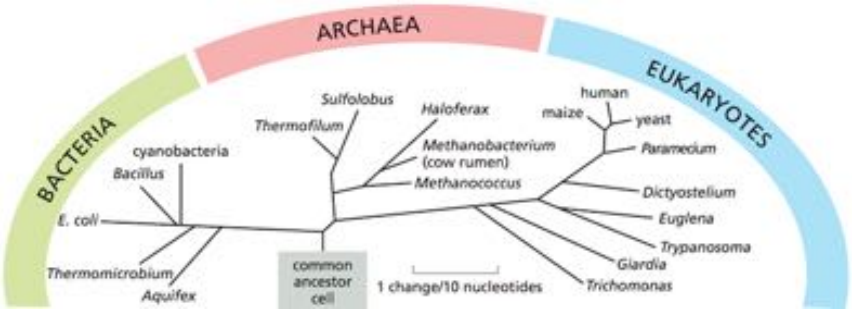
The Biology and Genetics of Cells and Organisms

Mikhail Dozmorov
Fall 2017

Phylogenetic tree



Overview

- Cells are the fundamental units of all living organisms.
- Each cell is a complex system consisting of many substructures.
- Types of organisms:
- **Viruses** are simplest organisms (~ 10,000 bp. long), which require a living host.
- **Prokaryotes** are simplest free-living organisms, e.g. bacteria (~ 1,000,000 bp. long).
- **Eukaryotes** have cells which contain internal structures such as a nucleus, e.g. yeast.
- **Multi-celled organisms** involve cell specialization, requiring differential gene expression and inter-cellular signaling.

Procaryotic cell

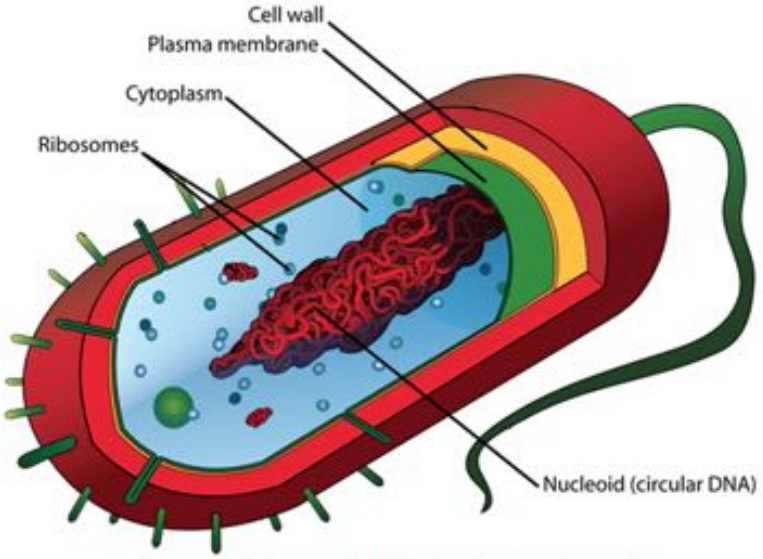


Figure 28: A Typical Prokaryotic Cell

Eukaryotic cell

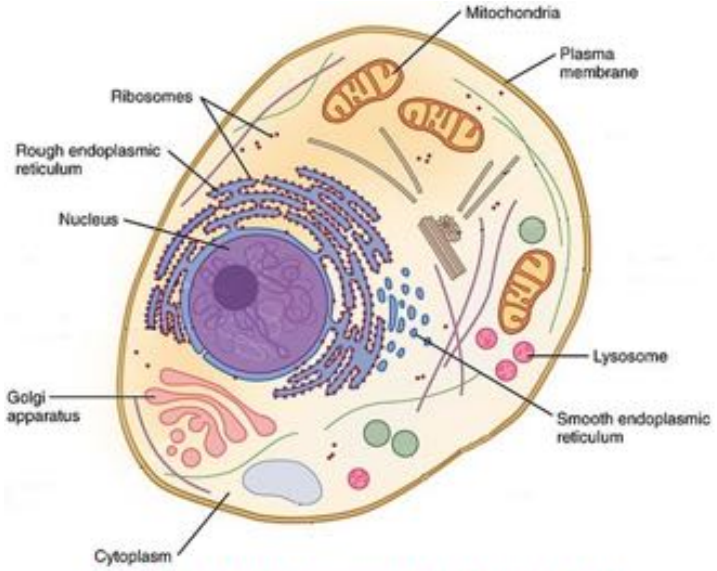


Figure 29: A Eukaryotic Cell (Animal Cell)

Differences between Prokaryotic and Eukaryotic Cells

Characteristic	Prokaryotic Cells	Eukaryotic Cells
Nucleus	No	Yes
Membrane-Bound Organelles	No	Yes
Size of Ribosomes	70S	80S
Cell Wall Composition	Peptidoglycan is Present	No Peptidoglycan
Mitotic Division	No	Yes
DNA Associated with Histones	No	Yes
Number of Chromosomes	One	More than One
Cell Membrane Composition	No Sterols (Except in Mycoplasmas)	Sterols Present
Number of Cells	Usually Unicellular	Usually Multicellular
Size of Cells	Smaller (1-5 µm)	Larger (10-100 µm)

http://www.getmededu.com/differences-between-prokaryotic-and-eukaryotic-cells.html

DNA: the secret of life

Your DNA, along with your environment and experiences, shapes who you are

- Height
- Hair, eye, skin color
- Broad/narrow, small/large features
- Susceptibility to disease
- Response to drug treatments
- Longevity and cognition

Physical traits tend to be strongly genetic, social characteristics tend to be strongly environmental, and everything else is a combination

The nucleus

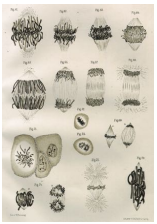
- The nucleus is a sub-compartment found only in eukaryotic cells, in which the organism's DNA resides.
- Enclosing the nucleus is the nuclear membrane, the protective wall that separates the nucleus from the rest of the cell, which is called the cytoplasm.
- The entire cell is enclosed by the plasma membrane.
- Embedded within this membrane is a variety of protein structures that act as channels and pumps to control movement into and out of the cell.

Cells and the genome

- Each cell contains a complete copy of an organism's genome, or blueprint for all cellular structures and activities.
- The genome is distributed along chromosomes, which are made of compressed and entwined DNA.
- Cells are of many different types (e.g. blood, skin, nerve cells), but all can be traced back to a single cell, the fertilized egg.

Discovery of chromosomes

- By the mid-1800s, microscopes were powerful enough to observe the presence of unusual structures called “chromosomes” that seemed to play an important role during cell division.
- It was only possible to see the chromosomes unless appropriate stains were used
- “Chromosome” comes from the Greek words meaning “color body”
- Today, we have much higher resolution microscopes, and a much richer varieties of dies and dying techniques so that we can visualize particular sequence elements.



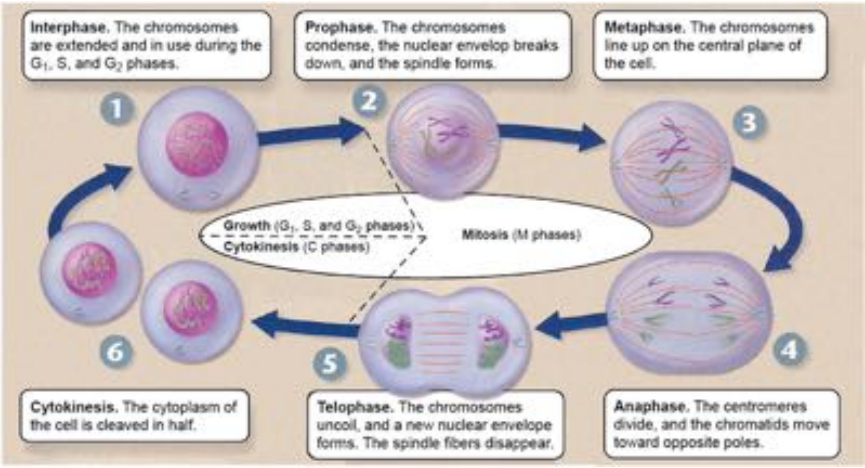
Drawing of mitosis by Walther Flemming. Flemming, W. Zellsubstanz, Kern und Zelltheilung (F. C. W. Vogel, Leipzig, 1882).

