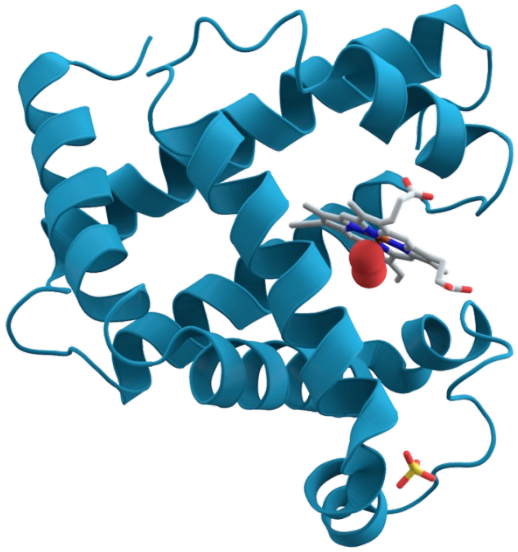


18.03 – 1.04.2026

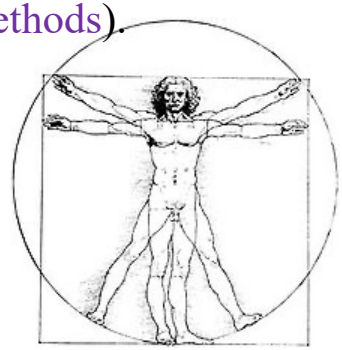
Curs 6 – Fișiere cu secvențe. Operații simple cu secvențe

De unde provin secvențele proteice?

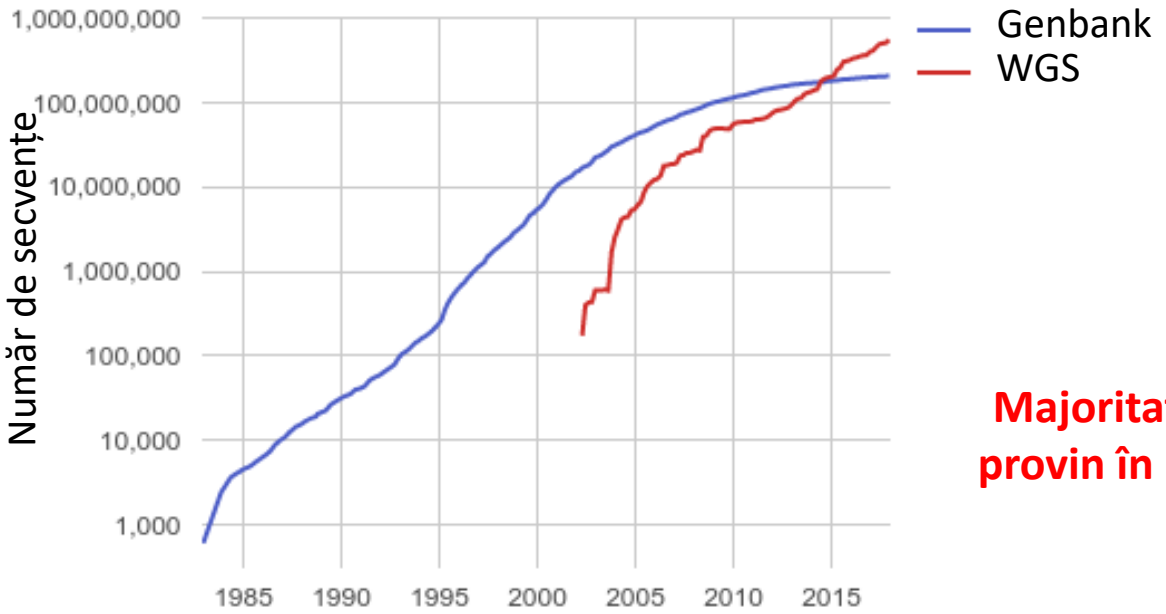
Spre deosebire de metodele de secvențiere a proteinelor, stabilirea ordinii nucleotidelor în acizii nucleici ADN și ARN este mult mai rapidă și ieftină. În ultimii ani metoda clasică Sanger a fost înlocuită de metode de `generație nouă`` (next-gen) ce au un randament și o viteză foarte mare (high-throughput sequencing methods).

