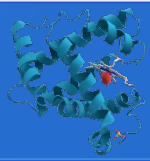


27.03. – 3.04.2024

Curs 6 – Fișiere cu secvențe. Operații simple cu secvențe

De unde provin secvențele proteice?



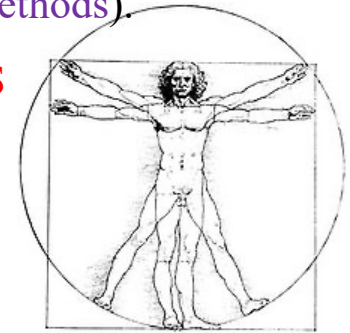
Spre deosebire de metodele de secvențiere a proteinelor, stabilirea ordinii nucleotidelor în acizii nucleici ADN și ARN este mult mai rapidă și ieftină. În ultimii ani metoda clasică Sanger a fost înlocuită de metode de ` generație nouă` (next-gen) ce au un randament și o viteză foarte mare (high-throughput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

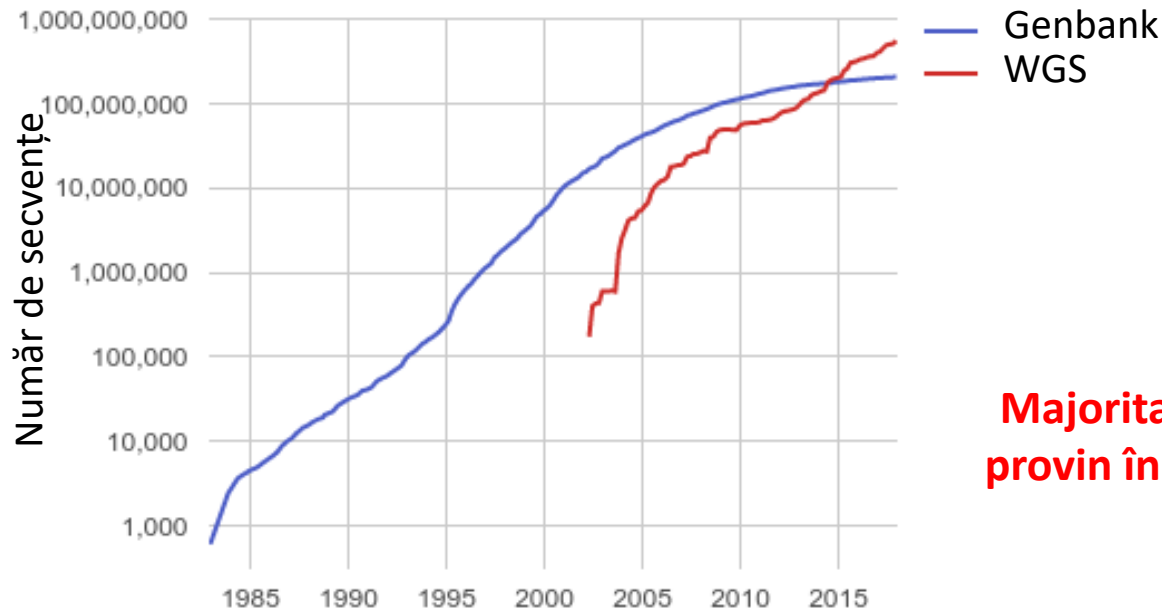
Secvențierea unui genom uman în 2015 - ~6 bilioane baze - ~4000 \$

Secvențierea unui genom uman în 2016 - ~6 bilioane baze - ~1000 \$



Human Genome Project (HGP, USA)

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



Mardis Genome Medicine 2010, 2:84
<http://genomemedicine.com/content/2/11/84>



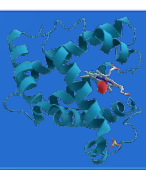
MUSINGS

The \$1,000 genome, the \$100,000 analysis?

Elaine R Mardis*

Majoritatea covârșitoare a secvențelor proteice provin în prezent din secvența de nucleotide prin *reducere 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.**

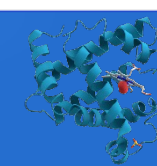
De-a lungul timpului au existat un număr mare de variante de fișiere/formate create pentru a înregistra secvențe, însă fișierul/formatul FASTA este cel ce s-a impus.

