B.S. Abdur Rahman ™
Crescent
Institute of Science & Technology
Deemed to be University u/s 3 of the UGC Act, 1956

CRESCENT
BIOINFORMATICS
ASSOCIATION

School of Life Sciences
LEADING LIFE THROUGH SCIENCE

VOL 2

2023

DECODING THE GENETIC POESY

BIOINFOLUCENT

A BIOINFORMATICS MAGAZINE

FOREWORD

President,CBA

Bioinformatics serves as a platform for exchanging information, knowledge, and expertise in computational molecular biology and genome bioinformatics, with a focus on pioneering algorithms and databases that drive advancements in bioinformatics and biomedical research. The Crescent Bioinformatics Association (CBA) continuously strives to harness the full potential of this field, relying on collaborative teamwork.

As a student-initiated association, their enthusiasm and efforts has always remained their hallmark.Over the past two years, Crescent Bioinformatics Association has utilized their online presence to the fullest, engaging in extensive interactions and successfully organizing a range of Intra college events and workshops. Their determination has enabled them to overcome challenges and steadily progress towards greater achievements.

Starting with a small group of students, the club's energy and dedication have played a significant role in bringing "BioInfoLucent" to life. This edition delves into the foundations of bioinformatics through insightful articles, facts, and news, infused with creative elements such as memes and games

I extend my heartfelt appreciation to the editorial team and contributors for their tremendous efforts in bringing this magazine to fruition. It is with immense pleasure that I extend my best wishes for their continued success and future endeavors.

Dr. S. Hemalatha
Professor & Dean
School of Life Sciences, BSACIST

Faculty Cordinator, CBA

In this second edition of "BioInfoLucent" magazine, we proudly present the continued efforts of the Crescent Bioinformatics Association in fostering diverse disciplines of bioinformatics among passionate biotechnology students.

Recent times have shown bioinformatics, especially when integrated with AI, has emerged as a prominent subject in many countries, establishing numerous partnerships with organizations to elevate its significance.

In conjunction with CBA, we aim to provide opportunities for aspiring students to interact, share knowledge, and overcome challenges. Our goal is to inspire not only students at B. S. Abdur Rahman Crescent Institute of Science and Technology, but also those from around the world.

I express my heartfelt gratitude for the unwavering support that has been provided by our Dean, Dr. S. Hemalatha Ma'am, in bringing forth the second edition of our magazine, "BioInfoLucent." This edition showcases enlightening yet captivating content that can be beneficial to any curious individual. I want to express my appreciation and best wishes to the editorial team and all the contributors who played a vital role in creating this remarkable magazine!

Dr. A Baskaran
Assistant Professor
School of Life Sciences, BSACIST

Contents

Editors NOTE

BioinfoLucent lives up to its name by shedding light on a range of concepts in the field of Bioinformatics.Within the pages of this edition, you will discover captivating content created by our dedicated members, who generously volunteered their time to craft this magazine.The magazine is artfully designed, demonstrating how design bridges the gap between information and comprehension.As we worked on this magazine, we grew closer as a team, deepened our understanding of the subject matter, and gained valuable knowledge from one another. This achievement is nothing short of a testament to our ongoing growth and development. We sincerely hope that you derive enjoyment from each and every page of this magazine.Happy reading and happy learning !

Lina Fatima M
Co-President , CBA

Thamarai G.R
Co-President , CBA

Welcome to the second edition of BioinfoLucent , where we embrace the fascinating nexus between biology, computer science, and data analysis.We are ecstatic to present to you yet another enthralling compilation of articles, insights, and news alongside a dash of fun.From machine learning to artificial intelligence, we have meticulously endeavoured to present an accurate grasp of all the different facets of bioinformatics.This edition serves as an affirmation to our dedication to encouraging student collaboration and we take great satisfaction in it. As always, we would like to express our gratitude to all our mentors, editors and designers who have devoted their time and expertise to enrich this edition of our magazine.We hope you find this edition of BioinfoLucent enlightening and thought-provoking.

CLUB'S ANNUAL REPORT

The Crescent Bioinformatics Association (CBA) is a dynamic community of students passionate about biology and the various applications of it in the bioinformatics field. This annual report highlights the club's activities, achievements, and contributions to the field of bioinformatics during the academic year 2022-2023.

In 2022, CBA played a vital role in fostering knowledge sharing, skill development, and fellowship among its members. We organized various events, workshops and meetings enabling members to enhance their expertise in bioinformatics and excel in their respective theory and practical knowledge (Curricular and extra curricular).

Cassandra Rifflin C.R
Technical Head,CBA

Membership Overview :

The club witnessed significant growth in membership throughout 2022. Starting with less than 50 members, we ended the year with over 190 active participants. This increase was a result of our efficient marketing team which was led by the co-presidents and the Public relations Team.

Social media Presence:

Throughout the year, our club remained highly active in utilizing social media platforms as a powerful tool for engagement and outreach. By leveraging the reach and influence of social media, we effectively connected with our target audience and shared valuable content related to our club's mission and activities. We consistently published a series of informative and engaging blogs, covering a wide range of topics within our field of interest. These blogs served as a platform for knowledge sharing, sparking discussions, and building a community of like-minded individuals. Through our active presence on social media and the regular publication of insightful blogs, we successfully disseminated information, fostered connections, and established ourselves as a trusted resource within our community.

Activities and Programs:

The Bioinformatics Club hosted a diverse range of activities and programs throughout 2022, aimed at enhancing members' skills and fostering collaboration.

HANDS-ON TRAINING ON EXTRACTION & PHYTOCHEMICAL ANALYSIS OF NATURAL COMPOUNDS-IN VIVO & INSILICO APPROACH:

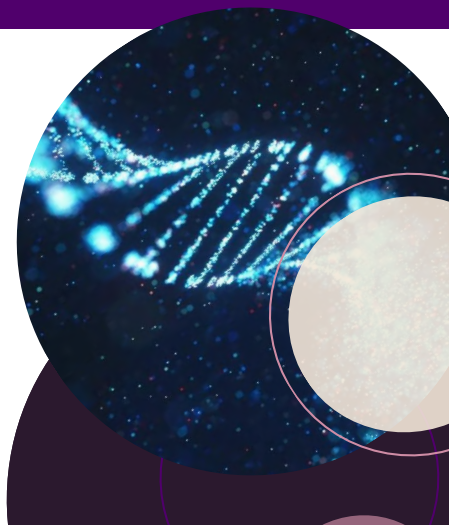
From September 21 to September 23,2022 our first batch of hands-on training was organised exclusively for the students of SLS. This workshop was successfully conducted as an interactive 3 day training programme where many students benefited through resourceful teachings and hands-on experience.

Day 1 of the workshop began with the inauguration where the Dean School of Life Sciences, Dr S.Hemalatha motivated the participants with her inspiring words. This was followed by an informative guest lecture on the uses and applications of natural products by Dr. Ashok Kumar, Associate Professor,SLS. The first day put a spotlight on how various phytochemicals are extracted and purified from plant sources, we also went on to teach the participants how to find the type of phytochemicals present in the particular species of plant using various research papers.

Day 2 commenced with Ms Shabnam Ameenudeen, a Research Scholar from SLS giving a lecture on the role of computational approach in drug designing and development. The day further proceeded with the students' learning how to use AutoDock and various softwares to analyse a compound's binding affinity to target molecules, this in-silico approach is expected to help the students in their future research.

Day 3 kept the students on their feet! Thanks to Dr. R. Rajesh Kannan , Scientist E and Professor from Sathyabama University, the guest speaker who delivered a lecture on zebrafish as an animal model and giving us a glimpse on his research on Zebrafish. The day proceeded with Organ harvesting of Zebrafish where each participant was provided with a zebrafish and the handling techniques were taught.

The facilitators , faculty co-ordinators and the office bearers played a significant role in the smooth conduct of the workshop and its end outcome. The workshop concluded with the Valedictory event where they were felicitated with certificates and feedbacks were given by the participants. In addition to this the office bearers were awarded their badges by the Dr S Hemalatha, Dean SLS.



The workshop was a massive success as it was evident after reviewing the feedback of the participants which were all positive. The faculty coordinators and the dean were all content as the workshop was highly beneficial to the students and acted as a foundation stone to their research careers.

BIOINFOTHON

A two-day event was organised by our club on the 17th and 18th of April, which was an engaging and comprehensive workshop that encompassed various activities and opportunities for learning and showcasing talent in the field of bioinformatics.

The first event of the day, Extempore,

encouraged spontaneity and quick thinking, allowing participants to demonstrate their ability to communicate effectively on bioinformatics-related topics.

Secondly, a stimulating quiz

that tested participants' knowledge and problem-solving skills was held in two levels with increasingly tougher questions based on bioinformatics and its applications, challenging the students to think critically in a competitive yet supportive environment. Other intriguing yet fun online events like Designing for the 'Gram and meme creation were also conducted and were judged based on creativity and innovation. Another interesting activity included "Find the Tool". Which involved a great learning experience for all the students who participated. Since it was a bioinformatic tool based search it was informative and enthralling at the same time.

On the second day,

we provided the students a platform to present their ideas and project ideas relating to bioinformatics, through a captivating poster presentation session. Moreover, we were honored to host a guest lecture by a distinguished expert in the field, Dr. Saleem Mohammed the founder & CEO of Xcode Life. He gave valuable insights and inspiration to all the attendees. He was relatable to almost all the students because he's a proud alumni of our own University!

The Valedictory was conducted at the end, in which the winners were felicitated with momentos and certificates by the Dean Dr S Hemalatha. Bioinfothon served as a comprehensive and multifaceted event, fostering collaboration, knowledge sharing, and personal growth for all involved.

FUTURE PLANS AND GOALS

Looking ahead to the academic year 2023-24, the Bioinformatics Club has set the following goals:

Establish a mentorship program connecting experienced members with newcomers to support their learning and research endeavors.

Expand the club's membership to make most of the SLS students join and participate in all future endeavours of CBA.

Establish a mentorship program connecting experienced members with newcomers to support their learning and research endeavors.

Plan, Execute and publish the next edition of BioinfoLucent!

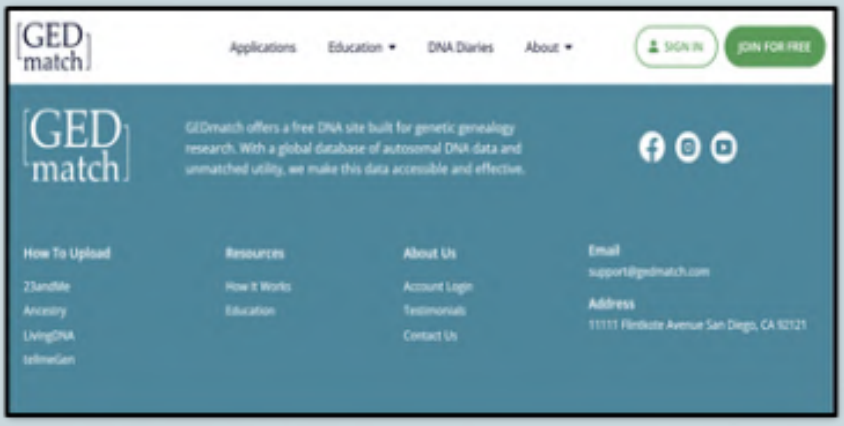
The past year has been a testament to the exceptional dedication, passion, and efforts within our club. We have seen great growth, both in terms of membership and the achievements. As we reflect on our accomplishments, we are excited to continue pushing the boundaries of bioinformatics, even though they're just the beginning! We extend our heartfelt gratitude to Our management, Respected Dean and our mentors and , and we eagerly look forward to another successful year ahead hoping for more progress!

GED MATCH

Kaviya A

1st year M.Tech Biotechnology

IN THIS ARTICLE



What is GED Match?

GEDmatch is an online tool for comparing autosomal DNA test results from different testing labs. For those interested in genetic genealogy, GEDmatch is a free DNA database. It makes this data accessible and useful with a global library of autosomal DNA data and incomparable utility. GEDmatch offers additional tools for comparing DNA test results with the largest possible worldwide population.

This website also offers some of the free tools listed below:

ONE-TO-MANY DNA COMPARISON

This tool helps compare one's genetic profile with those of other members of GEDmatch

How GED match works ?

Using GEDmatch, one can find potential relatives by uploading the results of their private autosomal DNA. Their RAW DNA data files can be uploaded using 23andMe, Ancestry, LivingDNA, or tellmeGen.

ONE-TO-ONE AUTOSOMAL DNA COMPARISON

This tool helps to narrow down the results of comparison based on the autosomal DNA of two individuals

ADMIXTURE

This tool helps identify one's biogeographical ancestry. It also shows the proportion of one's DNA based on a particular location.

Steps involved in GEDmatch

TEST

First, the user has to download their DNA data files from 23 andMe, Ancestry, or other tools.

LOGIN

The member then has to register and login to GEDmatch

UPLOAD

Then, they have to upload their DNA data file to GEDmatch for processing.

INFORMATION

They have to provide other information like mitochondrial haplogroups, Y haplogroups, and other information.

CONCENT

The members have to fill out their consent using the privacy option given on the website below. Then, they have to upload and press the finish button

COMPARE

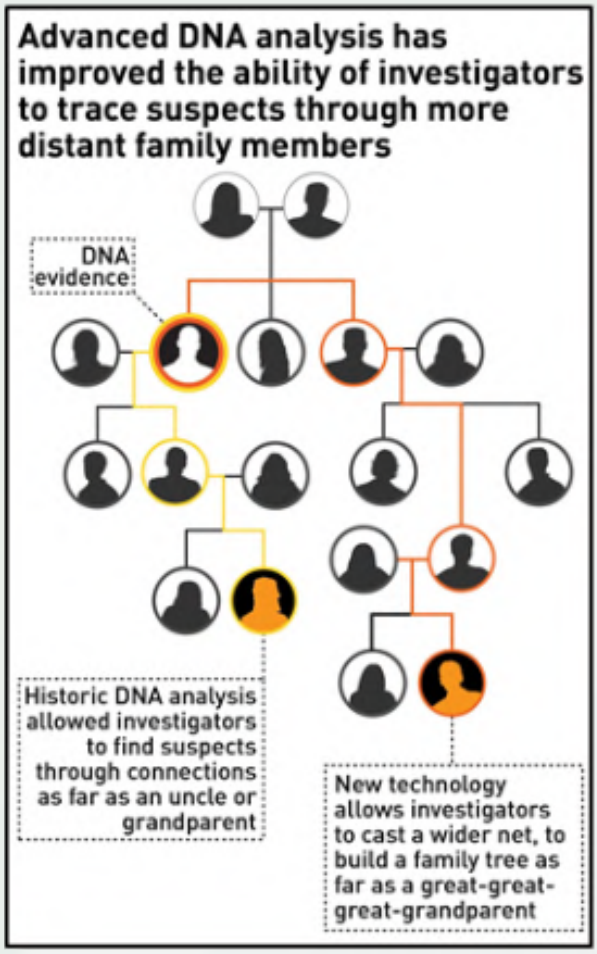
Now, they can explore matches, compare their DNA with others, and use other DNA tools.

