Journal of Chemical Health Risks

www.jchr.org

JCHR (2023) 13(6), 266-275 | ISSN:2251-6727

Exploring the Significant Contributions of Bioinformatics in Advancing Biotechnology

Morteza Okhovvat[1], Sedigheh Sadat Erfani*[2]

[1]Health Management and Social Development Research Center, Golestan University of Medical Sciences, Gorgan, Iran

[2*Corresponding author] Infectious Diseases Research Center, Golestan University of Medical Sciences, Gorgan, Iran

(Received: 07 October 2023 Revised: 12 November Accepted: 06 December)

KEYWORDS
Bioinformatic,
Biotechnology,
Genomics,
Omics

ABSTRACT:
In the past decade, bioinformatics and biotechnology have seen significant growth. Biotechnology has diverse applications in disease management, sustainable development, agriculture, and bioenergy. The combination of basic science research and data analysis is crucial for understanding key biotechnology questions and reaping economic benefits. In this article, bioinformatics has been tried to be explained in different fields of biotechnology. Applications and applications of bioinformatics in biotechnology is the main discussion of the present study. The primary objective of the present research is to elucidate the significant accomplishments of bioinformatics in the field of biotechnology. Specifically, the aim is to investigate both the fundamental and applied achievements of bioinformatics in biotechnology. This study is a review that encompasses all articles indexed in Ovid, PubMed, Scopus, Science of Web, ProQuest, Embase, and Cochran databases. The search for studies in this field was conducted using PubMed, ISI Web of Science, Scopus, Science Direct, Embase, and Cochrane Library. To facilitate this search, keywords were selected from the Medical Subject Headings (MeSH) database. It is evident that the impact of biotechnology in the modern world is still in its nascent stage. The development of technology capable of replacing current materials and fuels in a sustainable and economically viable manner has just begun. Countries with agriculture-based economies are poised to reap the greatest benefits from this new wave of biotechnology and bioenergy industries.

Introduction

Bioinformatics, an interdisciplinary field combining biological and computer sciences, is often referred to as "computational biology" nowadays. It integrates biology, computer science, and information technology to form a unified field that has expanded into various areas, including drug design, healthcare, plant and animal breeding, and biofuel production [1, 2]. With the advancements in molecular biology, data analysis methods have rapidly developed to interpret the vast amount of information generated by DNA sequencing technologies. This has led to an explosion of genomic and transcriptomic knowledge. High-throughput sequencing methods, such as Illumina sequencing technology and Pacific Biosciences' SMART platform, have revolutionized biological data analysis. Third-generation sequencing technologies, like Oxford Nanopore, have enabled the reading of lengthy sequences. These advancements, coupled with tools like electrophoresis, liquid chromatography (LC), and mass spectrometry (MS), have facilitated the expansion of proteomics and metabolomics.

The biotechnology industry has experienced unprecedented growth, with significant progress in molecular modeling, disease characterization, pharmaceutical discovery, clinical healthcare, forensics, and agriculture. As biotechnology gains recognition and development, bioinformatics has also reached new heights within the biological sciences. It offers numerous applications, including automated genome sequencing, gene identification, prediction of gene function and protein structure, phylogeny, drug design, organism identification, vaccine design, understanding gene and genome complexity, and comprehending protein structure and functionality.

In summary, bioinformatics plays a crucial role in diverse biotechnology domains such as drug design and development, genomics, proteomics, environmental biotechnology, and more. Its applications have far-reaching implications for global economic and social issues.

Methods

A narrative literature review was conducted by searching for relevant articles in Science Direct, PubMed, and Google Scholar using keywords such as bioinformatics and biotechnology. Initially, 100 research papers were identified as relevant studies in these databases. Subsequently, the titles and abstracts of the most relevant studies were screened, and the full text of 26 studies was reviewed.

Overview of achievements

Bioinformatics has emerged as a powerful tool with immense potential in various domains of biotechnology. The collaborative synergy between bioinformatics and biotechnology has significantly enhanced research across diverse areas. Figure 1 demonstrates the substantial impact of bioinformatics on the field of biotechnology, highlighting its main applications. This section offers detailed explanations for each of these applications and provides an overview of notable achievements in bioinformatics that have greatly influenced the field of biotechnology.

