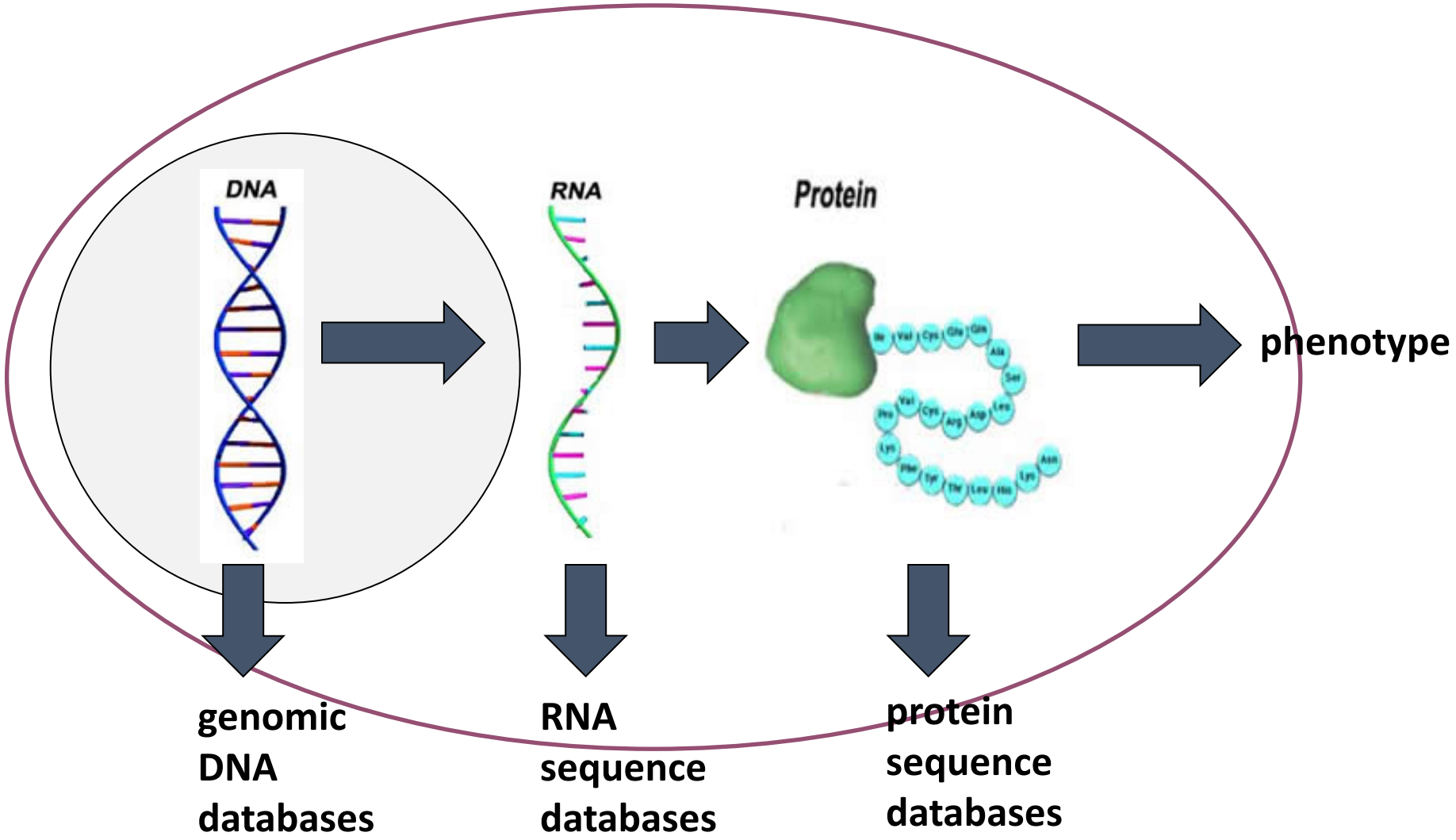
Bioinformatics

WEEK 2

NCBI

DATA


DNA
RNA
Protein
phenotype
genomic DNA databases
RNA sequence databases
protein sequence databases

