ISSN (O): 2581-8740 | ISSN (P): 2581-8732
Morbia *et al.*, 2023
DOI: 10.21276/SSR-IIJLS.2023.9.2.4

Review Article

open access

Review on Applicability of Bioinformatics in Current Research and Database Management

Ishani Morbia[1], Richa Dubey[2]*, Shivangi Mathur[3]

[1]Research Scholar, Department of Biotechnology, Indian Institute of Technology, Gandhinagar, Gujarat, India

[2]Assistant Professor, Department of Microbiology, President Science College, Affiliated to Gujarat University, Shayona Campus Ahmedabad, Gujarat, India

[3]Assistant Professor, Department of Biotechnology, President Science College, Affiliated to Gujarat University, Shayona Campus Ahmedabad, Gujarat, India

***Address for Correspondence:** Dr. Richa Dubey, Assistant Professor, Department of Microbiology, President Science College, Affiliated to Gujarat University, Shayona Campus Ahmedabad, Gujarat, India
E-mail: richa@presidentsciencecollege.org

Received: 23 Sep 2022/ Revised: 13 Nov 2022/ Accepted: 26 Jan 2023

ABSTRACT

A generation of new science has evolved with the development of bioinformatics and computational biology which have molecular biology as an integrated part. In the past decade, technological advances have promoted a prominent development in expertise and knowledge in the molecular basis of phenotypes. In Bioinformatics, biological data is evaluated by computational science and processed in a more statistical and meaningful way. It includes the collection classification storage and evaluation of biochemical and organic statistics using computers in particular as implemented in molecular genetics and genomics. Computational Biology and Bioinformatics are emerging branches of science and include the use of techniques and concepts from informatics statistics, mathematics, chemistry, biochemistry, physics, and linguistics. Therefore, bioinformatics and computational biology have sought to triumph over many challenges of which a few are listed in this overview. This evaluation intends to provide insight into numerous bioinformatics databases and their uses in the analysis of biological records exploring approaches emerging methodologies strategies tools that can provide scientific meaning to the information generated.

Key-words: Data analysis, Databases, Genomics, Sequence analyses, Systems biology

INTRODUCTION

Biological science has evolved unprecedently with advances in technology, which has generated a large amount of 'omic' data [1]. Making sense of this large amount of data is a great challenge. Bioinformatics aims at developing tools and databases to facilitate researchers in understanding the functionality of the raw data [2].

As the data that is generated is heterogenous, it becomes quite important to segregate it into different databases. Also, various tools need to be developed to search and mine these databases.

The application of computational tools is to organize, analyze, understand, visualize, and store information associated with biological macromolecules (Fig. 1). This review aims to present a brief overview of these tools and databases and their respective utilities in various aspects. We also seek to highlight various areas that bioinformatics has given rise to and aided too.

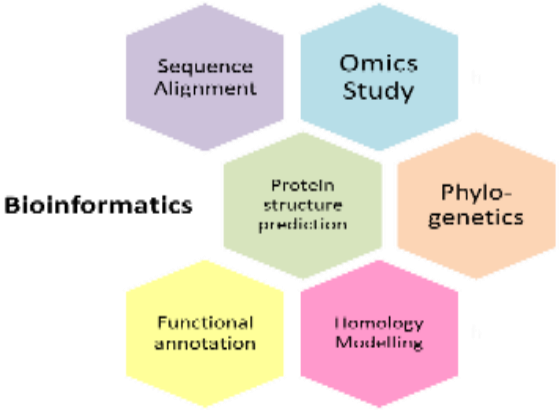

Fig. 1: Applied approaches of Bioinformatics

How to cite this article
Morbia I, Dubey R, Mathur S. Review on Applicability of Bioinformatics in Current Research and Database Management. SSR Inst. Int. J. Life Sci., 2023; 9(2): 3195-3205.

Access this article online
https://iijls.com/

ISSN (O): 2581-8740 | ISSN (P): 2581-8732
Morbia *et al*., 2023
DOI: 10.21276/SSR-IIJLS.2023.9.2.4

Organization of Information

Segregation into Databases- To make biological information (DNA, RNA, and Protein sequences) available for research, it is necessary to store them in an organized way. Primary databases are a collection of results of experimental databases, whereas Secondary databases are a compilation and interpretation of data obtained from primary databases [3]. GenBank at NCBI, DNA Database of Japan (DDBJ) and European Molecular Biology Laboratory (EMBL) are the main primary databases [2]. These databases share the deposited information with each other on daily basis [3]. Protein Information Resource (PIR), UniProt/ SwissProt, Protein Data Bank (PDB), and Prosite are secondary databases [4].

