

Graduate Computational Genomics

02-710 / 10-810 & MSCBIO2070

Elements of Molecular Biology

Takis Benos

Lecture #6a, February 1, 2007

Reading: hand-outs



Sequence analysis (6 lectures)

- A little biology...
- ...and statistics (conditional, Markov chains, HMMs)
- Biological sequence “matchmaking”
 - Evolution of DNA and protein sequences - Distances
 - Pairwise and multiple sequence analysis
 - Algorithms for database search
- Gene finding
- DNA motif discovery
 - *cis*-regulatory motifs and modules
 - microRNA genes



Sequence analysis (6 lectures)



- What we will **not** talk in these lectures
 - Genome sequencing assembly
 - Clustering/classification (K-means, SVMs, etc)
 - RNA folding

Benos 02-710/MSCBIO2070 1-FEB-2007

3

Outline of the biology part



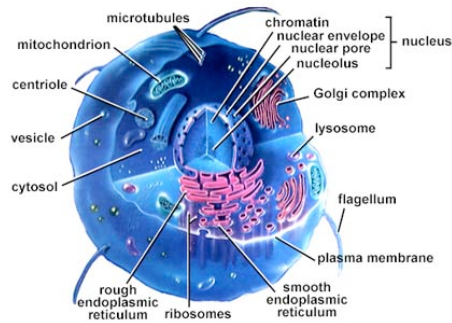
- Basic Definitions
- Cells' basic components
- Basic characteristics of DNA & Proteins
- Transcription and Translation: Central Dogma
- Other Features of Genetic Sequence
- Molecular Evolution

Benos 02-710/MSCBIO2070 1-FEB-2007

4

Cells' components

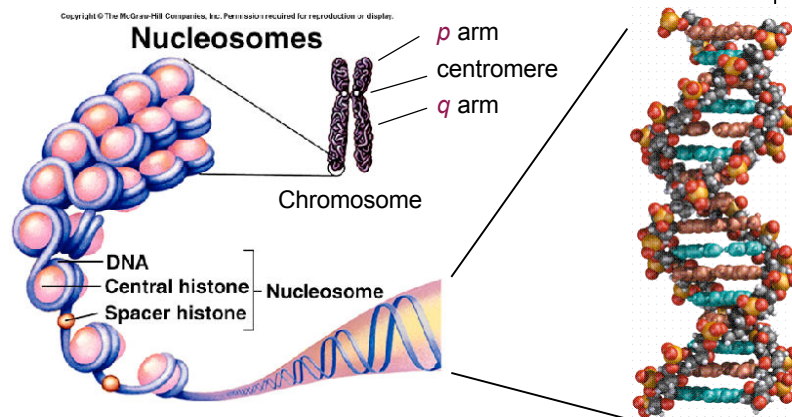
- Cells are complex
- We look at a simplified version:
 - Extracellular environment
 - Membrane
 - Cytoplasm
 - Nucleus (in eukaryotes)



Benos 02-710/MSCBIO2070 1-FEB-2007

5

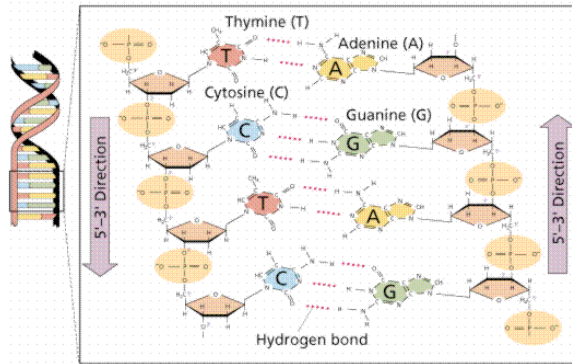
DNA - Chromosomes - Genes



Benos 02-710/MSCBIO2070 1-FEB-2007

6

DNA - Chromosomes - Genes



5' - **A T C G G T** - 3'
| | | | |
3' - **T A G C C A** - 5'
↓
5' - **A C C G A T** - 3'

Benos 02-710/MSCBIO2070 1-FEB-2007

7

What is a “gene”?

