

# Bioinformatics for Environmental Sciences

# What is this class about?

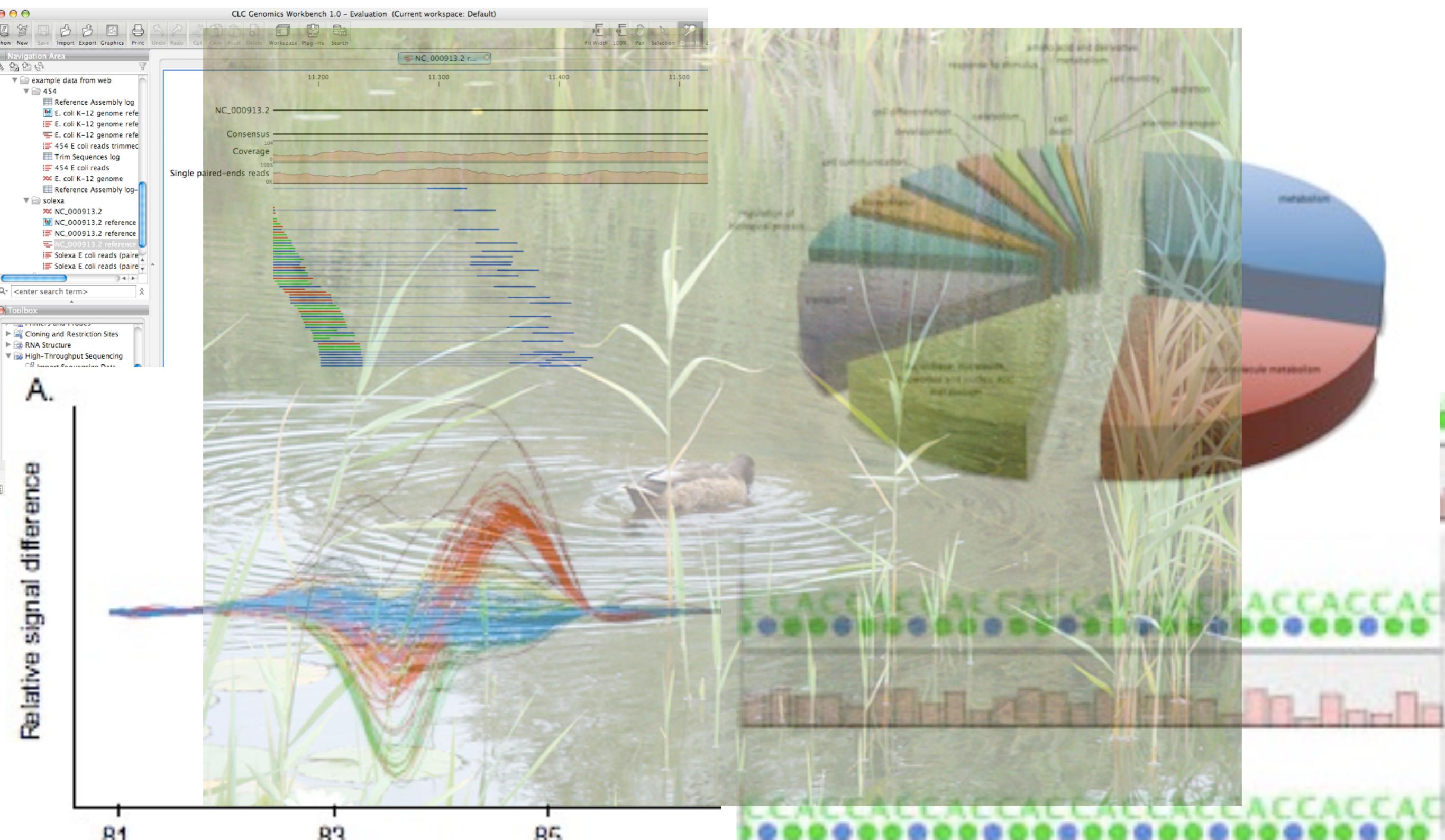


©2006 Jim Sutton

# What is this class about?



# What is this class about?



# What is this class about?

- data mining
- large datasets
- *in silico* analysis
- transcriptomics (physiology)
- pathogen discovery

SNPs (RAD/GBS)

lncRNA

Proteomics

DNA Methylation

# What is this class about?

EST

CLC

BLAST

GOOGLE

TIGR

BIOLOGY

GENE EXPRESSION

NCBI

NGS

ENSEMBLE

TECHNOLOGY

SNP

ACCESS

AMAZON

SOFTWARE

# Class Interest

# Biology

## Genetics



Are these populations reproductively isolated or is there successfully interbreeding?



## Transcriptomics

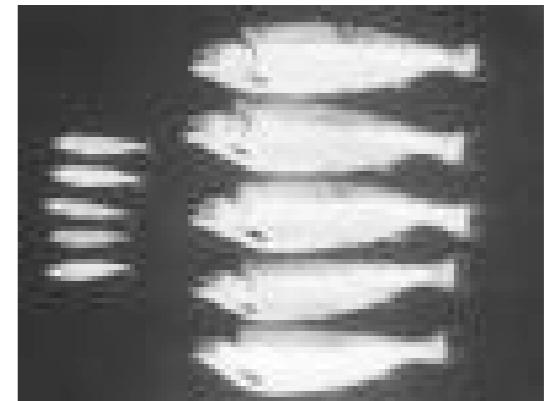
Has temperature stress compromised immune function?



