Molecular biology cheat sheet

Nucleotides A, C, G, T

gene

DNA

...TACCTACATCCACTCATC...AGCTACGTTCCCCGACTACGACATGGTGATT

5' ...ATGGATGTAGGTGAGTAG...TCGATGCAAGGGGCTGATGCTGTACCACTAA... 3'

exon intron exon

transcription

RNA

codon

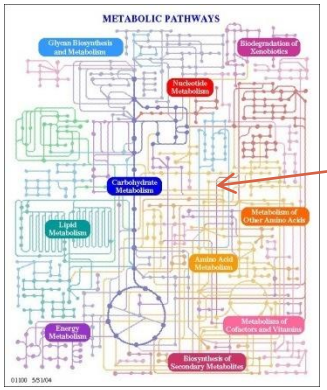
…AUGGAUGUAGAUGGGCUGAUGCUGUACCACUAA

translation

Protein

MDVDGLMLYH

Gene regulation



entsyme

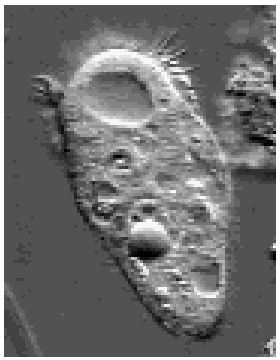
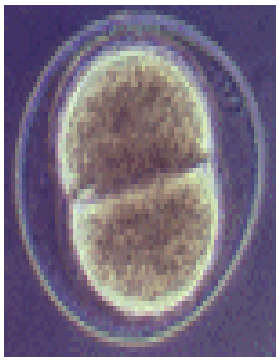
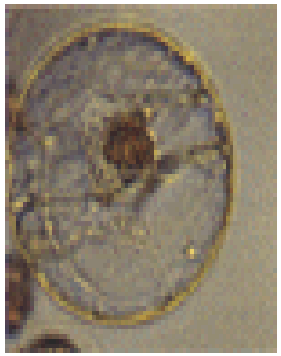
recombination

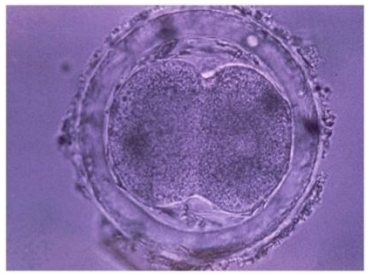
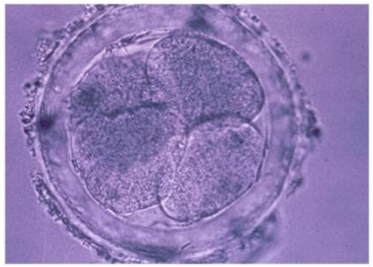
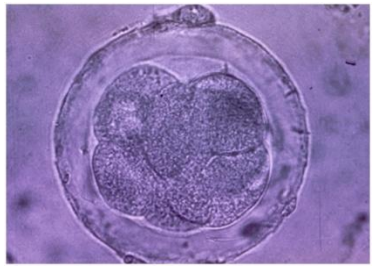
Mother DNA C G A A

Father DNA T C C T

Daughter DNA C C C A

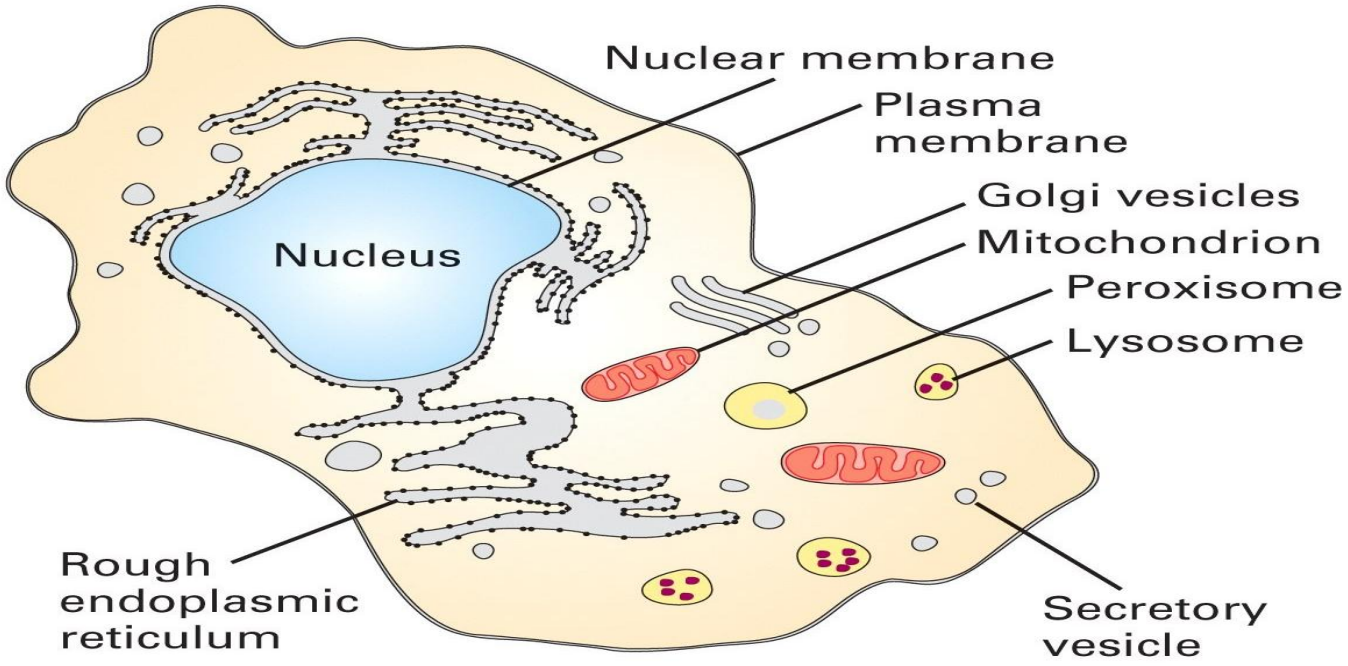
Molecular biology primer

Molecular Biology Primer *by Angela Brooks, Raymond Brown, Calvin Chen, Mike Daly, Hoa Dinh, Erinn Hama, Robert Hinman, Julio Ng, Michael Sneddon, Hoa Troung, Jerry Wang, Che Fung Yung*
Edited for Introduction to Bioinformatics (Autumn 2007, Summer 2008, Autumn 2008) by Esa Pitkänen

Life begins with Cell

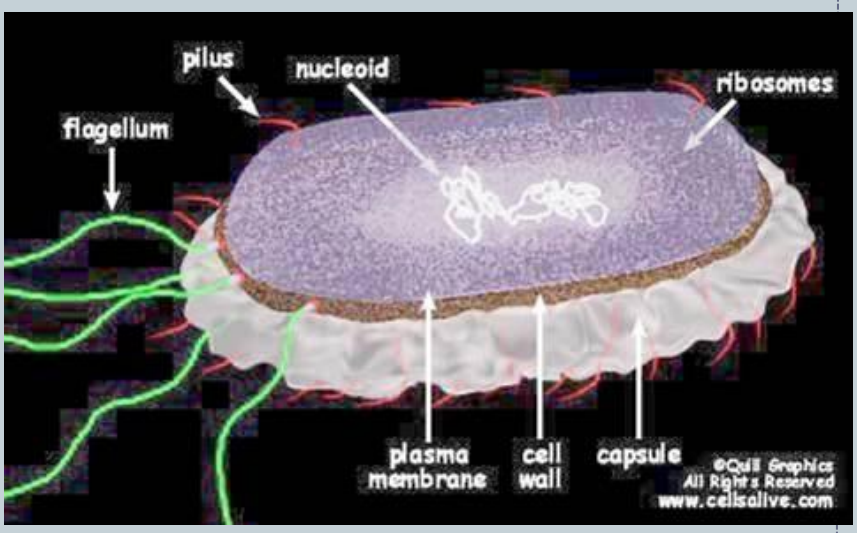


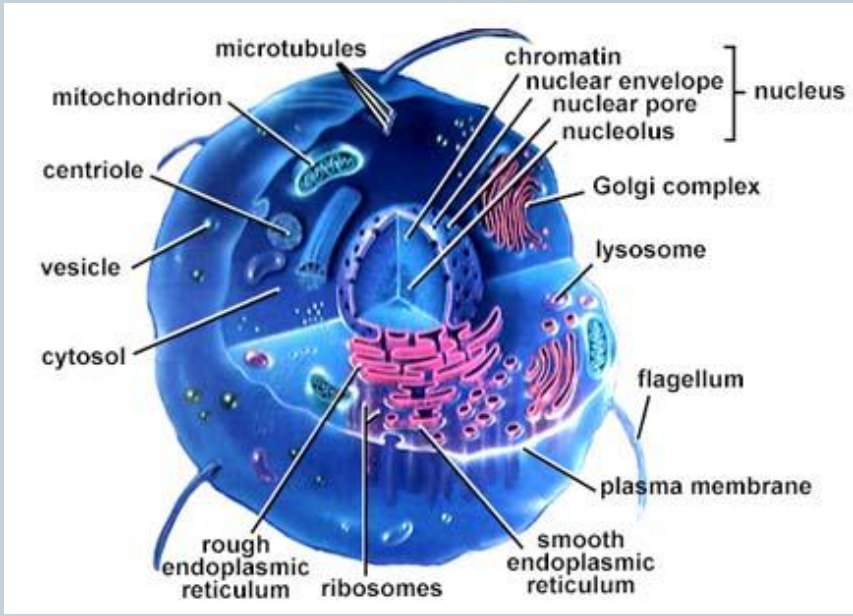
- A cell is a smallest structural unit of an organism that is capable of independent functioning
- All cells have some common features

Cells

- **Fundamental working units** of every living system.
- Every organism is composed of one of two radically different types of cells:
  - **prokaryotic** cells or
  - **eukaryotic** cells.
- **Prokaryotes** and **Eukaryotes** are descended from the same primitive cell.
  - All prokaryotic and eukaryotic cells are the result of a total of 3.5 billion years of evolution.