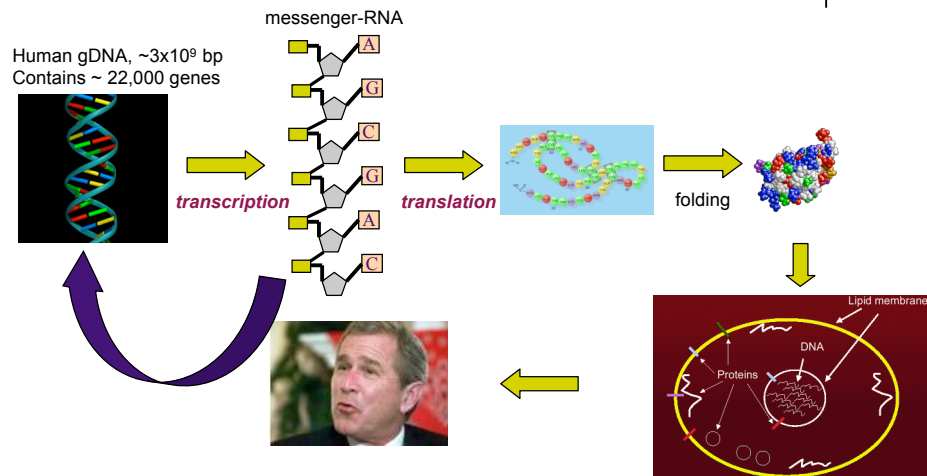
- We cannot define it (but we know it when we see it...)
- A loose definition:

“Gene” is a DNA information unit that is able to perform a *function* in a cellular *environment*

Benos 02-710/MSCBIO2070 1-FEB-2007

8

Central Dogma (and beyond)

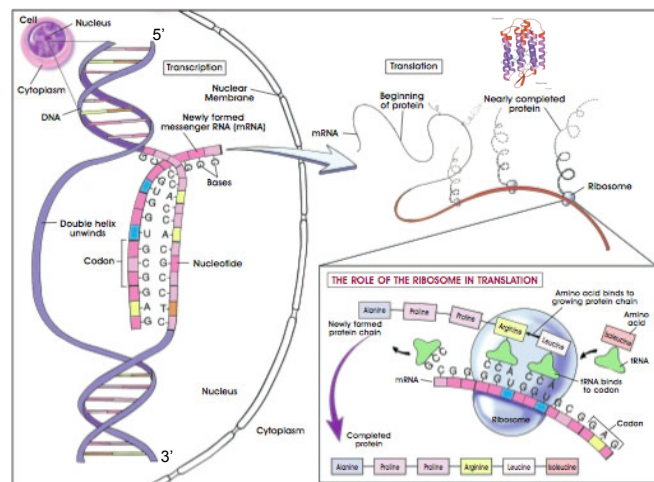


Benos 02-710/MSCBIO2070 1-FEB-2007

Slide courtesy: Serafim Batzoglou

9

Protein coding genes



Benos 02-710/MSCBIO2070 1-FEB-2007

10

Genes and Proteins

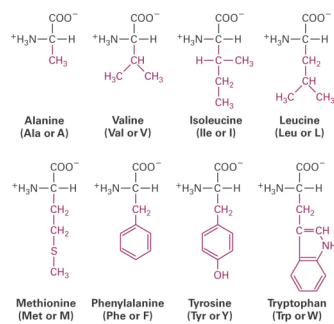
		Second letter				
		U	C	A	G	
First letter	U	UUU Phenyl-alanine UUC UUA Leucine UUG	UCU Serine UCC UCA UCG	UAU Tyrosine UAC UAA Stop codon UAG Stop codon	UGU Cysteine UGC UGA Stop codon UGG Tryptophan	U C A G
	C	CUU Leucine CUC CUA CUG	CCU Proline CCC CCA CCG	CAU Histidine CAC CAA Glutamine CAG	CGU Arginine CGC CGA CGG	U C A G
	A	AUU Isoleucine AUC AUA AUG Methionine; initiation codon	ACU Threonine ACC ACA ACG	AAU Asparagine AAC AAA Lysine AAG	AGU Serine AGC AGA Arginine AGG	U C A G
	G	GUU Valine GUC GUA GUG	GCU Alanine GCC GCA GCG	GAU Aspartic acid GAC GAA Glutamic acid GAG	GGU Glycine GGC GGA GGG	U C A G