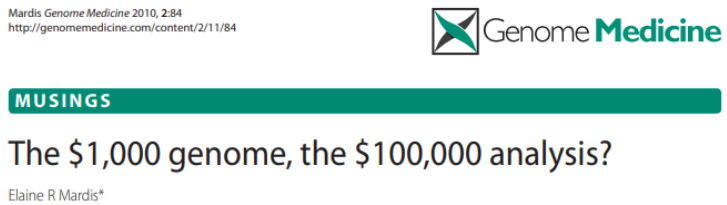
Secvențierea genomului uman - 3 bilioane de baze – 10 ani (1990 -2000) - **~2.7 bilioane $**
Secvențierea unui genom uman în 2006 - ~6 bilioane baze - **~14 milioane $**
Secvențierea unui genom uman în 2015 - ~6 bilioane baze - **~4000 $**
Secvențierea unui genom uman în 2016 - ~6 bilioane baze - **~1000 $**

https://www.genome.gov/27565109/the-cost-of-sequencing-a-human-genome/



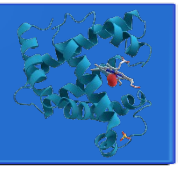
Human Genome Project (HGP, USA)





Majoritatea covârșitoare a secvențelor proteice provin în prezent din secvența de nucleotide prin treducere *in-silico*.

Fișiere FASTA - fișiere cu secvențe



O secvență, fie că este ADN, ARN sau proteine **reprezintă o înșiruire de litere** (A; T; G;C pentru ADN, A;U;G;C pentru ARN, cei 20/22 de aminoacizi notați cu o literă pentru proteine) **ce poate fi foarte ușor stocată într-un fișier text**.

În general o secvență este însoțită de o serie **de informații accesorii precum specia de la care provine secvența, gena, cromozomul, lucrarea în care este descrisă secvența**. **Modul în care sunt organizate și stocate aceste informații alături de secvența propriu-zisă reprezintă formatul sau standardul unui fișier cu secvențe**.

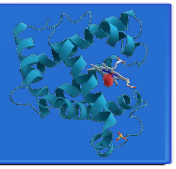
De-a lungul timpului au existat un număr mare de variante de fișiere/formate create pentru a înregistra secvențe, însă fișierul/formatul FASTA este cel ce s-a impus.

Fișierul FASTA – Fast-All

```
>AJ507836.1 Arthrobacter nicotinovorans pAO1 megaplasmid sequence, strain ATCC
49919
GATCCGGCGGTCCGCCGTCGTGGCGGCGCGGGCGGAGGTTGCCGGCGGCGGACCGCCGCCCGCACAAGAA
GGCCTTCGGGTTCCGGCAGGTGGCGCGGCCCCCGACACTGGTCCTCGCCTGGGGTGGAACGTGGGGTGCT
GGGTGTGGGCGGTTCGCCGGGCGAACCGGGAAAGGCGTCCACTCCTCTTCGCTTCCGTGGCCGGCGGTTG
GGGCCGGTCCCGTCGTCGCAGAGCTCCGCCGTGCCCGGCCCCGCCCGTTGTCTACGACGTCTTTTTGTGG
CTGTCCTTCGGATCATACGGTTCCGCTCCAGCTCAAGCTCCTGGCAGTCCGCAAAGCTCCGGTCTAGCAC
ACAGCGATGCGGGTAATGATGGCGACGAATTCGTCCCAGAGCGCTGGTGCGATTTCGTGGCCCATACCTG

>gi|5524211|gb|AAD44166.1| cytochrome b [Elephas maximus maximus]
LCLYTHIGRNIYYGSYLYSETWNTGIMLLLITMATAFMGYVLPWGQMSFWGATVITNLFSAIPYIGTNLV
EWIWGGFSVDKATLNRFFAFHFILPFTMVALAGVHLTFLHETGSNNPLGLTSDSDKIPFHPYYTIKDFLG
LLILILLLLLLALLSPDMLGDPDNHMPADPLNTPLHIKPEWYFLFAYAILRSVPNKLGGVLALFLSIVIL
GLMPFLHTSKHRSMMLRPLSQALFWTLTMDLLTLTWIGSQPVEYPYTIIGQMASILYFSIILAFLPIAGX
IENY
```

Fișiere FASTA - fișiere cu secvențe



Fișierul FASTA

- un fișier text ce poate **avea sau nu** extensia .fasta;
- conține secvențe de proteine sau de aminocizi stocate conform **standardului FASTA**:

Elementele formatului FASTA:

A.primul rând de text, marcat cu "**>**" (mai mare) conţine o serie de **informaţii cu caracter opţional**, - specia sau denumirea genei (proteinei);