Fișierul FASTA – Fast-All

```
>AJ507836.1 Arthrobacter nicotinovorans pAO1 megaplasmid sequence, strain ATCC
49919
GATCCGGCGGTCCGCCGTCGTGGCGGCGGGCGGAGGTTGCCGGCGGCGGACCGCCGCCCGCACAAGAA
GGCCTTCGGGTTCCGGCAGGTGGCGCGGCCCCCGACACTGGTCCTCGCCTGGGGTGGAACGTGGGGTGCT
GGGTGTGGGCGGTTCCGCCGGGCGAACCGGGAAAGGCGTCCACTCCTCTTCGCTTCCGTGGCCGGCGGTTG
GGGCCGGTCCCCTCGTTCGCAGAGCTCCGCCGTGCCCGGCCCGCCCGTTGTCTACGACGTCTTTTTGTGG
CTGTCTTCGGATCATAACGGTTCGCTCCAGCTCAAGCTCCTGGCAGTCCGCAAAGCTCCGGTCTAGCAC
ACAGCGATGCGGGTAATGATGGCGACGAATTCGTCCCAGAGCGCTGGTGCGATTTTCGTGGCCCATACCTG
```

```
>gi|5524211|gb|AAD44166.1| cytochrome b [Elephas maximus maximus]
LCLYTHIGRNIYYGSYLYSETWNTGIMLLLLITMATAFMGYVLPWGQMSFWGATVITNLFSaipyIGTNLV
EWIWGGFsvdkatlnrffafhfILPFTMVALAGVHLTFLHETGSNNPLGLTSDSDKIPFHPYYTIKDFLG
LLILILLLLLLLALLSPDMLGDPDNHMPADPLNTPHlKPEWYFLFAYAILRSVPNKLGGVLALFLSIVIL
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) conține o serie de **informații cu caracter opțional**, - specia sau denumirea genei (proteinei);

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

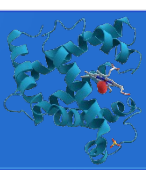
C. fiecare rând al secvenței are în general 80 de caractere (nu mai mult de 120)

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
Y	T/C (pYrimidine)		dimensiune
			intermediară

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

Standardul IUPAC pentru notarea nucleotidelor și aminoacizilor

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 PO₄ liberă, ultima literă de pe ultimul rând are gruparea OH 3' liberă. Moleculele circulare se reprezintă linear, prima nucleotidă fiind cel mai frecvent originea de replicare;

E. O secvență polipeptidică 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
QIKDLLVSSSTDLDTTLVLVNAIYFKGMWKTAFAEAEDTREMPEFHVTKQESKPVQMMCMNNSFNVATLPAE
KMKILELPPASGDLMLVLLPDEVSDLERIEKTINFELTEWTNPNTMEKRRVKVYLPQMKIEEKYNLTS
VLMALGMTDLFIPSANLTGISSAESLKISQAVHGAFMELSEDGIEMAGSTGVIEDIKHSPESQFRADHP
FLFLIKHNPTNTIVYFGRYWSP
```

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

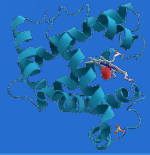
Cum scriu secvențele?

- Cu font **monospațiat**

TNR **Proportional**

Courier New **Monospace**

De unde pot obține secvențe?



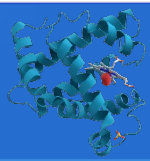
1. Un instrument de secvențiere a ADN-ului – cel mai frecvent;
2. Un instrument de secvențiere a proteinelor bazat pe degradarea Edman (cel mai puțin frecvent) sau un spectrometru de masă – destul de rar;
3. **O bază de date cu secvențe** – cel mai ușor de accesat; create de utilizatori ce au acces la 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 secvențe		
INSDC	http://www.insdc.org/	206 293 625 secvențe
DDBJ	http://www.ddbj.nig.ac.jp	Aceste baze de date sunt sincronizate zilnic între ele și conțin aceleași secvențe.
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 secvențe
nr	http://blast.ncbi.nlm.nih.gov	145 296 712 secvențe
UniProtKB/Swiss-Prot	http://www.uniprot.org/	556 568 secvențe
UniProtKB/TrEMBL	http://www.uniprot.org/	107 627 435 secvențe

