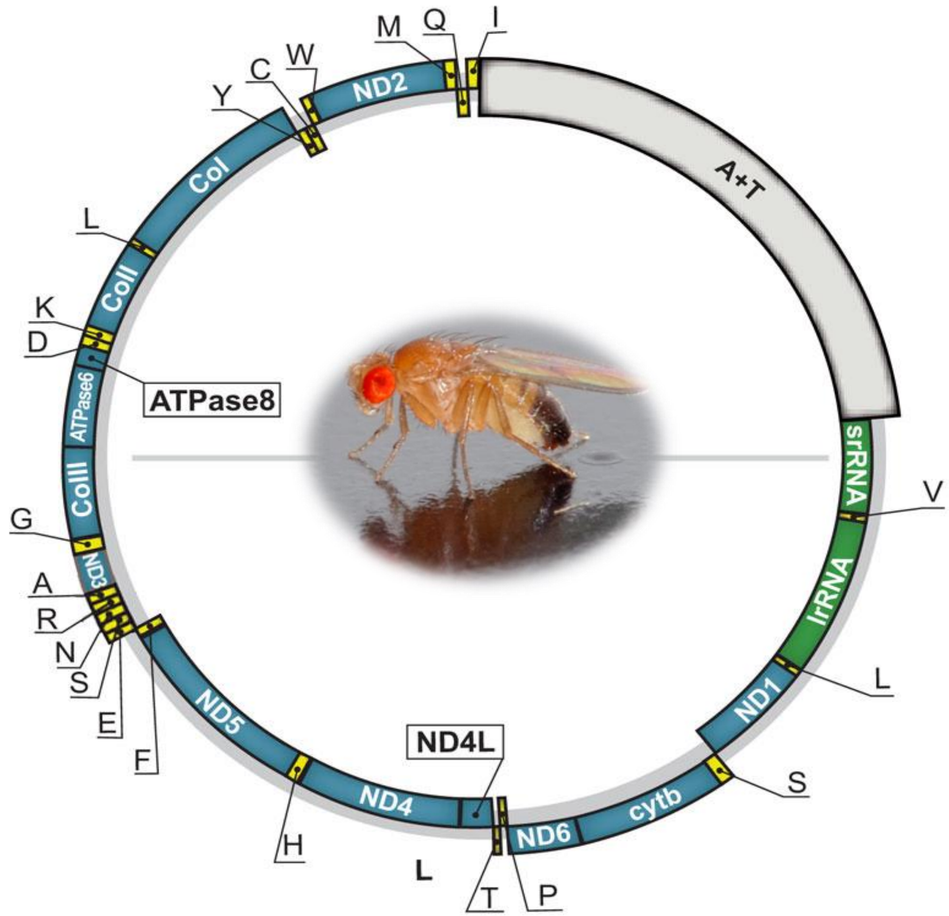
Ent-708

INSECT MOLECULAR BIOLOGY

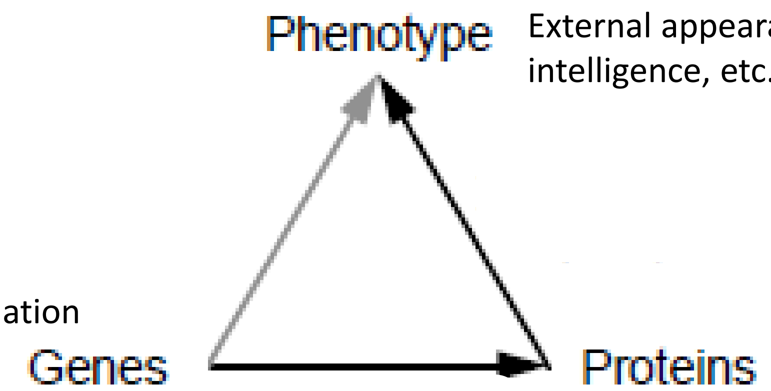


By: Dr. Sajjad Ali

Molecular Biology Origin

> "Molecular studies of the processes of replication, transcription and translation of the genetic (hereditary) material of the organism."

External appearance (size, color, intelligence, etc.)



Genes = digitally encoded information used to synthesis proteins

Polypeptide-complex

Classical genetics: Genes control the transmission of phenotype from one generation to the next.

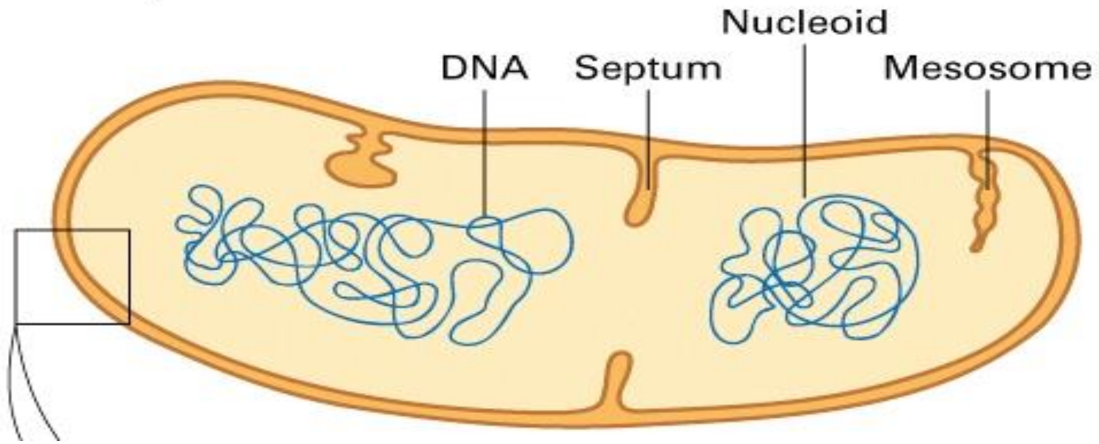
Biochemistry: Within one generation, proteins determine the phenotype.

Gene-Protein relation: How digital (genetic) code directs the creation of life is the goal of **Molecular biology**.

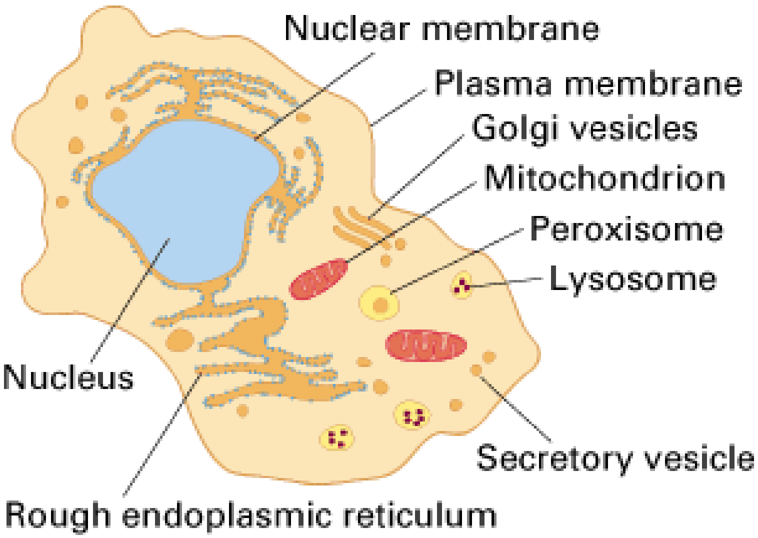
Robert J. Robbins (1995)

(a) Prokaryotic cell

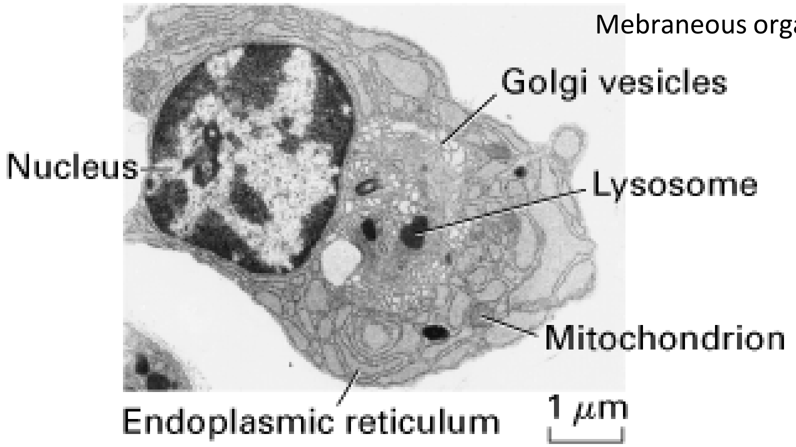
Single-celled organisms
1 micron diameter
1 membrane = single compartment
Single coiled DNA molecule



(b) Eukaryotic cell



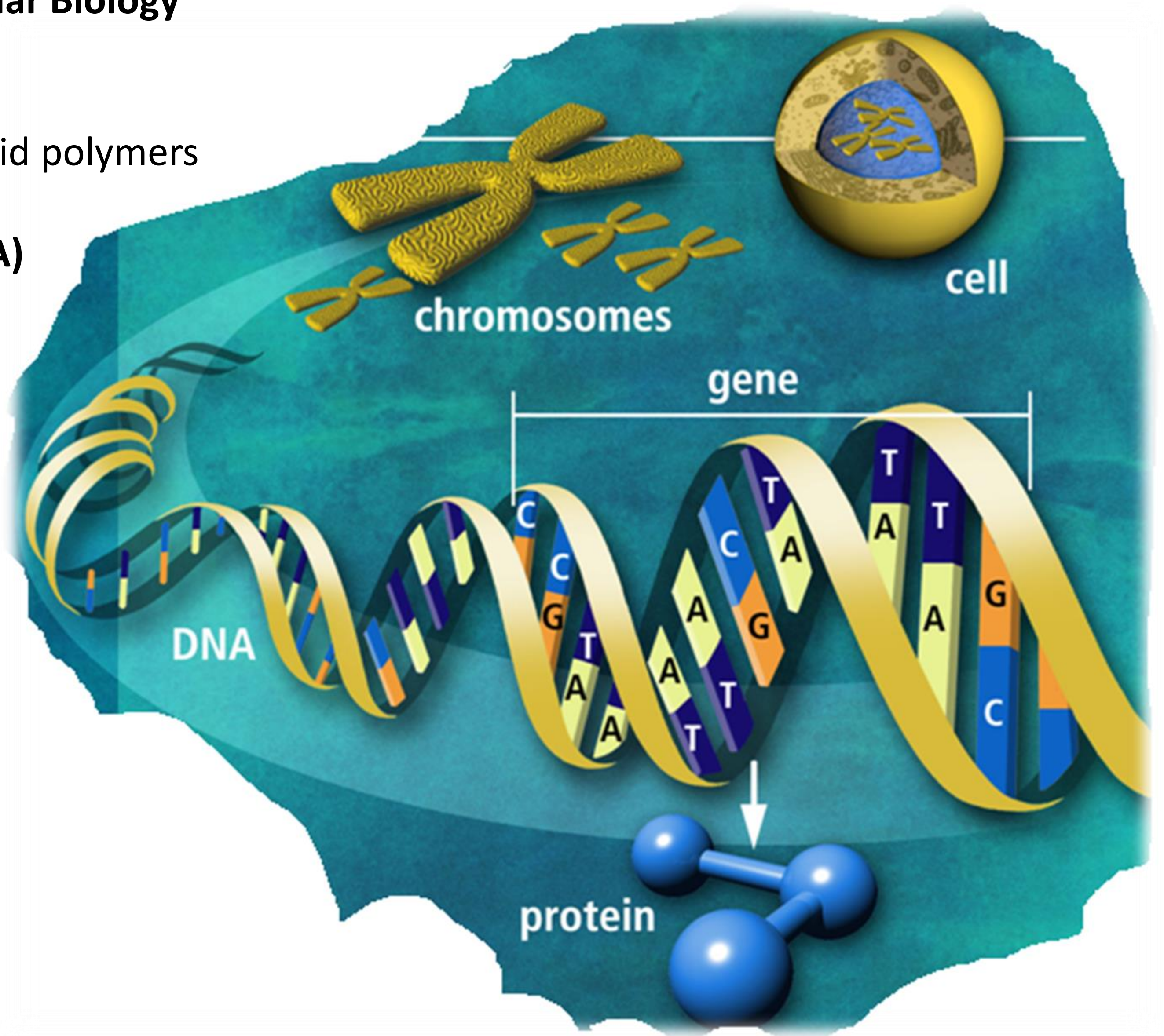
Single-celled organisms/plants/animals
10 micron diameter (variable)
Membrane-bound nucleaus containing DNA
Mebraneous organelles