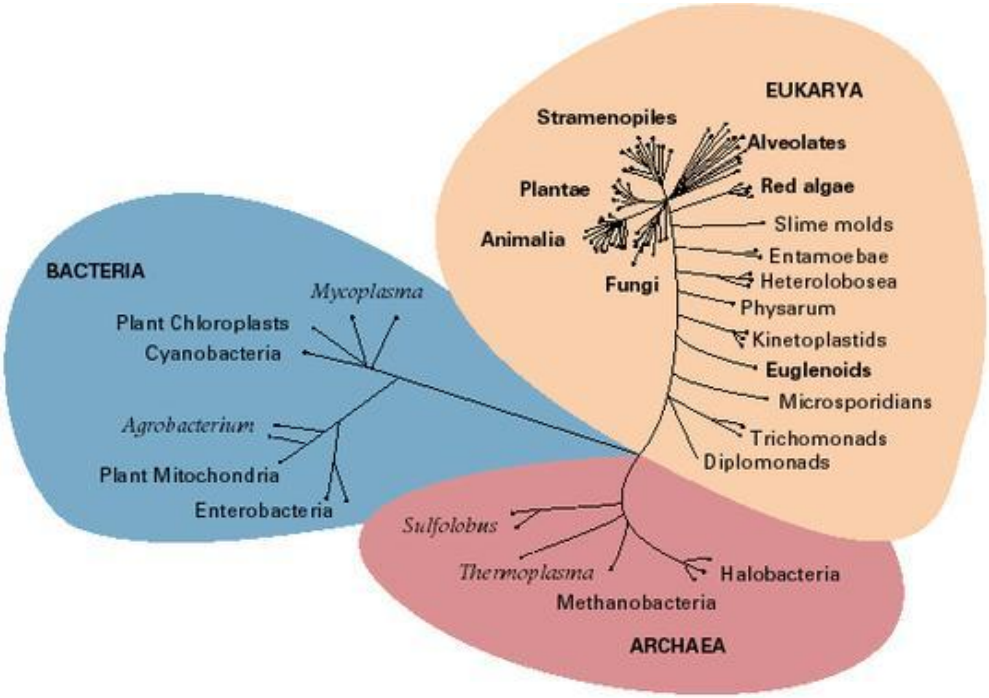
Two types of cells: Prokaryotes and Eukaryotes



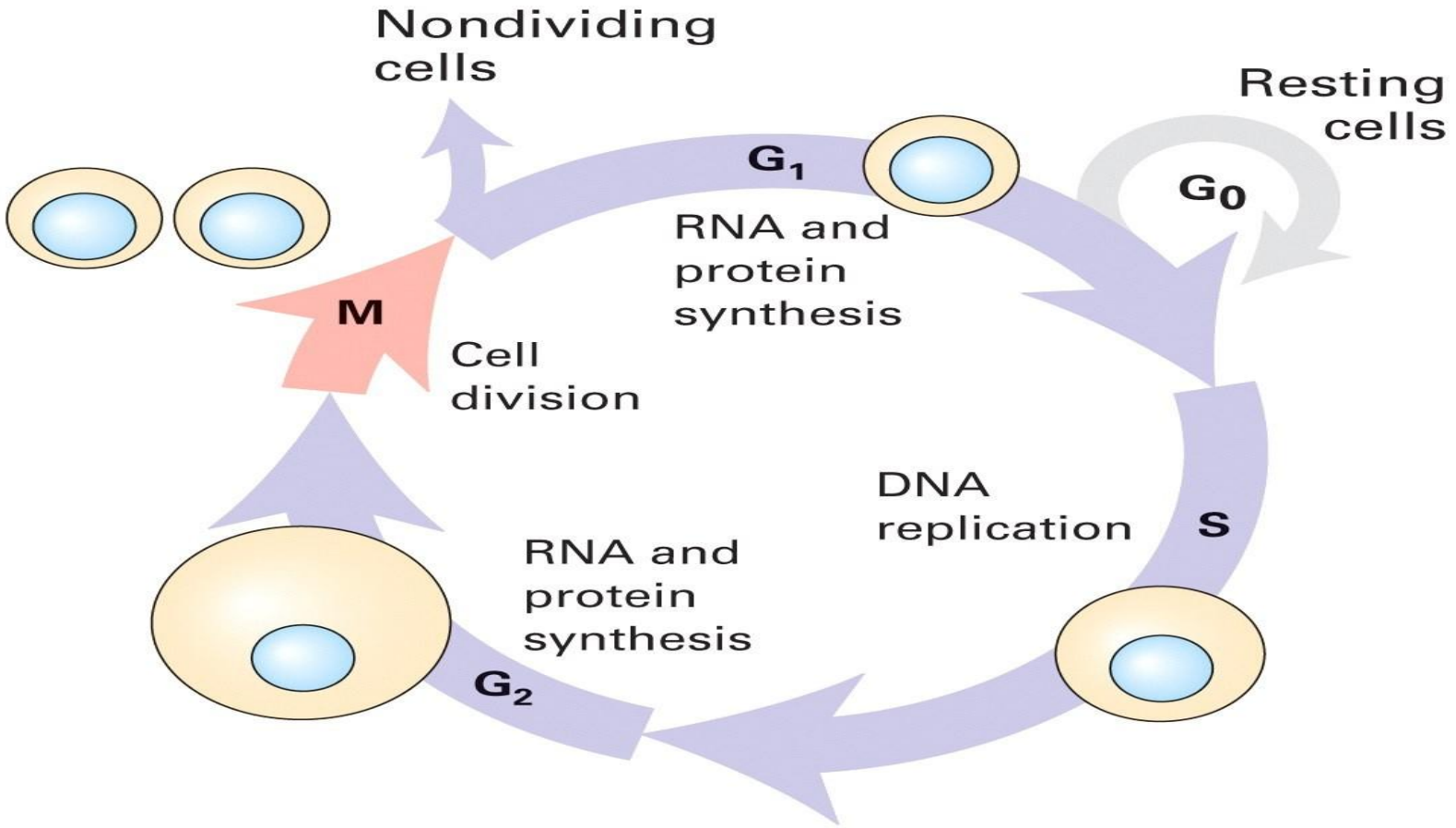


Prokaryotes and Eukaryotes

- According to the most recent evidence, there are three main branches to the tree of life
- Prokaryotes include Archaea (“ancient ones”) and bacteria
- Eukaryotes are kingdom Eukarya and includes plants, animals, fungi and certain algae



