GENETIC ARCHIVES AND LIFELINE: UNRAVELING THE UMBILICAL CORD'S BIOLOGICAL DATABASE

Dr CK Gomathy-Assistant Professor, Department of CSE, SCSVMV Deemed to be University, India

Mr.Rajasrivatsan Srinivasan, Ms.L Sai Purna Srinidhi Perumalla,-UG Scholars,

Department of CSE, SCSVMV deemed to be University, India.

Abstract:

Biological databases are critical components of the field of bioinformatics, which is dedicated to managing, organising, and making sense of the vast and diverse biological data generated by scientific research. These databases play a fundamental role in modern life sciences, allowing researchers, clinicians, and scientists to access, analyse, and interpret a wide range of biological information. These databases serve as invaluable resources, enabling researchers to access, manage, and analyse biological data, accelerating scientific discoveries and enhancing our understanding of life's complexities. There are various types of biological databases that serve as critical resources for researchers and scientists in the field of life sciences. These databases are categorised based on the types of biological data they store and manage.

Keywords: Bioinformatics, biological database, genetics, scientific research, genomics, medical advancements, data repositories

I. INTRODUCTION

From the mysteries locked within our DNA to the sprawling networks of proteins that orchestrate cellular functions, the biological world is a complex web waiting to be unravelled. These tools are biological databases, the unsung heroes of the genomic era, offering a treasure trove of information that has revolutionised our understanding of life on Earth. They house a staggering wealth of data, encompassing genetic sequences, protein structures, clinical insights, and much more. These databases have become the backbone of modern biological research, enabling scientists to explore, analyse, and interpret the immense complexity of life.

II. BIOLOGICAL DATABASES

A biological database is a large, organised body of persistent data, usually associated with computerised software designed to update, query, and retrieve components of the data stored within the system. A simple database might be a single file containing many records, each of which includes the same set of information. The chief objective of the development of a database is to organise data in a set of structured records to enable easy retrieval of information.

Components of Biological Databases:

- **Entity :** An entity refers to the thing we want to store in a database. For example, DNA sequence, genes, bibliography references, etc.
- **Fields :** The properties of an entity are called fields. For example, gene name, gene sequence, mutation(if any), etc.
- **Records :** A record typically refers to a combination of all the fields for a given entity. For example, a record for Genbank, stem cell - umbilical cord(umbilical cord that connects the baby to the mother's placenta).
- **Identifier :** The unique name which identifies a record.

III. IMPORTANCE OF DATABASE

Biological databases are of paramount importance in modern life sciences, serving as the linchpin of data-driven research and discovery. These repositories play a crucial role in organising, storing, and providing access to an ever-expanding wealth of biological information, from genomic sequences and protein structures to clinical data and pathways. Their significance lies in facilitating efficient data retrieval, enabling data integration, and promoting collaboration among researchers. They empower scientists to explore the genetic underpinnings of life, investigate disease mechanisms, and advance personalised medicine. These databases are not mere archives but dynamic catalysts for innovation, fostering a deeper understanding of the biological world and shaping the future of scientific exploration.

IV. TYPES OF DATABASE

Based on their contents, biological databases can be roughly divided into three categories as primary, secondary and derived (or) composite databases.

❖ ***Primary database:***

Primary databases are populated with experimentally derived data such as nucleotide sequence, protein sequence and macromolecular structure. Experimental results are submitted directly into the database by researchers and the data are essentially archival in nature. They are also called the Archival database.

- Nucleotide sequence is the most fundamental level of knowledge of a gene or genome.
- Protein sequencing is the practical process of determining the amino acid sequence of all or part of a protein. (Amino acids are molecules that combine to form protein.)
- Macromolecular structures are made up of basic molecular units. They include proteins (polymers of amino acids), nucleic acids (polymers of nucleotides), carbohydrates (polymers of sugars) and lipids (with a variety of modular constituents).