Karyotype

- **Cytogenetics** is the study of normal and abnormal chromosomes.
- The normal configuration of chromosomes is often termed the **euploid** karyotypic state.
- Euploidy implies that each of the autosomes is present in normally structured pairs and that the X and Y chromosome are present in normally structure pairs for the sex of the individual.
- Deviation from the euploid karyotype - the state termed **aneuploidy** - is some alteration in the overall chromosome structure, such as loss of entire chromosomes, the presence of extra copies of chromosomes, etc.

Karyotype

Cell cycle



https://www.youtube.com/watch?v=NR0mdDJMHIQ

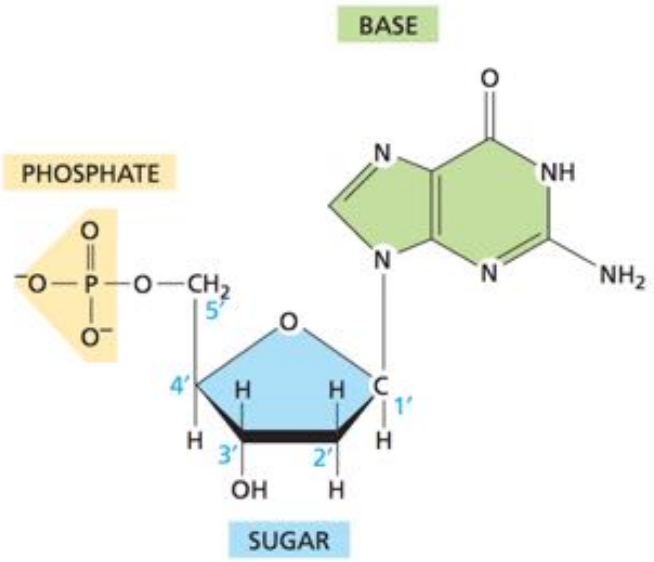
Genes and genome

- **Genes** are discrete hereditary units located on the chromosomes (DNA).
- Each gene provides a clear and unambiguous set of instructions for producing some property of its organism.
- The complete set of genes in an organism is referred to as its **genome**.

Building blocks of DNA

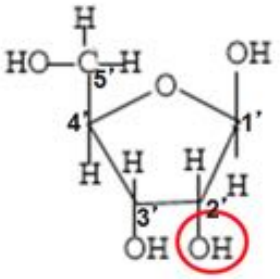
- The basic unit (**nucleotide**) is composed of an organic **base** attached to a deoxyribose **sugar**
- The **phosphate** group also attached to the sugar
- The **base** is one of cytosine (C), thymine (T), adenine (A), and guanine (G)

Nucleotide structure

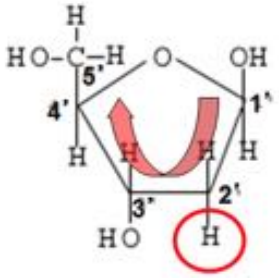


Nucleotide structure

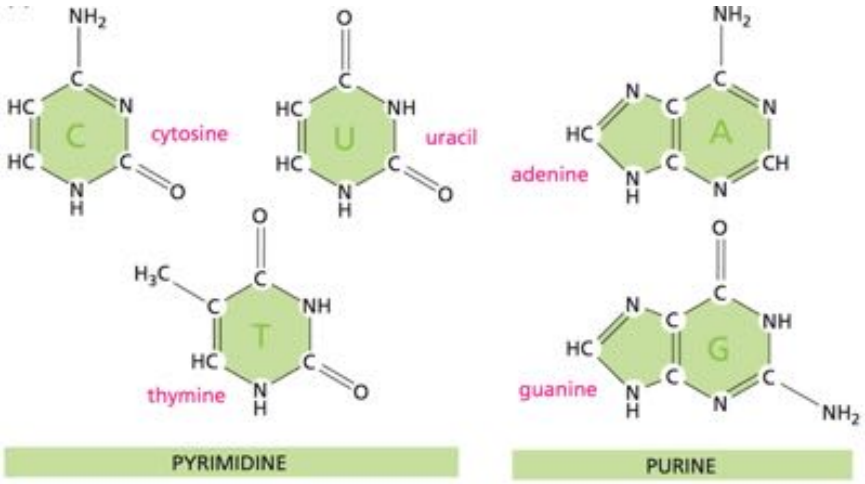
- Ribose Sugar



- Deoxyribose Sugar



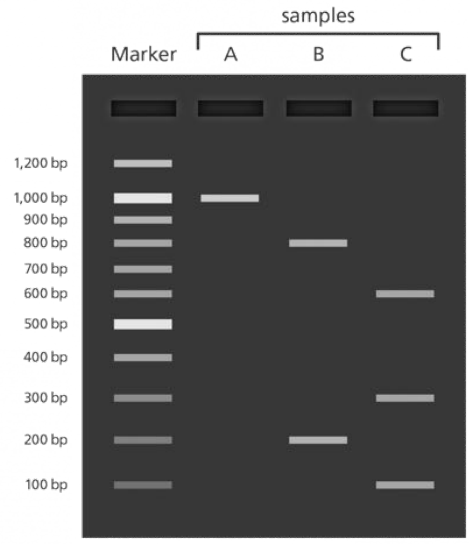
Bases



Getting to know DNA: Gel electrophoresis

- In the mid-1900's methods were developed to size separate and visualize DNA within an electrically charged gel (originally made from sugar)
- DNA is loaded into different "lanes" at the top of the gel, and the charge is applied
- Key idea is that smaller fragments of DNA would move faster through the gel and be towards the bottom than longer fragments that will be towards the top.
- Allows to test for presence/absence of DNA as well as compare relative lengths of molecules
- Will often reserve one of the lanes for a "DNA ladder" with fragments of a known size distribution

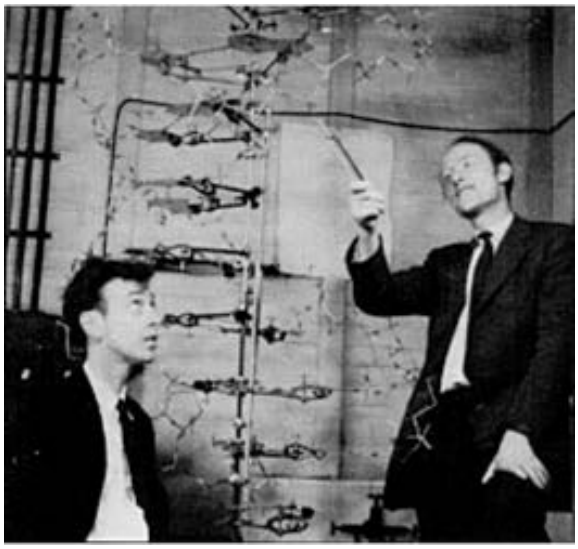
Getting to know DNA: Gel electrophoresis



http://www.yourgenome.org/facts/what-is-gel-electrophoresis

Discovery of double helix, 1953

- James Watson and Francis Crick



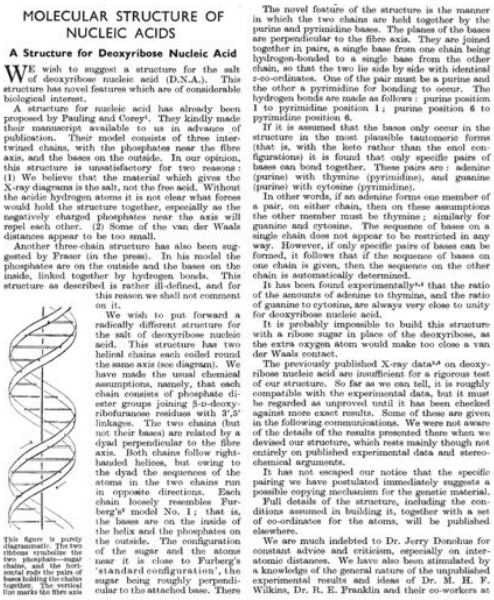
Before the discovery

Figure 1.2
Two scientists whose work was influential on James Watson and Francis Crick when they elucidated the structure of DNA.
(A) Maurice Wilkins.
(B) Rosalind Franklin. (A and B courtesy of Science Photo Library.)