All Cells have common Cycles



- Born, eat, replicate, and die

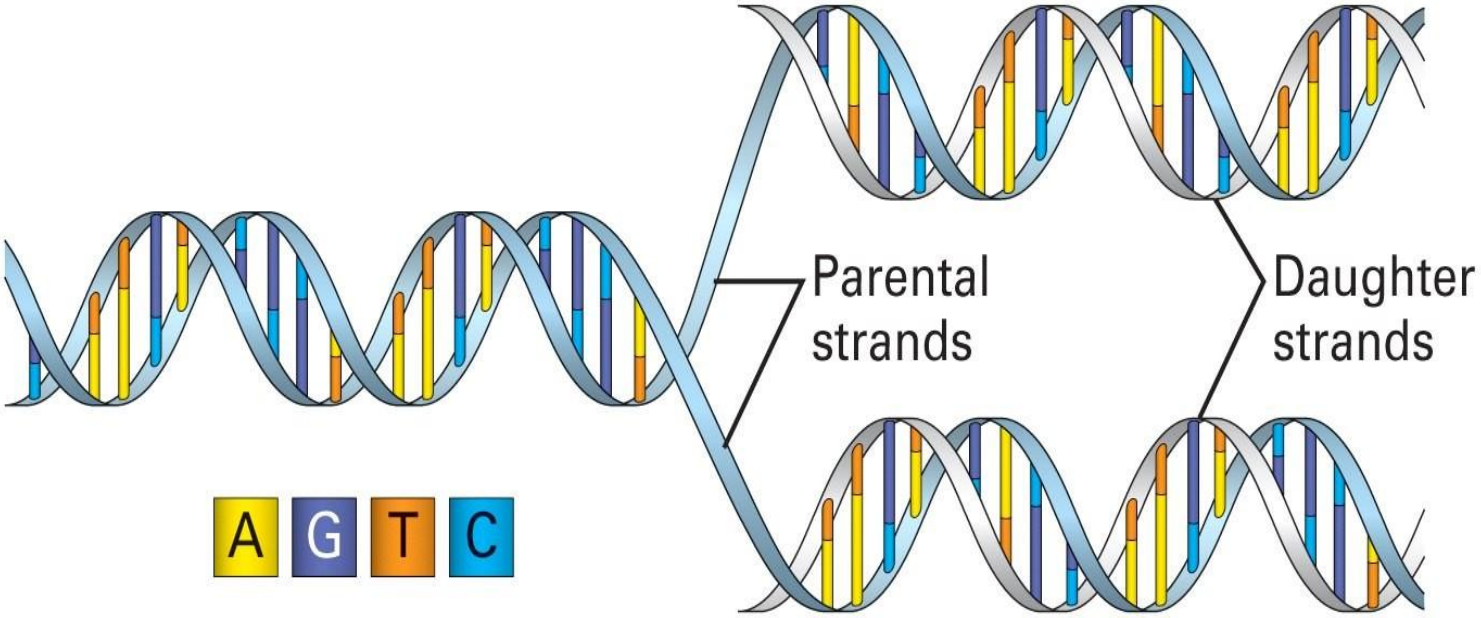
Common features of organisms

- Chemical energy is stored in ATP
- Genetic information is encoded by DNA
- Information is transcribed into RNA
- There is a **common triplet genetic code**
- Translation into proteins involves ribosomes
- Shared metabolic pathways
- Similar proteins among diverse groups of organisms

All Life depends on 3 critical molecules

- DNAs (Deoxyribonucleic acid)
  - Hold information on how cell works
- RNAs (Ribonucleic acid)
  - Act to transfer short pieces of information to different parts of cell
  - Provide templates to synthesize into protein
- Proteins
  - Form enzymes that send signals to other cells and regulate gene activity
  - Form body's major components (e.g. hair, skin, etc.)
  - "Workhorses" of the cell

DNA: The Code of Life

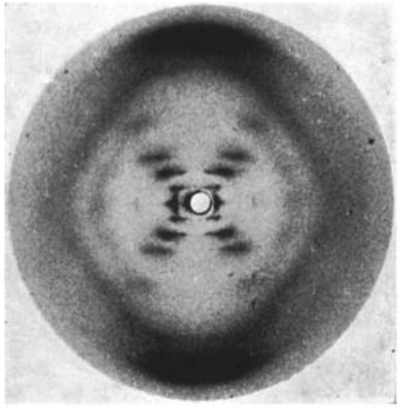


- The structure and the four genomic letters code for all living organisms
- Adenine, Guanine, Thymine, and Cytosine which pair A-T and C-G on complimentary strands.

Discovery of the structure of DNA

- **1952-1953** James D. Watson and Francis H. C. Crick deduced the double helical structure of DNA from X-ray diffraction images by Rosalind Franklin and data on amounts of nucleotides in DNA

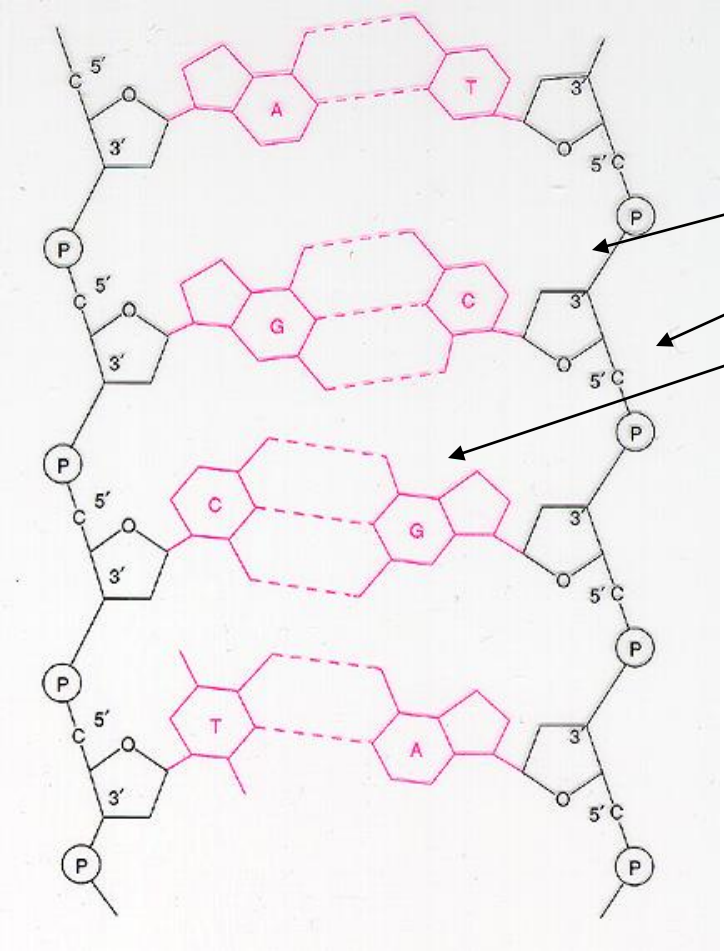
James Watson and Francis Crick



”Photo 51”

Rosalind Franklin

DNA, continued



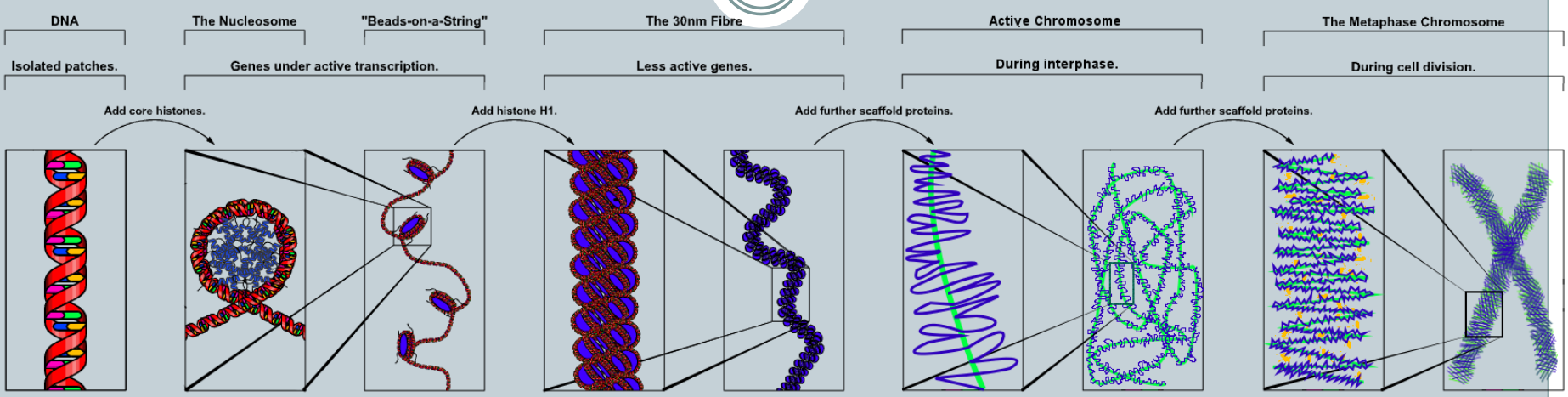
- DNA has a double helix structure which is composed of
  - sugar molecule
  - phosphate group
  - and a base (A,C,G,T)

- By convention, we read DNA strings in direction of transcription: from 5’ end to 3’ end

  5’ ATTTAGGCC 3’

  3’ TAAATCCGG 5’

DNA is contained in chromosomes



In eukaryotes, DNA is packed into *chromatids*

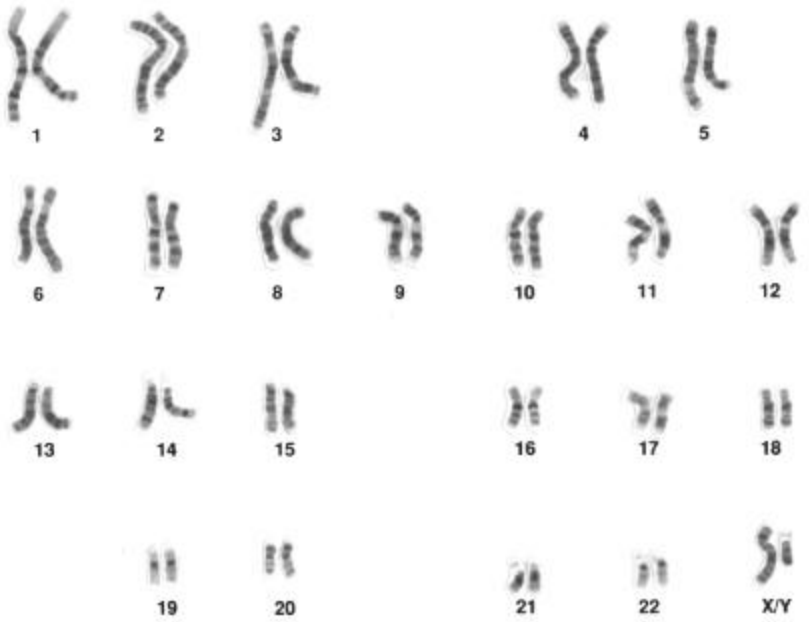
- In metaphase, the "X" structure consists of two identical chromatids

