Refresh

- Bioinformatics: introduction
- Features of bioinformatics
- Applications of bioinformatics
- Bioinformatics in different complexities of biological systems

Complementary strand

Although the two strands of a DNA molecule are complementary they are not in the same 5'/3' orientation.

Instead the two strands are said to be antiparallel.

5' ACGTTACG 3'

3' TGCAATGC 5'

5' CGTAACGT 3' (most cellular process involving DNA occur in the 5' to the 3' direction).

The two strands of double stranded DNA molecule are reverse complements of each other.

Example:

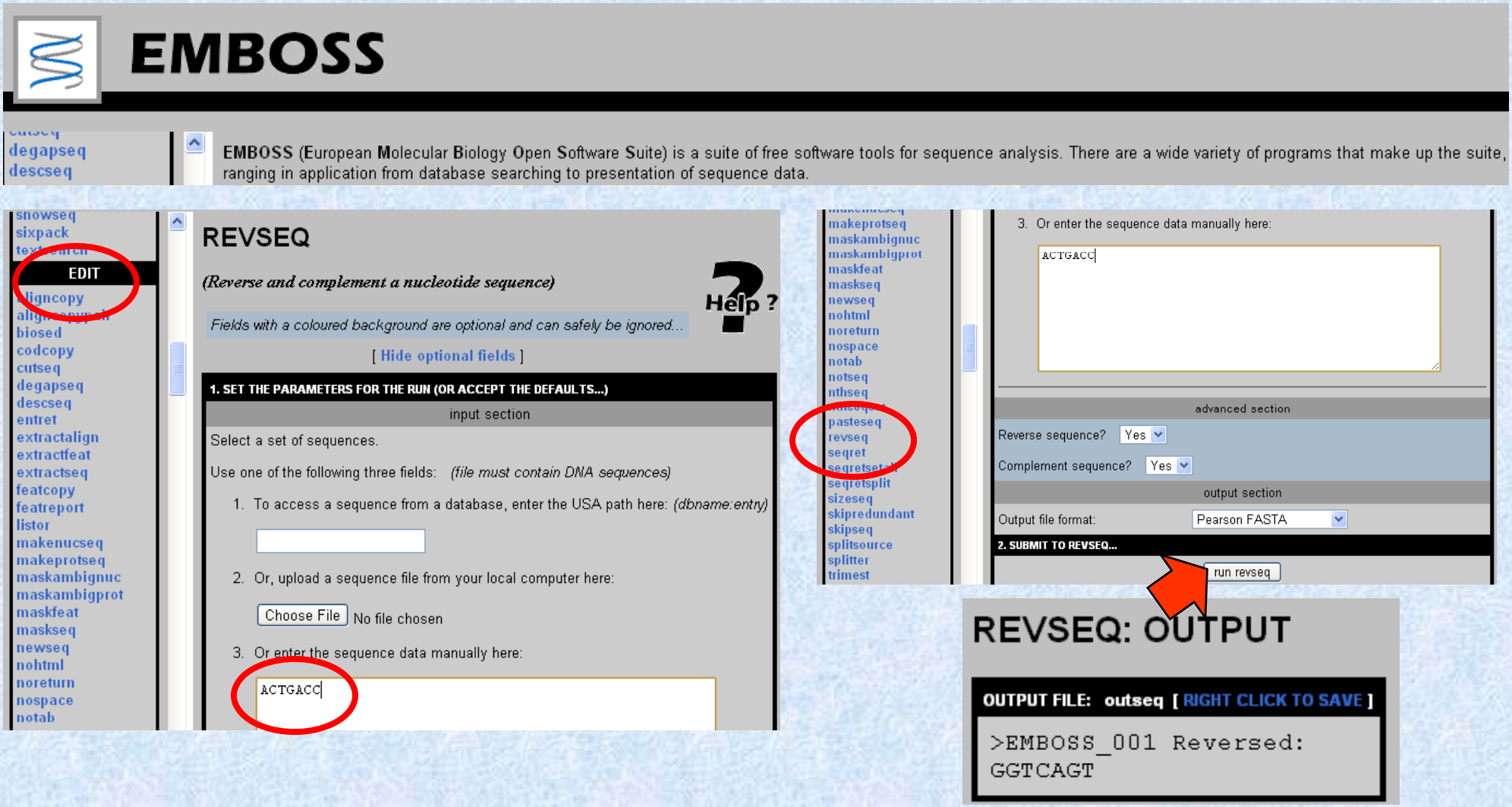
5' AGCCGTTAAGCTAATTCTGCTAGC 3'

Complementary strand is: ?

