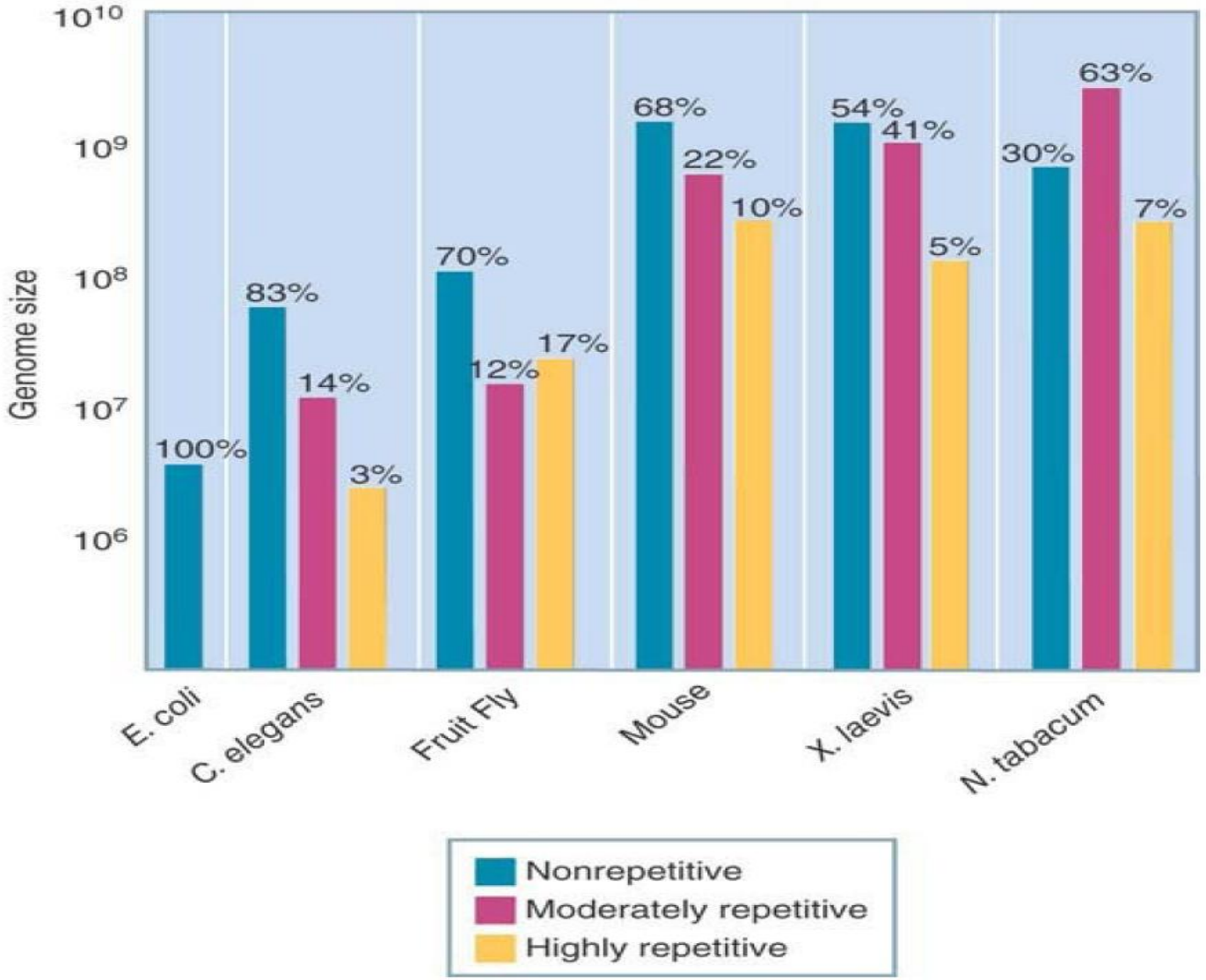


FIGURE 4.2 The proportions of different sequence components vary in eukaryotic genomes. The absolute content of nonrepetitive DNA increases with genome size but reaches a plateau at about *2* × 10^9 bp.

