BT3040 – BIOINFORMATICS – ASSIGNMENT 2

Submitted by Sahana (BE17B038)

Question 1 –

DDBJ – 31,894,999 results of which there are only 28,313,112 which exactly belong to Homo sapiens. Other sequences/results are for related organisms/life forms.

GenBank – 45,838,819 results of which there are only 27,598,713 which exactly belong to Homo sapiens. Other sequences/results are for related organisms/life forms.

EBI ENA - 27,695,793 results

Conclusion – All the available sequences throughout the three databases are similar.

Question 2 –

The given ID – AY410867 is for the Mus musculus TUBB gene, VIRTUAL TRANSCRIPT, partial sequence, genomic survey sequence.

GC content is calculated as = (number of occurrences of G + number of occurrences of C)/total number of nucleotides * 100

The GC content as calculated through EMBOSS = 0.58*100 = **58%.**

Question 3 –

Contents of DDBJ, GenBank, EMBL –

DDBJ –

1. Search and Analysis of sequences.
2. Submission of entries.
3. Download data.
4. Supercomputer
5. Statistics
6. Activities
7. Training
8. General info about them.

GenBank –

1. About GenBank
2. Submission Types.
3. Submission Tools.
4. Update GenBank records.
5. Search
6. BLAST
7. Statistics
8. Sample record

9. Sequence IDs
10. Additional contents –
    a. WGS
    b. TPA
    c. TSA
    d. INSDC

EMBL –

1. Submit entries
2. Search options –
    a. Text search
    b. Advanced search
    c. Sequence search
    d. Xref search
    e. Sequence versions
3. Rulespace
4. General info –
    a. EMBL ENA
    b. Research
    c. Industry
    d. Training

Question 4 –

As found in PubMed, there are 23 papers on the topic "discrimination of beta barrel membrane proteins".

They are –

1. The signal distinguishing between targeting of outer **membrane** β-**barrel** protein to plastids and mitochondria in plants. *Klinger A, Gosch V, Bodensohn U, Ladig R, Schleiff E.*
2. Nanoscale Probing of Informational Polymers with Nanopores. Applications to Amyloidogenic Fragments, Peptides, and DNA-PNA Hybrids. *Luchian T, Park Y, Asandei A, Schiopu I, Mereuta L, Apetrei A.*
3. Incorporating post translational modification information for enhancing the predictive performance of **membrane** transport **proteins**. *Le NQK, Sandag GA, Ou YY.*
4. Programmed Secretion Arrest and Receptor-Triggered Toxin Export during Antibacterial Contact-Dependent Growth Inhibition. *Ruhe ZC, Subramanian P, Song K, Nguyen JY, Stevens TA, Low DA, Jensen GJ, Hayes CS.*
5. Hitting with a BAM: Selective Killing by Lectin-Like Bacteriocins. *Ghequire MGK, Swings T, Michiels J, Buchanan SK, De Mot R.*
6. Going Outside the TonB Box: Identification of Novel FepA-TonB Interactions *In Vivo. Gresock MG, Postle K.*
7. Predicting **Beta Barrel** Transmembrane **Proteins** Using HMMs. *Tsaousis GN, Hamodrakas SJ, Bagos PG.*
8. Nucleobase Recognition by Truncated α-Hemolysin Pores. *Ayub M, Stoddart D, Bayley H.*

9. **Membrane** protein orientation and refinement using a knowledge-based statistical potential. *Nugent T, Jones DT.*
10. Physical determinants of β-**barrel membrane** protein folding in lipid vesicles. *Dewald AH, Hodges JC, Columbus L.*
11. Outer **membrane proteins** can be simply identified using secondary structure element alignment. *Yan RX, Chen Z, Zhang Z.*
12. Single-nucleotide **discrimination** in immobilized DNA oligonucleotides with a biological nanopore. *Stoddart D, Heron AJ, Mikhailova E, Maglia G, Bayley H.*
13. Molecular requirements for MHC class II alpha-chain engagement and allelic **discrimination** by the bacterial superantigen streptococcal pyrogenic exotoxin C. *Kasper KJ, Xi W, Rahman AK, Nooh MM, Kotb M, Sundberg EJ, Madrenas J, McCormick JK.*
14. TMBETADISC-RBF: **Discrimination** of **beta**-**barrel membrane proteins** using RBF networks and PSSM profiles. *Ou YY, Gromiha MM, Chen SA, Suwa M.*
15. Functional **discrimination** of **membrane proteins** using machine learning techniques. *Gromiha MM, Yabuki Y.*
16. Current developments on **beta**-**barrel membrane proteins**: sequence and structure analysis, **discrimination** and prediction. *Gromiha MM, Suwa M.*
17. A predictor of **membrane** class: Discriminating alpha-helical and **beta**-**barrel membrane proteins** from non-membranous **proteins**. *Taylor PD, Toseland CP, Attwood TK, Flower DR.*
18. A consensus algorithm to screen genomes for novel families of transmembrane **beta barrel proteins**. *Garrow AG, Westhead DR.*
19. PROFtmb: a web server for predicting bacterial transmembrane **beta barrel proteins**. *Bigelow H, Rost B.*
20. Prediction of **beta**-**barrel membrane proteins** by searching for restricted domains. *Mirus O, Schleiff E.*
21. TMB-Hunt: a web server to screen sequence sets for transmembrane **beta**-**barrel proteins**. *Garrow AG, Agnew A, Westhead DR.*
22. Predicting transmembrane **beta**-barrels in proteomes. *Bigelow HR, Petrey DS, Liu J, Przybylski D, Rost B.*
23. Odorant-binding **proteins**: structural aspects. *Pelosi P.*