* În Februarie 2018

Redundanță și incertitudine în bazele de date



Secvențele 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 apărea astfel 2 situații:

I

aceeași secvență este înregistrată cu nume diferite
aceeași proteină/genă este secvențiată de la specii diferite
aceeași secvență este înregistrată de autori diferiți

||

Mai multe ID pentru aceeași secvență

||
v

Redundanță

2 secvențe marcate ca diferite pot avea aceeași 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.

||

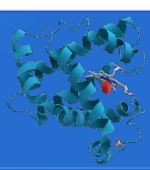
**Funcția secvențelor este stabilită automat și nu
este verificată experimental.**

||
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:

- 1. Denumirea
- 2. GenBank ID
- 3. Accesion number
- 4. Organism
- 5. Incadrarea taxonomică
- 6. Lucrări științifice
- 7. Autorul ce a solicitat înregistrarea și cel responsabil de secvență

LOCUS	AJ507836	165137 bp	DNA	linear	BCT 24-JUL-2016
DEFINITION	Arthrobacter nicotinovorans pA01 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 pA01 from Arthrobacter nicotinovorans and identification of a pA01-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	pA01 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.				

8. Gene identificate cu:

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

```
-----
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="pA01"
                     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="MPTTPSGIYWESQGPVAAPVLLIEGYTGQLIGWRDAFCDLLLA
                     QGLRVLRLMNRDVGLSRREDGNYMIADMDADDVIDVIADADLKGITIVGQSMGGLIAQH
                     TALGYPNMVTGLVLYFTTLPVLDIDDPGILTAEIPRPQHREDAIASFLEGDRATASPAW
                     GYDEAWKRELAGRMFDRAPDRSGLSRQRNAVALLPDLRPRTELTMPTVALIHGRNDAL
                     IRARGSLRIAEVVPQAEHLHYPGMGHEIAPALWDFVAIITRIAVC"|
ORIGIN
1  gatccggcgg  tccgctcgct  tggcggcgcg  ggcggaggtt  gccggcggcg  gaccgccgcc
61  cgcacaagaa  ggccttcggg  ttccggcagg  tggcggcgcc  cccgacactg  gtccctgcct
121  ggggtggaac  gtgggggtgt  ggggtggtgg  gtttcgccgg  gcgaaccggg  aaaggcgctc
181  actcctcttc  gcttcctggt  ccggcggttg  gggcggtccc  cgtcgtcgca  gagctccgcc
241  gtgcccgggc  ccgccctgtg  tctacgacgt  cttttgtgtg  ctgtccttcg  gatcatacgg
301  ttccgctcca  gctcaagctc  ctggcagtc  gccaaagctc  ggtctagcac  acagegatgc
361  gggtaagtgt  ggcgacgaat  tcgtcccaga  gcgctggtgc  gatttcgttg  cccatacctg
421  ggtagagggt  cagttcagct  tgggggacga  cttcggcgcc  cctcagggat  ccacgggctc
481  ggatgagggc  gtcgttcggg  ccatgtatga  gtgcgaccgt  catcgtgaat  tcagtaagtc
541  tcggccgcag  atcaggcaac  agtgaactgt  cgttcctttg  ccgggataag  ccgctgcggt
601  cggggggcac  gtcaaacatg  cgccctgcca  gttcccgcct  ccaggcctcg  tcgtaacccc
661  atgccgggtg  agcgtgtggc  cggtgccttt  ccaggaacga  ggcgatggca  tctcccggtt
721  attaaaaqca  caaaatctcc  acaatcaqaa  tcccaaatc  aatcatcaca  aatcaaaaaa
```