What makes a parasite virulent?



How are growth processes controlled?



# Biology

## Gene Expression is a Dynamic Process

**Genes are the expressed portion of the genome**

**Genomic DNA**

**Messenger RNA (mRNA)**

**mRNA codes for proteins**

- *Different proteins perform a wide variety of biological functions.*
- enzymes, which catalyze chemical reactions
- structural or mechanical roles,
- immune response
- storage and transport of chemical signals
- cell communication

# Biology

What can influence gene expression...

Environment (Temperature, Salinity, Oxygen)

Pollution (Environmental Estrogens, Pesticides)

Development and Everyday Biology

## Gene Expression is a Dynamic Process

**Genes are the expressed portion of the genome**

**Genomic DNA**

**Messenger RNA (mRNA)**

**mRNA codes for proteins**

- *Different proteins perform a wide variety of biological functions.*

- enzymes, which catalyze chemical reactions
- structural or mechanical roles,
- immune response
- storage and transport of chemical signals
- cell communication

# Biology

What can influence gene expression...

EVERYTHING

Environment (Temperature, Salinity, Oxygen)

Pollution (Environmental Estrogens, Pesticides)

Development and Everyday Biology

## Gene Expression is a Dynamic Process

**Genes are the expressed portion of the genome**

**Genomic DNA**

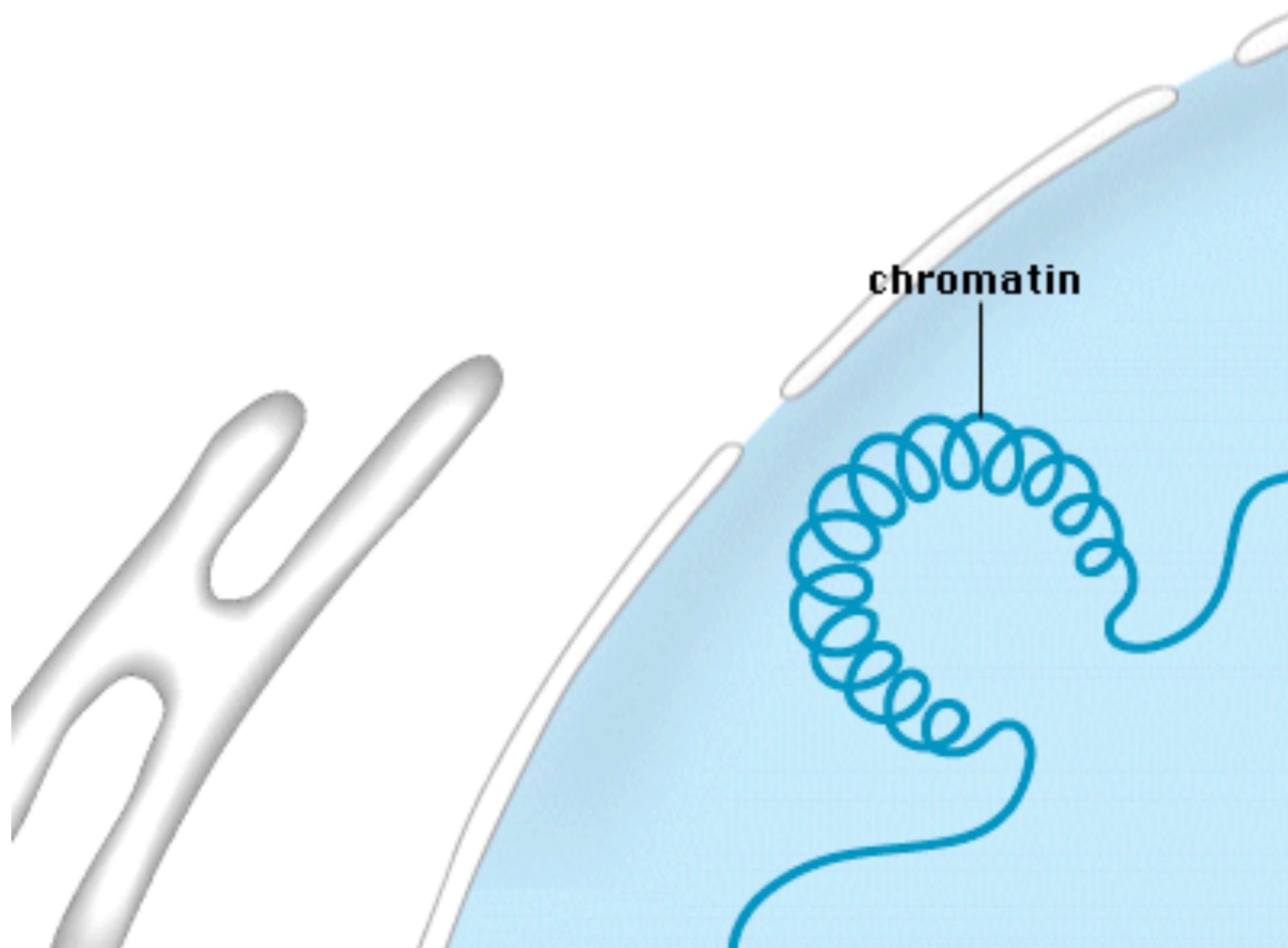
**Messenger RNA (mRNA)**

**mRNA codes for proteins**

- *Different proteins perform a wide variety of biological functions.*

- enzymes, which catalyze chemical reactions
- structural or mechanical roles,
- immune response
- storage and transport of chemical signals
- cell communication