Major public DNA databases

EMBL: Housed at European Bioinformatics Institute (EBI)
GenBank: Housed at National Center for Biotechnology Information (NCBI)
DDBJ: Housed in Japan

Secondary nucleotide sequence databases

UniGene
SGD
EMI Genomes
Genome Biology

Protein sequence databases

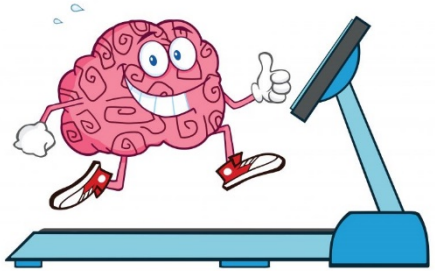
SwissProt
PIR

Protein structure databases

Protein Data Bank (PDB)
SCOP
CATH

Other relevant databases

Pfam
KEGG
PROSITE



Go to NCBI website!

http://www.ncbi.nlm.nih.gov/

ncbi.nlm.nih.gov

NCBI Resources How To Sign in to NCBI

NCBI
National Center for Biotechnology Information

All Databases | Search

COVID-19 is an emerging, rapidly evolving situation.
Get the latest public health information from CDC: https://www.coronavirus.gov .
Get the latest research from NIH: https://www.nih.gov/coronavirus.
Find NCBI SARS-CoV-2 literature, sequence, and clinical content: https://www.ncbi.nlm.nih.gov/sars-cov-2/.

- NCBI Home
- Resource List (A-Z)
- All Resources
- Chemicals & Bioassays
- Data & Software
- DNA & RNA
- Domains & Structures
- Genes & Expression
- Genetics & Medicine
- Genomes & Maps
- Homology
- Literature
- Proteins
- Sequence Analysis
- Taxonomy
- Training & Tutorials
- Variation

Welcome to NCBI

The National Center for Biotechnology Information advances science and health by providing access to biomedical and genomic information.

About the NCBI | Mission | Organization | NCBI News & Blog

Submit
Deposit data or manuscripts into NCBI databases

Download
Transfer NCBI data to your computer

Learn
Find help documents, attend a class or watch a tutorial

Develop
Use NCBI APIs and code libraries to build applications

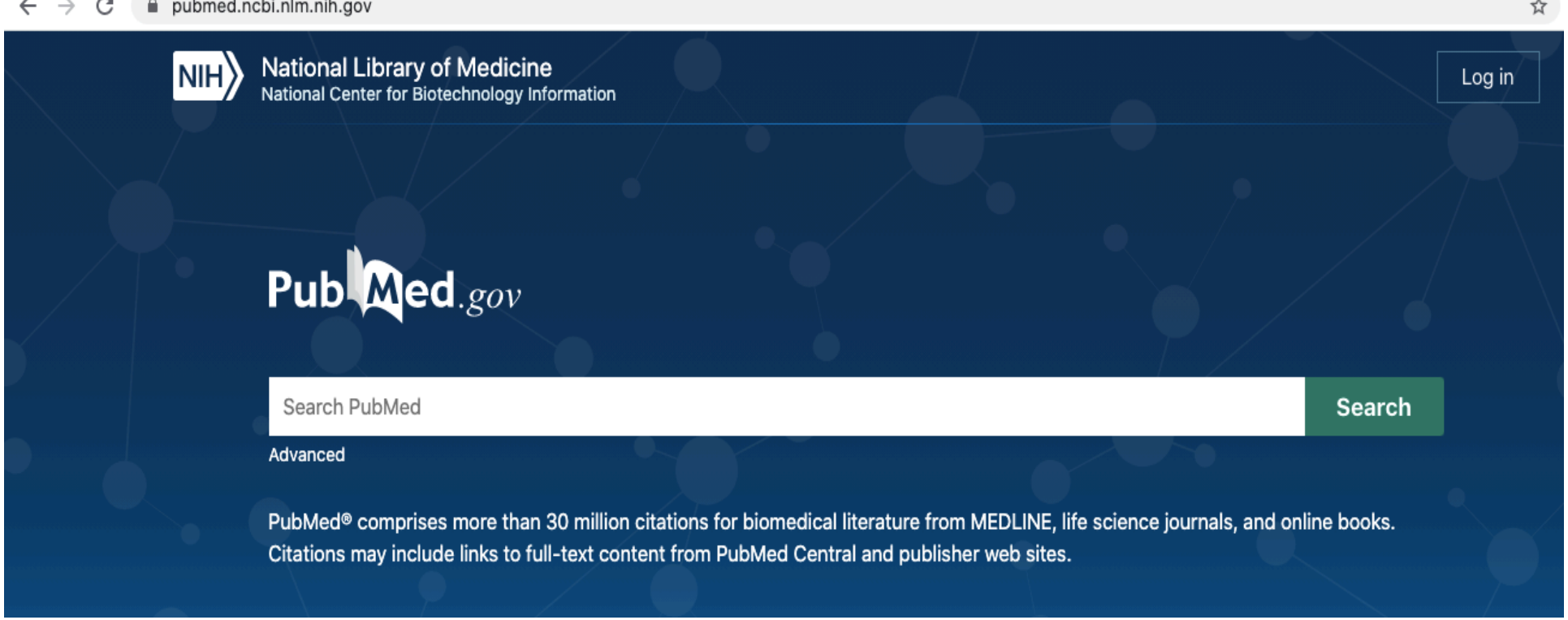
Analyze
Identify an NCBI tool for your data analysis task