Benos 02-710/MSCBIO2070 1-FEB-2007

11

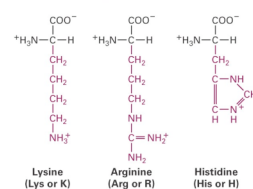
Amino acid varieties

HYDROPHOBIC AMINO ACIDS

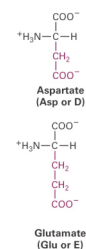


HYDROPHILIC AMINO ACIDS

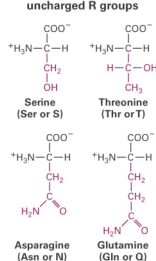
Basic amino acids



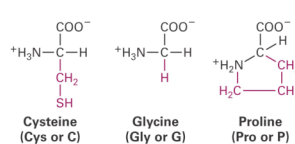
Acidic amino acids



Polar amino acids with uncharged R groups



SPECIAL AMINO ACIDS

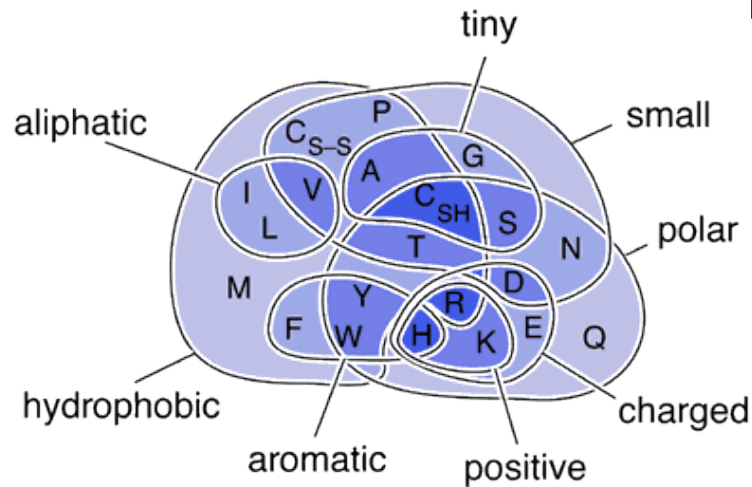


Benos 02-710/MSCBIO2070 1-FEB-2007

12

Slide courtesy: Serafim Batzoglou

Amino acid varieties (cntd)

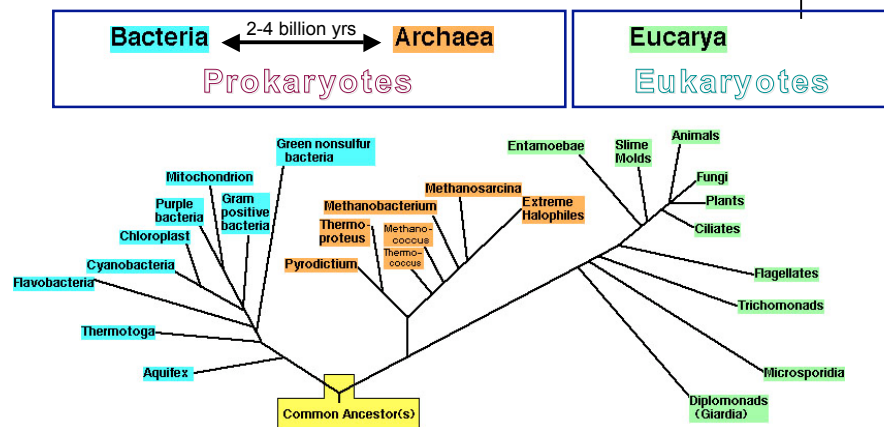


Benos 02-710/MSCBIO2070 1-FEB-2007

Slide courtesy: Serafim Batzoglou

13

Categories of living organisms



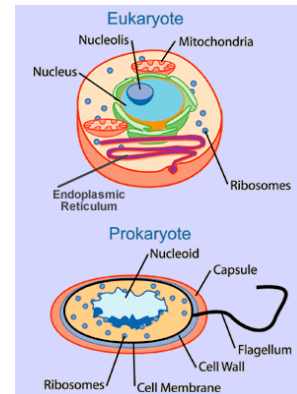
Source: <http://www.bmb.psu.edu/courses/micro401/default.htm>

Benos 02-710/MSCBIO2070 1-FEB-2007

14

Prokaryotes vs. Eukaryotes

	<i>Prokaryotes</i>	<i>Eukaryotes</i>
Cell	Cell walls, no nucleus/organelles	Cell membranes, nucleus/organelles
mRNA	Polycistronic	Monocistronic (mostly)
Gene	Simple (no introns, short UTRs)	Complex
Transcription	Simple	Complex