# Biology



[link](#)

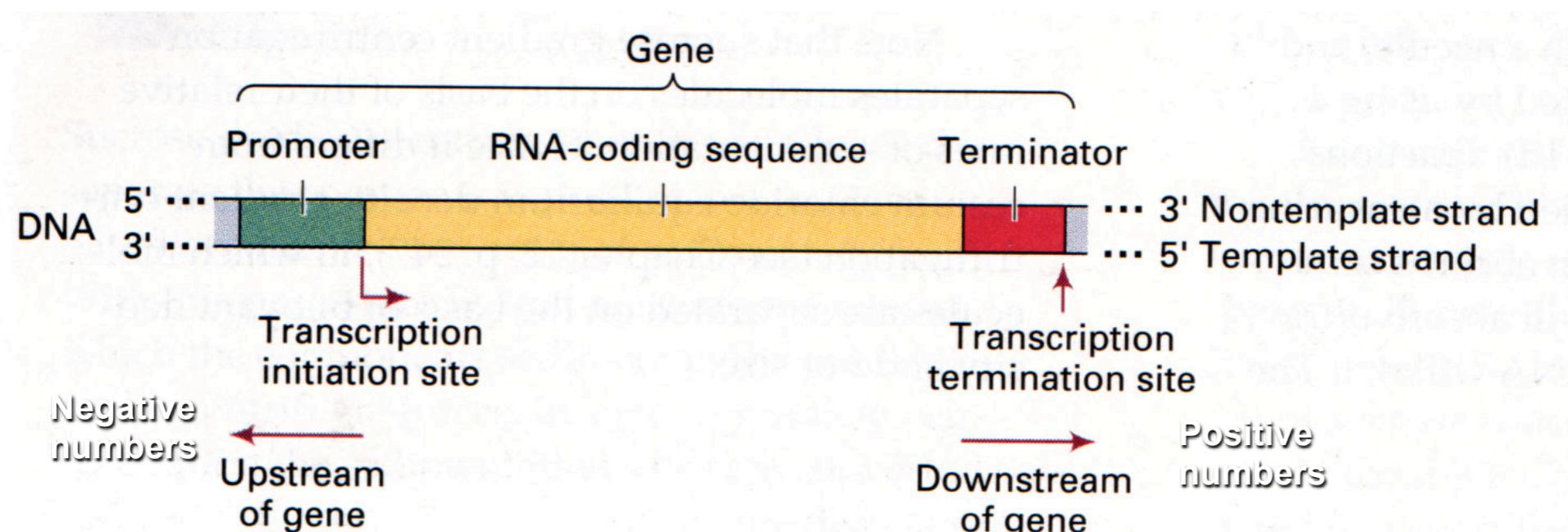
# Biology

## Transcription: organization of a gene

- Genes are flanked by regions called **Promoters**

**Promoters** – DNA sequences near the beginning of genes that signal RNA polymerase where to begin transcription

**Terminators** – Sequences in the RNA product that tells RNA polymerase where to stop (encoded by the DNA)



# Biology

## Eukaryotes :

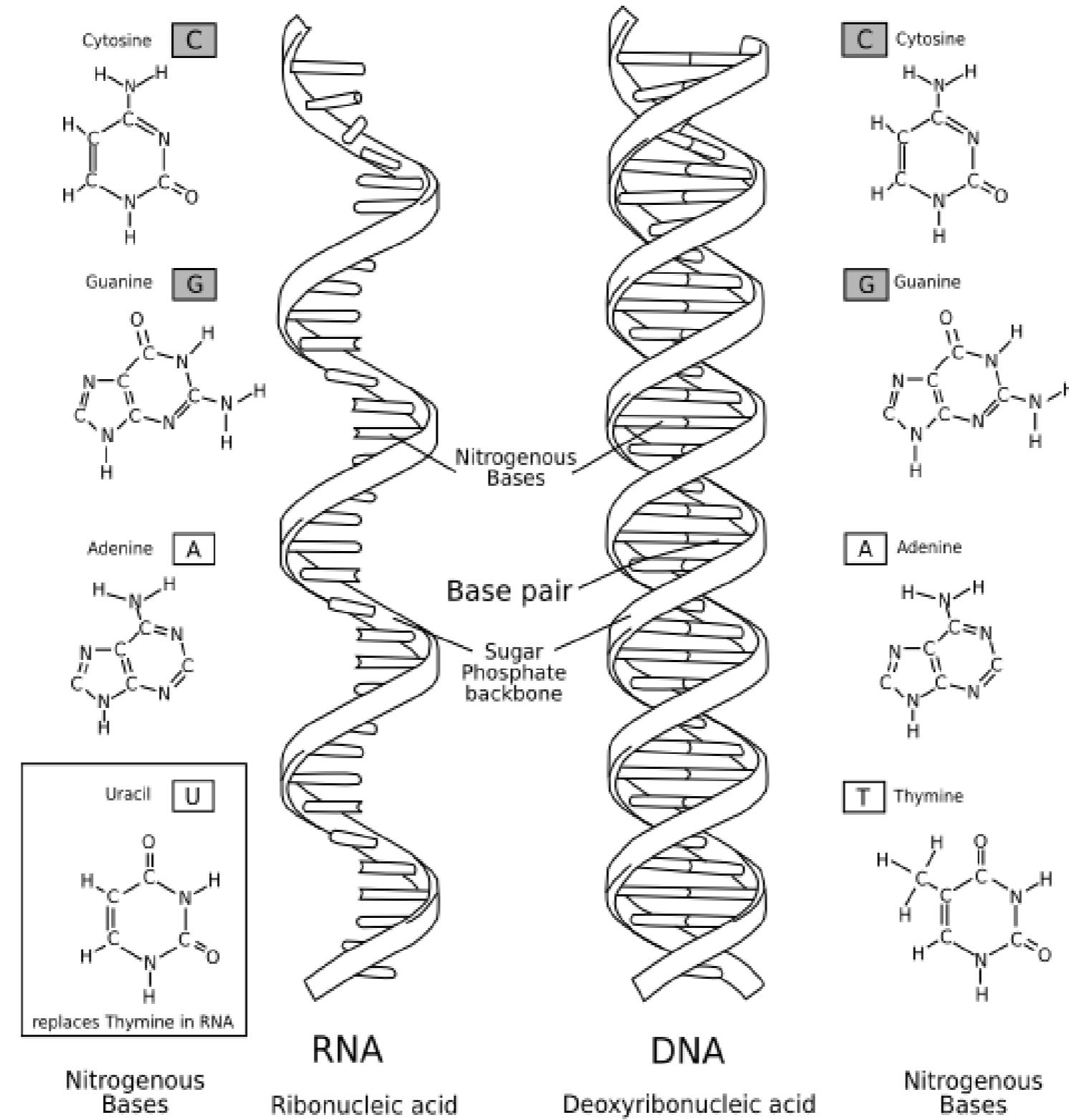
### Polyadenylation

- Addition of **poly-A tail** to 3' end ( $\approx$  100-200 A's)
- NO DNA template for poly-A tail (no strings of T's)
- RNAP runs past end of gene (no termination sites)
- Poly-A polymerase (PAP) adds As onto this new 3' end
