Bioinformatics (BT3040)-Practical 2

-Jagruthi Sarikonda (BE19B031)

1.

A. Number of "Homo sapiens" sequences deposited in DDBJ are 37,993,130. Where as in GenBank, there are 54,478,397 and in EMBL, there are 38,033,394.

2.

A. GC-content of the AY330867 is 46.53% {no: of G's = 129, no: of C's =72, Total no: of Nucleotides (A's + C's + G's + T's) = 432, GC content = ((129+72)/432) * 100 = 46.53}

3.

A. Contents of DDBJ, GenBank, EMBL (which are nucleotide sequence databases):

DDBJ	GenBank	EMBL
I. LOCUS II. DEFINITION III. ACCESSION IV. VERSION V. DBLINK VI. KEYWORDS VII. SOURCE a. ORGANISM VIII. REFERENCE a. AUTHORS b. TITLE c. JOURNAL IX. COMMENT X. FEATURES XI. BASE COUNT XII. ORIGIN	I. LOCUS II. DEFINITION III. ACCESSION IV. VERSION V. KEYWORDS VI. SOURCE a) ORGANISM VII. REFERENCE a) AUTHORS b) TITLE c) JOURNAL	The contents of EMBL will be nearly identical. Direct submissions from individual researchers, genome sequencing programmes, and patent applications are the primary sources of DNA and RNA sequences for EMBL. The most notable distinction is that GenBank does not display the base count, whereas DDBJ and EMBL show the base count

4

A. Papers about "discrimination of beta barrel membrane proteins"

- TMBETADISC-RBF: Discrimination of beta-barrel membrane proteins using RBF networks and PSSM profiles
- Current developments on beta-barrel membrane proteins: sequence and structure analysis, discrimination and prediction

5.

A. Author: Prof. Gromiha MM (https://pubmed.ncbi.nlm.nih.gov/?term=Gromiha+MM)
Number of papers published are 237.

6.

A. There are 28 similar articles in PUBMED for the paper, Cell 2008 Dec 26;135(7):1158-9 (Mitochondrial beta-barrel proteins, an exclusive club?) and it was cited by 6 articles.

7.
A. There are 401 papers published in the journal “Nature” for the year 2022 in PubMed and In Scopus, 353 papers are published in in the journal “Nature” for the year 2022

8.
A. h-index is 103 and number of citations are 47732 for “Prof. Burkhard Rost” which are given in Google scholar.

9.
A. The class of the enzyme EC 3.4.11.4 is Hydrolases.

EC Tree

- 3 Hydrolases(class).
  - 3.4 Acting on peptide bonds (peptidases)
    - 3.4.11 Aminopeptidases
      - 3.4.11.4 tripeptide aminopeptidase

and its function (reaction scheme) is release of the N-terminal residue from a tripeptide.

10.
A. The EC number of Asparagine synthetase is EC 6.3.5.4.
The Catalytic residue roles as given in catalytic site atlas are

UNIPROT	**PDB* (1CT9)**		
CYS2 (N-TERM)	Ala1A (N-term)	Acts as a general acid/base to activate the cysteine nucleophile.	Proton acceptor, proton donor
LEU51 (MAIN-C)	Leu50A (main-C)	Helps stabilise the reactive intermediates formed.	hydrogen bond acceptor, electrostatic stabiliser
THR322, ARG325	Thr321A, Arg324A	Bind and stabilise the phosphate groups of the ATP and reactive intermediates formed.	hydrogen bond donor, electrostatic stabiliser
CYS2	Ala1A	Acts as a catalytic nucleophile in the glutaminase domain reaction.	covalently attached, hydrogen bond acceptor, nucleofuge, nucleophile, proton acceptor, proton donor
GLY76 (MAIN-N), ASN75	Gly75A (main-N), Asn74A	Forms the oxyanion hole that stabilises the reactive intermediates and transition states formed.	hydrogen bond donor, electrostatic stabiliser

11.
A.

Organism	Scientific name	Taxonomy ID	Number of Chromosomes
Human	Homo sapiens	9606	23
Cat	*Felis catus*	9685	20
Dog	Canis lupus familiaris	9615	40
House mouse	Mus musculus	10090	40
Onion	Allium cepa. L	4679	9
Thale cress	Arabidopsis thaliana	3702	5

12.
A. The Entrez Programming Utilities (E-utilities) are a set of nine server-side programs that provide a stable interface into the Entrez query and database system at the National Center for Biotechnology Information (NCBI). The E-utilities use a fixed URL syntax that translates a standard set of input parameters into the values necessary for various NCBI software components to search for and retrieve the requested data. The E-utilities are therefore the structured interface to the Entrez system, which currently includes 38 databases covering a variety of biomedical data, including nucleotide and protein sequences, gene records, three-dimensional molecular structures, and the biomedical literature.

Syntax for fetching a record in FASTA format using E-utilities:

efetch.fcgi?db=<database>&id=<uid_list>&rettype=<retrieval_type>
&retmode=<retrieval_mode>

Input: List of UIDs (&id); Entrez database (&db); Retrieval type (&rettype); Retrieval mode (&retmode)

Output: Formatted data records as specified

Example: Download nuccore GIs 34577062 and 24475906 in FASTA format

https://eutils.ncbi.nlm.nih.gov/entrez/eutils/efetch.fcgi?db=nuccore&id=34577062,24475906&rettype=fasta&retmode=text

13.
A. List of two data bases under following categories:

- Protein properties:
    1. AAindex
    2. PFD-Protein Folding database
- Small molecules (Structure related):
    1. ChEBI-Chemical Entities of Biological Interest
    2. Hemolytik
- Cancer gene databases:
    1. Cancer3D
    2. CanGEM