Question 5 –

Author : Ramakrishnan V[Author]
(https://www.ncbi.nlm.nih.gov/pubmed/?term=Ramakrishnan%20V%5BAuthor%5D&cauthor=true&cauthor_uid=21914843)

Number of papers: 888

Question 6 –

For the paper - Cell. 2008 Dec 26;135(7):1158-9 - Mitochondrial beta-barrel proteins, an exclusive club?

There are 35 similar articles reported in PubMed and 100 related articled reported in Google Scholar. The discrepancy in the number of similar/related articles is because of the different search algorithms.

Question 7 –

There are 310 papers published in "Nature" of which 239 papers are reported in SCOPUS and 249 papers are reported in PUBMED.

Question 8 –

As reported in Google Scholar, Prof. Burkhard Rost has an h-index of 95 and has 41099 citations overall.

Question 9 –

EC Tree

└ 3 Hydrolases (Class)

   └ 3.4 Acting on peptide bonds (peptidases)

      └ 3.4.11 Aminopeptidases

         └ 3.4.11.13 Clostridial aminopeptidase

Reaction scheme/Function – Release of any N-terminal amino acid, including proline and hydroxy proline, but no cleavage of Xaa-Pro-.

Question 10 –

Asparagine synthetase – EC 6.3.5.4

Catalytic Residues in this enzyme are as follows -

UniProt	PDB		
Cys2 (N-term)	Ala1A (N-term)	Acts as a general acid/base to activate the cysteine nucleophile.	proton acceptor, proton donor
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
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

It has the following Catalytic CATH domains.

CATH Domains (8)

Domain ID	Date inserted into CATH	Superfamily	CATH Status
1ct9A01	05 Mar, 2006	3.60.20.10	Assigned ⊘
1ct9A02	05 Mar, 2006	3.40.50.620	Assigned ⊘
1ct9B01	05 Mar, 2006	3.60.20.10	Assigned ⊘
1ct9B02	05 Mar, 2006	3.40.50.620	Assigned ⊘
1ct9C01	05 Mar, 2006	3.60.20.10	Assigned ⊘
1ct9C02	05 Mar, 2006	3.40.50.620	Assigned ⊘
1ct9D01	05 Mar, 2006	3.60.20.10	Assigned ⊘
1ct9D02	05 Mar, 2006	3.40.50.620	Assigned ⊘

1. Superfamily 3.60.20.10 – Aminohydrolase, N-terminal nucleophile (Ntn) domain.
2. Superfamily 3.40.50.620 – HUPs

Question 11 –

1. Human –
    a. Scientific name – Homo sapiens
    b. Taxonomy ID – 9606
    c. Number of chromosomes – 23
2. Cat –
    a. Scientific name – Felis Catus
    b. Taxonomy ID – 9685
    c. Number of chromosomes – 20
3. Dog –
    a. Scientific name – Canis lupus familiaris
    b. Taxonomy ID – 9615
    c. Number of chromosomes – 40
4. House mouse –
    a. Scientific name – Mus Musculus
    b. Taxonomy ID – 10090
    c. Number of chromosomes – 40
5. Onion
    a. Scientific name – Allium cepa
    b. Taxonomy ID – 4679
    c. Number of chromosomes – 9
6. Thale cress
    a. Scientific name – Arabidopsis Thalliana
    b. Taxonomy ID – 3702
    c. Number of chromosomes – 5

Question 12 –

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.

Syntax for fetching a record in FASTA format –

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

Output: Formatted data records as specified

Code:
```
efetch.fcgi?db=<database>&id=<uid_list>&rettype=<retrieval_type>&ret
mode=<retrieval_mode>
```

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

Question 13 –

1. Protein properties –
   a. AAindex
   b. BindingDB
2. Small molecules (structure related) –
   a. ChEBI - Chemical Entities of Biological Interest
   b. PubChem
3. Cancer gene databases –
   a. Cancer3D
   b. Tumor Associated Gene database