EXPLORE

Finally, they will be able to find their relatives, build their family tree, and learn more about their own identities. This entire process may take up to 24 hours.

How are cases solved using GEDmatch?

Law enforcement organizations, like the FBI, were able to upload DNA profiles from suspects, arrestees, and crime sites to Family Tree DNA to help solve cold cases. GEDmatch may be particularly useful in assisting with the capture of interpreted suspects who might have previously avoided law enforcement.



California law enforcement used the DNA collected from the 1978 Golden State Killer case and uploaded it to the 23andMe and Ancestry.com websites. They were able to find the matching DNA against the DNA profile they had uploaded.The lead investigators used GEDmatch to construct a family tree using the DNA profile. From this tree, they were able to interpret the plausible suspect as Joseph James DeAngelo.

So they placed him under surveillance and collected his current DNA from the items he had discarded. Using this sample, they were able to match his DNA with the DNA collected in 1978, and then DeAngelo was arrested as the suspect in the Golden State Killer case.

Bioinformatics in Skincare and Cosmetics

Introduction

Cosmetics are a multi-billion-dollar industry that is constantly evolving to meet the changing needs and preferences of consumers. One of the areas that has seen significant growth in recent years is the application of bioinformatics in cosmetics. Bioinformatics is a branch of science that analyzes and interprets biological data by combining computer science, biology, and statistics. It has revolutionized many areas of science, including skincare and cosmetics. The use of bioinformatics in the development of skincare and cosmetic products has led to a more personalized approach that can address the individual needs of each person's skin. In this article, we will discuss the applications of bioinformatics in skincare and cosmetics, including the use of genomics, proteomics, and metabolomics.

Proteomics:

Proteomics is the study of an organism's proteins, which are the building blocks of cells and tissues. The application of proteomics in skincare and cosmetics involves analyzing the proteins that are present in the skin and how they interact with other molecules.

For example, the use of mass spectrometry can identify the proteins that are present in the skin, and this information can be used to develop products that target specific proteins, such as those that promote collagen production.

Metabolomics:

Metabolomics is the study of an organism's metabolites, which are the small molecules that are involved in cellular metabolism. The application of metabolomics in skincare and cosmetics involves analyzing the metabolites that are present in the skin and how they contribute to skin health. For example, the use of metabolomics can identify the metabolites that are involved in skin pigmentation, and this information can be used to develop products that target specific metabolites, such as those that reduce melanin production.

Identification of Active Ingredients:

One of the key areas where bioinformatics is being applied in cosmetics is the identification of active ingredients. Active ingredients are the compounds in cosmetic products that provide the intended therapeutic or cosmetic effect.The identification of active ingredients traditionally involved a trial-and-error approach, but with the use of bioinformatics, it is now possible to screen thousands of compounds and predict their efficacy based on their chemical structure and known biological activity.

Personalized Skincare and Cosmetics:

Another significant area where bioinformatics is being applied in skincare and cosmetics is the development of personalized products.

By analyzing an individual's genetic material, proteins, and metabolites, it is possible to identify the specific needs of their skin and develop products that address those needs. This personalized approach can lead to more effective cosmetic products that are tailored to the individual's needs.

Safety Evaluation:

The safety of cosmetic products is of paramount importance, and bioinformatics can be used to predict the toxicity of cosmetic ingredients based on their chemical structure and known toxicity profiles. This can help to identify potentially harmful ingredients before they are used in cosmetic products, ensuring the safety of consumers.

Regulatory Compliance:

Regulatory compliance is another area where bioinformatics is being applied in cosmetics. Regulatory agencies around the world require cosmetic companies to provide safety data on the ingredients used in their products. Bioinformatics can be used to generate this safety data by predicting the toxicity of cosmetic ingredients, reducing the need for animal testing, and providing a more ethical and cost-effective approach to safety evaluation.

Future Directions:

The application of bioinformatics in skincare and cosmetics is still in its early stages, and there is tremendous potential for growth and development.

With the continued development of new technologies, such as artificial intelligence and machine learning, it is possible to further enhance the analysis of biological data and improve the accuracy of personalized product development.Additionally, the use of nanotechnology in cosmetic products can improve the delivery of active ingredients to the skin, further enhancing their efficacy.

The use of genomics, proteomics, and metabolomics has led to the identification of specific genes, proteins, and metabolites that are involved in skin health, and this information can be used to develop products that target those molecules. With further development and advancements in technology, the approach of bioinformatics in skincare has the potential to improve the efficacy and safety of cosmetic products and enhance the overall health of our skin. The application of bioinformatics has the potential to revolutionize the skincare and cosmetics industry further, leading to even more effective products in the future.

Conclusion:

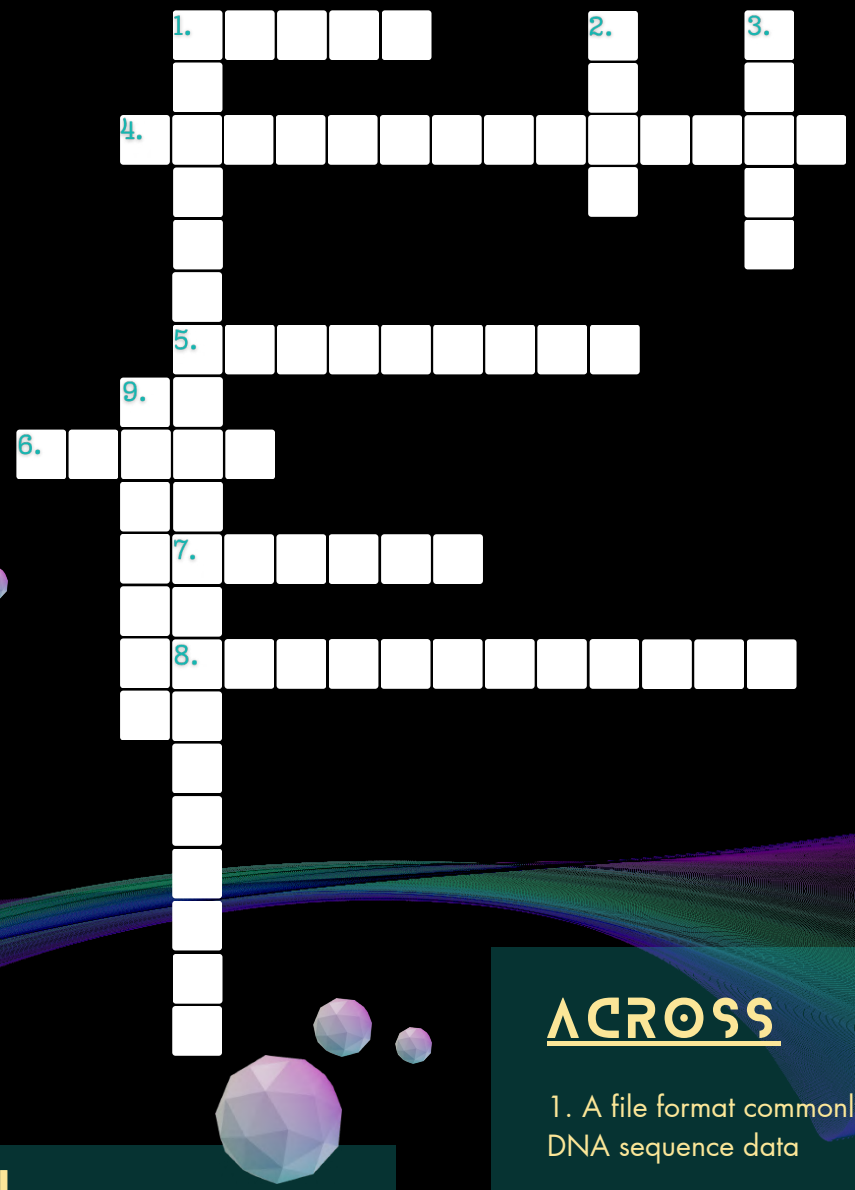
In conclusion, the application of bioinformatics has revolutionized the skincare and cosmetic industries by providing a more personalized approach to product development, improving safety evaluation, and reducing the need for animal testing.

Prawin M

1st Year M.Tech Biotechnology

VERBAL VORTEX

S. Sathiya, 2nd year B.Sc. Biotechnology



DOWN

1. A user can get files from another computer on the internet using.

2. A software that provides many tools for phylogenetic analysis.

3. A single piece of information in a database is called

9. The first bioinformatics database was created by

ACROSS

1. A file format commonly used to store DNA sequence data

4. The alignment procedure used to align the entire sequence

5. The step-wise method for solving problems in computer science is called

6. The program used for local similarity search

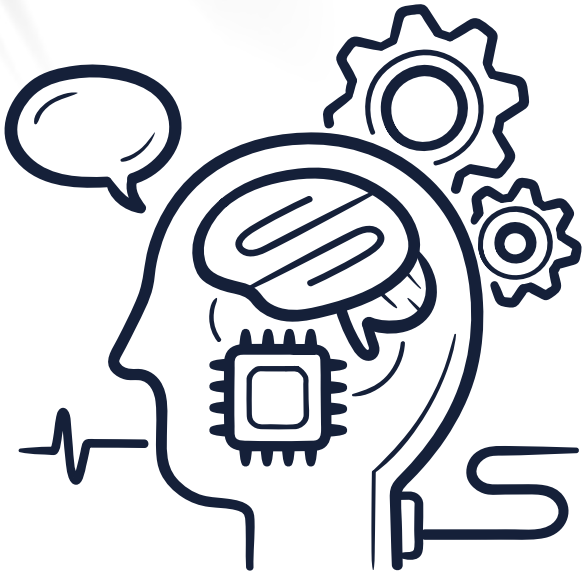
7. What is a molecular biology database and retrieval system developed by NCBI?

8. Name of the method used to study the evolutionary history of genes and genomes

Solutions on page 29

ALGORITHMS TO FIX

Artificial brains to detect and treat malfunctioned human brains !



Modern Evolution is not due to mutations but to artificial brains and algorithms. Technology growth took over within a nanosecond, and everywhere AI roams more than house flies: artificial assistants for business, for households to complete homework, and even for personal care. Why not for Healthcare? A decade ago, the bioinformatics field was equipped with specific tools and programmes to store, analyse, and retrieve data. The next level of bioinformatics provided a path for customised treatment called precision medicine.

As of now, to update the field, here comes the most attractive yet smart technology where you no longer spend years and energy to detect and diagnose the disease and even cure it. Deep learning and Artificial Neural Networks are here to specify the sample, detect the disorders, and even suggest treatments. Imagine that the human brain needs help from an artificial brain in the sense that malfunctioning human neurological networks need to be fixed with the help of artificial neural networks. At the beginning of this decade, scientists worked in this field and developed tools to detect Alzheimer's, Parkinson's, Ataxia, and Huntington's disorders.

Deep learning, which is a type of machine learning that uses artificial neural networks with multiple layers, has shown promise in the field of neurological disorders. These algorithms are particularly well-suited to handling large, complex datasets, such as those generated by brain imaging and genetic studies. One area where deep learning has been used is in the diagnosis of neurological disorders. For example, deep learning algorithms have been trained to analyse brain imaging data to identify patterns of abnormal activity associated with conditions such as Alzheimer's disease, Parkinson's disease, and multiple sclerosis. These algorithms may be able to detect subtle changes in brain function that are not visible to the human eye, allowing for an earlier and more accurate diagnosis.

" The neurological space is in desperate need of new drugs, but clinical successes have been extremely low. There is a need for target-based medicines, which can be succeeded by AI tools."

The application of deep learning to the early detection and automated classification of Alzheimer's disease (AD) has recently gained considerable attention as rapid progress in neuroimaging techniques has generated large-scale multimodal neuroimaging data. One of the challenges in diagnosing Alzheimer's is that the symptoms can be subtle and may overlap with other conditions. Deep learning models can analyse a wide range of data sources to help identify patterns and markers that may be indicative of Alzheimer's disease.

Convolutional neural networks (CNNs) are employed for brain imaging analysis. Researchers have used CNNs to analyse brain imaging data, such as MRI or PET scans, to identify patterns of brain atrophy or abnormal activity that are indicative of Alzheimer's disease.

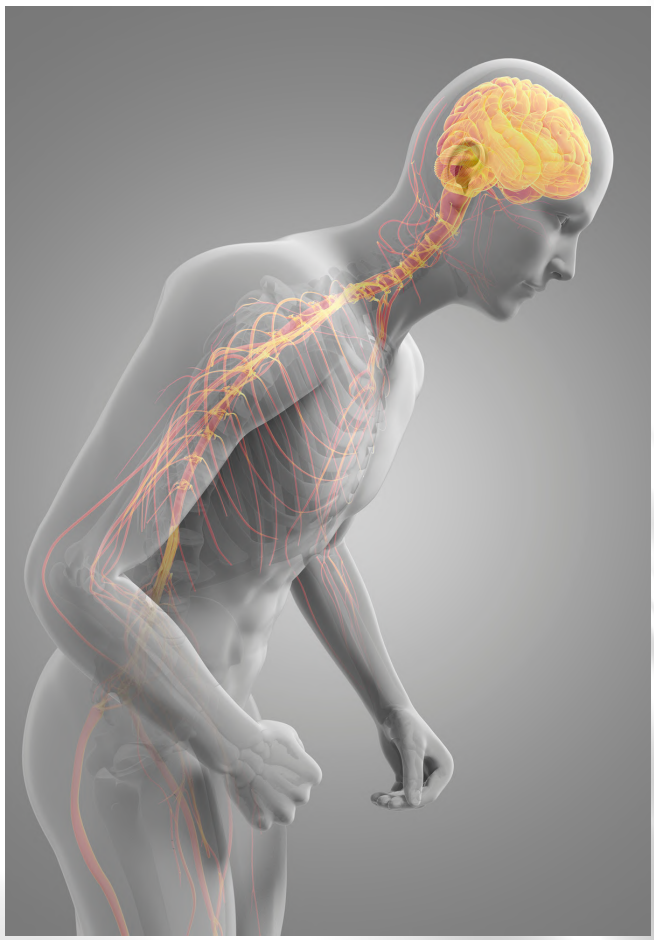
Recurrent neural networks (RNNs) for natural language processing are well-suited for analysing sequential data, such as spoken or written language. Researchers have used RNNs to analyse changes in speech patterns or word usage that may be indicative of cognitive decline. It's like our parents and best friends. When we stay silent, they will find there is a problem, and RNNs do the same for Alzheimer's patients.