Benos 02-710/MSCBIO2070 1-FEB-2007

15

Genome logistics: viruses and prokaryotes

	Organism	Size (bp x 10 ⁶)	No. of prot. genes
	<i>HIV-1</i>	0.1	8
	<i>λ phage</i>	0.05	71
	<i>E. coli</i>	4.7	3,200
	<i>H. influenza</i>	1.8	1,700

Benos 02-710/MSCBIO2070 1-FEB-2007

16

Genome logistics: eukaryotes

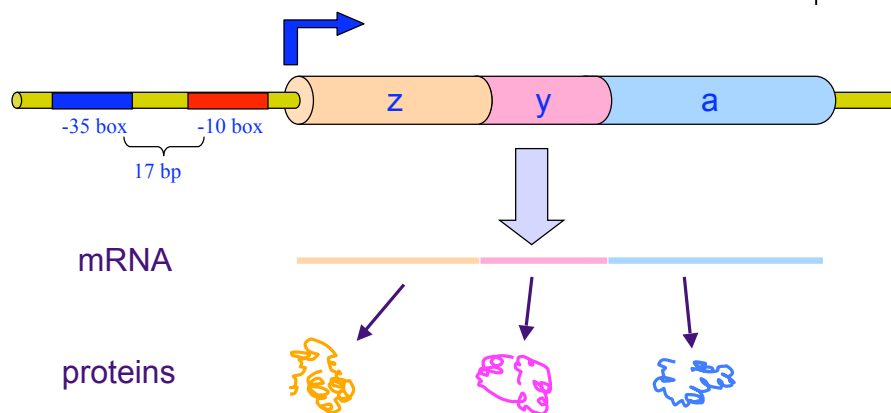


	Organism	Size (bp x 10 ⁶)	No. of prot. genes
	<i>S. cerevisiae</i>	12	6,300
	<i>C. elegans</i>	97	19,100
	<i>A. thaliana</i>	125	25,500
	<i>D. melanogaster</i>	180	13,600
	<i>H. sapiens</i>	2,900	25,000

