

---

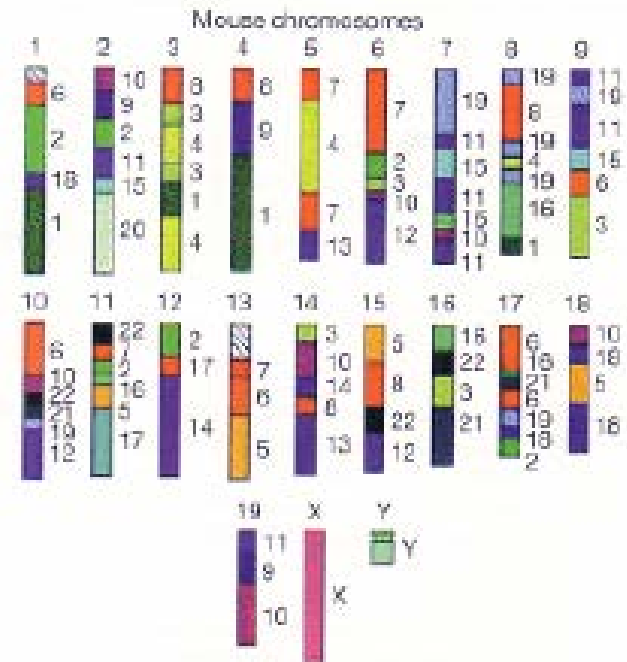
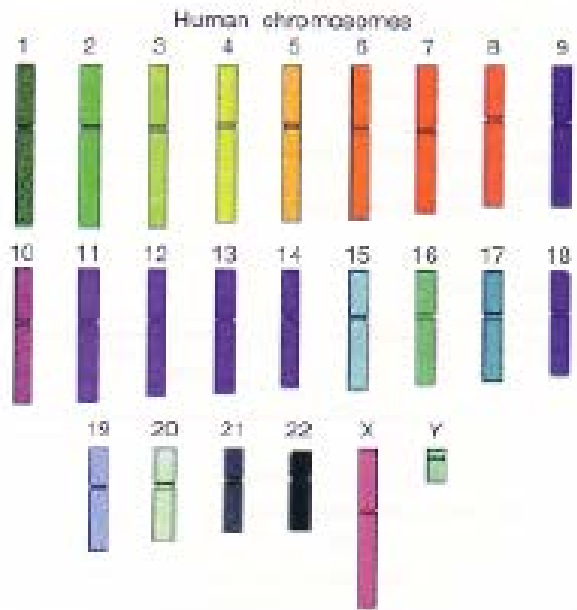
Curs1.

Biostatistică și bioinformatică. Introducere

---

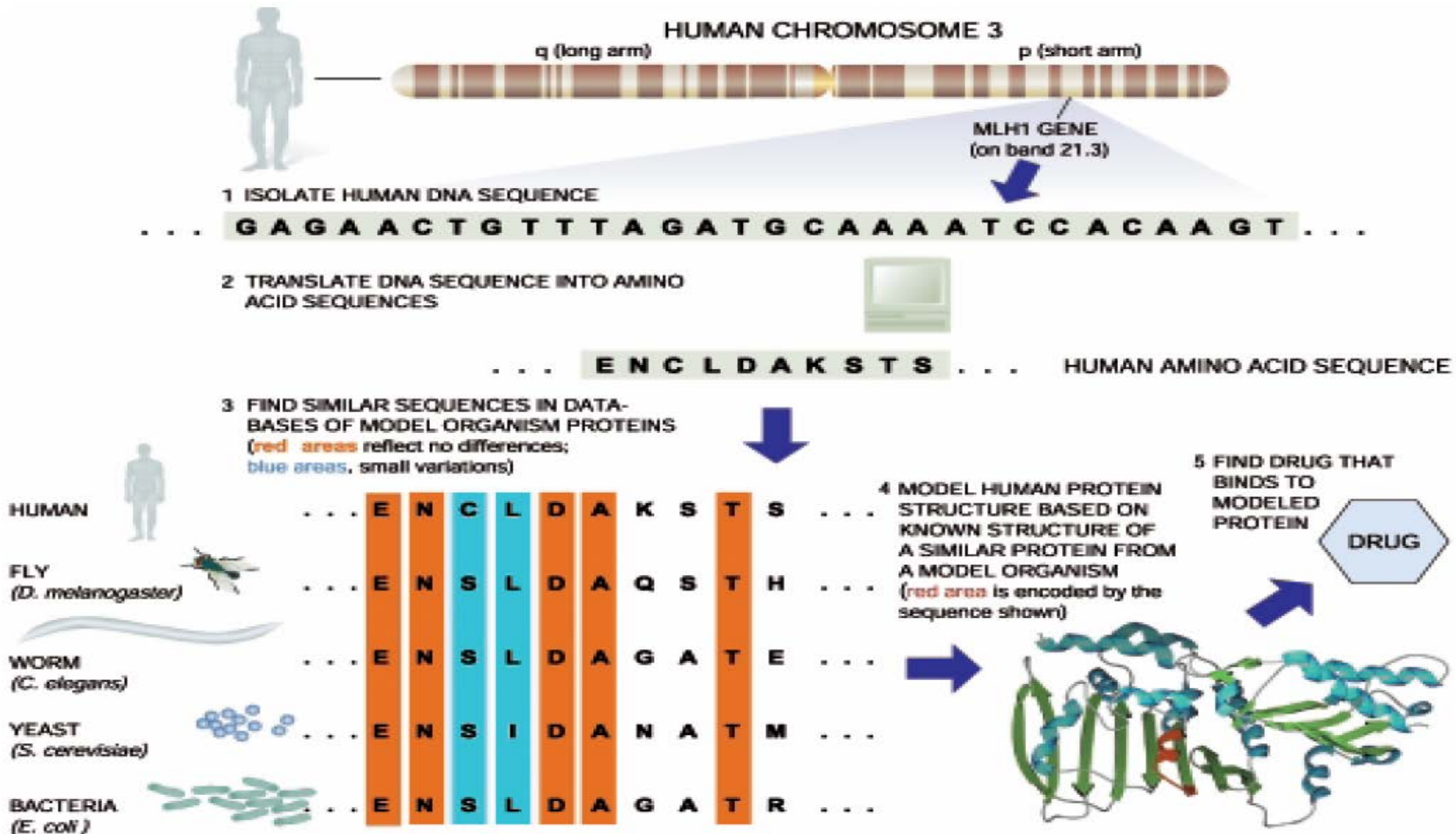
# Motivatie

- Cu milioane de ani în urmă diferențele dintre speciile de astăzi era aproape inexistentă
- În procesul de evoluție au intervenit mutații punctuale, inserări, ștergeri, rearanjări ale materialului genetic
- Întrebare: cât de mare este **similaritatea** între **genoamele** diferitelor specii și cum poate fi exploatată ?