5'

Public domain program

EMBOSS

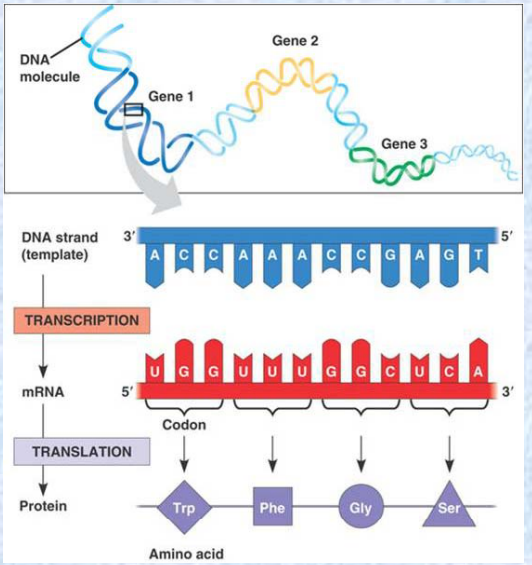


Protein synthesis: Central dogma in molecular biology

Information stored in DNA is used to make a more transient, single-stranded polynucleotide called RNA (ribonucleic acid) this is in turn used to make proteins. The process of making an RNA copy of a gene is called transcription and is accomplished through an enzymatic activity of an RNA polymerase.

There is one-to-one correspondence between the nucleotides to make RNA (G, A, U, uracil and C).

The process of converting that information from nucleotide sequences in RNA to amino acid sequences that make a protein is called translation: and it is performed by a complex of proteins and RNA called ribosomes.



DNA —(Transcription; RNA polymerase)→ RNA —(Translation; Ribosomes)→ Protein

Genetic code

Only 4 different nucleotides are used to make DNA/RNA molecules

20 different amino acids are used in protein synthesis .

There cannot be one-to-one correspondence between nucleotide and amino acid

Combination of 2 gives 4^2=16, which is less than 20.

Four nucleotides cannot be arranged in a total of 64 different combinations of three.

18 of the 20 amino acids are coded by more than one codon and this feature is called degeneracy.

It is possible to make the same amino acid sequence with changes in nucleotide.

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

DNA/RNA sequence to protein

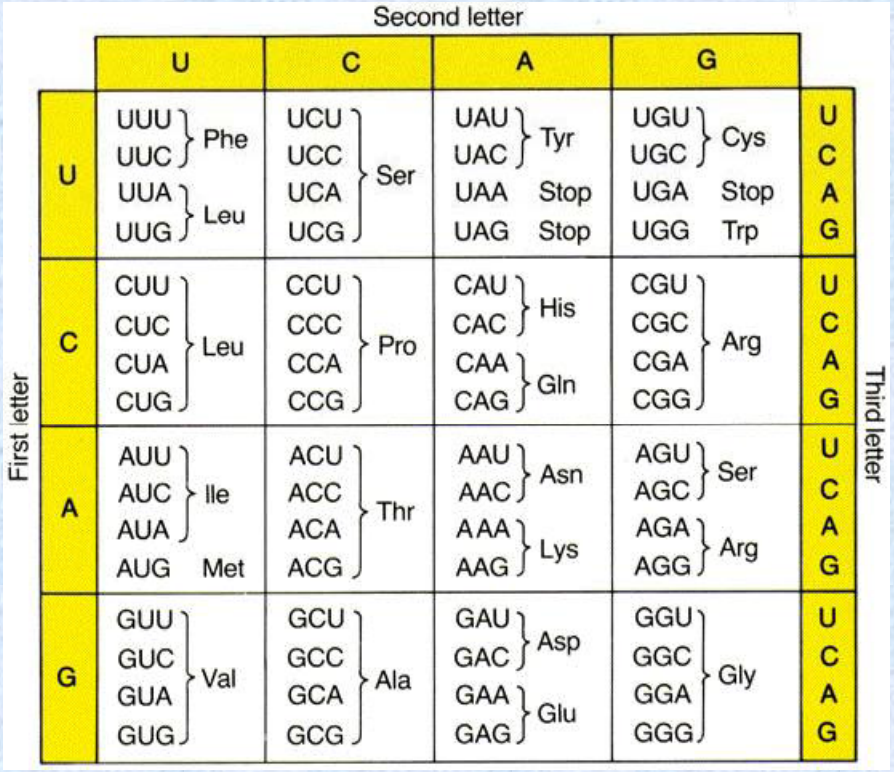
What sequence of amino acids would the following RNA sequence code for?