Tools and Database

Gene Identification and Sequence Analyses- Sequence analyses refer is the understanding of different aspects of biomolecules like nucleic acids or proteins, which gives unique function to it. First, the sequences of the respective molecule(s) are taken from public databases. They are then subjected to various tools for refinement and prediction of their features such as function, structure, evolutionary history, or identification of homologues [5]. The choice of tool to be used depends on the nature of the analysis to be done (Table 1).

Table 1: Primary sequence analyses tools

Tools	Utility
BLAST Basic Local Alignment Search Tool	It is an algorithm for comparing DNA, RNA, protein, or amino acid sequences based on identity. https://blast.ncbi.nlm.nih.gov/Blast.cgi
ORF Finder Open Reading Frame Finder	It is a program that identifies all open reading frames or the possible protein-coding regions in a sequence. https://www.ncbi.nlm.nih.gov/orffinder/
HMMER Hidden Markov Models	Identification of homologous protein and nucleotide sequences by performing sequence alignments. http://hmmer.org/
ProtParam	Various physico-chemical properties of proteins can be computed using this tool. https://web.expasy.org/protparam/
novoSNP Single Nucleotide Polymorphisms	Single nucleotide polymorphisms in the DNA can be found using this tool.
Clustal Omega	This tool enables us to perform multiple sequence alignments. https://www.ebi.ac.uk/Tools/msa/clustalo/
Sequerome	Sequence profiling can be performed using this tool. https://www.bioinformatics.org/sequerome/wiki/Main/HomePage
JIGSAW	Genes and predict the splicing sites can be found using this tool. http://www.cbcb.umd.edu/software/jigsaw/
Softberry	Animal, plant, and bacterial genomes can be annotated using this tool and the structure and function of RNA and proteins can also be predicted. http://www.softberry.com/
PPP Prokaryotic Promoter Prediction Tool	Promoter sequences lying upstream of bacterial genes can be predicted using this tool. http://bamics2.cmbi.ru.nl/websoftware/ppp/ppp_start.php
WebGeSTer Web Genome Scanner for Terminators	Transcription terminator sequences are contained in this database, which helps in the prediction of termination sites of the genes during transcription. http://pallab.serc.iisc.ernet.in/gester/dbsearch.php
Genscan	Predicts intron and exon sequences within the genome. http://hollywood.mit.edu/GENSCAN.html
Virtual Footprint	Allows recognition of single or composite DNA patterns. Enables prediction of genome-based regulons and analysis of individual promoter regions. http://www.prodoric.de/vfp/

ISSN (O): 2581-8740 | ISSN (P): 2581-8732
Morbia *et al.*, 2023
DOI: 10.21276/SSR-IIJLS.2023.9.2.4

Phylogenetic analyses- Phylogenetic analyses are used to infer evolutionary relationship among a group of related molecules or organisms, for the prediction of unknown functions, to determine gene flow, and to establish genetic relatedness. This can then be used in creating a phylogenetic tree. The principle of phylogeny is to group living organisms according to the degree of similarity: the higher the similarity, the closer the organisms would appear on the tree. A phylogenetic tree can be constructed by the following methods: distance methods, parsimony methods, and likelihood methods [6] (Table 2).

Table 2: Phylogenetic Analysis Tools

Tools	Utility
MOLPHY M**ol**ecular P**hy**logenetics	The tool is based on the maximum likelihood method for phylogenetic analyses. https://sbgrid.org/software/titles/molphy
PHYLIP P**hyl**ogeny Inference Package	It is a package of 35 portable computational phylogenetic programs. http://evolution.genetics.washington.edu/phylip/install.html
MEGA Molecular Evolutionary Genetic Analysis	This tool enables the construction of phylogenetic trees to find evolutionary relationships. https://www.megasoftware.net/
Treeview	Software to view the phylogenetic trees can be viewed with the help of this software, with an alternative of changing view. https://treeview-
PAML Phylogenetic Analysis by Maximum Likelihood	It analyzes phylogenetic relations based on maximum likelihood. http://abacus.gene.ucl.ac.uk/software/paml.html
Jalview	It helps in the refinement of multiple performed alignments. http://www.jalview.org/development/Version-Archive

Sequence Databases- With the advancement of high throughput sequencing techniques, a massive amount of data is generated every day. To make this data freely available to the scientific community, Primary, Secondary, or Composite databases are constructed. The data in a primary database is experimental, a secondary database contains curated information and a composite database contains information from different primary sources (Table 3).