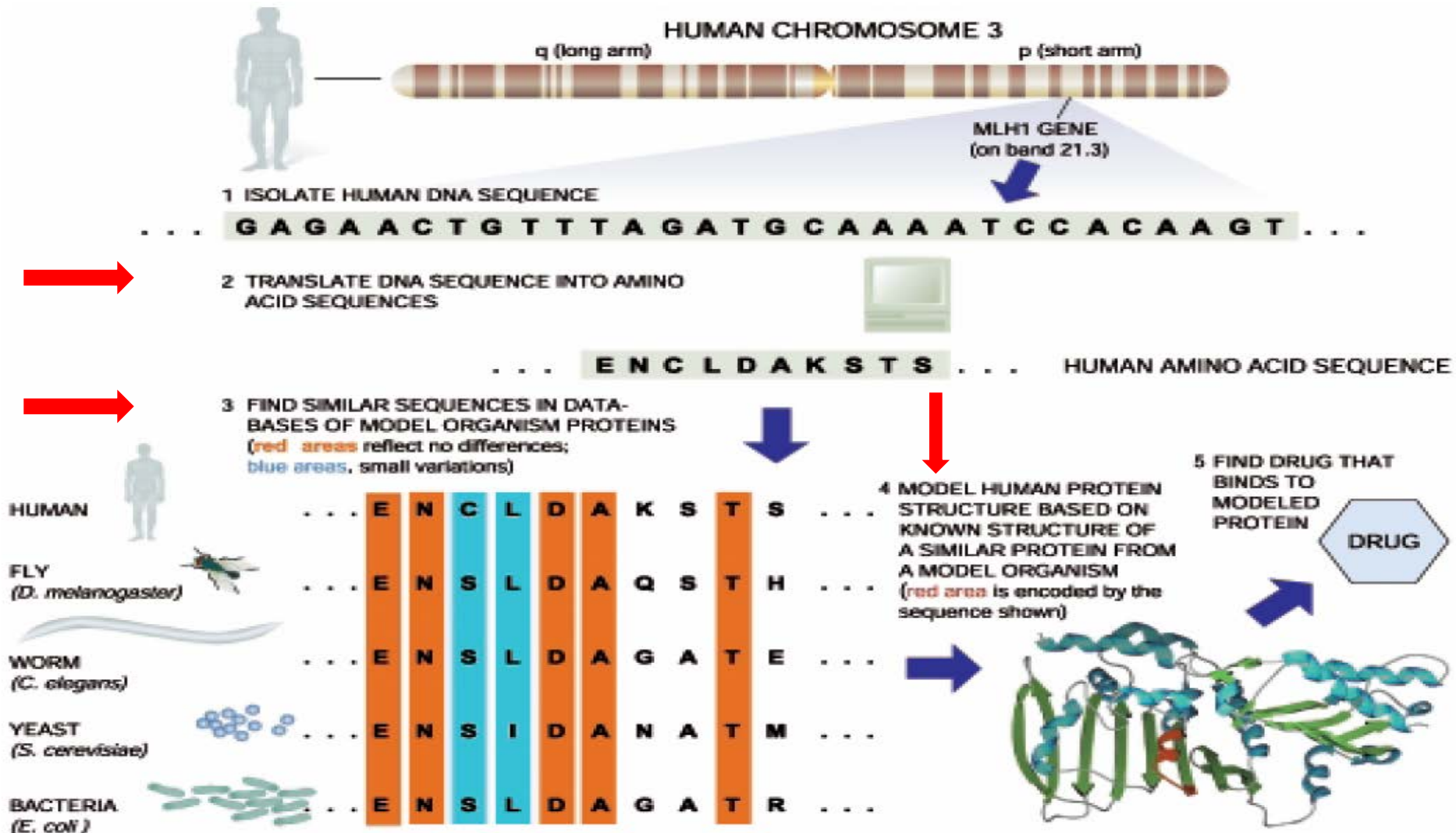
# Motivatie

- Scop final: tratamente personalizate [M.Luscombe et al. What is Bioinformatics?]



# Motivatie

- Scop final: tratamente personalizate [M.Luscombe et al. What is Bioinformatics?]



# Structura

- Biologie computațională: biostatistică și bioinformatică
- Scopul și structura cursului
- Scurtă introducere în biologia moleculară
- Probleme specifice în bioinformatică

# Biologie computațională: biostatistica și bioinformatică

- **Biologie computațională:** rezolvarea problemelor specifice biologiei (analiză, modelare și predicție pe baza datelor genomice) folosind metode și tehnici specifice matematicii și informaticii
- **Biostatistică:** utilizarea modelelor și metodelor statistice în analiza datelor biologice
- **Bioinformatică:** dezvoltarea de algoritmi pentru colectarea, stocarea și analiza datelor specifice biologiei moleculare

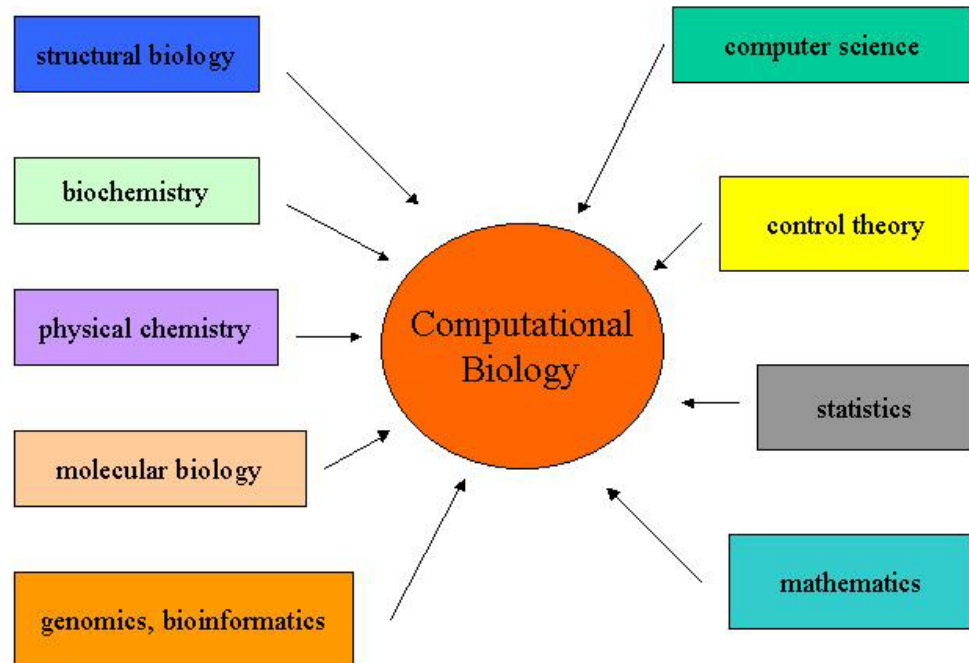
# Biologie computationala: biostatistica si bioinformatica

Este un domeniu interdisciplinar înrudit cu:

Discipline de biologie și chimie (structura genomului, procese de sinteză a proteinelor)

Informatica (tehnologii de stocare/regăsire a datelor biologice, algoritmi de analiză a secvențelor ADN etc.)

Statistica (testarea ipotezelor privind similaritatea dintre secvențele ADN)



Source: <http://ccb.wustl.edu/>

# Biologie computationala: biostatistica si bioinformatica

Este un domeniu în plină dezvoltare întrucât:

- Tehnicile de secvențiere a genomului generează în continuare un volum mare de date care așteaptă să fie analizate. Până în prezent au fost secvențiat genomul pentru peste 60000 de organisme (eucariote, procariote, virusi) în condițiile în care:
  - în 1995 se cunoștea genomul complet doar pt două organisme
  - în 2003 se cunoștea genomul complet pentru circa 100 de organisme (inclusiv genomul uman)

(<http://www.ncbi.nlm.nih.gov/genome/browse/>).

- Provocări principale:
  - Prelucrarea unui volum mare de date
  - Proiectarea unor sisteme complexe de analiză a datelor



# Scopul și structura cursului

## Scop:

- Familiarizarea cu metodele statistice și algoritmi destinați analizei datelor biologice
- Dobândirea de experiență în utilizarea pachetelor software folosite în bioinformatică

# Scopul si structura cursului

- Introducere în biologia moleculară. Noțiuni de bază. (cursul curent)
- Elemente de biostatistică. Distribuții utilizate în modelarea proceselor biologice. Teste statistice parametrice și neparametrice. Regresie liniară și neliniară.
- Algoritmi de identificare a șabloanelor (motivelor) în secvențele biologice.
- Algoritmi de aliniere a secvențelor biologice. Algoritmi de aliniere locală/globală/multiplă.
- Gruparea datelor biologice. Algoritmi de grupare partiționali și ierarhici.
- Algoritmi de construire a arborilor filogenetici
- Algoritmi aleatori. Aplicații în predicția structurii proteinelor.
- Analiza rețelelor biologice

# Bibliografie

- W. Ewens, G.R. Grant – Statistical Methods in Bioinformatics, Springer, 2005
- Neil C. Jones, Pavel A. Pevzner – An Introduction to Bioinformatics Algorithms, MIT Press, Cambridge, 2004
- R. Durbin, S. R. Eddy, A. Krogh, G. Mitchison – Biological Sequence Analysis. Probabilistic models of proteins and nucleic acids, Cambridge University Press, 2002
- W.K. Sung – Algorithms in Bioinformatics. A Practical Introduction, 2010

## Link-uri

- ISCB: <http://www.iscb.org/>
- NCBI (National Center for Biotechnology Information): <http://ncbi.nlm.nih.gov/>
- <http://www.expasy.org/>
- <http://www.bioinformatics.org/>
- [http://www.ornl.gov/sci/techresources/Human\\_Genome/posters/chromosome/tools.shtml](http://www.ornl.gov/sci/techresources/Human_Genome/posters/chromosome/tools.shtml)

Slide-uri: <http://kbrin.kwing.louisville.edu/~rouchka/CECS694/>  
<http://bix.ucsd.edu/bioalgorithms/>

---

# Materiale de curs si verificare

- Materiale de curs + lab:

