

## Practicals – 2

-BS19B032

-R. Vasantha Kumar

1) Number of Homo sapiens sequences in different databases are:

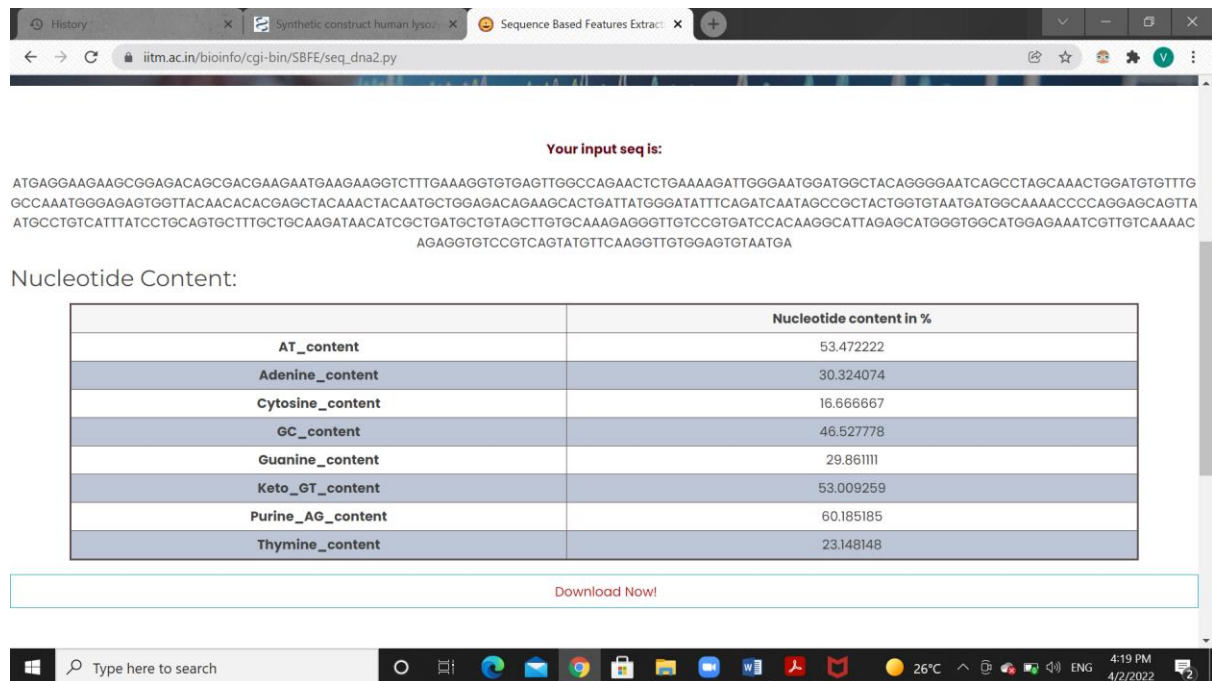
DDBJ: 18296211

GenBank: 189,232,925

EMBL: 129,325,718

2) AY330867 is the locus for synthetic construct human lysozyme mRNA, which I obtained from ncbi database. Then, by using the seq2feature tool, I found the GC content of the sequence.

The GC content of AY330867 is **46.527778**.



History: Synthetic construct human lysoz... Sequence Based Features Extrac...  
iitm.ac.in/bioinfo/cgi-bin/SBFE/seq\_dna2.py

Your input seq is:

ATGAGGAAGAAGCGGAGACAGCGACGAAGAATGAAGAAGGCTTTGAAAAGGTGTGAGTTGGCCAGAACTCTGAAAAGATTGGGAATGGATGGCTACAGGGGAATCAGCCTAGCAAACCTGGATGTGTTTG  
GCCAAATGGGAGAGTGGTTACAACACAGCTACAACTACAATGCTGGAGACAGAAGCACTGATTATGGGATATTCAGATCAATAGCCGCTACTGGTGAATGATGGCAAAACCCAGGAGCAGTTA  
ATGCCTGTCAATTTCCGAGTGCTTTGCTGCAAGATAACATCGCTGATGCTGTAGCTTGCAAGAGGGTTGTCCTGATCCACAAGGCATTAGAGCATGGGTGGCATGGAGAAATCGTTGTCAAAAC  
AGAGGTGTCCTCAGTATGTTCAAGGTTGTGGAGTGAATGA

Nucleotide Content:

|                   | Nucleotide content in % |
|-------------------|-------------------------|
| AT_content        | 53.472222               |
| Adenine_content   | 30.324074               |
| Cytosine_content  | 16.666667               |
| GC_content        | 46.527778               |
| Guanine_content   | 29.861111               |
| Keto_GT_content   | 53.009259               |
| Purine_AG_content | 60.185185               |
| Thymine_content   | 23.148148               |

Download Now!

Type here to search 26°C 4:19 PM 4/2/2022

3) DDBJ, EMBL and GenBank all are DNA databases collecting nucleic acid sequence data.

All the three databases have similar DNA data, but the differences come at the interface and the searching methods.

For example, the GenBank database allows us to search for specific sequences in two ways: one is the normal text search and the other is using molecular sequences comparing them using a blast. EMBL also has similar kind of interface.

