BIOINFORMATICS: DECODING LIFE'S DIGITAL SECRETS

Abstract

Bioinformatics and genomics are groundbreaking disciplines that combine molecular biology and computer science, revolutionizing our understanding of life's complexities. Bioinformatics leverages computational tools to analyze biological data, while genomics aims to unravel an organism's entire DNA sequence. These fields have witnessed significant milestones, from the establishment of the first DNA sequence database to the integration of artificial intelligence. Proteomics complements genomics by studying an organism's entire set of proteins, providing insights into cellular processes and disease mechanisms. The global genomic market is experiencing rapid growth, driven by advancements in DNA sequencing, precision medicine, and pharmaceutical applications. These interdisciplinary fields continue to shape modern biology, paving the way for transformative contributions to medicine, agriculture, and environmental research.

Keywords: DNA, NCBI, Genomics, Proteomics

Authors

Ankur Dahayat
Tropical Forest Research Institute
P.O. – RFRC
Jabalpur,Madhya Pradesh, India.
ankurdahayat1990@gmail.com

Raj Singh Yadav
Tropical Forest Research Institute
P.O. – RFRC
Jabalpur,Madhya Pradesh, India.

I. INTRODUCTION

Bioinformatics stands as a groundbreaking convergence of 20th-century molecular biology and computer revolutions. This burgeoning discipline centers on harnessing the power of computer databases and advanced algorithms to dissect and interpret the intricate realms of proteins, genes, and an organism's complete DNA composition, commonly known as its genome (Boguski, 1994). The formidable challenge in biology lies in comprehending the sheer magnitude of sequence and structural data emanating from genome-sequencing initiatives, proteomics, and other large-scale molecular biology endeavors. In response, bioinformatics equips researchers with a repertoire of sophisticated computer programs that unveil the fundamental mechanisms underpinning biological enigmas, ranging from macromolecular structure-function relationships to intricate biochemical pathways, disease mechanisms, and the marvels of evolution (Luscombe et al. 2001).Following the National Institutes of Health (NIH) definition, bioinformatics represents a realm of research and development that leverages computational tools and approaches to extend the frontiers of biological, medical, behavioral, and health data utilization. This encompasses the acquisition, storage, organization, analysis, and visualization of such data, revolutionizing the way we perceive life's complexities. Computational biology, close kin to bioinformatics, advances the application of data-analytical and theoretical methods, mathematical modeling, and computational simulations in studying the multifaceted systems of biology, behavior, and society (Goodman, 2002).

While bioinformatics primarily focuses on molecular sequence analysis, genomics, and functional genomics form essential allies in this quest for biological understanding. Genomics, at its core, seeks to unravel and explore the entire DNA sequence of an organism, its genome, a repository of genetic information that encodes genes. These genes, in turn, have the potential to be transcribed into ribonucleic acid (RNA) transcripts, and in many instances, translated into the intricate machinery of proteins. Functional genomics, on the other hand, leverages genome-wide assays to illuminate the roles and functions of genes and proteins, thereby painting a holistic picture of biological complexity (Henikoff, 2002; Yu et al. 2004).

This chapter is to elucidate both the theoretical underpinnings and the practical applications of bioinformatics and genomics. Designed with the biology student in mind, our approach facilitates the seamless integration of computer programs and databases as indispensable tools in tackling biological enigmas related to proteins, genes, and genomes. As an integrative discipline, bioinformatics embarks on an ambitious journey to unravel profound questions in biology, such as the intricate interplay between structure and function, the enigmatic realms of development and disease, and the marvels of life's evolutionary tapestry (Chen et al. 2019). Moreover, for the computer scientist, here offer compelling insights into the motivations behind the development and utilization of algorithms and databases, empowering them to make transformative contributions at the interface of biology and technology.

Bioinformatics is an interdisciplinary field that applies computational and statistical techniques to analyze and interpret biological data. It involves the development and application of algorithms, software tools, and databases to store, retrieve, and analyze biological information. Bioinformatics plays a crucial role in managing and interpreting the vast amounts of data generated from high-throughput sequencing technologies, such as DNA sequencing, RNA sequencing, and proteomics (Rather et al. 2023; Varshney et al. 2022).