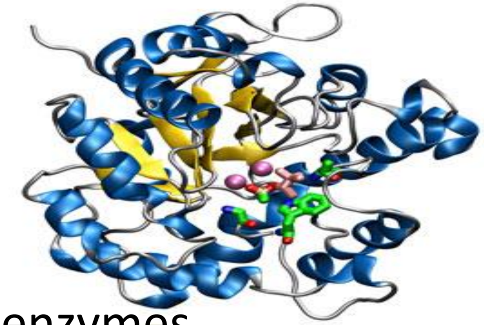
Basic components of Molecular Biology

Two major types of nucleic acid polymers

- **Deoxyribonucleic acid (DNA)**
- **Ribonucleic acid (RNA)**
- **Proteins**



Proteins:



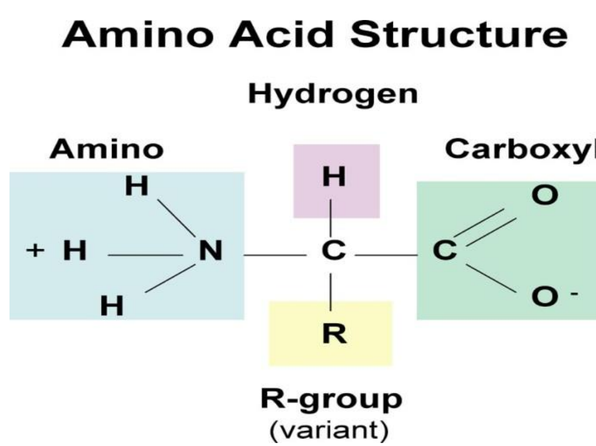
Involved in;

- Rearranging different chemical bonds by acting as enzymes/co-enzymes.
- Carrying signals to and from the outside of the cell, and within the cell.
- Transporting small molecules.
- Forming many of the cellular structures.
- Regulating different cell processes, turning them on and off and controlling their rates.

Protein = 3-dimensional structures of polypeptide chains of amino acids (which make backbones of protein structure) and some co-factors (heme, vitamins)

Polypeptide = linear chains of amino acids

Amino acid = an amine group ($-NH_2$), a carboxylic group and a side chain group (R-group), determining type of AA



20 different amino acids: Each has unique properties.

4 categories (based on major chemical properties);

1. P**ositively charged** (and therefore basic) amino acids (3).

Arginine Arg R   Histidine His H   Lysine Lys K

2. **Negatively charged** (and therefore acidic) amino acids (2).

Aspartic acid Asp D   Glutamic acid Glu E

3. **Polar amino acids** (7). Though uncharged overall, these amino acids have an uneven charge distribution. Because of this uneven charge distribution, these amino acids can form hydrogen bonds with water. As a consequence, polar amino acids are often found on the outer surface of folded proteins, in contact with the watery environment of the cell, in which case they are called **hydrophilic**.

Asparagine Asn N   Cysteine Cys C   Glutamine Gln Q   Serine Ser S   Threonine Thr T
Tryptophan Trp W   Tyrosine Tyr Y

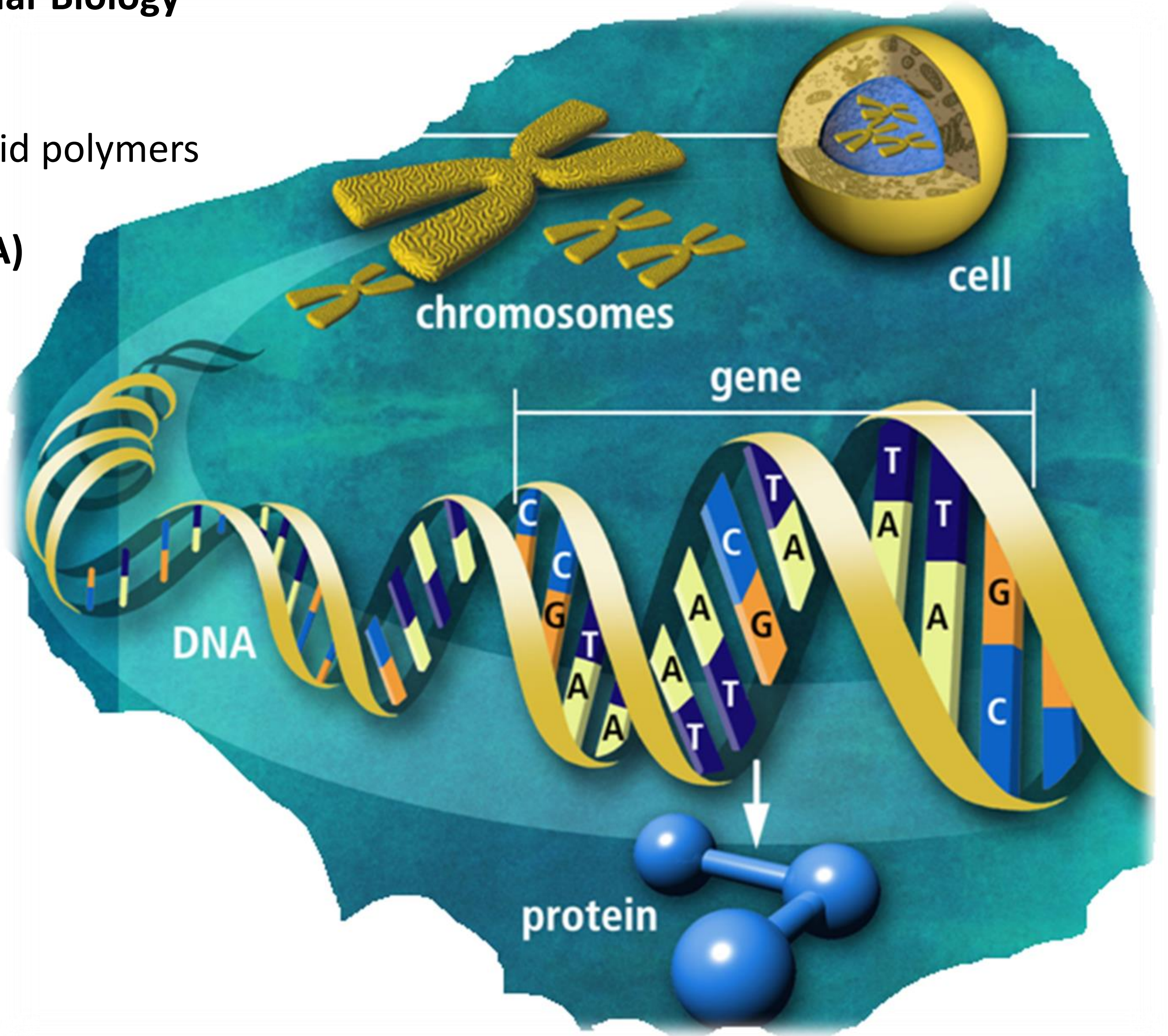
4. **Nonpolar amino acids** (8). These amino acids are uncharged and have a uniform charge distribution. Because of this, they do not form hydrogen bonds with water, and tend to be found on the inside surface of folded proteins, in which case they are called **hydrophobic**.

Alanine Ala A   Isoleucine Ile I   Glycine Gly G   Leucine Leu L   Methionine Met M
Phenylalanine Phe F   Proline Pro P   Valine Val V

Basic components of Molecular Biology

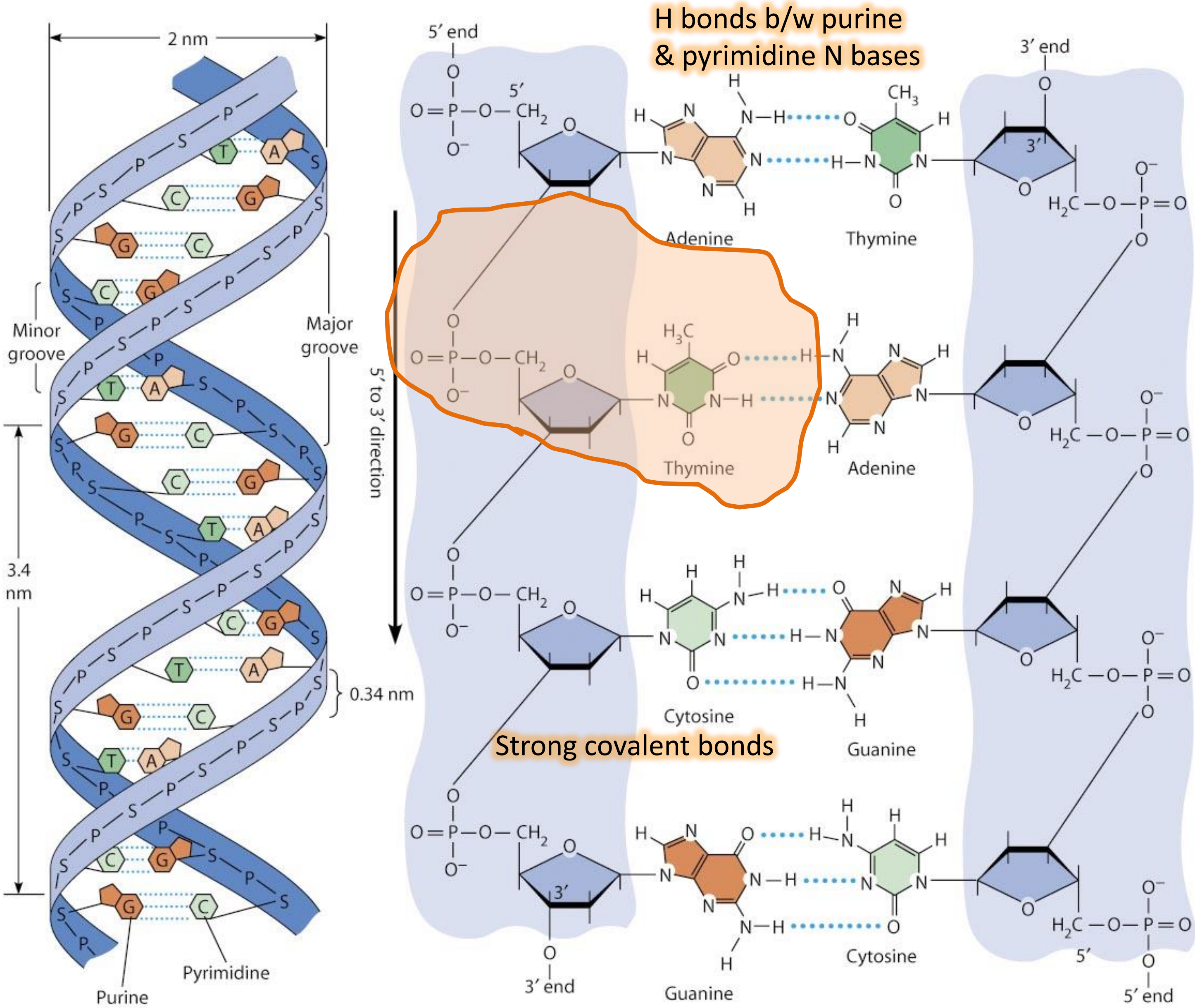
Two major types of nucleic acid polymers