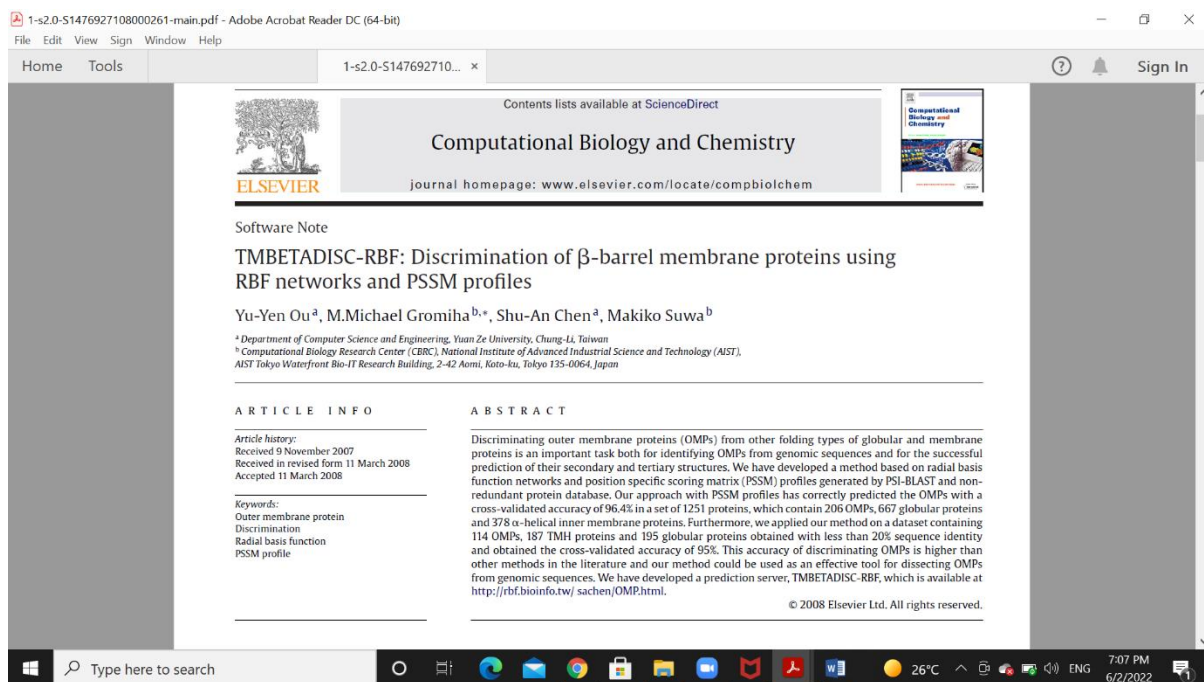
In DDBJ, anyone from any country can add DNA data, but the GenBank database only allows the author to alter or add DNA data.

4) From pubmed, I got the papers about discrimination of beta barrel membrane proteins.

The screenshot shows a web browser window with the URL [pubmed.ncbi.nlm.nih.gov/?term=discrimination+of+beta+barrel+membrane+proteins&schema=alltitle](https://pubmed.ncbi.nlm.nih.gov/?term=discrimination+of+beta+barrel+membrane+proteins&schema=alltitle). The page is from the National Center for Biotechnology Information (NCBI) and features the PubMed logo. The search term 'discrimination of beta barrel membrane proteins' is entered in the search bar. Below the search bar, there are buttons for 'Save', 'Email', and 'Send to', and a 'Search' button. The results are sorted by 'Best match' and displayed on page 1 of 1. On the left side, there is a 'MY NCBI FILTERS' section with a 'RESULTS BY YEAR' chart showing a single bar for the year 2007. Below the chart, there is a 'TEXT AVAILABILITY' section with checkboxes for 'Abstract' and 'Free full text'. The main results area shows two results:

- ☐ **TMBETADISC-RBF: Discrimination of beta-barrel membrane proteins using RBF networks and PSSM profiles.**  
Cite: Ou YY, Gromiha MM, Chen SA, Suwa M.  
Comput Biol Chem. 2008 Jun;32(3):227-31. doi: 10.1016/j.compbiolchem.2008.03.002. Epub 2008 Mar 18.  
Share: PMID: 18434251
- ☐ **Current developments on beta-barrel membrane proteins: sequence and structure analysis, discrimination and prediction.**  
Cite: Gromiha MM, Suwa M.  
Curr Protein Pept Sci. 2007 Dec;8(6):580-99. doi: 10.2174/138920307783018712.  
Share: PMID: 18220845 Review.

The bottom of the screenshot shows a Windows taskbar with various application icons, a search bar, and system information including the date and time (7:04 PM, 6/2/2022).



5) *Using the google scholar database, I found the articles published by Nanjiang Shu.*

- Topology of membrane proteins—predictions, limitations and variations
- The TOPCONS web server for consensus prediction of membrane protein topology and signal peptides
- ProQ3D: improved model quality assessments using deep learning
- Inclusion of dyad-repeat pattern improves topology prediction of transmembrane β-barrel proteins
- Improved topology prediction using the terminal hydrophobic helices rule
- SubCons: a new ensemble method for improved human subcellular localization predictions
- Improved protein model quality assessments by changing the target function
- Large tilts in transmembrane helices can be induced during tertiary structure formation
- KalignP: Improved multiple sequence alignments using position specific gap penalties in Kalign2
- The evolutionary history of topological variations in the CPA/AT superfamily

Public access

scholar.google.com/citations?view\_op=list\_mandates&hl=en&user=1a9GihIAAAJ&tzom=-330

Public access

Nanjiang Shu

Articles with public access mandates [Learn more](#)

**Not available: 1**

Topology of membrane proteins—predictions, limitations and variations  
 KD Tsingos, S Govindarajan, C Bassot, A Västermark, J Lamb, N Shu, ...  
 Current opinion in structural biology 50, 9-17, 2018  
 Mandates: Knut and Alice Wallenberg Foundation, Swedish Research Council [REVIEW](#)

**Available: 12**

The TOPCONS web server for consensus prediction of membrane protein topology and signal peptides [\[HTML\] oup.com](#)  
 KD Tsingos, C Peters, N Shu, L Käll, A Elofsson  
 Nucleic acids research 43 (W1), W401-W407, 2015  
 Mandates: Swedish Research Council [REVIEW](#)

ProQ3D: improved model quality assessments using deep learning [\[PDF\] oup.com](#)  
 K Uziela, D Menéndez Hurtado, N Shu, B Wallner, A Elofsson  
 Bioinformatics 33 (10), 1578-1580, 2017  
 Mandates: Swedish Research Council [REVIEW](#)

Type here to search

26°C 1:31 PM 5/2/2022

6) The number of related articles that are listed in PUBMED for the paper - Cell 2008 Dec 26;135(7):1158-9 is **28**.