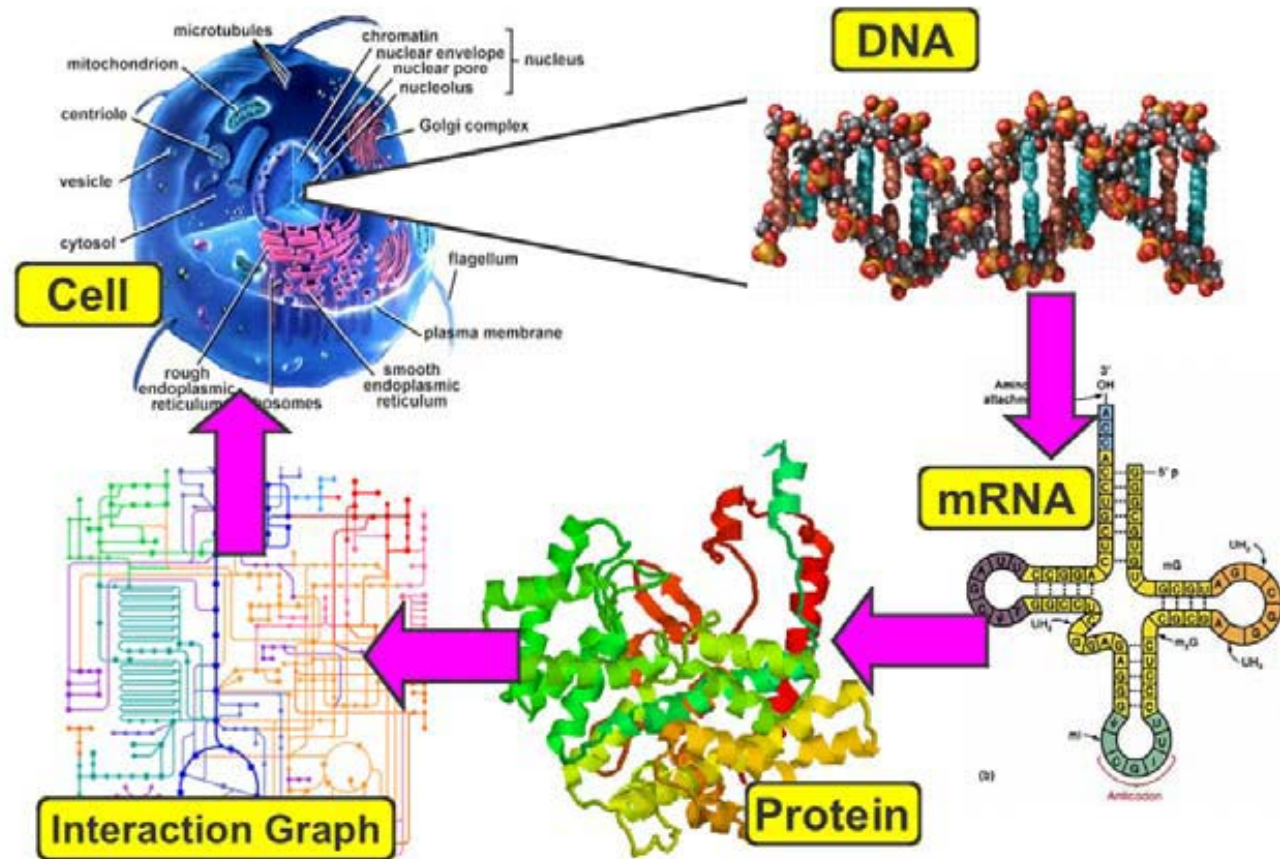
<http://www.info.uvt.ro/~dzaharie/bioinfo2016>

- Proiect (referat+implementare algoritm /studiu produs software existent) 60%
- Test scris cu acces la materiale bibliografice 20%
- Activitate laborator (+teme) 20%

# Scurtă introducere in biologia moleculară.

## Notiuni de baza

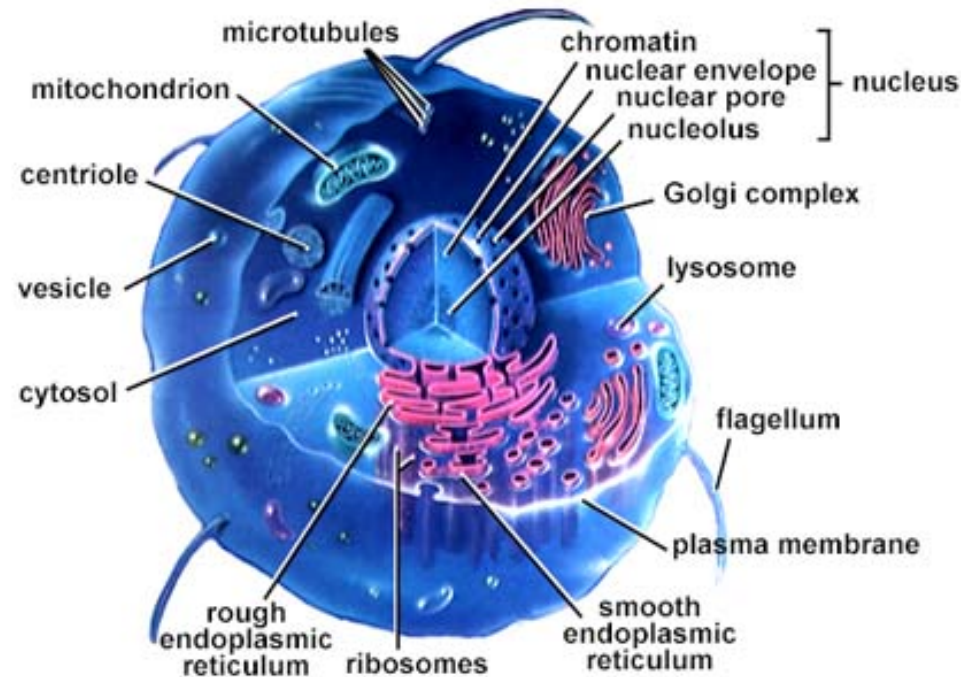
- celula
- cromozom
- ADN
- ARN
- aminoacizi
- proteine



[Gunnar Raetsch - <http://www.fml.mpg.de/~raetsch>] – Machine learning in bioinformatics]

# Celula

- Sistem complex delimitat de o membrană
- Organismele pot fi unicelulare (ex: bacterii) sau multicelulare
- Celulele pot fi: procariote (fără nucleu) sau eucariote (cu nucleu)
- Om:
  - 60 trilioane de celule
  - 320 tipuri de celule