Genomic DNA

- DNA is a **double helix**, with bases to the center (like rungs on a ladder) and sugar-phosphate units along the sides of the helix (like the sides of a twisted ladder).
- The strands are **complementary** (Watson-Crick base pairing rules)
- A (purine) pairs with T (pyrimidines) C (pyrimidines) pairs with G (purine)
- The pairs held together by hydrogen bonds. The helix is caused by the use of the hydrogen bonds between the single-strands

Genomic DNA



MOLECULAR STRUCTURE OF NUCLEIC ACIDS

A Structure for Deoxyribose Nucleic Acid

WE wish to suggest a structure for the salt of deoxyribose nucleic acid (D.N.A.). This structure has novel features which are of considerable biological interest.

A structure for nucleic acid has already been proposed by Pauling and Corey[1]. They kindly made their manuscript available to us in advance of publication. Their model consists of three intertwined chains, with the phosphates near the fibre axis, and the bases on the outside. In our opinion, this structure is unsatisfactory for two reasons: (1) We believe that the material which gives the X-ray diagrams is the salt, not the free acid. Without the acidic hydrogen atoms it is not clear what forces would hold the structure together, especially as the negatively charged phosphates near the axis will repel each other. (2) Some of the van der Waals distances appear to be too small.

Another three-chain structure has also been suggested by Fraser (in the press). In his model the phosphates are on the outside and the bases on the inside, linked together by hydrogen bonds. This structure as described is rather ill-defined, and for this reason we shall not comment on it.

We wish to put forward a radically different structure for the salt of deoxyribose nucleic acid. This structure has two helical chains each coiled round the same axis (see diagram). We have made the usual chemical assumptions, namely, that each chain consists of phosphate diester groups joining β-D-deoxyribofuranose residues with 3',5' linkages. The two chains (but not their bases) are related by a dyad perpendicular to the fibre axis. Both chains follow right-handed helices, but owing to the dyad the sequences of the atoms in the two chains run in opposite directions. Each chain loosely resembles Furberg's[2] model No. 1; that is, the bases are on the inside of the helix and the phosphates on the outside. The configuration of the sugar and the atoms near it is close to Furberg's 'standard configuration', the sugar being roughly perpendicular to the attached base. There

This figure is purely diagrammatic. The two ribbons symbolize the two phosphate–sugar chains, and the horizontal rods the pairs of bases holding the chains together. The vertical line marks the fibre axis

The novel feature of the structure is the manner in which the two chains are held together by the purine and pyrimidine bases. The planes of the bases are perpendicular to the fibre axis. They are joined together in pairs, a single base from one chain being hydrogen-bonded to a single base from the other chain, so that the two lie side by side with identical z-co-ordinates. One of the pair must be a purine and the other a pyrimidine for bonding to occur. The hydrogen bonds are made as follows: purine position 1 to pyrimidine position 1; purine position 6 to pyrimidine position 6.

If it is assumed that the bases only occur in the structure in the most plausible tautomeric forms (that is, with the keto rather than the enol configurations) it is found that only specific pairs of bases can bond together. These pairs are: adenine (purine) with thymine (pyrimidine), and guanine (purine) with cytosine (pyrimidine).

In other words, if an adenine forms one member of a pair, on either chain, then on these assumptions the other member must be thymine; similarly for guanine and cytosine. The sequence of bases on a single chain does not appear to be restricted in any way. However, if only specific pairs of bases can be formed, it follows that if the sequence of bases on one chain is given, then the sequence on the other chain is automatically determined.

It has been found experimentally[3,4] that the ratio of the amounts of adenine to thymine, and the ratio of guanine to cytosine, are always very close to unity for deoxyribose nucleic acid.

It is probably impossible to build this structure with a ribose sugar in place of the deoxyribose, as the extra oxygen atom would make too close a van der Waals contact.

The previously published X-ray data[5,6] on deoxyribose nucleic acid are insufficient for a rigorous test of our structure. So far as we can tell, it is roughly compatible with the experimental data, but it must be regarded as unproved until it has been checked against more exact results. Some of these are given in the following communications. We were not aware of the details of the results presented there when we devised our structure, which rests mainly though not entirely on published experimental data and stereochemical arguments.

It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material.

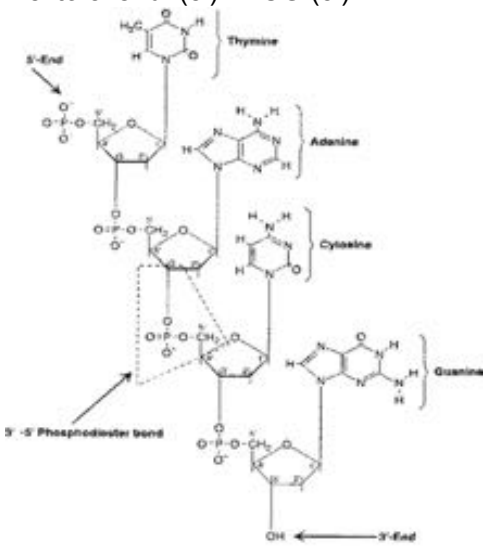
Full details of the structure, including the conditions assumed in building it, together with a set of co-ordinates for the atoms, will be published elsewhere.

We are much indebted to Dr. Jerry Donohue for constant advice and criticism, especially on interatomic distances. We have also been stimulated by a knowledge of the general nature of the unpublished experimental results and ideas of Dr. M. H. F. Wilkins, Dr. R. E. Franklin and their co-workers at

https://www.nature.com/nature/dna50/watsoncrick.pdf

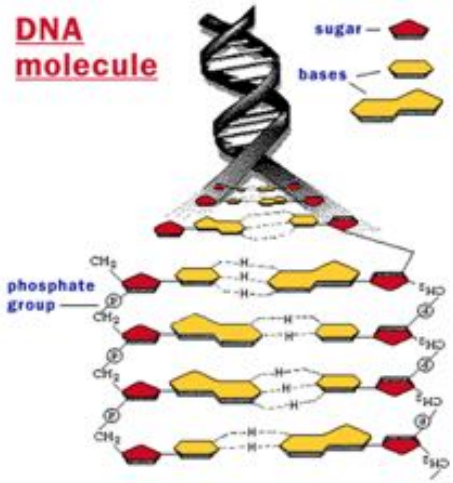
Nucleic acid strand

- Strand synthesis - from 5' to 3' end: (5') TACG (3')



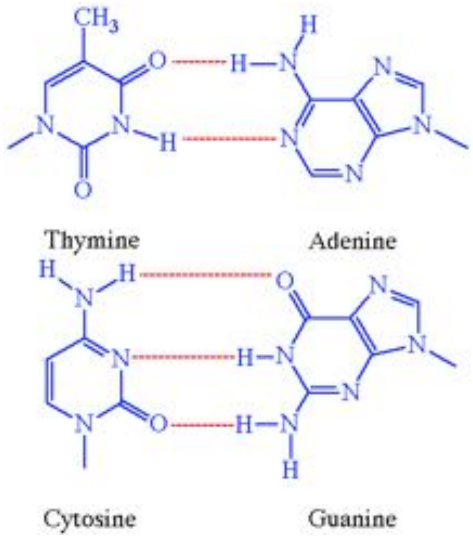
Double strand: DNA base pairing

- The two strands are held together by hydrogen bonds between nitrogen bases



Rules of base pairing

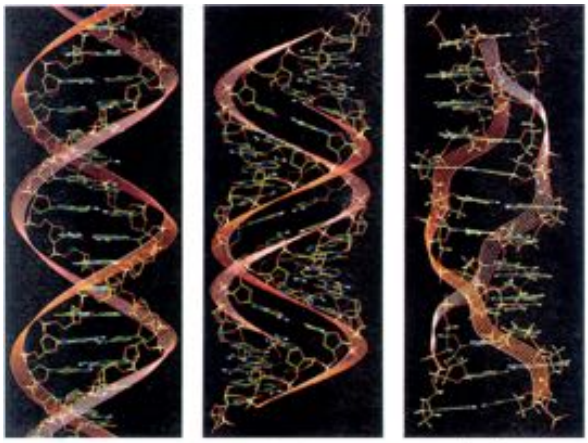
- Rules of base pairing: A-T(U), C-G (and G-U in RNA)