Similar articles for PMID: 19109882

pubmed.ncbi.nlm.nih.gov/?linkname=pubmed\_pubmed&from\_uid=19109882

NIH National Library of Medicine  
 National Center for Biotechnology Information

Log in

PubMed.gov

Advanced Search

User Guide

Save Email Send to

Sorted by: Best match Display options

MY NCBI FILTERS

RESULTS BY YEAR

2007 2022

TEXT AVAILABILITY

☐ Abstract

Similar articles for PMID: 19109882

28 results

Page 1 of 3

☐ Mitochondrial beta-barrel proteins, an exclusive club?  
 1 Imai K, Gromiha MM, Horton P.  
 Cite Cell. 2008 Dec 26;135(7):1158-9; author reply 1159-60. doi: 10.1016/j.cell.2008.12.017.  
 Share PMID: 19109882 **Free article.** No abstract available.

☐ Evolution. Tinkering inside the organelle.  
 2 Alcock F, Clements A, Webb C, Lithgow T.  
 Cite Science. 2010 Feb 5;327(5966):649-50. doi: 10.1126/science.1182129.

Type here to search

26°C 4:36 PM 4/2/2022

7) The number of research articles published by nature journal, according to their website in 2022 is 86.

The screenshot shows the Nature Research articles website. The header includes the Nature logo, navigation links (View all journals, Search, Login), and a top bar with links (Explore content, About the journal, Publish with us, Subscribe, Sign up for alerts, RSS feed). The main content area is titled "Research articles" and features a filter section with "Article Type" set to "All" and "Year" set to "2022 (86)". Below the filters, three articles are listed:

- Article**  
03 Feb 2022  
**ACE2 binding is an ancestral and evolvable trait of sarbecoviruses**  
Tyler N. Starr, Samantha K. Zepeda ... Jesse D. Bloom
- Article**  
03 Feb 2022  
**Structure of the NLRP3 decamer bound to the cytokine release inhibitor CRID3**  
Inga V. Hochheiser, Michael Pils ... Matthias Geyer
- Matters Arising**  
**No freshwater-filled glacial Arctic Ocean**

*According to the SCOPUS website, the number of documents published by nature journal in 2022 is 243.*

The screenshot shows the SCOPUS website with search results for Nature journal in 2022. The left sidebar shows "Refine results" with "Open Access" filters and "Year" filters. The main content area displays a table of search results:

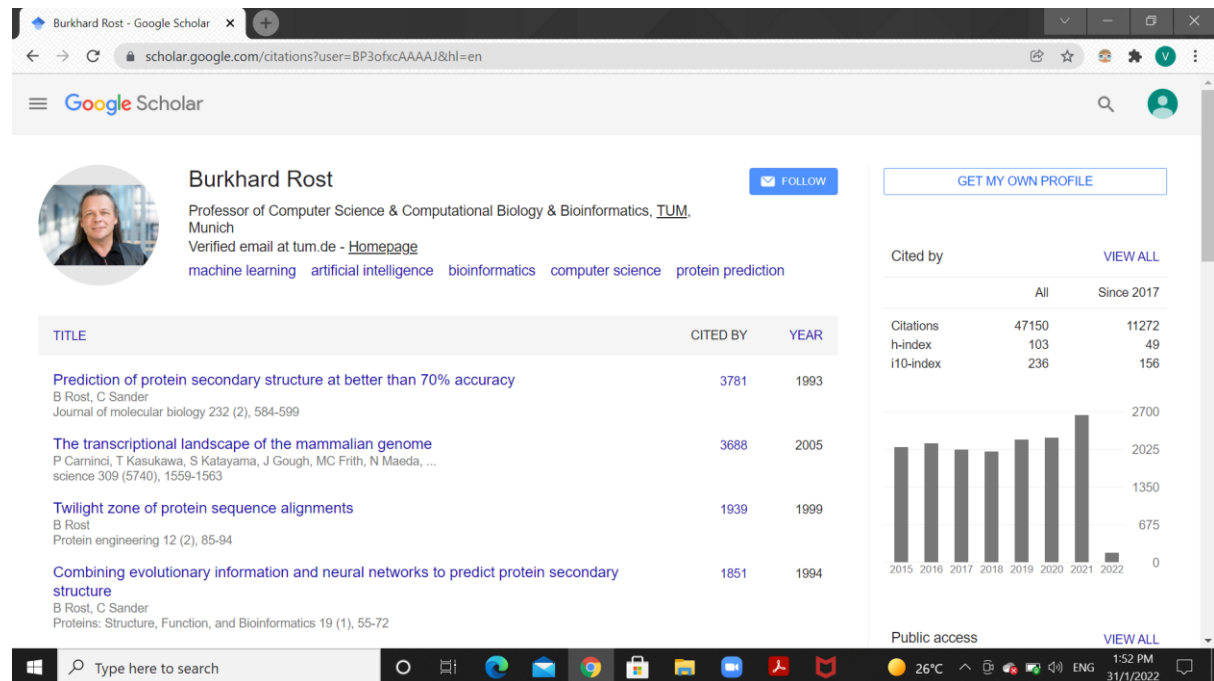
|                            | Document title  | Authors   | Year | Source                        | Cited by |
|----------------------------|---|---|------|-------------------------------|----------|
| <input type="checkbox"/> 1 | Emergent interface vibrational structure of oxide superlattices<br><i>Open Access</i> | Hoglund, E.R., Bao, D.-L., O'Hara, A., (...), Hopkins, P.E., Howe, J.M.           | 2022 | Nature 601(7894), pp. 556-561 | 0        |
| <input type="checkbox"/> 2 | Do not downplay biodiversity loss   | Loreau, M., Cardinale, B.J., Isbell, F., (...), O'Connor, M.I., de Mazancourt, C. | 2022 | Nature 601(7894), pp. E27-E28 | 1        |
| <input type="checkbox"/> 3 | A radio transient with unusually slow periodic emission                               | Hurley-Walker, N., Zhang, X., Bahramian, A., (...), Heald, G.H., Galvin, T.J.     | 2022 | Nature 601(7894), pp. 526-530 | 0        |

*Using the ISSN number of nature journal(1476-4687), I got the articles published by nature journal. Total articles of nature journal for year 2022 is 41.*

8) In the google scholar website, I found the h-index and number of citations for “Burkhard Rost”.