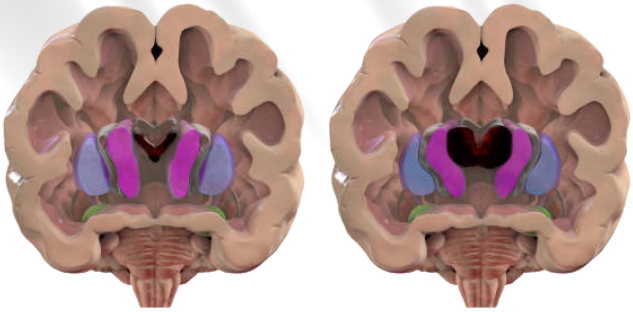
Ataxia is a neurological disorder characterized by a lack of coordination of voluntary movements, and it can be caused by a variety of underlying conditions. Deep learning methods have been used in ataxia research to aid in the diagnosis, classification, and prediction of disease progression.

ARTIFICIAL NEURAL NETWORKS (ANNS)

have been used in ataxia diagnosis with promising results. ANNs can be used to predict the efficacy of a drug for treating ataxia. This can be done by training an ANN on preclinical data, such as animal models, and using the ANN to predict how the drug will perform in humans. For example, a study used ANNs to predict the efficacy of a drug for treating ataxia based on preclinical data. To analyse large datasets of drug properties and predict which drugs may be effective in treating ataxia. Uploading the approved drugs and using the ANN to predict which drugs may have efficacy in ataxia.

Huntington's disease (HD) is caused by a genetic mutation in the HTT gene, but there are also other genes that can contribute to disease onset and progression. Machine learning techniques are designed to extract data about ongoing cognitive changes present in these patterns of coactivation, which can be very subtle. This involved building a high-resolution network of the brain from resting-state activity, where functional connections are defined by high correlations between network nodes. Then, maps are built that rank the centrality or importance of each node in each individual brain. These maps provide a reliable neurological signature, even when remeasured one or two years later.





Using interpretable AI models, we can identify that these patterns can indicate how fast cognitive changes are beginning to take place. By applying a simple logistic regression model modified to select from the whole brain only the most relevant nodes. These models could reliably identify extreme subgroups of patients that were either stable or showing rapid decline.

Learning about the utmost and greatest organ of our body, the brain, led to the creation of an artificial brain, which in turn serves the master who designed it! The revolutionized world of AI has taken off to care for human brains using deep learning tools.

Diagnosis, prediction of disease progression, classification of subtypes, and identification of precise medications with the help of trained artificial neural networks with specified data ANNs are only as good as the quality and quantity of the data they are trained on, and further research is needed to validate their use in clinical settings. Even though it is best in all senses, it does have some limitations, but it never fails to serve human healthcare!

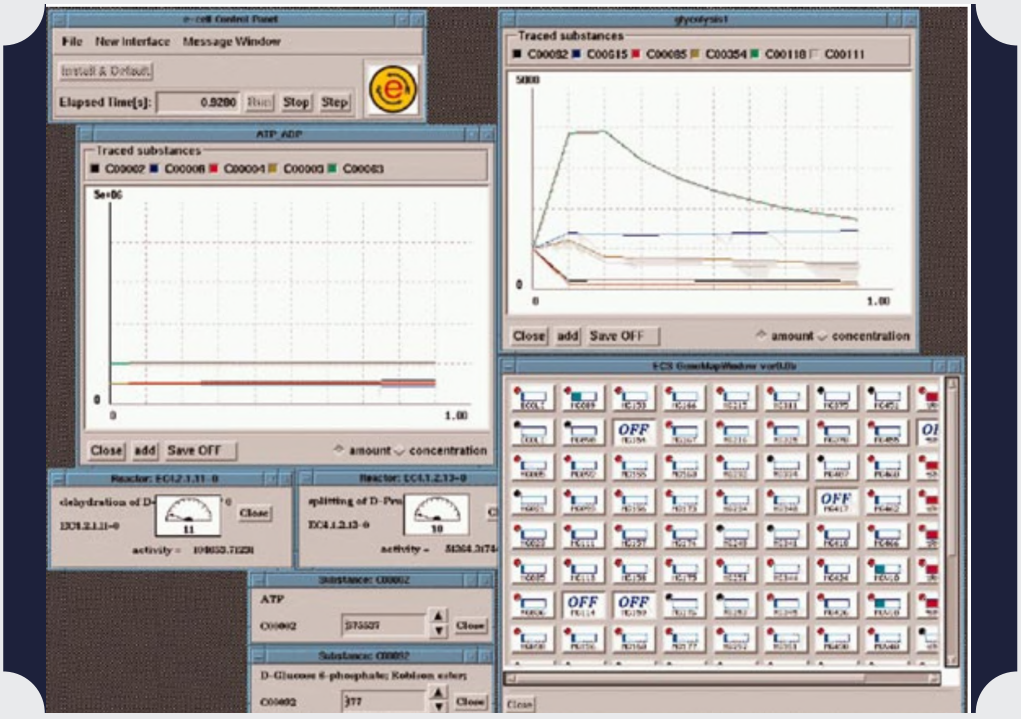
HARINE A

3rd year B.Tech Biotechnology

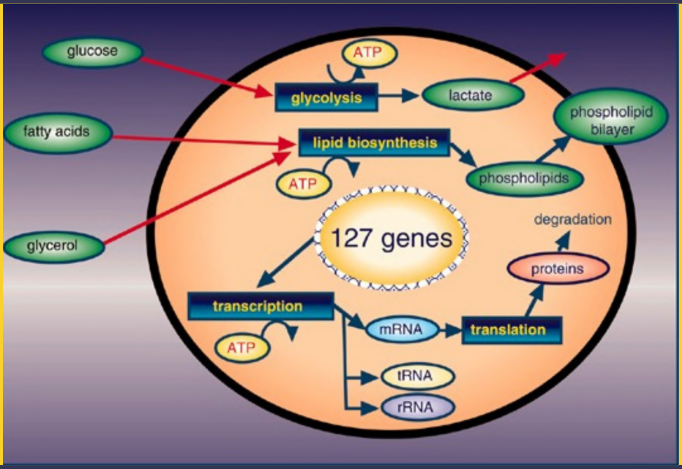
E-CELL

SOFTWARE ENVIRONMENT FOR WHOLE-CELL SIMULATION

E-Cell System is a software platform for modeling, simulation and analysis of complex, heterogeneous and multi-scale systems like the cell.It was launched in the year 1996 by the Laboratory for Bioinformatics at Keio University SFC. Genome sequencing projects and further systematic functional analysis of complete gene sets produced an unprecedented mass of molecular information for a wide range of model organisms. This provided a detailed account of the cell, which were required to build models for simulating intracellular molecular processes to predict the dynamic behavior of living cells. Thus the students were motivated to develop a software environment for building integrative models based on gene sets, and running simulations to conduct experiments in-silico.



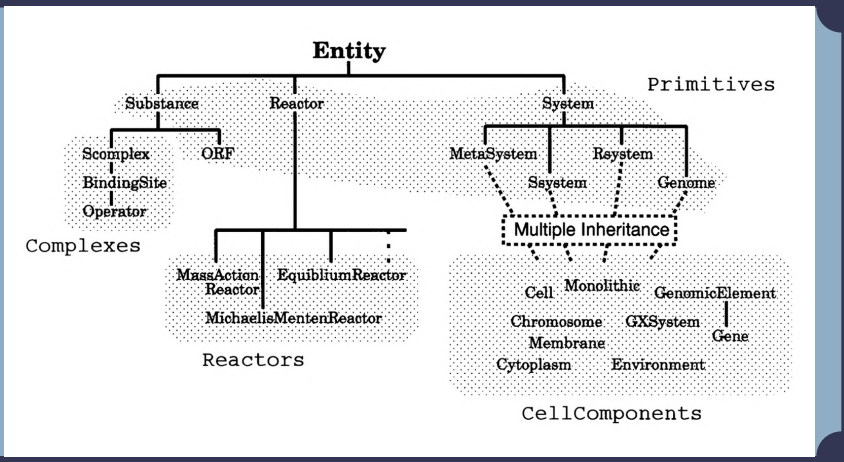
A snapshot of user interfaces of the E-CELL system



Metabolism overview of the model cell. It has pathways for glycolysis and phospholipid biosynthesis, as well as transcription and translation metabolisms.

The E-CELL employs a structured Substance-Reactor model to construct and simulate the cell model. The structured Substance-Reactor model consists of three classes of objects- Substance, Reactor, and System, representing molecular species, reactions and functional/physical compartments, respectively. To model chromosomes and other genetic materials, the system also has sophisticated data structures such as Genome, Genomic Element, and Gene.

Using the E-CELL system, the first virtual cell with 127 genes sufficient for "self-support"was constructed. The gene set was selected from the genome of Mycoplasma genitalium, the organism having the smallest known genome. The set includes genes for transcription, translation, the glycolysis pathway for energy production, membrane transport, and the phospholipid biosynthesis pathway for membrane structure.blts latest version, E-Cell4, accepts multi-algorithms, multi-timescales an multi-spatial-representations as its central feature. E-Cell4 is a free and open-source software licensed under the GNU General Public License version 3.



Ontology structure of the E-CELL system.

There are three fundamental classes: Substance, Reactor and System.

The challenge created by genomics is to understand how all the cellular proteins work collectively as a living system. The development of sufficiently refined cell models which allow predictions of such behavior would complement the experimental efforts now being made systematically to modify and engineer entire genomes.

– KAMALALAKSHMI V
2nd year B.Tech Biotechnology

DRUG DISCOVERY IN BIOINFORMATICS

Priyadharshini M
1st year B.Sc Biotechnology

Historically, drugs were either accidentally discovered, like penicillin or by identifying the active component in conventional treatments. More recently, chemical libraries of synthetic small molecules, natural products, or extracts were screened in intact cells or whole organisms to identify substances that had a desirable therapeutic effect, in a process known as classical pharmacology. After sequencing, the human genome allows for rapid cloning and synthesis of large quantities of purified proteins.

It has become common practice to use high-throughput screening of large compound libraries against isolated biological targets that are hypothesized to be disease-modifying in a process known as reverse pharmacology. Hits from these screens are then tested in cells and then in animals for efficacy.

Modern drug discovery involves the identification of screening hits in medicinal chemistry and the optimization of those hits to increase affinity, selectivity (to reduce the potential of side effects), efficacy or potency, metabolic stability (to increase the half-life), and oral bioavailability. Once a compound that satisfies all of these requirements has been identified, the process of drug development can continue. If successful, clinical trials will be developed.

Unlocking Medical Mysteries

Modern drug discovery is thus usually a capital-intensive process that involves large investments by pharmaceutical industry corporations as well as national governments (who provide grants and loan grantees). Despite advances in technology and understanding of biological systems, drug discovery is still a lengthy, "expensive, difficult, and inefficient process" with a low rate of new therapeutic discovery.

In 2010, each new molecular entity's research and development cost was about US$1.8 billion. Governments and charitable organizations have attempted to address this issue by providing funding for drug research and development, but there is still a need for more efficient and cost-effective methods. One potential solution is the use of artificial intelligence (AI) and machine learning algorithms to aid in drug discovery.

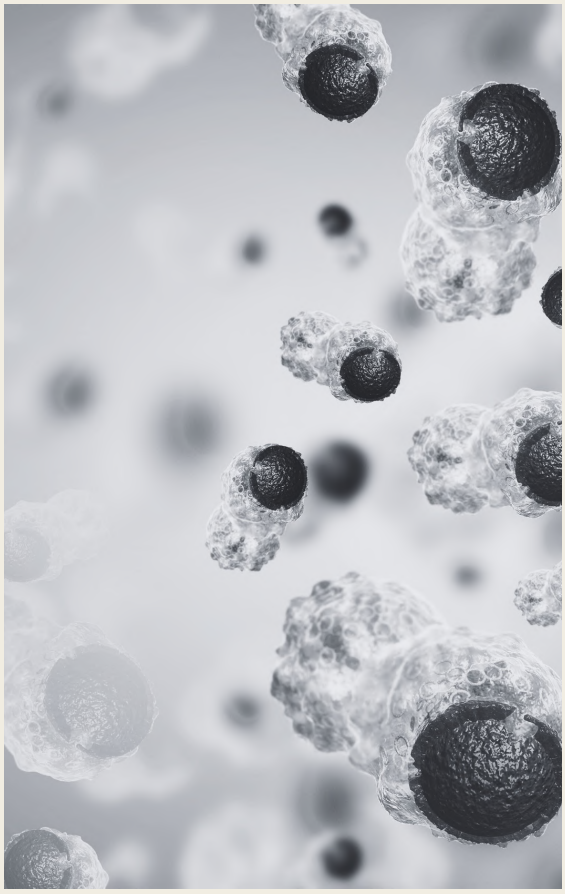
These technologies can analyze vast amounts of data and identify potential drug candidates much faster than traditional methods. Additionally, AI can help optimize clinical trials by identifying patient populations that are most likely to benefit from a particular drug. While there are still challenges to overcome, such as the need for large amounts of high-quality data, the use of AI in drug discovery shows great promise for improving the efficiency and effectiveness of this critical process. They primarily fund basic discovery research in the 21st century, whereas pharmaceutical companies or venture capitalists primarily fund late-stage development. To be allowed to come to market, drugs must undergo several successful phases of clinical trials and pass through a new drug approval process, called the new drug application in the United States.

Discovering drugs that may be commercial or public health successes involves a complex interaction between investors, industry, academia, patent laws, regulatory exclusivity, marketing, and the need to balance secrecy with communication. Meanwhile, for disorders whose rarity means that no large commercial success or public health effect can be expected, the orphan drug funding process ensures that people who experience those disorders can have some hope of pharmacotherapeutic advances.