4.5 Eukaryotic Protein-Coding Genes Can Be Identified by the Conservation of Exons and of Genome Organization

- Researchers can use the conservation of exons as the basis for identifying coding regions as sequences that are present in multiple organisms.
- Methods for identifying functional genes are not perfect and many corrections must be made to preliminary estimates.
- Pseudogenes must be distinguished from functional genes.
- There are extensive syntenic relationships between the mouse and human genomes, and most functional genes are in a syntenic region.

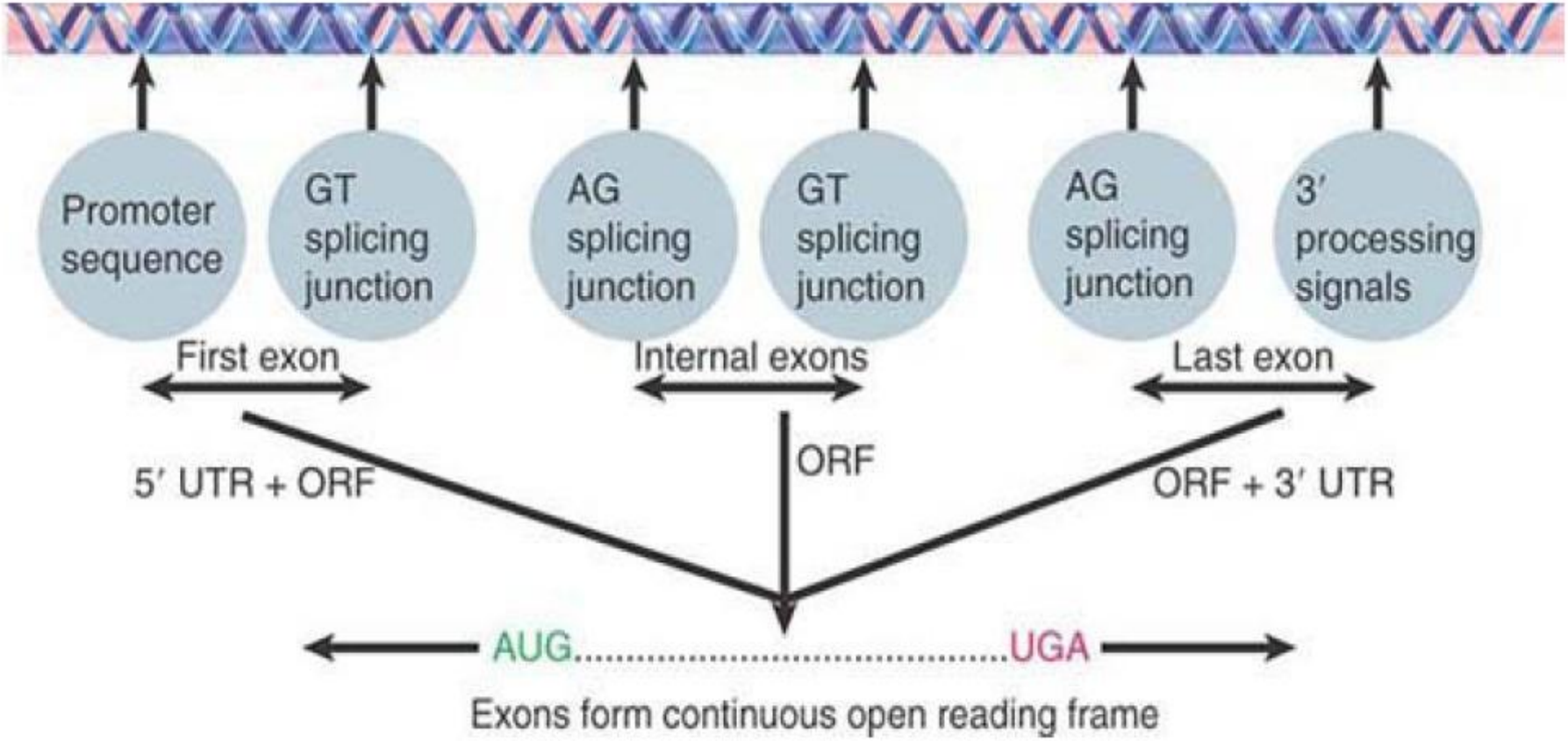


FIGURE 4.3 Exons of protein-coding genes are identified as coding sequences flanked by appropriate signals (with untranslated regions at both ends). The series of exons must generate an ORF with appropriate initiation and termination codons.