Key areas of bioinformatics include:

- **Sequence Analysis:** Analyzing DNA, RNA, and protein sequences to identify genes, regulatory elements, and functional motifs.
- **Structural Bioinformatics:** Predicting and analyzing the three-dimensional structures of proteins and nucleic acids.
- **Comparative Genomics:** Comparing genomes of different organisms to understand evolutionary relationships and identify conserved regions.
- **Functional Annotation:** Assigning biological functions to genes and proteins based on experimental and computational evidence.
- **Systems Biology:** Integrating and analyzing data from multiple levels (genomic, transcriptomic, proteomic) to understand complex biological systems.

II. REVIEW OF LITERATURE

Bioinformatics, an interdisciplinary field, emerged in the 1960s when biology intersected with computing. Early pioneers developed tools for DNA and protein sequence analysis. GenBank's 1982 establishment marked a turning point, offering a central repository for DNA sequences. The 1990s saw a genomic data explosion with projects like the Human Genome Project, driving advanced sequence tools. Structural bioinformatics emerged, focusing on biomolecule 3D structures. The post-genomic era shifted attention to gene functions and systems biology. High-throughput technologies generated omics data in the mid-2000s, demanding bioinformatics for analysis. Personalized medicine relied on bioinformatics in the 2010s to customize treatments using genetic insights. Presently, machine learning revolutionizes bioinformatics, aiding protein folding, drug discovery, and disease understanding. As bioinformatics continues evolving, it remains crucial in diverse domains, including medicine, agriculture, and environmental science, unraveling life's molecular intricacies.

1. **Claverie, (2000)**underscores the limitations of current statistical and sequence similarity-based methods and calls for a focus on contextual recognition, incorporating biological and biochemical knowledge. The author urges a new generation of computational biologists with interdisciplinary expertise to steer bioinformatics toward genuine success in deciphering genomic functions in their biological context.This critique questions the direction of modern bioinformatics, noting the paradox of gene count uncertainty despite genomic sequencing completion. It highlights the field's shift towards a phenomenological approach that lacks deeper theoretical exploration. The analogy of DNA as a decipherable text is questioned, advocating for a paradigm shift towards understanding biological meaning through intricate cellular interactions.

2. **Schleif et al. (2006)**demonstrated the small journey of bioinformatics and its applications. Bioinformatics holds promise for innovative medical diagnostics through novel approaches. Preprocessing and advanced analysis methods are essential. This overview spotlights machine learning and neural networks, addressing issues like dimensionality. It showcases mass spectrometry's cancer detection potential. Despite numerous techniques and successes, bioinformatics is in the early stages of integrating diverse subjects: biology, medicine, computer science, chemistry, physics, and mathematics yielding a vibrant research landscape.

3. **Krawetz, (2009)**emphasized the importance of statistics in analyzing vast datasets and shared ground between various fields. "Bioinformatics for Systems Biology" captures the life sciences revolution akin to physics' transformative era. Rapid data generation fuels insights into human biology's systems realm, guided by bioinformatics. The text unifies diverse disciplines, introducing fundamental biological principles for physical and computer scientists while emphasizing statistical importance in data analysis. The shared ground is explored, showcasing tools, databases, and data presentation strategies. Culminating in personalized medicine applications, this text's interdisciplinary approach drives the uncharted frontier of Systems Biology.

4. **Ouzounis, (2012)**examined the historical foundations and future trajectories, highlighting evolving questions, expanding applications across biology, and hinting at potential redefinitions of objectives and boundaries.Bioinformatics and computational biology have evolved significantly in 15 years, becoming pivotal in modern biology. Amid disruptive shifts in science and technology, the field's growth, though fatigue-inducing, is unignorable.