ACGUGCGCAUGCAACCGAAUGA

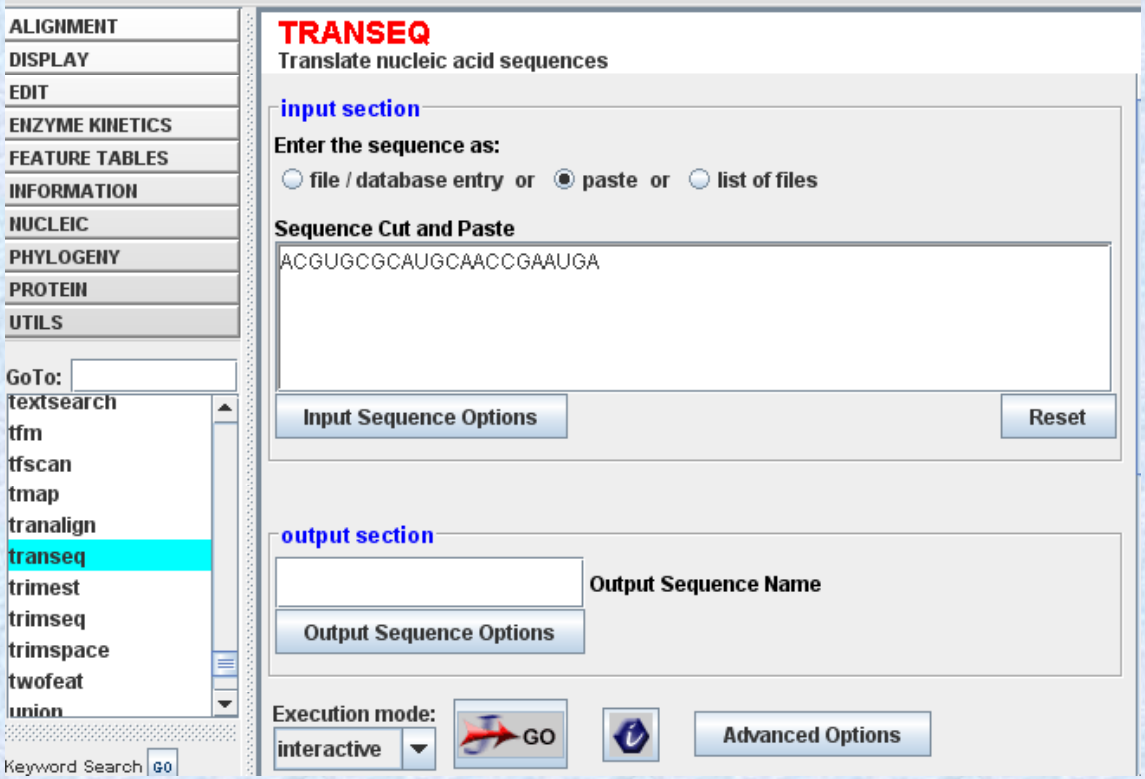
TCACNRMX

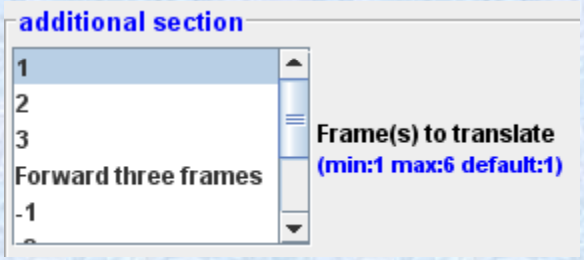
What will happen if the first nucleotide "A" is deleted?