Some of the examples of the primary databases are as follows:

1. **GenBank :** GenBank is a widely used database of nucleotide sequences, particularly DNA sequences. Maintained by the National Center for Biotechnology Information (NCBI), GenBank is a comprehensive repository of nucleotide sequences, including DNA and RNA sequences.
2. **Protein Data Bank (PDB):** The PDB is a comprehensive repository of three dimensional structural data of biological macromolecules, such as proteins and nucleic acids.
3. **DNA Data Bank of Japan (DDBJ) :** The DDBJ is one of the three major databases in the world that store and manage nucleotide sequence data. The other two are GenBank in the United States and the European Nucleotide Archive (ENA) in Europe. It is a crucial resource for molecular biologists and bioinformaticians, as it provides a platform to deposit, search, and access DNA and RNA sequences.

4. **European Nucleotide Archive (ENA) :** The ENA is a comprehensive database for storing and sharing nucleotide sequence data, particularly DNA and RNA sequences.
5. **NCBI's PubMed :** PubMed is a widely used online database that provides access to a vast collection of scientific articles in the field of life sciences and biomedicine. It is maintained by the National Center for Biotechnology Information (NCBI), a part of the United States National Library of Medicine (NLM).

❖ ***Secondary database:***

Secondary databases are critical resources that collect, curate, and organise data derived from primary biological databases. These secondary databases often provide more specialised and integrated information for specific research purposes. Secondary databases are often referred to as curated databases, but this is a bit of a misnomer because primary databases are also curated to ensure that the data in them is consistent and accurate.

Secondary databases often draw upon information from numerous sources, including other databases (primary and secondary), controlled vocabularies and the scientific literature. They are highly curated, often using a complex combination of computational algorithms and manual analysis and interpretation to derive new knowledge from the public record of science.

The below mentioned are some of the examples of secondary databases:

1. **Swiss-Prot :** SWISS-PROT is a well-known and widely used secondary database of protein sequences that provides detailed annotation, including information on structure, function, and protein family assignment.
2. **Prosite :** ProSite is a database of protein families, domains, and functional sites that contains manually curated information on amino acid patterns and profiles of proteins. It is a secondary protein database that provides tools for the analysis of protein sequences and the identification of motifs.
3. **Pfam :** Pfam is a database of protein families and domains, making it a valuable resource for identifying conserved protein sequences and domains.

4. **UniProt :** UniProt is a primary database for protein sequences, but it also serves as a secondary database by providing comprehensive annotations and cross-references for proteins, including information on function, domains, pathways, and literature citations.
5. **InterPro :** InterPro is an integrated resource that combines information from several primary databases to offer protein domain and functional information. It helps in the annotation and classification of protein sequences.
6. **miRbase :** miRBase is a specialised database for microRNAs, providing sequences and annotations for these small regulatory RNA molecules. This database is responsible for assigning the gene names to novel microRNAs.
7. **NCBI's RefSeq :** While NCBI's GenBank is a primary database for sequences, RefSeq is a curated and well-annotated secondary database that provides comprehensive sequence and functional information for genes, transcripts, and proteins.

V. USES OF BIOLOGICAL DATABASES

- It helps researchers to study the available data and form a new thesis, anti-virus, helpful bacteria, medicines, etc.
- It helps scientists to understand the concepts of biological phenomena.
- The database acts as a storage of information.
- It help remove the redundancy of data

Table 1: Database and its info

Database Name	Information Contained
GenBank	Nucleotide sequences (DNA and RNA)
UniProt	Protein sequences and functional data
PDB (Protein Data Bank)	3D structures of biological molecules
ENA (European Nucleotide Archive)	Nucleotide sequences