B.următoarele rânduri conţin secvenţa propriu-zisă, în care nucleotidele/aminoacizii sunt reprezentați folosind codul standard IUPAC cu o singură literă;

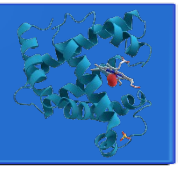
C.fiecare rând al secvenței are în general 80 de caractere (nu mai mult de 120)

A	adenozină	**M**	A/C (amino)
T	timină	**W**	A/T (weak)
C	citidină	**R**	G/A (puRine)
U	uracil	**B**	G/T/C
G	guanină	**D**	G/A/T
N	A/G/C/T (oricare)	**H**	A/C/T
K	G/T (keto)	**V**	G/C/A
S	G/C (strong)	**-**	spatiu de dimensiune intermediară
Y	T/C (p**Y**rimidine)		

Standardul IUPAC pentru notarea nucleotidelor și aminoacizilor

A	alanină	**P**	prolină
B	aspartat/asparagină	**Q**	glutamină
C	cistină	**R**	arginină
D	aspartat	**S**	serină
E	glutamat	**T**	threonină
F	fenilalanină	**U**	selenocisteină
G	glicină	**V**	valină
H	histidină	**W**	triptofan
I	isoleucină	**Y**	tirosină
K	lisină	**Z**	glutamat/glutamin
L	leucină	**X**	oricare
M	metionină	*****	stop
N	asparagină	**-**	spațiu

Fișiere FASTA - fișiere cu secvențe



Fișierul FASTA

D. O secvență de nucleotide este notată în direcția 5' → 3‘ : prima literă de pe primul rând corespunzător secvenței este nucleotida 1 și are o grupare PO_4 liberă, ultima literă de pe ultimul rând are gruparea OH 3´ liberă. Moleculele circulare se reprezentate linear, prima nucleotidă fiind cel mai frecvent originea de replicare;

E. O secvență polipeptdică este notată în sensul sintezei, de la capătul N terminal spre cel C terminal; prima literă din primul rând reprezintă aminoacidul 1 din secvență - aminoacidul N-terminal; ultima literă reprezintă aminoacidul C-terminal;

F. Poziția literelor poate fi sau nu numerotată; în cazul în care literele din secvență sunt numerotate, numeroatarea se face la începutul fiecărui rând și se include un spațiu după fiecare a 10-a literă.

G. **Literele ce desemnează secvența pot fi sau nu scrise cu majuscule**; indiferent de tipul de scriere, semnificația este aceeași;

H. Unele programe nu acceptă caracterul ‘–’ (spațiul în secvență) se indică cu un sir de N pentru nucleotide sau X pentru aminoacizi; **spațiul între litere este ignorat**;

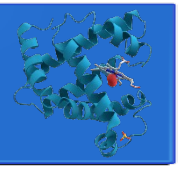
```
>secventa peptidica necunoscuta 1 fara numere
QIKDLLVSSSTDLDTTLVLVNAIYFKGMWKTAFNAEDTREMPFHVTKQESKPVQMMCMNNSFNVATLPAE
KMKILELPFASGDLSMLVLLPDEVSDLERIEKTINFEKLTEWTNPNTMEKRRVKVYLPQMKIEEKYNLTS
VLMALGMTDLFIPSANLTGISSAESLKISQAVHGAFMELSEDGIEMAGSTGVIEDIKHSPESEQFRADHP
FLFLIKHNPTNTIVYFGRYWSP

>secventa peptidica necunoscuta cu numere
1   qikdllvsss tdldttlvlv naiyfkgmwk tafnaedtre mpfhvtkqes kpvqmmcmnn
61  sfnvatlpae kmkilelpfa sgdlsmlvll pdevsdleri ektinfeklt ewtnpntmek
121 rrvkvylpqm kieekynlts vlmalgmtdl fipsanltgi ssaeslkisq avhgafmels
181 edgiemagst gviedikhsp eseqfradhp flflikhnpt ntivyfgryw sp
```

Cum scriu secventele?
- Cu font **monospatiat**

TNR Proportional
Courier New Monospace

De unde pot obține sevențe?