RAHATE*

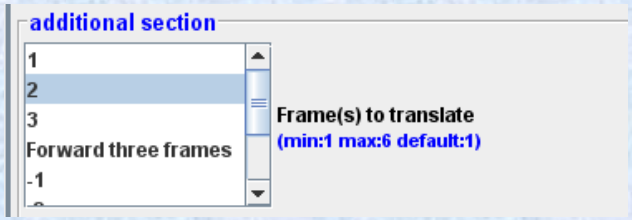


Resources





TCACNRMX



RAHATE*

Reading frames

e.g. 5' CAATGGCTAGGTACTATGTATGAGATCATGATCTTTACAAATCCGAG 3' DNA

Forward Frames

```
      CAA TGG CTA GGT ACT ATG TAT GAG ATC ATG ATC TTT ACA AAT CCG AG DNA
       Q   W   L   G   T   M   Y   E   I   M   I   F   T   N   P     Amino Acids

    C AAT GGC TAG GTA CTA TGT ATG AGA TCA TGA TCT TTA CAA ATC CGA G  DNA
       N   G   *   V   L   C   M   R   S   *   S   L   Q   I   R     Amino Acids

   CA ATG GCT AGG TAC TAT GTA TGA GAT CAT GAT CTT TAC AAA TCC GAG    DNA
       M   A   R   Y   Y   V   *   D   H   D   L   Y   K   S   E     Amino Acids
```

Reverse frames

Reverse Frames

Here we take the reverse/complimentry (bottom) strand and reverse it so it starts with the 5' end.

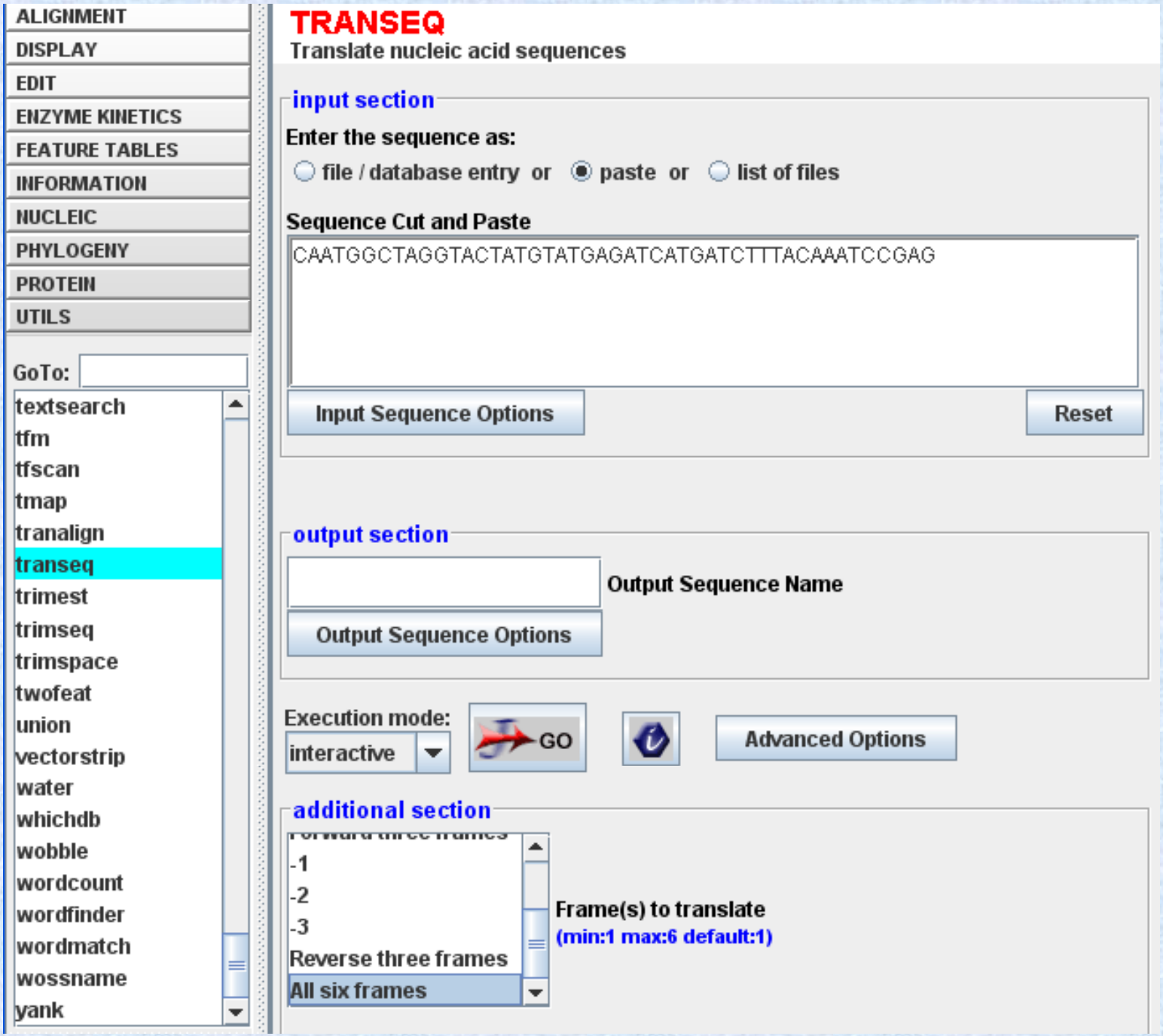
```
   5' CAATGGCTAGGTACTATGTATGAGATCATGATCTTTACAAATCCGAG  3'  DNA Top Strand
      ||||||||||||||||||||||||||||||||||||||||||||||
   3' GTTACCGATCCATGATACATACTCTAGTACTAGAAATGTTTAGGCTC  5'  DNA Bottom (Complimentry) Strand

   5' CTCGGATTTGTAAAGATCATGATCTCATACATAGTACCTAGCCATTG  3'  DNA Bottom (Complimentry) Strand Reversed


       CTC GGA TTT GTA AAG ATC ATG ATC TCA TAC ATA GTA CCT AGC CAT TG DNA
        L   G   F   V   K   I   M   I   S   Y   I   V   P   S   H   X Amino Acids

     C TCG GAT TTG TAA AGA TCA TGA TCT CAT ACA TAG TAC CTA GCC ATT G  DNA
        S   D   L   *   R   S   *   S   H   T   *   Y   L   A   I  X  Amino Acids

    CT CGG ATT TGT AAA GAT CAT GAT CTC ATA CAT AGT ACC TAG CCA TTG    DNA
        R   I   C   K   D   H   D   L   I   H   S   T   *   P   L     Amino Acids
```