In prokaryotes, DNA is usually contained in a single, circular chromosome

http://en.wikipedia.org/wiki/Image:Chromatin_Structures.png

Human chromosomes

- Somatic cells in humans have 2 pairs of 22 chromosomes + XX (female) or XY (male) = total of 46 chromosomes
- Germline cells have 22 chromosomes + either X or Y = total of 23 chromosomes



Karyogram of human male using Giemsa staining (http://en.wikipedia.org/wiki/Karyotype)

Length of DNA and number of chromosomes

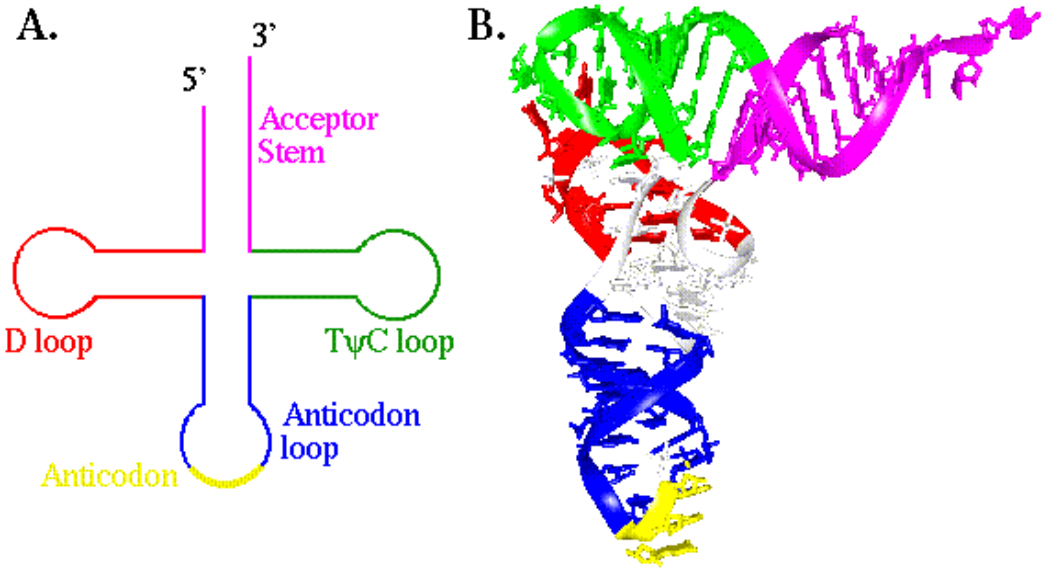
Organism	#base pairs	#chromosomes (germline)
Prokayotic		
Escherichia coli (bacterium)	$4x10^6$	1
Eukaryotic		
Saccharomyces cerevisia (yeast)	$1.35x10^7$	17
Drosophila melanogaster (insect)	$1.65x10^8$	4
Homo sapiens (human)	$2.9x10^9$	23
Zea mays (corn / maize)	$5.0x10^9$	10

Hepatitis delta virus, complete genome

```
   1 atgagccaag ttccgaacaa ggattcgcgg ggaggataga tcagcgcccg agaggggtga
  61 gtcggtaaag agcattggaa cgtcggagat acaactccca agaaggaaaa aagagaaagc
 121 aagaagcgga tgaatttccc cataacgcca gtgaaactct aggaagggga aagagggaag
 181 gtggaagaga aggaggcggg cctcccgatc cgaggggccc ggcggccaag tttggaggac
 241 actccggccc gaagggttga gagtacccca gagggaggaa gccacacgga gtagaacaga
 301 gaaatcacct ccagaggacc ccttcagcga acagagagcg catcgcgaga gggagtagac
 361 catagcgata ggaggggatg ctaggagttg ggggagaccg aagcgaggag gaaagcaaag
 421 agagcagcgg ggctagcagg tgggtgttcc gccccccgag aggggacgag tgaggcttat
 481 cccggggaac tcgacttatc gtccccacat agcagactcc cggacccccc ttcaaagtga
 541 ccgaggggggg tgactttgaa cattggggac cagtggagcc atgggatgct cctcccgatt
 601 ccgcccaagc tccttccccc caagggtcgc ccaggaatgg cgggacccca ctctgcaggg
 661 tccgcgttcc atcctttctt acctgatggc cggcatggtc ccagcctcct cgctggcgcc
 721 ggctgggcaa cattccgagg ggaccgtccc ctcggtaatg gcgaatggga cccacaaatc
 781 tctctagctt cccagagaga agcgagagaa aagtggctct cccttagcca tccgagtgga
 841 cgtgcgtcct ccttcggatg cccaggtcgg accgcgagga ggtggagatg ccatgccgac
 901 ccgaagagga aagaaggacg cgagacgcaa acctgcgagt ggaaacccgc tttattcact
 961 ggggtcgaca actctgggga gaggagggag ggtcggctgg gaagagtata tcctatggga
1021 atccctggct tcccccttatg tccagtccct ccccggtccg agtaaagggg gactccggga
1081 ctccttgcat gctggggacg aagccgcccc cgggcgctcc cctcgttcca ccttcgaggg
1141 ggttcacacc cccaacctgc gggccggcta ttcttctttc ccttctctcg tcttcctcgg
1201 tcaacctcct aagttcctct tcctcctcct tgctgaggtt ctttcccccc gccgatagct
1261 gctttctctt gttctcgagg gccttccttc gtcggtgatc ctgcctctcc ttgtcggtga
1321 atcctccccc ggaaggcctc ttcctaggtc cggagtctac ttccatctgg tccgttcggg
1381 ccctcttcgc cgggggagcc ccctctccat ccttatcttt ctttccgaga attcctttga
1441 tgtttcccag ccagggatgt tcatcctcaa gtttcttgat tttcttctta accttccgga
1501 ggtctctctc gagttcctct aacttctttc ttccgctcac ccactgctcg agaacctctt
1561 ctctcccccc gcggtttttc cttccttcgg gccggctcat cttcgactag aggcgacggt
1621 cctcagtact cttactcttt tctgtaaaga ggagactgct ggccctgtcg cccaagttcg
1681 ag
```

RNA

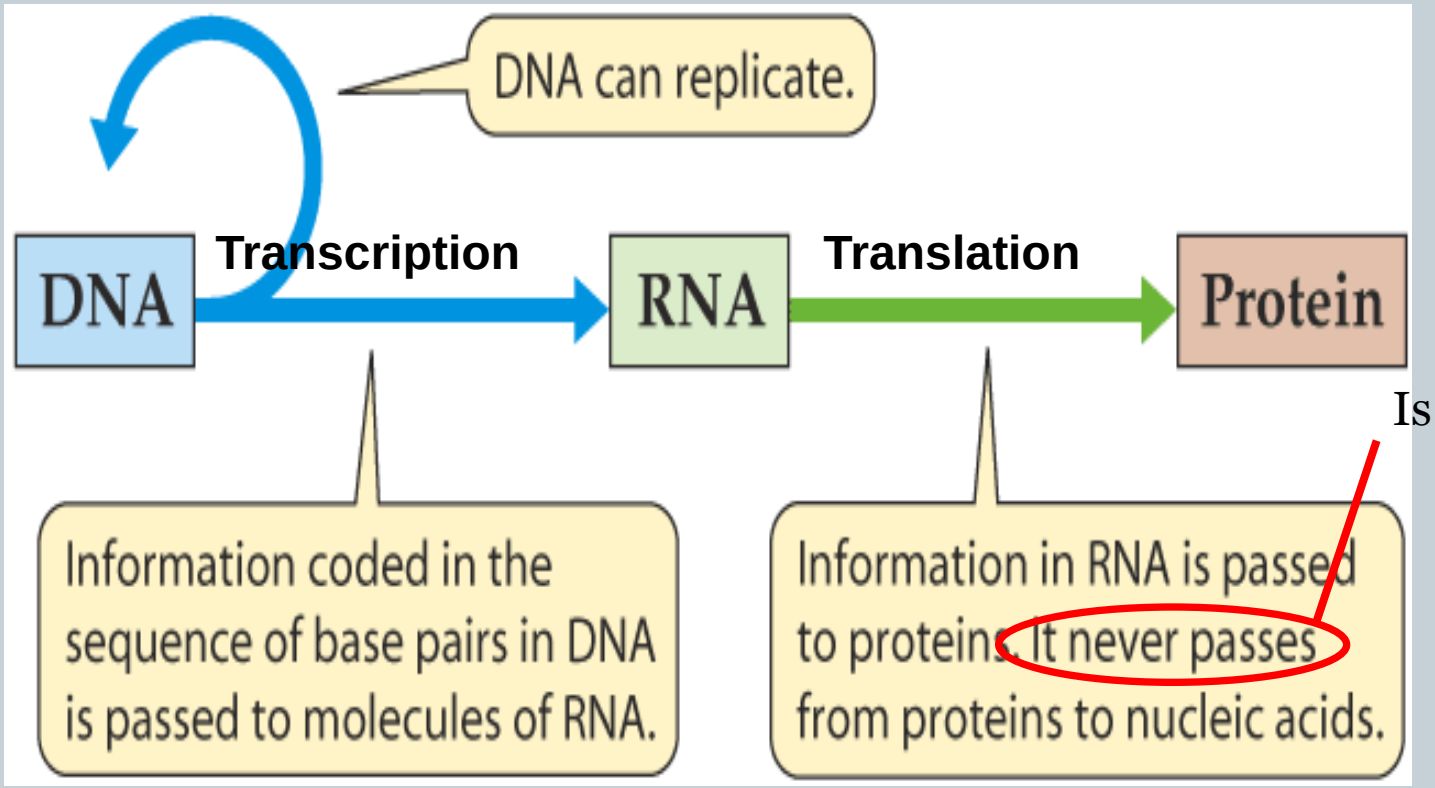
- RNA is similar to DNA chemically. It is usually only a single strand. T(hyamine) is replaced by U(racil)
- Several types of RNA exist for different functions in the cell.