Genome Sequence Databases- The GenBank, built by the NCBI, collects genome sequences of over 2,50,000 species. Each sequence carries information about the literature, bibliography, organism, and a set of various other features, which include coding regions, promoters, untranslated regions, terminators, exons, introns, repeat regions, and translations (Table 4).

Table 3: Nucleotide Sequence Databases

Databases	Utility
DDBJ DNA Data Bank of Japan	It is an integral member of the International Nucleotide Sequence Database Collaboration (INSDC) that collects DNA sequences. https://www.ddbj.nig.ac.jp/index-e.html
GenBank	It is a member of the International Nucleotide Sequence Database Collaboration (INSDC) and is an annotated collection of all publically available nucleotide sequences. https://www.ncbi.nlm.nih.gov/genbank/
European Nucleotide Archive	It is a collection of information related to experimental workflows based on nucleotide sequencing and a comprehensive record of sequence assembly information and functional annotation. https://www.ebi.ac.uk/ena
Rfam RNA Families	A collection of RNA families, each represented by multiple sequence alignments, consensus secondary structures and covariance models. https://rfam.xfam.org/

ISSN (O): 2581-8740 | ISSN (P): 2581-8732
Morbia *et al*., 2023
DOI: 10.21276/SSR-IIJLS.2023.9.2.4

Table 4: Genome Sequence Databases

Databases	Utility
Ensemble	It contains annotated genomes of eukaryotes including humans, vertebrates, and other model organisms. https://m.ensembl.org/index.html
PIR Protein Information Resource	It is the largest, most comprehensive, annotated protein sequence database in the public domain. https://proteininformationresource.org/

Protein Sequence Databases- The most significant protein sequence databases are SWISS-PROT (Swiss Protein) Databank, TrEMBL (translation of DNA sequences in EMBL), UniProt (Universal Protein Resource), PIR (Protein Information Resource) and wwPDB (worldwide Protein DataBank) [7] (Table 5).

Table 5: List of protein sequence databases

Databases	Utility
SWISS PROT	It is a part of UniProt knowledgebase that consists of annotated protein sequences. http://www.ebi.ac.uk/swissprot/
Protein Data Bank	It consists of experimentally-determined structures of nucleic acids and proteins. https://www.rcsb.org/
Uniprot	It is one of the biggest collections of protein sequences. https://www.uniprot.org/
Prosite	Collection of protein families, conserved domains, and actives sites of proteins. http://www.expasy.org/prosite/
PRIDE PRoteomics IDEntification Database	It is a public data repository of mass spectrometry-based proteomics data, containing functional characterization and post-translation modification of proteins and peptides. https://www.ebi.ac.uk/pride/
Pfam Protein Families	It is a database of protein families. https://pfam.xfam.org/
InterPro	Collection of protein families, domains and functional sites for the functional characterization of new protein sequences. https://www.ebi.ac.uk/interpro/

Table 6: Miscellaneous Databases

Databases	Utility
Reactome	It is a database of reactions, pathways and biological processes largely focused on humans and certain specific organisms. https://reactome.org/
TAIR The Arabidopsis Information Resource	It is a community resource and online model organism database of genetic and molecular biology data for the model plant *Arabidopsis thaliana*. https://www.arabidopsis.org/
Medherb	It is an interactive database and analysis resource for medicinally important herbs.
Textpresso	It is an online literature search and curation platform that enables biocurators to mine full-text literature searches of model organism research and to identify new allele and gene names and human disease gene orthologs. http://www.textpresso.org/tpc
DictyBase	Database for *Dictyostelium discoideum*. http://dictybase.org/

ISSN (O): 2581-8740 | ISSN (P): 2581-8732
Morbia *et al*., 2023
DOI: 10.21276/SSR-IIJLS.2023.9.2.4