FIGURE 4.4 shows the relationship between mouse chromosome 1 and the human chromosomal set. Twenty-one segments in this mouse chromosome that have syntenic counterparts in human chromosomes have been identified. The extent of reshuffling that has occurred between the genomes is shown by the fact that the segments are spread among six different human chromosomes.

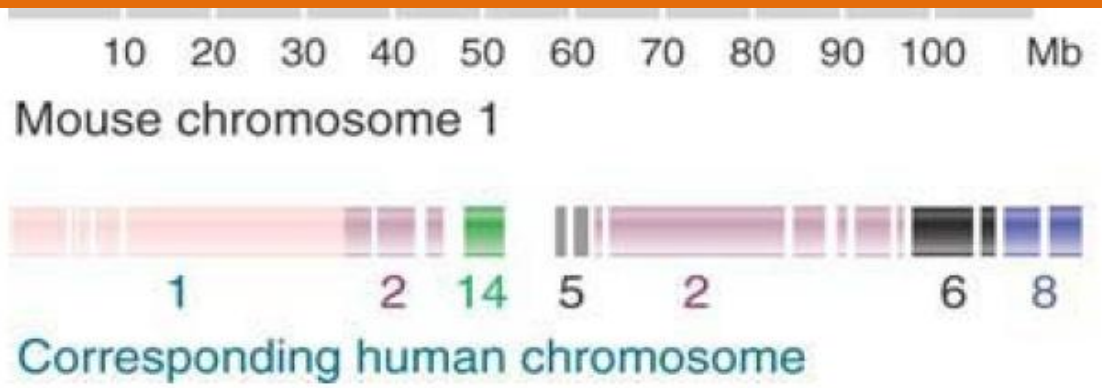


FIGURE 4.4 Mouse chromosome 1 has 21 segments between 1 and 25 Mb in length that are syntenic with regions corresponding to parts of six human chromosomes.

4.6 Some Eukaryotic Organelles Have DNA

- Mitochondria and chloroplasts have genomes that show non-Mendelian inheritance. Typically they are maternally inherited.
- Organelle genomes can undergo somatic segregation in plants.
- Comparisons of human mitochondrial DNA suggest that it is descended from a single population that existed approximately 200,000 years ago in Africa.