tRNA linear and 3D view: http://www.cgl.ucsf.edu/home/glasfeld/tutorial/trna/trna.gif

DNA, RNA, and the Flow of Information

”The central dogma”



Is this true?

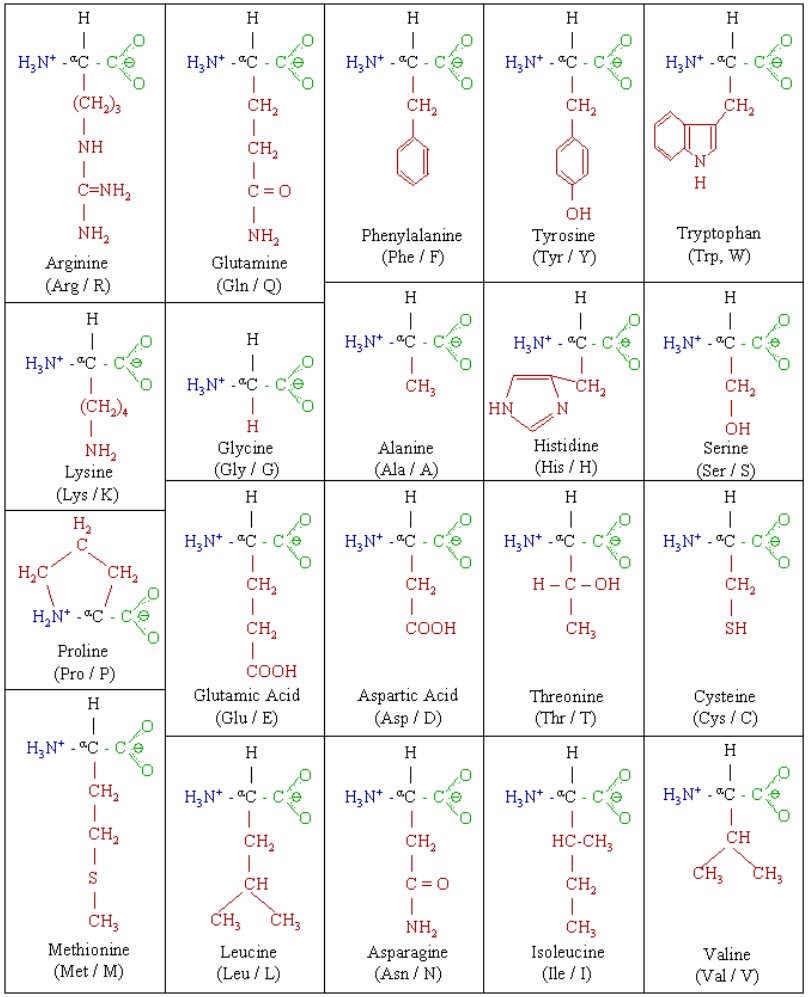
Proteins

- Proteins are polypeptides (strings of amino acid residues)
- Represented using strings of letters from an alphabet of 20: AEGLV…WKKLAG
- Typical length 50…1000 residues

Urease enzyme from Helicobacter pylori

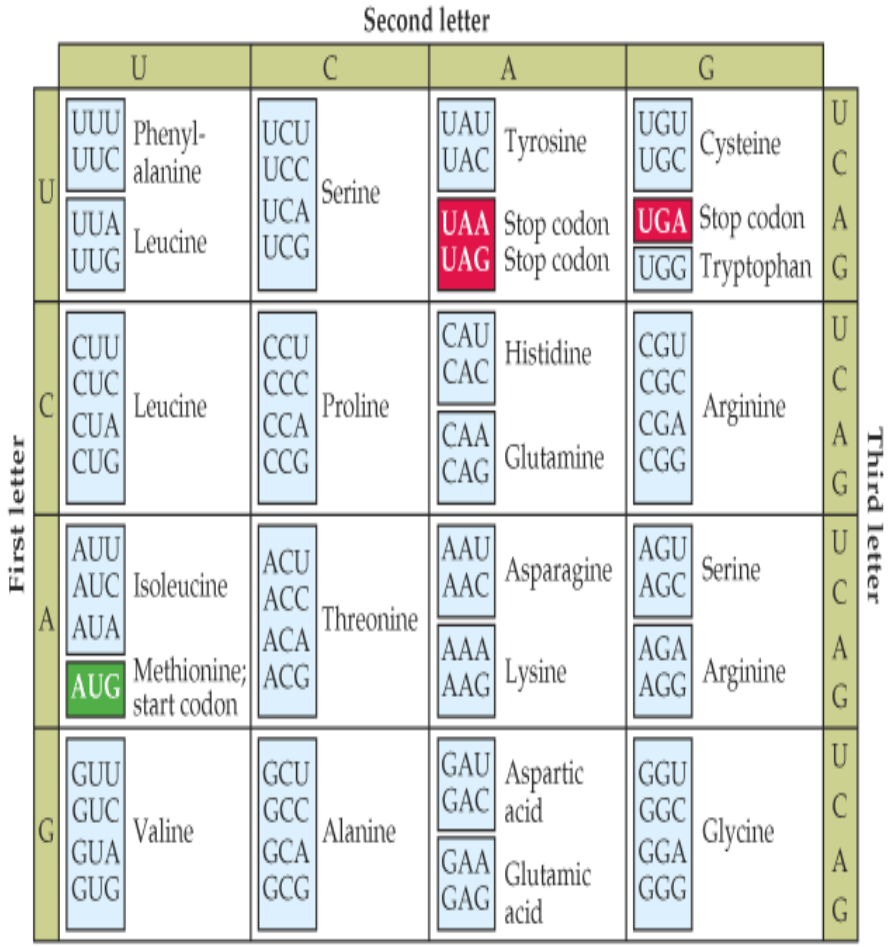
http://upload.wikimedia.org/wikipedia/commons/c/c5/Amino_acids_2.png

Amino acids



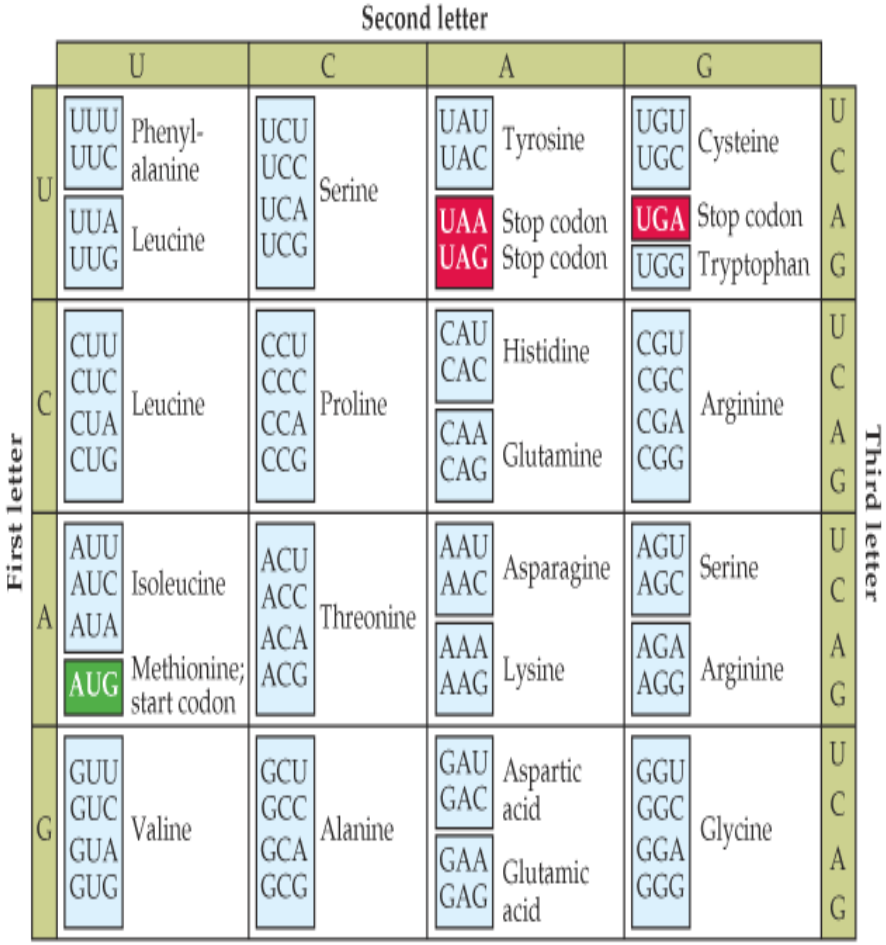
How DNA/RNA codes for protein?