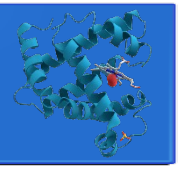
1. Un instrument de secvențiere a ADN-ului – cel mai frecvent;
2. Un instrument de secvențiere a proteinelor bazat pe degradarea Edman (cel mai putin frecvent) sau un spectrometru de masă – destul de rar;
3. **O bază de date cu secvențe** – cel mai ușor de accesat; create de utilizatori ce au acces la instrumentele 1 și 2 și deci pot determina experimental o secvență;

O bază de date cu secvențe reprezintă **o colecție de secvențe de acizi nucleici sau aminoacizi ce au fost stabilite experimental și care au fost depozitate în formă digitalizată pe un server central într-un format tip specific**. Fiecărei secvențe i se alocă un **identificator unic** – **ID** – **o combinație unică de litere și cifre ce poate fi folosită pentru a regăsi fără echivoc secvența în respectiva bază de date**. În general, accesul la bazele de date cu secvențe este gratuit.

Bază de date	Site web	Dimensiune *
Baze de date cu secvenţe		
INSDC	http://www.insdc.org/	206 293 625 secvenţe Aceste baze de date sunt sincronizate zilnic între ele şi conţin aceleaşi secvenţe.
DDBJ	http://www.ddbj.nig.ac.jp	
EMBL Nucleotide Sequence Database	http://www.ebi.ac.uk/embl/	
GenBank	http://www.ncbi.nlm.nih.gov/genbank/	
European Nucleotide Archive (ENA)	http://www.ebi.ac.uk/ena/	
KEGG	http://www.genome.jp/kegg/	25 679 056 secvenţe
nr	http://blast.ncbi.nlm.nih.gov	145 296 712 secvenţe
UniProtKB/Swiss-Prot	http://www.uniprot.org/	556 568 secvenţe
UniProtKB/TrEMBL	http://www.uniprot.org/	107 627 435 secvenţe

* În Februarie 2018

Redundanță și incertitudine în bazele de date



Secvenţele sunt înregistrate în bazele de date de către o multitudine de utilizatori, independenți unul de celălalt. În momentul înregistrării unei secvențe aceasta primește un ID unic și este tratată ca entitate de sine stătătoare. Pot apare astfel 2 situații:

I

aceeaşi secvență este înregistrată cu nume diferite

aceeaşi proteină/genă este secvențiată de la specii diferite

aceeaşi secvență este înregistrată de autori diferiţi

II

Mai multe ID pentru aceeaşi secvenţă

II

V

Redundanță

2 secvențe marcate ca diferite pot avea aceeși funcție

II

Datele de secvențe sunt înregistrate automat. Metodele moderne de secvențiere produc o cantitate mare de date ce sunt analizate și înregistrate în bazele de date automat.

II

Funcția secvențelor este stabilită automat și nu este verificată experimental.

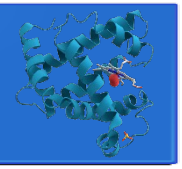
II

V

Nesiguranță

1 secvență poate fi identificată ca având o anumită funcție în mod eronat.

Principalele baze de date cu secvențe



GenBank https://www.ncbi.nlm.nih.gov/genbank/

- o bază de date cu secvențe ADN creată și întreținută de National Institutes of Health (NIH), USA;
- conține **secvențe adnotate** – pe lângă secvența propriu-zisă, fiecare ID în baza de date conține și o serie de informații accesorii precum:

D36–D42 Nucleic Acids Research, 2013, Vol. 41, Database issue *Published online 27 November 2012*
doi:10.1093/nar/gks1195

GenBank

Dennis A. Benson, Mark Cavanaugh, Karen Clark, Ilene Karsch-Mizrachi, David J. Lipman, James Ostell and Eric W. Sayers*

National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Building 38A, 8600 Rockville Pike, Bethesda, MD 20894, USA