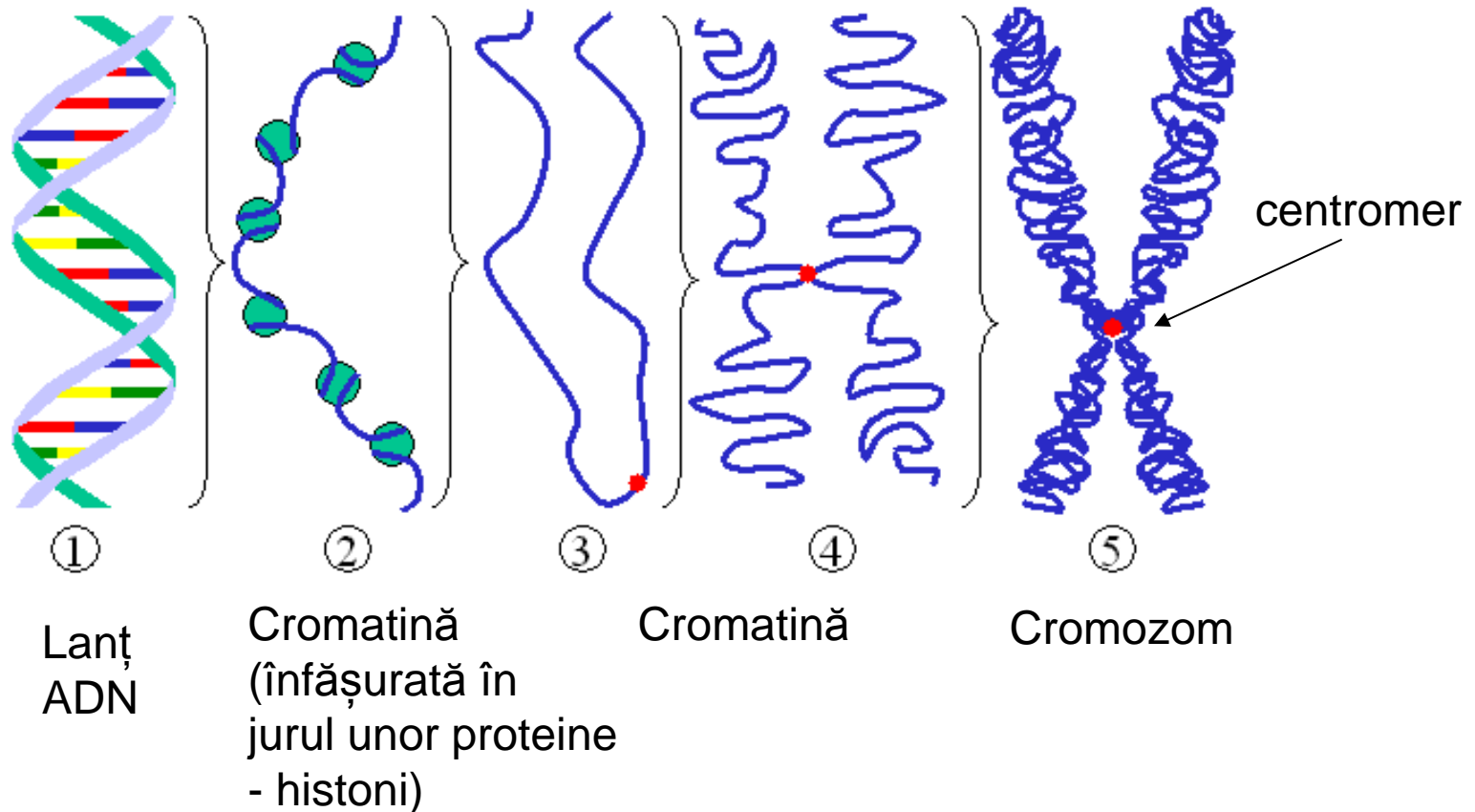


Exemplu de celulă animală

[www.ebi.ac.uk/microarray/biology\\_intro.htm](http://www.ebi.ac.uk/microarray/biology_intro.htm)

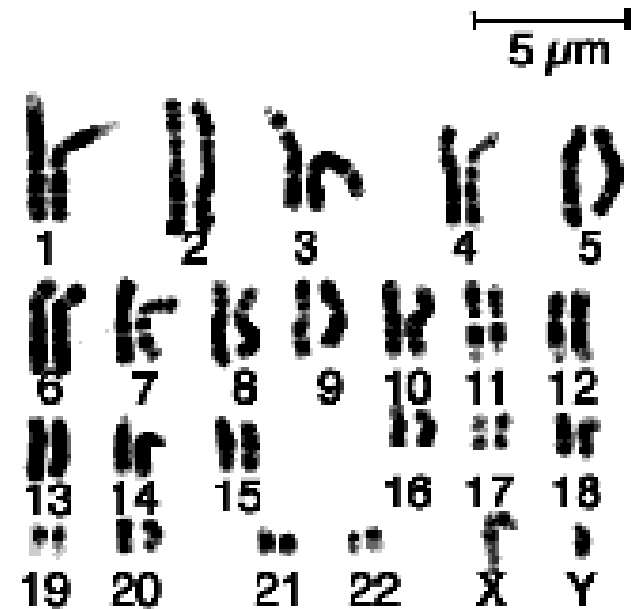
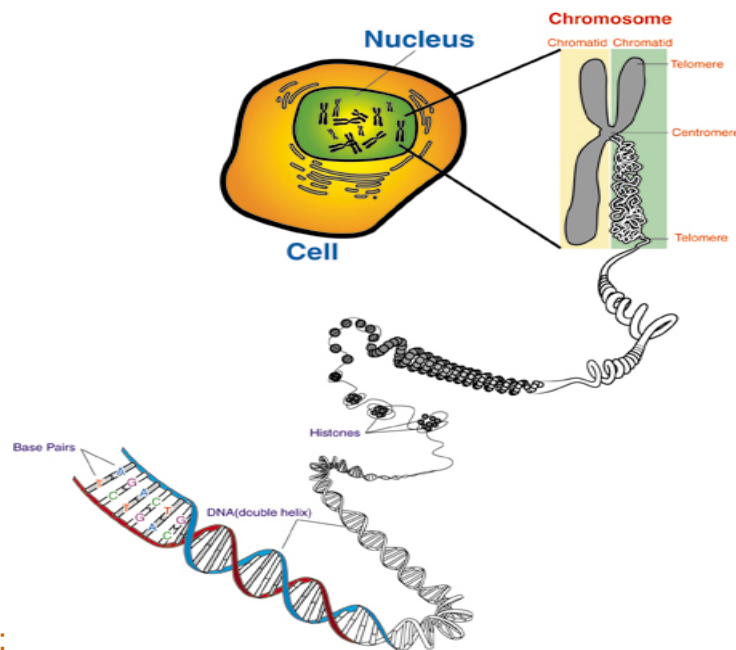
# Cromozom

- Nucleul celulelor conține unul sau mai multe lanțuri duble de ADN (acid dezoxiribonucleic) organizate în structuri numite cromozomi



# Cromozom

- Nucleul celulelor conține unul sau mai multe lanțuri duble de ADN organizate în cromozomi
- Om:
  - 22 perechi de autosomim + 1 pereche heterozomă (perechea care definește sexul XX sau XY)



Cariotipul uman

Sursa:  
[www.biotec.or.th/Genome/whatGenome.html](http://www.biotec.or.th/Genome/whatGenome.html)

<http://avery.rutgers.edu/WSSP/StudentScholars/Session8/Session8.html>



# ADN

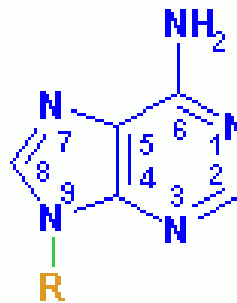
- ADN = acid dezoxiribonucleic = lanț de nucleotide (circa  $10^9$  nucleotide la plante și mamifere)
- nucleotidă = baza azotată + zahar + grup fosfat
- 4 tipuri de baze azotate (definesc cele 4 tipuri de nucleotide):
  - adenina (A)
  - citozina (C)
  - guanina (G)
  - timina (T)

grupate în două categorii: purine (A și G) și pirimidine (C și T)

Obs: în acidul ribonucleic (ARN) în loc de timină este uracil (U)

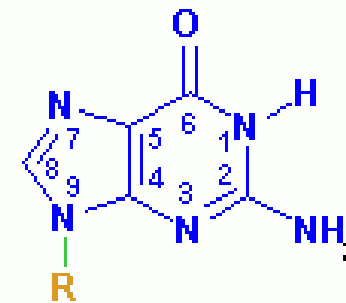
# ADN

- Structura chimică a bazelor azotate



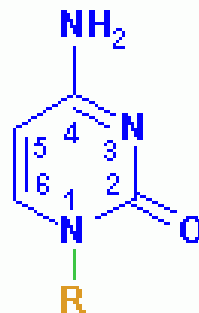
Adenine

## Purines

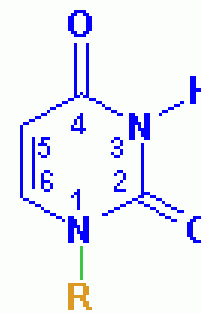


Guanine

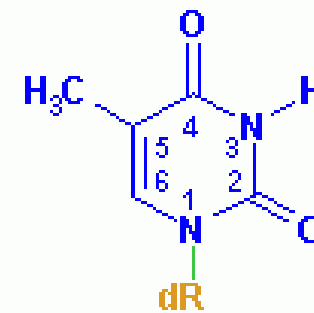
## Pyrimidines



Cytosine



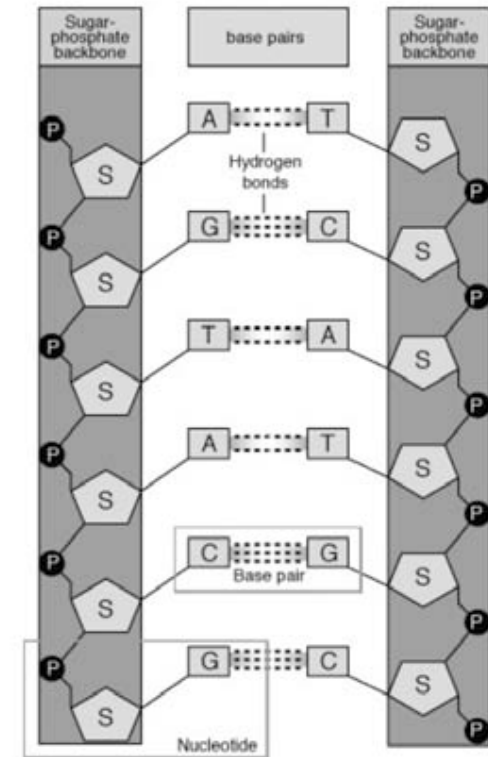
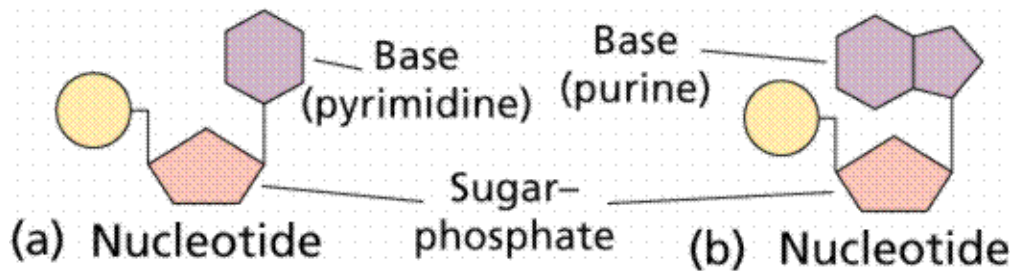
Uracil