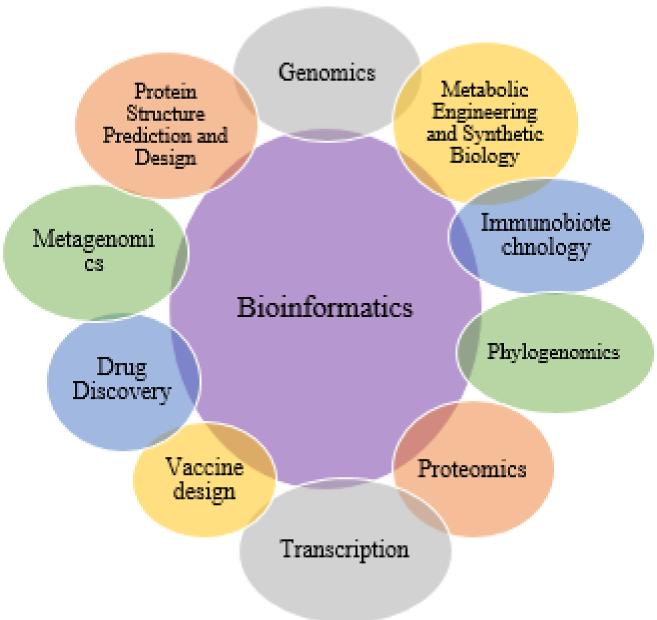

Figure 1: The Utilization of Bioinformatics in Biotechnology

Genomics

Genomics entails the comprehensive study of an organism's entire genome and incorporates genetic elements within its scope. It employs a combination of non-redundant DNA, DNA sequencing methods, and bioinformatics to determine the sequence, assembly, and analysis of genome structure and function. This field generates an immense volume of data encompassing gene sequences, their interactions, and their functions. In managing such extensive and diverse information, bioinformatics assumes a pivotal role.

Through the complete genome sequencing of numerous organisms, bioinformatics has provided both conceptual frameworks and practical methods for deciphering the systemic functional behaviors of cells and organisms [4]. Genomic analysis relies on various DNA extraction, processing, and sequencing methods, including:

1) Whole-genome sequencing, which involves extracting DNA from millions of cells or an individual.

2) Exome sequencing, which focuses solely on the coding regions of DNA targeted by exons.

3) Genotyping-by-sequencing (GBS), which reduces genome complexity using restriction enzymes and selects single nucleotide polymorphisms (SNPs).

4) Epigenomics, which examines inherited transcriptional activation through methylation sequencing.

The choice of sequencing strategy depends on project goals, genome characteristics (such as size and measurement techniques), and population diversity. It is also crucial to select the most suitable sequencing method based on factors like read depth, read length, and cost. Commonly used technologies include Illumina, Pacific Biosciences (PacBio), and Nanopore [7].

Illumina platforms, particularly Hiseq and Miseq, generate paired-end reads ranging from 100 to 250 base pairs and up to 300 base pairs, respectively. Miseq is suitable for small genomes (<20 megabases). On the other hand, PacBio is a third-generation technology that performs real-time sequencing by detecting nucleotide incorporation in a molecule, but it exhibits a higher error rate. In comparison, second-generation technologies demonstrate error rates below 2%. PacBio can generate reads ranging from 2.5 kilobases to 80 kilobases and is often used for sequencing complex genomes in conjunction with other technologies. However, several articles have reported satisfactory results using only PacBio reads for bacterial assembly.

Oxford Nanopore is developing MinION, an affordable and highly promising device capable of producing more than 2 megabases of reads and raw data up to 20 gigabases [5].

The quality of sequencing output is contingent upon the DNA sequencing technology employed. Presently, Illumina stands out as the producer of the highest-quality reads, boasting an average PHRED score of approximately 40. However, it is common to detect low-quality base pairs (PHRED score below 20) at the end of reads. Various tools, including FastQC, are readily available to evaluate the quality of Illumina reads [6].

Comparative genomics, which centers on the study of evolutionary relationships between species through genome comparisons, is particularly concentrated in specific areas of biotechnology. Its primary objective is to identify valuable genetic resources that can support biotechnological advancements. Major gene discovery programs frequently prioritize wild-type or evolutionarily adapted species of interest, specifically targeting them for industrial applications. This approach serves as a potent tool for pinpointing specific traits within a particular species [8].

The urgent need to address climate change and population growth has highlighted the importance of creating sustainable agricultural products. Omics-based research has proven instrumental in enhancing plant breeding techniques. Moreover, animal production for food has the potential to meet the demands of the growing human population, thanks to advancements in our understanding of animal species. Current and emerging methods in animal species, which leverage empirical data or bioinformatics-driven studies, play a crucial role in unravelling the genetic systems underlying complex traits. These methods enable us to make meaningful and accurate biological predictions. Lastly, the next-generation omics tools and methods utilized in various biological sciences can also find valuable applications in the field of veterinary sciences [9].

Metagenomics

Metagenomics is a widely used technique for directly sequencing DNA from an entire microbial community in its natural environment. Whole shotgun metagenomic sequencing (WSM) allows for the comprehensive study of any microbiome, including its population structure and function. The exploration of this diversity is particularly intriguing for uncovering new genes and microorganisms with diverse metabolic capabilities and biological functions that may have relevance in biotechnology.