ALIGNMENT
DISPLAY
EDIT
ENZYME KINETICS
FEATURE TABLES
INFORMATION
NUCLEIC
PHYLOGENY
PROTEIN
UTILS
GoTo:
textsearch
tfm
tfscan
tmap
tranalign
transeq
trimest
trimseq
trimspace
twofeat
union
vectorstrip
water
whichdb
wobble
wordcount
wordfinder
wordmatch
wossname
yank
TRANSEQ
Translate nucleic acid sequences
input section
Enter the sequence as:
file / database entry or paste or list of files
Sequence Cut and Paste
CAATGGCTAGGTACTATGTATGAGATCATGATCTTTACAAATCCGAG
Input Sequence Options
Reset
output section
Output Sequence Name
Output Sequence Options
Execution mode:
interactive
GO
Advanced Options
additional section
-1
-2
-3
Reverse three frames
All six frames
Frame(s) to translate
(min:1 max:6 default:1)

transeq044069.pep

```
>_1
QWLGTMYEIMIFTNPX
>_2
NG*VLCMRS*SLQIRX
>_3
MARYYV*DHDLYKSE
>_4
RICKDHDLIHST*PL
>_5
SDL*RS*SHT*YLAIX
>_6
LGFVKIMISYIVPSHX
```

>_1
QWLGTMYEIMIFTNPX
>_2
NG*VLCMRS*SLQIRX
>_3
MARYYV*DHDLYKSE
>_4
RICKDHDLIHST*PL
>_5
SDL*RS*SHT*YLAIX
>_6
LGFVKIMISYIVPSHX

Programming biological problems

1. Find the complementary strand of DNA

2. Convert DNA sequence into protein sequence

3. Number of nucleotides in a DNA sequence.

4. Pair preference

Computer programming

Computer hardware requires an operating system for its function.