BIOINFOBUZZ

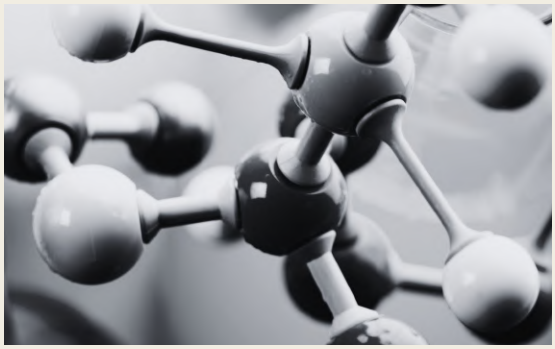
Sameera Siddique, 2nd year B.Sc Biotechnology

BU study reveals crucial signaling pathway that drives head and neck cancers



Despite the various advances in defining the genomic characteristics of head and neck cancers, the malignancies continue to rank among the deadliest cancers with few targeted therapies (currently available). An important challenge in designing effective treatments is the intratumor heterogeneity, it is defined as the presence of multiple subpopulations of cells with distinct genomic and molecular alterations, with some cells inherently more resistant to certain treatments. A new study from researchers at Boston University Chobanian and Avedisian School of Medicine applied advanced bioinformatics and machine learning approaches to the analysis of large multi-omics head and neck cancer datasets and found activation of mTORC1 by b-catenin/CBP as an upstream driver of the malignancy-associated partial epithelial-mesenchymal transition (p-EMT) phenotype.

A super-resolution strategy for mass spectrometry imaging



High-spatial-resolution mass spectrometryimaging (HSR-MSI)

provides precise spatial information on thousands of biomolecules without labelling across a tissue section. Deep learning methods, trained on large numbers of images, can be used to further improve resolution. Nonetheless, the limited amount of HSR-MSI data that are publicly available makes that super-resolution reconstruction of images obtained by MSI a difficult endeavor.

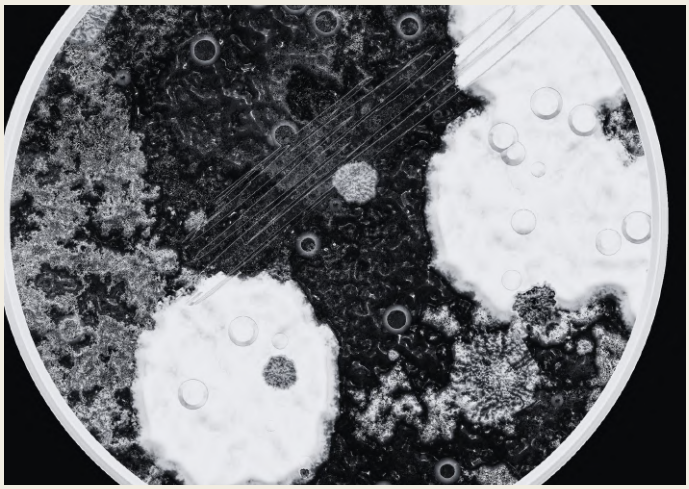
New pangenome reference could deepen our understanding of human biology and disease

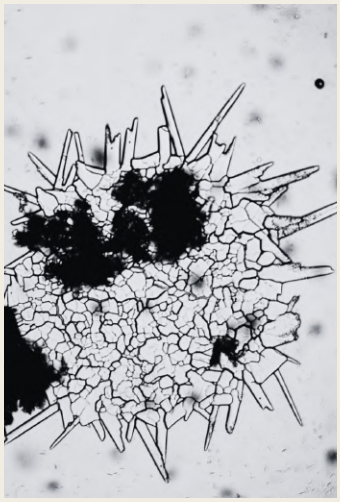
The first generation of a new type of reference genome, called a pangenome, that represents 47 individuals that are as genetically diverse as possible, from Africa, Asia, and the Caribbean, to name just a few regions.

The researchers revealed that every time they sequenced a human genome, they found pieces of human DNA that were unique to each individual. The pangenome allows the integration of this new information into the reference. In the long run, this development could potentially have substantial implications for the advancement of personalized medical treatments tailored to the specific DNA sequences unique to individuals or specific populations.

Using nanopore sequencing to identify fungi from clinical samples with high phylogenetic resolution

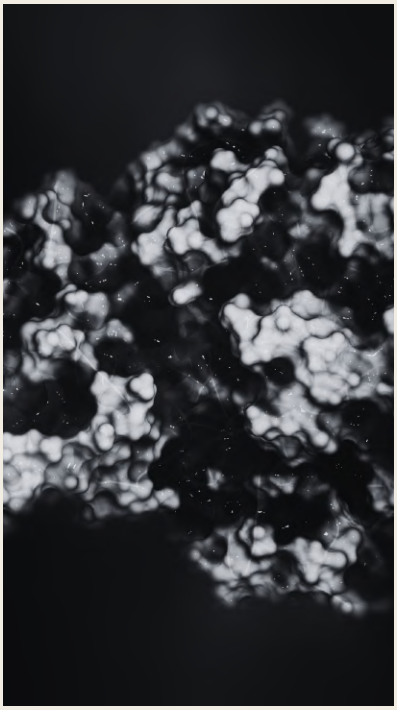
The study of microbiota has been revolutionized by the development of DNA metabarcoding. This sequence-based approach enables the direct detection of microorganisms without the need for culture and isolation, which significantly reduces analysis time and offers more comprehensive taxonomic profiles across broad phylogenetic lineages. While there has been an accumulating number of researches on bacteria, molecular phylogenetic analysis of fungi still remains challenging due to the lack of standardized tools and the incompleteness of reference databases, which restrict the accurate and precise identification of fungal taxa As a result, the process of identifying and classifying fungi remains difficult.





Puzzling Biochemists for Decades:

Reconstruction of Two-Billion-Year-Old Enzyme Solves a Long-Standing Mystery



The research team reconstructed an ancestral enzyme by searching databases for corresponding modern enzymes, using the obtained sequences to calculate the original sequence, and introducing the corresponding gene sequence into lab bacteria to produce a desired protein. The enzyme was then studied in detail and compared to modern enzymes.

They focused on enzymes called tRNA nucleotidyltransferases, which attach three nucleotide building blocks in the sequence C-C-A to small RNAs (transfer RNAs) in cells. These RNAs are subsequently used to supply amino acids for protein synthesis. Using phylogenetic reconstructions, the team reconstructed a candidate for an ancestral enzyme that existed in bacteria around billion years ago and compared it to a modern bacterial enzyme.

Insilico announces new multimodal transformer-based aging clock for processing diverse data sets

Clinical stage generative artificial intelligence (AI)-driven drug discovery company, Insilico Medicine ("Insilico") has announced a new multimodal transformer-based aging clock that is capable of processing diverse data sets and providing insights into biomarkers for aging, mapping them to genes relevant to both aging and disease, and discovering new therapeutic targets designed to slow or reverse both aging and aging-related diseases. Dubbed Precious1GPT, this aging clock pays homage to the influential "One Ring" from the Lord of the Rings. These findings were published in the June 13 issue of the journal Aging.

Insilico has been at the forefront of both generative AI and aging research, and began publishing studies on biomarkers of aging using advanced bioinformatics in 2014. Later, the company trained deep neural networks (DNNs) on human "multi-omics" longitudinal data and retrained them on diseases to develop its end-to-end Pharma.AI platform for target discovery, drug design, and clinical trial prediction.

BIOMETANET

Metabolomics study using bioinformatic tool

BioMetaNet is an innovative and comprehensive approach to studying metabolism using metabolomics and bioinformatics tools and techniques.

This advanced approach involves the integration and analysis of complex metabolic data sets to gain a deeper understanding of biological systems and metabolic processes.

It utilises high-throughput methods to obtain large amounts of data on the metabolites present in a biological sample and then uses bioinformatics tools to analyse this data and generate insights into the metabolic pathways and networks at play. This approach has wide-ranging applications in fields such as medicine, environmental science, and agriculture and has the potential to provide valuable insights into the functioning of complex biological systems. Through the use of sophisticated data analysis techniques, BioMetaNet is poised to revolutionise our understanding of metabolism and its role in health, disease, and the environment.

One of the major applications of metabolomics is understanding disease mechanisms. Metabolic variations are known to play a critical role in the development and progression of various diseases, including cancer, cardiovascular disease, diabetes, and neurodegenerative diseases.

Metabolomics has been used to identify specific metabolic signatures associated with these diseases, which can serve as potential biomarkers for early detection, diagnosis, and prognosis. By utilising advanced technologies such as mass spectrometry and nuclear magnetic resonance (NMR) spectroscopy, metabolomics enables the identification, quantification, and analysis of thousands of metabolites in various biological samples, including blood, urine, tissue, and saliva.

BioMetaNet employs a range of sophisticated tools and techniques to analyse and integrate large-scale metabolic data sets, providing a comprehensive understanding of metabolic pathways and networks in biological systems. There are various statistical and bioinformatics tools used in BioMetaNet for the analysis and interpretation of large-scale metabolic data.

PCA is a statistical technique used to reduce the complexity of large datasets by identifying patterns and correlations among variables. In BioMetaNet, PCA is used to identify metabolic profiles and visualise metabolic differences between samples. Partial Least Squares-Discriminant Analysis (PLS-DA) is a machine learning algorithm used to identify metabolites that are differentially expressed between groups of samples. This technique is commonly used in biomarker discovery and to identify metabolic pathways that are altered in disease.

To identify and annotate metabolic pathways and to identify significant changes in pathway activity between different samples or groups.

Commonly used pathway analysis tools in BioMetaNet include MetaboAnalyst, KEGG, and Reactome.

To create visual representations of metabolic data, allowing for easier interpretation and identification of metabolic patterns.

Examples of data visualisation software used in BioMetaNet include R, Python, and Cytoscape.

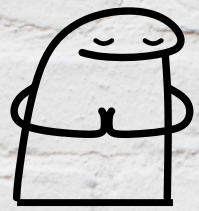
To identify and visualise metabolic networks and identify key metabolites and pathways that are critical to network function.

Examples of network analysis tools used in BioMetaNet include MetScape, MetNet, and MetaMapR.

These tools allow researchers to identify metabolic pathways and networks and discover biomarkers and metabolic signatures associated with specific diseases and conditions. By integrating metabolomics data with other omics data sets and using sophisticated statistical and bioinformatics tools, BioMetaNet has the potential to provide a comprehensive understanding of biological systems at the molecular level. This approach has numerous applications in fields such as medicine, environmental science, and agriculture and has the potential to transform our understanding of metabolism and its role in health and disease. The continued development and refinement of statistical and bioinformatics tools will be critical for the success of BioMetaNet and the advancement of precision medicine.

Aravindhan Vijayakumar

3rd Year B.Tech Biotechnology

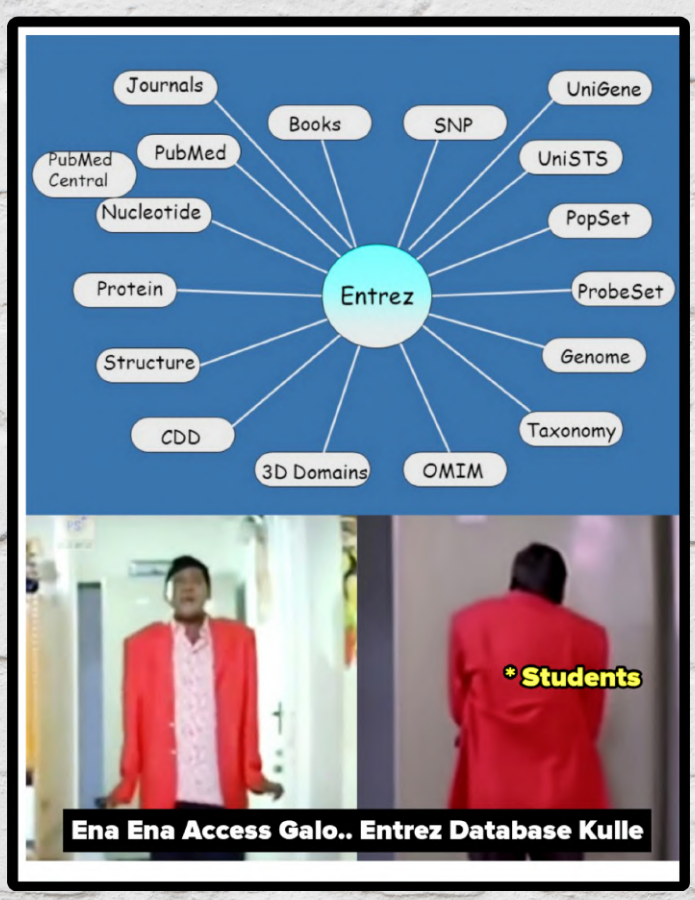


MEMES

K Dhanya
2nd year B.Sc Biotechnology

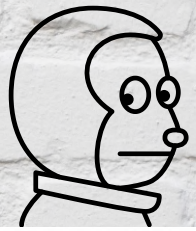
S Sujith
2nd year B.Tech Biotechnology

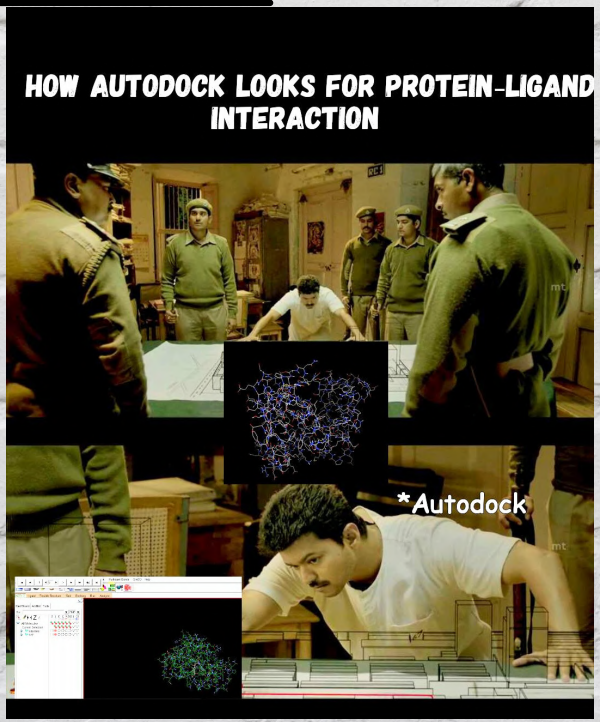
Abdullah Naveed Khan
1st year M.Tech Biotechnology



Meenakshi Sundaram L
1st year B.Tech Biotechnology

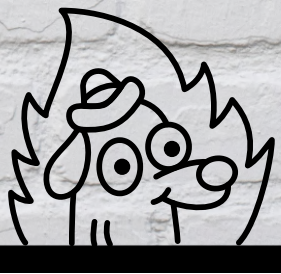
Dhanushwr K
1st year M.Tech Biotechnology

