

1. The _____ is a collection of functions built on the MATLAB® technical computing environment.

Ans :- toolbox

2. What is the purpose of Matlab Bioinformatics Toolbox

Ans:- Bioinformatics Toolbox provides sequence analysis and visualization tools for genomic and proteomic sequence data. You can perform a variety of analyses, including multiple sequence alignments and the building and interactive viewing and manipulation of phylogenetic trees.

3. The latest version of Gene Ontology database, available from <http://www.geneontology.org> is downloaded over the Web into MATLAB using the function **geneont('LIVE', true)** when the LIVE parameter is true
4. Write the commands for downloading a file from the nucleotide databases GenBank and EMBL and the protein database GenPept

To use rsync

Replace the "ftp:" at the beginning of the FTP path with "rsync:"

rsync --copy-links --recursive --times --verbose (rsync:// followed by FTP path)

To use HTTPS

Replace the "ftp:" at the beginning of the FTP path with "https:"

Also append a '/' to the path if it is a directory.

wget --recursive -e robots=off --reject "index.html" --no-host-directories --cut-dirs=6

(https:// followed by FTP path)

To use FTP

Append a '/' to the path if it is a directory

wget --recursive --no-host-directories --cut-dirs=6 (ftp:// followed by FTP path)

5. Describe these commands - nt2aa, aa2nt, dna2rna & rna2dna with an example

Ans nt2aa :- Convert nucleotide sequence to amino acid sequence

nt2aa('agttgccgacgcgcncar','ACGTOnly', false)

ans = SCRRQA

aa2nt :- Convert amino acid sequence to nucleotide sequence

aa2nt('TYNYMRQLVVDVVITNHYSV')

ans

=

'ACATATAACTACATGAGACAGCTTGTAGTTGACGTTGTCATTACTAACCA
CTATAGCGTT'

dna2rna :- Convert DNA sequence to RNA sequence

dna2rna('TTATGACGTTATTCTACTTTGATTGTGCGA')

ans = 'UUAUGACGUUAUUCUACUUUGAUUGUGCGA'

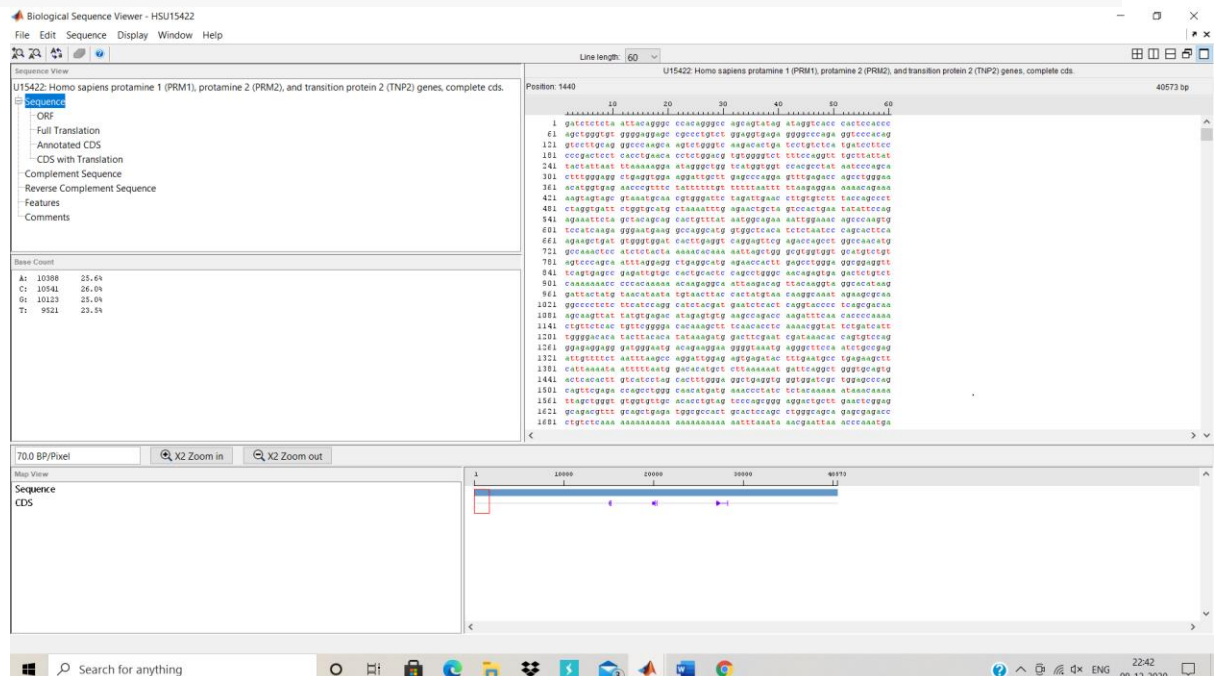
rna2dna:- Convert RNA sequence to DNA sequence