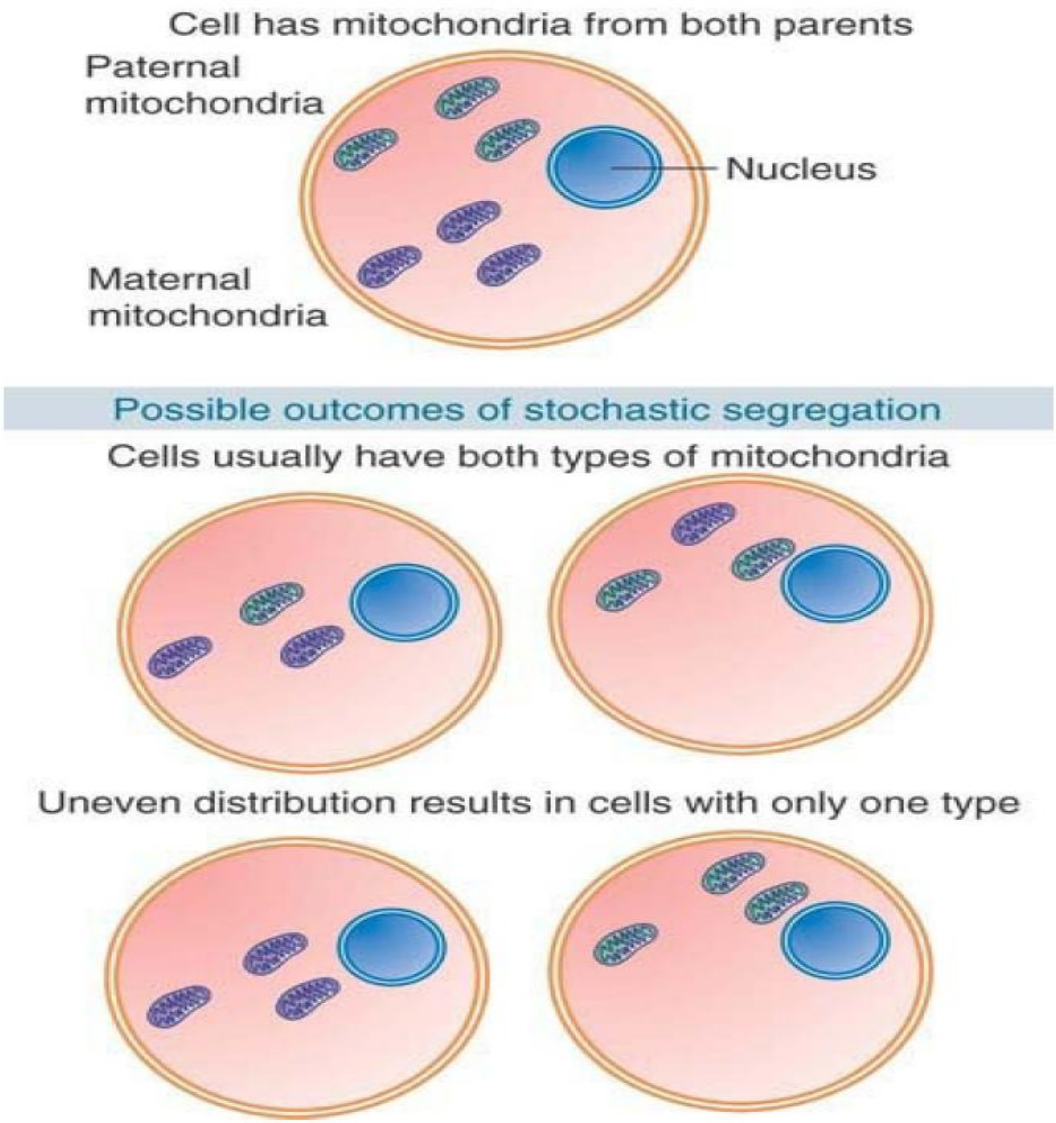


FIGURE 4.5 When mitochondria are inherited from both parents and paternal and maternal mitochondrial alleles differ, a cell has two sets of mitochondrial DNAs. Mitosis usually generates daughter cells with both sets. Somatic variation can result if unequal segregation generates daughter cells with only one set.