- **Deoxyribonucleic acid (DNA)**
- **Ribonucleic acid (RNA)**
- Proteins



Double-helical structure of DNA

Nucleotides = Building blocks of DNA molecule
Anti-parallel Polynucleotides (complementary)



(a) Double helix (b) Antiparallel orientation of strands

Nucleotide

- A sugar
- A phosphate group
- A nitrogenous base

In polynucleotide
Linear order = No rule
Pairing oder = Complementary rule

A binds with T (U)
G binds with C
(via H-bonds)

More GC content = more thermo-stable

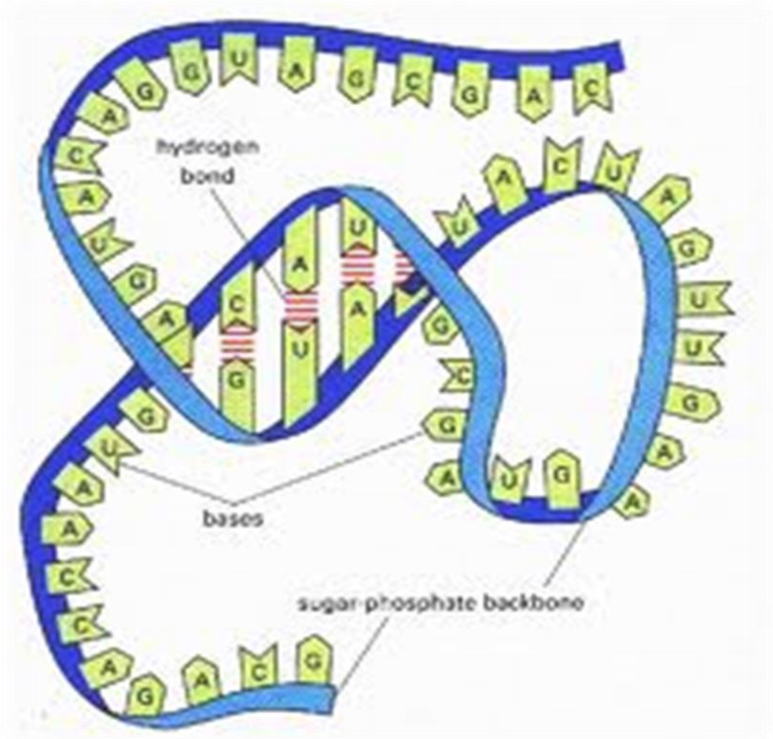
Actinobacteria (*Streptomyces coelicolor)*
G+C content = 72%

Human genome = 41%

Plasmodium sp. = ~ 20%

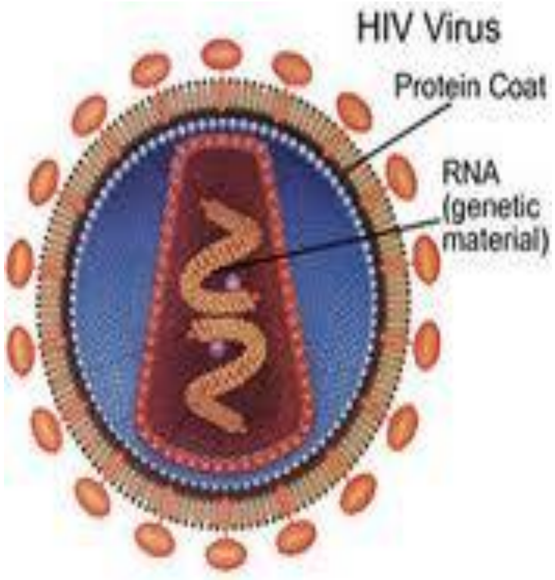
Ribonucleic acid (RNA)

- **RNA** is a biologically important type of molecule that consists of a long chain of nucleotide units.
- Each nucleotide consists of a nitrogenous base, a ribose sugar, and a phosphate.
- Base-pairing capabilities (like DNA) and sometimes form interamolecular hydrogen bonds, leading to a complex 3-dimentional shapes.



Double-stranded RNA

- Double-stranded RNA (dsRNA) is RNA with two complementary strands, similar to the DNA found in all cells.
- dsRNA forms the genetic material of some viruses (double-stranded RNA viruses).

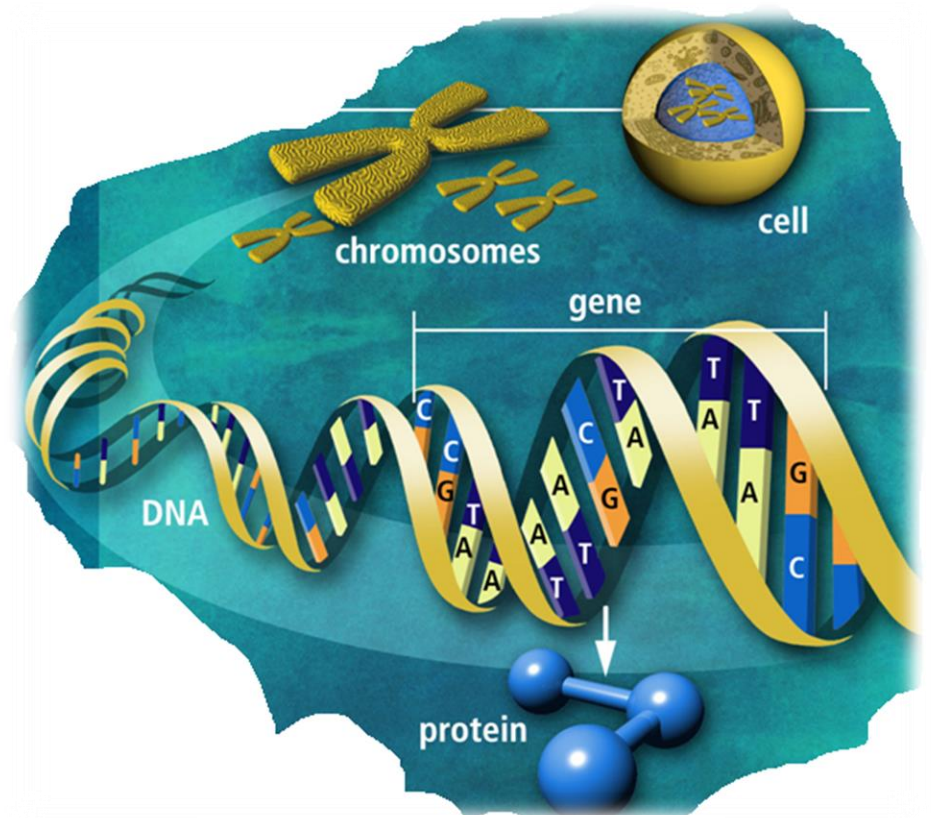


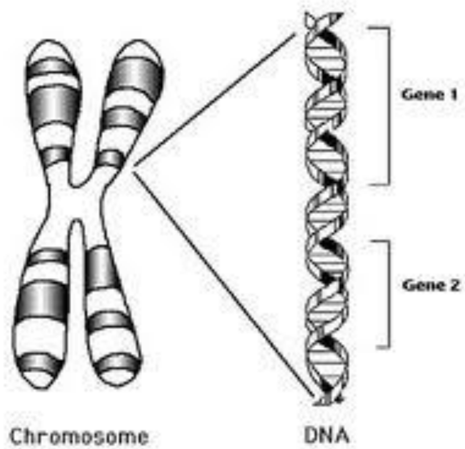
Genes: The units of heredity

•DNA segments that carry the genetic information.

•A stretch (sequence) of <u>DNA</u> that codes for a particular <u>protein (polypeptide)</u> or for a specific RNA type.

•The genetic information to build and maintain an organism's cells and pass genetic traits to offspring.