5. **Idrees and Ashfaq (2014)** sketched the protein structure elucidation underscoring its pivotal role in disease comprehension and therapeutic innovation. A testament to its profound impact on life sciences.Bioinformatics, an interdisciplinary marvel, illuminates hidden biological phenomena and protein complexities. Integrating Biotechnology, Information Technology, and sciences like chemistry and math, it unveils molecular insights. Protein exploration spans primary, secondary, and tertiary levels, utilizing bioinformatics tools and databases.

6. **Abbass and Nebel (2020)** discussed the evolution of the chronicle, emphasizing its milestone achievements, notably in CASP assessments, exemplifying its enduring significance in deciphering the complex world of protein structures.for two decades, Rosetta has pioneered protein structure prediction, evolving from an algorithm for ab initio prediction to a comprehensive package for diverse macromolecular modeling tasks. Originally coined in CASP3, it's akin to the Rosetta stone in decoding genetic intricacies. Baker's team expanded into protein design, but Rosetta's essence lies in fragment-assembly prediction, grounded in sequence-structure correlation and fragment utilization.

7. **Chavda et al. (2021)** discussed the modern drug design's interplay in the evolving landscape of drug discovery.In the burgeoning biologics era, with 1,600+ under evaluation and 250+ approved drugs, genomic mapping aided by bioinformatics drives transcriptome and protein engineering studies. Protein drug discovery embarks on a path

from virtual screening to optimization and commercial viability. Rational and computer-aided drug design unite through molecular docking, forming in silico strategies. Ligand-based and structure-based screenings converge, shaping an integrated approach for target discovery and lead optimization.

8. **Tiwary, (2022)**This review underscores the symbiotic blend of sciences, emphasizing how computational skills empower effective exploration in bioinformatics and computational biology, driving innovation at the nexus of biology and technology.Bioinformatics and computational biology epitomize interdisciplinary synergy, demanding adeptness in biological and physical sciences. Proficiency in computational thinking is pivotal for cultivating research prowess in these fields.

9. **Iacoangeli et al. (2022)** covered the rich expanse of Structural Bioinformatics, echoing the profound resonance of Cyrus' research. Beyond his scientific accomplishments, Cyrus' ethos of valuing the journey as much as the destination resonates. He was a true visionary in his open-minded collaborations across backgrounds, nationalities, and disciplines. This collection's contributors extend their heartfelt appreciation to Cyrus for his monumental impact on Bioinformatics.

III. BIOINFORMATICS GATEWAY

Bioinformatics, a captivating discipline at the forefront of modern biological research, finds its core strength in the vast expanse of the Internet an indispensable realm that grants access to a treasure trove of sequence data, molecular analysis tools, and a seamless integration of diverse biological resources. Within this illuminating journey through the chapters of this book, we embark on a voyage that commences with the venerated National Center for Biotechnology Information (NCBI), steward of the esteemed GenBank, and purveyor of an array of captivating bioinformatics tools. Our quest then leads us to the European Bioinformatics Institute (EBI), guardian of the complementary DNA database (EMBL) from the revered European Molecular Biology LaboratoryAccompanying this revelation is the unveiling of the DNA Database of Japan (DDBJ), harmoniously connected with its GenBank and EMBL counterparts through a daily exchange of precious sequence data.Throughout our odyssey, we find ourselves in awe of the ingenious University of California, Santa Cruz (UCSC) genome browser—a beacon of knowledge that guides our understanding of the intricate genetic landscapes. As we delve further into this transformative realm of bioinformatics, we discover the tapestry of interconnected databases—where a rich and expansive network of over 1,000 additional websites reveal their relevance, providing profound insights into the complexities of life's molecular wonders. Each digital thread woven into this vast web of knowledge contributes to our collective pursuit of unlocking nature's secrets and enhancing our understanding of the intricacies of life itself.

IV. BIOINFORMATICS TOOLS

Listing and discussing all bioinformatics tools is an enormous task, as the field of bioinformatics is vast and continuously evolving with new tools and software being developed regularly. Instead, here provide an overview of the major categories of bioinformatics tools and some examples within each category (Brenner, 1998; Gerstein et al. 2007).