- DNA alphabet contains four letters but must specify protein, or polypeptide sequence of 20 letters.
- Dinucleotides are not enough: $4^2 = 16$ possible dinucleotides
- Trinucleotides (triplets) allow $4^3 = 64$ possible trinucleotides
- Triplets are also called *codons*



How DNA/RNA codes for protein?

- Three of the possible triplets specify ”stop translation”
- Translation usually starts at triplet AUG (this codes for methionine)
- Most amino acids may be specified by more than triplet
- How to find a gene? Look for start and stop codons (not that easy though)

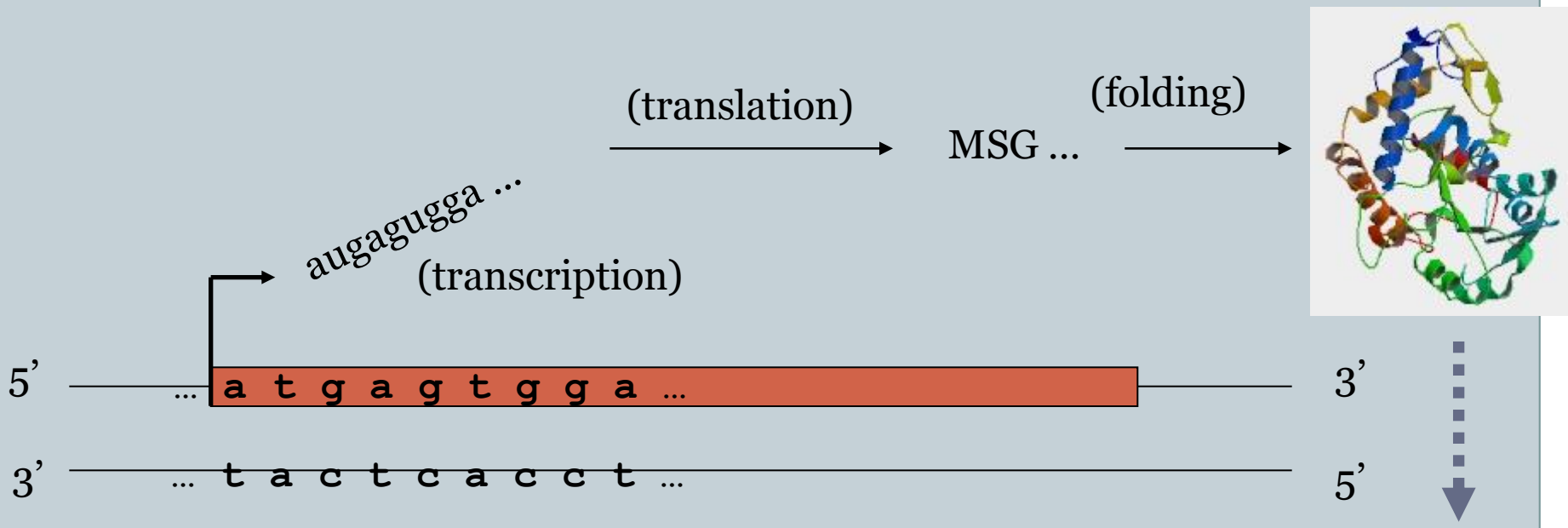


Proteins: Workhorses of the Cell

- 20 different **amino acids**
  - different chemical properties cause the protein chains to fold up into specific three-dimensional structures that define their particular functions in the cell.
- Proteins do all essential work for the cell
  - build cellular structures
  - digest nutrients
  - execute metabolic functions
  - mediate information flow within a cell and among cellular communities.
- Proteins work together with other proteins or nucleic acids as "molecular machines"
  - structures that fit together and function in highly specific, lock-and-key ways.

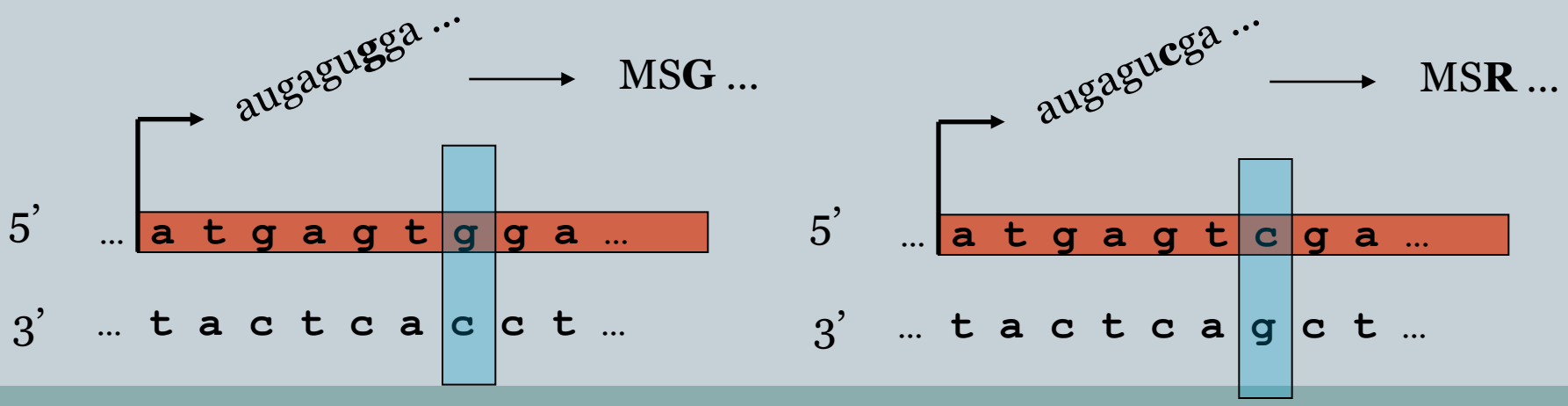
Genes

- "A gene is a union of genomic sequences encoding a coherent set of potentially overlapping functional products" --Gerstein et al.
- A DNA segment whose information is expressed either as an RNA molecule or protein