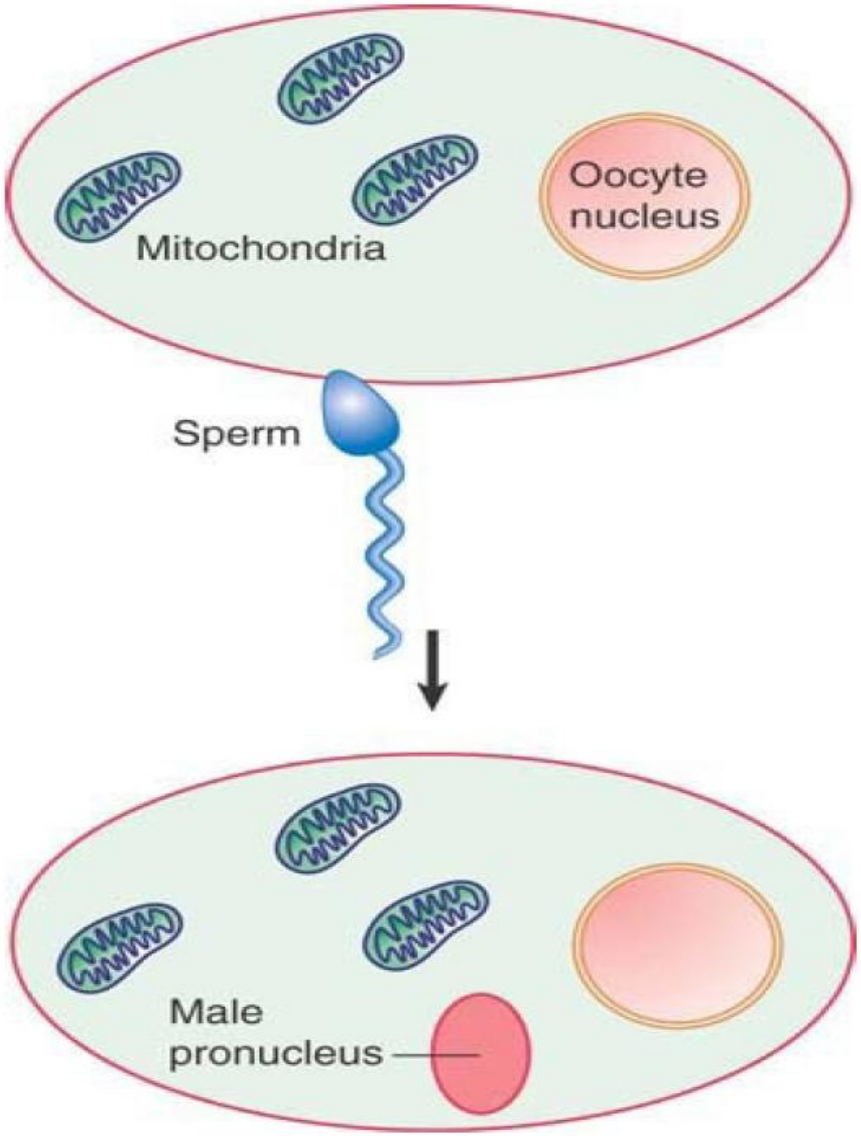


FIGURE 4.6 In animals, DNA from the sperm enters the oocyte to form the male pronucleus in the fertilized egg, but all the mitochondria are provided by the oocyte.

4.7 Organelle Genomes Are Circular DNAs That Encode Organelle Proteins

- Organelle genomes are usually (but not always) circular molecules of DNA.
- Organelle genomes encode some, but not all, of the proteins used in the organelle.
- Animal cell mitochondrial DNA is extremely compact and typically encodes 13 proteins, 2 rRNAs, and 22 tRNAs.
- Yeast mitochondrial DNA is five times longer than animal cell mtDNA because of the presence of long introns.