However, to successfully extract these valuable insights, the utilization of bioinformatics tools is of paramount importance in order to make sense of the vast amount of data generated by metagenomic projects. A WSM project typically begins by generating data from a targeted microbial community. High-throughput sequencing technologies, such as Illumina sequencing, are commonly employed. More recently, long reads from platforms like Oxford NanoPore and Pacific Biosciences have been utilized to enhance the quality of metagenome assembly, thereby enabling the recovery of more complete genomes from the target metagenome.

As previously mentioned, each sequencing platform possesses a certain error rate, underscoring the criticality of quality control and assessment (QC) in metagenomic projects. These projects generally focus on two main aspects: taxonomic characterization, which involves determining the microbial community composition and the relative abundance of each organism within a sample, and functional characterization, which explores the genetic potential of the community [13].

The 16S rRNA target sequencing approach is generally more cost-effective compared to WSM, but it does have limitations [14]. These limitations include low taxonomic resolution, primarily at the genus level, and limited insights into the functional profile of microorganisms. On the other hand, WSM, although more expensive, provides direct information about the genetic and functional potential of the microbiome. It is particularly interesting for identifying new genes and genomes.

Numerous WSM projects have focused on discovering new enzymes and novel biocatalysts with specific properties relevant to biotechnological applications. Various environments, such as marine microbiomes, bovine rumen, soil microbiomes, and other harsh conditions, have been extensively analyzed to uncover new enzymes. These environments are exposed to diverse conditions like pressure, temperature, pH, salinity, and nutrient availability.

Metagenome-derived enzymes have been described for applications in biotechnology, including biofuels, food, chemical, and pharmaceutical industries. Several patents have been filed for the identification and application of enzymes from metagenomes. Examples include nitrile hydratase (EP 1730300 B1), cellulases (WO 2017/081705 A2), amylases (CN 103290039 A), phosphatases (US 8647854 B2), shows esterases (WO 2013/009018 A1), lipases (KR 100997735 B1), and more.

Metagenomic analysis of the Castor candadensis faecal microbiome has revealed novel polysaccharide utilization sites (PUL) related to xylan degradation, specifically related to GH43 enzymes that can be applied to biomass degradation. Additionally, a new Bacteroidetes family consisting solely of microorganisms recovered from ruminants shows great potential for lignocellulose degradation through a group of multi-modular catalytic CAZymes.

Relevant enzymes for biotechnological applications can also be obtained through the enrichment of microbial communities, encouraging resulting consortia to produce higher amounts of specific enzymes of interest. For

Journal of Chemical Health Risks
www.jchr.org
JCHR (2023) 13(6), 266-275 | ISSN:2251-6727

instance, genome recovery of a compost-derived aerobic cellulolytic consortium revealed a multidomain gene cluster of glycosyl hydrolases responsible for 70% of the cellulase activity within the microbial community.

The utility of metagenomics in biotechnology is evident as it enables the identification and exploitation of the applied capacities of microbial communities, transforming them into industrial bioproducts [15].

Protein Structure Prediction and Design

Protein structure prediction and design are essential in bioinformatics and molecular biology research. Understanding the three-dimensional structure of proteins is crucial for uncovering their functions and developing therapeutic interventions. Protein structure prediction focuses on determining the spatial arrangement of atoms within a protein, while protein design aims to modify or create protein structures with specific desired properties.

Experimental methods used for protein structure determination include X-ray crystallography, which involves growing protein crystals and analyzing the diffraction patterns produced by X-rays. Nuclear Magnetic Resonance (NMR) spectroscopy provides valuable insights into protein structures by studying the interactions between atomic nuclei and magnetic fields. Cryo-Electron Microscopy (Cryo-EM) is a cutting-edge technique that allows researchers to visualize protein structures at near-atomic resolution using frozen samples.

Computational methods play a crucial role in protein structure prediction. Homology modeling utilizes known protein structures as templates to predict the structure of a target protein that shares a high degree of sequence similarity. Ab initio methods, on the other hand, predict protein structures from first principles and available experimental data, without relying on known templates. Hybrid approaches combine experimental and computational techniques to improve prediction accuracy.

Protein design encompasses rational design and de novo design. Rational design involves modifying existing protein structures to enhance their properties or create new functions. Computational tools and algorithms aid in identifying key regions for modification and predicting the impact of mutations on protein structure and function. De novo design, on the other hand, involves designing novel proteins from scratch based on desired functions or structures. Computational approaches, such as fragment assembly and optimization algorithms, play a pivotal role in de novo protein design.

Despite significant progress, protein structure prediction and design face challenges. Protein folding remains a complex process that is difficult to accurately model. The vast conformational space and the need to consider environmental factors further complicate the prediction process. However, advancements in machine learning and deep learning techniques have shown promise in improving prediction accuracy and design capabilities.

Metabolic Engineering and Synthetic Biology

Metabolic engineering is a multidisciplinary field that seeks to optimize cellular metabolism for the production of valuable compounds, including biofuels, pharmaceuticals, and industrial chemicals. Bioinformatics plays a vital role in metabolic engineering by offering tools and techniques to analyze, model, and manipulate metabolic pathways. This passage will outline some of the commonly used bioinformatics tools and techniques in the field of metabolic engineering.