Base pairing

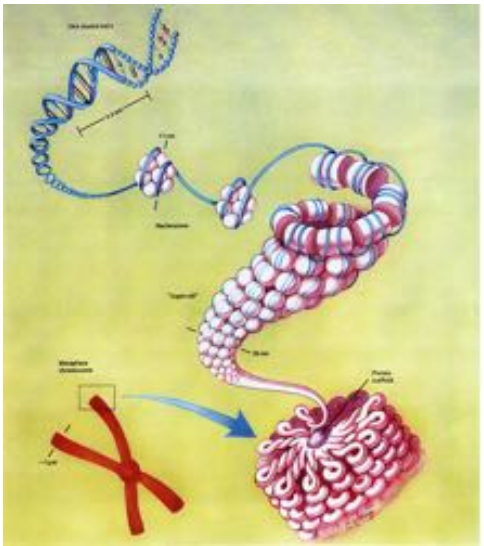
- The force that holds a base pair together is a weak hydrogen bond.
- Although each individual bond is weak, their cumulative effect along the strands is strong enough to bind the two strands tightly together.
- As a result, DNA is chemically inert and is a stable carrier of genetic information.

Helixes conformation



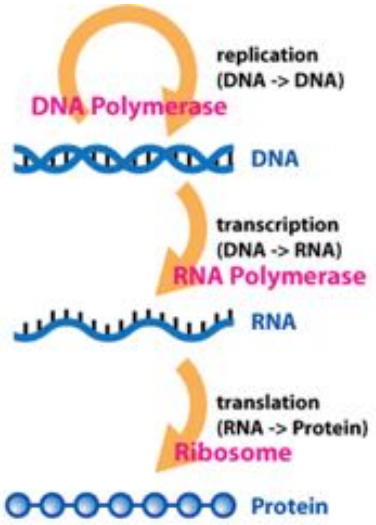
- A: B-form: Right-handed, 3.4 nm between bases, 10 bases per turn
- B: A-form: Right-handed, 2.3 nm between bases, 11 bases per turn
- C: Z-form: Left-handed

Eukaryotic DNA packaging



http://www.hhmi.org/biointeractive/dna-packaging

The central dogma of molecular biology

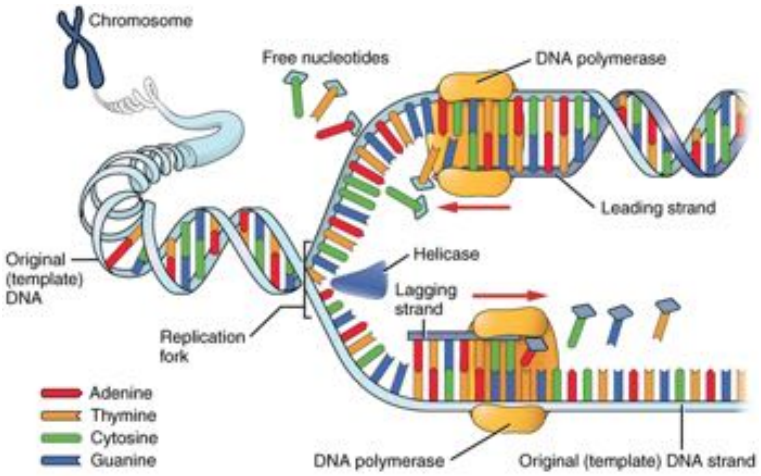


- Formulated by Francis Crick in 1956
- **DNA makes RNA and RNA makes protein**
- **Transcription** is the making of an RNA molecule off a DNA template.

DNA replication

- In DNA replication, the DNA molecule unwinds and the "ladder" unzips, thereby disrupting the weak bonds between the base pairs and allowing the strands to separate.
- Nucleotides have to be assembled and available in the nucleus, along with energy to make bonds between nucleotides.
- DNA polymerases unzip the helix by breaking the H-bonds btw bases
- Once the polymerases have opened the molecule, an area known as the replication bubble forms (always initiated at a certain set of nucleotides, the origin of replication).
- New nucleotides are placed in the fork and link to the corresponding parental nucleotide already there (A with T, C with G).

DNA replication



https://www.youtube.com/watch?v=TNKWgcFPHqw

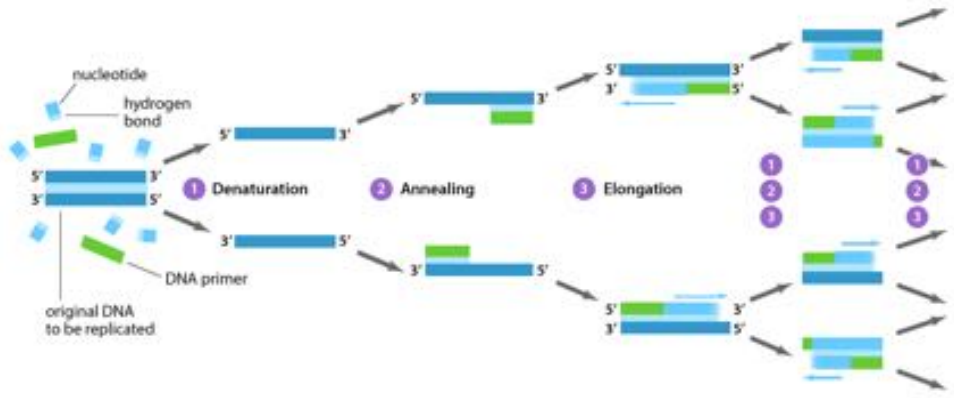
Biotechnologies

- Amazing biotechnologies for manipulating DNA molecules are used as building blocks for even more powerful technologies.
- *DNA synthesis machines* enable one to grow short DNA molecules of a specified sequence.
- The *Polymerase chain reaction (PCR)* enables one to make many copies of a particular DNA sequence anywhere in solution given only the starting/ending sequences (primers).

Polymerase Chain Reaction (PCR)

- Developed in 1983 by Kary Mullis, PCR allows for the amplification of DNA fragments that are flanked by known “primer” sequences.
- Exquisitely sensitive and specific, can amplify a single molecule in a sample into billions of copies with nearly perfect fidelity
- Uses naturally occurring polymerase enzymes that copy DNA by adding free nucleotides to a single-stranded template

Polymerase Chain Reaction (PCR)



https://www.abmgood.com/marketing/knowledge_base/polymerase_chain_reaction_introduction.php

Polymerase Chain Reaction (PCR)

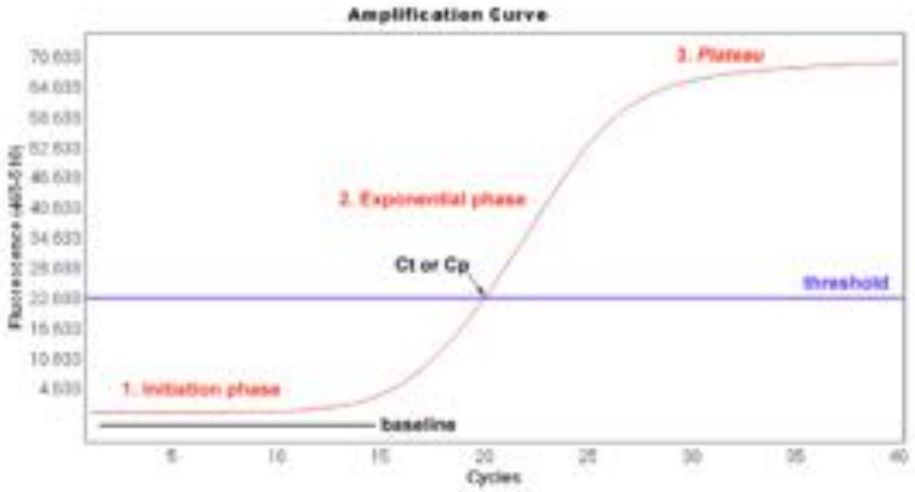
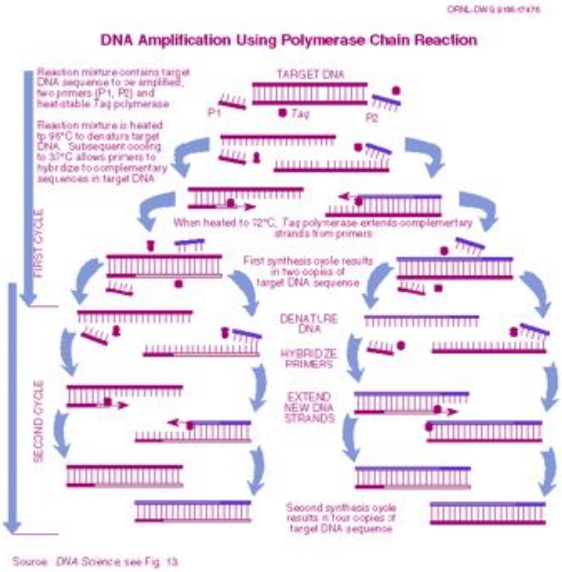