Thymine

# ADN

- Nucleotidele consecutive sunt conectate printr-o legătură între zahar și grupul fosfat
- Nucleotidele din cele două lanțuri sunt grupate după regula: **A** cu **T**, **C** cu **G** (nucleotidele din familia purinelor pot fi conectate doar cu cele din familia pirimidinelor)



Sursa: [www.ebi.ac.uk/microarray/biology\\_intro.htm](http://www.ebi.ac.uk/microarray/biology_intro.htm)

# ADN

- Cele 4 tipuri de nucleotide pot fi interpretate ca un alfabet
- Pornind de la acest alfabet se pot crea cuvinte suficient de lungi (lanțuri ADN) pentru a codifica informația specifică unor organisme complexe
- Acest alfabet joacă un rol similar alfabetului binar  $\{0,1\}$  folosit pentru a reprezenta informația în memoria unui calculator

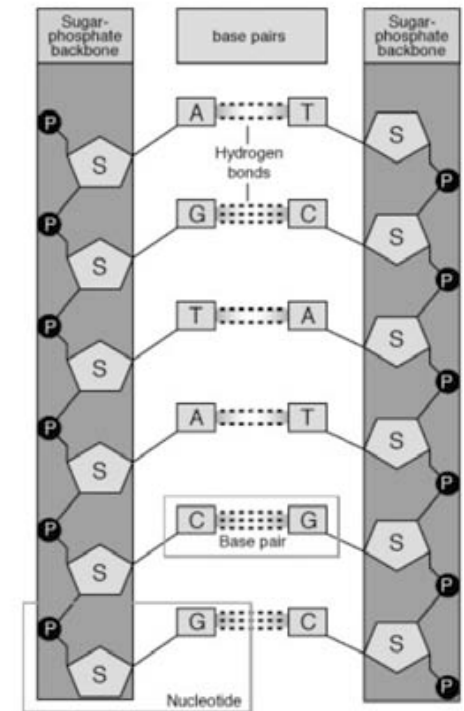
# Lant ADN

- Succesiunea în care bazele apar în lanțul ADN reprezintă de fapt informația genetică
- În lanțul de nucleotide există o polaritate chimică astfel că informația este codificată și decodificată într-o singură direcție (de la 5' la 3')

5' G→T→A→A→A→G→T→C→C→C→G→T→T→A→G→C 3'

# Lant ADN dublu

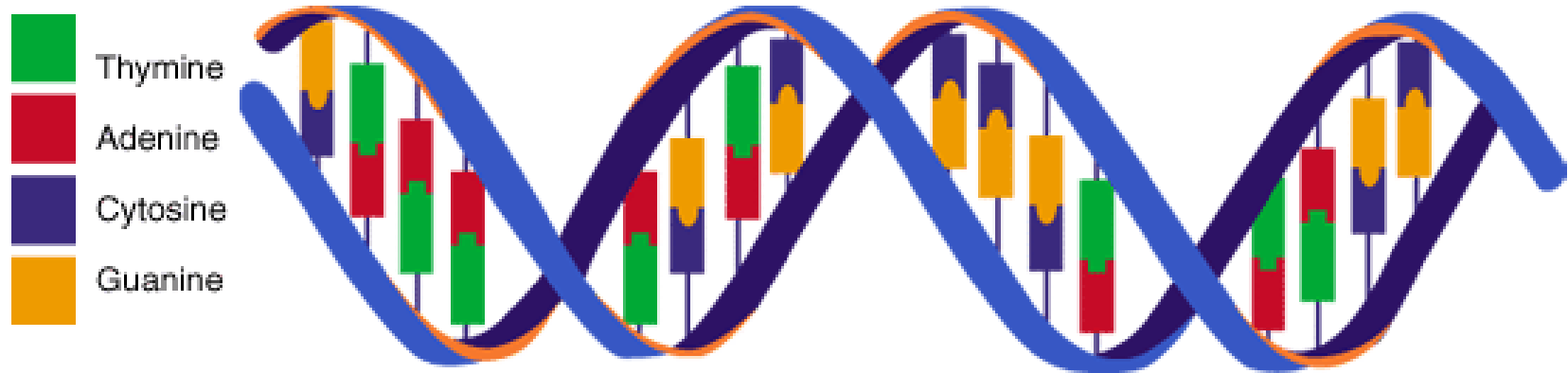
- Lanțul ADN poate fi simplu sau dublu
- In cadrul lanțului dublu cele două componente sunt complementare în sensul că se leagă între ele doar baze complementare (o purină cu o pirimidină):
  - A cu T
  - C cu G
- Dacă direcția de codificare a unui lanț este de la 5' la 3' direcția lanțului complementar este de la 3' la 5'.





# Structura de elice

- Cele două lanțuri complementare formează o structură stabilă de elice



Obs: structura spiralată a ADN-ului a fost descoperită în 1953 de către Crick și Watson (premiu Nobel)



# Rol ADN

- In lanțul ADN există:
  - Porțiuni care codifică informație utilizată în sinteza proteinelor (**exoni**) și care sunt organizate în gene
  - Porțiuni al căror rol este necunoscut încă (**introni – intragenic regions**) – informație redundantă
- Informațiile codificate in ADN sunt decodificate și utilizate în procesul de sinteză a proteinelor
- Ipotezele privind procesele de decodificare și sinteză reprezintă așa-numita “**dogma centrală a biologiei**”

# Dogma centrala a biologiei

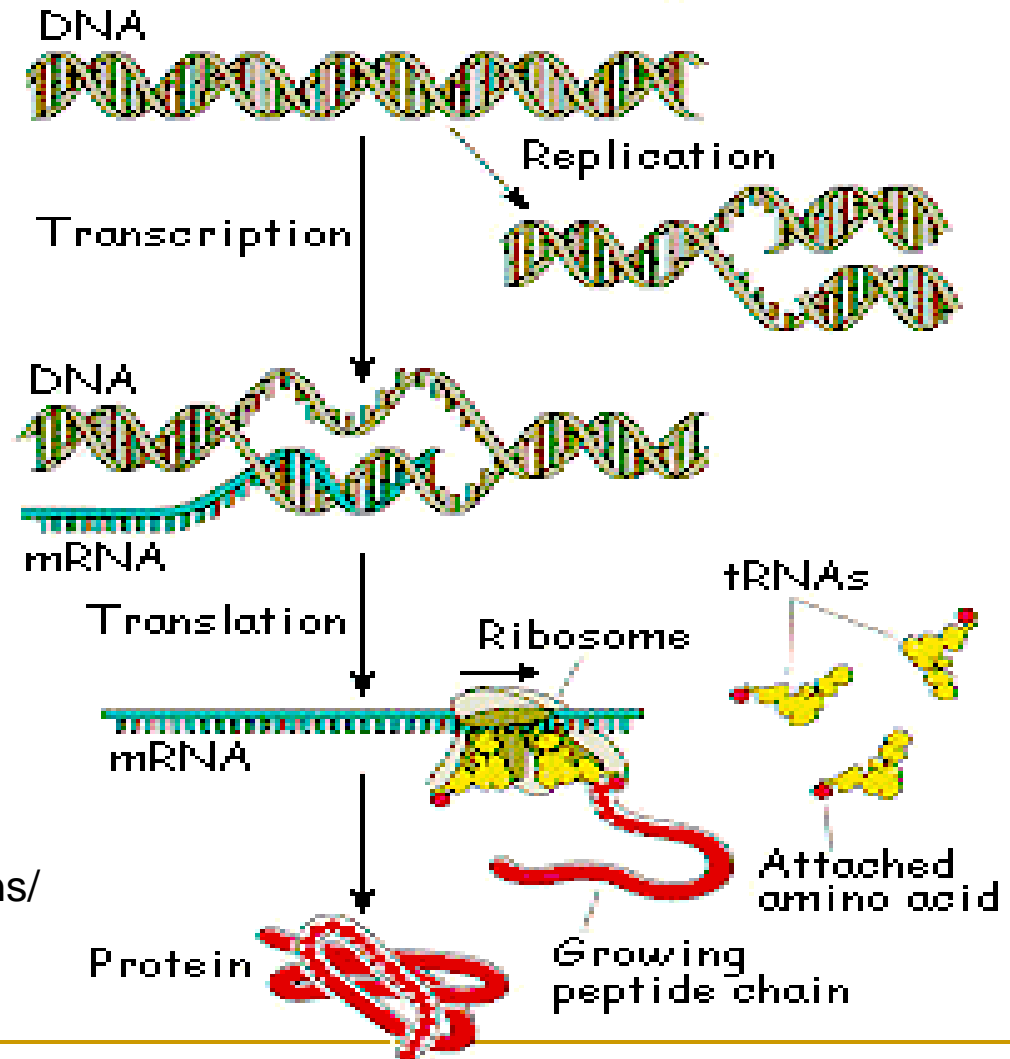
ADN

↓ (transcriere /  
copiere)