Table 7: Signaling and Metabolic pathway Databases

Databases	Utility
CMAP Complement Map Database	It is a resource that uses transcriptional expression data to probe the relationship between diseases, cell physiology and therapeutics and thus generate gene expression profiles. http://gmod.org/wiki/CMap
PID Pathway Interaction Database	It is a growing collection of human signalling and regulatory pathways curated from peer-reviewed literature. It can be used to study various cellular pathways, especially those related to cancer. http://pid.nci.nih.gov
KEGG Kyoto Encyclopedia of Genes and Genomes	It is a collection of manually drawn pathway maps representing molecular interaction, reaction and relation networks for metabolism, cellular processes, human diseases, drug development, organismal processes, environmental information processing and genetic information processing. https://www.genome.jp/kegg/pathway.html
HMDB Human Metabolome Database	It contains detailed information about small molecule metabolites found in the human body. It is intended to be used in applications in metabolomics, clinical chemistry, and biomarker discovery. The database is designed to contain or link three kinds of data: 1) chemical data, 2) clinical data and 3) molecular biology/biochemistry data. http://www.hmdb.ca/
SGMP Signalling Gateway Molecule Pages	It provides structured data on proteins which exist in different functional states participating in signal transduction pathways. www.signaling-gateway.org/molecule

Protein structure and function prediction Databases- Proteins must fold up into a three-dimensional (3D) structure to become biologically active. So, insight into protein 3D structure is required to know its function. 3D structures are normally determined by X-ray crystallography or NMR. But as these techniques are costly, difficult, and time-consuming, a protein's 3D structure can be predicted using various bioinformatics tools. These approaches help in the easy identification of the secondary structure of protein sequences like helices, sheets, domains, strands, and coils. The most widely used approach to predict the 3D structure of a protein molecule is comparative modelling. In this approach, a related known sequence (with at least 30% sequence identity with target protein) is selected to predict the unknown structure [8]. The below given link is a list of protein prediction tools, http://www.biologie.unihamburg.de/bonline/library/genomeweb/GenomeWeb/prot-2-struct.html (Table 8).

Table 8: Protein structure and function prediction tools

Tools	Utility
PHD	It is a neural network system to predict protein secondary structure, relative solvent accessibility and transmembrane helices. https://npsa-prabi.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSAHLP/npsahlp_secpredphd.html
MODELLER	It is used for homology or comparative modelling of protein 3-D structures. https://salilab.org/modeller/
RaptorX	It facilitates secondary, tertiary and contact prediction for protein sequences without close homologs in the Protein Data Bank. http://raptorx.uchicago.edu/
CATH	Based on Class, Architecture, Topology & Homology, it is a hierarchical domain classification of protein structures in the PDB. https://www.cathdb.info/

ISSN (O): 2581-8740 | ISSN (P): 2581-8732
Morbia *et al.*, 2023
DOI: 10.21276/SSR-IIJLS.2023.9.2.4

Phyre & Phyre 2 Protein Homology/Analogy Recognition Engine	It investigates known homologues, builds a hidden Markov model (HMM) of the targeted sequence based on the detected homologues and scans it against a database of HMMs of known protein structures. http://www.sbg.bio.ic.ac.uk/~phyre2/html/page.cgi?id=index
JPred	It is a protein secondary structure prediction server. Also, it predicts solvent accessibility and coiled regions. http://www.compbio.dundee.ac.uk/jpred/
HMMSTR Hidden Markov Model for local sequence STRucture	It is a hidden Markov model to predict sequence-structure correlations in proteins. http://www.bioinfo.rpi.edu/~bystrc/hmmstr/server.php
APSSP 2 Advanced Protein Secondary Structure Prediction Server	Predicts the secondary structure of proteins from their amino acid sequence. http://crdd.osdd.net/raghava/apssp/

Molecular interactions Databases- Discovering interaction among molecules is important to elucidate their biological function. Protein-protein interactions are vital for cellular activities like signalling, transportation, metabolism, etc. Bioinformatics can predict protein-protein interactions without the involvement of costly, and time-consuming methods like X-ray crystallography and Nuclear Magnetic Resonance (NMR) spectroscopy. The parameters influencing protein-protein interactions are then studied [9]. A list of selected tools to study protein-protein interactions is given in Table 9.