Genome-Scale Metabolic Models

Genome-scale metabolic models (GEMs) are computational representations of an organism's metabolism that integrate genomic, biochemical, and physiological information. GEMs offer a comprehensive perspective on cellular metabolism and enable the prediction of metabolic phenotypes under varying conditions. Tools such as Constraint-Based Reconstruction and Analysis (COBRA) and the Systems Biology Markup Language (SBML) aid in the construction, simulation, and analysis of GEMs.

Pathway Databases and Resources

Bioinformatics databases, such as the Kyoto Encyclopedia of Genes and Genomes (KEGG) and MetaCyc, offer valuable information on metabolic pathways, enzymes, and metabolites. These databases enable researchers to explore and identify relevant pathways for metabolic engineering projects. Furthermore, resources like BiGG Models and ModelSEED provide curated metabolic models and tools for metabolic reconstruction, aiding in pathway analysis and design.

Metabolic Pathway Design and Optimization

Bioinformatics tools aid in the design and optimization of metabolic pathways for the production of desired compounds. These tools utilize algorithms such as OptKnock, OptStrain, and OptGene to identify genetic modifications that maximize the production of target compounds. They take into account factors such as enzyme kinetics, thermodynamics, and regulatory constraints to guide genetic engineering strategies.

Flux Balance Analysis (FBA)

Flux Balance Analysis (FBA) is a commonly employed computational approach in metabolic engineering. It utilizes Genome-scale Metabolic Models (GEMs) to predict metabolic fluxes and optimize cellular metabolism for specific objectives, such as maximizing product yield or minimizing by-product formation. FBA takes into account stoichiometric constraints, reaction kinetics, and nutrient availability to determine the optimal distribution of fluxes within metabolic networks.

High-Throughput Omics Data Analysis

High-throughput omics data analysis plays a crucial role in bioinformatics and has made significant contributions to the field of biotechnology. The main aspects include:

a) Data processing and analysis: High-throughput omics technologies generate vast amounts of biological data, including genomics, transcriptomics, proteomics, and metabolomics. Bioinformatics tools and algorithms enable the processing, integration, and analysis of these datasets, allowing researchers to extract meaningful insights and identify patterns or correlations.

b) Biomarker discovery: High-throughput omics data analysis has facilitated the identification of biomarkers, which are specific molecules or genetic variations associated with diseases or biological processes. By analyzing large-scale datasets, bioinformatics helps identify potential biomarkers that can be used for diagnostic, prognostic, or therapeutic purposes in biotechnology.

c) Drug discovery and development: Omics data analysis plays a crucial role in drug discovery and development. By analyzing genomic, proteomic, and metabolomic data, bioinformatics helps identify drug targets, predict drug efficacy, and optimize drug design. This enables the development of more effective and personalized therapies.

d) Systems biology: High-throughput omics data analysis allows researchers to study biological systems as a whole, rather than focusing on individual components. By integrating data from multiple omics levels, bioinformatics helps unravel complex biological networks and understand how different components interact. This systems-level understanding is essential for advancing biotechnology applications.

e) Precision medicine: Bioinformatics analysis of omics data has paved the way for precision medicine, which aims to tailor medical treatments to individual patients based on their genetic makeup, lifestyle, and environmental factors. By analyzing large-scale patient data, bioinformatics helps identify genetic variations associated with diseases and predict treatment responses, leading to more personalized and effective therapies.

High-throughput omics data analysis in bioinformatics has revolutionized biotechnology by enabling the processing, integration, and analysis of large-scale biological datasets. It has significantly contributed to biomarker discovery, drug development, systems biology, and precision medicine, ultimately advancing our understanding and application of biotechnology in various fields.

Transcription

The study of the collection of all messenger RNA molecules is known as transcriptomics. RNA molecules play a crucial role in mediating information between the genome and proteome. They can function as messenger RNAs (mRNAs), regulating gene expression as non-coding RNAs (ncRNAs), or acting as catalysts in biochemical reactions (ribozymes). These molecules are vital components of all living cells and can be studied using high-throughput RNA sequencing technologies (RNA-Seq), microarrays, and other methods. The microarray technique generates a substantial amount of data, with each run producing thousands of data points, and an experiment typically requiring hundreds of runs. To analyze such extensive data, various software packages are utilized [10].

In transcriptomics, DNA microarrays are used to determine the level of mRNA expression in a cell population. The microarray technique generates a large amount of data, and the analysis of such extensive data is performed using several software packages. In this way, bioinformatics is used to analyze the transcriptome, where the mRNA level can be specifically expressed [11].