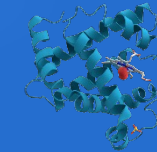
a. Poziția pe secvență și catena codificatoare

b. denumire genei

c. secvența proteinei codificate

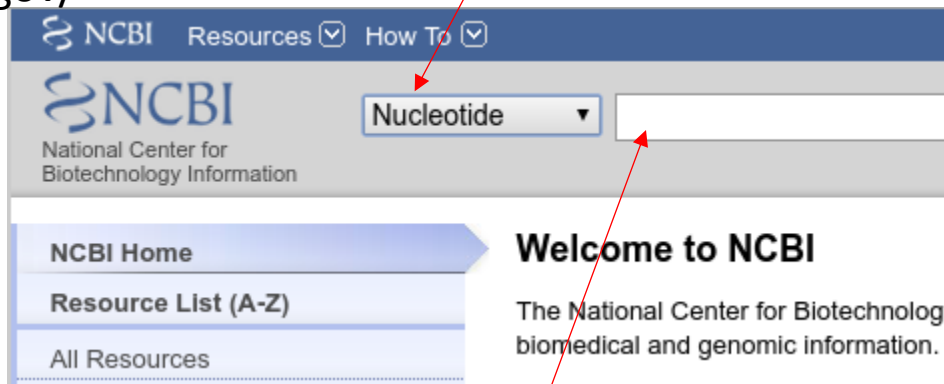
9. Secvența de ADN propriu-zisă

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

GenBank

Arthrobacter nicotinovorans pAO1 megaplasmid sequence, strain ATCC 49919

GenBank: AJ507836.1

[FASTA](#) [Graphics](#)

Go to:

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 1 (bases 91550 to 92827)

AUTHORS Brandsch,R., Hinkkanen,A.E., Mauch,L., Nagursky,H. and Decker,K.

TITLE 6-Hydroxy-D-nicotine oxidase of Arthrobacter oxidans. Gene structure of the flavoenzyme and its relationship to 6-hydroxy-L-nicotine oxidase

JOURNAL Eur. J. Biochem. 167 (2), 315-320 (1987)

PUBMED [3622516](#)

REFERENCE 2 (bases 98647 to 99762)

AUTHORS Menendez,C., Igloi,G., Henninger,H. and Brandsch,R.

TITLE A pAO1-encoded molybdopterin cofactor gene (moaA) of Arthrobacter

Send to: shown

Complete Record
 Coding Sequences
 Gene Features

Choose Destination

File Clipboard
 Collections Analysis Tool

Download 1 items.

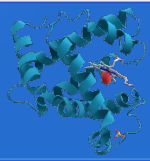
Format
FASTA

Show GI

Create File

Protein
PubMed
Taxonomy
BioCollections
Component Of
Full text in PMC

Principalele baze de date cu secvențe



UniProtKB/TrEMBL - <http://www.uniprot.org/>

- conține toate secvențele de nucleotide din GenBank translate în secvențe de aminoacizi;
- are un nivel foarte mare de redundanță;

UniProtKB/SwissProt - <http://www.uniprot.org/>

- toate secvențele conținute sunt adnotate manual – siguranță crescută legată de funcțiile secvențelor conținute;
- nivelul de redundanță este minim;
- adăugarea unei noi secvențe în SwissProt este un proces care implică prelucrarea manuală a datelor corespunzătoare secvenței respective => SwissProt nu conține cele mai noi secvențe.