*h-index: 103(since 2017-49)*

*number of citations: 47150(since 2017-11272)*



9) EC 3.4.11.4 is actually tripeptide aminopeptidase enzyme, in the brenda-enzymes database.

*This enzyme belongs to the class of enzymes called Hydrolases.*

*This enzyme acts on peptide bonds. Its main function is to release the N-terminal residue from the tripeptide.*



BRENDA - Information on EC 3.4.11.4

W Tripeptide aminopeptidase - Wikipedia

brenda-enzymes.org/enzyme.php?ecno=3.4.11.4

Login

BRENDA The Comprehensive Enzyme Information System

Classic view All enzymes Enzyme history BRENDA support

BRENDA home History

show all | hide all No of entries

- Enzyme Nomenclature 47
- Enzyme-Ligand Interactions 238
- Diseases 57
- Functional Parameters 129
- Organism related information 65
- General Information 0
- Enzyme Structure 13k
- Molecular Properties 41
- Applications 0
- References 43
- External Links

### Information on EC 3.4.11.4 - tripeptide aminopeptidase

for references in articles please use BRENDA:EC3.4.11.4

EC Tree

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

Specify your search results

Mark a special word or phrase in this record:  Mark!

Search Reference ID:  Search

Search UniProt Accession:  Search

Select one or more organisms in this record: ☐

All organisms

Arabidopsis thaliana

Bacillus subtilis

Bos taurus

Caenorhabditis elegans

Submit

Word Map hide

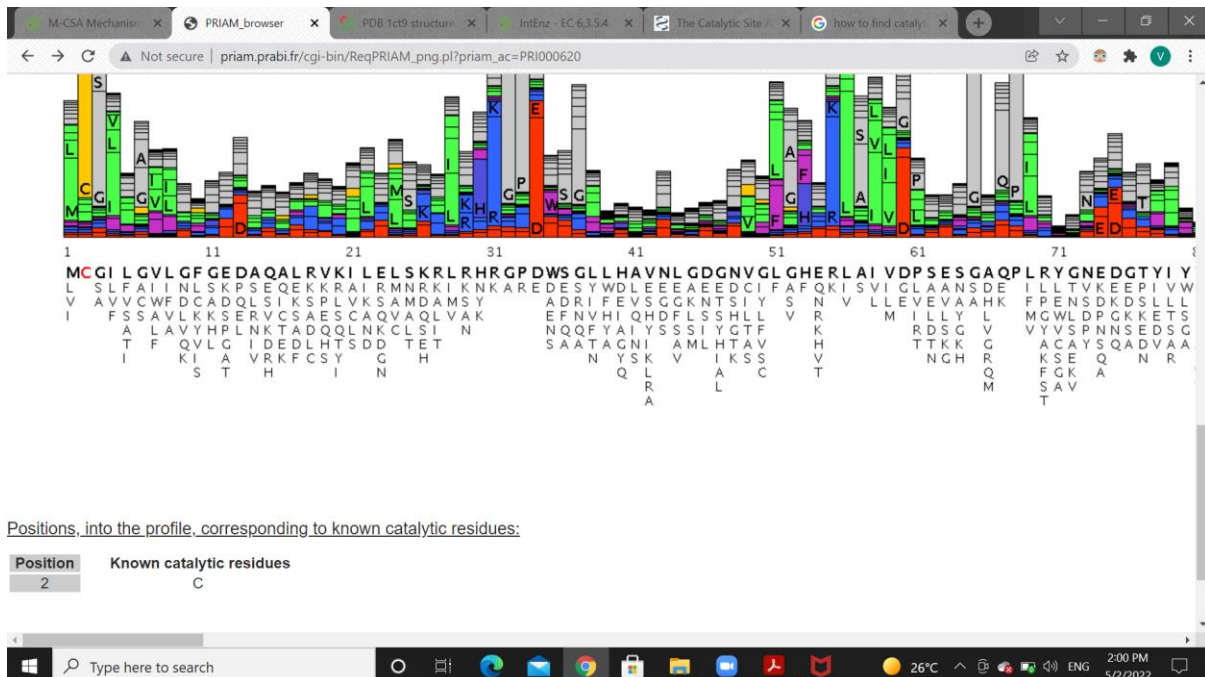
proteasome dipeptidase 3.4.11.4 tripeptidyl-peptidase tripeptides

lipolucipos oligopeptidase channa aminopeptidases caseinolytic glycolytic late infantile ceroid extralysosomal

4:49 PM 4/2/2022

10) From the link given, I found the EC number of Asparagine synthetase, that is, **6.3.5.4**.

Then in the catalytic site atlas I found the catalytic residues:  
It is found in site 2, as catalytic residue C.