Benos 02-710/MSCBIO2070 1-FEB-2007

17

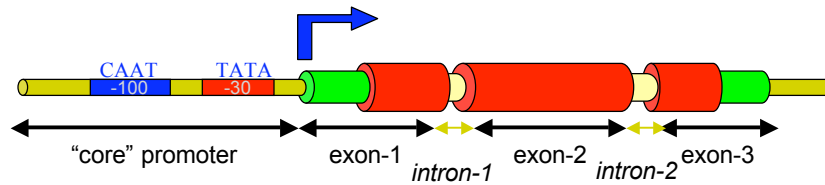
Gene structure: prokaryotes



Benos 02-710/MSCBIO2070 1-FEB-2007

18

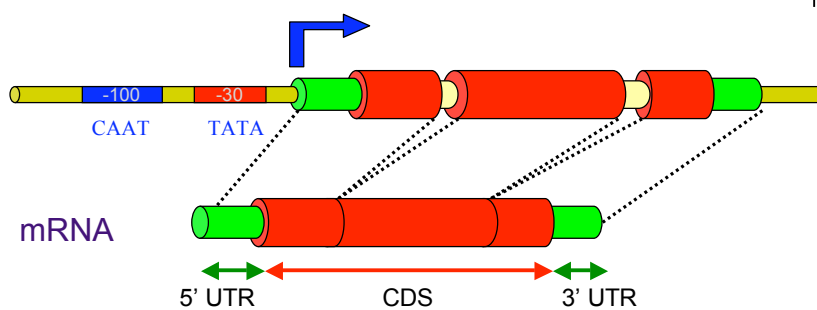
Gene structure: eukaryotes



Benos 02-710/MSCBIO2070 1-FEB-2007

19

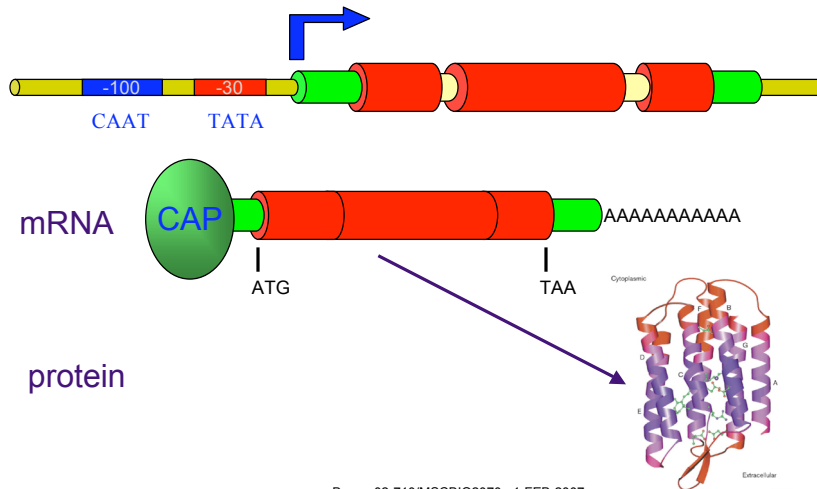
RNA Splicing



Benos 02-710/MSCBIO2070 1-FEB-2007

20

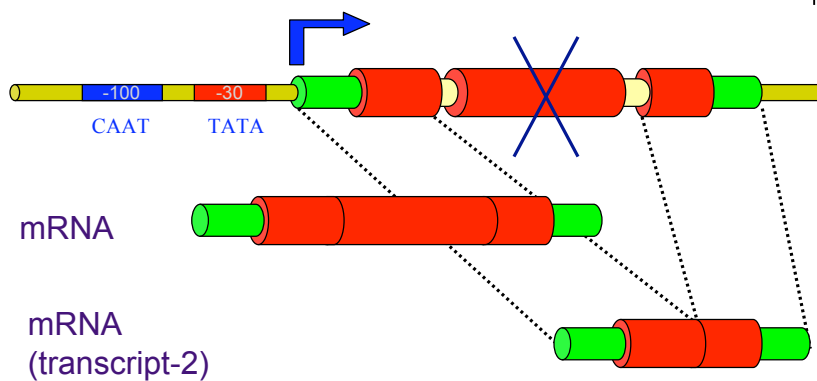
Translation



Benos 02-710/MSCBIO2070 1-FEB-2007

21

Alternative splicing

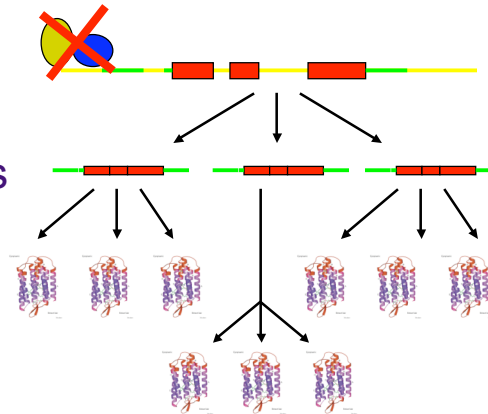


Benos 02-710/MSCBIO2070 1-FEB-2007

22

Transcription regulation

- promoter region
- expression levels
- degradation
- post modifications



Benos 02-710/MSCBIO2070 1-FEB-2007

23

Transcription regulation

- promoter region
- expression levels
- degradation
- post modifications

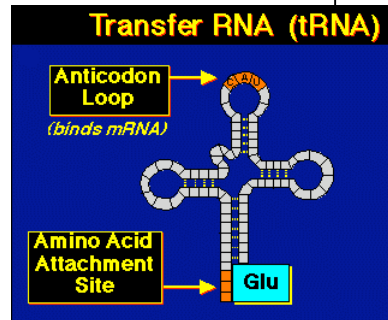


Benos 02-710/MSCBIO2070 1-FEB-2007

24

Non-coding genes

- tRNA^(*)
- ribosomal RNA^(*)
- snoRNA^(*)
- microRNA
- etc



Source: <http://www.emc.maricopa.edu/faculty/farabee/BIOBK/BioBookPROTSYn.html>

^(*)Not to be discussed in this course.

Benos 02-710/MSCBIO2070 1-FEB-2007

25

Other DNA elements

- transposable elements^(*)
- repetitive DNA^(*)
- “junk” DNA^(*)

^(*)Not to be discussed in this course.

Benos 02-710/MSCBIO2070 1-FEB-2007

26