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

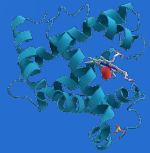
Protein Identification Resource PIR

- o colecție completă de secvențe lipsită de redundanță;
- secvențele identice sau foarte similare sunt grupate 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) și 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 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
ctactaacacctcagcgccgccccttcaacccttttctctgaggaaatatgactaaaaca
gcgattgtccgctcgccatgaatggcatcaccggccggatgggctaccgccagcacctg
ccggagcggcgggcttaccctcgaagacggcaccagggtccagatcgaaccgatcct
cgtaggccgcaacgaagccaagatccgcgaactcggcgagaagcacaaggttccgagtg
gagcagggacctggactcggctcgtcaacgacccaccgtcgacatcatctcgacgctc
catgaccagcctccgcccgcaccctgaagaaggcagctgctggccggcaagcacatctt
caccgagaagcccaccgcgaaaccctggaagaggccatgaaactggcccgcacgcgcaa
gcaggcagcgctcaccgagggcgtgtacacgacaagctgtacctcccgggctggctcaa
gctccgcccctgggtggacgaaggcttctcggccgcacatcctgccatccgcggtgagt
cggctactgggtcttgaaggtgacgttcaggcagcacagcggcctctggaactaccg
caaggaagacggcgggtggaatgaccacggacatgtctgccactggaactacgtccttga
aggcatcatcggcaaggtcaagagcgtcaacgccaagaccgacgacacatcccaccg
ctgggacgaagccggcaaggagtacaaggcaacggctgatgacgcttctacggcatctt
cgagcttgaaacccgggcggcgacgacgtcatcggccagatcaactcttctgggccgt
ccgctctaccgcgacgaactcgtcgaatccagggtggacggcaccacggctccgcccgt
tgcggcctgaacaagtgcgtcgccagcagcgcgacacaccccccaagccggtctggaa
ccctgacctgccatcaccgaatccttcccgaccagtggcaggaagtcccgcacaacc
```

Please check the [browser compatibility page](#) before using this program.

- Translate in on the strand.
- Use the genetic code.

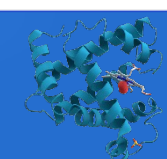
Căsuță text pentru secvență FASTA

*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

Operații simple cu secvențe – SMS2



Denumire operație*	Descriere
Convertire între diverse formate	
Combine FASTA	– permite combinarea a două sau a mai multe secvențe în format FASTA și obținerea unei singure secvențe
EMBL to FASTA	– permite convertirea unei secvențe din formatul EMBL în formatul FASTA; funcția este utilă în situația în care se dorește eliminarea rapidă a informațiilor care nu au legătură cu secvența de ADN dintr-un fișier EMBL
Filter DNA	– elimină caracterele care nu corespund codului standard IUPAC pentru ADN-ul dintr-o secvență (inclusiv spații, numere, etc)
Filter Protein	– elimină caracterele ce nu corespund cu codul standard IUPAC de o literă pentru aminoacizi dintr-o secvență (inclusiv spații, numere etc.)
GenBank to FASTA	– permite convertirea unei secvențe din formatul GenBank în formatul FASTA; funcția este utilă în situația în care se dorește eliminarea rapidă a informațiilor care nu au legătură cu secvența de ADN dintr-un fișier GenBank
One to Three	– permite convertirea unei secvențe de aminoacizi din codul standard IUPAC de o literă în cel de trei litere
Analiza secvențelor	
Codon Plot	– analizează frecvența de utilizare a codonilor dintr-o secvență de ADN și generează un grafic cu aceste frecvențe
Codon Usage	– analizează frecvența de utilizare a codonilor dintr-o secvență de ADN și poate fi utilizat pentru a identifica preferința pentru un anumit codon sinonim
DNA Molecular Weight	– calculează masa moleculară a uneia sau a mai multor secvențe de ADN
DNA Stats	– calculează frecvența fiecărei nucleotide într-o secvență dată, rezultatele fiind exprimate în procente
PCR Primer Stats	– analizează și calculează proprietățile specifice ale unui set indicat de primeri, inclusiv temperatura de topire și procentul de GC
PCR Products	– realizează amplificarea virtuală prin PCR a unei secvențe date folosind un set de primeri indicat de utilizator
Protein Isoelectric Point	– calculează valoarea teoretică a punctului izoelectric pentru o secvență dată
Protein Molecular Weight	– calculează valoarea teoretică a masei moleculare pentru una sau mai multe secvențe de aminoacizi
Protein Stats	– calculează frecvența fiecărui aminoacid într-o secvență dată, rezultatele fiind exprimate în procente
Translate	– realizează translația și transcripția virtuală a unei secvențe de ADN date, utilizatorul având posibilitatea de a alege cadrul de citire specific
Reverse Translate	– realizează procesul invers față de Translate, acceptând o secvență de aminoacizi și generând secvența ADN corespunzătoare

* corespunzătoare cu denumirea din suita de programe SMS2