The combination of genome-wide experiments with in silico approaches paves the way for a more systematic understanding of the molecular mechanisms of transcriptional regulation. Various bioinformatics tools have been developed to help uncover these mechanisms by processing data at different stages: from data collection and storage to the identification of molecular targets, and from DNA motif detection in the regulatory sequences of function-related genes to the identification of relevant regulatory networks. Additionally, the recent generation of a large amount of genome-wide data has attracted experts from different backgrounds to this advanced field of molecular biology [12].

Phylogenomics

Phylogenetics is the study of the evolutionary relationship between individuals or a group of organisms. Taxonomy involves finding these relationships using different

Journal of Chemical Health Risks
www.jchr.org
JCHR (2023) 13(6), 266-275 | ISSN:2251-6727

anatomical methods, which can be time-consuming. Bioinformatics allows us to determine the phylogenetic tree based on sequence order using various methods. Several algorithmic methods have been developed to construct phylogenetic trees for different evolutionary lineages [1].

Unfortunately, the most commonly used phylogenetic methods in biotechnology-based articles are outdated. Previous studies reviewed the phylogenetic inference methods by searching the Web of Science (WoS) using terms such as "bioenergy," "biofuels," "biomass," and "phylogeny." Over the years, different genetic markers have been used depending on the subject of the study. For fungi and bacteria analysis, most strategies utilize 16S, 18S, or ITS (Internal transcribed spacer) sequences. Recent studies have taken a comparative genomic approach, using a large number of phylogenomic inferences of single-copy orthologous genes. However, caution should be exercised when using ribosomal sequences to differentiate species, as a small number of nucleotide differences may not be sufficient. In crop plant reconstruction, distance-based methods like Neighbor-Joining and UPGMA are still widely used, although some studies employ maximum likelihood and Bayesian methods. Coalescent Multispecies methods have been used in only one biotechnology paper for bioenergy applications. In plant studies, even if some inferences are made using maximum likelihood, the Neighbor-Joining approach remains dominant. The prevalence of Neighbor-Joining methods in this field indicates a greater focus on molecular and phenotypic characteristics, while phylogenetic analysis may suffer [16].

Recent papers have analyzed the genome of the yeast Saccharomycotina in a comparative phylogenomic manner, identifying genes and gene families related to fermentation pathways that can enhance the performance of industrial yeast for 2G ethanol production. Most biotechnology articles are related to plants or cell wall genes. Yeast genes associated with distinct metabolic pathways and their evolution are recurring topics. Although these topics may seem unrelated at first glance, they are all aimed at solving problems or improving industrial biofuel production processes. Despite some flawed methods, there is a growing effort to address bottlenecks in plant cell wall processing and yeast fermentation, setting a standard for high-quality work that integrates bioinformatics tools to solve real-world problems. The works of Wohlbach, Riley, and colleagues serve as excellent examples of phylogenetic programs that yield industry-focused results. Wohlbach et al. conducted an exploratory search of various yeast genomes, identifying and testing genes with potential applications in the ethanol industry, such as aldo-keto reductase from Candida tenuis. Subsequently, Riley et al. explored more complex patterns of xylose utilization, suggesting that this pattern is a single-gene trait. They also tested other genes and studied the biological diversity of tRNAs, confirming xylose-fermenting yeasts as part of a clade with transfer tRNAs different from the universal standard for a particular amino acid. Understanding the phylogenetic relationships among the studied targets facilitates the interpretation of common genes, traits, and heterologous expression, as the distance between organisms is related to the adaptations necessary for effective expression. Additionally, evolutionary processes such as positive selection should be considered for rational modifications [17].

Proteomics

The study of protein structure, function, and interactions within a specific cell, tissue, or organism is known as proteomics. It is investigated using genetic, biochemical, and molecular biology techniques. Advanced techniques in biology have led to the accumulation of vast amounts of data on protein-protein metabolism, protein profiles, and protein activity patterns. This data can be effectively managed and accessed using bioinformatics tools, software, and databases. Several algorithms have been developed in the field of proteomics, including two-dimensional gel image analysis, peptide mass fingerprinting, and peptide fragmentation.

To investigate the proteome, various equipment and techniques such as electrophoresis, liquid chromatography (LC), and mass spectrometry (MS) have been developed and continuously improved. These techniques enable quantitative and qualitative analysis, as well as the study of protein structure and interactions. They can be combined to generate large amounts of data. Bioinformatics software and algorithms, when used in conjunction with proteomics techniques, are evolving and must be carefully selected for different purposes to extract the required information quickly and accurately. As a result, various proteomics methods from different fields of biotechnology have been presented.