Table 9: Molecular Interactions study tool

TOOLS	UTILITY
PathBLAST	It is a network alignment and search tool for comparing protein interaction networks across species to identify protein pathways and complexes that have been conserved by evolution. http://www.pathblast.org/
AutoDock	It predicts protein-ligand interaction. http://autodock.scripps.edu/
STRING Search Tool for the Retrieval of Interacting Genes/Proteins	It is a database of known and predicted protein-protein interactions. https://string-db.org/
BIND Biomolecular Interaction Network Database	It defines the molecular interaction of proteins and bio-complexes. http://bind.ca
IntAct	It is a database for the storage, presentation, and analysis of protein interactions, both in textual and graphical formats. https://www.ebi.ac.uk/intact/
CFinder	It is a program for locating and visualizing overlapping, densely inter-connected groups of nodes in undirected graphs and allowing the user to easily navigate between the original graph and the web of these groups. It can be used to predict the function of a single protein and to discover novel modules. http://www.cfinder.org/
HADDOCK High Ambiguity Driven DOCKing	It can deal with multiple molecules (for docking), a capability that will be required to build large macromolecular assemblies. https://haddock.science.uu.nl/
MOE	It is an integrated drug discovery software. It tracks design ideas and ligand

ISSN (O): 2581-8740 | ISSN (P): 2581-8732
Morbia *et al*., 2023
DOI: 10.21276/SSR-IIJLS.2023.9.2.4

Molecular Operating Environment	modifications with property models, produces correlation plots to visualize structure, property, activity relationships and visualize hydrophobic and charged protein surface to study aggregation-prone regions. https://www.chemcomp.com/Products.htm
MIMO Molecular Interaction Maps Overlap	It offers a flexible and efficient graph-matching tool for comparing complex biological pathways.
Gremlin	It can be used for multiple network alignment that allows the generalization of existing alignment scoring schemes and the location of conserved network topologies. http://gremlin.bakerlab.org/index.php
SMART Simple Modular Architecture Research Tool	Used for the identification and analysis of protein domains within protein sequences. http://smart.embl-heidelberg.de/
MCODE Molecular COmplex Detection	It is a graph theoretic clustering algorithm that detects densely connected regions in large protein-protein interaction networks that may represent molecular complexes. https://baderlab.org/Software/MCODE

Drug designing Databases- As the traditional process of drug discovery is quite slow and expensive, bioinformatics tools have been developed to achieve the same. The process can be divided into four different steps: identification of drug target, validation of target, lead identification, and lead optimization [10]. The target is a small biomolecule upon which the drug molecule acts to produce a desired effect. So, the first step in the drug-designing process is the identification of a target. Many databases have been developed for the search for new drug targets. After the selection of potential targets, the role of those targets in a particular disease is studied. This is called target validation. Bioinformatics tools for modelling enable the prediction of the efficiency of compounds to bind at a particular site [11]. Then a certain compound-lead compound is to be found which can alter the action of the target. Bioinformatics tools allow the virtual screening of a large number of compounds that could manipulate a protein. Many times, the identified compound does not have the required properties, but it can be 'refined' to produce the desired effect with reduced side effects. This process is called 'lead optimization' (Table 10).

Table 10: Drug-Target interaction study databases

Databases	**Utility**
Therapeutic Target Database	It is a database to provide information about known and explored therapeutic protein and nucleic acid targets, the targeted disease, pathway information and corresponding drugs directed at each of these targets. http://bidd.nus.edu.sg/group/cjttd/
Drug Bank	It is a comprehensive database containing information on drugs and drug targets. It combines detailed drug data i.e. chemical, pharmacological and pharmaceutical with comprehensive drug target information i.e. sequence, structure and pathway. https://www.drugbank.ca/
DrugPort	It provides an analysis of the structural information available in the PDB, relating to drug molecules and their protein targets. https://www.ebi.ac.uk/thornton-srv/databases/drugport/
chEMBL	It is a manually curated database of bioactive molecules with drug-like properties. It brings together chemical, bioactivity and genomic data to aid the translation of genomic information into effective new drugs. https://www.ebi.ac.uk/chembl/

ISSN (O): 2581-8740 | ISSN (P): 2581-8732
Morbia *et al.*, 2023
DOI: 10.21276/SSR-IIJLS.2023.9.2.4

MATADOR Manually Annotated Targets And Drugs Online Resorce	It is a database for protein-chemical interactions. It differs from DrugBank in its inclusion of as many direct and indirect interactions as we could find. DrugBank usually contains only the main mode of interaction. http://matador.embl.de/
TDR Target Database Tropical Disease Research	It facilitates rapid identification and prioritization of molecular targets for drug development, focusing on pathogens responsible for neglected human diseases. It integrates pathogen-specific genomic information with functional data i.e. expression, and phylogeny for genes collected from various sources. https://tdrtargets.org/
TB Drug Target Database	It contains information on anti-tubercular drugs and target proteins for the treatment of Tuberculosis. https://www.bioinformatics.org/tbdtdb/
PDTD Potential Drug Target Database	It associates informatics data with structural database of known and potential drug targets. It focuses principally on drug targets with known 3-D structures.