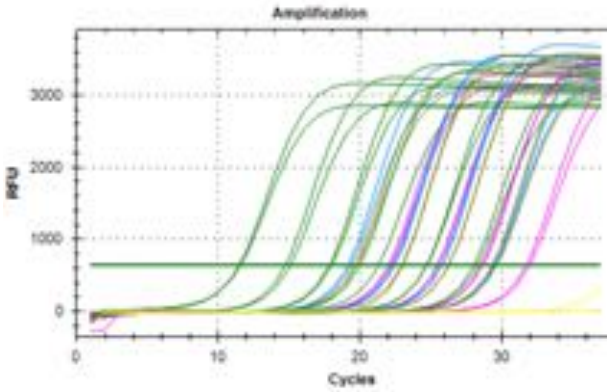
Figure 11. Phases of a PCR amplification curve. Blue: amplification curve of a positive sample. Red: negative control.

DNA amplification using PCR



https://www.animalgenome.org/edu/doe/pcr.html

Quantitative PCR



- Quantitation requires normalization (comparison to standard curves)
- Normalization is based on assumptions

RNA vs. DNA: Single vs. double strands

- **RNA - ribonucleic acid**:

1. Single strand.
2. Ribose sugar.
3. AUCG nucleotides

- **DNA - deoxyribonucleic acid**:

1. Double strand.
2. Deoxyribose sugar.
3. ATCG nucleotides

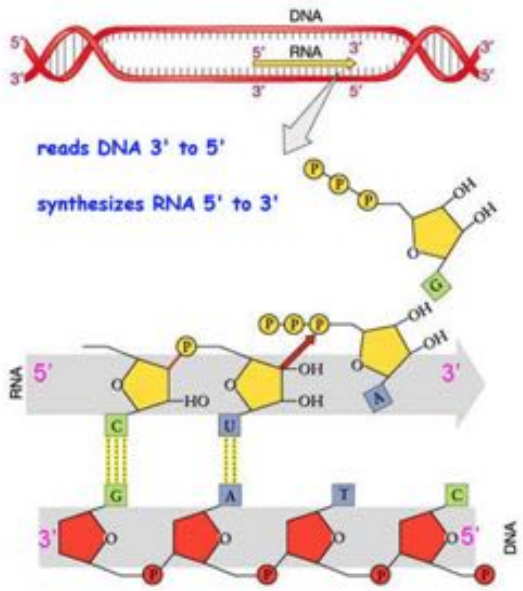
Nucleic acids - RNA

- **Messenger RNA (mRNA)** - carrier of genetic information
- **Transfer RNA (tRNA)** - deliver amino acids for protein synthesis
- **Ribosomal RNA (rRNA)** - central component of ribosome, protein manufacturing machinery
- **Small RNA (siRNA, miRNA, snRNA, piwiRNA)** - regulation of transcription/translation

Transcription

- In transcription, the DNA double helix opens along its length
- One strand of the open helix remains inactive, while the other strand acts as a template against which a complementary strand of mRNA forms
- The sequence of bases along the mRNA strand is identical to the sequence of bases along the inactive DNA strand, except uracil (U) replaces T. Also RNA has ribose sugar instead of deoxyribose sugar.
- RNA (single-stranded) moves out into the cytoplasm.

RNA transcribed from 5' to 3' end

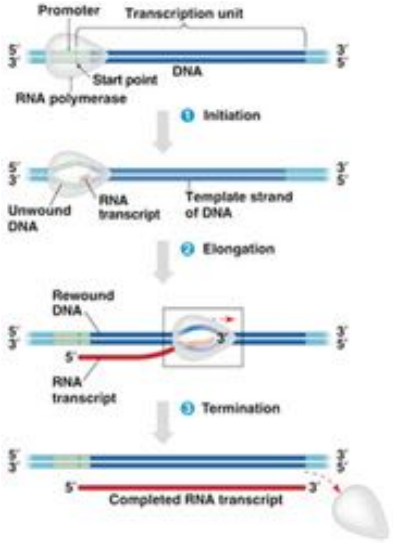


Three RNA polymerases

- RNA polymerase is an enzyme that produces RNA

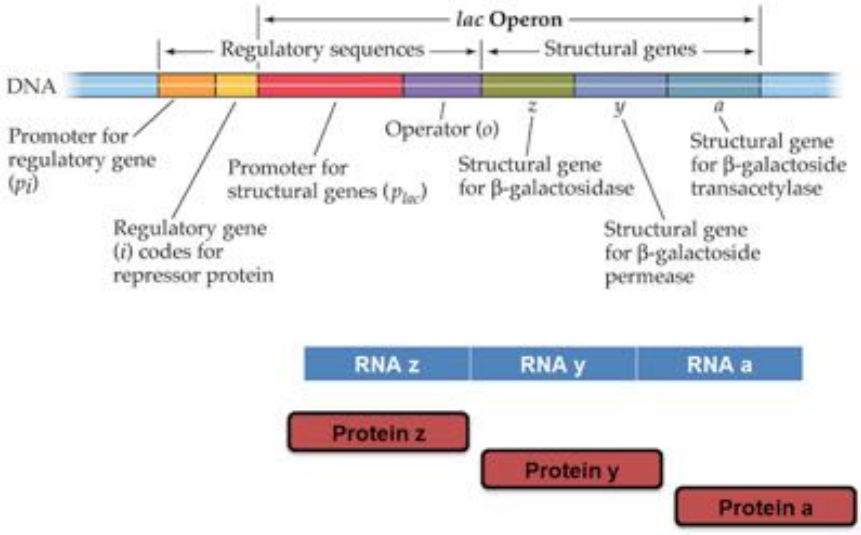
1. RNA Pol I - transcription of ribosomal RNA (not the 5S subunit)
2. RNA Pol II - mRNA, snRNA, microRNA
3. RNA Pol III - tRNA, 5S rRNA, small RNA

Three stages of transcription



http://vcell.ndsu.nodak.edu/animations/transcription/movie-flash.htm

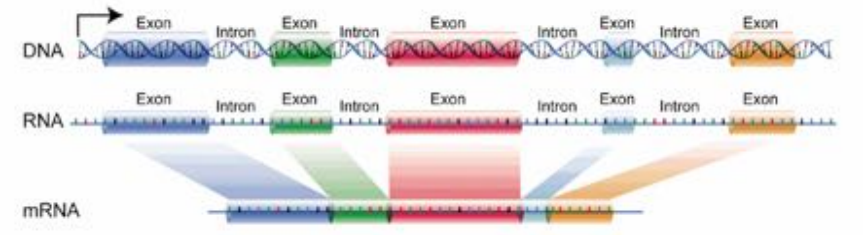
Gene structure in prokaryotes



Gene structure in eukaryotes

- Non-coding interruptions are known as intervening sequences or **introns**.
- Coding sequences that are expressed are **exons**.
- Most, but not all structural eukaryote genes contain introns. Although transcribed, these introns are excised (cut out) before translation.