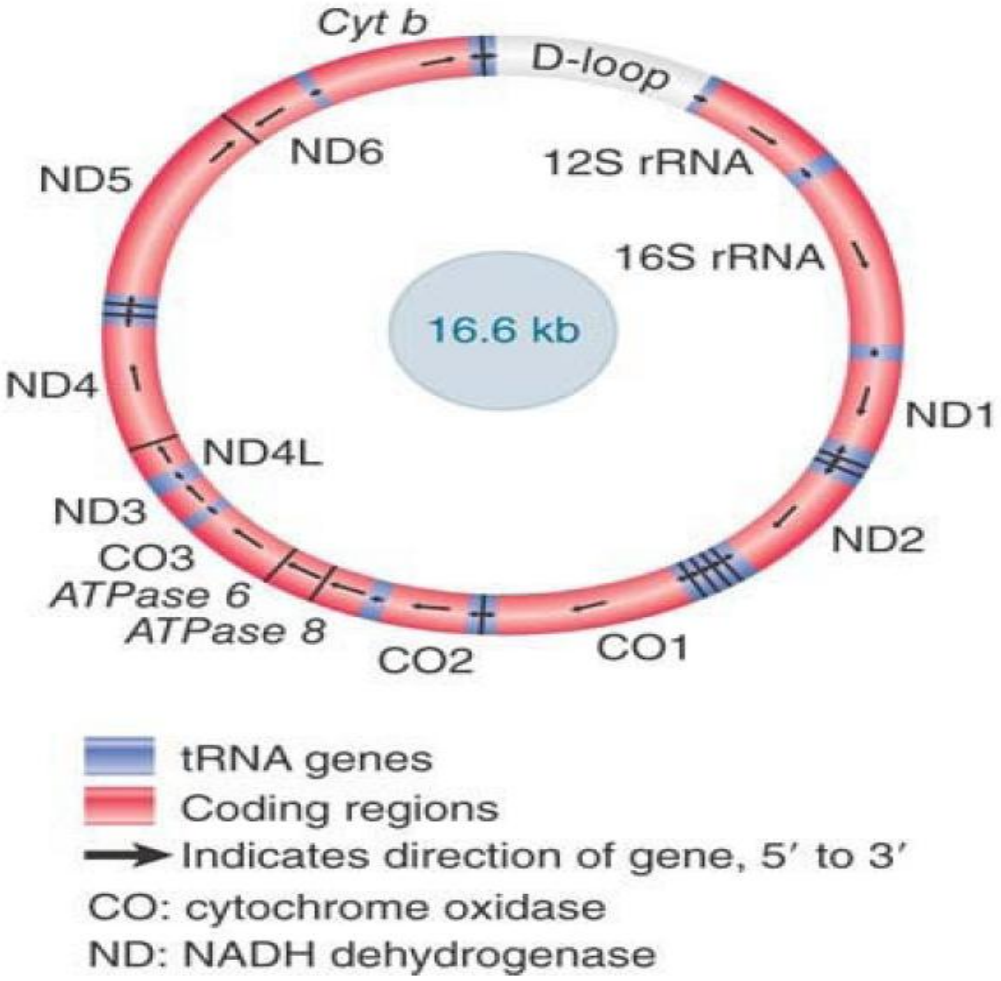


FIGURE 4.7 Human mitochondrial DNA has 22 tRNA genes, 2 rRNA genes, and 13 protein-coding regions. Fourteen of the 15 protein-coding and rRNA-coding regions are transcribed in the same direction. Fourteen of the tRNA genes are expressed in the clockwise direction and 8 are read counterclockwise.

TABLE 4.1 Mitochondrial genomes have genes encoding (mostly complex I–IV) proteins, rRNAs, and tRNAs.

Species	Size (kb)	Protein-Coding Genes	RNA-Coding Genes
Fungi	19–100	8–14	10–28
Protists	6–100	3–62	2–29
Plants	186–366	27–34	21–30
Animals	16–17	13	4–24