Molecular dynamic simulation databases- Biological activities occur due to molecular interactions in a time-dependent manner. The time dependency of a molecule can be studied bioinformatics tools called Molecular Dynamics Simulations (MDS). These tools provide detailed information on fluctuations, dynamic cellular processes, and conformational changes of proteins and nucleic acids. They also help in determining structures from experimental approaches like XRD and NMR spectroscopy [12] (Table 11).

Table 11: Molecular Simulation study tools

TOOLS	UTILITY
Discovery Studio	It is a suite of software for simulating small molecules and macromolecular systems, ligand design, pharmacophore modelling, structure-based design, macromolecule design and validation, macromolecule engineering and predictive toxicity. https://www.3dsbiovia.com/
FoldX	It can be used for the prediction of the effect of point mutations or human SNPs on protein stability or protein complexes and to design proteins to improve stability or modify affinity or specificity. http://foldxsuite.crg.eu/
Abalone	It is a molecular modelling program for performing biomolecular dynamics simulations of proteins, DNA, and ligands. http://www.biomolecular-modeling.com/Abalone/index.html
AMBER Assisted Model Building with Energy Refinement	It is a set of molecular mechanical force fields for the simulation of biomolecules. https://ambermd.org/
Ascalaph	It is a program for molecular building, graphics, dynamics, and optimization, with an interface to quantum chemistry. http://www.biomolecular-modeling.com/Ascalaph/Ascalaph_Designer.html

Applications of Bioinformatics Databases

Human Genome Project- Human Genome Project (HGP) was aimed towards sequencing the human genome and mapping every gene on every chromosome and developing tools for storing and analyzing this information. HGP employed the shotgun sequencing technique for whole genome sequencing. The enormous amount of data that was generated during this process was segregated, curated, and stored in various functional bioinformatics databases.

e.g. Functional Mapping: Agricultural, evolutionary, and biomedical genetic research is requiring the knowledge of genetic controls governing various phenotypes. Quantitative trait loci (QTLs) responsible for a complex trait can be known using a statistical mapping framework, called functional mapping [8,13].

ISSN (O): 2581-8740 | ISSN (P): 2581-8732
Morbia *et al*., 2023
DOI: 10.21276/SSR-IIJLS.2023.9.2.4

Oncology- Oncology is the study of tumour cells and tumour environment. It is a big challenge to discover the molecular and cellular mechanisms underlying tumour metastasis. Analysing alterations of protein levels in the tumour and correlating it to metastasis helps in facilitating the development of therapeutic strategies and clinical management of cancer. Biomarker prediction and discovery also remain an important aspect here [14].

e.g. **The Cancer Genome Atlas:** The Cancer Genome Atlas (TCGA) holds tumour gene expression data, along with clinical information, which enables researchers to gather information on prominent genomic alterations occurring during the development and metastasis of a tumour.

Gene therapy- Gene therapy is a method of efficient introduction of a functional gene into the cells of the patient to cure diseases related to the deficiency or over-production of that gene product. These procedures primarily require knowledge of the organism's annotated genome, which is provided by bioinformatics [15].

SNP Detection- A single nucleotide polymorphism (SNP) results due to variation of a single nucleotide at a particular position in the genome. It has been established that SNPs are associated with the susceptibility of the individual to specific diseases. Human genome sequences shed light on such SNP data associated with certain diseases and have led towards the development of predictive preventive personalized medicine [8].

Personal medicine- Personalized medicine is based upon an individual's genetic makeup to decide the amount and type of medications to be prescribed for the prevention and treatment of disease [16]. Translational bioinformatics is a field which deals with this area of healthcare. Research in personalized medicine aims to discover solutions based on the susceptibility profile of everyone [17].