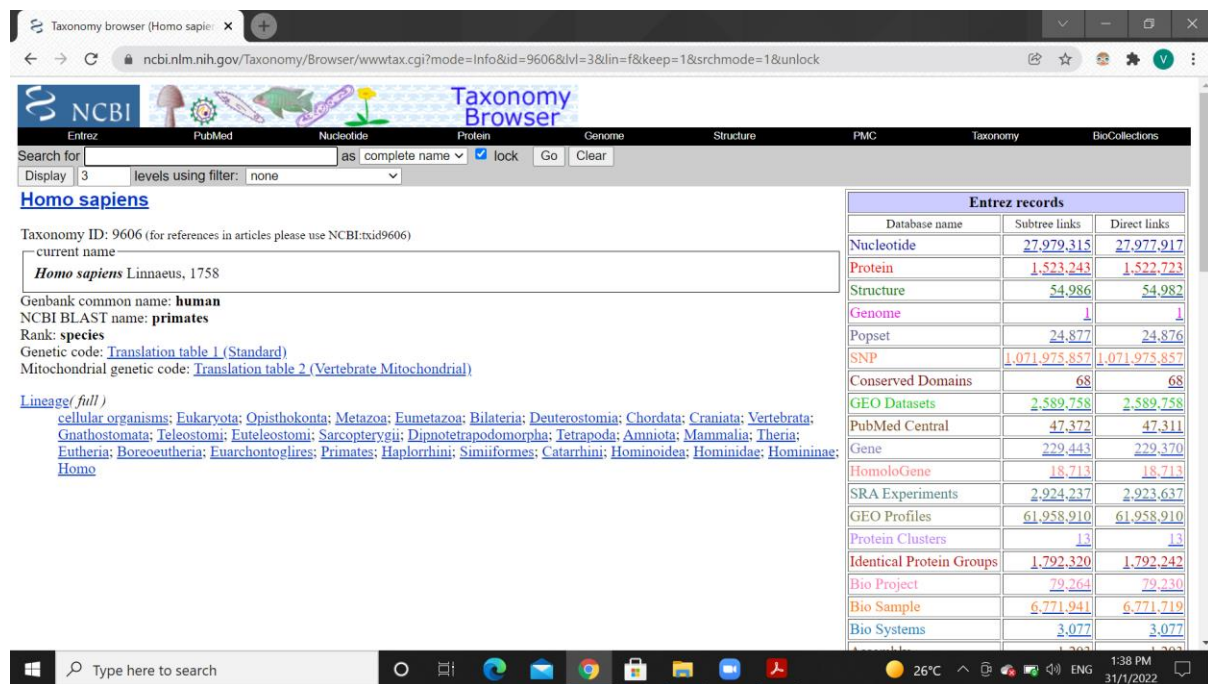
11) From the ncbi database, and using its taxonomy browser, I found the taxonomy IDs of given organisms.

- Human:

Scientific name: *Homo sapiens*

Taxonomy ID : 9606

No. of chromosomes: 46(23 pairs)



Search for  as  complete name ☒ lock Go Clear

Display 3 levels using filter: none

**Homo sapiens**

Taxonomy ID: 9606 (for references in articles please use NCBI:txid9606)

current name  
*Homo sapiens* Linnaeus, 1758

Genbank common name: **human**  
NCBI BLAST name: **primates**  
Rank: **species**  
Genetic code: [Translation table 1 \(Standard\)](#)  
Mitochondrial genetic code: [Translation table 2 \(Vertebrate Mitochondrial\)](#)

[Lineage \(full\)](#)  
cellular organisms; Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Deuterostomia; Chordata; Craniata; Vertebrata; Gnathostomata; Teleostomi; Euteleostomi; Sarcopterygii; Dipnotetrapodomorpha; Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Boreoeutheria; Euarchontoglires; Primates; Haplorrhini; Simiiformes; Catarrhini; Hominoidea; Hominidae; Homininae; Homo

| Database name            | Subtree links                 | Direct links                  |
|--------------------------|-------------------------------|-------------------------------|
| Nucleotide               | <a href="#">27,979,315</a>    | <a href="#">27,977,917</a>    |
| Protein                  | <a href="#">1,523,243</a>     | <a href="#">1,522,723</a>     |
| Structure                | <a href="#">54,986</a>        | <a href="#">54,982</a>        |
| Genome                   | <a href="#">1</a>             | <a href="#">1</a>             |
| Popset                   | <a href="#">24,877</a>        | <a href="#">24,876</a>        |
| SNP                      | <a href="#">1,071,975,857</a> | <a href="#">1,071,975,857</a> |
| Conserved Domains        | <a href="#">68</a>            | <a href="#">68</a>            |
| GEO Datasets             | <a href="#">2,589,758</a>     | <a href="#">2,589,758</a>     |
| PubMed Central           | <a href="#">47,372</a>        | <a href="#">47,311</a>        |
| Gene                     | <a href="#">229,443</a>       | <a href="#">229,370</a>       |
| HomoloGene               | <a href="#">18,713</a>        | <a href="#">18,713</a>        |
| SRA Experiments          | <a href="#">2,924,237</a>     | <a href="#">2,923,637</a>     |
| GEO Profiles             | <a href="#">61,958,910</a>    | <a href="#">61,958,910</a>    |
| Protein Clusters         | <a href="#">13</a>            | <a href="#">13</a>            |
| Identical Protein Groups | <a href="#">1,792,320</a>     | <a href="#">1,792,242</a>     |
| Bio Project              | <a href="#">79,264</a>        | <a href="#">79,230</a>        |
| Bio Sample               | <a href="#">6,771,941</a>     | <a href="#">6,771,719</a>     |
| Bio Systems              | <a href="#">3,077</a>         | <a href="#">3,077</a>         |

- Cat:

Scientific name: *Felis catus*

Taxonomy ID : 9685

No. of chromosomes: 38(18 pairs, 1X, 1Y=20)



Taxonomy browser (Felis catus) x

ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi

NCBI Taxonomy Browser

Search for **cat** as complete name lock Go Clear

Display 3 levels using filter: none

**Felis catus**

Taxonomy ID: 9685 (for references in articles please use NCBI:txid9685)

current name

**Felis catus** Linnaeus, 1758

homotypic synonym: **Felis silvestris catus**

includes: **Korat cats** L.