Research
Explore NCBI research and collaborative projects

Popular Resources
- PubMed
- Bookshelf
- PubMed Central
- BLAST
- Nucleotide
- Genome
- SNP
- Gene
- Protein
- PubChem

NCBI News & Blog
Primer-BLAST now designs primers for a group of related sequences
25 Sep 2020
Primer-BLAST now has a "Primers common for a group of sequences"

Improved chromosome searching in Genome Browsers
18 Sep 2020
Are you interested in searching for a



PubMed

PubMed is a free resource supporting the search and retrieval of biomedical and life sciences literature with the aim of improving health–both globally and personally.

The PubMed database contains more than 30 million citations and abstracts of biomedical literature. It does not include full-text journal articles; however, links to the full text are often present.

ncbi.nlm.nih.gov/Web/Search/entrezfs.html

NCBI

Entrez Molecular Sequence Database System

PubMed | Entrez | BLAST | OMIM | Taxonomy | Structure

NCBI

SITE MAP

Introduction

Entrez is a molecular biology database system that provides integrated access to nucleotide and protein sequence data, gene-centered and genomic mapping information, 3D structure data, PubMed MEDLINE, and more. The system is produced by the National Center for Biotechnology Information (NCBI) and is available via the Internet.

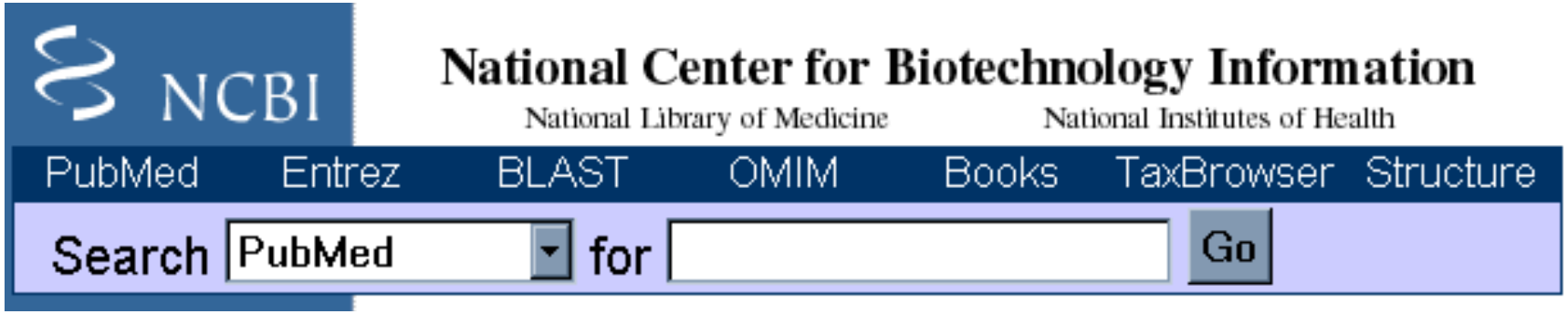
Entrez Databases and Retrieval System

Entrez covers over 20 databases including the complete protein sequence data from PIR-International, PRF, Swiss-Prot, and PDB and nucleotide sequence data from GenBank that includes information from EMBL and DDBJ.

The Entrez retrieval system uses an intuitive user interface for rapidly searching sequence and bibliographic data. A unique feature of the system is its use of precomputed similarity searches for each record to create links to "neighbors" or related records in other Entrez databases. These links facilitate integrated access across the various databases. An Entrez global query provides search capability for a subset of Entrez databases at one time. Results may be viewed in various formats inlcuding FlatFile, FASTA, XML, and others. A graphical interface provides easy visualization of complete genomes or chromosomes, as well as biological annotation on individual sequences. Entrez also allows Batch downloads of large search results.

Internet Access to Entrez

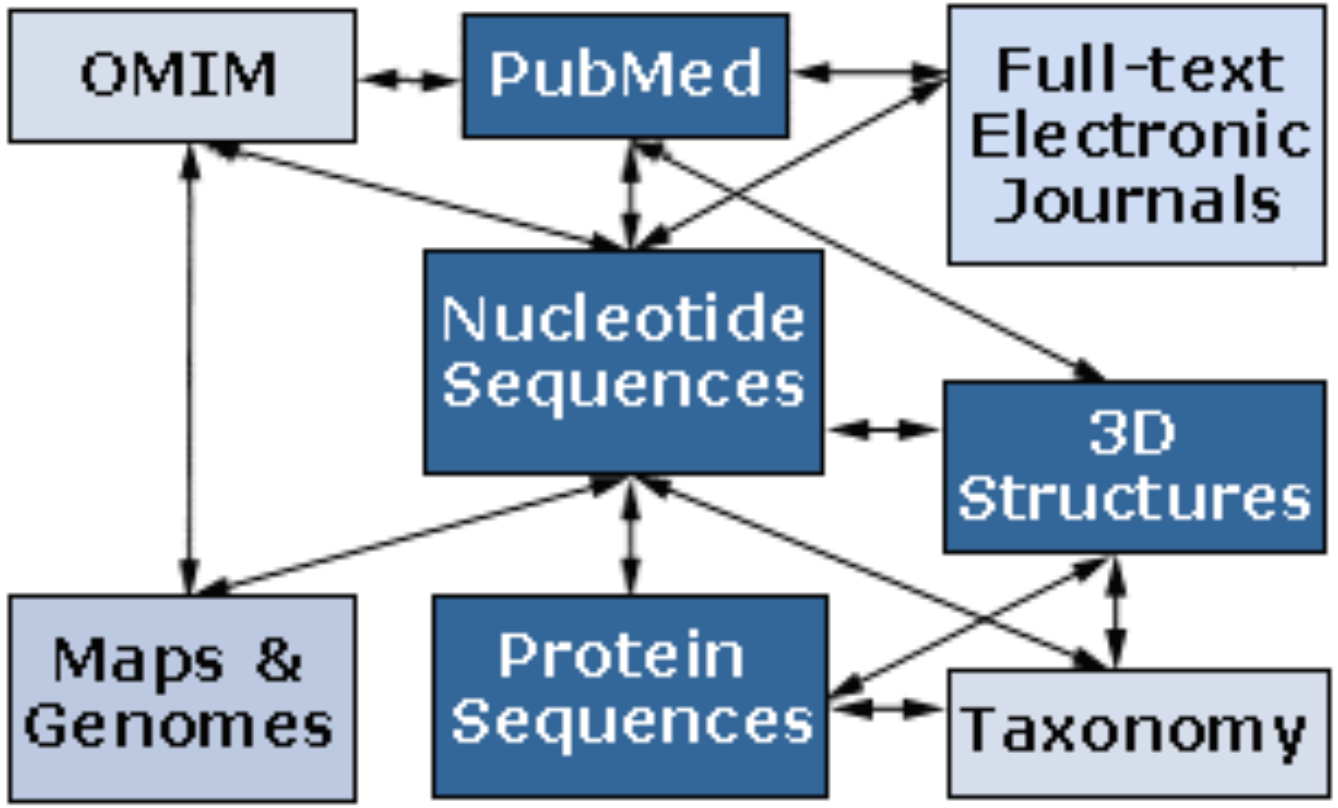
Entrez is available via the World Wide Web at http://www.ncbi.nlm.nih.gov/Entrez/.