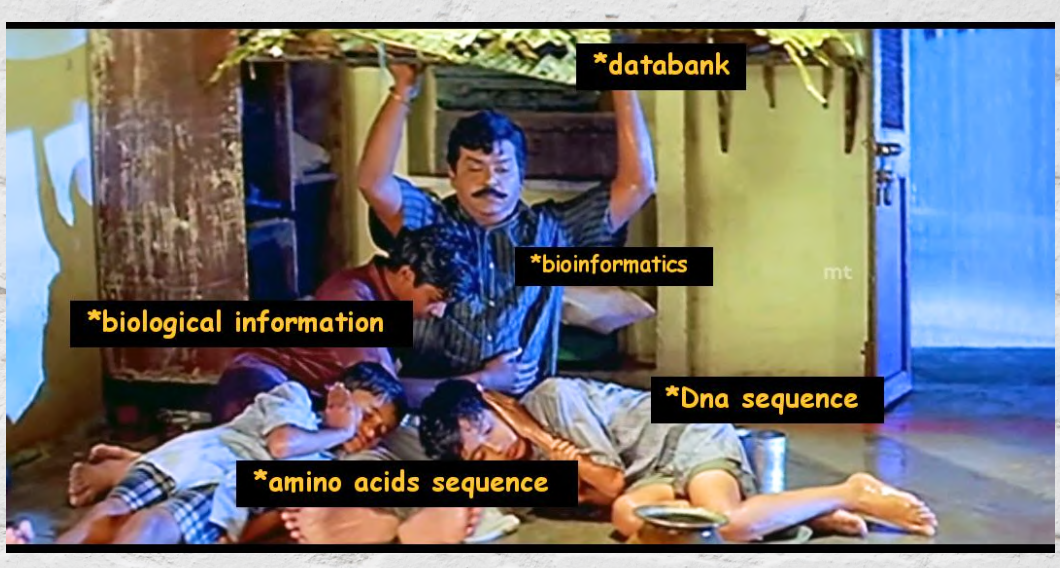




Rakshana B
3rd year B.Tech Biotechnology

Rasitha Arafa R
2nd year B.Tech Biotechnology





Venkatesan S
1st year M.Sc Biotechnology

ANSWER KEY

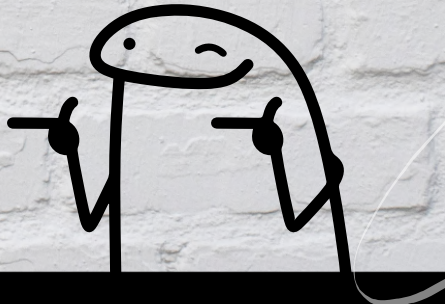
VERBAL VORTEX

ACROSS

1. FASTA
4. GLOBAL ALIGNMENT
5. ALGORITHM
6. BLAST
7. ENTREZ
8. PHYLOGENETIC

DOWN

1. FILE TRNSFER PROTOCOL
2. MEGA
3. FIELD
9. DAYHOFF



CD CODE

CrowDsourcing COndensate Database and Encyclopedia

Biomolecular condensates revolutionized intracellular molecule compartmentalization. The crowdsourced condensate database and encyclopedia integrate multidisciplinary scientific information regarding biomolecular condensate function and composition (cd-code.org). CD-CODE is a community-editable platform with a literature-based biomolecular condensates database, a scientific terms encyclopedia, and a crowdsourcing online application. This platform will speed up biomolecular condensate discovery, validation, and disease and therapeutic target research.

Membraneless biomolecular condensates selectively concentrate proteins and nucleic acids in the cell. They were recently implicated in various biochemical processes in in physiology and disease. Biomolecular condensates are increasingly therapeutic targets. Discoveries in basic research and drug development rely on quick access to relevant data. Like with any new paradigm, new phrases and ideas arise and adapt as the area advances. Hence, condensates and their component proteins and RNAs are defined differently in existing databases of proteins implicated in condensate production. These phase-separated protein databases are great. LLPSDB7 and PhaSePro6 gather proteins hypothesized to cause liquid-liquid phase separation. The former only collects in vitro data.

These databases do not address biomolecular condensate questions: When were the biomolecular condensates found and validated to date? What proteins are known? Proteins belong to which condensates? What experiments prove a condensate exists? We want to address these and other critical questions and establish a community-editable database for dynamic data updates. So a condensate-centric database based on scientific literature that contains experimental evidence, ratings, and references for each condensate-protein connection was created. Contributors dynamically update this database to reflect field knowledge.

CD-CODE is a "live database" where experienced researchers may quickly upload and evaluate condensate and protein data. The user management system supports viewers, contributors, and maintainers. Read and download the collected information. Contributors may suggest condensate and protein entries and modifications. Maintainers, and members of the development team curate changes and accept or reject contributor recommendations, which are subsequently notified and discussed. The crowdsourcing platform aggregates condensate biological scientific results to keep up with fast-changing terminologies, nomenclature, and empirical evidence.

CD-CODE (cd-code.org) now links 9,861 proteins to 244 biomolecular condensates and 375 in vitro synthetic condenses from 49 species. As contributors submit and evaluate data, these values change. CD-CODE, a semi-manually curated and annotated resource, collects primary literature and other databases. Protein entries are cross-referenced with UniProt9, Ensembl10, and the Human Protein Atlas (proteinatlas.org) for simple integration. Graphically displaying disorder score 12 and amino acid composition helps identify locations that may induce condensate partitioning.

An ontology from the literature to standardize condensate names and classify them by function to show their progression was created. Most condensates are mammalian and clade-specific. CD-CODE can help investigate the evolutionary genesis of condensates, as our information is limited and biased.

It is unknown which proteins produce condensates in cells and which condenses they split into. To investigate protein condensate-specificity, all known condenses a protein were identified, and the experimental evidence for each protein's relationship with each condensate was curated (confidence score, equating to zero to five stars): 1 star (literary evidence, PubMed ID; 2 stars, high-throughput; 3 stars, in vitro; 4 stars, in cellulo; 5 stars, in vivo) was gathered. Zero- or one-star condensates and proteins have not been carefully curated.

As predicted, many proteins partition into various condensates in dynamic cellular compartments, and condensate proteomes overlap significantly. Certain proteins must be present in all condensates (drivers). We identified 205 driver proteins in condensates using experimental data. G3BP1, a driver of stress granules, is also found in processing bodies (P-bodies) and neuronal ribonucleoprotein particle granules. CD-CODE will help us understand condensate-specific driver behaviour and if a driver protein may be used as a condensate "marker" in research.

Marker proteins identify condensates and guide drug screening procedures. They are uniquely connected with condensate and utilized in microscopy to view condensates, such as in colocalization tests to verify protein localization. Many marker proteins in our database are not condensate-specific. DCP1A marks P-bodies, stress granules, and nucleoli. Understanding protein components helps design experiments to identify condensates accurately.

CD-CODE answers the first question: P-granules, the germ granules of Caenorhabditis elegans, contain 190 protein components, one of which, PGL-3 (PGL3 CAEEL), is a driver for P-granule production and is supported by in vivo experimental findings (5 stars). Pgl-3 is a marker protein unique to P-granules.

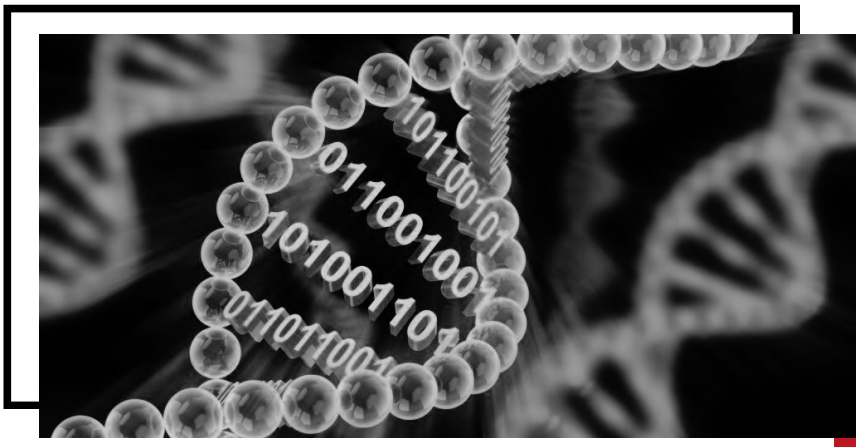
Databases of proteins experiencing liquid-liquid phase separation have allowed machine learning systems to predict phase separation and identify protein features that cause it. Next is which biomolecular condensate a protein belongs to. This condensate proteome database may aid protein recruitment studies. This site can provide high-quality benchmarking data for condensate protein prediction machine learning methods.

Drug seekers may use this complete curation of condensate kinds, their composition in numerous species, and experimental assistance to create assays and screening pipelines. For high-content imaging phenotypic tests, the protein or protein combination should be selective for the target condensate. CD-CODE facilitates and speeds the nomination of novel condensate-associated pharmacological targets via community database updates and publication curation.

Biomolecular condensates are extremely transdisciplinary and constantly evolving, requiring community agreement on definitions and terminologies. The standalone wiki encyclopedia collects condensate research information. As the study area evolves, we will add new features and data points regularly and annually.

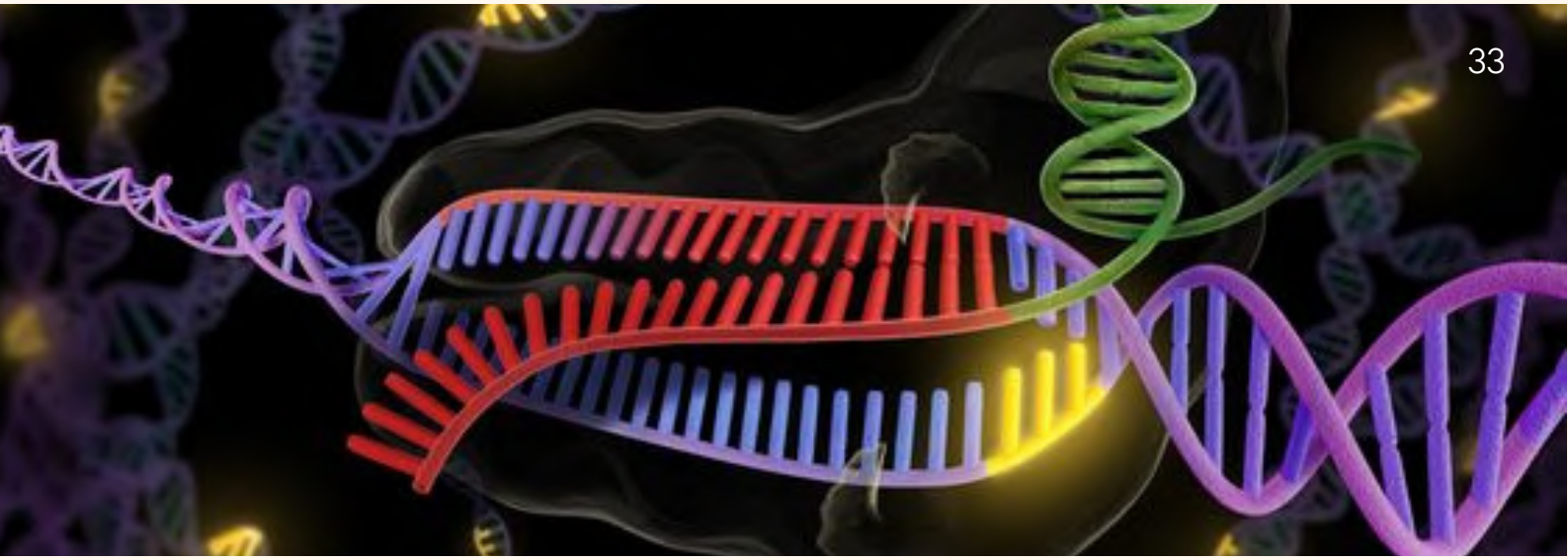
Experimentally confirmed entries define CD-CODE. Users should be cautious when interpreting a protein, condensate, or species' lack of data, since it may just represent the community's predisposition towards specific model systems and biological pathways.Any missing information might indicate that,

(1) The protein or condensate has not been investigated;
(2) There is a study article, but the material has not been contributed to the database;
(3) the condensate does not exist; or
(4) The protein does not belong to a certain condensate.



CD-CODE highlights field unknowns to drive future research topics to address gaps. Computational predictions beyond CD-CODE can fill these experimental gaps. New experimental evidence will enhance high-scoring condensate entries in the CD-CODE database. The crowdsourcing tool lets the community review definitions and evidence as the topic advances. This will guarantee that new condensate research information is quickly added to the database and encyclopedia.

DHANUSHWR K
1st Year M.Tech Biotechnology

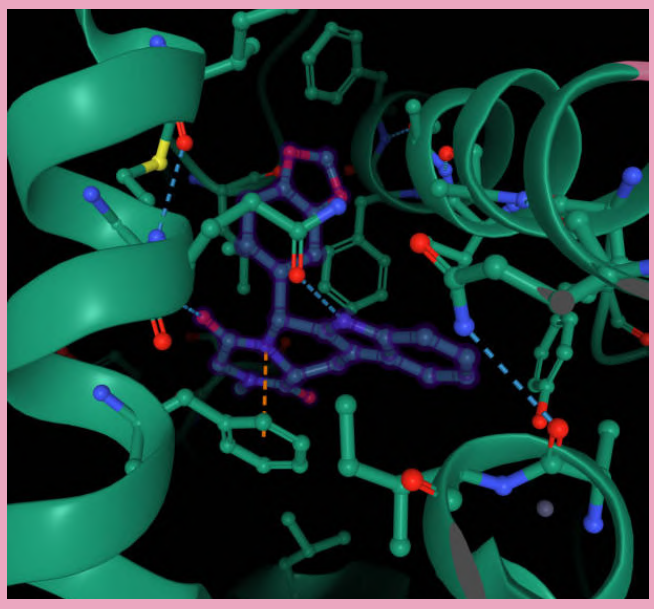


MACHINE LEARNING IN BIOLOGY AND BIOINFORMATICS

Machine learning (ML) is a wide collection of algorithms that predict and perform tasks based on a data set. Models are the famous algorithms that we use today. For an existing problem, the model is selected based on the available data and the desired outcomes. Reinforcement learning, supervised learning, and unsupervised learning are the three main branches of ML models.

In reinforcement learning techniques, the device uses trial-and-error methods to give accurate solutions by observing and interpreting the environment of a complex problem.

In supervised learning, labelled data sets are used to teach the machine to classify and predict outcomes. In unsupervised learning, the models work on their own to understand hidden data patterns.