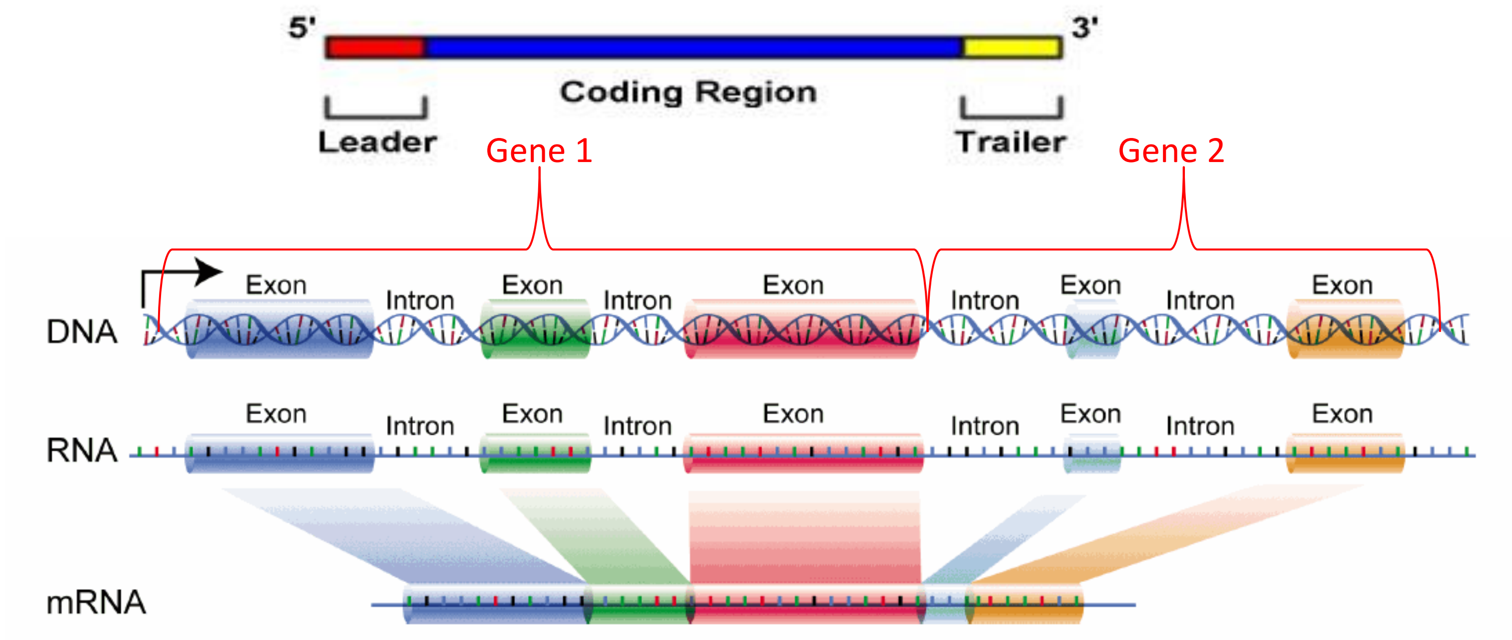




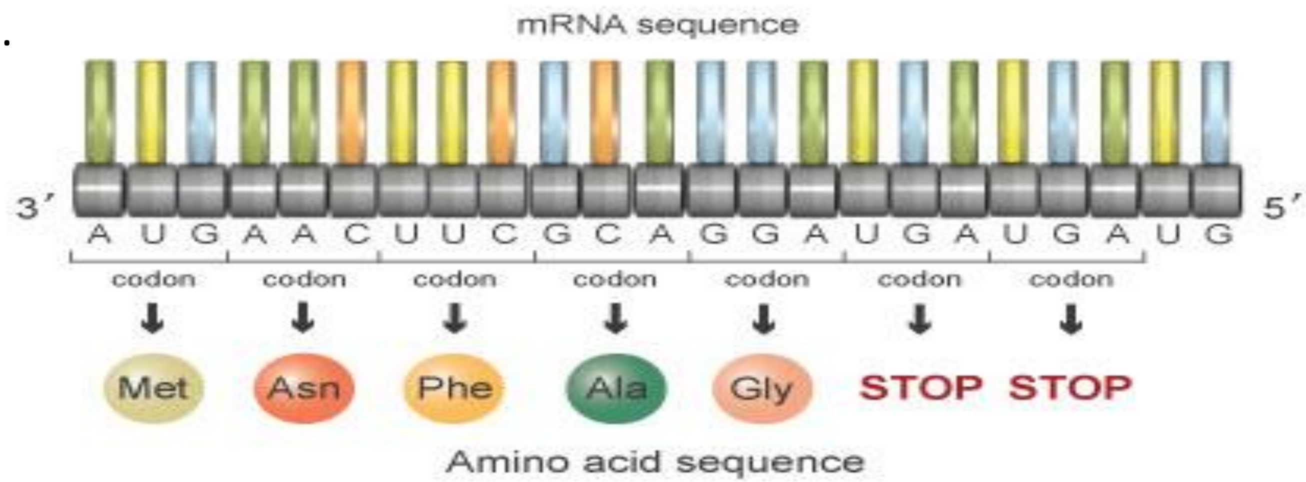
Genes

Fundamentals of Molecular Biology

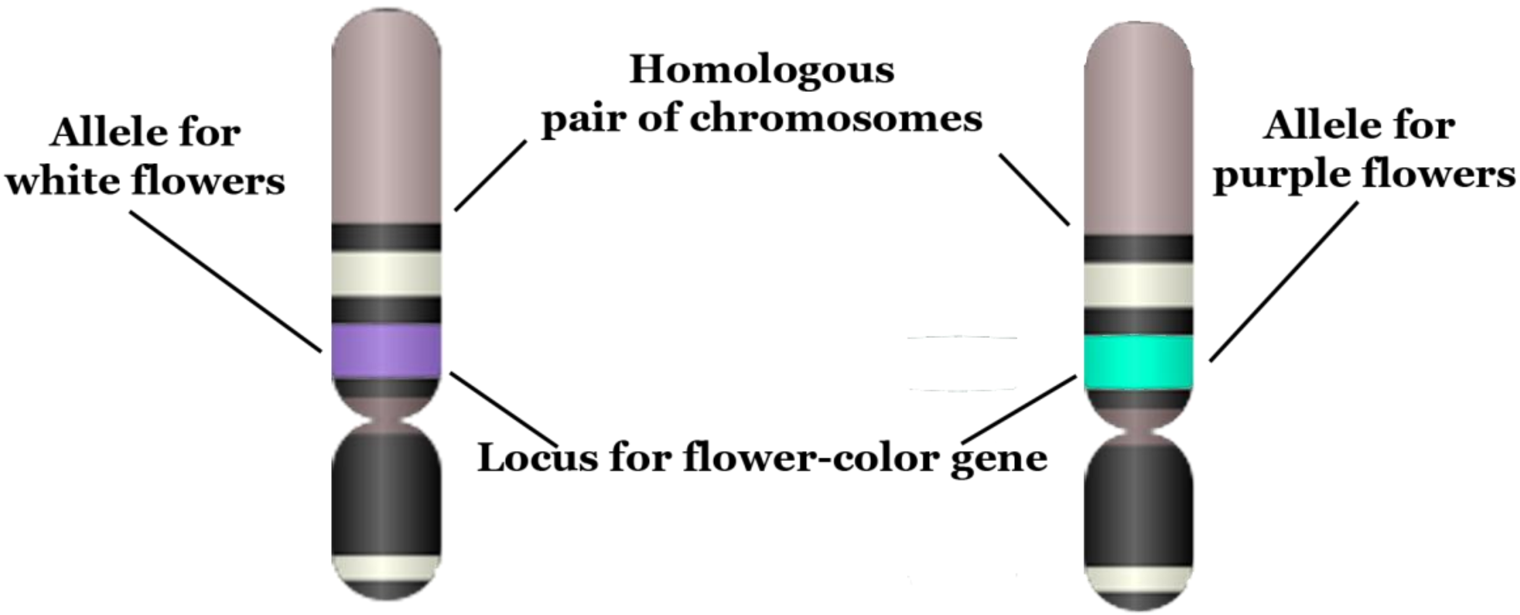
Gene: includes regions preceding and following the coding region (leader and trailer) as well as intervening sequences (introns) between individual coding segments (exons).



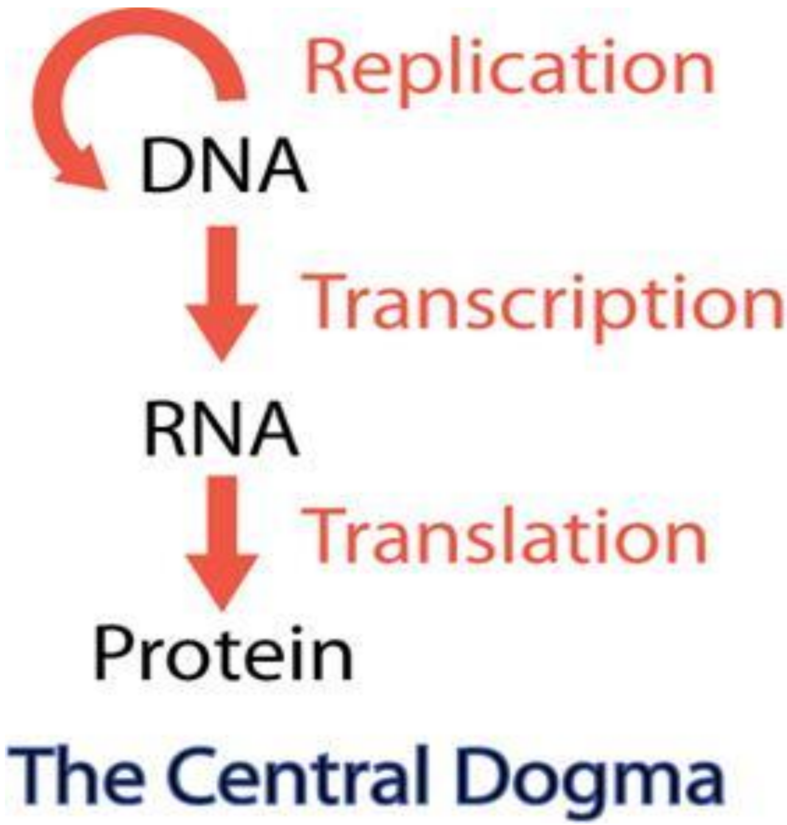
Codon: Three successive nucleotides on an mRNA that encode a specific amino acid in the polypeptide.



Allele: is one of several alternative forms of a gene occupying a given locus on a chromosome.

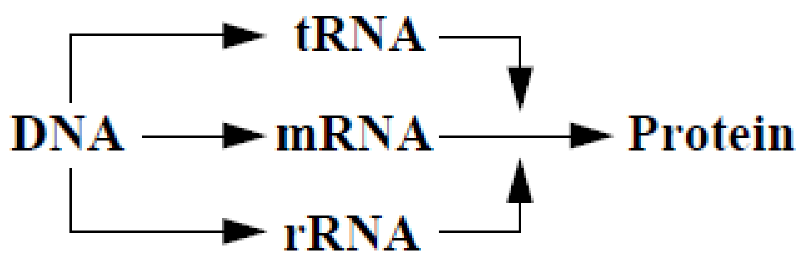


Locus: is the position on a chromosome at which the gene for a particular trait resides; locus may be occupied by any one of the alleles for the gene.


Replication
DNA
Transcription
RNA
Translation
Protein
The Central Dogma

The Fundamental Dogma

Information coded in DNA directs the synthesis of different RNA molecules.

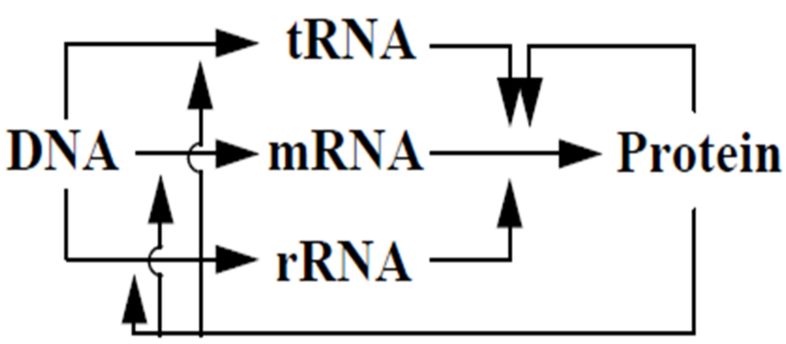


<u>RNA molecule types:</u>

rRNA: *ribosomal RNA that is required for building ribosomes, which* are structures necessary for protein synthesis.

tRNA: *transfer RNA that serves to transfer individual amino acid* molecules from the general cytoplasm to their appropriate location in a growing polypeptide during protein synthesis.