Entrez integrates…

- the scientific literature;
- DNA and protein sequence databases;
- 3D protein structure data;
- population study data sets;
- assemblies of complete genomes

Entrez is a search and retrieval system that integrates NCBI databases



ncbi.nlm.nih.gov/omim

NCBI Resources How To Sign in to NCBI

OMIM OMIM Search

Limits Advanced Help

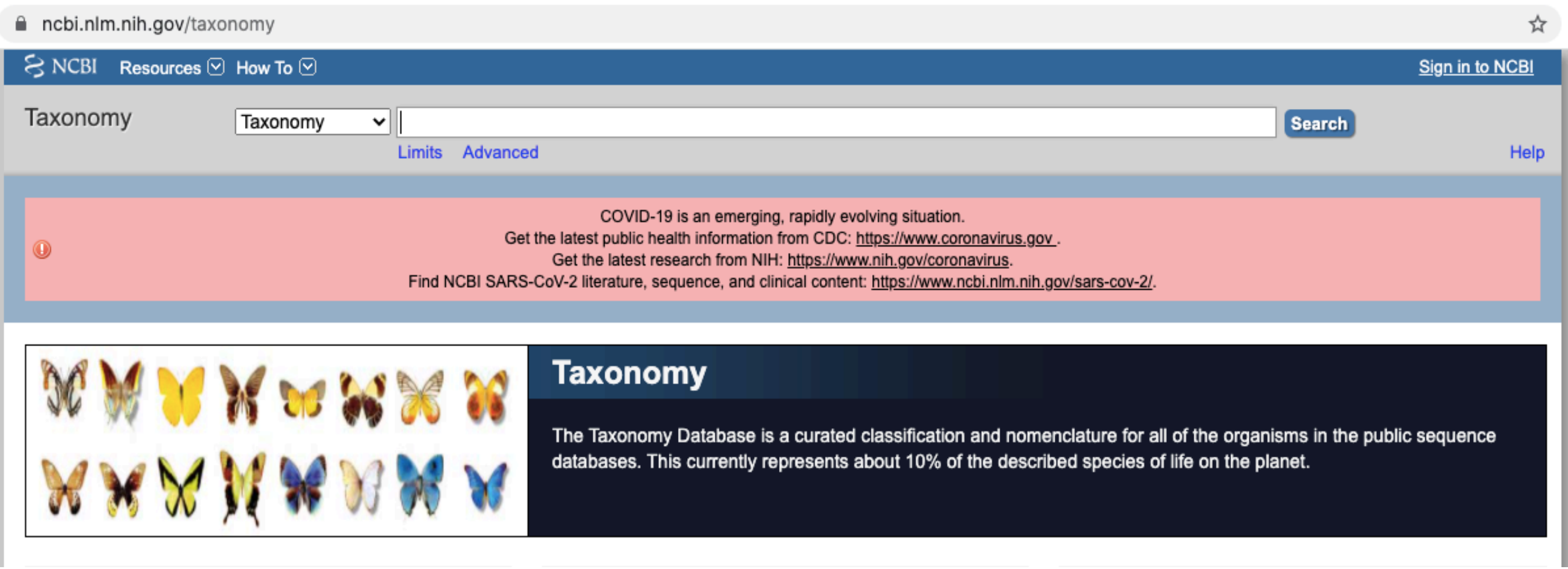
COVID-19 is an emerging, rapidly evolving situation.
Get the latest public health information from CDC: https://www.coronavirus.gov .
Get the latest research from NIH: https://www.nih.gov/coronavirus.
Find NCBI SARS-CoV-2 literature, sequence, and clinical content: https://www.ncbi.nlm.nih.gov/sars-cov-2/.