Received September 28, 2012; Revised and Accepted October 29, 2012

1. Denumirea
2. GenBank ID
3.
4.
5. Incadrarea taxonomica
6. Lucrări științifice

```
LOCUS       AJ507836              165137 bp    DNA     linear   BCT 24-JUL-2016
DEFINITION  Arthrobacter nicotinovorans pAO1 megaplasmid sequence, strain ATCC
            49919.
ACCESSION   AJ507836
VERSION     AJ507836.1
KEYWORDS    .
SOURCE      Paenarthrobacter nicotinovorans
  ORGANISM  Paenarthrobacter nicotinovorans
            Bacteria; Actinobacteria; Micrococcales; Micrococcaceae;
            Paenarthrobacter.
REFERENCE   6  (bases 1 to 165137)
  AUTHORS   Igloi,G.L. and Brandsch,R.
  TITLE     Sequence of the 165-kilobase catabolic plasmid pAO1 from
            Arthrobacter nicotinovorans and identification of a pAO1-dependent
            nicotine uptake system
  JOURNAL   J. Bacteriol. 185 (6), 1976-1986 (2003)
   PUBMED   12618462
REFERENCE   18 (bases 27948 to 117790)
  AUTHORS   Mihasan,M. and Brandsch,R.
  TITLE     pAO1 of Arthrobacter nicotinovorans and the spread of catabolic
            traits by horizontal gene transfer in gram-positive soil bacteria
  JOURNAL   J. Mol. Evol. 77 (1-2), 22-30 (2013)
   PUBMED   23884627
REFERENCE   19 (bases 1 to 165137)
  AUTHORS   Igloi,G.L. and Brandsch,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-SEP-2002) Institute of Biochemistry and Molecular
            Biology, University of Freiburg, Hermann-Herder-Str. 7, Freiburg
            D-79104, GERMANY
REFERENCE   20 (bases 1 to 165137)
  AUTHORS   Mihasan,M. and Brandsch,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (23-DEC-2013) Biochemistry and Molecular Biology
            Laboratory, Faculty of Biology, Alexandru Ioan Cuza University,
            Copou Bvd, No.22B, Iasi 700506, ROMANIA
REMARK      revision by M. Mihasan, authorised by R. Brandsch.
```

7. Autorul ce a solicitat înregistrarea și cel responsabil de secvență

8. Gene identificate cu:

```
FEATURES             Location/Qualifiers
     source          1..165137
                     /organism="Paenarthrobacter nicotinovorans"
                     /mol_type="genomic DNA"
                     /strain="ATCC 49919"
                     /culture_collection="ATCC:49919"
                     /db_xref="taxon:29320"
                     /plasmid="pAO1"
     CDS             complement(344..1138)
                     /experiment="non-experimental evidence, no additional
                     details recorded"
                     /note="ORF2"
                     /codon_start=1
                     /transl_table=11
                     /product="putative hydrolase"
                     /protein_id="CAD47860.1"
                     /db_xref="GOA:Q8GAP3"
                     /db_xref="InterPro:IPR000073"
                     /db_xref="UniProtKB/TrEMBL:Q8GAP3"
                     /translation="MPTTPSGIYWESQGPVAAPAVLLIEGYTGQLIGWRDAFCDLLLA
                     QGLRVLRMDNRDVGLSRREDGNYMIADMADDVIDVIADADLGKITIVGQSMGGLIAQH
                     TALGYPNMVTGLVLFYTTPVLDDIDPGILTAEIPRPQHREDAIASFLEGDRATASPAW
                     GYDEAWKRELAGRMFDRAPDRSGLSRQRNAVALLPDLRPRLTELTMPVALIHGRNDAL
                     IRARGSLRIAEVVPQAELHLYPGMGHEIAPALWDEFVAIITRIAVC"
ORIGIN
        1 gatccggcgg tccgccgtcg tggcggcgcg ggcggaggtt gccggcggcg gaccgccgcc
       61 cgcacaagaa ggccttcggg ttccggcagg tggcgcggcc cccgacactg gtcctcgcct
      121 ggggtggaac gtggggtgct gggtgtgggc ggttcgccgg gcgaaccggg aaaggcgtcc
      181 actcctcttc gcttccgtgg ccggcggttg gggccggtcc cgtcgtcgca gagctccgcc
      241 gtgcccggcc ccgcccgttg tctacgacgt ctttttgtgg ctgtccttcg gatcatacgg
      301 ttccgctcca gctcaagctc ctggcagtcc gcaaagctcc ggtctagcac acagcgatgc
      361 gggtaatgat ggcgacgaat tcgtcccaga gcgctggtgc gatttcgtgg cccatacctg
      421 ggtagaggtg cagttcagct tgggggacga cttcggcgat cctcagggat ccacgggctc
      481 ggatgagggc gtcgttccgg ccatgtatga gtgcgaccgg catcgtgagt tcagtaagtc
      541 tcggccgcag atcaggcaac agtgcaactg cgttcctttg ccgggataag ccgctgcggt
      601 cgggggcacg gtcaaacatg cggcctgcca gttcccgctt ccaggcctcg tcgtaacccc
      661 atgccggtga ggctgtggca cggtcgcctt ccaggaacga ggcgatggca tcctcccggt
      721 gttgaaggcg cggaatctcc gcggtcagga tcccgggatc gatgtcgtcg agtacgggcg
```