- Thought to **stabilize mRNA** from degradation
- **Aids in efficiency of translation**

# Biology

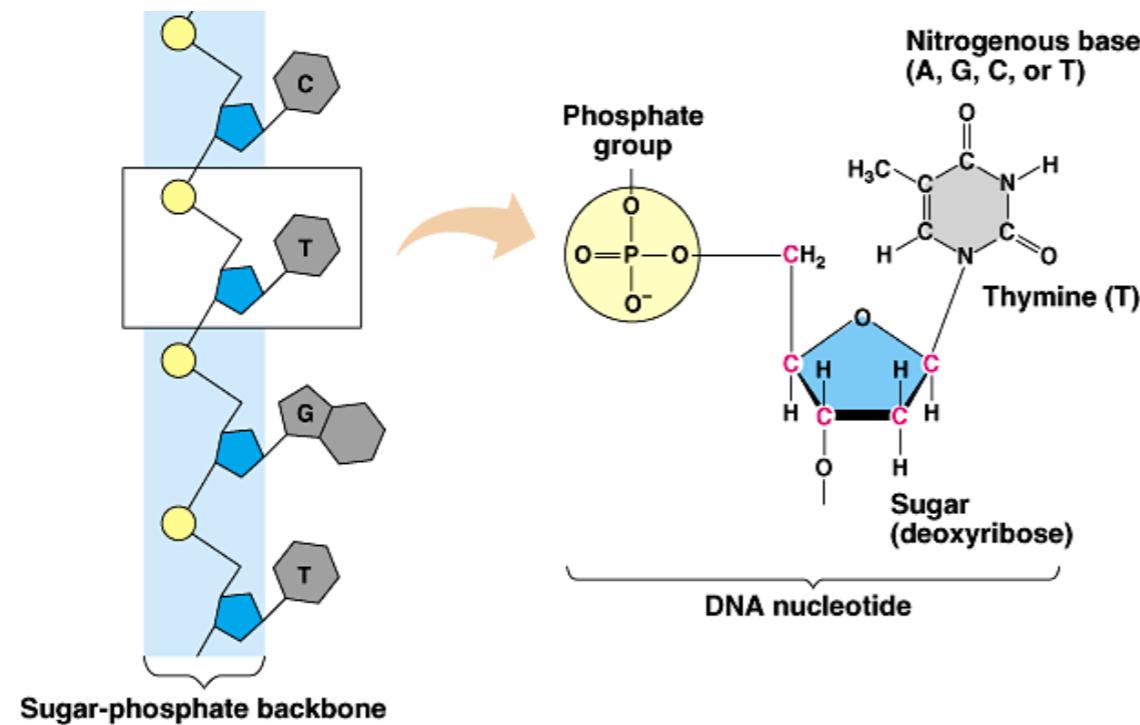
		Second letter					
		U	C	A	G		
First letter	U	UUU UUC UUA UUG	UCU UCC UCA UCG	UAU UAC UAA UAG	Tyr Stop Stop	UGU UGC UGA UGG	Cys Stop Trp
	C	CUU CUC CUA CUG	CCU CCC CCA CCG	CAU CAC CAA CAG	His Gln	CGU CGC CGA CGG	Arg
	A	AUU AUC AUA AUG	ACU ACC ACA ACG	AAU AAC AAA AAG	Asn Lys	AGU AGC AGA AGG	Ser Arg
	G	GUU GUC GUA GUG	GCU GCC GCA GCG	GAU GAC GAA GAG	Asp Glu	GGU GGC GGA GGG	Gly
		Third letter					
		U	C	A	G		

# Biology



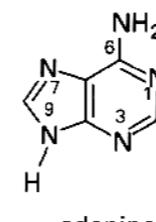
# Biology

## Structure of DNA

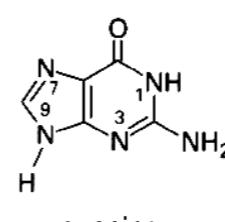


, publishing as Benjamin Cummings.