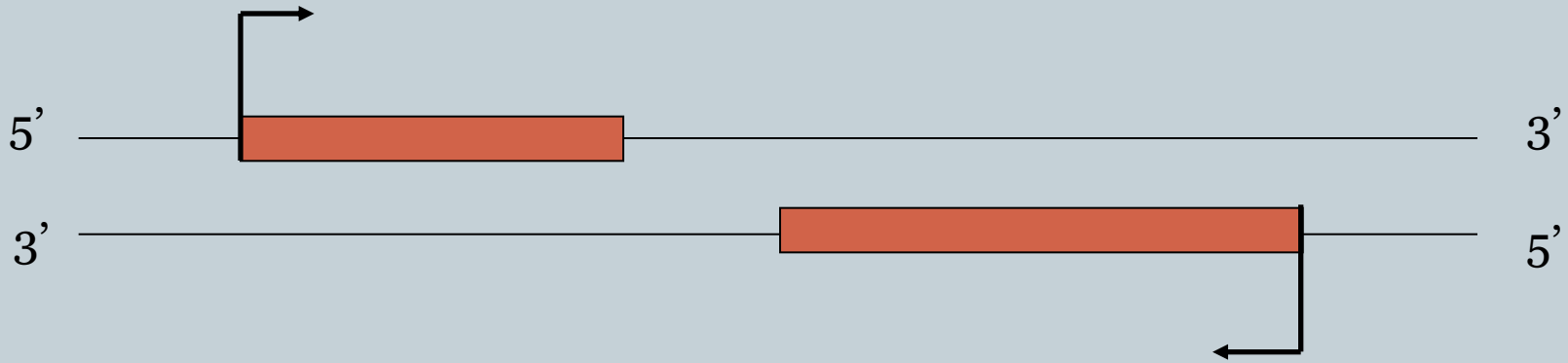


Genes & alleles

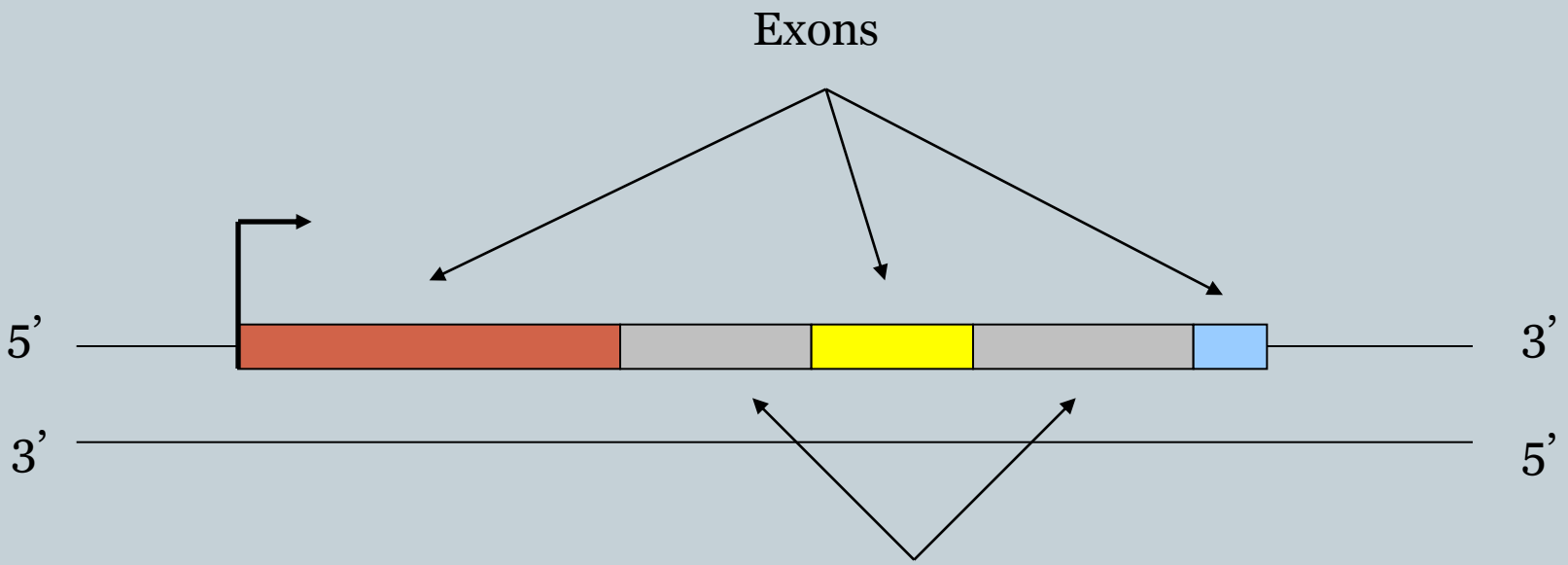
- A gene can have different variants
- The variants of the same gene are called *alleles*



Genes can be found on both strands



Exons and introns & splicing



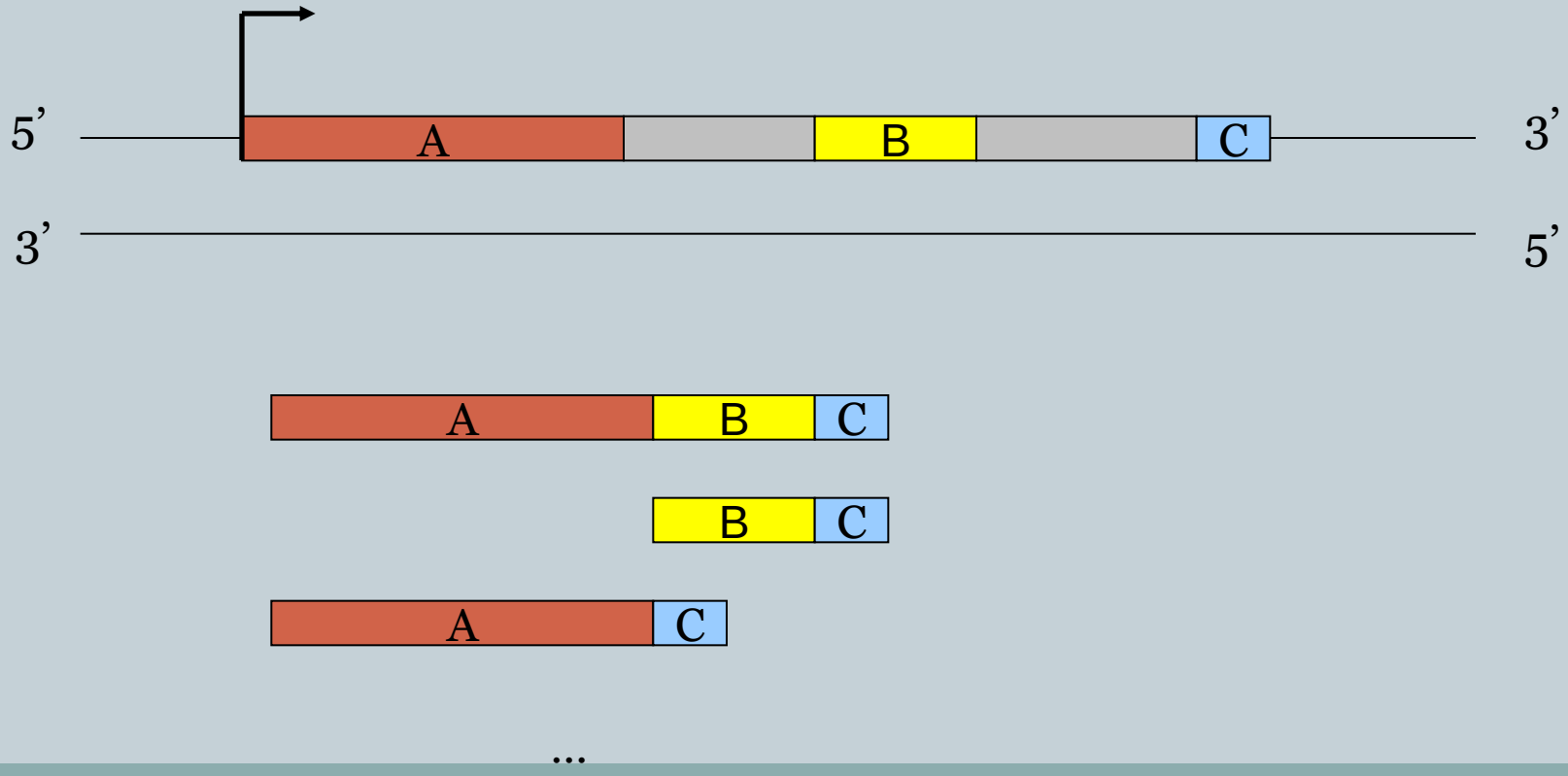
Introns are removed from RNA after transcription

Exons are joined:

This process is called *splicing*

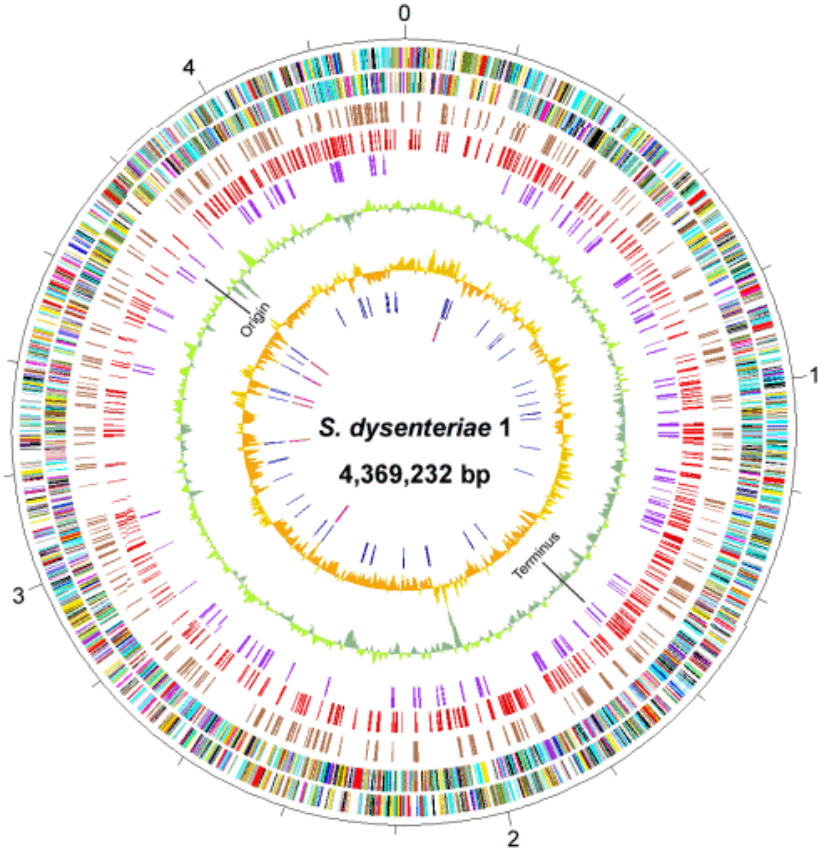
Alternative splicing

Different *splice variants* may be generated



Where does the variation in genomes come from?