Table 1: Bioinformatics tools (Edwards et al. 2009; Chojnacki et al. 2017; Idrees and Ashfaq 2014).

| Tool	Programm	Application
Sequence Analysis Tools	BLAST	For sequence similarity searches
	Clustal Omega	Multiple sequence alignment tool
	HMMER	Profile hidden Markov model searches
Structural Bioinformatics Tools	PyMOL	Visualization and analysis of protein structures
	SWISS-MODEL	Homology modeling of protein structures
	GROMACS	Molecular dynamics simulations
Genomics Tools	Ensembl Genome Browser	Comprehensive genome annotation and visualization
	BEDTools	Genomic interval operations and analysis
	ANNOVAR	Functional annotation of genetic variants
Transcriptomics Tools	DESeq2	Differential gene expression analysis
	StringTie	Transcript quantification and differential expression analysis
	Cufflinks	Transcript assembly and abundance estimation from RNA-Seq data
Proteomics Tools	MaxQuant	Identification and quantification of proteins from mass spectrometry data
	Skyline	Targeted proteomics data analysis
	Proteome Discoverer	Proteomic data analysis suite
Metabolomics Tools	MetaboAnalyst	Comprehensive metabolomics data analysis and visualization
	XCMS	Data processing and peak detection for LC-MS metabolomics
	MassBank	Mass spectral database for metabolite identification
Systems Biology Tools	Cytoscape	Visualization and analysis of biological networks
	BioPAX	Exchange format for biological pathway data
	SBML	Standard for representing computational models in systems biology
Phylogenetic Tools	PhyloBayes	Bayesian inference of phylogenetic trees
	RAxML	Maximum likelihood phylogenetic analysis
	MEGA	Phylogenetic analysis software with a user-friendly interface
ChIP-Seq Analysis Tools	MACS2	Peak calling for ChIP-Seq data
	SICER	Identification of broad domains in ChIP-Seq data
	HOMER	Motif discovery and ChIP-Seq data analysis
Epigenetics Tools	Bismark	Bisulfite mapping and methylation analysis
	DMAP	Differential methylation analysis pipeline
	Segtools	Segmentation analysis of epigenomic data

V. GENOMICS
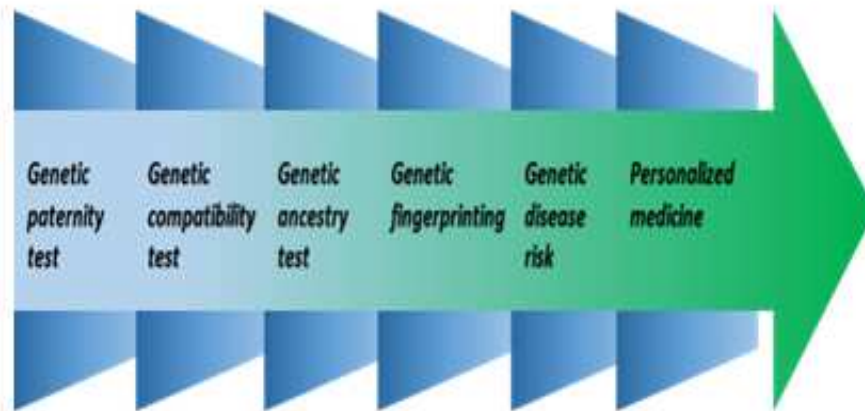
Genomics is the study of an organism's entire DNA sequence, including all of its genes and non-coding regions. It encompasses various approaches to understanding the structure, function, evolution, and organization of genomes. Genomic research has been instrumental in advancing our understanding of genetics, evolutionary biology, and human health (Ideker et al. 2001; Searls, 2000).