### Purines

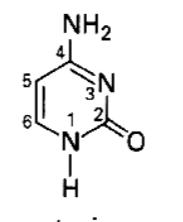


adenine

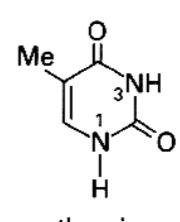


guanine

### Pyrimidines



cytosine



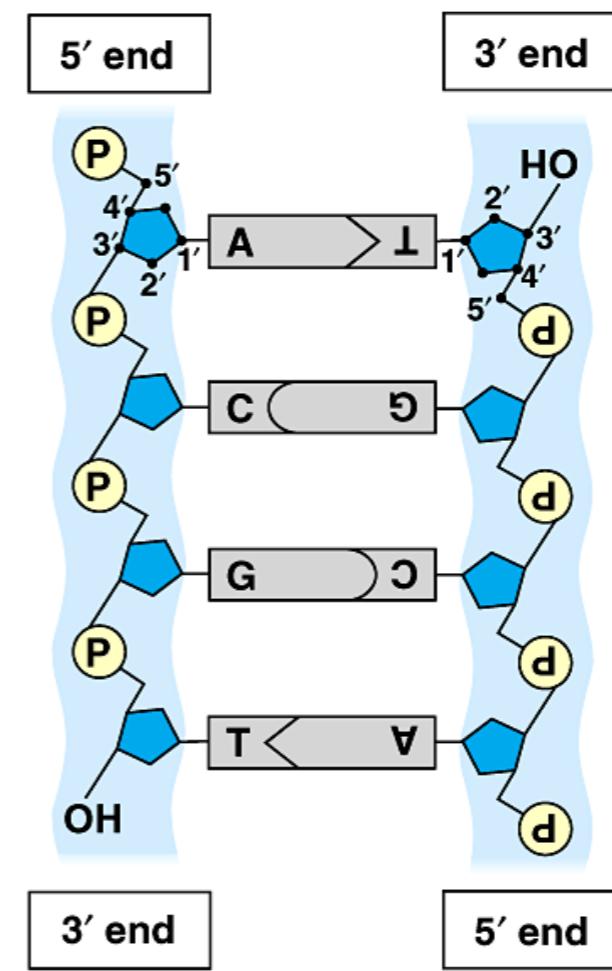
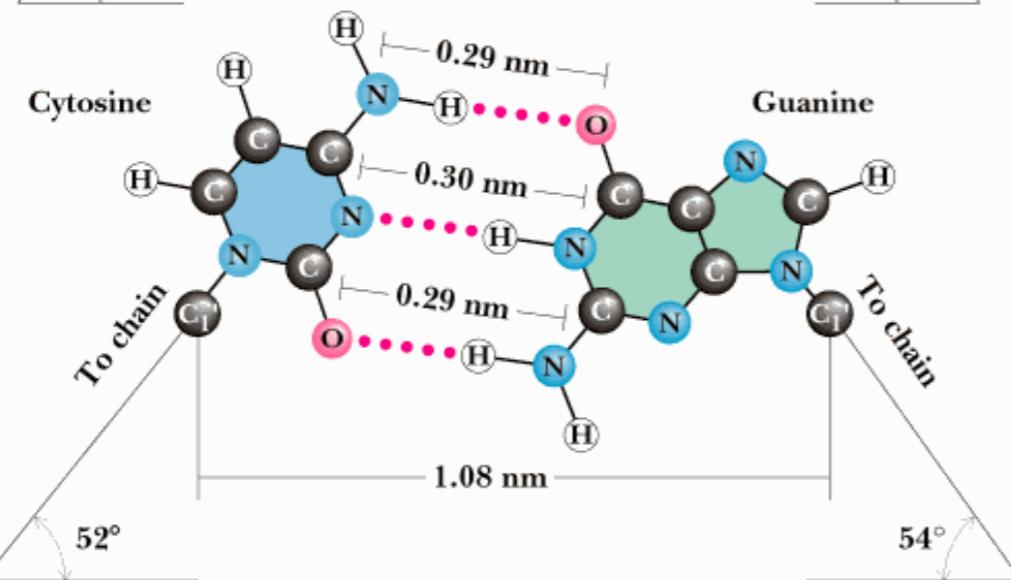
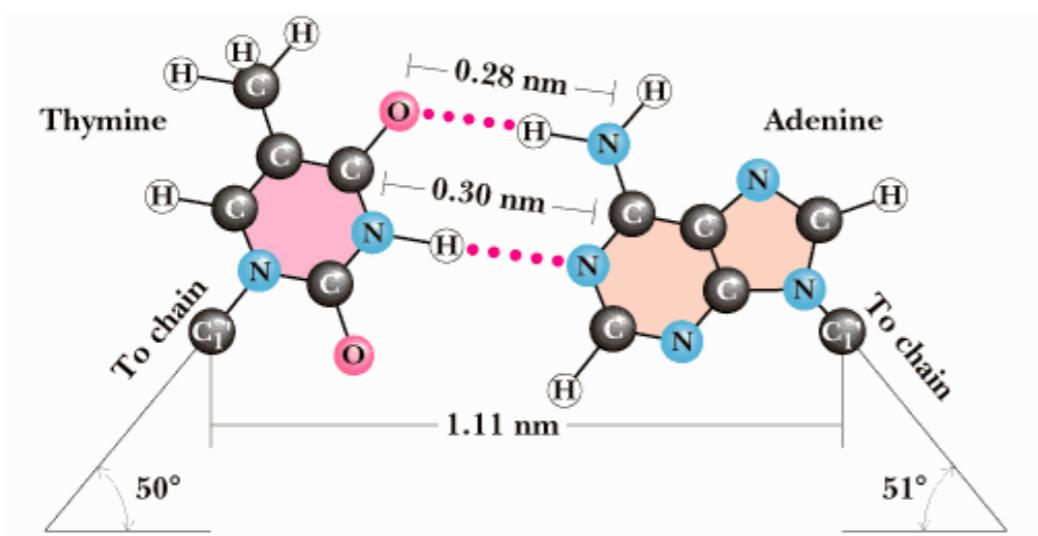
thymine