General operating systems are Windows, MacOS and Unix

Linux is an open source version of Unix

Many libraries and tools in bioinformatics and computational biology are available in linux platforms.

The commonly used programming languages are FORTRAN, C, C++, JAVA, PERL, Python etc.

The unix environment is very convenient to write computer programs

(i) Set up unix operating system

(ii) Use unix operating system for programming.

When you log on Unix, it prompts with $ for any command

Unix commands

Files

- **ls** --- lists your files
  ls -l --- lists your files in 'long format', which contains lots of useful information, e.g. the exact size of the file, who owns the file and who has the right to look at it, and when it was last modified.
  ls -a --- lists all files, including the ones whose filenames begin in a dot, which you do not always want to see.
  There are many more options, for example to list files by size, by date, recursively etc.
- **vi *filename*** --- is an editor that lets you create and edit a file.
- **mv *filename1 filename2*** --- moves a file
- **cp *filename1 filename2*** --- copies a file
- **rm *filename*** --- removes a file. It is wise to use the option rm -i, which will ask you for confirmation before actually deleting anything.
- **diff *filename1 filename2*** --- compares files, and shows where they differ
- **wc *filename*** --- tells you how many lines, words, and characters there are in a file
- **chmod *options filename*** --- lets you change the read, write, and execute permissions on your files.

Finding things

grep *string filename(s)* --- looks for the string in the files.

Unix commands

- **File Compression**

  gzip *filename* --- compresses files, so that they take up much less space.

  gunzip *filename* --- uncompresses files compressed by gzip.

  tar –cvf home.tar home/

  tar –xvf home

  lpr *filename* --- print.

  lpq --- check out the printer queue

  Passwd --- change password

Directories

- **mkdir *dirname*** --- make a new directory
- **cd *dirname*** --- change directory.
- **pwd** --- tells you where you currently are.

Communicating with other computers

1. Web (Netscape, Firefox, Explorer, Chrome etc.)

The computers recognize each other by their internet protocol (IP) addresses. IP addresses consists of four numbers separated by dots (e.g., 10.93.219.140). These are interpreted as directions to the host by network software.

Computers also have hostnames biotech.iitm.ac.in

2. Telnet

Usage: telnet hostname

It opens a shell on a remote Unix machine and one has to give username and password for the unix machine. Then one can use the computer from remote.

3. ftp: file transfer protocol

It is a method for transferring files from one computer to another

E.g. ftp ftp.wwpdb.org (for downloading Protein Structure Data)

ftp commands

bin: ***to set the mode of file transfer to binary mode***

bye: ***to exit the FTP environment***

cd: ***to change directory on the remote machine***

get: ***to copy one file from the remote machine to the local machine***

mget*: to copy multiple files from the remote machine to the local machine;*

mput: ***to copy multiple files from the local machine to the remote machine;***

put: ***to copy one file from the local machine to the remote machine***

lcd: ***to change directory on your local machine (same as UNIX cd)***

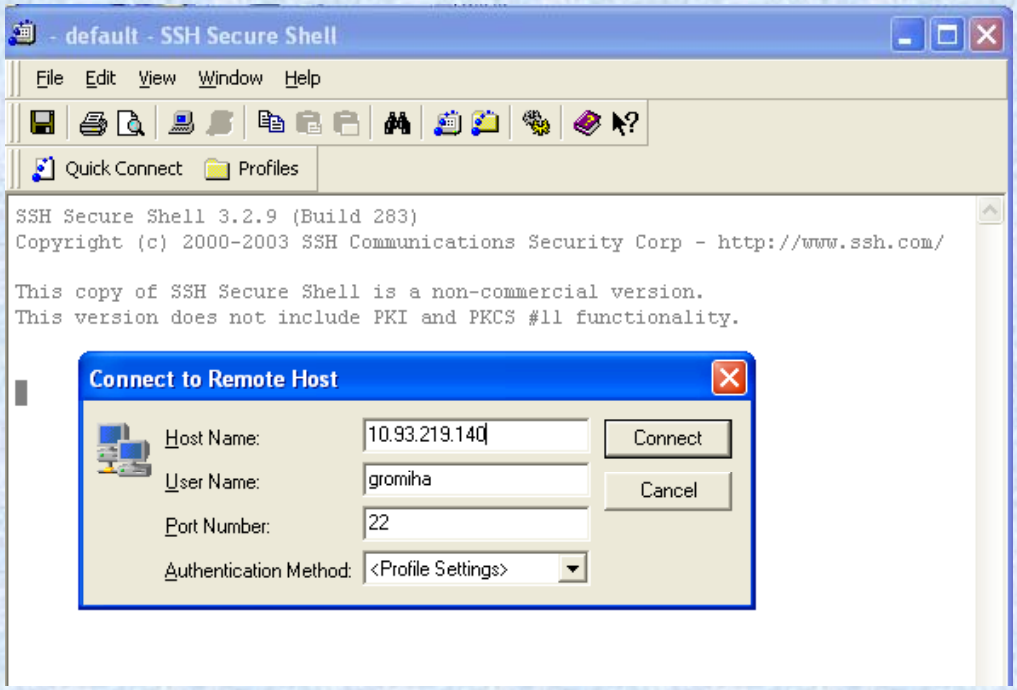
ls: ***to list the names of the files in the current remote directory***