RNA Sequencing- Genome-wide gene expression and regulatory mechanisms underlying basic physiological traits of various human pathologies are nowadays studied using RNA-Seq experiments. But, as these are complex analyses, the processing of the obtained data requires the assistance of various bioinformatics tools [8].

BBB Permeation- Prediction of blood-brain barrier (BBB) permeation is vital for designing drug molecules acting on the central nervous system (CNS). The process of permeation is complicated as compounds can cross the BBB both by passive diffusion and/or active transport. Hence, as an alternative to invasive animal experiments, in silico-screening methods have been developed for designing central nervous system active drugs by establishing their BBB permeation [8].

Agriculture- Stressful conditions lead to reduced plant growth, delayed seed germination, and decreased crop yield. Organ-specific proteomic analyses can be used to identify proteins that accumulate in plants under such conditions [18]. These conditions can then be subjected to genetic engineering to produce stress-resistant plant varieties [19].

Insect Resistance- Insect resistance was introduced in many plants by incorporating certain genes. An insect-killing gene was isolated from the genome of a bacteria called Bacillus thuringiensis and was incorporated into plants to make them insect-resistant. Corn, cotton, brinjal, soybean and potatoes have been made insect resistant so far.

Nutritional Quality- Increasing population demands a higher supply of food, but as agricultural land is limited, the solution to overcome this issue is to produce nutritionally enriched and enhanced food [20]. Golden rice is an important achievement in this area. Here the genes to increase Vitamin A levels are increased in the crop. This has solved the problem of malnutrition quite well [21].

Radioactive waste clean-up- Bioinformatics tools are important to understand various metabolic pathways [22]. The bio-degradative pathways in the bacteria Deinococcus radiodurans were explored using these tools. It was then used to break down organic chemicals, solvents, and heavy metals in radioactive waste sites.

Forensic Science- Forensic science includes the study regarding identification and relatedness of individuals. Conventional techniques include fingerprinting and others. These have now advanced to DNA fingerprinting techniques, which use bioinformatics tools and techniques [23]. DNA fingerprinting works on the principle of comparison of repetitive DNA sequences which are

ISSN (O): 2581-8740 | ISSN (P): 2581-8732
Morbia *et al.*, 2023
DOI: 10.21276/SSR-IIJLS.2023.9.2.4

unique to everyone. Criminal databases store DNA profiles of respective individuals to be compared [24].

Bioenergy/Biofuels- Bioinformatics aids in the understanding of biofuel-producing pathways. Recent studies in algal genomics, along with other 'omics' approaches, have proved to be potential targets in the development of genetically engineered microalgal strains producing biofuels [25].

Antibiotic resistance- *Enterococcus faecalis* is known to cause infection, attributing to a virulence region comprising of antibiotic-resistant genes contributing to the bacterium's transformation from a harmless gut bacterium to a pathogen. The Discovery of such useful biomarkers for detecting pathogenic strains can establish controls to prevent the spread of infection.

CONCLUSIONS

Bioinformatics aids modern-day biology by sorting big biological data into functional databases and uncovers various aspects of different biomolecules. It provides scopes for the development of crucial fields such as drug development and screening, genetic engineering, genome annotation and others.

There is merely any area which remained untouched by bioinformatics and computational biology and thus the bright future of Biology will have a lot to owe to it.

Acknowledgement

The authors gratefully acknowledge guides and mentors from President Science College for their valuable guidance and support.

CONTRIBUTION OF AUTHORS

Research article concept- Dr. Richa Dubey
Research design- Ms. Ishasni Morbia
Supervision- Dr. Shivangi Mathur
Data analysis and interpretation- Ms. Ishasni Morbia
Literature search- Ms. Ishasni Morbia
Writing article- Ms. Ishasni Morbia
Critical review- Dr. Shivangi Mathur
Article editing- Dr. Richa Dubey
Final approval- Dr. Shivangi Mathur