Key areas of genomics include:

1. **Genome Sequencing:** Determining the complete DNA sequence of an organism's genome.
2. **Genome Assembly:** Piecing together short DNA sequences obtained from sequencing into a complete genome.
3. **Comparative Genomics:** Studying similarities and differences between the genomes of different species to gain insights into evolutionary processes and gene function.
4. **Functional Genomics:** Investigating gene functions, expression patterns, and regulatory mechanisms on a genome-wide scale.
5. **Population Genomics:** Analyzing genetic variation within and between populations to understand the genetic basis of traits and diseases.
6. **Structural Genomics:** Studying the three-dimensional structures of proteins and other macromolecules encoded by the genome.
7. **Cancer Genomics:** Studying the genomic alterations associated with cancer development and progression.

The integration of bioinformatics and genomics has revolutionized many areas of life sciences, including personalized medicine, agriculture, biotechnology, and environmental research. By leveraging computational tools and big data analysis, researchers can gain valuable insights into the genetic basis of various biological phenomena, leading to numerous practical applications and advancements in our understanding of life itself.

VI. PROTEOMICS
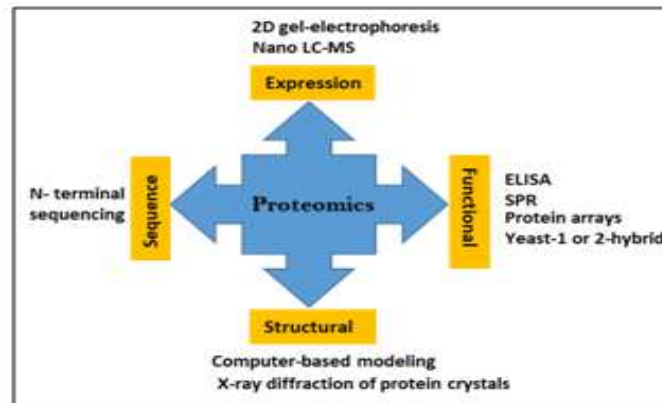
Proteomics is the study of the entire set of proteins produced or expressed by a cell, tissue, or organism. It is a branch of molecular biology that aims to understand the structure, function, and interactions of proteins.Proteins play a crucial role in virtually all cellular processes, including enzymatic reactions, signal transduction, gene regulation, cell structure, and immune responses (Schaffer et al. 2001; Altschul et al. 2005; Yu and Altschul 2005).

Key aspects of proteomics include:

1. **Protein Identification:** Proteomics involves the identification of proteins present in a biological sample. This is typically done using techniques such as mass spectrometry, which can accurately measure the mass and charge of proteins, allowing for their identification and quantification.
2. **Protein Characterization:** Proteomics aims to understand the properties and functions of identified proteins. This includes studying post-translational modifications (PTMs), such as phosphorylation, glycosylation, and acetylation, which can greatly affect a protein's activity and stability.
3. **Protein-Protein Interactions:** Proteomics helps identify protein-protein interactions, revealing how proteins interact and form functional complexes within cells. This information is crucial for understanding cellular pathways and regulatory networks.
4. **Functional Proteomics:** Functional proteomics involves studying how proteins contribute to cellular processes and responses to different stimuli or conditions. This may include investigating changes in protein expression levels under specific conditions or disease states.
5. **Structural Proteomics:** This area focuses on determining the three-dimensional structures of proteins, which is essential for understanding their function and how they interact with other molecules.
6. **Clinical Proteomics:** Proteomics has important applications in clinical research and diagnostics. By analyzing protein expression patterns in biological samples, such as blood or tissue, researchers can identify biomarkers associated with diseases and monitor treatment responses.
7. **Systems Biology:** Integrating proteomics data with genomics, transcriptomics, and other omics data allows for a comprehensive understanding of biological systems, leading to the emerging field of systems biology.

Proteomics plays a significant role in various fields, including medicine, drug discovery, agriculture, and biotechnology. It provides valuable insights into disease mechanisms, drug targets, and the development of personalized medicine approaches. By understanding the complex interactions and functions of proteins within biological systems, researchers can gain a deeper understanding of cellular processes and potentially find solutions to various human health and environmental challenges.