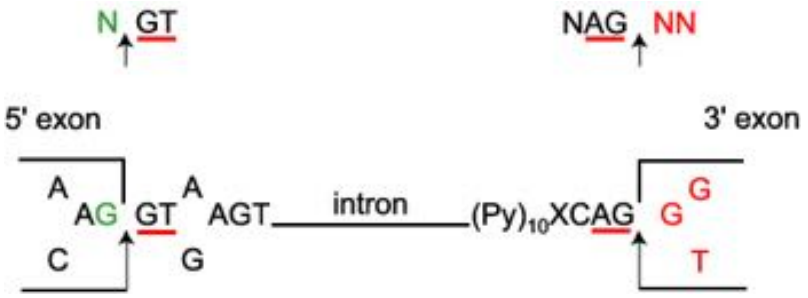
Gene structure in eukaryotes



- Exon - EXpressed regiON
- Intron - INTragenic regiON
- mRNA splicing - variants of mRNA assembly

Intron boundaries

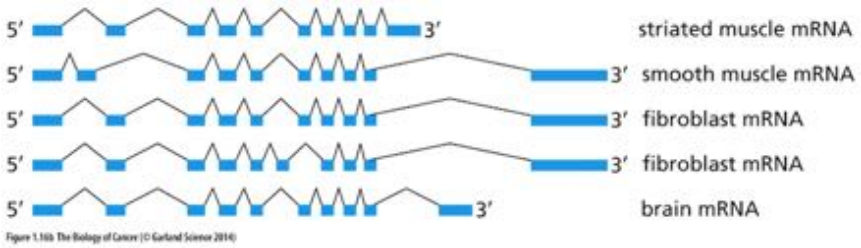
- Introns always have two distinct nucleotides at either end.
- At the 5' end the DNA nucleotides are GT [GU in the premessenger RNA (pre-mRNA)]; at the 3' end they are AG.
- These nucleotides are part of the splicing sites.



The GT/AG mRNA processing rule is applicable for almost all eukaryotic genes

http://www.imgt.org/IMGTeducation/Aide-memoire/_UK/splicing/

Alternative splicing

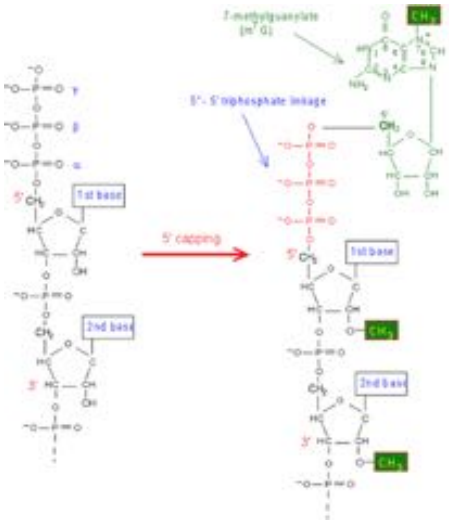


- Tissue specific alternative splicing patterns of the α**-tropomyosin** pre-mRNA molecule
- Exons are blue rectangles
- Introns are black carets

RNA processing

- After eukaryotes transcribe an RNA, the RNA transcript is extensively modified before export to the cytoplasm.
- A **cap of 7-methylguanine** (a series of an unusual base) is added to the 5' end of the mRNA. This cap is essential for binding the mRNA to the ribosome.
- A **string of adenines** (as many as 200 nucleotides known as poly-A) is added to the 3' end of the mRNA after transcription. The function of a poly-A tail is not known, but it can be used to capture mRNAs for study.
- Introns are cut out of the message and the exons are spliced together before the mRNA leaves the nucleus.

RNA processing

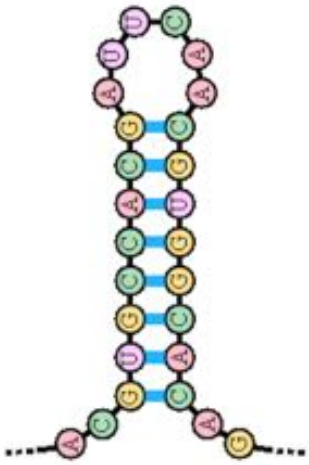


- 5' cap - 7-methylguanylate (m7G), 5'-5' triphosphate linkage
- 3' tail - poly (A) tail, 100-250 bases of adenylic acid
- m7G, 1st and 2nd riboses are methylated

http://vcell.ndsu.nodak.edu/animations/mrnaproflash.htm

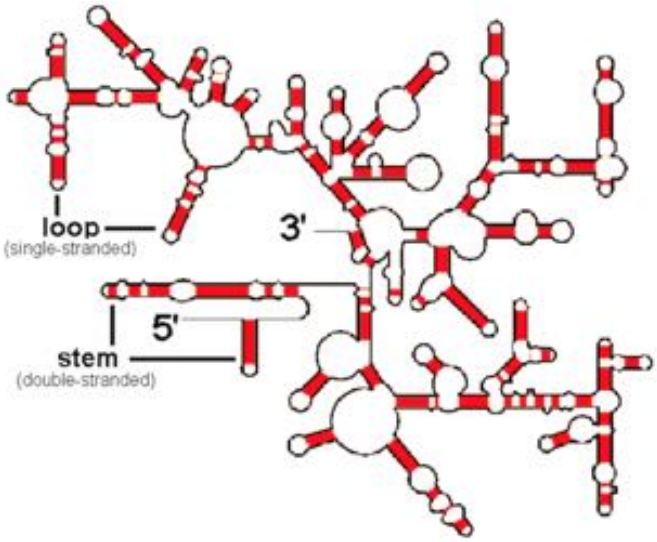
RNA strand structure

- Single stranded
- Hairpin structure



RNA strand structure

- Transfer RNA (tRNA) structure

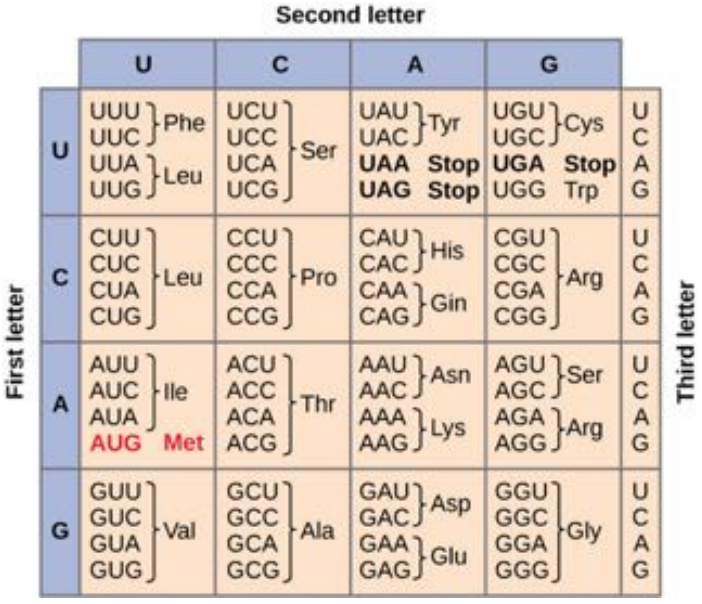


Translation

- In translation, the mRNA serves as a template for protein synthesis.
- The sequence of bases along the mRNA is thus converted into a string of amino acids.
- Consecutive non-overlapping triplets of bases (called codons) act as the code to specify the particular amino acids
- There are 64 possible codons but only 20 amino acids.
- There is room for redundancy - this provides a safeguard against small errors that might occur during transcription.

http://vcell.ndsu.nodak.edu/animations/translation/movie-flash.htm

Translation code



Exercise: Transcribe and translate

- http://learn.genetics.utah.edu/content/basics/transcribe/TranscribeTranslate.swf
- http://sepuplhs.org/high/sgi/teachers/genetics_act16_sim.html

Gene expression

- Each cell contains a complete copy of the organism's genome. A gene that is transcribed is said to be expressed
- Not all cells express the same genes which is why different cells perform different functions
- Even within the same cell different genes will be expressed at different times and perhaps at different levels

Housekeeping genes

- **Housekeeping genes** are genes that are required for the maintenance of basal cellular functions that are essential for the existence of a cell, regardless of its specific role in the tissue or organism.
- They are expected to be expressed in all cells of an organism under normal conditions, irrespective of tissue type, developmental stage, cell cycle state, or external signal.
- Can be used as internal controls in gene expression studies

Housekeeping genes

- Typical examples:
- glyceraldehyde- 3-phosphate dehydrogenase (GAPDH)
- tubulins (beta-tubulin TUBB)
- cyclophilin (cyclophilin A CYPA)
- albumin (ALB)
- actins (beta-actin ACTB)
- 18S rRNA or 28S rRNA.