OMIM

OMIM is a comprehensive, authoritative compendium of human genes and genetic phenotypes that is freely available and updated daily. OMIM is authored and edited at the McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, under the direction of Dr. Ada Hamosh. Its official home is omim.org.

OMIM

Online Mendelian Inheritance in Man (OMIM®) is a continuously updated catalog of human genes and genetic disorders and traits, with particular focus on the molecular relationship between genetic variation and phenotypic expression.



TaxBrowser is…

- browser for the major divisions of living organisms (archaea, bacteria, eukaryota, viruses)
- taxonomy information such as genetic codes
- molecular data on extinct organisms

Similarity Search: BLAST

A tool for searching gene or protein sequence databases for related genes of interest

Alignments between the query sequence and any given database sequence, allowing for mismatches and gaps, indicate their degree of similarity

The structure, function, and evolution of a gene may be determined by such comparisons.

% identity

```
CATTATGATA
   |||||||
GTTTATGATT
```

70%

```
MRCKTETGAR
|||  |||||||
MRCGTETGAR
```

90%

Strengths:

Accessibility

Growing rapidly

User friendly

Weaknesses:

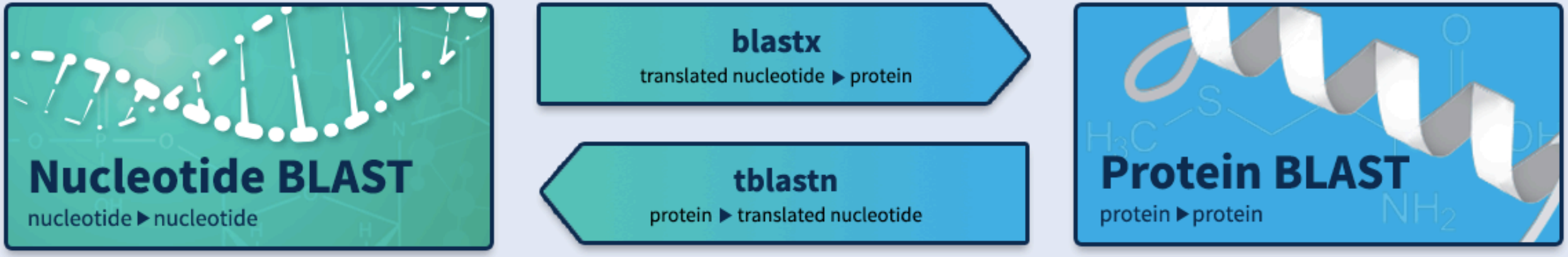
Sometimes not up-to-date

Limited possibilities

Limited comparisons and information

Not always accurate

Web BLAST



BLAST is…

- Basic Local Alignment Search Tool
- NCBI's sequence similarity search tool
- supports analysis of DNA and protein databases

BLAST's Types:

- BLASTN programs search nucleotide databases using a nucleotide query.
- BLASTP programs search protein databases using a protein query.
- BLASTX search protein databases using a translated nucleotide query.
- TBLASTN search translated nucleotide databases using a protein query.
- TBLASTX search translated nucleotide databases using a translated nucleotide query.

NIH U.S. National Library of Medicine | NCBI National Center for Biotechnology Information

BLAST® » blastn suite

COVID-19 is an emerging, rapidly evolving situation.
Get the latest public health information from CDC: https://www.coronavirus.gov .
Get the latest research from NIH: https://www.nih.gov/coronavirus.
Find NCBI SARS-CoV-2 literature, sequence, and clinical content: https://www.ncbi.nlm.nih.gov/sars-cov-2/.