NCBI's PubMed	Scientific literature in life sciences
NCBI's RefSeq	Curated and annotated sequences
GO (Gene Ontology)	Gene and gene product attributes
KEGG (Kyoto Encyclopedia of Genes and Genomes)	Pathways and functional data
Reactome	Detailed biological pathway data and reactions
InterPro	Protein domains and functional data
Pfam	Protein families and domains
STRING	Protein-protein interactions and networks
miRBase	MicroRNA sequences and annotation
dbSNP	Information about genetic variations
OMIM (Online Mendelian Inheritance in Man)	Human genes and genetic disorders
DrugBank	Drug information and interactions
TAIR (The Arabidopsis Information Resource)	Information on Arabidopsis thaliana
FlyBase	Genetic and molecular data for Drosophila
WormBase	Information on Caenorhabditis elegans
MINT (Molecular INTeraction Database)	Molecular interactions

This table includes a selection of databases from various domains within biology, including genomics, proteomics, structural biology, genetics, and more. Many other specialised databases are available for specific organisms, diseases, or research areas.

VI. UMBILICAL CORD

The umbilical cord is a crucial structure during pregnancy that connects the developing foetus to the placenta, which in turn is attached to the mother's uterus. It's a vital organ that connects a developing foetus to the placenta in the womb. It's a tube that contains three blood vessels: one vein and two arteries. The vein carries oxygen and nutrients from the placenta to the baby, while the arteries carry waste products from the baby back to the placenta.

It contains two umbilical arteries and one umbilical vein. The umbilical arteries carry deoxygenated blood from the foetus to the placenta, while the umbilical vein carries oxygenated blood from the placenta to the foetus. The length of the umbilical cord is approximately equal to the crown-rump length of the foetus throughout pregnancy.

VII. CORD BLOOD

Cord blood, also known as umbilical cord blood, is the blood that remains in the umbilical cord and placenta following the birth of a baby. It is a valuable source of stem cells, particularly hematopoietic stem cells, which are responsible for forming blood cells in the body. Cord blood is rich in these stem cells, and it can be collected and stored for various medical purposes. Here are some important aspects of cord blood:

- **Collection :** Cord blood is collected immediately after a baby's birth. The process is safe, painless, and non-invasive. After clamping and cutting the umbilical cord, the blood is collected from the placental side of the cord. It is then placed in a sterile bag for transport and processing.
- **Stem Cells :** Cord blood contains hematopoietic stem cells, which have the potential to develop into various types of blood cells, including red blood cells, white blood cells, and platelets. These stem cells are important for treating a variety of medical conditions.
- **Medical Uses :** Cord blood can be used in medical treatments, particularly in stem cell transplantation. Hematopoietic stem cell transplants from cord blood are used to treat certain blood disorders, such as leukaemia, lymphoma, and anaemia. These stem cells can also be used for some non-blood-related conditions, including immune system disorders.
- **Compatibility :** One of the advantages of cord blood is that it is often easier to find a suitable match for a transplant, as there is less strict matching required compared to bone marrow or

peripheral blood stem cell transplants. This can be especially important in cases where a matching donor within the family is not available.

- **Research :** Cord blood is also valuable for scientific research. It can be used to study various aspects of stem cell biology, genetics, and regenerative medicine.

VIII. CORD BLOOD BANKING

Cord blood banking, also known as umbilical cord blood banking, is the process of collecting and storing blood from the umbilical cord of a newborn baby. This blood, which is rich in hematopoietic stem cells, can be used in medical treatments and research. Cord blood can be stored for future use through cord blood banking. There are two main types of cord blood banking: public cord blood banking and private cord blood banking.

❖ ***Public Cord Blood Banking***:
  - In a public cord blood bank, parents voluntarily donate their baby's cord blood to a publicly accessible and nonprofit repository.
  - The donated cord blood is tested, processed, and stored for potential use by anyone in need of a compatible stem cell transplant. It is made available to patients with specific medical conditions, such as leukaemia, lymphoma, and other blood disorders, who require a stem cell transplant but do not have a suitable match within their family.

❖ ***Private Cord Blood Banking:***
  - In private cord blood banking, parents pay to have their baby's cord blood collected, processed, and stored for the exclusive use of their family.
  - This allows the family to have a readily available source of stem cells if a family member requires a stem cell transplant in the future. It is often used as a form of biological insurance.