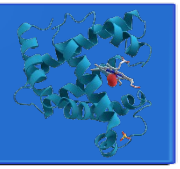
a. Poziția pe secvență și catena codificatoare

b. denumire genei

c. secvența proteinei codificate

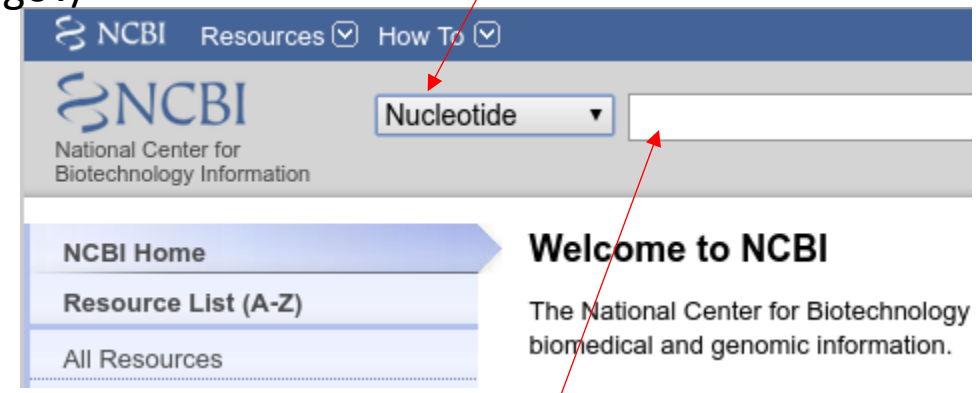
9. Secvența de ADN propriu-zisă

Obținerea unei secvențe din GenBank



1. Accesează https://www.ncbi.nlm.nih.gov/

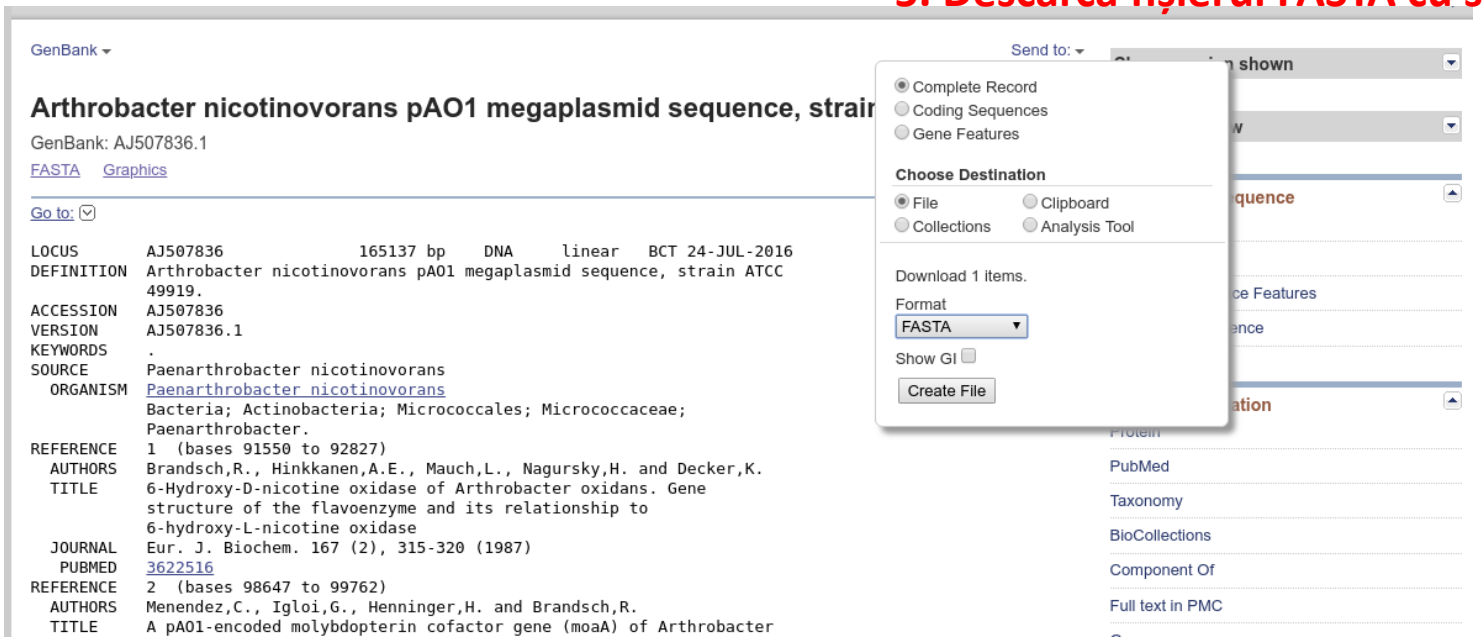
2. Selectează Nucleotide din meniu



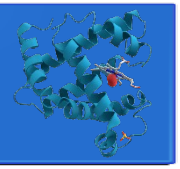
3. Scrie termenul de căutare

4. Selectează secvența dorită

5. Descarcă fișierul FASTA cu secvența dorită



Principalele baze de date cu secvențe



UniProtKB/TrEMBL - http://www.uniprot.org/
- conţine toate secvenţele de nucleotide din GenBank translate în secvenţe de aminoacizi;
- are un nivel foarte mare de redundanță;

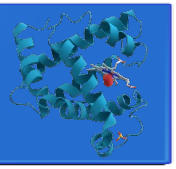
UniProtKB/SwissProt - http://www.uniprot.org/
-toate secvenţele conţinute sunt adnotate manual – siguranță crescută legată de funcțiile secvențelor conținute;
-nivelul de redundanţă este minim;
-adăugarea unei noi secvenţe în SwissProt este un proces care implică prelucrarea manuală a datelor corespunzătoare secvenţei respective => SwissProt nu conține cele mai noi secvențe.

TrEMBL este completă dar incorectă, SwissProt este mai corectă, dar incompletă.

Protein Identification Resource PIR
-o colecţie completă de secvenţe lipsită de redundanţă;
-secvenţele identice sau foarte similare sunt gruparea într-o singură intrare în baza de date;

nr