Standard Nucleotide BLAST

blastn | blastp | blastx | tblastn | tblastx

BLASTN programs search nucleotide databases using a nucleotide query. more...

Enter Query Sequence

Enter accession number(s), gi(s), or FASTA sequence(s)

Clear

Query subrange

From

To

Or, upload file

Choose File No file chosen

Job Title

Enter a descriptive title for your BLAST search

Align two or more sequences

Choose Search Set

Database

Standard databases (nr etc.): rRNA/ITS databases Genomic + transcript databases Betacoronavirus

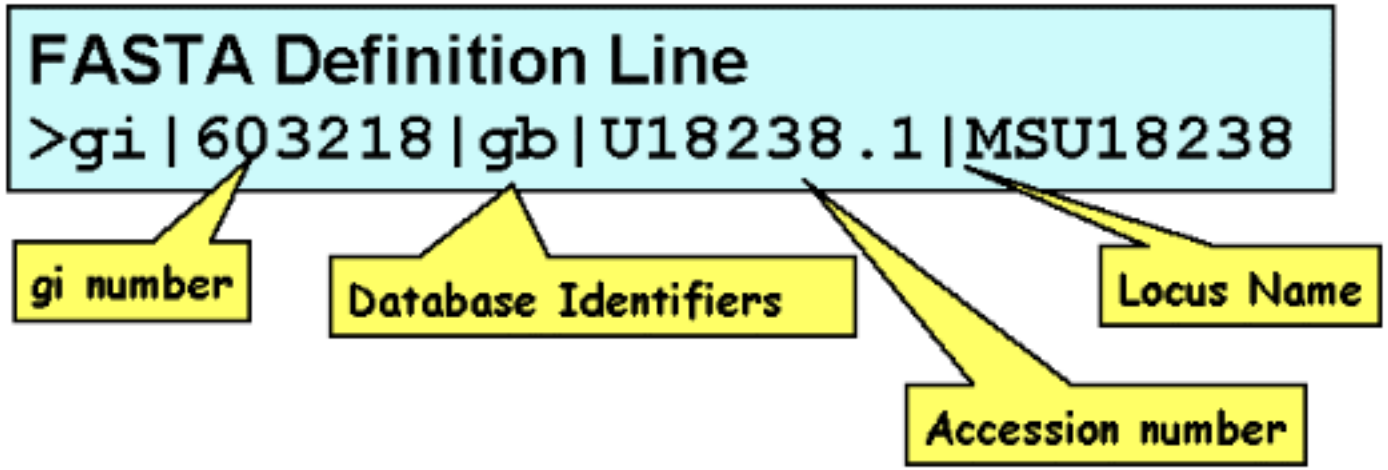
Nucleotide collection (nr/nt)

Organism
Optional

Enter organism name or id--completions will be suggested

exclude +

Enter organism common name, binomial, or tax id. Only 20 top taxa will be shown



What is FASTA format?

FASTA format is a text-based format for representing either nucleotide sequences or peptide sequences, in which base pairs or amino acids are represented using single-letter codes. A sequence in FASTA format begins with a single-line description, followed by lines of sequence data. The description line is distinguished from the sequence data by a greater-than (">") symbol in the first column. An example sequence in FASTA format is:

>gi|186681228|ref|YP_001864424.1| phycoerythrobilin:ferredoxin oxidoreductase
MNSERSDVTLYQPFLDYAIAYMRSRLDLEPYPIPTGFESNSAVVGKGKNQEEVVTTSYAFQTAKLRQIRA
AHVQGGNSLQVLNFVIFPHLNYDLPFFGADLVTLPGGHLIALDMQPLFRDDSAYQAKYTEPILPIFHAHQ
QHLSWGGDFPEEAQPFFSPAFLWTRPQETAVVETQVFAAFKDYLKAYLDFVEQAEAVTDSQNLVAIKQAQ
LRYLRYRAEKDPARGMFKRFYGAEWTEEYIHGFLFDLERKLTVVK

Dinlediğiniz için Teşekkürler…