Genbank common name: **domestic cat**

NCBI BLAST name: **carnivores**

Rank: **species**

Genetic code: [Translation table 1 \(Standard\)](#)

Mitochondrial genetic code: [Translation table 2 \(Vertebrate Mitochondrial\)](#)

Other names:

heterotypic synonym

**Felis domesticus**

common name(s)

**cat, cats**

[Lineage/ full \)](#)

[cellular organisms](#); [Eukaryota](#); [Opisthokonta](#); [Metazoa](#); [Eumetazoa](#); [Bilateria](#); [Deuterostomia](#); [Chordata](#); [Cranata](#); [Vertebrata](#); [Gnathostomata](#); [Teleostomi](#); [Euteleostomi](#); [Sarcopterygii](#); [Dipnotetrapodomorpha](#); [Tetrapoda](#); [Amniota](#); [Mammalia](#); [Theria](#); [Eutheria](#); [Boreoeutheria](#); [Laurasiatheria](#); [Carnivora](#); [Feliformia](#); [Felidae](#); [Felinae](#); [Felis](#)

| Database name            | Direct links |
|--------------------------|--------------|
| Nucleotide               | 115,198      |
| Protein                  | 75,013       |
| Structure                | 21           |
| Genome                   | 1            |
| Popset                   | 207          |
| GEO Datasets             | 394          |
| PubMed Central           | 3,618        |
| Gene                     | 54,020       |
| SRA Experiments          | 2,649        |
| Protein Clusters         | 12           |
| Identical Protein Groups | 56,717       |
| Bio Project              | 121          |
| Bio Sample               | 1,891        |
| Bio Systems              | 495          |
| Assembly                 | 10           |
| Probe                    | 2,877        |
| PubChem BioAssay         | 1,118        |
| Taxonomy                 | 1            |

- **Dog:**

Scientific name: **Canis lupus familiaris**

Taxonomy ID : 9615

No. of chromosomes: 78(38 pairs, 1X, 1Y=40)

Taxonomy browser (Canis lupus) x

ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi

NCBI Taxonomy Browser

Search for **dog** as complete name lock Go Clear

Display 3 levels using filter: none

**Canis lupus familiaris**

Taxonomy ID: 9615 (for references in articles please use NCBI:txid9615)

current name

**Canis lupus familiaris** Linnaeus, 1758

homotypic synonym: **Canis familiaris** Linnaeus, 1758

includes: **beagle dog**

Genbank common name: **dog**

NCBI BLAST name: **carnivores**

Rank: **subspecies**

Genetic code: [Translation table 1 \(Standard\)](#)

Mitochondrial genetic code: [Translation table 2 \(Vertebrate Mitochondrial\)](#)

Other names:

heterotypic synonym

**Canis canis**

heterotypic synonym

**Canis domesticus**

common name(s)

**dogs**

[Lineage/ full \)](#)

[cellular organisms](#); [Eukaryota](#); [Opisthokonta](#); [Metazoa](#); [Eumetazoa](#); [Bilateria](#); [Deuterostomia](#); [Chordata](#); [Cranata](#); [Vertebrata](#); [Gnathostomata](#); [Teleostomi](#); [Euteleostomi](#); [Sarcopterygii](#); [Dipnotetrapodomorpha](#); [Tetrapoda](#); [Amniota](#); [Mammalia](#); [Theria](#); [Eutheria](#); [Boreoeutheria](#); [Laurasiatheria](#); [Carnivora](#); [Caniformia](#); [Canidae](#); [Canis](#)

| Database name            | Direct links |
|--------------------------|--------------|
| Nucleotide               | 1,663,640    |
| Protein                  | 156,072      |
| Structure                | 143          |
| Genome                   | 1            |
| Popset                   | 374          |
| Conserved Domains        | 1            |
| GEO Datasets             | 12,805       |
| PubMed Central           | 6,389        |
| Gene                     | 69,500       |
| HomoloGene               | 17,434       |
| SRA Experiments          | 20,622       |
| GEO Profiles             | 205,497      |
| Protein Clusters         | 13           |
| Identical Protein Groups | 68,659       |
| Bio Project              | 732          |
| Bio Sample               | 20,020       |
| Bio Systems              | 1,984        |
| Assembly                 | 157          |

- **House mouse:**

Scientific name: **Mus musculus**

Taxonomy ID : 10090

No. of chromosomes: 40(20 pairs)

Taxonomy browser (Mus musculi) x

ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?mode=Info&id=10090&lvl=3&lin=f&keep=1&srchmode=1&unlock

NCBI Taxonomy Browser

Search for  as complete name ☒ lock Go Clear

Display 3 levels using filter: none

**Mus musculus**

Taxonomy ID: 10090 (for references in articles please use NCBI:txid10090)

current name

**Mus musculus** Linnaeus, 1758

includes: **LK3 transgenic mice**

Genbank common name: **house mouse**

NCBI BLAST name: **rodents**

Rank: **species**

Genetic code: [Translation table 1 \(Standard\)](#)

Mitochondrial genetic code: [Translation table 2 \(Vertebrate Mitochondrial\)](#)

Other names:

common name(s)

**mouse**

**Entrez records**

| Database name            | Subtree links | Direct links |
|--------------------------|---------------|--------------|
| Nucleotide               | 11,063,116    | 10,453,119   |
| Protein                  | 348,571       | 340,771      |
| Structure                | 7,631         | 7,627        |
| Genome                   | 1             | 1            |
| Popset                   | 8,618         | 8,374        |
| Conserved Domains        | 23            | 23           |
| GEO Datasets             | 1,451,327     | 1,450,473    |
| PubMed Central           | 32,722        | 32,347       |
| Gene                     | 212,430       | 212,335      |
| HomoloGene               | 19,032        | 19,032       |
| SRA Experiments          | 1,470,340     | 1,459,554    |
| GEO Profiles             | 50,177,624    | 50,177,624   |
| Protein Clusters         | 13            | 13           |
| Identical Protein Groups | 203,743       | 201,183      |
| Bio Project              | 53,967        | 53,747       |
| Bio Sample               | 1,494,371     | 1,483,849    |
| Bio Systems              | 2,369         | 2,369        |
| Assembly                 | 52            | 45           |