Proteomics studies in biotechnology have been applied to improve processes and optimize functions in industrial strains. By understanding cellular functions in response to changing production parameters, proteomics helps validate and describe final products. This approach provides valuable biochemical information about known industrial microorganisms. For example, the yeast Saccharomyces cerevisiae, which plays a crucial role in alcoholic

Journal of Chemical Health Risks
www.jchr.org
JCHR (2023) 13(6), 266-275 | ISSN:2251-6727

fermentation, was the first eukaryote to have its complete genome sequenced. Proteomics studies have been instrumental in understanding cell dynamics and communication. Proteomics is also used to determine secretory products in filamentous fungi. For instance, Aspergillus oryzae, commonly used in the food and beverage industry, has cellulosic capacity in association with extracellular enzymes. Proteomics is used to map protein level changes and increase the strain's secretory capacity. Similarly, investigations on Trichoderma reesei, a filamentous fungus widely used in various industries, have focused on understanding enzyme regulation, secretion mechanisms, and product identification. Comparative proteomics is performed among different races of fungi, different carbon sources, or using discovery-based proteomics in various works involving Aspergillus, Penicillium, Fusarium, and others. Another aspect of proteomics analysis is the study of protein glycosylation, which is crucial for protein stability, secretion, and localization.

In biomass resources such as sugarcane, rice, corn, or wheat, proteomics is employed to understand parameters that can affect productivity, such as biotic and abiotic stress, cell wall composition, structural proteomics, and organelle-specific proteomics. The diverse approaches offered by proteomics contribute to the optimization and improvement of many basic industrial processes related to the bioeconomy [18]. However, integrating data from different findings is essential to enrich the generated information and draw relevant biological conclusions.

Bioinformatics plays a crucial role in advancing our understanding of biological processes and the treatment and prevention of genetic diseases. Modern biology and related sciences increasingly rely on bioinformatics techniques. Molecular dynamics (MD) simulations, for example, help us comprehend protein structure at the molecular level, which is not possible through other means [19]. Therefore, bioinformatics holds great potential for future developments in science and technology.

Immunobiotechnology

Immunobiotechnology is one of the fastest growing fields in biotechnology. Biotech-careers.org's Digital World Biology database of biotech employers (over 6,800) has nearly 700 organizations involved in some way with immunology. With the advent of next-generation DNA sequencing and other technologies, immunobiotechnologies require the use of computational technology to decipher large data sets and predict interactions between immune receptors (antibodies/MHC/T cell receptors) and their targets has increased attention. The use of new technologies such as immune profiling, in which large numbers of immune receptors are massively sequenced, and targeted cancer therapies, in which researchers create, engineer, and grow modified T cells to attack tumors, have led to Career advancement and demand for new skills and knowledge in biomanufacturing, quality systems, bioinformatics, and cancer biology. In response to these new demands, Shoreline Community College has begun developing a Biotechnology Safety Certificate. Part of this certificate includes a five-week course (30 hours of hands-on computer lab) in bioinformatics. The immunobioinformatics course includes exercises in immune profiling, vaccine development, and running bioinformatics programs using a command line interface. In immunoprofiling, students will examine T-cell receptor datasets from early-stage breast cancer samples using the immunoSEQ Analyser public server from Adaptive Biotechnologies (Seattle, WA) to learn how T cells communicate between normal tissue, blood, and tumors. are different Then, they used IEBD (Immune Epitope Database) together with Molecular World (Digital World Biology) to predict antigens from sequences and confirm the results to learn the differences between contiguous and non-contiguous epitopes recognized by the T cell receptor. And antibodies are detected, used. Finally, students will use cloud computing (cyVerse) and IgBLAST (NCBI) to explore data from an immune profiling experiment [21].

Vaccine design

Viral diseases and epidemics are increasingly prevalent worldwide, with both new and existing viruses becoming more active. Traditional drugs and vaccines face limitations such as lengthy development times, allergic reactions, the emergence of resistant strains, and other challenges, making them inadequate in combating this threat. However, advancements in computer science, technology, genetics, immunology, and the field of bioinformatics offer a more focused approach to vaccine design. This new paradigm challenges traditional methods and holds the promise of greater effectiveness. The new approach aims to reduce development time, preselect peptide antivirals to avoid allergic reactions, enhance reliability against mutational changes in viral strains, and develop community-specific vaccines by considering immunological status. While bioinformatics provides tools for analysing diverse sequence data, practical constraints currently limit its full potential.

Journal of Chemical Health Risks
www.jchr.org
JCHR (2023) 13(6), 266-275 | ISSN:2251-6727

Bioinformatics plays a crucial role in identifying potential epitope regions in pathogens that elicit the most effective immune response and provide broad population coverage without triggering autoimmunity. In contrast, traditional vaccine design methods take years to achieve similar results and often lack the same level of population coverage as bioinformatics studies.

However, transitioning from computer-based predictions to market-ready vaccines requires further efforts. The theoretically derived peptides must be synthesized and tested in the laboratory using mouse models to confirm the predicted immune responses. Administering synthetic peptides requires determining the optimal carrier proteins and ensuring they do not accumulate in the body or hinder antibody-antigen binding. Adjuvants, such as AS04 in the Cerverix HPV vaccine, can be used to enhance vaccine effectiveness. Despite significant progress, bioinformatics in vaccine design is still in its early stages, and more work is needed. Assigning epitopes for T and B cells heavily relies on accurate genomic and proteomic data, which is not always guaranteed. Reliable data is crucial for successful vaccine target identification.