ARN

↓ (translație/  
traducere)

PROTEINA



Ilustrare prin animatie:

<http://vcell.ndsu.nodak.edu/animations/>

# ARN

- ARN = acid ribonucleic
- Structura ARN este similară cu cea a ADN însă:
  - Conține riboză în loc de dezoxiriboză
  - În locul timinei (T) apare uracil (U)
  - De regulă constă într-un singur lanț de nucleotide
- ARN joacă un rol cheie în sinteza proteinelor: realizează transferul informației codificate în ADN
- există entități (cum sunt virusii) a căror celule nu conțin ADN ci doar ARN (ca efect multiplicarea acestora este mult mai rapidă)

# ARN

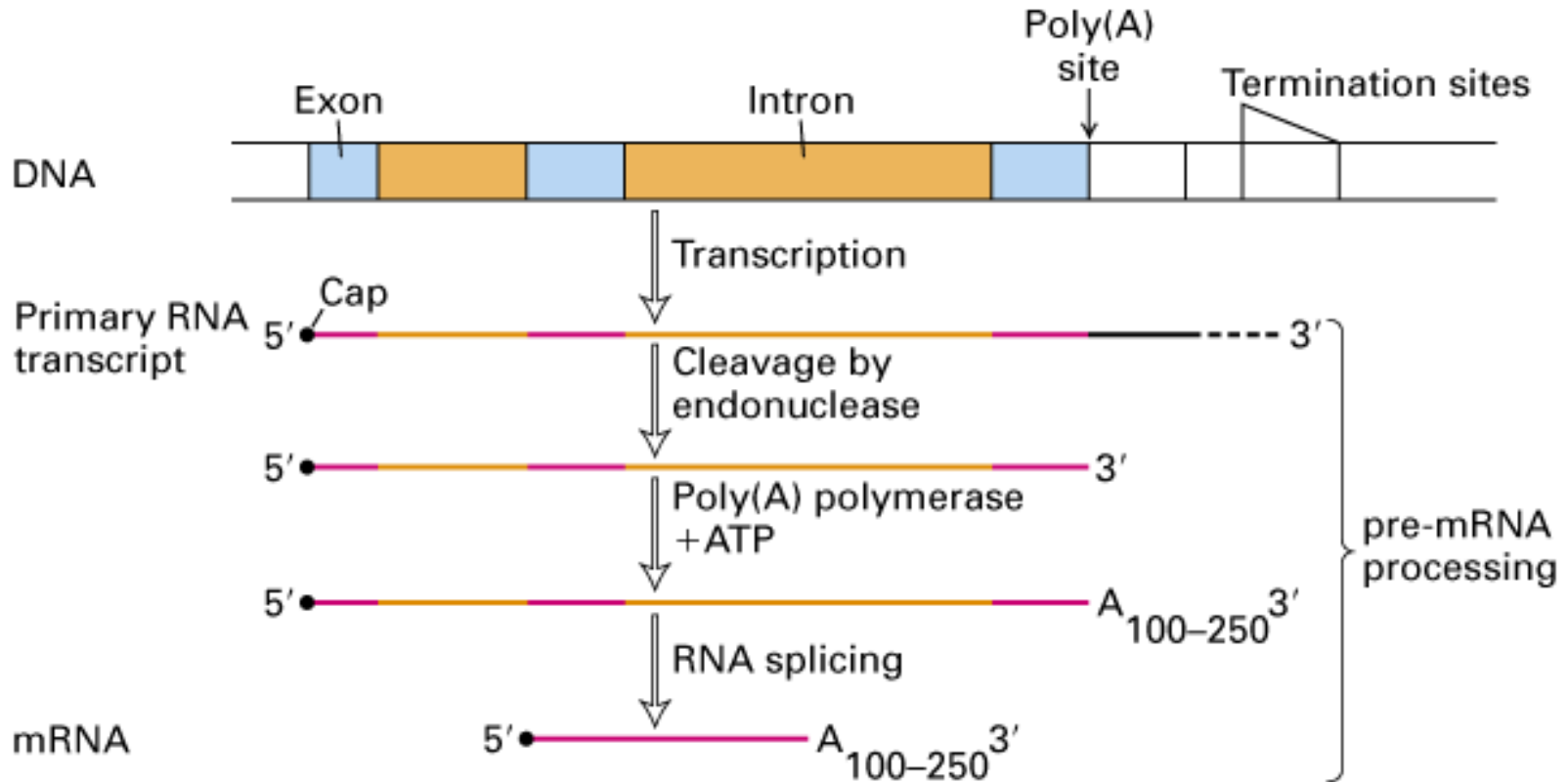
Tipuri de ARN:

- mARN = **ARN mesager**: joacă rolul unui șablon care preia informația din ADN necesară sintezei proteinelor
- tARN = **ARN de transfer**: este o secvență scurtă de nucleotide (mai puțin de 100) care nu codifică informație ci este utilizat de către ribozom în procesul de translație a codonilor (triplete de nucleotide) în aminoacizi
- dsARN = ARN constituit din două lanțuri de nucleotide (structura similară cu ADN-ul); întâlnit la unii viruși
- micro ARN (secvența de 20-25 nucleotide utilizate în construirea dsARN); rol: influență în expresia genică

# ARN messenger

- Moleculă care codifică informația copiată din moleculele ADN
- La generarea ARN-ului messenger porțiunile necodante (introni) sunt eliminate iar cele codante (exoni) sunt concatenate
- **Transcriere/copiere:** procesul prin care ADN este copiat într-o molecula de ARN messenger
- **Translație/traducere:** procesul de sinteză a proteinelor pornind de la ARN messenger

# ARN mesager



Sursa : [http://departments.oxy.edu/biology/Stillman/bi221/111300/processing\\_of\\_hnrnas.htm](http://departments.oxy.edu/biology/Stillman/bi221/111300/processing_of_hnrnas.htm)

# Cod genetic

- 4 tipuri de baze azotate (A, C, G, U)
- Codon = triplet de baze care codifică un aminoacid (aminoacizii sunt molecule constituite din grupuri amino și carboxil și sunt elementele constitutive ale proteinelor; proteinele joacă un rol esențial în procesele de dezvoltare celulară)
- Proteina este codificată printr-o secvență de codoni care începe cu un **codon de start** și se termină cu un **codon de stop**
- **Codon start:** AUG, CUG
- **Codon stop:** UAA, UAG, UGA
- $4 * 4 * 4 = 64$  codoni posibili
- Excluzând codonii de stop rămân 61 codoni pt. a codifica aminoacizii (inclusiv codonii de start)
- 20 amino acizi → codul genetic este redundant

# Aminoacizi - codificare

- Glycine (G, GLY)
- Alanine (A, ALA)
- Valine (V, VAL)
- Leucine (L, LEU)
- Isoleucine (I, ILE)
- Phenylalanine (F, PHE)
- Proline (P, PRO)
- Serine (S, SER)
- Threonine (T, THR)
- Cysteine (C, CYS)
- Methionine (M, MET)
- Tryptophan (W, TRP)
- Tyrosine (T, TYR)
- Asparagine (N, ASN)
- Glutamine (Q, GLN)
- Aspartic acid (D, ASP)
- Glutamic Acid (E, GLU)
- Lysine (K, LYS)
- Arginine (R, ARG)
- Histidine (H, HIS)
- START: AUG
- STOP: UAA, UAG, UGA

		Pozitia 2				Poz. 3
		U	C	A	G	
Pozitia 1	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