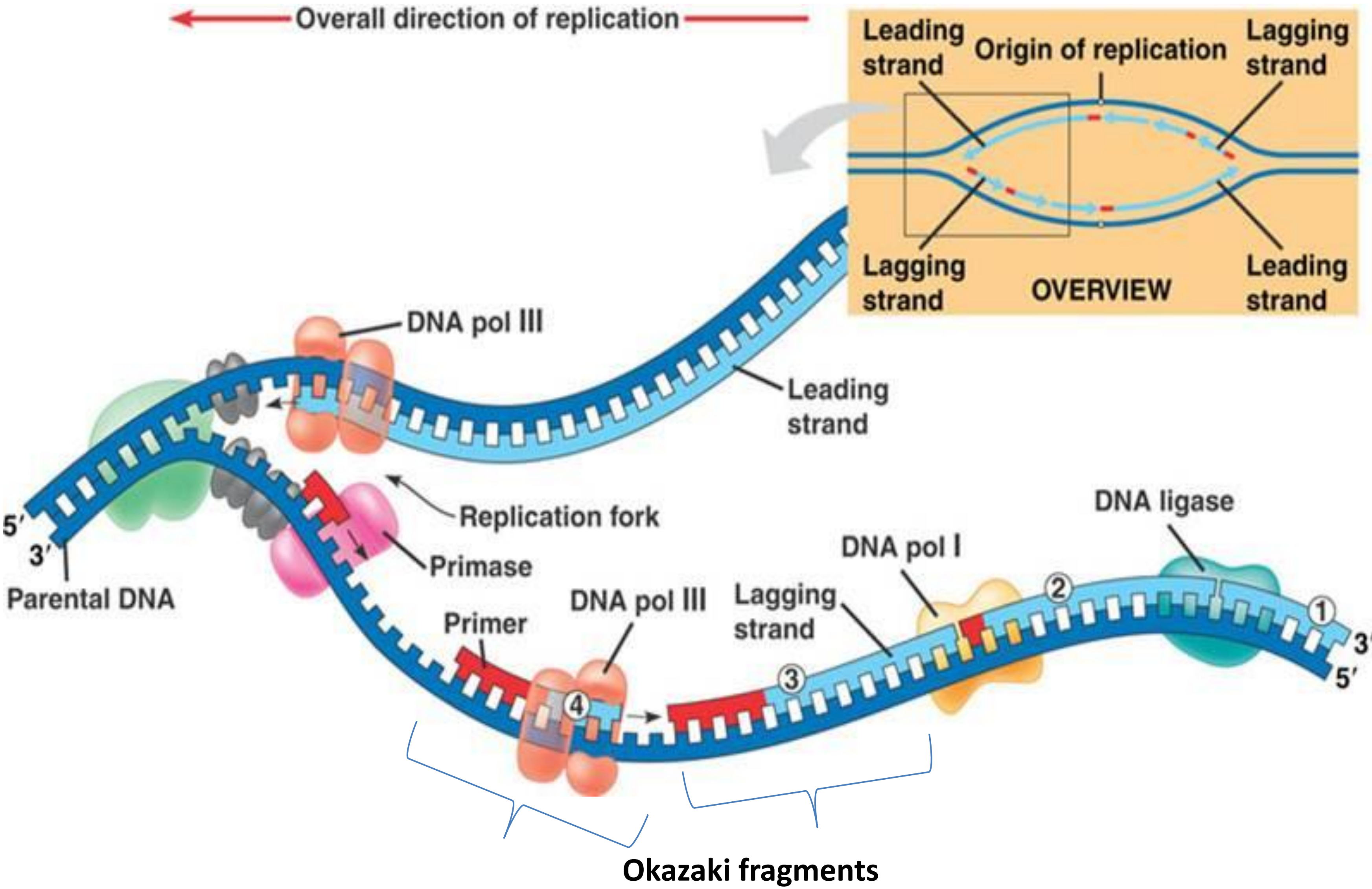
mRNA: *messenger RNA that carries the specific instructions for building* a specific protein.



DNA Replication

- Main enzyme: **DNA polymerase**. Several other enzymes also involved (see below)
- Replication is semi-conservative:
  - DNA helix is opened up and unwound by a helicase.
  - Each old strand gets a new strand built on it.
  - DNA polymerase can only add bases to the 3’ –OH group on a pre-existing nucleic acid that is base-paired with the template strand it is copying. This means that DNA synthesis starts with the enzyme primase synthesizing a short RNA primer. DNA polymerase then adds bases to this primer.
  - DNA polymerase can only add new bases to 3' end, so one strand is synthesized continuously (leading strand) and the other is built up of short fragments: discontinuous synthesis on the lagging strand.
  - The short (100-1000 bp) DNA fragments, called Okazaki fragments, are built in the opposite direction of fork movement and then ligated together (by DNA ligase by formation of covalent bonds).
  - The whole process starts at several points on each chromosome and goes in both directions. Takes 8 hr to complete.

DNA Replication



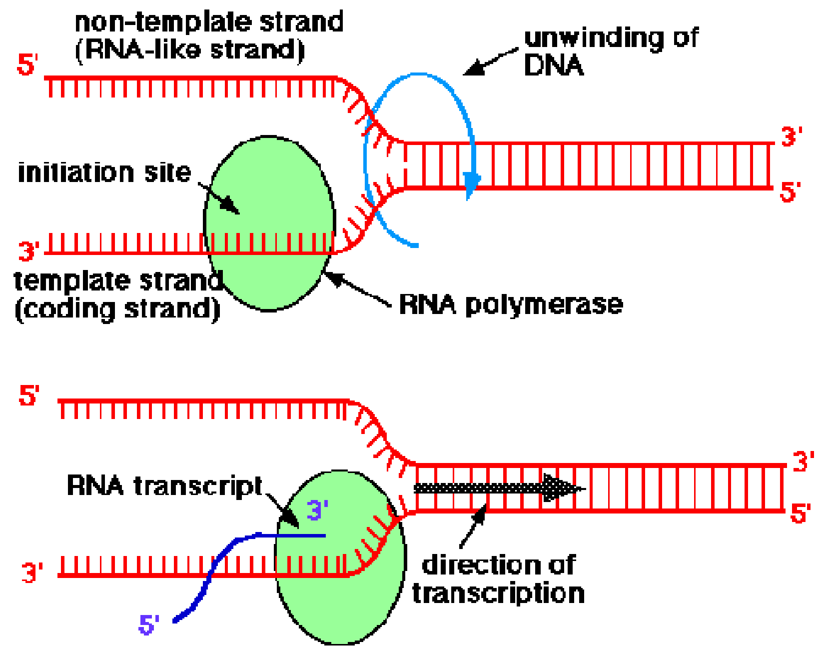
Each protein that an organism can produce is encoded in a piece of the DNA called a "gene"

Nucleic acids Transcription

- Transcription is making an RNA copy of a short region of DNA.
- Only part of the DNA is transcribed. A transcribed region is called a transcription unit, which is approximately equivalent to "gene".
  - most transcription units code for proteins
  - some code for functional RNAs that never get translated into proteins (RNA genes).
- When transcription starts, the DNA double helix is unwound and only one strand is used as a template for the RNA.
  - the template DNA strand is called the antisense strand, and the other DNA strand, not used in transcription is called the sense strand. This is because the sense strand has the same base sequence as the RNA transcript. Gene sequences are generally written as the sense strand.
  - Genes are oriented from 5' to 3' based on transcription direction (even though the template DNA is read 3' to 5'). Thus, 5' end of a gene is where transcription starts. Upstream and downstream also relate to this direction.

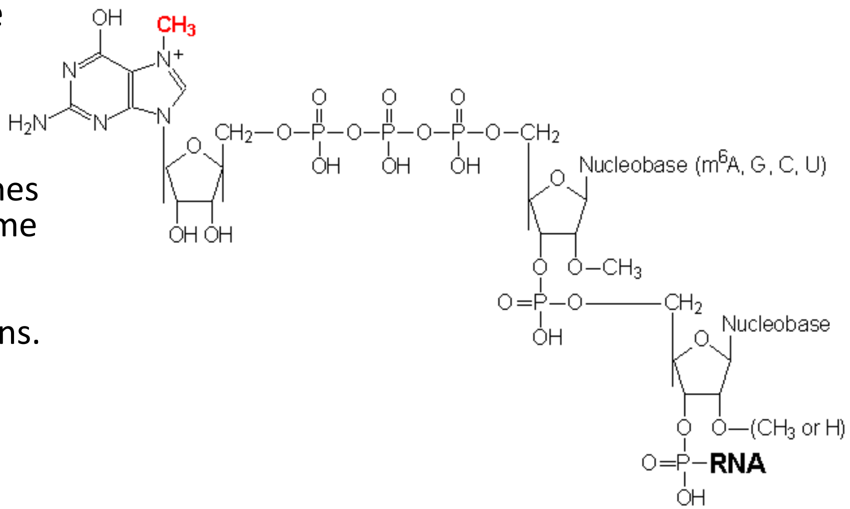
Transcription Process

- The primary enzyme used for transcription is **RNA polymerase**
  - There are 3 types of RNA polymerase: pol2 does the protein coding genes, while pol1 and pol3 do ribosomal RNA and other structural RNA genes.
- RNA polymerase binds to a promoter sequence just upstream from the transcription starting point, with the help of several proteins called transcription factors.
  - some transcription factors are used for all transcriptions, but others are very specific for cell type, hormonal stimulus, developmental time, etc.
- RNA polymerase then moves in a 3' direction, adding new RNA nucleotides to the growing RNA molecule.



RNA Processing

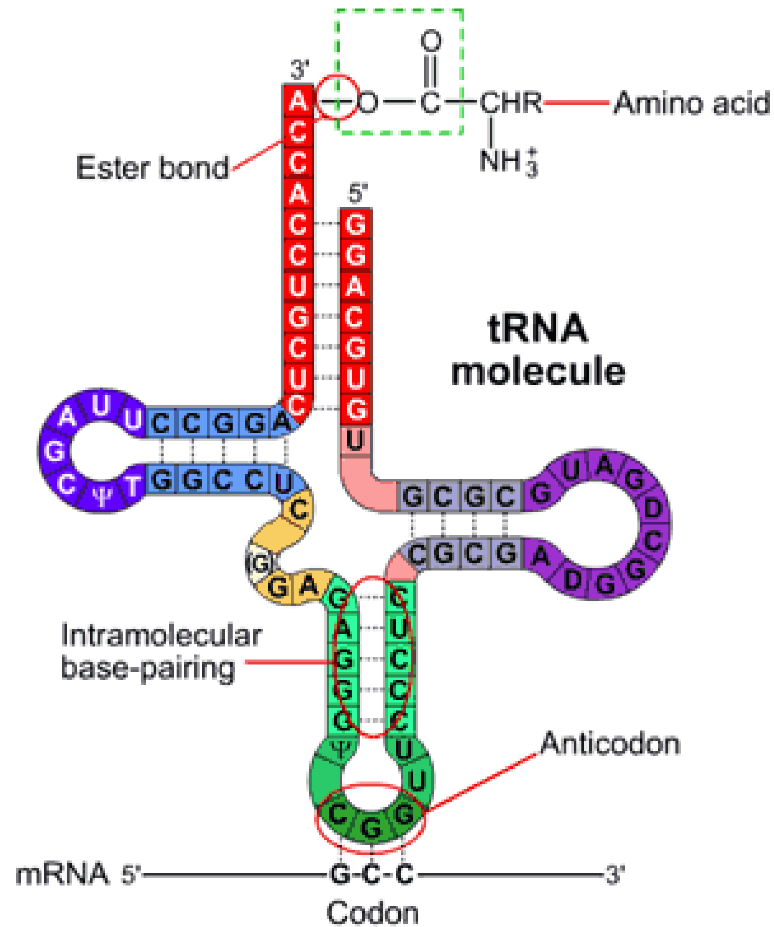
- The RNA molecule that results for RNA polymerase transcribing a gene is called a primary transcript that is exact copy of the DNA. Before it can be translated into protein, it must be processed, then transported to the cytoplasm.
- RNA processing has 3 steps:
  1. Splicing out of introns, which are non-protein coding regions in the middle of protein-coding genes. Most genes are interrupted by introns: up to 99% of the gene in some cases. Exons are the regions of genes that code for protein. Primary transcript contains introns, but spliceosomes (RNA/protein hybrids) splice out the introns. There are signals on the RNA for this, but it can vary between tissues (alternative splicing).
  2. 5' cap: a 7-methyl guanine linked 5’ to 5’ with the first nucleotide of the RNA.
  3. 3' poly A tail: several hundred adenosines added to 3’ end. Signal for poly A marks end of gene, but transcription continues past this without having a definite end point. All except histone genes have poly A. Stability of mRNA is probable reason for it.



- After processing, the RNA is called messenger RNA (mRNA) , and it gets transported to the cytoplasm.