- o bază de date non-redundantă completă, cuprinzând intrări din GenBank, SwissProt, PIR, Protein Research Foundation (PRF) şi Protein data Bank (PDB).

Operații simple cu secvențe – SMS2



http://www.bioinformatics.org/sms2/

https://mail.uaic.ro/~marius.mihasan/research/mirrored_sites_tools/sms2/index.html

Operații cu secvențe selectabile în SMS2

SMS

Format Conversion
-Combine FASTA
-EMBL to FASTA
-EMBL Feature Extractor
-EMBL Trans Extractor
-Filter DNA
-Filter Protein
-GenBank to FASTA
-GenBank Feature Extractor
-GenBank Trans Extractor
-One to Three
-Range Extractor DNA
-Range Extractor Protein
-Reverse Complement
-Split Codons
-Split FASTA
-Three to One
-Window Extractor DNA
-Window Extractor Protein

Sequence Analysis
-Codon Plot
-Codon Usage
-CpG Islands
-DNA Molecular Weight
-DNA Pattern Find
-DNA Stats
-Fuzzy Search DNA
-Fuzzy Search Protein
-Ident and Sim
-Multi Rev Trans
-Mutate for Digest
-ORF Finder
-Pairwise Align Codons
-Pairwise Align DNA
-Pairwise Align Protein
-PCR Primer Stats
-PCR Products
-Protein GRAVY
-Protein Isoelectric Point
-Protein Molecular Weight
-Protein Pattern Find
-Protein Stats
-Restriction Digest
-Restriction Summary
-Reverse Translate
-Translate

Sequence Figures
-Color Align Conservation
-Color Align Properties
-Group DNA
-Group Protein

Sequence Manipulation Suite:

Translate

Translate accepts a DNA sequence and converts it into a protein in the reading frame you specify. Translate supports the entire IUPAC alphabet and several genetic codes.