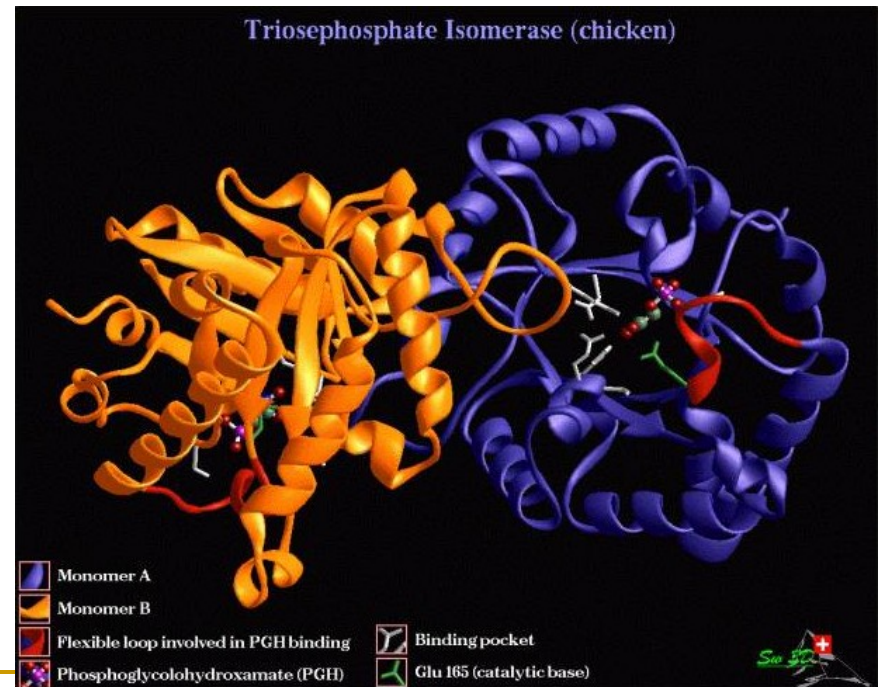
# Aminoacizi

- Sunt componentele de bază ale proteinelor
- Tipuri: hidrofili (solubili în apă), hidrofobi
- **peptida**: lanț scurt de aminoacizi (< 30)
- **polipeptida**: lanț lung de aminoacizi (până la 4000).

Obs. Proteinele sunt polipeptide cu o structura geometrică specifică (nu sunt structuri liniare ci au o structură sofisticată – sunt “împachetate”)

# Proteine

- Structura spațială a proteinelor (rezultatul procesului de “împachetare”) este foarte importantă și este rezultatul unei proceduri complicate influențate de interacțiunile dintre aminoacizi și apă
- Erori la “împachetare” conduc la boli ca Alzheimer, Parkinson



# Sumar:

- ADN
  - Purtător al informației referitoare la modul în care operează celulele
- ARN
  - Permite transferul informației între diferite părți ale celulei
    - ADN-ul se află în nucleu
    - Procesul de sinteză se desfășoară în afara nucleului
  - Furnizează un șablon pentru sintetizarea proteinelor
- Proteine
  - Stau la baza constituirii principalelor elemente ale organismului
  - Formează enzime care trimit semnale către alte celule și reglează activitatea genelor

# Gena

- **Gena** = unitatea fizică și funcțională a eredității care permite transferul de informație de la o generație la alta.
- Gena = secvență ADN necesară în general pentru sinteza unei proteine (obs: există și gene care controlează sinteza mai multor proteine) și care conține câteva regiuni fiecare având un rol specific:
  - **Regiunea reglatoare** (promotor) – reglează procesul de transcriere
  - **Regiunea codantă** (numită și **Open Reading Frame (ORF)**) – conține tripletele folosite la sinteza aminoacizilor ce intervin în lanțul peptidic care formează proteina)
  - Regiuni care rămân “netraduse”
- Obs:
  - genele reprezintă partea codantă a ADN-ului (la om reprezintă cca 2-3% din ADN, la E. Coli reprezintă cca 90% din ADN)
  - omul are cca 30000 gene

# Genom

- Reprezintă ADN-ul unui organism structurat în cromozomi
- Numărul de cromozomi și dimensiunea genomului variază de la o specie la alta
- Dimensiunea genomului și numărul de gene nu determină în mod necesar complexitatea organismului.
- Variațiile existente între indivizii aceleiași specii sunt cauzate în principal de diferențe izolate care apar la nivelul unei nucleotide. Identificarea pozițiilor unde apar astfel de diferențe (**SNP – Single Nucleotide Polymorphism**) reprezintă una dintre problemele importante în bioinformatică. În cadrul genomului uman, SNP-urile controlează culoarea părului, tipul de sânge etc.

# Diferențe între dimensiunea genomului

ORGANISM	CROMOZOMI	DIM. GENOM (perechi baze)	GENE
<u><i>Homo sapiens</i></u> Om	23	3.200.000.000	~ 30.000
<u><i>Mus musculus</i></u> Șoarece	20	2.600.000.000	~30.000
<u><i>Drosophila melanogaster</i></u> (musculita de otet)	4	180.000.000	~18.000
<i>Escherichia coli</i>	1 (un cromozom circular)	4.600.000	~42.000
<i>Saccharomyces cerevisiae</i> (drojdie)	16	12.000.000	~6.200

# Analiza genomului

## Scurt istoric al genomicii:

- 1866 enunțarea legilor eredității (Mendel)
- 1869 descoperirea nucleinei (acizi nucleici) (Miescher)
- 1909 reguli ale geneticii deduse prin experimente asupra musculiței de oțet – 4 perechi de cromozomi (Morgan)
- 1944 argumente bazate pe ADN privind ereditatea (Avery)
- 1953 structura spiralată a ADN-ului (Crick and Watson)
- 1966 codul genetic (codificarea aminoacizilor) (Nirenberg)
- 1977 dezvoltarea primei metode pentru secvențierea ADN-ului (Sanger)
- 1982 crearea bazei de date Genbank
- 1990 lansarea proiectului destinat secvențierii genomului uman
- 2003 finalizarea secvențierii genomului uman

# Analiza genomului

Tipuri de prelucrări (biologice):

**Tăiere:**

- folosind enzime de restricție se secționează lanțul ADN în fragmente
  - de exemplu enzima EcoRI fragmentează lanțul în toate pozițiile unde apare **GAATTC**) – secționarea se realizează între G și A
  - Se cunosc cca 300 de astfel de enzime
- aleator (shotgun method) – se aplică vibrații care fragmentează moleculele

**Copiere:** se multiplică fragmentele pentru a fi mai ușor de detectat (PCR – Polymerase Chain Reaction)

Motivație: pt. identificarea unui tip de moleculă prin măsurători sunt necesare milioane de copii

**Citire:** se “citesc” fragmentele folosind tehnici chimice (exemplu: măsurarea lungimii ADN: electroforeza in gel)

**Asamblare:** se assemblează fragmentele (difícil de realizat; poate fi formulată ca o problema de “shortest common superstring”)

<http://www.youtube.com/watch?v=-gVh3z6MwdU>

<https://www.youtube.com/watch?v=MvuYATh7Y74>

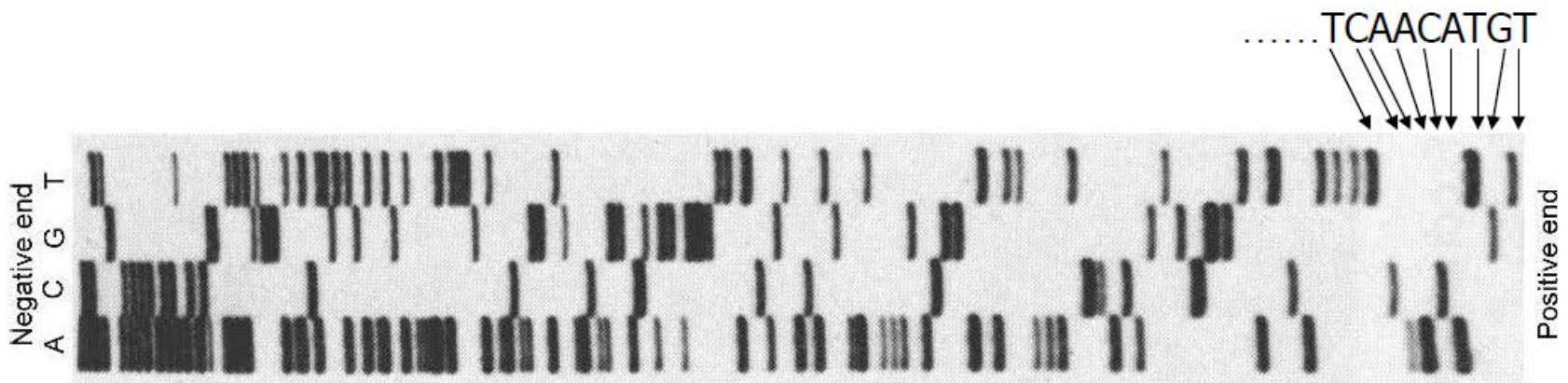


# Analiza genomului

**Citire:** se “citesc” fragmentele folosind tehnici chimice  
(exemplu: măsurarea lungimii ADN: electroforeza in gel)

**Asamblare:** se assemblează fragmentele (difícil de realizat; poate fi formulată ca o problema de “shortest common superstring”)