RNA Translation

- In cytoplasm, messenger RNA molecules are translated into polypeptides by ribosomes (RNA/protein hybrids).
- Starts at first AUG in the messenger RNA, goes to first stop codon. (So, only one polypeptide per messenger RNA.) Open reading frame (ORF): region between a start codon and a stop codon.
- Protein is synthesized from N terminus to C-terminus, based on free NH2 and COOH groups on the polypeptide. These correspond to the ribosome moving down the messenger RNA from 5' end to 3' end.
- There are 5' and 3' untranslated regions (UTRs) on the RNA.
- Most mRNA molecules are translated multiple times.
- transfer RNA: short RNA molecules with several modified bases that act as adapters between codons on the mRNA and the amino acids.



Genetic coding

- Three bases of DNA or RNA code for 1 amino acid = codon.
- 4 bases = 4^3 = 64 codons. 61 code for amino acids and the last 3 are stop codons that end the translation process.
- Most amino acids have more than 1 possible codon: Most variation is in third position of codon.
- Nearly all organisms use the same code, with minor variations mostly in mitochondria and chloroplasts.
- All translations start with ATG/AUG, methionine.

mRNA to Amino Acid Dictionary

5´	U	C	A	G	3´
U	phe	ser	tyr	cys	U
	phe	ser	tyr	cys	C
	leu	ser	STOP	STOP	A
	leu	ser	STOP	trp	G
C	leu	pro	his	arg	U
	leu	pro	his	arg	C
	leu	pro	gln	arg	A
	leu	pro	gln	arg	G
A	ile	thr	asn	ser	U
	ile	thr	asn	ser	C
	ile	thr	lys	arg	A
	met	thr	lys	arg	G
G	val	ala	asp	gly	U
	val	ala	asp	gly	C
	val	ala	glu	gly	A
	val	ala	glu	gly	G

Life is specified by Genomes.
Every organism (bacteria, virus, fungi, plants, animals) possess genomes_containing all The biological information required to construct and mainatin a living example of that organism.

Genomes _made up of DNA in most life forms
_of RNA in some viruses (HIV)
DNA polymers_several millions of nucleotides
RNA polymers_rarely few thousands of nucleotides.

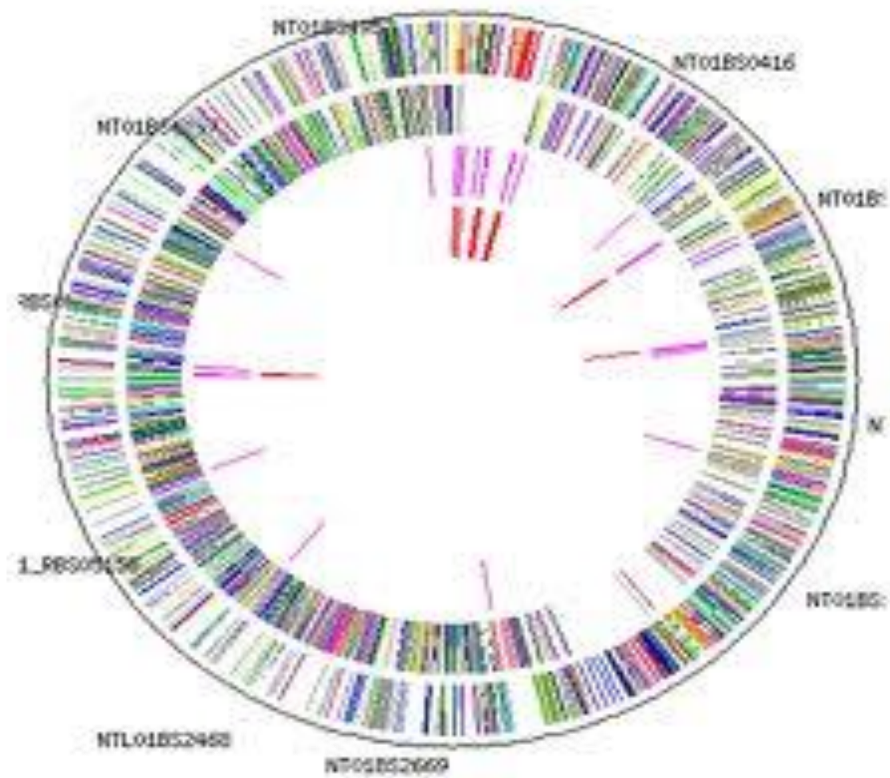
Genetic information_in a genome is encoded in the nucleotide sequences in the form of discrete units (Genes).
Information is translated by gene present at specific locations of a genome into proteins through a serious of biochemical reactions **(Gene expression).**
Gene expression: Transcription and Translation
Simple in simpler organisms like bateria but process is complex in higher organism Due to many points of gene expression pathway at which the information flow is regulated.

A complete copy of a genome develops as the cell divides_Extremely accurate with no muntations
If some occur (in DNA repli., due to certain environmental factors) corrected by DNA repaire enzymes
Those errors which escape become permanent features of lineage descending from that original mutated genome.

- **Genome** is the entirety of an organism's hereditary information.
- It is encoded either in DNA or, for many types of virus, in RNA.
- The genome includes both the genes and the non-coding sequences of the DNA.

A **genome** is “all the DNA contained in an organism or a cell, which includes both the chromosomes within the nucleus and the DNA in mitochondria… all our genes together.” (Source: National Human Genome Research Institute)

Two types of Genomes_Nuclear genome, Mitochondrial genome



Genome size is the total amount of DNA contained within one copy of a single genome.

Typically measured in terms of mass in picograms (trillionths (10^{-12}) of a gram, abbreviated pg)

or as the total number of nucleotide base pairs typically in megabases (millions of base pairs, abbreviated Mb or Mbp). One picogram equals 978 megabases.[1]

In diploid organisms, genome size is used interchangeably with the term C-value. An organism's complexity is not directly proportional to its genome size; some single cell organisms have much more DNA than humans

(The term **C-value** refers to the amount, in picograms, of DNA contained within a haploid nucleus (e.g. a gamete) or one half the amount in a diploid somatic cell of a eukaryotic organism. In some cases (notably among diploid organisms), the terms C-value and genome size are used interchangeably, however in polyploids the C-value may represent two or more genomes contained within the same nucleus.)

Eukaryotic genome size = 10 Mb to 100,000 Mb
Simple and smallest genome = fungi
Large and complex genomes = vertebrates and flowering plants

Comparative genome sizes of organisms

organism	Size (Mbp)	gene number	average gene density	chromosome number
Homo sapiens (human)	3000	~50-80,000	1 gene /100,000 bases	46
Mus musculus (mouse)	2600	~25,000	1 gene /100,000 bases	40
Drosophila melanogaster (fruit fly)	140	13,000	1 gene / 9,000 bases	8
Arabidopsis thaliana (plant)	100	25,000	1 gene / 4000 bases	10
Bombyx moori	43	19,000	1 gene / 5000 bases	12
Saccharomyces cerevisiae (yeast)	12.1	6000	1 gene / 2000 bases	32
Escherichia coli (bacteria)	4.6	3200	1 gene / 1400 bases	1
H. influenzae (bacteria)	1.8	1700	1 gene /1000 bases	1

Genomics is “the study of functions and interactions of all the genes in the genome, including their interactions with environmental factors.

(Source: Collins, Francis, and Alan Guttmacher. “Genomic Medicine—A Primer,” NEJM, Vol. 347:1512-1520.)

The human Genome Project (1990 -2003)

- A public-private collaborative, supported by the National Institutes of Health and the U.S. Department of Energy, that is mapping all humangenes.
- The final draft of the genome was published in April 2003.

IN HUMAN GENOME

Nuclear genome = 3000000000 bp = 3000000 kbp = 3000 Mbp = 300 million bp = 50,000 genes

24 linear DNA molecules (chromosomes, 22 autosomes), 2 sex chromosomes) = shortest molecule = 55 Mbp, longest = 250 Mpb

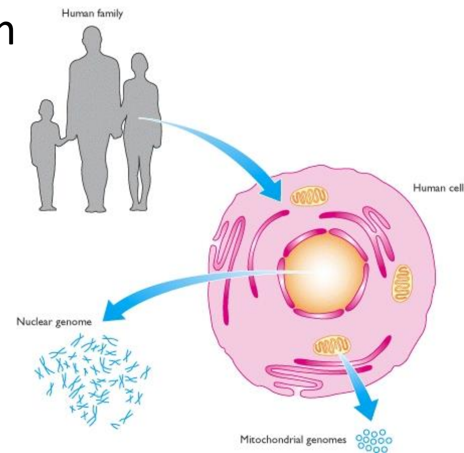
Mitochondrial genome = circular DNA molecule of 16569 bp = 16.5 Kbp (m single mitochondria) = 8000 copies = 10 copies per mitochondria

Each body cell_its own genome copy (except RBCs which lack nucleous)

Somatic cells (diploid chromosomes number)

Gametes (haploid chromosome number)

Human genome length = 5000 Km = 3000 copies of book in hand



Human nuclear genom = only 3% carries genetic information
Example:
50 Kb segment of Chromosome 7 = specifies human beta-T-cell receptor locus” of acutal size 685 Kb, involved in synthesis of proteins in the immune response

1. **One gene (TRY4)** coding Trypsinogen (inactive precursor of digestive ezyme pepsin)
2. **Two gene segments** (V28 and V29-1) coding for Beta T-cell receptor proteins
3. **One pseudogene (TRY5)** non-functional copy of gene
4. **52 genome-wide repeat sequences** (covers 39% of the segment)
5. **2 Microsatellites** (sequences containing short motifs like GAGAGAGAGAGA and TATTTATTTATT)
6. **50% non-genic** single copy DNA of no known function