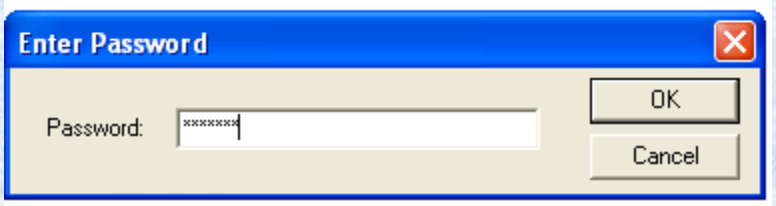
help: ***to request a list of all available FTP commands***

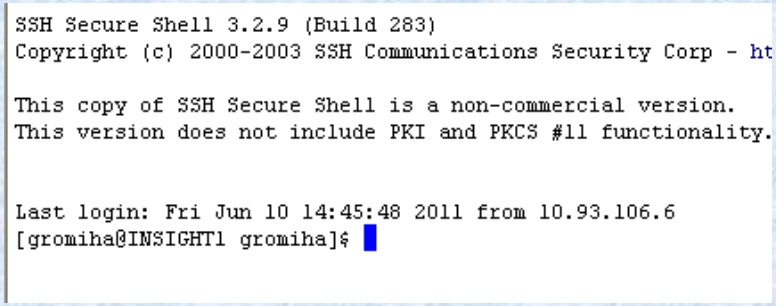
Connecting with remote computer

SSH is secure shell protocol for file transfer or connecting to remote computer.

It is available for Windows, Unix and MacOS.