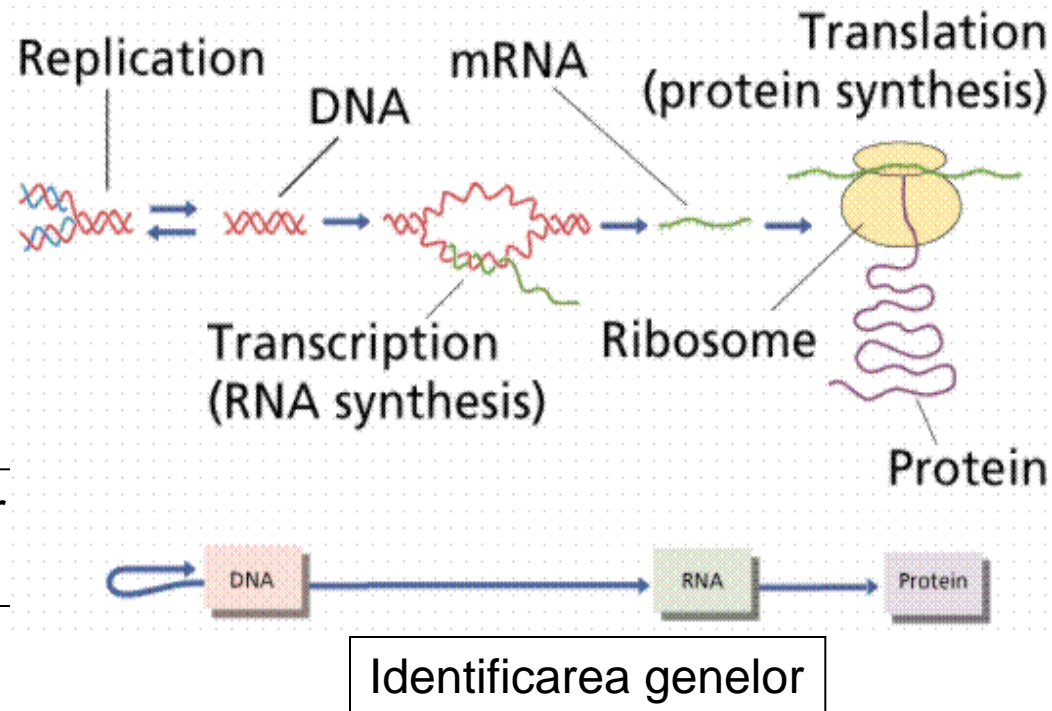
<http://www.youtube.com/watch?v=-gVh3z6MwdU>



W.K. Sung – Algorithms in Bioinformatics. A Practical Introduction, 2010

# Analiza genomului

- Prelucrări



Asamblare/  
secvențiere

Analiza secvențelor  
ADN

Identificarea genelor

Analiza/  
predicția  
structurii  
proteinelor

# Analiza genomului

Rolul bioinformaticii:

- Extragerea de informații din ADN
- Identificarea rolului unor gene prin analiza similarității cu gene al căror rol este cunoscut
- Predicția structurii proteinelor în vederea identificării rolului lor
- Efectuare de comparații între genomul corespunzător organismelor din aceeași specie sau din specii diferite

# Rezultate ale bioinformaticii

Alinierea și compararea secvențelor ADN

## BLAST (Basic Local Alignment Search Tool)

- instrument ce permite compararea dintre secvențe de interogare și elemente din bazele de date biologice
- utilizat pentru a prezice funcția unei secvențe ADN pe baza similarităților acesteia cu gene cunoscute
- stă la baza funcțiilor de căutare de secvențe în baze de date biologice (ex: GenBank)

# Probleme specifice in bioinformatica

- Aliniere secvențe ADN = stabilirea “distanței” dintre două secvențe, introducerea de spații între elemente pentru a maximiza numărul de elemente corespondente identice

- Ex: Secvențe inițiale                      Secvențe aliniate  
ATGTTAT                                      AT-GTTAT-  
ATCGTAC                                      ATCGT-A-C

- Problemă de optimizare: determinarea numărului minim de inserări, ștergeri sau înlocuiri pentru a ajunge de la un șir la alt șir
- Tehnici: programare dinamică, metode euristice

- Alt exemplu de secvențe ADN aliniate:

```
--T--CC-C-AGT--TATGT-CAGGGGACACG-A-GCATGCAGA-GAC  
AATTGCCGCC-GTCGT-T-TTCAG-----CA-GTTATG-T-CAGAT--C
```

# Probleme specifice in bioinformatica

- Stabilire similaritate între două secvențe ADN aliniate

GGAGACTGTAGACAGCTAATGCTATA  
GAACGCCCTAGCCACGAGCCCTTATC

- Sunt coincidențele dintre cele două secvențe doar întâmplătoare sau exprimă o similaritate semnificativă?
- **Tehnici:** statistice (estimare și verificare ipoteze)

# Probleme specifice in bioinformatica

- Mapare ADN = asociere între porțiuni de ADN și cromozomi (gene)
- Problema de căutare în șiruri a unor subșiruri
- Tehnici: algoritmi de căutare subșiruri (pattern matching)

# Probleme specifice in bioinformatica

- Secvențiere ADN = identificarea ordinii nucleotidelor
- Problema de optimizare: determinarea celui mai scurt șir ce conține un set de subșiruri (nu neapărat distincte)

Ex: 000, 001, 010, 011, 100, 111

“Suprașir” obținut prin concatenare: 00000101001100111

Cel mai scurt “suprașir”: 0001110100

- Instrumente: algoritmi de tip greedy, algoritmi de prelucrare a grafurilor



# Probleme specifice in bioinformatica

- Identificare “șabloane” ADN = identificarea unor secvențe scurte care apar frecvent în lanțul ADN
- Problema de căutare: se caută un șablon necunoscut dar frecvent
- Tehnici: căutare exhaustivă (ineficientă), algoritmi de tip branch and bound

# Baze de date biologice

## Exemple de baze de date biologice

NCBI GenBank <http://ncbi.nih.gov>  
(colecție mare de baze de date, ex: baza cu secvențe de nucleotide)

EMBL <http://www.embl-heidelberg.de>  
(baza de date cu nucleotide)

Protein Data Bank <http://www.pdb.org>  
(baza de date cu structuri ale proteinelor)

SWISSPROT <http://www.expasy.org/sprot/>  
(baza de date cu secvențe adnotate ale proteinelor)

PROSITE <http://kr.expasy.org/prosite>  
(baza de date cu pattern-uri asociate proteinelor)

# Instrumente software

- BLAST [<http://www.ncbi.nlm.nih.gov/blast/Blast.cgi>] – instrument de căutare în bazele de date care utilizează tehnici de aliniere locală
- EMBOSS [<http://emboss.sourceforge.net>] – pachet (open source) de prelucrare a datelor din biologia moleculară)
- ClustalW [<ftp://ftp-igbmc.ustrasbg.fr/pub/ClustalW/>] – instrument pentru alinierea multiplă a secvențelor

# Extensii ale limbajelor de programare

- BioPerl [<http://www.bioperl.org/>] - conține module care facilitează scrierea unor scripturi în Perl destinate rezolvării unor probleme din bioinformatică
- BioJava [[http://biojava.org/wiki/Main\\_Page](http://biojava.org/wiki/Main_Page)] – framework Java pt prelucrarea datelor biologice
- BioPython [<http://biopython.org/wiki/Biopython>] - set de instrumente dezvoltate în Python pentru prelucrarea datelor biologice

# Instrumente software

- Bioconductor [<http://www.bioconductor.org/>] – extensie pentru bioinformatică a pachetului statistic R [<http://www.r-project.org/>]
- BioWeka [[http://bioweka.sourceforge.net/index.php/Main\\_Page](http://bioweka.sourceforge.net/index.php/Main_Page)] – extensie pentru bioinformatică a pachetului Weka utilizat în data mining
- MatLab Bioinformatics ToolBox [<http://www.mathworks.com/>]