Mathematical and statistical techniques are fundamental in bioinformatics. To overcome the effects of mutational variation, identifying conserved sequences and regions is essential. Various techniques exist to facilitate this search, but their robustness requires careful analysis. Combining these techniques can yield more accurate and reliable results.

There is a need for more robust techniques that can identify vaccine targets by analysing fewer sequences than current methods. It is fair to say that the world is not yet fully prepared to effectively manage epidemics.

Drug Discovery

Bioinformatics has become increasingly vital in all aspects of drug discovery, assessment, and development. Its significance stems not only from its ability to handle vast amounts of data but also from the utility of bioinformatics tools in predicting, analysing, and interpreting clinical and preclinical findings [22]. Traditional approaches to drug discovery, based on pharmacology and chemistry, face numerous challenges in finding new drugs. The mounting pressure to generate a greater number of drugs within shorter timeframes and with lower risks has sparked significant interest in bioinformatics. This has led to the emergence of a distinct field called computer-aided drug design (CADD). Bioinformatics offers extensive support in overcoming cost and time constraints through various means [23]. It provides a wide range of drug-related databases and software that can be utilized for diverse purposes in the drug design and development process [24].

Discussion

Bioinformatics and its applications have grown significantly since the last decade. Likewise, biotechnology has implications in many industrial fields such as disease management, sustainable development, agriculture, bioenergy, and others. Developing basic science research combined with data analysis in biotechnology studies is essential to understanding important questions in the field, while also providing many economic benefits. Future developments in "omics" technologies, such as readout and metabolite measurement, and computational capabilities, such as artificial intelligence and quantum computing, show promising prospects. In addition, biotechnology has also continuously grown due to the progress in genetic engineering and the quality of DNA synthesis in the past years. The knowledge of the studied organisms has a lot of room for growth, and with most of the biodiversity being unknown, the possibility of new applications in this field is limitless. We believe that the impact of biotechnology in the modern world has just begun, the production of technology capable of replacing all existing materials and fuels in a sustainable and economically appropriate way. Countries that have an economy based on agriculture benefit the most from this new generation of biotechnology and bioenergy industries. Brazil in particular has an excellent opportunity to lead this revolution in biomass exploitation with a focus on first and second-generation bioethanol, as well as the development of new industrial processes that can strengthen the national bioeconomy.

Conclusion

In conclusion, the impact of biotechnology in the modern world is still in its early stages, but the potential it holds is immense. With ongoing advancements, there is a growing possibility of developing technologies that can replace current materials and fuels in a sustainable and economically viable manner. This has far-reaching implications for various industries and sectors, including agriculture, energy, healthcare, and environmental sustainability. Countries that rely heavily on agriculture as a significant part of their economy stand to benefit the most from this new generation of biotechnology and bioenergy industries. By leveraging biotechnological advancements, such nations can enhance their agricultural practices, improve crop yields, develop disease-resistant varieties, and

Journal of Chemical Health Risks
www.jchr.org
JCHR (2023) 13(6), 266-275 | ISSN:2251-6727

optimize resource utilization. Additionally, the integration of bioenergy industries can provide alternative and renewable sources of energy, reducing dependence on fossil fuels and mitigating environmental impact. One such country that holds great potential in this regard is Brazil. With its abundant natural resources, vast agricultural lands, and expertise in bioethanol production, Brazil has a unique opportunity to lead the revolution in biomass exploitation. By focusing on the development of first and second-generation bioethanol, as well as investing in innovative industrial processes, Brazil can strengthen its national bioeconomy and position itself as a global leader in sustainable biotechnology and bioenergy. Overall, the transformative power of biotechnology is still unfolding, and its impact on various sectors and economies is poised to grow exponentially. By embracing and investing in this new generation of biotechnology and bioenergy, countries can pave the way for a more sustainable and prosperous future.

Declarations

This study was approved by the institutional review board (IRB) of Golestan University of Medical Sciences (GOUMS) on 2 August 2022 and the research ethics committee (REC) on 21 August 2021 (approval no. IR.GOUMS.REC.1401.216). All methods were carried out in accordance with relevant guidelines and regulations.