Bioinformatics is a multidisciplinary approach to the life sciences. It solves biological issues by evaluating, analyzing, and interpreting biological data sets. In bioinformatics codes, algorithms, and models are developed that record and store data. Molecular medicine, drug development, and genome application are a few fields where bioinformatics is used.

MACHINE LEARNING
APPLICATIONS

GENOME SEQUENCING

ML refers to DNA sequence techniques such as next-generation sequencing, which has made human genome sequencing easier.

GENE EDITING

CRISPR is a technology used to insert, delete, or replace a DNA sequence in order to manipulate the genetic composition of an organism. Here, ML identifies the target audience, reducing the cost and time required for gene editing.

PROTEOMICS

It is the study of protein components, their interactions, and their role in an organism. ML is used to identify biomarkers in samples without identifying the proteins and peptides present. It just identifies the peak with high signal intensities. Prosit is a software developed recently to recognize protein sequences without errors. Machine learning can be used to solve complex biological problems using various algorithms and models.

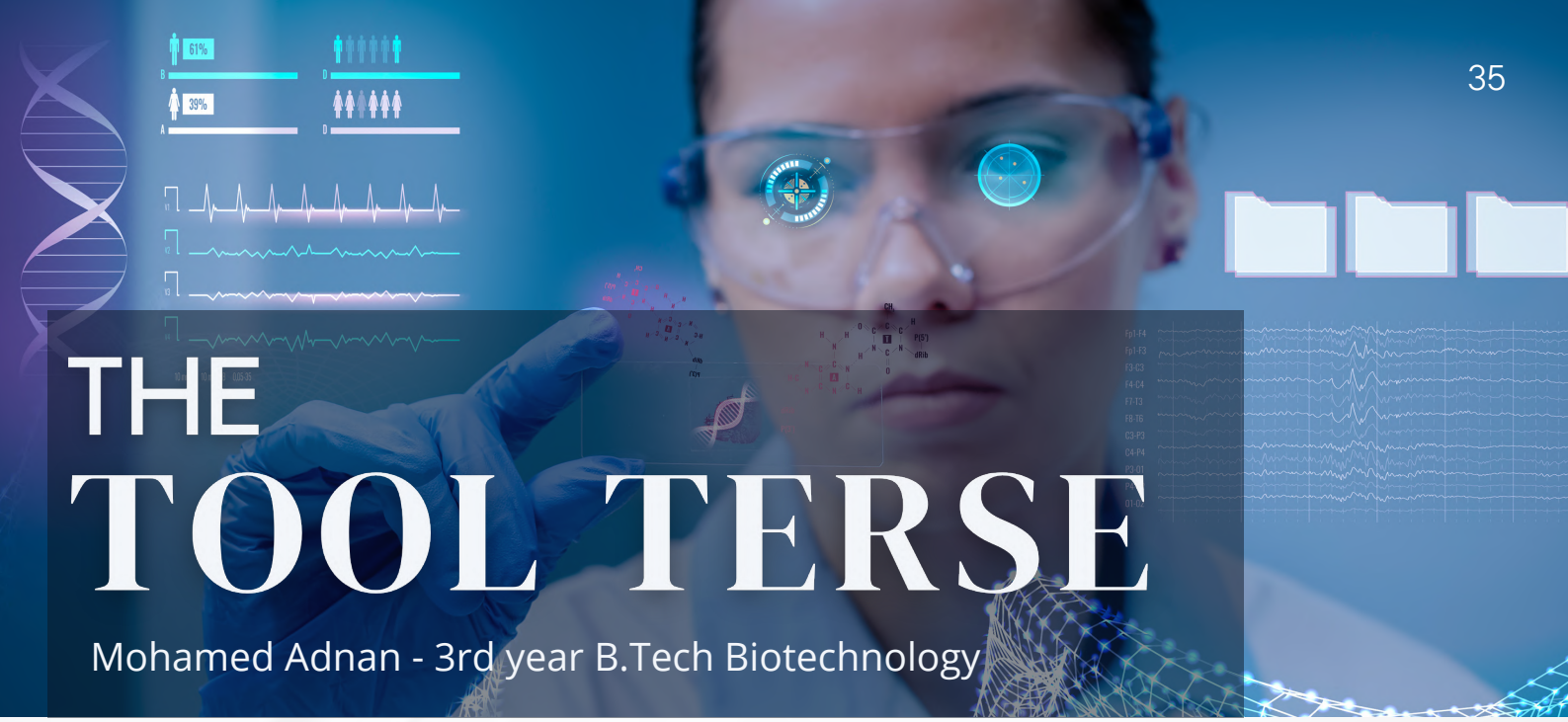
CLINICAL WORKFLOW

Accessing patients' medical data has always been a problem for healthcare workers. The ML analytical toolkit has now made things easier for keeping records.

GENOMICS

A genome is the complete set of genetic material. The main focus of genomics is the mapping, evolution, and editing of the genome. In regulatory genomics, ML provides transcription factors and RNA-binding proteins in order to predict and classify the expression of genes. In structural genomics, ML helps to classify the 1°, 2°, and 3° structures of proteins. In functional genomics, ML helps with mutation and localization of subcellular proteins.

Megasri S
3rd year B.Tech Biotechnology

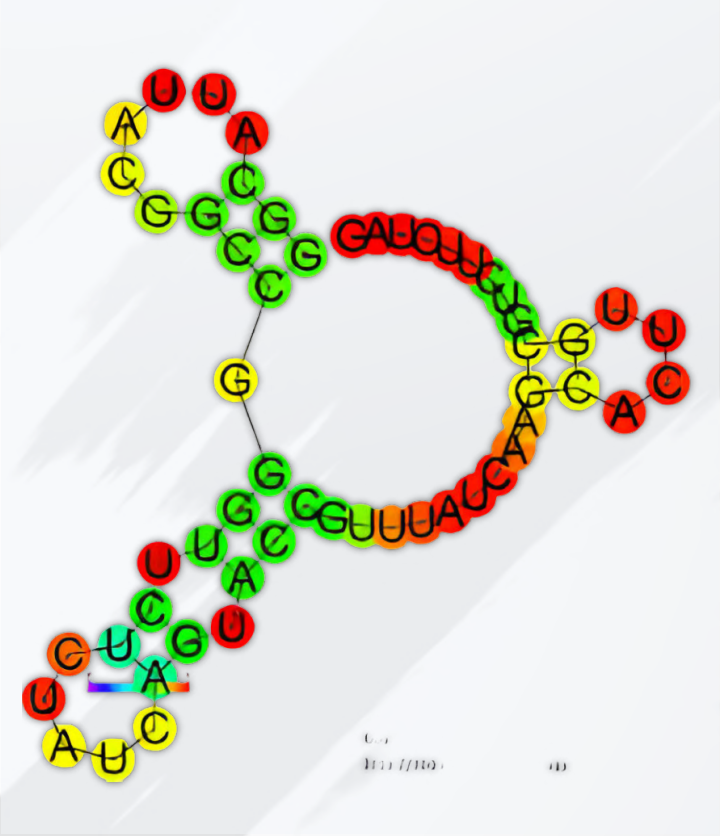


THE TOOL TERSE

Mohamed Adnan - 3rd year B.Tech Biotechnology

1. RNAfold cgi

This tool is mainly used for secondary structure prediction of ssRNA and DNA sequences. It requires the FASTA sequence of any desired RNA sequence whose structure needs to be predicted. It's a user friendly software which can be used by anyone easily with minimal practice and efforts, however, the person must be clear in what type of results are wanted. The software has a fast execution time and can provide results in form of secondary structures, graphical plots and also provide details about the minimum free energy and thermodynamic ensemble. We also get the minimum free energy (MFE) and centroid secondary structures.

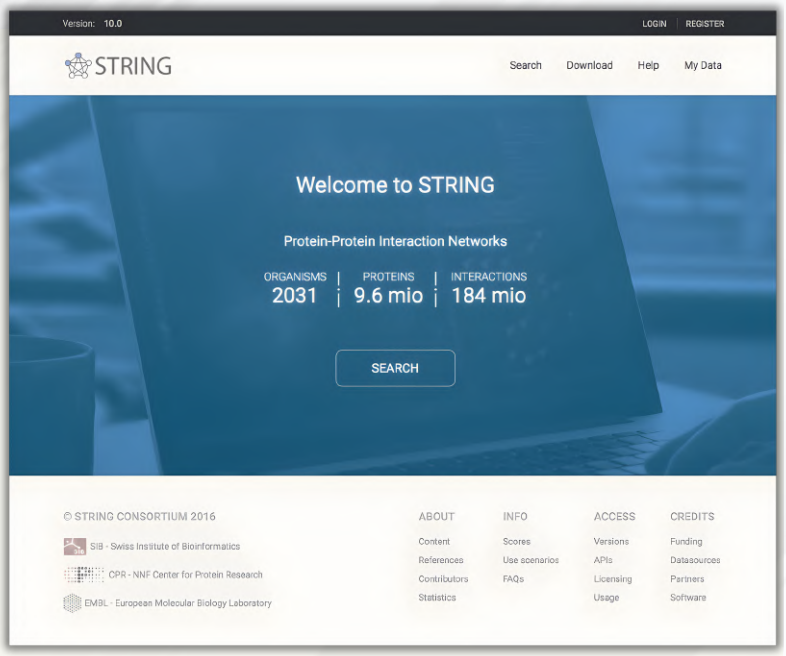


2. PROSITE

This tool focuses on the functions of proteins, it determines possible functions of newly found proteins or determines the previously unknown activity of older discovered proteins. This is another software that can be accessed via ExPASy. This software accepts input in the form of accession ids or FASTA sequences. The results are shown in a list wherein the primary function performed is displayed along with other essential factors associated with that function.

3. STRING

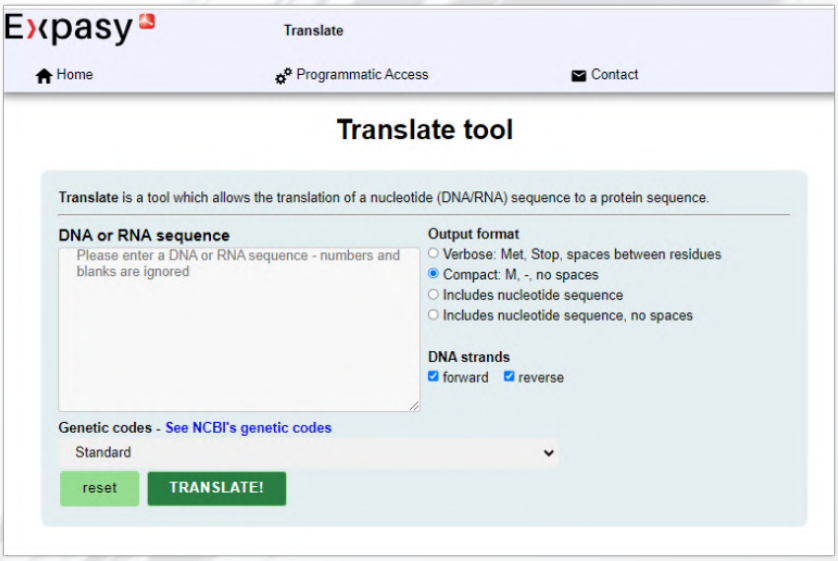
This is a software which can be accessed through ExPASy, this tool mainly focuses on protein-protein interaction at every level of complexity.



A newer update has made it very easy to use this software, all you need to do is select the protein which must be analysed and the organism which usually expresses the protein. The search results are displayed within a short amount of time with beautiful multi-coloured graphics. Every interacted protein is displayed as a sphere with different colours and the interactions are displayed in the form of a coloured line.

4. Translate

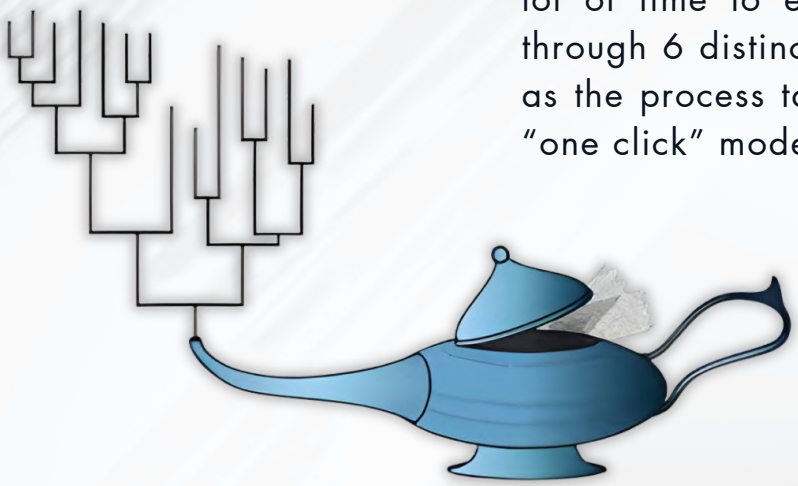
Translate is a widely used tool which converts the nucleotide sequence into a protein sequence and also tells about the open reading frames present. You can either use DNA or RNA sequences as the input depending on the organism of target. Running the program will give us both reverse and forward results and will also highlight the open reading frames and the start codons.



This software usually gives us 3 different frames which can then be used as per choice.

5. Phylogeny.fr

This is an independent tool which provides us with phylogenetic trees of organisms whose FASTA sequence must be provided. It accepts FASTA sequences of both proteins and genomes, however we must not mix up the sequences and must focus either on proteins or on genomes. This tool takes a lot of time to execute its operation as it must go through 6 distinct steps which will be showed to you as the process takes place. We usually focus on the "one click" mode for our results.

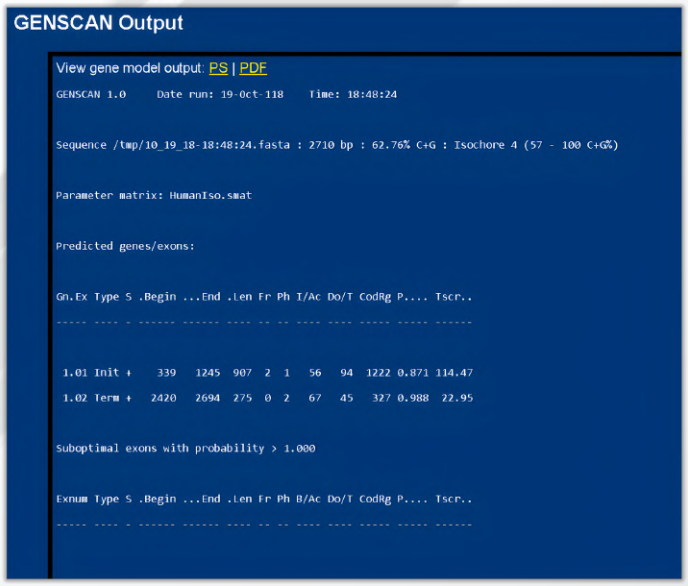


It is an efficient software which also gives us the evolutionary distance values. The results can be downloaded in the format of your choice as this tool also supports the TGF and Newick formats.