Housekeeping genes

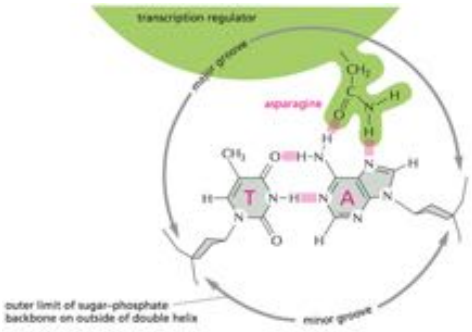
- Should they be expressed at constant level?
- How to account for alternative splicing?

3,804 housekeeping genes (and exons) defined from Human BodyMap project gene expression data http://www.tau.ac.il/~elieis/HKG/

Eisenberg, Eli, and Erez Y. Levanon. “Human Housekeeping Genes, Revisited.” Trends in Genetics: TIG 29, no. 10 (October 2013): 569–74. doi:10.1016/j.tig.2013.05.010.

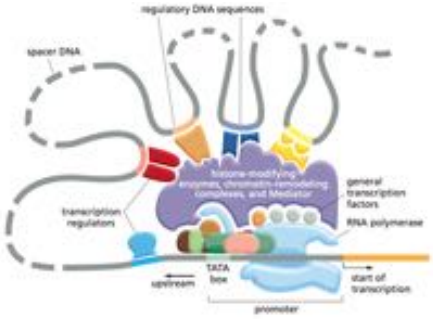
Transcription factors

- Transcription factors (TFs) are proteins that bind to specific DNA sequences in the control region of each gene and determine whether or not the gene will be transcribed.
- The specific stretch of nucleotide sequence to which the TFs bind, often called a sequence motif, is usually quite short, typically 5-10 nucleotides long.



Transcription factors

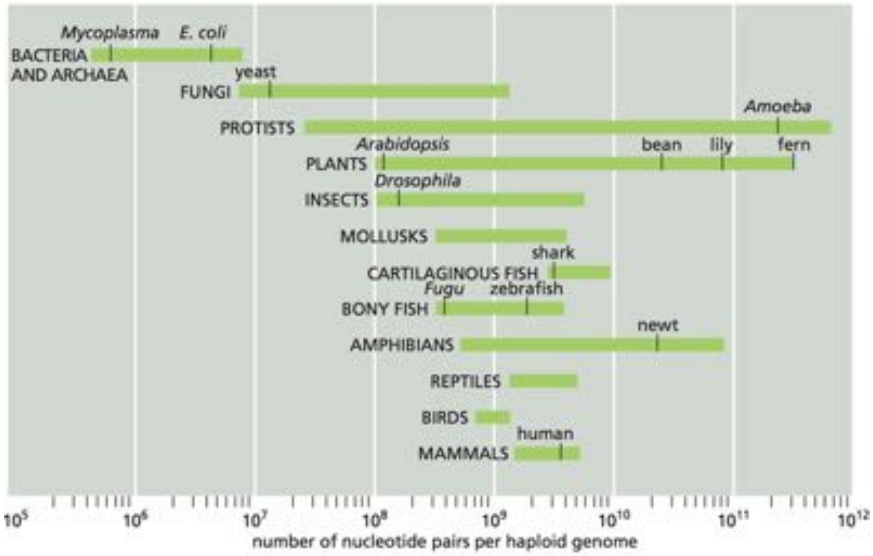
- Some TFs provide the RNA polymerase enzyme with access to the gene while other TFs block such access to ensure the gene is transcriptionally repressed
- Histone modifications may also affect transcription by RNA polymerases of specific regions of chromosomal DNA. Methylation of CpG sites and microRNAs also affect gene expression.



Human Genome Project

- Computational biology attempts to use genome sequence to ascertain function of genes.
- Although genomes vary slightly from person to person, it seemed reasonable to try to establish a consensus human genome sequence.
- Robert Sinsheimer, chancellor of UC Santa Cruz, proposed to sequence the human genome in 1984.
- After much debate, the human genome project started in October 1990.

Genome sizes compared



http://www.hhmi.org/biointeractive/coding-sequences-dna

The advent of genomics

- In the 1860's while studying peas, Gregor Mendel observed that genetic information is passed in particulate form from an organism to its offspring.
- He found that the heritable material controlling the smoothness of peas behaved independently of the material governing plant height or flower color. He deduced there are two copies of a gene for flower color and two copies of a gene for pea shape.

Versuche uber Pflanzen-Hybriden. Verh. Naturforsch (Experiments in Plant Hybridization) Mendel, G. (1866). Ver. Brünn 4: 3–47 (in English in 1901, J. R. Hortic. Soc. 26: 1–32).

http://www.indiana.edu/~p1013447/dictionary/mendel.htm

Mendel's theory of inheritance

	Seed shape	Seed color	Flower color	Flower position	Pod shape	Pod color	Plant height
One form of trait (dominant)	round	yellow	violet-red	axial	inflated	green	tall
A second form of trait (recessive)	wrinkled	green	white	terminal	pinched	yellow	short

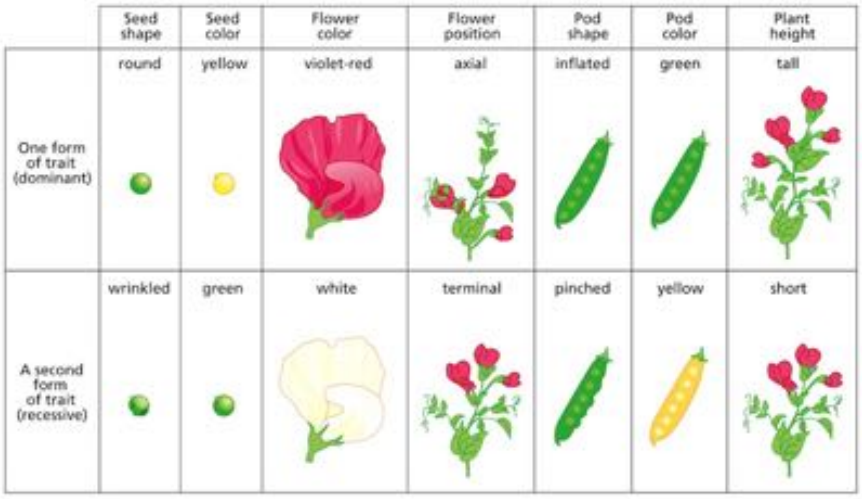


Figure 1.2 The Biology of Cancer (© Garland Science 2014)

The advent of genomics

- Mendel's work implied that the entire repertoire of an organism's genetic information - its genome - is organized as a collection of discrete, separable information packets, now called genes.
- His research implied that the genetic constitution of an organism (its genotype) could be divided into hundreds, perhaps thousands of discrete information packets
- The observable outward appearance of an organism (its phenotype) could be subdivided into a large number of discrete physical or chemical traits.

Genetic definitions

- **genotype**: The genetic (alleleic) makeup of an organism with regard to an observed trait. The sum total of sequence variations (polymorphisms and mutations) present in a genome.
- **phenotype**: The observed properties or outward appearance of a trait.

The first genetic map

- Mendel's Second Law (The Law of Independent Assortment) states alleles of one gene sort into gametes independently of the alleles of another gene.
- However, Morgan and his student Sturtevant noticed that for certain traits the probability of having one trait given another was not 50/50 – those traits are genetically linked

http://www.caltech.edu/news/first-genetic-linkage-map-38798

The first genetic map

- Sturtevant realized the probabilities of co-occurrences could be explained if those alleles were arranged on a linear fashion: traits that are most commonly observed together must be locates closest together
- Today genetic maps are routinely generated by measuring the rates of polymorphic markers in large populations of individuals

The Linear Arrangement of Six Sex-Linked Factors in Drosophila as shown by their mode of Association. Sturtevant, A. H. (1913) Journal of Experimental Zoology, 14: 43-59 https://www.nature.com/scitable/content/the-linear-arrangement-of-six-sex-linked-16655

Genetic variations

- Though DNA is stable, the genome is corruptible, in other words, the genetic code can be changed.
- An allele that is present in the great majority of individuals within a species is termed **wild type** (naturally present in large numbers of apparently health organisms).
- **Mutations** are when one allele is converted into another allele or an allele is created. The collection of alleles present in the genomes of all members of a species is the **gene pool** for the species.