Paste a raw sequence or one or more FASTA sequences into the text area below. Input limit is 200000 characters.

```
>Arthrobacter nicotinovorans orf388 sequence
ctactaaccctcagcgccgccccttcaaccctttttctctgaggaaaatatgactaaaaca
gcgattgtccgcgtcgccatgaatggcatcaccggccggatgggctaccgccagcacctg
ccgcgacgccggcggcttcaccctcgaagacggcaccaaggtccagatcgaaccgatcct
cgtaggccgcaacgaagccaagatccgcgaactcgccgagaagcacaaggttgccgagtg
gagcacggacctggactcggtcgtcaacgaccccaccgtcgacatcatcttcgacgcctc
catgaccagcctccgcgccgccaccctgaagaaggcgatgctggccggcaagcacatctt
caccgagaagcccaccgcggaaaccctggaagaggccattgaactggcccgcatcggcaa
gcaggcaggcgtcaccgcaggcgtcgtacacgacaagctgtacctcccgggcctggtcaa
gctccgccgcctggtggacgaaggcttcttcggccgcatcctgtccatccgcggtgagtt
cggctactgggtctttgaaggtgacgttcaggcagcacagcgcccgtcctggaactaccg
caaggaagacggcggtggaatgaccacggacatgttctgccactggaactacgtccttga
aggcatcatcggcaaggtcaagagcgtcaacgccaagaccgccacgcacatccccacccg
ctgggacgaagccggcaaggagtacaaggcaacggctgatgacgcttcctacggcatctt
cgagcttgaaaccccgggcggcgacgacgtcatcggccagatcaactcttcctgggccgt
ccgcgtctaccgcgacgaactcgtcgaattccaggtggacggcacccacggctccgccgt
tgccggcctgaacaagtgcgtcgcccagcagcgcgcacacacccccaagccggtctggaa
ccctgacctgcccgtcaccgagtccttccgcgaccagtggcaggaagtccccgccaacgc
```

Please check the browser compatibility page before using this program.

Submit | Clear | Reset

- Translate in reading frame 1 on the direct strand.
- Use the standard (1) genetic code.

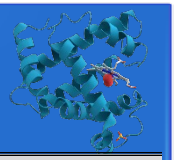
*This page requires JavaScript. See browser compatibility.
*You can mirror this page or use it off-line.

new window | home | citation

Fri Mar 30 17:46:57 2012

Căsuță text pentru secvență FASTA

Operații simple cu secvențe – SMS2



Denumire operaţie*	Descriere
Convertire între diverse formate	
Combine FASTA	– permite combinarea a două sau a mai multe secvenţe în format FASTA şi obţinerea unei singure secvenţe
EMBL to FASTA	– permite convertirea unei secvenţe din formatul EMBL în formatul FASTA; funcţia este utilă în situaţia în care se doreşte eliminarea rapidă a informaţiilor care nu au legătură cu secvenţa de ADN dintr-un fişier EMBL
Filter DNA	– elimină caracterele care nu corespund codului standard IUPAC pentru ADN-ul dintr-o secvenţă (inclusiv spaţii, numere, etc)
Filter Protein	– elimină caracterele ce nu corespund cu codul standard IUPAC de o literă pentru aminoacizi dintr-o secvenţă (inclusiv spaţii, numere etc.)
GenBank to FASTA	– permite convertirea unei secvenţe din formatul GenBank în formatul FASTA; funcţia este utilă în situaţia în care se doreşte eliminarea rapidă a informaţiilor care nu au legătură cu secvenţa de ADN dintr-un fişier GenBank
One to Three	– permite convertirea unei secvenţe de aminoacizi din codul standard IUPAC de o literă în cel de trei litere
Analiza secvenţelor	
Codon Plot	– analizează frecvenţa de utilizare a codonilor dintr-o secvenţă de ADN şi generează un grafic cu aceste frecvenţe
Codon Usage	– analizează frecvenţa de utilizare a codonilor dintr-o secvenţă de ADN şi poate fi utilizat pentru a identifica preferinţa pentru un anumit codon sinonim
DNA Molecular Weight	– calculează masa moleculară a uneia sau a mai multor secvenţe de ADN
DNA Stats	– calculează frecvenţa fiecărei nucleotide într-o secvenţă dată, rezultatele fiind exprimate în procente
PCR Primer Stats	– analizează şi calculează proprietăţile specifice ale unui set indicat de primeri, inclusiv temperatura de topire şi procentul de GC
PCR Products	– realizează amplificarea virtuală prin PCR a unei secvenţe date folosind un set de primeri indicat de utilizator
Protein Isoelectric Point	– calculează valoarea teoretică a punctului izoelectric pentru o secvenţă dată
Protein Molecular Weight	– calculează valoarea teoretică a masei moleculare pentru una sau mai multe secvenţe de aminoacizi
Protein Stats	– calculează frecvenţa fiecărui aminoacid într-o secvenţă dată, rezultatele fiind exprimate în procente
Translate	– realizează translaţia şi transcripţia virtuală a unei secvenţe de ADN date, utilizatorul având posibilitatea de a alege cadrul de citire specific
Reverse Translate	– realizează procesul invers faţă de Translate, acceptând o secvenţă de aminoacizi şi generând secvenţa ADN corespunzătoare

* corespunzătoare cu denumirea din suita de programe SMS2