Its best to run this software on google chrome as sometimes there are issues faced while running it in browsers such as Mozilla Firefox.

6. GENSCAN

This software is used to predict genes in eukaryotes by using the Hidden Markov Models (HMM). All you have to do for this program is to download the FASTA format of the nucleotide sequence and paste it. The operation gives us results in a fraction of seconds. The results predict the functions of genes present in the sequence pasted and also predicts the position of these genes along with the length of the gene and possible protein sequence. C+G percentages are also calculated in the process



BIOINFORMATICS IN HOMELAND SECURITY

In today's culture, bioinformatics is a vital tool. Medicine, biotechnology, agriculture, and even national security are just a few of its uses. Homeland security is one such area where bioinformatics has become essential. Identification and defence against biological threats that may be a result of terrorism, hostilities, or even natural calamities depend heavily on bioinformatics.

The goal of homeland security is to protect the nation from any threat that could endanger its infrastructure or its population. One of the top issues on the minds of homeland security organizations is the biological threat. Viral, bacterial, and toxin-based biological agents can spread quickly and seriously harm both people and animals. Therefore, having a reliable and effective system to detect, identify, and react is crucial.

Bioinformatics provides a powerful tool for the detection and identification of biological agents.

The genome sequencing of microorganisms can be used to identify specific pathogens and their genetic characteristics. The use of databases such as GenBank, which contain a vast amount of genomic data, can help identify an unknown pathogen based on its genetic sequence. Additionally, the creation of vaccinations and medications to combat biological threats relies heavily on bioinformatics.

The discovery of a pathogen's genetic sequence can help in the creation of vaccinations and medications that can target particular elements of the disease. Drug development can be accelerated through the use of bioinformatics in drug discovery by anticipating a drug candidate's effectiveness prior to laboratory testing.

The tracking of a disease's transmission can also be made easier with the use of bioinformatics. A specific infection's origin and transmission can be tracked by identifying genetic markers that are distinctive to that pathogen. Utilizing this knowledge will enable policy makers to control the disease's spread and stop further outbreaks.

Forensic analysis is a vital area where bioinformatics plays a significant role in homeland security.

In forensic investigations, DNA analysis is an effective tool. Even in minute quantities, a person or an entity can be recognised by the detection of particular genetic markers. Using bioinformatics techniques, DNA samples can be analyzed and specific genetic markers that can be used in forensic investigations can be found.

In summary, bioinformatics is essential to homeland security. Its uses include identifying biological agents, creating medications & vaccinations, monitoring the spread of diseases, and even performing forensic investigation. The use of bioinformatics in homeland security can assist in defending the country from biological threats and secure its infrastructure and population. To stay ahead of the nation's biological dangers, which are always changing, it is crucial to sustain funding this research.

DNA analysis revolutionizes homeland security investigations, leveraging genetic samples to identify individuals involved in security incidents, aiding in solving cases, combating terrorism, and strengthening prosecution efforts.

DECIPHERING THE LANGUAGE OF LIFE TO SAFEGUARD OUR NATION FROM BIOLOGICAL PERILS.

- Jeffrey Jackson C
3rd year B.Tech Biotechnology

FROM THEORY TO PRACTICE

INTERVIEWER

As the Public Relations Executive of the Crescent Bioinformatics Association, Annish Shabiya M is our adept interviewer in this edition. Driven by a passion for bridging theoretical knowledge to practical applications, she delves deep into scholars' groundbreaking research in the field of bioinformatics. The insightful inquiries shed light on the cutting-edge discoveries shaping science's future.

Annish Shabiya M,
4th year B.Tech Biotechnology

INTERVIEWEES

Ms. Arunika Krishnan,
Research Scholar, School of Life Sciences

Arunika Krishnan is a dynamic researcher in bioinformatics and genomics. Since embarking on her journey in January 2019, she's been passionate about unraveling the mysteries of Mycobacterium tuberculosis. Through groundbreaking genomic analysis, she seeks novel therapeutic targets to combat tuberculosis, offering hope in the fight against infectious diseases.

Ms. Shabnam Ameenudeen,
Research Scholar, School of Life Sciences

Shabnam Ameenudeen is a research scholar specializing in the captivating fields of bioinformatics and neuroscience. Having embarked on her doctoral journey on the day of 26th August 2021, she is making significant strides in her pursuit of unraveling and understanding the complexities of the nervous system disorder via bioinformatics.

Interviewer: Welcome, dear readers, to this exciting edition of "FROM THEORY TO PRACTICE." Today, we are privileged to engage with two exceptional research scholars, Ms. Arunika Krishnan, and Ms. Shabnam Ameenudeen, from the School of Life Sciences, B S Abdur Rahman Crescent Institute of Science and Technology. Let's explore their intriguing journeys in the field of bioinformatics.

Interview with Ms. Arunika Krishnan, Research Scholar, School of Life Sciences

Interviewer: **Ms. Arunika Krishnan, In bioinformatics research, what are some common challenges or limitations that researchers often encounter, and how have you personally navigated them in your own work?**

Ms. Arunika Krishnan: *There are multiple different tools for various purposes in bioinformatics. One major challenge is finding the right tool. Choose the tool, based on the operating system, whether the tool is free or paid, and choose the most cited tool in research articles.*

Interviewer: **That makes sense,Can you discuss a specific instance in your research where the initial expectations or hypotheses differed from the actual outcomes? How did you handle this situation, and what did you learn from it?**

Ms. Arunika Krishnan: *Definitely, The hypothesis I had framed initially failed and gave me opposing results. One of the ways to approach this problem would be to look at the various publications, looks at the pattern of results obtained. By intensive literature you can try to understand the scenario.*

Interviewer: **Bioinformatics research often involves working with large and complex datasets. Could you shed some light on the practical difficulties researchers face in terms of data collection, processing, and analysis?**

Ms. Arunika Krishnan: *Availability of computational facilities is the major challenge. Next difficulty is the availability of the right tool. Many tools in bioinformatics are commercial ones, finding the alternate free version is the key.*

Interviewer: **How do you handle issues related to reproducibility and transparency in your bioinformatics research? What steps do you take to ensure that your methods and results can be validated and replicated by other researchers?**

Ms. Arunika Krishnan: Ensure that you clearly mention the tool used and cite the actual paper.

Interviewer: **Bioinformatics often requires integrating data from multiple sources and applying various algorithms. Can you share your experiences with the practical challenges associated with data integration and algorithm selection?**

Ms. Arunika Krishnan: *It is a little tricky when it comes to finding the right algorithm. Python and R programming codes are available online, adequate knowledge is required to convert the codes according to your requirement.*

Interviewer: **Collaboration is essential in many bioinformatics projects. Can you describe a collaboration or teamwork experience that had a significant impact on your research? What were the key factors that contributed to its success or challenges?**

Ms. Arunika Krishnan: *I did not have any collaborative work. I took the help of experts in the field to understand the various aspects of my work.*

Interviewer: **Bioinformatics research often involves working with cutting-edge technologies and tools. Could you elaborate on the real-world complexities associated with implementing and utilizing these technologies in your research?**

Ms. Arunika Krishnan: *Availability of computational resources and availability of commercial tools are the deciding factors. Most of them are commercial tools.*

Interviewer: **Can you discuss the role of data quality and annotation in bioinformatics research? What steps do you take to ensure the accuracy and reliability of the data you work with?**

Ms. Arunika Krishnan: *The possible approach would be performing the experiments in triplicates to ensure accuracy and reliability.*

Interviewer: **How do you navigate the process of troubleshooting and problem-solving in bioinformatics research, especially when faced with unexpected results or technical issues?**

Ms. Arunika Krishnan: *I immediately look into the literature. Since this is bioinformatics, looking for a paper which has performed laboratory experiments, highlighting or objecting your results are key. Based on the former or latter, you can decide on whether the results are valid or need an upgrade.*

Interviewer: **Bioinformatics research often requires making assumptions and simplifications to model complex biological phenomena. Can you elaborate on the potential impact of these assumptions on the interpretation and generalizability of research findings?**

Ms. Arunika Krishnan: *These assumptions are actually very helpful, especially in a condition where similar organisms are involved.*

Interviewer: **Time management and meeting deadlines can be stressful in research. How do you handle the pressure of meeting research milestones and completing tasks on time without compromising your emotional well-being?**

Ms. Arunika Krishnan: *You tend to slightly lose your cool, when you are trying to figure out something. Giving enough breaks has helped me significantly.*

Interviewer: **The nature of bioinformatics research often involves uncertainty and ambiguity. How do you manage the emotional aspects of working with incomplete or inconclusive data, and how do you stay motivated during such times?**

Ms. Arunika Krishnan: *Getting the right answer is supremely important. So in spite of getting unexpected results, understanding the experimental system and fixing issues are primary and that urge to find out the right answer makes all the difference.*

Interviewer: **Imposter syndrome is a common phenomenon among researchers. Have you ever experienced self-doubt or a feeling of not being competent enough in your bioinformatics work? How do you combat imposter syndrome?**

Ms. Arunika Krishnan: *Each and every work is different. I have experienced a lot of self doubt. But looking at the positive milestones that you have achieved will take you out of that self.*

Interviewer: **Research projects can sometimes face unexpected challenges or roadblocks. Can you discuss a situation where your research encountered a significant obstacle, and how did you handle the emotional impact of that setback?**

Ms. Arunika Krishnan: *I faced a major roadblock as I got opposing results to my hypothesis.I tried to figure out the issue by checking if any literature is available supporting that result. Yes, it was a really difficult time for me as I wasn't sure how to take it forward from there. Brainstorming and discussion with my supervisor really helped.*

Interviewer: **Receiving feedback and criticism is an integral part of the research process. How do you deal with constructive criticism or negative feedback on your bioinformatics work, and how has it influenced your emotional well-being?**

Ms. Arunika Krishnan: Criticism is actually important. They act as checkpoints in your research progress. You get to learn a lot, while trying to address the above.

Interviewer: **Can you discuss any interesting or unexpected findings you have come across during your Ph.D. research and how they have influenced your work?**

Ms. Arunika Krishnan: *I work on mutation and its association with drug resistance. In one such instance, the drug that I was working on totally had opposing results compared to the hypothesis. I started exploring more and tried to find out interesting scenarios from those results.*

Interviewer: **How do you stay updated with the latest advancements and trends in bioinformatics, and how have these influenced your research approach?**

Ms. Arunika Krishnan: *Reading latest research articles, attending conferences and talks has helped me in this.*

Interviewer: **How do you plan to disseminate your research findings to the scientific community and beyond?**

Ms. Arunika Krishnan: *Publishing my research in journals that are specific to the area of research. In that way you can increase the chances of viewership.*

Interview with Ms. Shabnam Ameenudeen, Research Scholar, School of Life Sciences

Interviewer: **Bioinformatics research often involves working with large and complex datasets. Could you shed some light on the practical difficulties researchers face in terms of data collection, processing, and analysis?**

Ms. Shabnam Ameenudeen:
In bioinformatics research, besides defining the problem statement the selection of dataset is the most common challenge as there are a lot of biological data available with different sample size, its important the chosen data is in correspondence with the objective of the study and the sample size is sufficient.

Interviewer: **Bioinformatics often requires integrating data from multiple sources and applying various algorithms. Can you share your experiences with the practical challenges associated with data integration and algorithm selection?**

Ms. Shabnam Ameenudeen: *In my case, troubleshooting the errors is the biggest challenge. An algorithm worked for one researcher may not work for us. In such cases, identifying the logic behind the application of each algorithm is important.*

Interviewer: **In bioinformatics research, what are some common challenges or limitations that researchers often encounter, and how have you personally navigated them in your own work?**

Ms. Shabnam Ameenudeen: *Defining a problem statement is the most crucial step, till we get a solid result whatever the work proposed or research being carried out is hypothetical. One can easily deviate from the objective once when we start focusing on the minute details. Hence it is important that equal emphasis is given on understanding both the problem and the steps of addressing it.*

Interviewer: **How do you handle issues related to reproducibility and transparency in your bioinformatics research? What steps do you take to ensure that your methods and results can be validated and replicated by other researchers?**

Ms. Shabnam Ameenudeen:
Extensive reading is the key to ensure that the research carried out and results obtained are reproducible enough. It is important that we understand the research which was carried out previously this helps in better interpretation and analysis of the results and to make defending statements.

Interviewer: **Can you discuss a specific instance in your research where the initial expectations or hypotheses differed from the actual outcomes? How did you handle this situation, and what did you learn from it?**

Ms. Shabnam Ameenudeen: *Sure. Besides giving me the opposing results the hypothesis I had framed initially did not provide a brief picture of my objective. One advice I received to handle such a problem is looking into research publications is fine but "always go back to the basics". As research publications are part of something and someone's objective, they don't give the background information.*

Interviewer: **Collaboration is essential in many bioinformatics projects. Can you describe a collaboration or teamwork experience that had a significant impact on your research? What were the key factors that contributed to its success or challenges?**

Ms. Shabnam Ameenudeen: *Definitely. Interdisciplinary research always has its perks, as they help in knowledge exchange and open the doors of broader possibilities and applications. In Bioinformatics research, most of the tools and software are made using Python, R, Linux, etc., and require knowledge of the same. I had the opportunity to work with Professors from a Computer science background. It helped me with carrying out efficient research.*

Interviewer: **Bioinformatics research often involves working with cutting-edge technologies and tools. Could you elaborate on the real-world complexities associated with implementing and utilizing these technologies in your research?**

Ms. Shabnam Ameenudeen: *The cutting-edge technologies and tools which fascinates every researcher comes with its own complexities. To mention a few, the need of GPU's as data is huge and data analytics with those datasets is the most laborious process. Next would be the understanding or knowledge of different programming languages. Since post purchase of the resources, their application plays an important role.*

Interviewer: **Time management and meeting deadlines can be stressful in research. How do you handle the pressure of meeting research milestones and completing tasks on time without compromising your emotional well-being?**

Ms. Shabnam Ameenudeen: *To me research is like a marathon, you have to be consistent with whatever you are doing. Not every research and every researcher will work at the same pace, hence its better not to compare. Prioritizing and planning work accordingly helps in completion of tasks on time. Moreover, one has to understand that in research a little progress is also a progress, there is no right or definite measure.*

Interviewer: **Imposter syndrome is a common phenomenon among researchers. Have you ever experienced self-doubt or a feeling of not being competent enough in your bioinformatics work? How do you combat imposter syndrome?**

Ms. Shabnam Ameenudeen: *Whenever I feel anxious about not being competent, I give myself a break. Its not like stopping whatever I'm doing but changing the way I do it because the pattern I once followed could have failed or might not support me in further progress. There is no fixed way. Until and unless the work is done self-experimentation always works best.*

Conclusion:

Interviewer: Throughout the interview, Ms. Arunika Krishnan and Ms. Shabnam Ameenudeen shared valuable insights into the challenges and experiences they encountered as Ph.D. scholars in the field of bioinformatics. They highlighted the importance of selecting appropriate tools, managing large datasets, ensuring reproducibility and transparency, and addressing unexpected results. They also discussed the significance of collaboration, staying updated with advancements, managing time and emotional well-being, and dealing with imposter syndrome.