# Biology

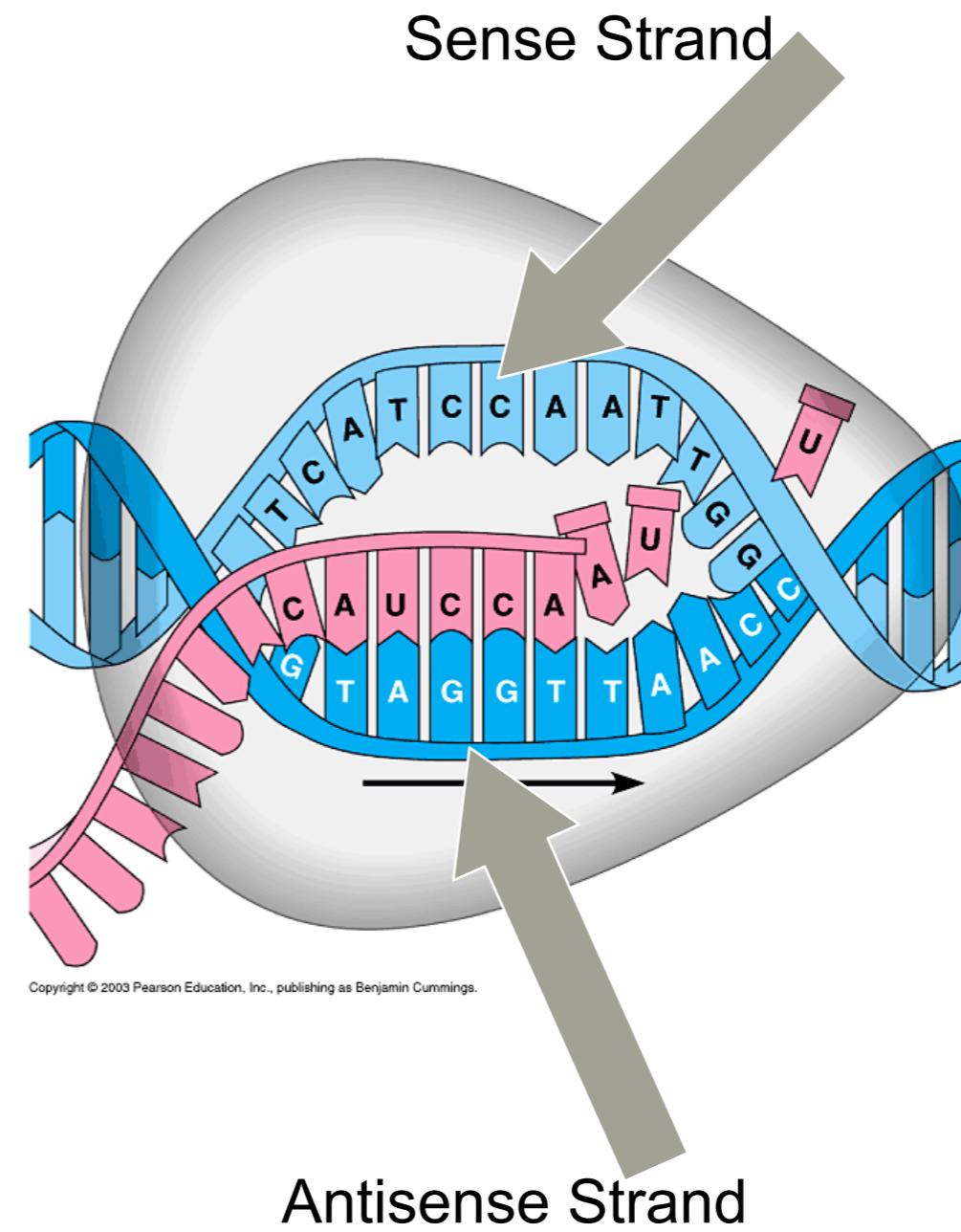
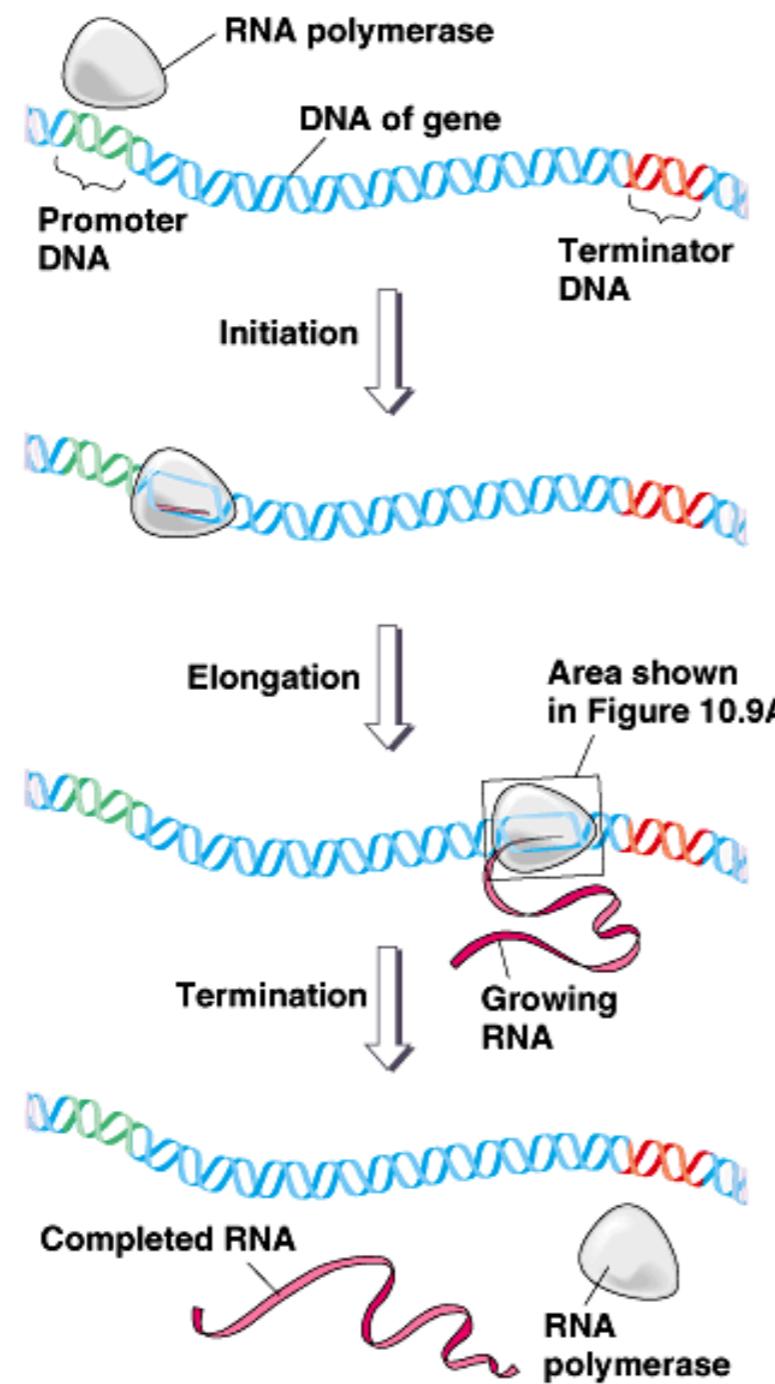
Complimentary binding  
Hydrogen bonds  
Purine with Pyrimidine