Homo/heterozygosity

- The two copies of a gene could convey different, possibly conflicting information. The different versions of a gene is called an allele.
- Organisms with two *identical* alleles of a gene are **homozygous**
- Organisms with two *different* alleles of a gene are **heterozygous**.

Homo/heterozygosity

- When a gene is **heterozygous**, the observed phenotype encoded by one allele of a gene is **dominant** with respect to the phenotype encoded by another allele, the **recessive** one.
- The alleles of some genes may be **co-dominant**, wherein a blend of the two alleles result in a phenotype.
- **Incomplete penetrance** is when a dominant allele is present but the phenotype is not manifested because of the actions of other genes in the organism's genome.

Patterns of inheritance

- Autosomal dominant
- Autosomal recessive
- X-linked dominant
- X-linked recessive
- Mitochondrial
- Non-Medelian (e.g., imprinting)

Evolution

- Evolutionary change happens because of changes in genomes due to *mutations* and *recombination*.
- *Mutations* are rare events, sometimes single base changes, sometimes larger events.
- *Recombination* is how your genome was constructed as a mixture of your two parents.
- Through *natural selection*, favorable changes tend to accumulate in the genome.

Homology

- Evolution motivates *homology* (similarity) search, because different species are assumed to have common ancestors. Thus DNA/amino acid sequences for a given protein (e.g. hemoglobin) in two species or individuals should be more similar the closer the ancestry between them.
- The genetic variation between different people is surprisingly small, perhaps only 1 in 1000 base-pairs.
- Homology searches can often detect similarities between extremely distant organisms (e.g. humans and yeast).

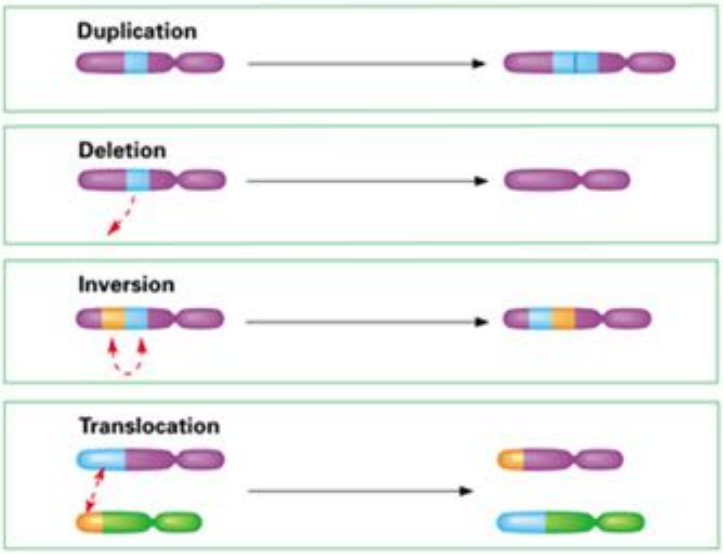
Phylogenies

- *Phylogenic trees* based on gene homologies provide an independent confirmation of those proposed by taxonomists. This is convincing evidence of evolution.
- A host of interesting computational problems arise in trying to reconstruct evolutionary history.

DNA alterations

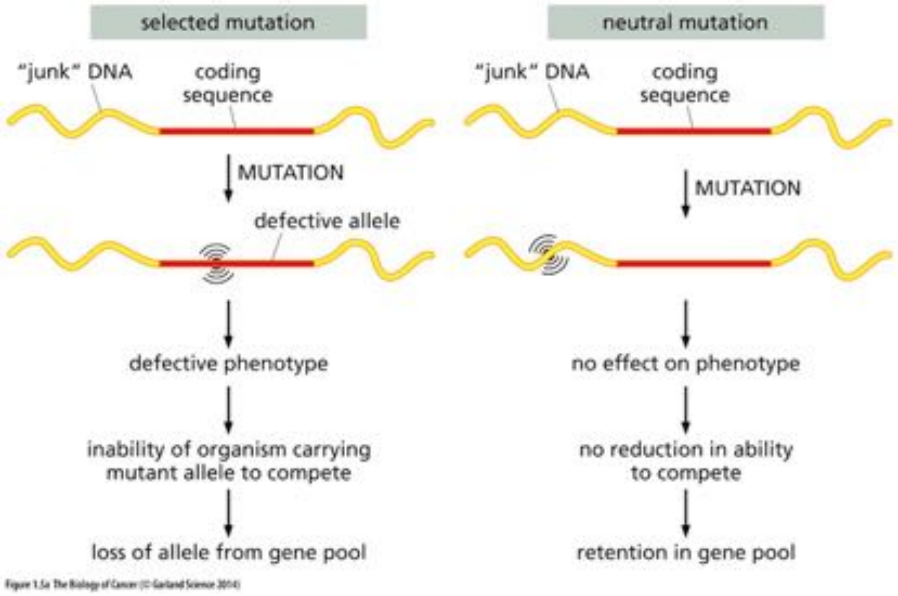
- One base being replaced by another (**substitution**)
- A base being excised (**deletion**)
- A base being added (**insertion**)
- A small subsequence of bases being removed and reinserted in the opposite direction (**inversion**)
- A small subsequence of bases being removed and reinserted in a different place (**translocation**)
- Since DNA is information, and information typically has a beginning point, an inversion would produce an inactive or altered protein.
- Likewise deletion or duplication will alter the gene product.

Changes to chromosome's structure



Neutral mutations are "silent"

but may alter regulatory sites



Germline and somatic mutations

- Transmission of a mutation from one generation to the next, by the germ cells (sperm and egg), is said to occur via the **germ** line.
- Mutations affecting the genomes of cells everywhere else in the body, which constitute the soma, have no prospect of being transmitted to offspring and are called **somatic** mutations.

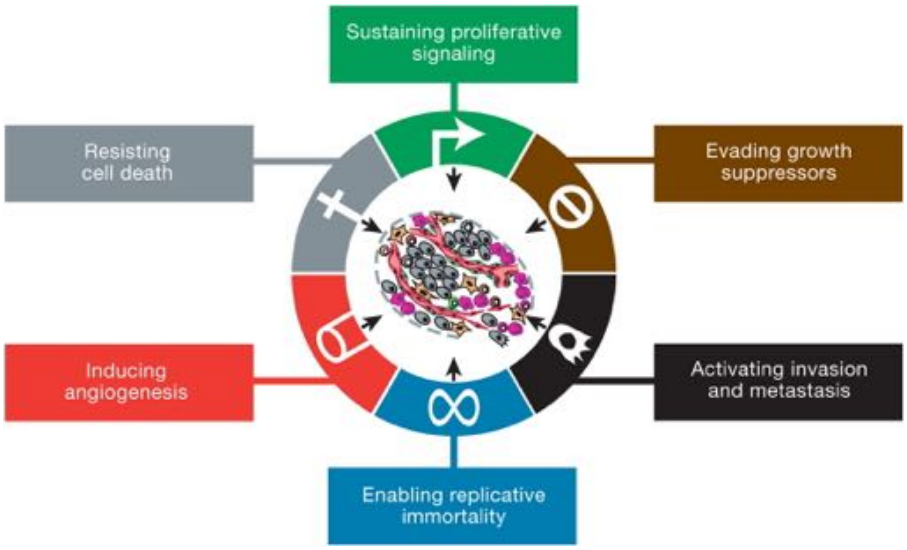
Chromosomal abnormalities

Chromosome disorders
- Congenital (7 per 1000 newborns, 50% of spontaneous first trimester abortions)
- Acquired (cancer)

Single-gene disorders Individually rare
- As a group affect ~ 2% of population over lifespan

Multifactorial or complex disorders
- A result of combination of genes
- May affect ~ 60% of entire population

Hallmarks of cancer



http://www.sciencedirect.com/science/article/pii/S0092867400816839

http://www.sciencedirect.com/science/article/pii/S0092867411001279