Here are some key points to consider regarding cord blood banking:

- **Collection :** The process of collecting cord blood is non-invasive and painless. It occurs immediately after the baby is born and the umbilical cord is clamped and cut. The blood is collected from the remaining portion of the cord, typically into a sterile bag.
- **Processing :** Once collected, the cord blood is processed to isolate and freeze the stem cells. This process involves removing red blood cells and other components to preserve the valuable stem cells.
- **Storage :** Cord blood can be stored for an extended period, often decades, in specialised cryogenic storage tanks. The stem cells are preserved at very low temperatures to maintain their viability.
- **Cost :** Private cord blood banking involves a fee for collection, processing, and ongoing storage. Public cord blood banking is generally free for donors.
- **Availability :** The main advantage of private cord blood banking is that the cord blood is reserved exclusively for the donor family's use. However, the likelihood of a family member needing their own cord blood is relatively low, and it may not be a cost-effective choice for everyone.
- **Donation :** Public cord blood banks rely on voluntary donations, and the cord blood is made available to a wider range of patients in need. It can potentially save lives by providing matches for individuals with rare or difficult-to-match tissue types.

The decision to collect and bank cord blood is a personal one, and it depends on factors such as family medical history and the potential need for stem cell transplantation. Public donation of cord blood can also be a valuable contribution to the community and may help save lives. It's important to discuss your options with your healthcare provider and make an informed decision based on your individual circumstances.

Some of the major examples of cord blood databases are the following:

- **WMDA's Search & Match Service :** A global database of cord blood units available for transplantation
- **WMDA Share :** Provides address details of cord blood banks that list their cord blood units for public use
- **FDA's website :** Has a searchable database that includes information on registered cord blood banks
- **ClinicalTrials.gov :** An online registry of medical research worldwide that is a searchable database
- **Stem Cell Therapeutic Outcomes Database (SCTOD) :** A scientific database that collects information on patients who received stem cell
- **GII Research :** A global cord blood industry database with over 450 cord blood banks worldwide
- **Bio Information :** A global database with hundreds of active cord blood banks across 98 different countries

IX. CONCLUSION

In conclusion, biological databases are fundamental tools in the field of biology and life sciences, serving as centralised repositories of diverse and structured biological data. They play a crucial role in facilitating research, enabling data-driven discoveries, and advancing our understanding of various aspects of life on Earth. cord blood databases have the potential to save lives and offer opportunities for medical treatments and research. Whether stored in a private bank or donated to a public bank, cord blood can make a significant difference in the health and well-being of individuals and the broader community. The decision to bank cord blood is a personal one, influenced by a balance of medical, financial, and ethical considerations.

X. REFERENCES

1. Y. Guo, F. -L. Chung, G. Li and L. Zhang, "Multi-Label Bioinformatics Data Classification With Ensemble Embedded Feature Selection," in IEEE Access, vol. 7, pp. 103863-103875, 2019, doi: 10.1109/ACCESS.2019.2931035.

2. F. Gurcan and N. E. Cagiltay, "Exploratory Analysis of Topic Interests and Their Evolution in Bioinformatics Research Using Semantic Text Mining and Probabilistic Topic Modeling," in IEEE Access, vol. 10, pp. 31480-31493, 2022, doi: 10.1109/ACCESS.2022.3160795.

3. A. Xiaohui and C. Guoping, "Characterization of adherent nonhematopoietic cells derived from human umbilical cord blood," in Tsinghua Science and Technology, vol. 8, no. 4, pp. 412-416, Aug. 2003.

4. Dr.C K Gomathy and et al, Machine Learning-Based Clinical Decision Support System, International Journal of Scientific Research in Engineering and Management (IJSREM) Volume: 06 Issue: 10 | October - 2022 Impact Factor: 7.185 ISSN: 2582-3930

5. .Dr.C K Gomathy et al,Web Service Composition In A Digitalized Health Care Environment For Effective Communications, Published by International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 5 Issue 4, April 2016, ISSN: 2278 – 1323.