The interviews emphasized the critical role of meticulous data collection, processing, and analysis and the challenges associated with algorithm selection and data integration. Both scholars acknowledged the need for clear problem statements, extensive literature reviews, and adaptability in research. They emphasized the significance of feedback, constructive criticism, and maintaining motivation during setbacks. The interviews also shed light on the impact of interdisciplinary collaborations and the complexities of implementing cutting-edge technologies in bioinformatics research.

These discussions with Ms. Arunika Krishnan and Ms. Shabnam Ameenudeen provide valuable insights into bioinformatics research's challenges, strategies, and emotional aspects. Their experiences and learnings contribute to the overall understanding of the field and guide aspiring bioinformatics researchers.

Can **ChatGPT** replace Bioinformatics work?

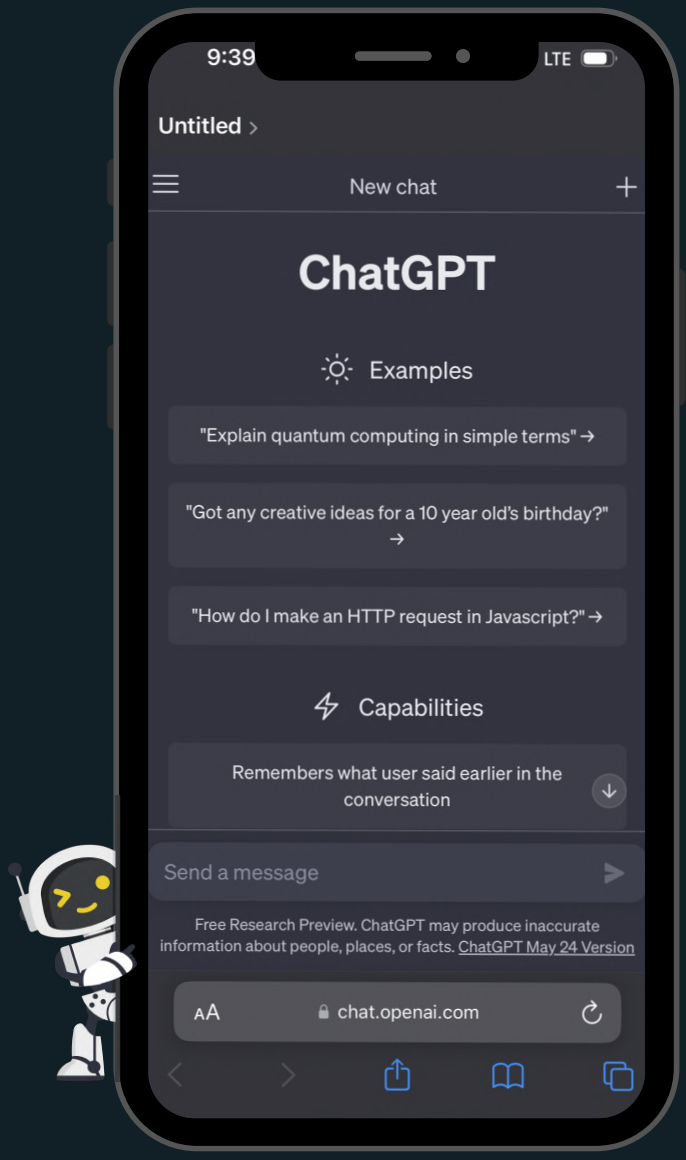
Introduction

Bioinformatics is a multidisciplinary field that combines biology, computer science, and statistics to analyse and interpret biological data. With the rapid advancement of artificial intelligence (AI), there has been speculation about the potential of ChatGPT, developed by OpenAI, to replace bioinformatic work. However, while ChatGPT has shown great promise in many areas, there are limitations to its ability to fully replace bioinformatics work.

Use Cases of ChatGPT in Bioinformatics.

Key strengths of ChatGPT are its ability to process and analyse vast amounts of text data, including scientific literature, research articles, and databases. ChatGPT can quickly sift through large datasets and extract relevant information, making it a powerful tool for literature reviews, data mining, and data extraction tasks in bioinformatics. Additionally, ChatGPT can generate hypotheses, predict protein structures, and perform gene expression analysis, which can aid researchers in formulating research questions and designing experiments.

ChatGPT also had the potential to facilitate collaboration among bioinformatics researchers by serving as a virtual assistant that could answer questions, provide suggestions, and help with data analysis. It can also assist in the interpretation of complex genetic data, such as identifying gene mutations associated with diseases or predicting drug targets. This can save researchers time and effort in their data analysis and interpretation tasks and potentially accelerate the pace of research.

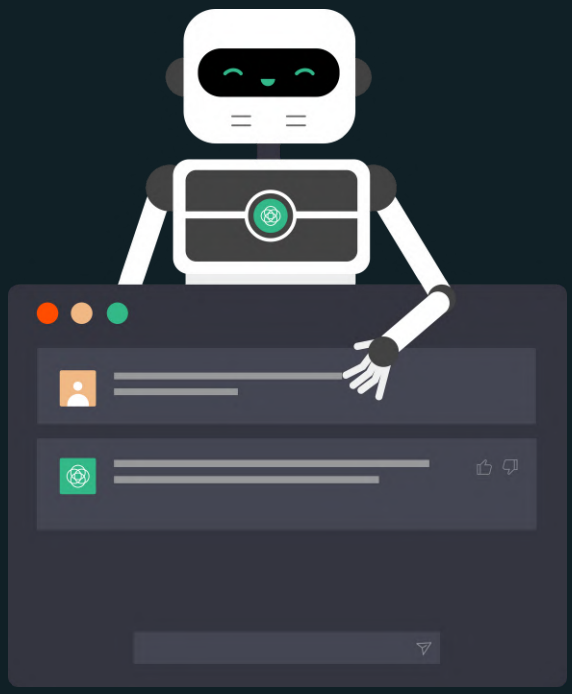


Limitations of ChatGPT in Bioinformatics

However, despite its strengths, ChatGPT has limitations in replacing bioinformatic work. One of the major challenges is the lack of domain-specific knowledge. While ChatGPT can generate text based on patterns learned from vast amounts of data, it may not have a deep understanding of the underlying biological concepts, mechanisms, and intricacies that are essential for the accurate analysis and interpretation of bioinformatics data. Bioinformatics requires a comprehensive understanding of molecular biology, genetics, genomics, and other specialised fields, which may go beyond the capabilities of ChatGPT.

Another limitation is the lack of context awareness. Bioinformatics data is complex and often requires deep contextual understanding to make meaningful interpretations. ChatGPT may not be able to fully grasp the contextual nuances of biological data, leading to inaccurate or incomplete results. Moreover, ChatGPT's inability to perform experiments or generate new data limits its ability to validate hypotheses or generate novel findings, which are crucial in bioinformatics research.

Jhawahar M.R.
2nd year M.Sc Biotechnology



Conclusion

In conclusion, while ChatGPT has shown promising capabilities in aiding bioinformatic work by processing large amounts of text data, generating hypotheses, and assisting in data analysis, it has limitations that prevent it from fully replacing human expertise in bioinformatics research. Bioinformatics is a complex and multidisciplinary field that requires a deep understanding of biological concepts, contextual awareness, and experimental validation, which may go beyond the capabilities of ChatGPT. Additionally, concerns related to data privacy and security should be carefully addressed when using ChatGPT for bioinformatics tasks involving sensitive data.

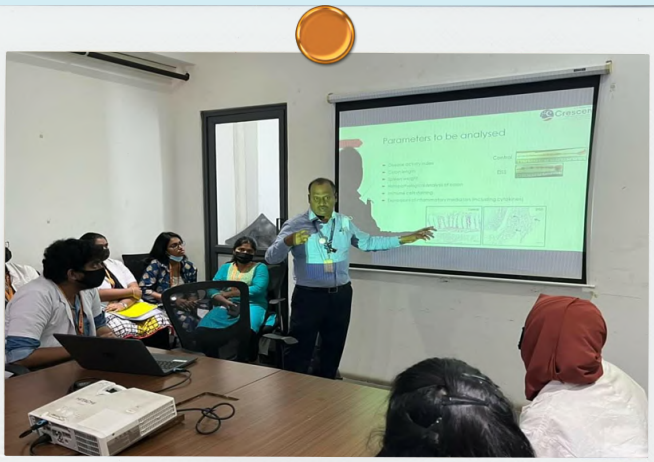
ChatGPT and other AI tools can be valuable aids in supporting bioinformatics research, but they should be used in conjunction with human expertise, critical thinking, and rigorous experimental validation to ensure accurate and reliable results. Researchers should continue to rely on their domain-specific knowledge, experience, and judgement in conjunction with AI tools to achieve the best possible outcomes in their bioinformatics work.

Frames of the year

Day-1

WORKSHOP : HANDS-ON TRAINING ON THE EXTRACTION AND PHYTOCHEMICAL ANALYSIS OF NATURAL COMPUNDS-IN-VIVO AND IN-SILICO APPROACH

QUALITATIVE ANALYSIS OF PHYTOCHEMICAL ANALYSIS



GUEST LECTURER: DR. ASHOK KUMAR (LECTURE ON NATURAL PRODUCTS; USES AND APPLICATIONS)

Day-2

DAY-2: MOLECULAR MODELING (IN-SILICO ANALYSIS)

GUEST LECTURER: MS. SHABNAM AMEENUDEEN (LECTURE ON ROLE OF COMPUTATIONAL APPROACH IN DRUG DESIGNING AND DEVELOPMENT)

GUEST LECTURER : DR.RAJESH KANAN (LECTURE ON ZEBRAFISH AS ANIMAL MODEL)

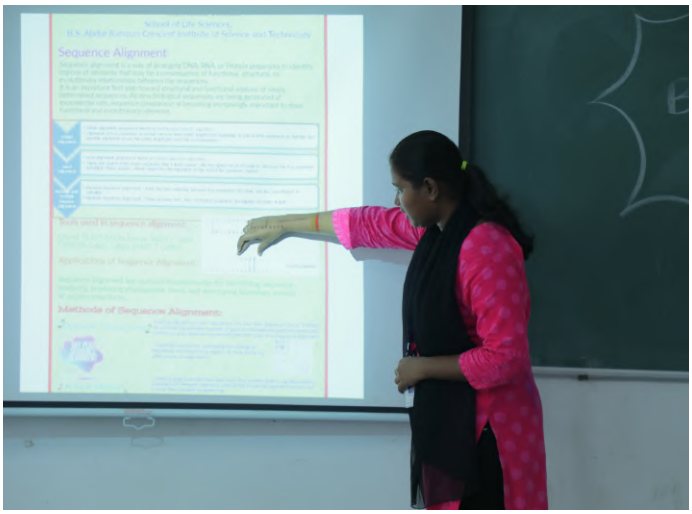
DAY-3: HANDS-ON TRAINING ON ZEBRAFISH HANDLING

PARTICIPANTS OF THE WORKSHOP

BIOINFOTHON CLICKS

POSTER PRESENTATION

ORGANIZERS AND FACILITATORS OF THE WORKSHOP



EXTEMPORE

QUIZ

FIND THE TOOL

GUEST LECTURER : DR. SALEEM MOHAMMED, PROVIDED INSIGHTS ON BIOINFORMATICS

PARTICIPANTS OF BIOINFOTHON

OFFICE BEARERS OF
CRESCENT BIOINFORMATICS ASSOCIATION 2022-23

Photography courtesy:
Jazeera Mariyam
2nd year, M.Sc Biotechnology

Muhammad Waleed
3rd year, B.Tech Biotechnology

Raykanth R
1st year, B.Sc Biotechnology

OFFICE BEARERS

2022 - 2023

DR. A BASKARAN
Faculty Coordinator

SHABNAM AMEENUDEEN
Club Advisor

LINA FATIMA M & THAMARAI G R
Co-Presidents

MOHAMED JUVAD
Student Coordinator

RIFAT AARA
Operations Head

RONISHA S
Design Head

MOHAMED ADNAN
Public Relations Head

NOORLIZA FATHIMA & PRATHIMA LAKSHMI
Operations Executives

ANNISH SHABIYA M & SUVETHA P
Public Relation Executives

SAKTHIVEL J & SANDHIYA DEVI P
Treasurers

CREDENTIALS

Editors-in-chief

Lina Fatima M , 4th year B.Tech Biotechnology
Thamarai G.R , 3rd year B.Tech Biotechnology

Editors

Annish Shabiya M , 4th year B.Tech Biotechnology
Cassandra Rifflin C.R , 3rd year B.Tech Biotechnology
Mohamed Adnan , 3rd year B.Tech Biotechnology
Suvetha P , 3rd year B.Tech Biotechnology

Head Designer

Ronisha S , 3rd year B.Tech Biotechnology

Designers

Deshmukhi K.R , 3rd year B.Tech Biotechnology
Navandikka R , 3rd year B.Tech Biotechnology
Rasitha Arafa R , 2nd year B.Tech Biotechnology
Sathiya S , 2nd year B.Sc Biotechnology
Shinoz T , 3rd year B.Tech Biotechnology
Yuvaraj S , 1st year M.Sc Biotechnology

UNDER THE GUIDANCE OF

Dr. A. Baskaran

FACULTY COORDINATOR


School of Life Sciences
LEADING LIFE THROUGH SCIENCE
CRESCENT
BIOINFORMATICS
ASSOCIATION
B.S. Abdur Rahman ™
Crescent
Institute of Science & Technology
Deemed to be University u/s 3 of the UGC Act, 1956