rna2dna('ACGAUGAGUCAUGCUU')

ans = ACGATGAGTCATGCTT

6. Use matlab bioinformatics toolbox function to download the GenBank sequence with accession number U15422. Extract out the raw sequence data from the above sequence object and perform a display of the nucleotides.

Ans Data = getgenbank('U15422')
seqviewer(Data)



7. Which of the following is not an alignment between sequences s = ATTACG and t = TTAG?

(a)	(b)	(c)	(d)
AT-TACG	--ATTACG--	ATACG	ATTACG
:		:	
TTAG---	TT-----AG	TTAG-	-TTA-G

Ans A and C

8. Perform a BLAST query for a nucleotide sequence (GCGCGTCTGTTCTGTGGAACAGGAGGCAGTTGTTTTCCGTCCGGCT) in the nr database using the blastn program and retrieve the results from NCBI

```
Ans id = blastncbi("GCGCGTCTGTTCTGTGGAACAGGAGGCAGTTGTTTTCCGTCCGGCT", 'blastn')
Id = 'X3NT6KJN01R'
```

9. Give a brief description of BLOSUM with Matlab

Return BLOSUM scoring matrix

Matrix = blosum(*Identity*) returns a BLOSUM (Blocks Substitution Matrix) scoring matrix with a specified percent identity. The default ordering of the output includes the extended characters B, Z, X, and *.

A R N D C Q E G H I L K M F P S T W Y V B Z X *

[*Matrix*, *MatrixInfo*] = blosum(*Identity*) returns *MatrixInfo*, a structure of information about *Matrix*, a BLOSUM matrix. *MatrixInfo* contains the following fields:

- Name
- Scale
- Entropy
- ExpectedScore
- HighestScore
- LowestScore
- Order

... = blosum(*Identity*, ...'*PropertyName*', *PropertyValue*, ...) calls blosum with optional properties that use property name/property value pairs. You can specify one or more properties in any order. Each *PropertyName* must be enclosed in single quotation marks and is case insensitive. These property name/property value pairs are as follows:

... = blosum(*Identity*, ...'*Extended*', *ExtendedValue*, ...) controls the listing of extended amino acid codes. Choices are true (default) or false. If *ExtendedValue* is false, returns the scoring matrix for the standard 20 amino acids. Ordering of the output when *ExtendedValue* is false is

A R N D C Q E G H I L K M F P S T W Y V

... = blosum(*Identity*, ...'*Order*', *OrderValue*, ...) returns a BLOSUM matrix ordered by *OrderValue*, a character vector or string containing legal amino acid characters that specifies the order amino acids are listed in the matrix. The length of the character vector or string must be 20 or 24.

10. Phylogenetics is _____

- A) The grouping of organisms by their physical characteristics
- B) The study of evolutionary relationships in organisms
- C) The study of gene expression in organisms
- D) The extraction of phylo from genetic sequences

Ans = B) The study of evolutionary relationships in organisms

11. What is the correct order of the steps the ClustalW algorithm uses for Multiple Sequence Alignment:

1. ClustalW constructs a distance matrix of $N(N-1)/2$ pairs of sequences by pairwise alignment of the sequences
2. ClustalW builds a guide tree from the distance matrix using the clustering method (neighbor-joining) by Saitou and Nei
3. ClustalW will convert the similarity scores to evolutionary distances based on the model by Kimura

Correct order of processing steps:

- A) 1,2,3
- B) 2,3,1
- C) 1,3,2
- D) 2,1,3

Ans A) 1,2,3

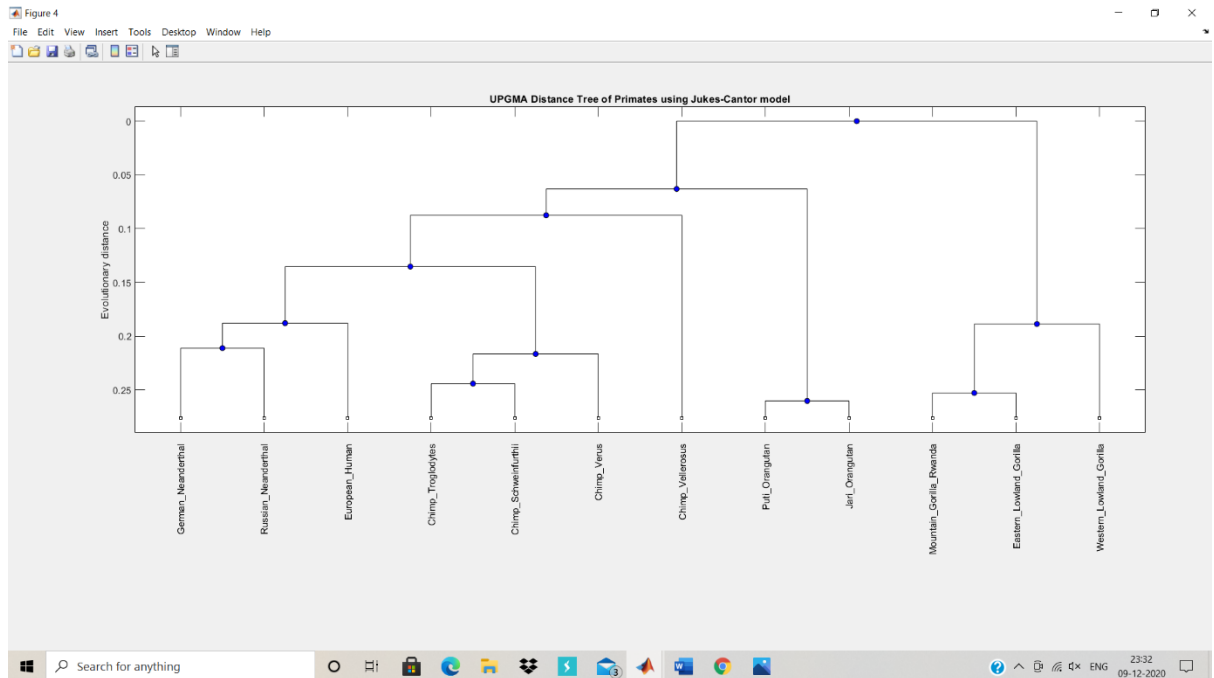
12. Consider the following piece of matlab code:

```
data = {'German_Neanderthal' 'AF011222'; 'Russian_Neanderthal' 'AF254446';  
'European_Human' 'X90314' ; 'Mountain_Gorilla_Rwanda' 'AF089820';  
'Chimp_Troglodytes' 'AF176766'; };  
for ind = 1:5 seqs(ind).Header = data{ind,1};  
seqs(ind).Sequence = getgenbank(data{ind,2},... 'sequenceonly', true);  
end distances = seqpdist(seqs,'Method','Jukes-Cantor','Alphabet','DNA');  
tree = seqlinkage(distances,'UPGMA',seqs)
```

Answer the following questions:

(a) Comment on what is stored in the data array.

(b) Run the program and examine the output.



13. What is the output of excuting the following matlab code:

```
>> cellseq = { 'TTGGTT' , 'TTTGGG' , 'TGGTTGGT' , 'GGGTTT' } ;
>> comp = regexpi ( cellseq , 'TTT' ) ;
>> ind = find (~ cellfun ( 'isempty' , comp ) ) ;
>> cellseq (ind)
cellseq
comp = regexpi ( cellseq , 'TTT' ) ;
comp
ind = find (cellfun( 'isempty' , comp ) ) ;
celsq = cellseq (ind )
```

Ans

cellseq =

1×4 cell array

{'TTGGTT'} {'TTTGGG'} {'TGGTTGGT'} {'GGGTTT'}

comp =

1×4 cell array

```

{0×0 double}  {[1]}  {0×0 double}  {[4]}
celsq =
1×2 cell array
{'TTGGTT'}  {'TGGTTGGT'}

```

14. What is the purpose of Microarray Box Plot?

Ans :- Box plots allow you to assess if the scale and distribution of the data on different arrays is comparable. Differences in shape or center of the boxes indicate that normalization of the data is required.

15. Show the value of the variable seqs upon executing the following matlab commands:

```

>> seq 1 = 'ATTA' ;
>> seq 2 = 'ATTTA' ;
>> seq 3 = 'ATTTAA' ;
>> seqs = char ( seq1 , seq2 , seq3 );

```

Ans seqseq =

3×6 char array

```

'ATTA '
'ATTTA '
'ATTTAA'

```