References

1. Heather J.M., Chain B., 2016. The sequence of sequencers: the history of sequencing DNA. Genomics. 1-8.
2. Baxevanis A. D., Bader G. D., Wishart D. S., 2020. Bioinformatics. John Wiley & Sons. 75-143.
3. Rao VS., Das SK., Rao VJ., et al. Recent developments in life sciences research: Role of bioinformatics. 2008. Afr J Biotechnol. 7, 495-503.
4. Muthamilarasan M., Venkata Suresh B., Pandey G., Kumari K., Parida S.K., Prasad M. 2014. Development of 5123 intron-length polymorphic markers for largescale genotyping applications in foxtail millet. DNA Res. 41–52.
5. Singh A., Rao P., Mishra A.K., Yadav Maurya R., Kaur S., Tandon G., 2018. Bioinformatics in next-generation genome sequencing. Current Trends in Bioinformatics: an Insight, Springer, Singapore. 27–38.
6. Hackl T., Hedrich R., Schultz J., Förster F., 2014. large-scale high-accuracy PacBio correction through iterative short read consensus. Bioinformatics. 3004-3011.
7. David M. Kelly E., Kelly S., 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biology. 16, 150-157.
8. Weijde T., Kamei C.L.A., Severing E.I., Torres A.F., Gomez L.D., Dolstra O., Maliepaard C.A., McQueen-Mason S.J., Visser R.G.F., Trindade L.M., 2017. Genetic complexity of miscanthus cell wall composition and biomass quality for biofuels, BMC Genomics. 18, 400-406.
9. Ewels P. A., Peltzer A., Fillinger S., Patel H., Alneberg J., Wilm A., Nahnsen S., 2020. The nf-core framework for community-curated bioinformatics pipelines. Nature biotechnology. 276-278.
10. Conesa A., Madrigal P., Tarazona S., Gomez-Cabrero D., Cervera A., McPherson A., Wojciech M., Daniel J., Gaffney Laura L., Elo Zhang X., Mortazavi A., 2016. A survey of best practices for RNA-seq data analysis. Genome Biology. 13-17.
11. Vyatkina K., Dekker L.J.M., Wu S., VanDuijn M.M., Liu X., Tolić N., Luider T.M., Paša-Tolić L., 2017. De novo sequencing of peptides from high-resolution bottom-up tandem mass spectra using top-down intended methods. Proteomics. 17, 23–24.
12. Altobelli G., 2012. Bioinformatics applied to gene transcription regulation. Journal of Molecular Endocrinology. 49, R51–R59.
13. Frank J.A., Pan Y., Tooming-Klunderud A., Eijsink V.G.H., McHardy A.C., Nederbragt A.J., Pope P.B., 2016. Improved metagenome assemblies and taxonomic binning using long-read circular consensus sequence data. Sci. Rep. 6, 25373.
14. Mitchell A.L., Scheremetjew M., Denise H., Potter S., Tarkowska A., Qureshi M., Salazar G.A., Pesseat S., Boland M.A., Hunter F.M.I., Ten Hoopen P., Alako B., Amid C., Wilkinson D.J., Curtis T.P., Cochrane G., Finn R.D., 2018. EBI Metagenomics in: enriching the analysis of microbial communities, from sequence reads to assemblies, Nucleic Acids Res. 46, D726–D735.
15. Kolinko S., Tachea Y.W., Wu F., Denzel E., Hiras J., Gabriel R., Bäcker N., Chan L.J.G., Eichorst S.A., Frey D., Chen Q., Azadi P., Adams P.D., Pray T.R., Tanjore D., Petzold C.J., Gladden J.M., Simmons, B.A., Singer S.W., 2018. A bacterial pioneer produces cellulase complexes that persist through community. Succession. Nat. Microbiol. 3, 99–107.

16. Shen X., Zhou X., Kominek J., Kurtzman C.P., Hittinger C.T., Rokas A., 2016. Reconstructing the backbone of the Saccharomycotina yeast phylogeny using genome-scale data, G3: Genes|Genomes|Genetics. 109-116.
17. Muthamilarasan M., Prasad M., 2015. Advances in Setaria genomics for genetic improvement of cereals and bioenergy grasses. Theor. Appl. Genet. 128, 1-14.
18. Fonseca J. G., Calderan-Rodrigues M.J., de Moraes F.E., Cataldi T.R., Jamet E., Labate C.A., 2018. Cell wall proteome of sugarcane young and mature leaves and stems. Proteomics 18, 1700129.
19. Bhardwaj R., SharmaM N., Agrawal N., 2021. Bioinformatics and Its Application Areas. Computation in BioInformatics: Multidisciplinary Applications. 121-137.
20. Alonso A., Marsal S., Julià A., 2015. Analytical methods in untargeted metabolomics: state of the art in 2015, Front. Bioeng. Biotechnol. 23.
21. Smith T., Porter S., Kovarik D., 2019. Immuno-Biotechnology and Bioinformatics in Community Colleges. Journal of Biomolecular Techniques: JBT. 20-45.
22. Nandy A., 2019. Bioinformatics in Design of Antiviral Vaccines. Encyclopedia of Biomedical Engineering. 280–290.
23. Wishart DS., 2005. Bioinformatics in drug development and assessment. Drug Metabol Rev. 37, 279-310.
24. Katara P., 2013. Role of bioinformatics and pharmacogenomics in drug discovery and development process. Netw Model Anal Health Inform Bioinform. 2, 225-230.