VII. GLOBAL GENOMIC MARKET

The global genomic market was experiencing rapid growth and was expected to continue expanding in the coming years. The genomics market encompasses various technologies, products, and services related to the study and analysis of genomes, including DNA sequencing, gene editing, and other genomic tools and application(databridgemarketresearch.com; fortunebusinessinsights.com).

Some key factors driving the growth of the global genomic market include:

1. **Advancements in DNA Sequencing:** Technological advancements in DNA sequencing techniques, such as next-generation sequencing (NGS), have significantly reduced the cost and time required for genome sequencing. This has made genomic analysis more accessible and affordable, leading to increased adoption in research and clinical applications.
2. **Precision Medicine and Personalized Genomics:** The growing emphasis on precision medicine, which tailors medical treatments based on individual genetic profiles, has fueled the demand for genomic testing and analysis. Personalized genomics aims to provide customized healthcare solutions, including disease risk assessment and treatment selection.
3. **Rising Prevalence of Genetic Disorders:** The increasing prevalence of genetic and hereditary disorders has led to a higher demand for genetic testing and screening. Genomic technologies play a vital role in identifying disease-causing genetic mutations and enabling early diagnosis and intervention.
4. **Biotechnology and Pharmaceutical Applications:** Genomic tools are widely used in biotechnology and pharmaceutical research for drug discovery, target identification, and biomarker development. These applications contribute to the growth of the genomic market.
5. **Agricultural and Environmental Genomics:** Genomics is applied in agriculture to improve crop yields, develop genetically modified organisms (GMOs), and enhance livestock breeding. Additionally, environmental genomics helps in studying biodiversity and understanding ecosystems.
6. **Government Initiatives and Funding:** Many governments around the world have invested in genomics research and initiatives to promote innovation, healthcare advancements, and economic growth.
7. **Direct-to-Consumer (DTC) Genetic Testing:** The availability of DTC genetic testing kits has allowed individuals to access information about their ancestry, health traits, and potential genetic risks, contributing to the expansion of the genomic market.

VIII. CONCLUSION

Bioinformatics, genomics, and proteomics synergize biology and computational science, catalyzing transformative insights. Bioinformatics blends data analysis with biology, unveiling genetic mysteries. Genomics deciphers entire DNA sequences, from genes to non-coding regions, underpinning evolution, genetics, and health. Proteomics studies proteins' roles, interactions, and structures, unraveling cellular intricacies. Their fusion revolutionizes precision medicine, agriculture, and biotech, empowered by global genomics market growth. DNA sequencing leaps, personalized medicine, and CRISPR gene editing fuel their ascent. Computational tools from NCBI, EBI, and beyond drive discovery, from gene functions to protein structures. This triumvirate forms a transformative beacon, fusing biology and technology to illuminate life's complex tapestry, reshaping healthcare, ecology, and innovation.

REFERENCES

[1] Abbass, J., &Nebel, J. C. (2020). Rosetta and the journey to predict proteins' structures, 20 years on. Current Bioinformatics, 15(6), 611-628.

[2] Altschul, S. F., Wootton, J. C., Gertz, E. M., Agarwala, R., Morgulis, A., Schaffer, A. A., and Yu, Y.-K. Protein database searches using compositionally adjusted substitution matrices. FEBS J. 272, 5101–5109 (2005).

[3] Boguski, M. S. Bioinformatics. Curr. Opin. Genet. Dev. 4, 383–388 (1994).

[4] Brenner, S. E. Practical database searching. Bioinformatics: A Trends Guide 1998, 9–12 (1998).

[5] Chavda, V., Sheta, S., Changani, D., &Chavda, D. (2021). New bioinformatics platform-based approach for drug design. Computation in Bioinformatics: Multidisciplinary Applications, 101-120.

[6] Chen, L., Heikkinen, L., Wang, C., Yang, Y., Sun, H. and Wong, G. Trends in the development of miRNA bioinformatics tools. Briefings in bioinformatics, 20(5), 1836-1852 (2019).