A – T

G – C



# Biology



Copyright © 2003 Pearson Education, Inc., publishing as Benjamin Cummings.

# Biology

## *Working with RNA*

Transcription changes over time  
RNA is less stable than DNA



### **RNA Isolation**

Process immediately  
Preserve it  
Disrupt sample to release RNA  
Extract and purify  
acid phenol/choloroform  
glass filter  
oligo dT binding  
Store

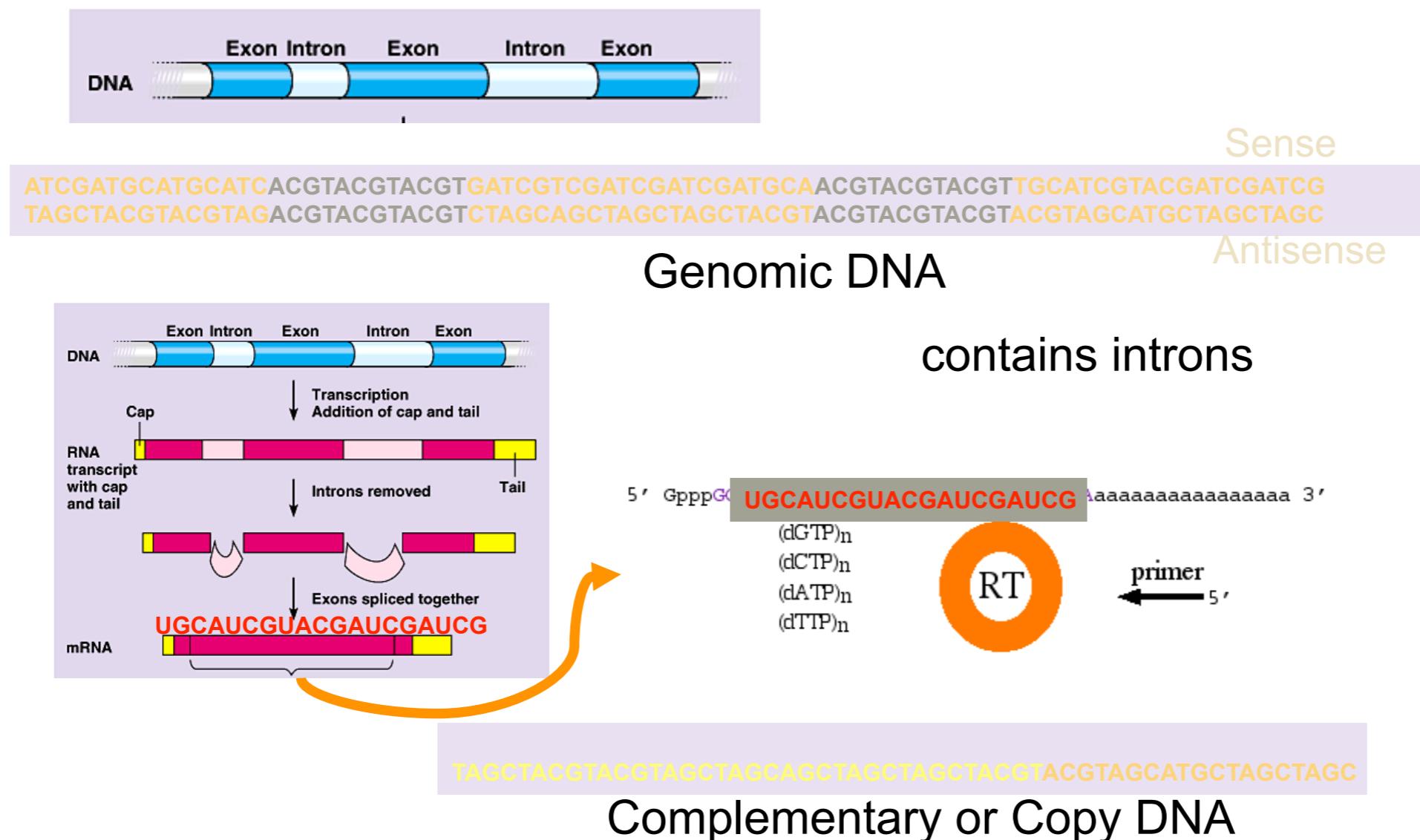
### **Reverse Transcribe RNA into complementary DNA**

This allows PCR to be performed on RNA



# Biology

# What is the Difference between DNA and cDNA?



# NCBI

**Resources**

- NCBI Home
- All Resources (A-Z)
- Literature
- DNA & RNA
- Proteins
- Sequence Analysis
- Genes & Expression
- Genomes
- Maps & Markers
- Domains & Structures
- Genetics & Medicine
- Taxonomy
- Data & Software
- Training & Tutorials

**Genotype and Phenotype**

Data from Genome Wide Association studies that links genes and diseases. See study variables, protocols, and analysis.

1 2 3

**How To...**

- Obtain the full text of an article
- Retrieve all sequences for an organism or taxon
- Find a homolog for a gene in another organism
- Find genes associated with a phenotype or disease
- Design PCR primers and check them for specificity
- Find the function of a gene or gene product
- Find syntenic regions between the genomes of two organisms

**DNA & RNA**

**Resources** **How To**

**DATABASES**

BioSystems Database that groups biomedical molecules, and sequence data relationships.

Database of Expressed Sequence Tags A division of GenBank that contains reads of cDNA (transcript) sequences searched directly through the NCBI.

Database of Genome Survey Sequences A division of GenBank that contains reads of genomic DNA. dbGSS through the Nucleotide GSS Database.

GenBank The NIH genetic sequence database collection of all publicly available sequence data. GenBank is part of the International Nucleotide Sequence Database Collaboration.

**Quick Links**

BLAST (Basic Local Alignment Search Tool)