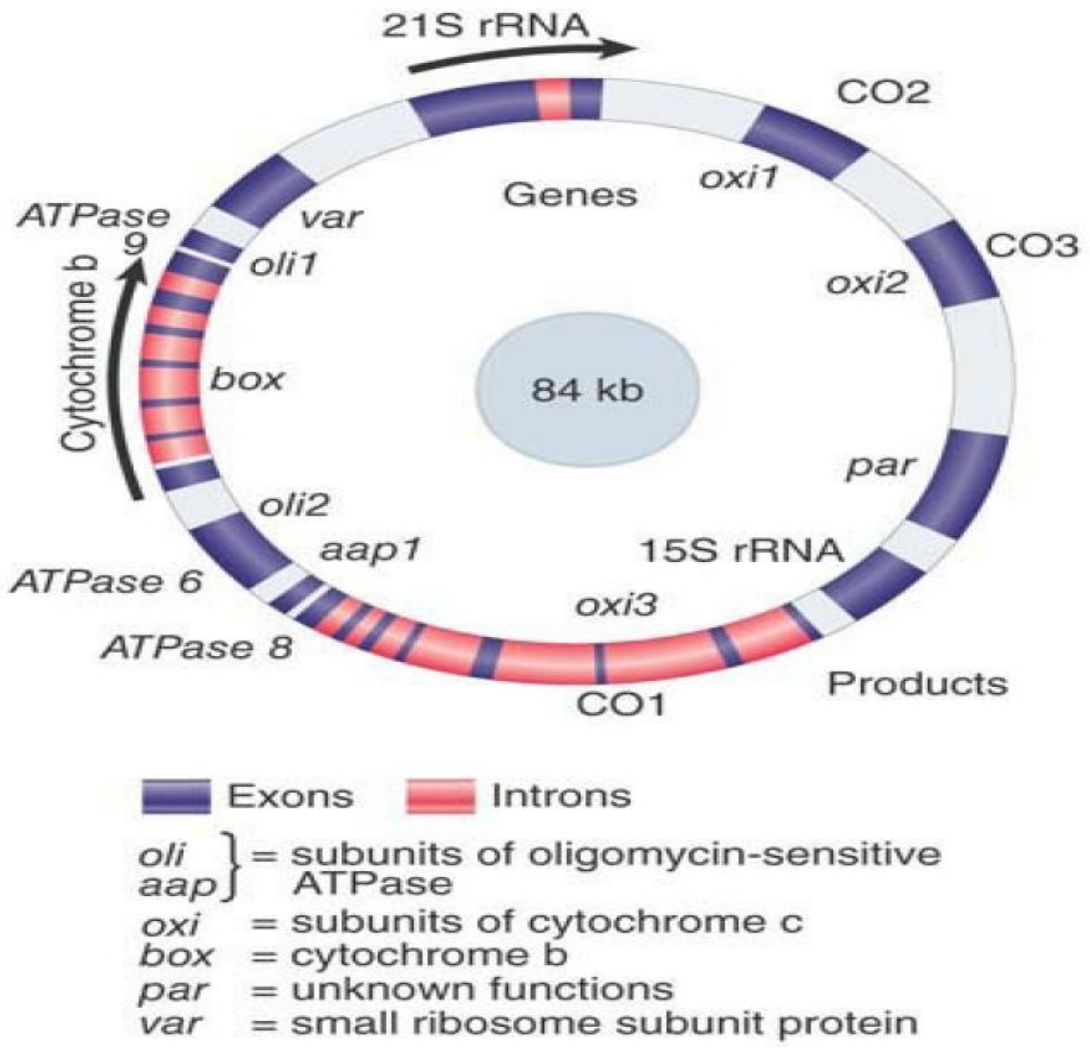


FIGURE 4.8 The mitochondrial genome of *S. cerevisiae* contains both interrupted and uninterrupted protein-coding genes, rRNA genes, and tRNA genes (positions not indicated). Arrows indicate direction of transcription.

4.8 The Chloroplast Genome Encodes Many Proteins and RNAs

- Chloroplast genomes vary in size, but are large enough to encode 50 to 100 proteins as well as the rRNAs and tRNAs.

What genes are carried by chloroplasts? Chloroplast DNAs vary in length from about 120 to 217 kb (the largest in geranium). The sequenced chloroplast genomes (more than 200 in total) have 87 to 183 genes. **TABLE 4.2** summarizes the functions encoded by the chloroplast genome in land plants. There is more variation in the chloroplast genomes of algae.

TABLE 4.2 The chloroplast genome in land plants encodes 4 rRNAs, 30 tRNAs, and about 60 proteins.

Genes	Types
RNA coding	
16S rRNA	1
23S rRNA	1
4.5S rRNA	1
5S rRNA	1
tRNA	30–32
Gene expression	
Proteins	20–21
RNA polymerase	3
Others	2
Chloroplast functions	
Rubisco and thylakoids	31–32
NADH dehydrogenase	11
Total	105–113

4.9 Mitochondria and Chloroplasts Evolved by Endosymbiosis

- Both mitochondria and chloroplasts are descended from bacterial ancestors.
- Most of the genes of the mitochondrial and chloroplast genomes have been transferred to the nucleus during the organelle's evolution.

How is it that an organelle evolved so that it contains genetic information for some of its functions, whereas the information for other functions is encoded in the nucleus? **FIGURE 4.9** shows the endosymbiotic hypothesis for mitochondrial evolution, in which primitive cells captured bacteria that provided the function of cellular respiration and over time evolved into mitochondria. At first, the proto-organelle must have contained all of the genes needed to specify its functions. A similar mechanism has been proposed for the origin of chloroplasts.