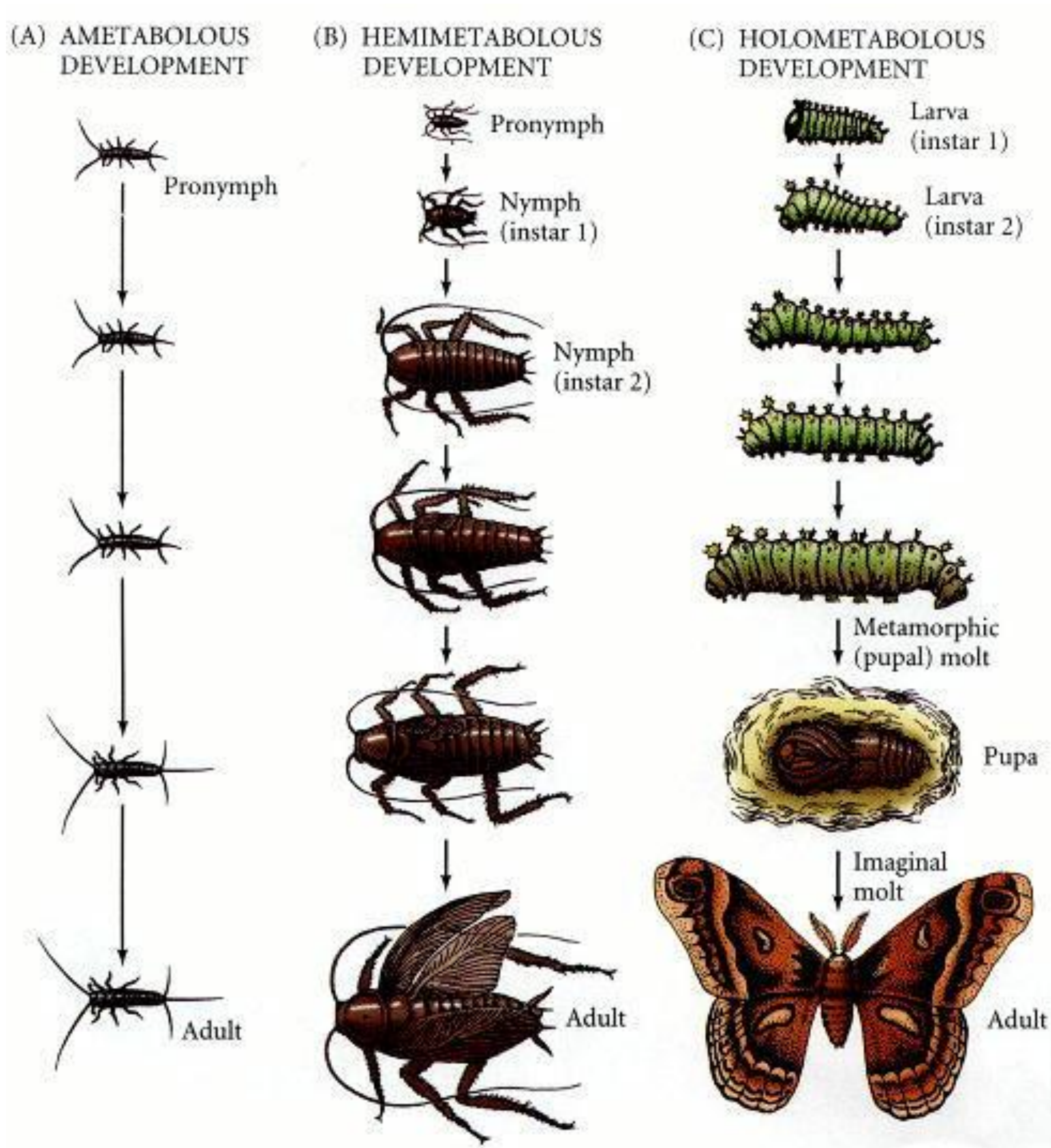
Why Genome analysis ?

Benefit of Genomics

- The prediction of genes in uncharacterised genomic sequences.
- To obtain the complete sequences of as many genomes as possible.
- For Genetic modification (Genetic engineering, recombinant DNA techn.)
- Genetic modification to develop new varieties quickly as BT cotton and BT brinjal.

- **Genetic testing**
  - To detect mutations
    - For disease diagnosis and prognosis
    - For the prediction of disease risk in individuals or families
- Several hundred genetic tests are in use.
  - Rare genetic disorders (muscular dystrophies, cystic fibrosis, Huntington's disease)
  - Complex conditions (breast, ovarian, and colon cancers)

Molecular Basis of Insect Metamorphosis

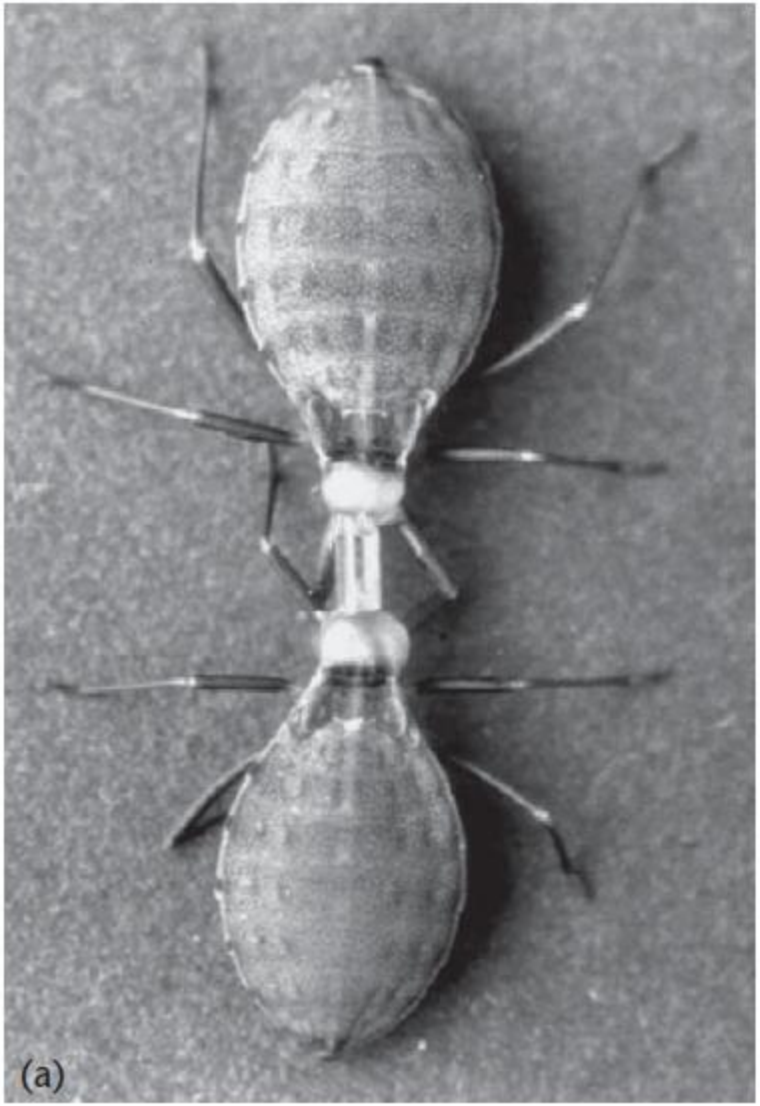


English entomologist
(1899-1994)

Elaboration of role of hormones in insect growth, metamorphosis, and reproduction

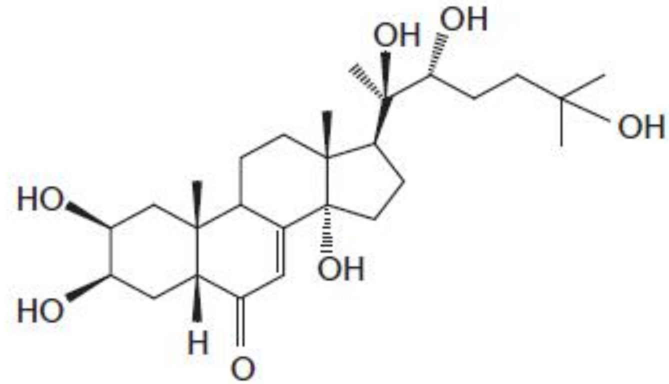
Endocrine control of metamorphosis _ Parabiosis experiment

Transporting Factors

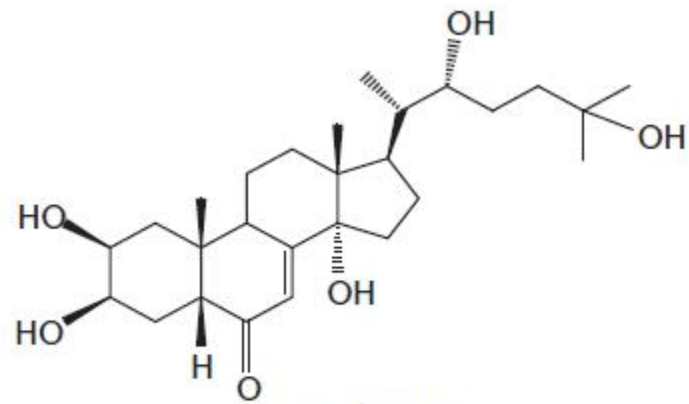


(a)

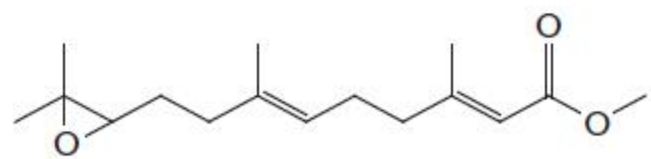
Kissing bug (*Rhodnius prolixus*)



20-Hydroxyecdysone



Ecdysone

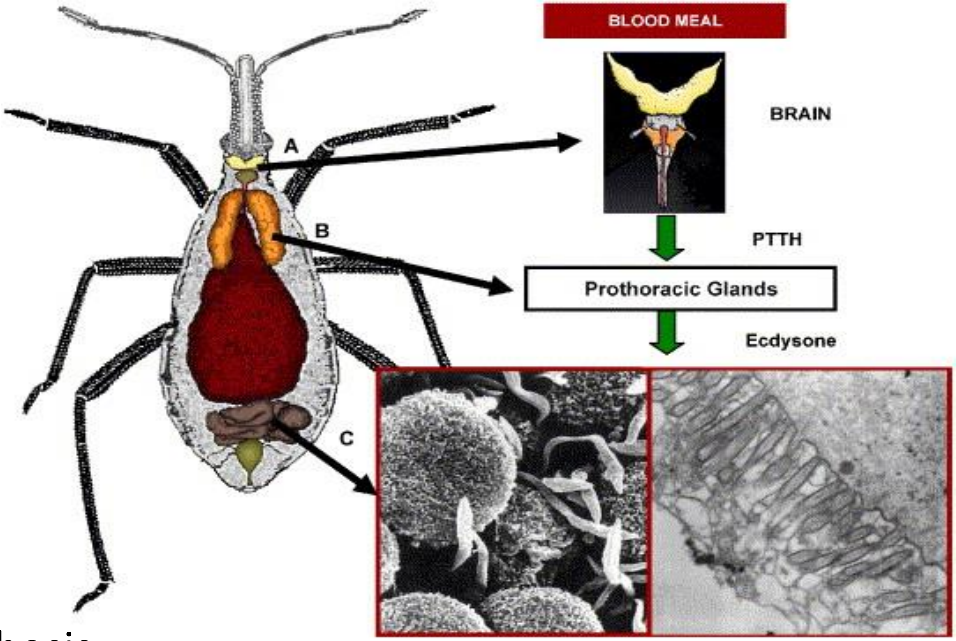


(b) Juvenile hormone III

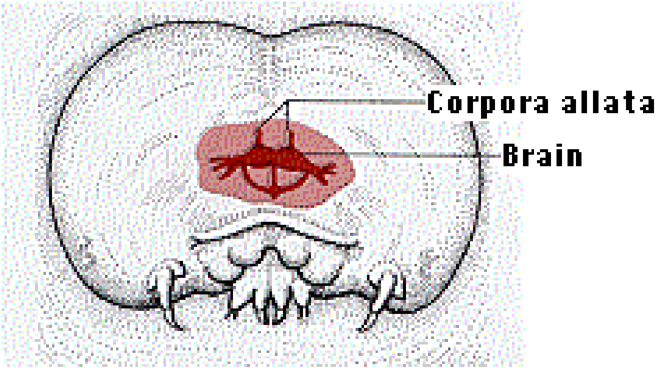
Moulting Hormone (20E)_an ecdysteroid

Produced by prothoracic glands

Regulated by prothoracicotropic hormones (PTTH)



Moulting _ a basic process mediating metamorphosis



Juvenile Hormone (JH)_a terpenoid

Produced by Corpora allata

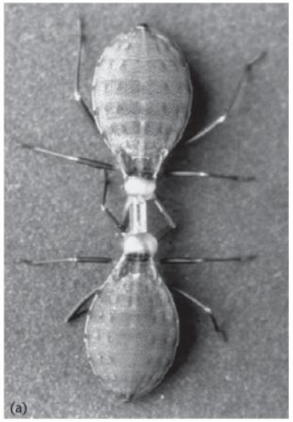
Regulated by allatotropic and allatostatic peptides