Lineage (full)

cellular organisms: Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Deuterostomia; Chordata; Craniata; Vertebrata; Gnathostomata; Teleostomi; Euteleostomi; Sarcopterygii; Dipnotetrapodomorpha; Tetrapoda; Amniota; Mammalia; Theria; Eutheria; Boreoeutheria; Euarchontoglires; Glires; Rodentia; Myomorpha; Muroidea; Muridae; Murinae; Mus; Mus

26°C 1:39 PM 31/1/2022

- Onion:

Scientific name: *Allium cepa*

Taxonomy ID : 4679

No. of chromosomes: 16(8 pairs)

Taxonomy browser (Allium cepa) x

ncbi.nlm.nih.gov/Taxonomy/Browser/wwwtax.cgi?mode=Info&id=4679&lvl=3&lin=f&keep=1&srchmode=1&unlock

NCBI Taxonomy Browser

Search for  as complete name ☒ lock Go Clear

Display 3 levels using filter: none

**Allium cepa**

Taxonomy ID: 4679 (for references in articles please use NCBI:txid4679)

current name

**Allium cepa** L.

Genbank common name: **onion**

NCBI BLAST name: **monocots**

Rank: **species**

Genetic code: [Translation table 1 \(Standard\)](#)

Mitochondrial genetic code: [Translation table 1 \(Standard\)](#)

Plastid genetic code: [Translation table 11 \(Bacterial, Archaeal and Plant Plastid\)](#)

**Entrez records**

| Database name            | Subtree links | Direct links |
|--------------------------|---------------|--------------|
| Nucleotide               | 609,418       | 609,403      |
| Protein                  | 1,383         | 1,370        |
| Structure                | 6             | 6            |
| Genome                   | 1             | 1            |
| Popset                   | 116           | 116          |
| GEO Datasets             | 63            | 63           |
| PubMed Central           | 3,378         | 3,378        |
| Gene                     | 179           | 179          |
| SRA Experiments          | 982           | 976          |
| Identical Protein Groups | 509           | 499          |
| Bio Project              | 50            | 46           |
| Bio Sample               | 953           | 946          |
| Assembly                 | 1             | 1            |
| Probe                    | 237           | 237          |
| PubChem BioAssay         | 47            | 47           |
| Taxonomy                 | 2             | 1            |

Lineage (full)

cellular organisms: Eukaryota; Viridiplantae; Streptophyta; Streptophytina; Embryophyta; Tracheophyta; Euphyllophyta; Spermatophyta; Magnoliopsida; Mesangiospermae; Liliopsida; Petrosavidae; Asparagales; Amaryllidaceae; Allioidae; Allieae; Allium

Comments and References:

26°C 1:40 PM 31/1/2022

- Thale cress:

Scientific name: *Arabidopsis thaliana*

Taxonomy ID : 3702

No. of chromosomes: 5

Search for **thale cress** as complete name lock Go Clear

Display 3 levels using filter: none

**Arabidopsis thaliana**

Taxonomy ID: 3702 (for references in articles please use NCBI:txid3702)

current name  
**Arabidopsis thaliana** (L.) Heynh., 1842  
 basionym: *Arabis thaliana* L., 1753

Genbank common name: **thale cress**  
 NCBI BLAST name: **eudicots**  
 Rank: **species**  
 Genetic code: [Translation table 1 \(Standard\)](#)  
 Mitochondrial genetic code: [Translation table 1 \(Standard\)](#)  
 Plastid genetic code: [Translation table 11 \(Bacterial, Archaeal and Plant Plastid\)](#)  
 Other names:  
 common name(s)  
**mouse-ear cress, thale-cress**

[Lineage \(full\)](#)  
 cellular organisms; Eukaryota; Viridiplantae; Streptophyta; Streptophytina; Embryophyta; Tracheophyta; Euphyllophyta; Spermatophyta; Magnoliopsida; Mesangiospermae; eudicotyledons; Gunneridae; Pentapetalae; rosids; malvids; Brassicales; Brassicaceae; Camelineae; Arabidopsis

| Database name            | Direct links              |
|--------------------------|---------------------------|
| Nucleotide               | <a href="#">2,700,232</a> |
| Protein                  | <a href="#">370,512</a>   |
| Structure                | <a href="#">1,702</a>     |
| Genome                   | <a href="#">1</a>         |
| Popset                   | <a href="#">1,312</a>     |
| Conserved Domains        | <a href="#">51</a>        |
| GEO Datasets             | <a href="#">83,296</a>    |
| PubMed Central           | <a href="#">79,286</a>    |
| Gene                     | <a href="#">44,111</a>    |
| HomoloGene               | <a href="#">10,445</a>    |
| SRA Experiments          | <a href="#">104,664</a>   |
| GEO Profiles             | <a href="#">2,014,308</a> |
| Protein Clusters         | <a href="#">19,225</a>    |
| Identical Protein Groups | <a href="#">134,574</a>   |
| Bio Project              | <a href="#">6,959</a>     |
| Bio Sample               | <a href="#">126,260</a>   |
| Bio Systems              | <a href="#">1,028</a>     |
| Assembly                 | <a href="#">109</a>       |

12) NCBI E-utilities are nine set of server side programs, which provide stable interface for Entrez query and for accessing the database in NCBI. The E-utilities are set of simple server syntax codes which helps us to retrieve or fetch the requested data.

This is an example syntax for fetching homo sapiens related sequences, from ncbi in fasta format.