More...

**DNA & RNA**

**Resources** **How To**

- Download a large, custom set of records from NCBI
- View/download features around an object or between two objects on a chromosome
- Link from an object on a map to another resource
- Obtain a genomic DNA clone for a gene
- Retrieve all sequences for an organism or taxon
- Find a curated version of a sequence record (NCBI Reference Sequence)
- Find transcript sequences for a gene
- Design PCR primers and check them for specificity
- Save a text search and/or receive regular search results by e-mail

**How To: Retrieve all sequences for an organism or taxon**

Starting with an organism or taxon name

- Search the Taxonomy database with the organism name. Accepted common names usually work at all taxonomic levels. Use the scientific name or formal name if no results are obtained with the common name.
- Click on the desired taxon name in the results. For terminal taxa - generally subspecies, species, or strains - this link leads directly to the summary page. For higher taxa this link will lead to the Taxonomy Browser showing the lower taxa contained within the higher taxon.
- If necessary, click on the desired taxon link in the Taxonomy Browser to reach the summary page.
- The number of records in each database are linked in the Entrez records table on the taxon summary page. Click the linked number of records in the table to retrieve all records from the chosen sequence database (Nucleotide, Nucleotide EST, Nucleotide GSS, Protein).

**Popular Resources**

- PubMed
- PubMed Central
- Bookshelf
- BLAST
- Gene
- Nucleotide
- Protein
- GEO
- Conserved Domains
- Structure
- PubChem

**NCBI News**

NCBI News - September 05 Oct 2009  
The September 2009 issue of the NCBI News is available ...

NCBI News - August 19 Aug 2009  
The August 2009 issue of the NCBI News is available online. ...

NCBI News - July 2009 17 Jul 2009  
The July 2009 issue of the NCBI News is now available online...

More...