REFERENCES

[1] Ritchie MD, Holzinger ER, Li R, Pendergrass SA, et al. Methods of integrating data to uncover genotype–phenotype interactions. Nat Rev Genet., 2015; 16(2): 85-97.
[2] Pevsner J. Pairwise sequence alignment. Bioinformatics and functional genomics, 2nd edition. Hoboken: John Wiley & Sons, 2009; pp. 47-97.
[3] Prosdocimi F. Introdução à bioinformática. Curso Online, 2010.
[4] Luscombe NM, Greenbaum D, Gerstein M. What is bioinformatics? An introduction and overview. Yearb Med Inform, 2001; 10(01): 83-100.
[5] Pevsner J. Bioinformatics and functional genomics. John Wiley & Sons, 2015.
[6] Allaby RG, Woodwark M. Phylogenetics in the bioinformatics culture of understanding. Int J Genomics, 2004; 5(2): 128-46.
[7] Chou KC. Progress in protein structural class prediction and its impact to bioinformatics and proteomics. Curr Protein Pept Sci., 2005; 6(5): 423-36.
[8] Sousa SA, Leitão JH, Martins RC, Sanches JM, Suri JS, et al. Bioinformatics applications in life sciences and technologies. BioMed Res int., 2016.
[9] Vinayagam A, Zirin J, Roesel C, Hu Y, Yilmazel B, Samsonova AA, et al. Integrating protein-protein interaction networks with phenotypes reveals signs of interactions. Nat Methods, 2014; 11(1): 94.
[10] Katara P. Role of bioinformatics and pharmacogenomics in drug discovery and development process. Network Modeling Analysis Health Informatics Bioinformatics, 2013; 2(4): 225-30.
[11] Murray-Rust P. Bioinformatics and drug discovery. COBIOT, 1994; 5(6): 648-53.
[12] Kannan S, Zacharias M. Role of tryptophan side chain dynamics on the Trp-cage mini-protein folding studied by molecular dynamics simulations. PloS One, 2014; 9(2): e88383.
[13] Wani M, Ganie NA, Rani, S, Mehraj S, Mi MR, et al. Advances and applications of bioinformatics in various fields of life. Int J Fauna Biol Stud., 2018; 5(2): 3-10.
[14] Lancashire LJ, Lemetre C, Ball GR. An introduction to artificial neural networks in bioinformatics-application to complex microarray and mass spectrometry datasets in cancer studies. Brief. bioinformatics, 2009;10(3): 315-29.
[15] Hack C, Kendall G. Bioinformatics: Current practice

ISSN (O): 2581-8740 | ISSN (P): 2581-8732
Morbia *et al*., 2023
DOI: 10.21276/SSR-IIJLS.2023.9.2.4

and future challenges for life science education. Biochem Mol Bio Educ., 2005; 33(2): 82-85.

[16]Tiwari A. Applications of Bioinformatics tools to combat the Antibiotic Resistance. In 2015 International Conference on Soft Computing Techniques and Implementations (ICSCTI), 2015; pp. 96-98.

[17]Zhang L, Hong H. Genomic discoveries and personalized medicine in neurological diseases. Pharm., 2015; 7(4): 542-53.

[18]Komatsu S, Hossain Z. Organ-specific proteome analysis for identification of abiotic stress response mechanism in crop. Front Plant Sci., 2013; 4: 71.

[19]Jacoby RP, Millar H, Taylor NL. Application of selected reaction monitoring mass spectrometry to field-grown crop plants to allow dissection of the molecular mechanisms of abiotic stress tolerance. Front Plant Sci., 2013; 4: 20.

[20]Subramaniam S, Fahy E, Gupta S, Sud M, Byrnes RW, Cotter D, et al. Bioinformatics and systems biology of the lipidome. Chem Rev., 2011; 111(10): 6452-90.

[21]Desiere F, German B, Watzke H, Pfeifer A, et al. Bioinformatics and data knowledge: the new frontiers for nutrition and foods. Trends Food Sci Tech, 2001;12(7): 215-29.

[22]Sadraeian M, Molaee Z. Bioinformatics Analyses of *Deinococcus radiodurans* in order to Waste clean-up. In 2009 Second Inter Conference Environ Computer Sci., 2009; pp. 254-58.

[23]Krane DE, Ford S, Gilder JR, Inman K, Jamieson A, et al. Sequential unmasking: a means of minimizing observer effects in forensic DNA interpretation. J Front Sci., 2008; 53(4): 1006-07.

[24]Bianchi L, Lio P. Forensic DNA and bioinformatics. Briefin Bioinformatics, 2007; 8(2): 117-28.

[25]Misra N, Panda PK, Parida BK. Agrigenomics for microalgal biofuel production: an overview of various bioinformatics resources and recent studies to link OMICS to bioenergy and bioeconomy. Omics: J Integ Boil., 2013; 17(11): 537-49. doi: 10.1089/omi.2013.0025.