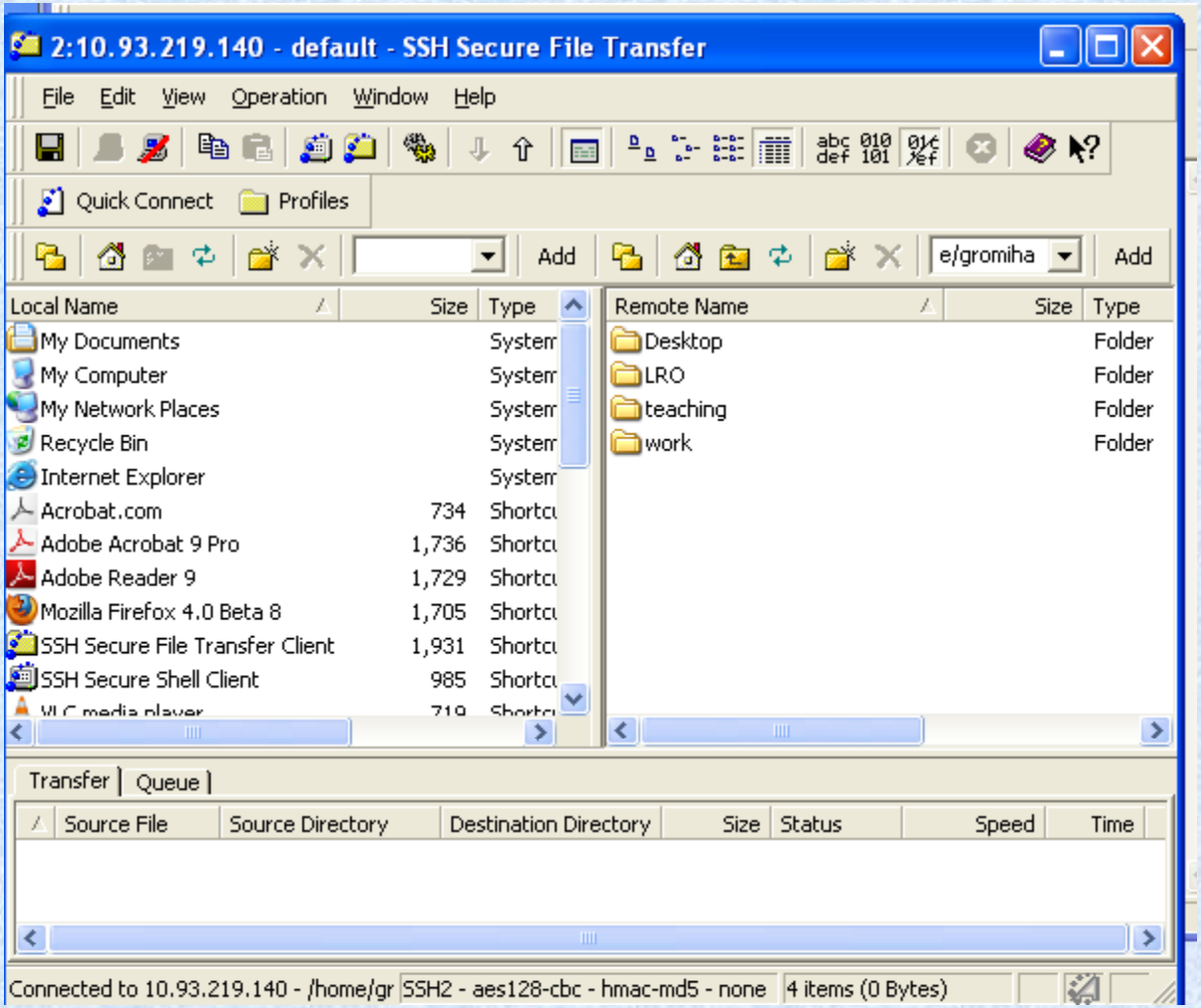




10.21.48.99

bioinfo
btclass

File transfer using SSH



Basic vi commands (Edit)

Start vi

vi filename: *edit filename starting at line 1*

vi -r filename: *recover filename that was being edited when system crashed*

Exit vi

:wq<Return>: *quit vi, writing out modified file to file named in original invocation*

:q<Return>: *quit (or exit) vi*

:q!<Return>: *quit vi even though latest changes have not been saved for this vi call*

Moving cursor

j [*or* down-arrow]: *move cursor down one line;* **k [*or* up-arrow]**: *move cursor up one line*

h [*or* left-arrow]: *move cursor left one character;* **l [*or* right-arrow]**: *move cursor right one character*

0 (zero): *move cursor to start of current line (the one with the cursor)*

$: *move cursor to end of current line*; **W:** *move cursor to beginning of next word*

B: *move cursor back to beginning of preceding word;* **:0<Return>:** *move cursor to first line in file*

:n<Return> *move cursor to line n;* **:$<Return>:** *move cursor to last line in file*

Basic vi commands (Edit)

Insert text

i: *insert text before cursor, until <Esc> hit*

a: *append text after cursor, until <Esc> hit*

Deleting text

x: *delete single character under cursor*

Dd: *delete entire current line*

Ndd *: delete N lines, beginning with the current line; e.g., 5dd deletes 5 lines*

Cutting and pasting

Nyy: *copy (yank, cut) the next N lines, including the current line, into the buffer*

p: *put (paste) the line(s) in the buffer into the text after the current line*

Searching text

/string: *search forward for occurrence of string in text*

?string: *search backward for occurrence of string in text*

n: *move to next occurrence of search string*

N: *move to next occurrence of search string in opposite direction*

Details about unix and vi commands

http://unlser1.unl.csi.cuny.edu/tutorials/viunix/unixman.html

http://unix.t-a-y-l-o-r.com/index.html

PDF and ps files