*esearch -db ncbi -query "homo sapiens sequences"*

*elink -receptor gene*

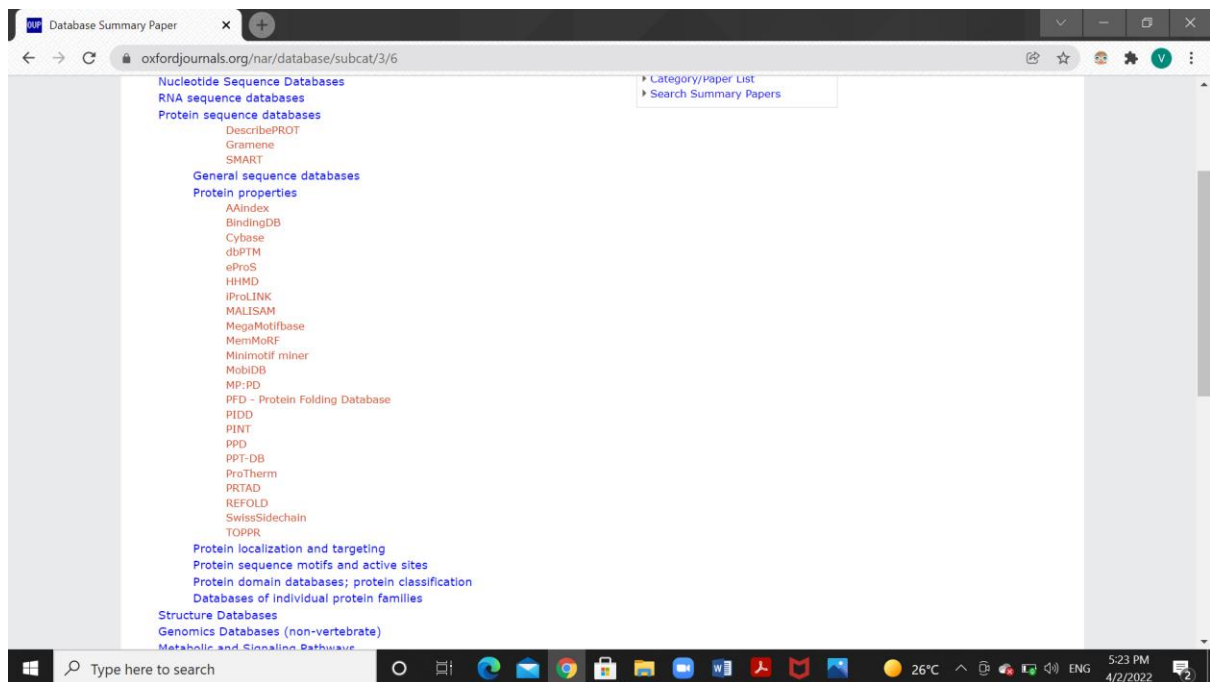
*efetch -format fasta*

13) Using Nucleic Acids Research (NAR), database category list, I found different databases for given categories:

a) Protein properties:

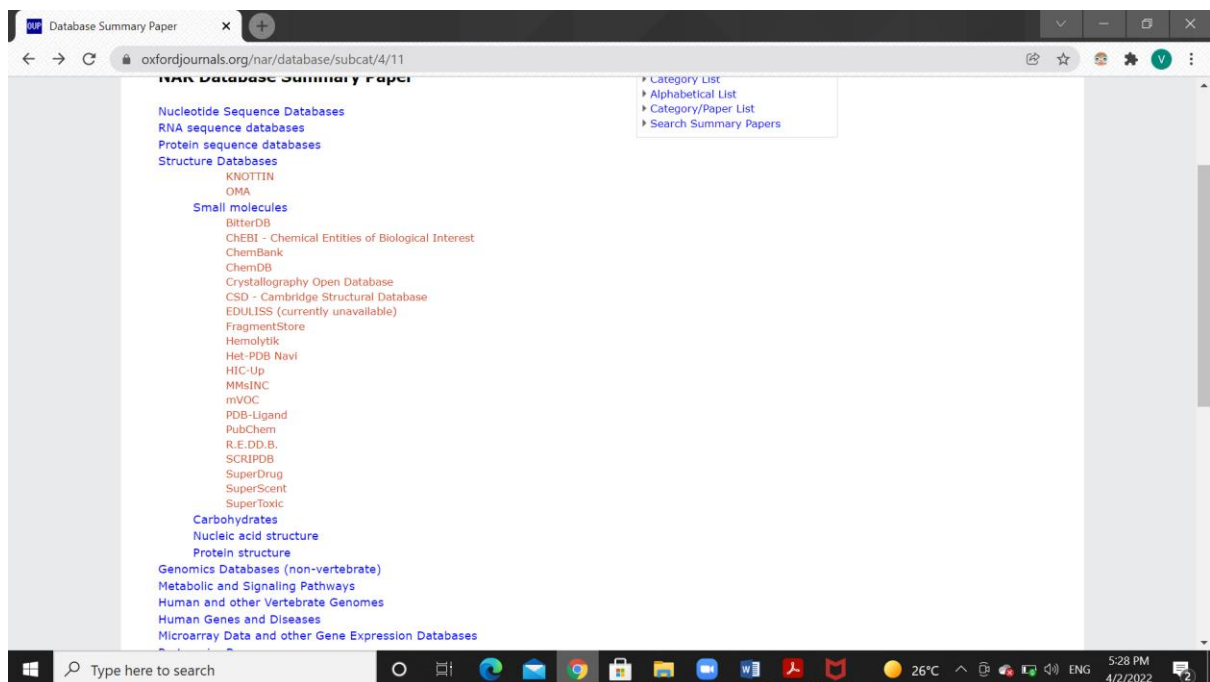
- HHMD – Human Histone Modification Database

- *PIDD – Database for protein modelling*



*b) Small molecules (Structure related):*

- *BitterDB – For bitter taste molecules and receptors.*
- *Hemolytik – For hemolytic and non-hemolytic peptides.*



*c) Cancer gene databases:*

- *BCCTBbp - Breast Cancer Campaign Tissue Bank bioinformatics portal*
- *CanGEM - Gene copy number changes in cancer*