6. Vishnupriya C K and et al,Dimensional and Morphologic Variations of palatal Rugae-a hospital based study among Chennai populations, International Journal Of Science Research, ISSN No: 2277-8179 Volume 7, Issue 7,P.No-19-20,July ’2018

7. Dr.C K Gomathy et al, Machine Learning-Based Clinical Decision Support System, International Journal of Scientific Research in Engineering and Management (IJSREM) Volume: 06 Issue: 10 | October - 2022 Impact Factor: 7.185 ISSN: 2582-3930

8. Dr.C K Gomathy et al,A Review On IOT Based Covid-19 Patient Health Monitor In Quarantine, International Research Journal of Engineering and Technology (IRJET),e-ISSN: 2395-0056 Volume: 08 Issue: 09 | Sep 2021 www.irjet.net p-ISSN: 2395-0072`

9. Dr.C K Gomathy, et al, A Medical Information Security Using Cryptosystem For Wireless Sensor Networks, International Journal Of Contemporary Research In Computer Science And Technology (Ijcrcst) E-Issn: 2395-5325 Volume3, Issue 4, P.No-1-5,April ’2017

10. Dr.C K Gomathy and et al, The Parkinson’s Disease Detection Using Machine Learning Techniques, International Research Journal of Engineering and Technology (IRJET), Volume: 08 Issue: 10 | Oct 2021, e-ISSN: 2395-0056, p-ISSN: 2395-0072.

11. Dr.C K Gomathy, V Geetha , T.Jayanthi, M.Bhargavi, P.Sai Haritha: A Medical Information Security Using Cryptosystem For Wireless Sensor Networks, International Journal Of Contemporary Research In Computer Science And Technology (Ijcrcst) E-Issn: 2395-5325 Volume3, Issue 4, P.No-1-5,April '2017

12. Dr.C K Gomathy and et.al, The Smart Stick Assistant For Visually Challenged People Using Ai Image Recognition, International Research Journal of Engineering and Technology (IRJET), Volume: 08 Issue: 9 | Sep 2021, e-ISSN: 2395-0056, p-ISSN: 2395-0072.

13. Dr. C.K. Gomathy , Dr. V.Geetha ,G.S.V.P.Praneetha , M.Sahithi sucharitha. (2022). Medicine Identification Using Open Cv. Journal of Pharmaceutical Negative Results, 3718–3723. https://doi.org/10.47750/pnr.2022.13.S09.457

14. Dr. V.Geetha ,Dr. C.K. Gomathy , Kommuru Keerthi , Nallamsetty Pavithra. (2022). Diagnostic Approach To Anemia In Adults Using Machine Learning. Journal of Pharmaceutical Negative Results, 3713–3717. https://doi.org/10.47750/pnr.2022.13.S09.456

AUTHOR'S PROFILE:

L SAI PURNA SRINIDHI PERUMALLA, UG Scholar, B.E, Computer Science and Engineering in Sri Chandrasekharendra Saraswathi Viswa Maha Vidyalaya (SCSVMV Deemed to be University). Her area of interests are Cloud computing Web development, App development, Artificial intelligence, Bioinformatics (Email id: srinidhiperumalla22248@gmail.com).

RAJASRIVATSAN SRINIVASAN, UG Scholar, B.E, Computer Science and Engineering in Sri Chandrasekharendra Saraswathi Viswa Maha Vidyalaya (SCSVMV Deemed to be University). His area of interests are Artificial intelligence, Data science, Machine learning, Web development, Bioinformatics (Email id: srivatsansridevi320@gmail.com).

 Dr. C.K Gomathy, M.E (CSE), M.B.A (IT and Management), Ph.D. (CSE), Assistant Professor in CSE, SCSVMV Deemed to be University, Her Area of Interest lies in Software Engineering, Web Service, Machine Learning, Medical informatics,IOT and Bioinformatics.