- Prokaryotes are typically haploid: they have a single (circular) chromosome
- DNA is usually inherited vertically (parent to daughter)
- Inheritance is clonal
  - Descendants are faithful copies of an ancestral DNA
  - Variation is introduced via mutations, transposable elements, and horizontal transfer of DNA



Chromosome map of *S. dysenteriae*, the nine rings describe different properties of the genome
http://www.mgc.ac.cn/ShiBASE/circular_Sd197.htm

Causes of variation

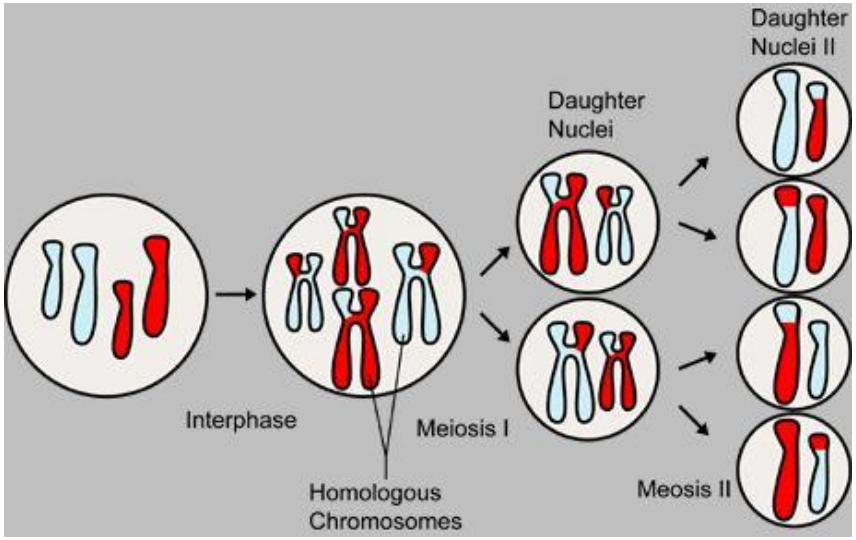
- Mistakes in DNA replication
- Environmental agents (radiation, chemical agents)
- Transposable elements (transposons)
  - A part of DNA is moved or copied to another location in genome
- Horizontal transfer of DNA
  - Organism obtains genetic material from another organism that is not its parent
  - Utilized in genetic engineering

Biological string manipulation

- Point mutation: substitution of a base
  - …ACG**G**CT… => …ACG**C**CT…
- Deletion: removal of one or more contiguous bases (substring)
  - …TT**GA**TCA… => …TTTCA…
- Insertion: insertion of a substring
  - …GGCTAG… => …GG**TCAAC**TAG…

Meiosis

- Sexual organisms are usually diploid
  - Germline cells (gametes) contain N chromosomes
  - Somatic (body) cells have 2N chromosomes
- Meiosis: reduction of chromosome number from 2N to N during reproductive cycle
  - One chromosome doubling is followed by two cell divisions



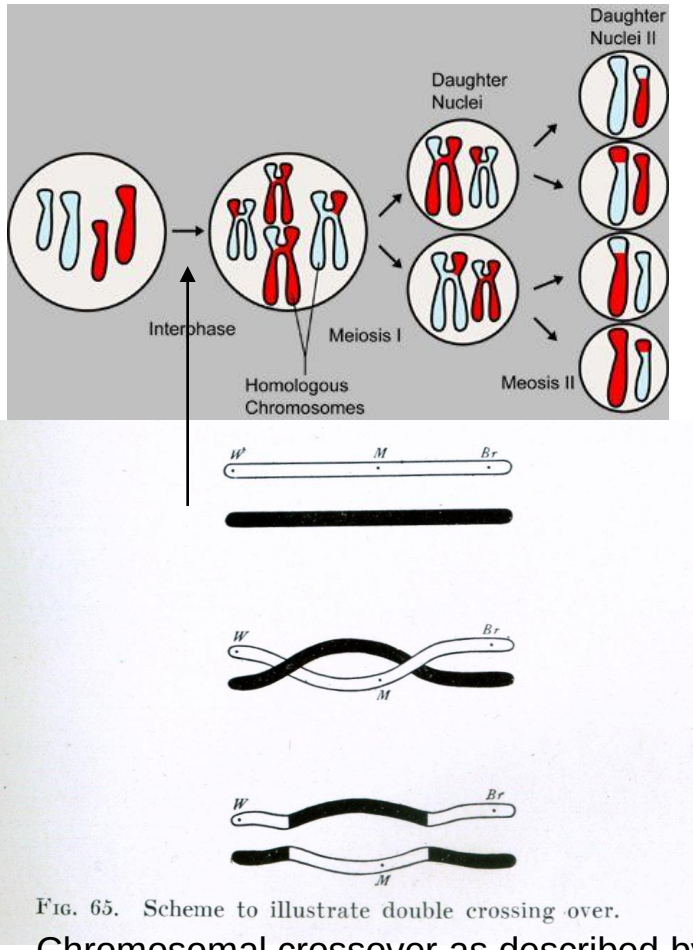
Major events in meiosis

http://en.wikipedia.org/wiki/Meiosis

http://www.ncbi.nlm.nih.gov/About/Primer

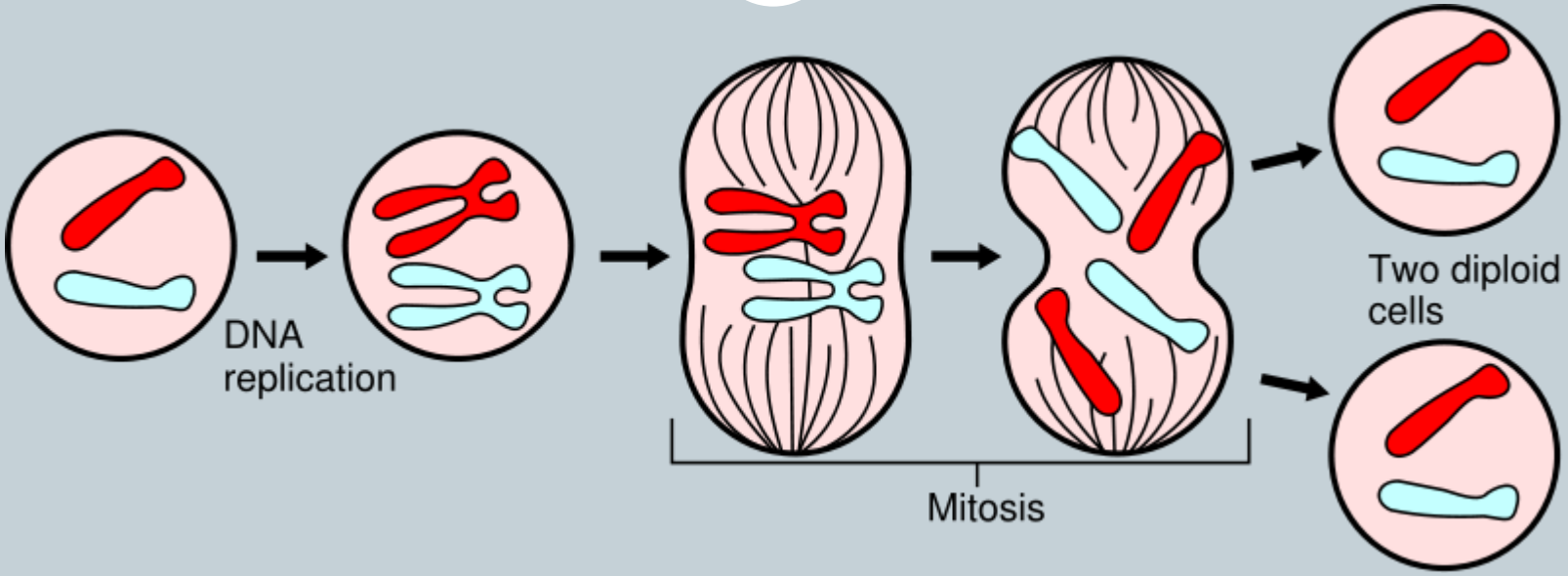
Recombination and variation

- Recap: Allele is a viable DNA coding occupying a given locus (position in the genome)
- In recombination, alleles from parents become suffled in offspring individuals via chromosomal crossover over
- Allele combinations in offspring are usually different from combinations found in parents
- Recombination errors lead into additional variations



Chromosomal crossover as described by T. H. Morgan in 1916

Mitosis



- Mitosis: growth and development of the organism
  - One chromosome doubling is followed by one cell division

http://en.wikipedia.org/wiki/Image:Major_events_in_mitosis.svg