[7] Chojnacki, S., Cowley, A., Lee, J., Foix, A. and Lopez, R. Programmatic access to bioinformatics tools from EMBL-EBI update: 2017. Nucleic acids research, 45(1), 550-553 (2017).

[8] Claverie, J. M. (2000). From bioinformatics to computational biology. Genome research, 10(9), 1277-1279.

[9] Edwards, D., Stajich, J. and Hansen, D. eds. Bioinformatics: tools and applications. springer science & business media (2009).

[10] Gerstein, M., Greenbaum, D., Cheung, K., and Miller, P. L. An interdepartmental Ph.D. program in computational biology and bioinformatics: The Yale perspective. J. Biomed Inform. 40, 73–79 (2007).

[11] Goodman, N. Biological data becomes computer literate: Newadvances in bioinformatics. Curr. Opin. Biotechnol. 13, 68–71 (2002).

[12] Henikoff, S. Beyond the central dogma. Bioinformatics 18, 223–225 (2002).

[13] Iacoangeli, A., Marcatili, P., Deane, C., Lesk, A. M., Pastore, A., &Teichmann, S. A. (2022). A Journey Through 50 Years of Structural Bioinformatics in Memoriam of Cyrus Chothia. Frontiers in Molecular Biosciences, 9, 885318.

[14] Ideker, T., Galitski, T., and Hood, L. A new approach to decoding life: Systems biology. Annu. Rev. Genomics Hum. Genet. 2, 343–372 (2001).

[15] Idrees, S., &Ashfaq, U. A. (2014) Journey of Protein from Sequence to Structure: An Innovative Bioinformatics Approach. Bioinformatics and Computational Biology, Biotechnology (6): 129-142.

[16] Krawetz, S. A. (2009). Bioinformatics for systems biology. S. A. Krawetz (Ed.). New York: Humana Press.

[17] Luscombe, N. M., Greenbaum, D., and Gerstein, M. What is bioinformatics? A proposed definition and overview of the field. Methods Inf. Med. 40, 346–358 (2001).

[18] Rather, M.A., Agarwal, D., Bhat, T.A., Khan, I.A., Zafar, I., Kumar, S., Bhat, R.A., Sundaray, J.K. and Qadri, T. Bioinformatics approaches and big data analytics opportunities in improving fisheries and aquaculture. International Journal of Biological Macromolecules, 123549 (2023).

[19] Scha¨ffer, A. A., Aravind, L., Madden, T. L., Shavirin, S., Spouge, J. L., Wolf, Y. I., Koonin, E.V., and Altschul, S. F. Improving the accuracy of PSI-BLAST protein database searches with composition-based statistics and other refinements. Nucleic Acids Res. 29, 2994–3005 (2001).

[20] Schleif, F. M., Villmann, T., Elssner, T., Decker, J., &Kostrzewa, M. (2006). Machine learning and soft-computing in bioinformatics-a short journey. In Applied Artificial Intelligence (pp. 541-548).

[21] Searls, D. B. Bioinformatics tools for whole genomes. Annu. Rev. Genomics Hum. Genet. 1, 251–279 (2000).

[22] Tiwary, B. K. (2022). Introduction to Bioinformatics and Computational Biology. Bioinformatics and Computational Biology: A Primer for Biologists, 1-9.

[23] Varshney, S., Bharti, M., Sundram, S., Malviya, R. and Fuloria, N.K. The Role of Bioinformatics Tools and Technologies in Clinical Trials. In Bioinformatics Tools and Big Data Analytics for Patient Care (pp. 1-16). Chapman and Hall/CRC (2022).

[24] Yu, U., Lee, S. H., Kim, Y. J., and Kim, S. Bioinformatics in the post-genome era. J. Biochem. Mol. Biol. 37, 75–82 (2004).

[25] Yu, Y., K., and Altschul, S. F. The construction of amino acid substitution matrices for the comparison of proteins with nonstandard compositions. Bioinformatics 21, 902–911 (2005).