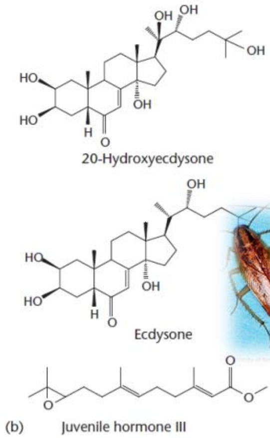
A repressing agent of metamorphosis

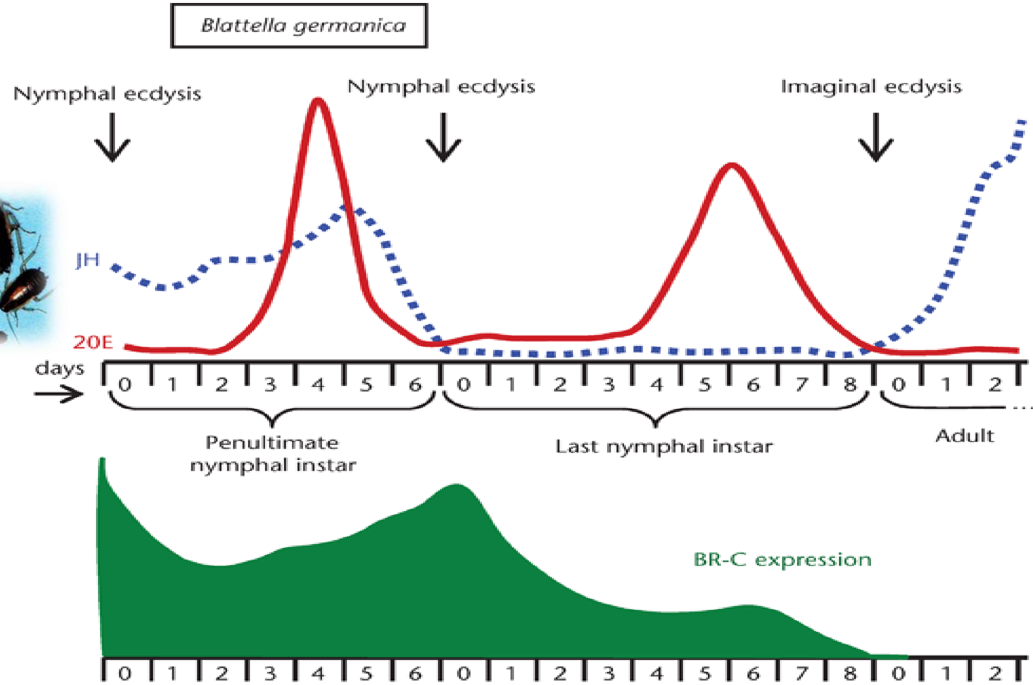
Peptide hormones

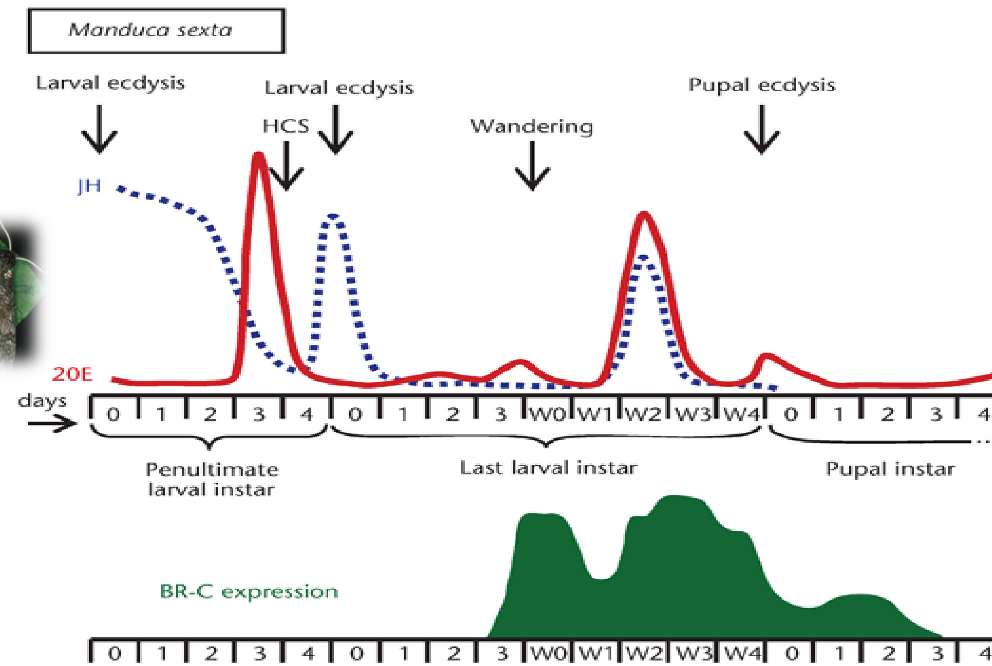
Produced by Inka cells (in tracheal systems)

Controls ecdysis process


(a)


(b)
20-Hydroxyecdysone
Ecdysone
Juvenile hormone III


Blattella germanica
Nymphal ecdysis
Nymphal ecdysis
Imaginal ecdysis
JH
20E
days
0 1 2 3 4 5 6 0 1 2 3 4 5 6 7 8 0 1 2
Penultimate nymphal instar
Last nymphal instar
Adult
BR-C expression
0 1 2 3 4 5 6 0 1 2 3 4 5 6 7 8 0 1 2


Manduca sexta
Larval ecdysis
Larval ecdysis
HCS
Wandering
Pupal ecdysis
JH
20E
days
0 1 2 3 4 0 1 2 3 W0 W1 W2 W3 W4 0 1 2 3 4
Penultimate larval instar
Last larval instar
Pupal instar
BR-C expression
0 1 2 3 4 0 1 2 3 W0 W1 W2 W3 W4 0 1 2 3 4

Molecular control of metamorphosis

Ecdysteroid 20-hydroxyecdysone (20E)

Ecdysone receptor (EcR)
Ultraspiracle (USP)

Signaling and activation

Genetic cascade of transcription factors
(HR3, HR4, HR39, E75, E78, FTZ-F1, etc.)

Expression of the target genes underlying the cellular changes associated to moulting and metamorphosis

Juvenile Hormone (JH)

JH receptor (Met)

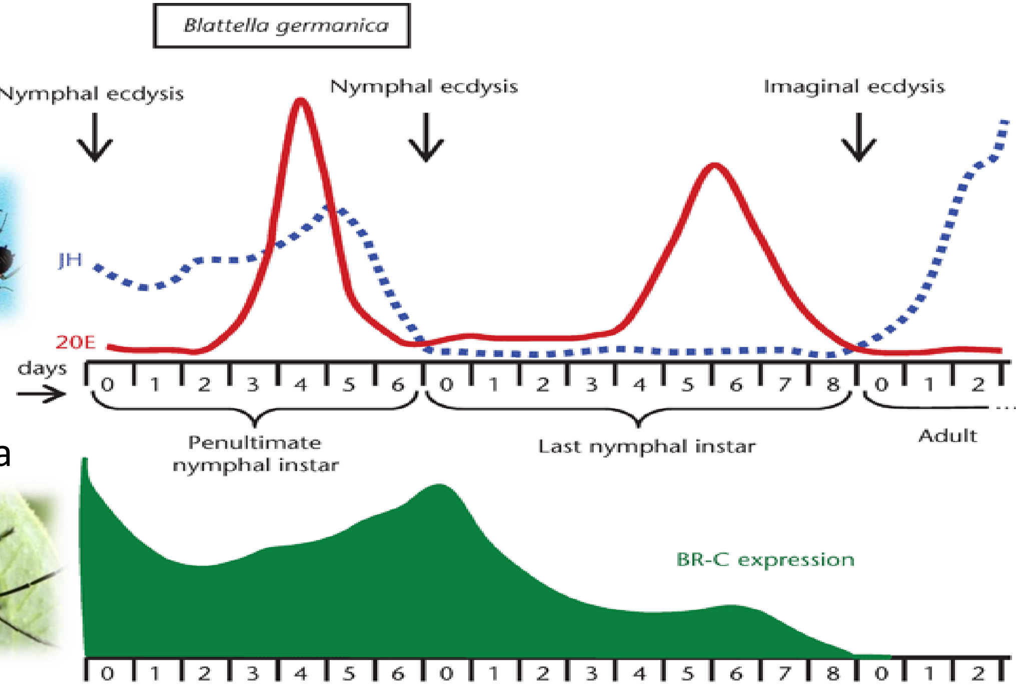
Signaling and activation

Methoprene tolerant (Met)
KrÜppel homolog 1 (Kr-h1)

Micro ribonucleic acids (RNAs)_22 nucleotide length
Regulation of mRNA repression
Confirmed in B. germinaca through gene silencing (RNAi)

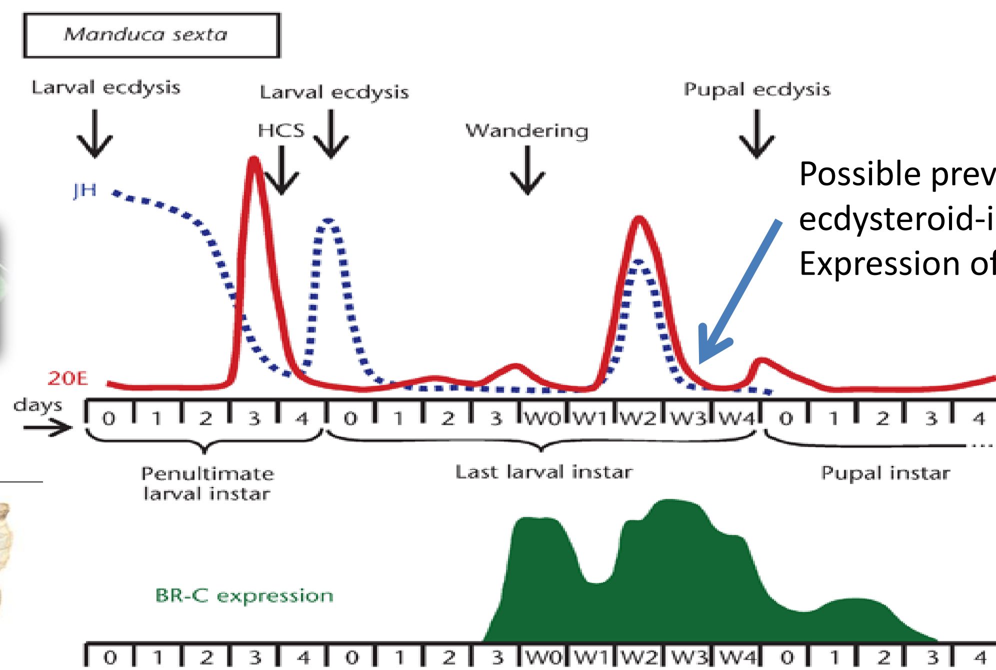
Broad-Complex Transcription Factors

An key family of transcription factors (BR-C)



In holometabolans:
Regulation of larval trnasition to pupa

In hemimetabolans:
Regulation of morphogenesis during nymphal moults



Possible prevention of ecdysteroid-induced Expression of BR-C (by JH)

function in the adult moult