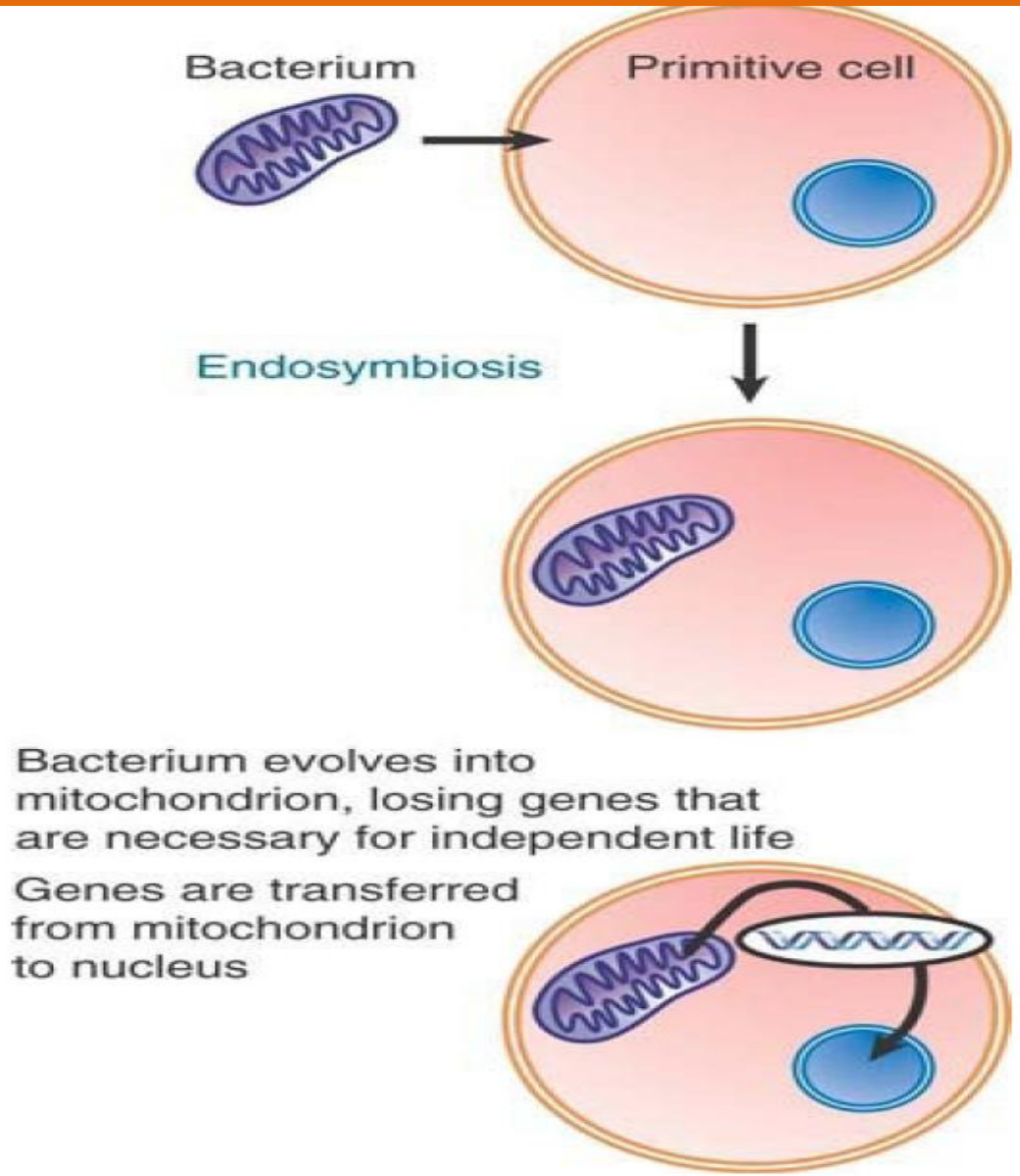


FIGURE 4.9 Mitochondria originated by an endosymbiotic event when a bacterium was captured by a eukaryotic cell.